A CONTEXTUAL BAYESIAN USER EXPERIENCE MODEL FOR SCHOLARLY RECOMMENDER SYSTEMS

ZOHREH DEHGHANI CHAMPIRI

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

A CONTEXTUAL BAYESIAN USER EXPERIENCE MODEL FOR SCHOLARLY RECOMMENDER SYSTEMS

ZOHREH DEHGHANI CHAMPIRI

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Zohreh Dehghani Champiri

Matric No: WHA130071

Name of Degree: Doctor of Philosophy

Title of Thesis: A Contextual Bayesian User Experience Model for Scholarly Recommender Systems

Field of Study: Human Computer Interaction - Computer Science

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

A CONTEXTUAL BAYESIAN USER EXPEREINCE MODEL FOR SCHOLARLY RECOMMENDER SYSTEMS

ABSTRACT

Scholarly recommender systems attempt to narrow down the number of research resources and predict availability of unknown resources to assist scholars with their scholarly tasks. Studies point out that the embedding of the recommending methods in the user experience dramatically affects the value to the users. Besides, researchers state that factors such as personal and situational characteristics, mostly considered as contextual data, affect the user experience of recommender systems. They started to improve classical recommending methods by modelling contextual data. It has been emphasised, contextual modelling plays a crucial role in recommendations because it can present the status of people, places, objects and devices in the environment. Hence, incorporating contextual data is an effective approach to enhance personalisation, which results in higher efficiency levels of user experience. However, it is not easy to decide which contextual data must be incorporated into scholarly recommender systems. The irrelevant contextual data might have a negative impact and lead to false reasoning models and irrelevant recommendations. Consequently, users lose their trust and stop using the system. Moreover, using too much contextual data leads to computational complexity and ambiguity in the system. Therefore, it requires formulating informed estimations about the influence of certain contexts before exploiting the naturalistic environments. This research aims to first investigate how contexts influence users' experience of scholarly recommenders and predict the relevant contexts, and then exploits the predicted contexts to develop a Bayesian user model which can be embedded in the recommending process. Additionally, a user interface for recommendation presentation is designed. Finally, the proposed user model and user interface are evaluated. The empirical results showed that there is strong relation between the user's contexts and paper quality and user interface

design adequacy as well as user interaction design adequacy. The empirical results have been exploited to develop a suitable Bayesian user model and user interface for scholarly recommender systems.

Keywords: Scholarly recommender systems, Research paper recommender systems, User eXperience, Bayesian Networks, Human-Computer Interaction

MODEL PENGALAMAN PENGGUNA KONTEKSTUAL BAYESIAN UNTUK SISTEM PENGESYORAN ILMIAH

ABSTRAK

Sistem pengesyoran ilmiah cuba mengecilkan jumlah sumber penyelidikan dan meramalkan ketersediaan sumber yang tidak diketahui untuk membantu para ilmuan dengan tugas ilmiah mereka. Kajian menunjukkan bahawa penyematan kaedah yang disyorkan dalam pengalaman pengguna secara dramatik menjejaskan nilai kepada pengguna. Selain itu, para penyelidik menyatakan bahawa faktor-faktor seperti ciri peribadi dan keadaan, yang kebanyakannya dianggap sebagai data kontekstual, mempengaruhi pengalaman pengguna sistem pengesyorkan. Mereka mula memperbaiki kaedah mengesyorkan klasik dengan memodelkan data kontekstual. Ia telah dipertekankan, pemodelan kontekstual memainkan peranan penting dalam pengesyoran kerana ia dapat mempersembahkan status orang, tempat, objek dan peranti di alam sekitar. Oleh itu, menggabungkan data kontekstual merupakan pendekatan yang berkesan untuk meningkatkan keperibadian, yang menjadikan tahap kecekapan yang tinggi dalam pengalaman pengguna. Walau bagaimanapun, tidak mudah untuk menentukan data konteks mana yang mesti dimasukkan ke dalam sistem pengesyoran ilmiah. Data kontekstual yang tidak relevan mungkin mempunyai kesan negatif dan membawa kepada model penaakulan palsu dan cadangan tidak relevan. Akibatnya, pengguna kehilangan kepercayaan mereka dan berhenti menggunakan sistem. Selain itu, menggunakan terlalu banyak data kontekstual membawa kepada kerumitan komputasi dan kekaburan dalam sistem. Oleh itu, ia memerlukan perumusan maklumat mengenai pengaruh konteks tertentu sebelum mengeksploitasi persekitaran semula jadi. Penyelidikan ini bertujuan untuk menyiasat terlebih dahulu bagaimana konteks mempengaruhi pengalaman pengguna terhadap pengkaji-pengkaji ilmiah dan meramalkan konteks yang relevan, dan kemudian mengeksploitasi konteks yang diramalkan untuk membangunkan model pengguna Bayesian yang dapat disematkan dalam proses mengesyorkan. Di samping itu, antara muka pengguna untuk persembahan cadangan dibuat. Akhirnya, model pengguna yang dicadangkan dan antara muka pengguna dinilai. Keputusan empirical menunjukkan bahawa terdapat hubungan yang kuat antara konteks pengguna dan kualiti kertas penyelidikan dan kecerdasan reka bentuk antara muka pengguna serta kecukupan reka bentuk interaksi pengguna. Keputusan empirikal telah dieksploitasi untuk membangunkan model pengguna Bayesian yang sesuai dan antara muka pengguna untuk sistem pengesyoran ilmiah.

Kata kunci: Sistem pengesyoran ilmiah, sistem pengesyoran kertas penyelidikan, Pengalaman Pengguna, Rangkaian Bayesian, Interaksi Insani-Komputer

DEDICATION

To my parents

ACKNOWLEDGEMENT

I particularly thank my supervisor Prof Siti Salwah Salim for expanding my intellectual research horizons in the field of Human-Computer Interaction (HCI). All her critical and fastidious comments on my work have taught me how to do scientific research successfully and stimulated me to widen my research prospects. I have been lucky to have a supervisor who cares so much about my work and responds to my questions so promptly and insightfully. I am also very thankful to my co-supervisor Prof. Loo Chu Kiong for his indispensable and valuable guidance, support and knowledge.

I am also especially grateful to Prof. Luanne Freund, Prof. Kellogg Booth, and Dr. Narges Mahyar from the University of British Columbia (UBC), Canada, for their generosity, time, support and knowledge. My work and research at UBC and HCI@UBC have been remarkable encounters, which have provided me finer insight into HCI research. It has been a pleasure collaborating with them and I hope many more years will follow.

My deepest gratitude to Dr. Seyed Reza Shahamiri (Manuka Institute of Technology, Auckland, New Zealand), who encouraged me to pursue my PhD studies at the University of Malaya (UM). His motivation is most likely the reason I began my PhD at UM. It has been wonderful working together. I also sincerely appreciate Dr. Shahram Khadivi (Research Scientist, eBay) and Dr. Mahmoud Danaee (UM) for their guidance and mentorship.

This research would not have been possible without the tremendous involvement of RSs, UX/UI experts and participants, especially in Malaysia and Canada. Hence, I express my profound gratitude for their time and feedback.

My thanks to the Norsys and SmartPLS companies for providing free licenses to facilitate parts of this research.

I would like to thank my special Malay friends Kamal, Chong, Norlida (Nat), Khairul Sabri (Kay) and Nor'ain, who have cared about me tremendously. Thank you Kamal for your support and encouragement that always came at the right time to help me. Thank you Chong for standing by my side day by day throughout my life in Malaysia and creating many unforgettable memories. I would also like to thank Dr. Shirin Fassihi, Dr. Danial Hooshyar, Dr. Khubaib Amjad Alam, Dr. Ghulam Mujtaba Shaikh, Dr. Vala Ali Rohani, Moein Fathi, and Mozhgan as well as my research mates at the HCI lab.

Finally and most importantly, my warmest regards to my beloved parents Eidimohammad and Shahnaz, my lovely sisters Zahra and Elham, and my dear brothers Reza and Amir for their unconditional love, heartfelt sympathy and motivation, particularly during the past months with little sleep and lengthy working hours. If I have learned anything while being away from my family, it is that you are my biggest source of motivation and inspiration. I am so blessed to have been born and grown up in such an amazing family, in a place that I cannot find any words to describe other than paradise. Thank God for having you!

I am also grateful to my dear sister and brothers-in-law Fariba, Hesam and Mehdi as well as my sweet nieces Raha and Yasamin for expanding my family and creating many happy and wonderful moments.

Baba and Maman, you have always been my life force. Thank you for believing and having confidence in me to be an independent thinker. Baba, thanks for teaching me to not be afraid of challenges, no matter how hard the challenges are. Maman, thanks for teaching me the importance of family, trust and love. I am highly grateful to you for all you have done for me. I dedicate this thesis to you and I love you both!

Research Grants

The following table shows the research grants which I received during my studies at the University of Malaya:

No	*Title*	*Institute*	*Grant Number*
1	University of Malaya Research Grant (UMRG)	University of Malaya, Malaysia	RP003B-13ICT
2	High Impact Research (HIR) Grant	Ministry of Higher Education, Malaysia	UM.C/625/1/HIR/MOE/FCSIT/05
3	Postgraduate Research Management and Monitoring (IPPP) Grant	University of Malaya, Malaysia	RP025B-15HNE
4	Postgraduate Research Management and Monitoring (IPPP) Grant	University of Malaya, Malaysia	RP061A-18SBS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Symbols and Abbreviations

CSM	:	Average of CS rates
RDM	:	Average RD rates
BN	:	Bayesian Network
BNUXM	:	Bayesian Network User Experience Model
$X \rightarrow Y$	:	Cause and effect relationship: variable X *effects variable* Y
CTR	:	Click-Through Rate
R^2	:	Coefficient of Determination
CF	:	Collaborative Filtering
CSV	:	Comma-Separated Values
CS	:	Consistency
CBF	:	Content-Based Filtering
CARSs	:	Context-Aware Recommender Systems
CBN	:	Contextual Bayesian Network
$\rho(X, Y)$	:	Correlation between two variables X, Y
DSRM	:	Design Science Research Methodology
DAG		Directed Acyclic Graph
f^2	:	Effect Size
EMs	:	Empirical Methods
$X \perp Y$	:	Event X is independent from event Y
$\forall x$	:	For all elements of x
FT	:	Friedman Test
GQM	:	Goal, Question, Metric
GS	:	Grow- Shrink algorithm
HCI	:	Human-Computer Interaction

~	:	is similar to (tilde)
KFCV		K Folds Cross Validation
KCC	:	Kendall's Coefficient of Concordance test
KBF	:	Knowledge Based Filtering
KW	:	Kruskal Wallis test
LINQ	:	Language Integrated Query
LR	:	Literature Review
ML	:	Machine Learning
MW	:	Mann- Whitney test
MMHC	:	Max-Min Hill-Climbing algorithm
MXE	:	Mean Cross Entropy
MSE	:	Mean Square Error
Obj	:	Objective
OL	:	Outer Load
OW	:	Outer Weight
PLS-SEM	:	Partial Least Square- Structural Equation Modeling
+→	:	Positive relationship
$P(X\|Y)$	:	Probability event *X* given event *Y*
$P(X)$	:	Probability of event *X*
$\prod_{i=1}^{n}, \prod_{i \in I} 3$	:	Product from *i*=1 to *n* or over all elements *i* in set I
RD	:	Recommendation Display
RSs	:	Recommender Systems
RQ	:	Research Question
SRSs	:	Scholarly Recommender Systems
r	:	Spearman correlation coefficient (Spearman's rho)

SRMR	:	Standardized Root Mean Square Residual
SRQ	:	Sub-Research Question
SLR	:	Systematic Literature Review
ANOVA	:	Two- way Analysis of Variance test
UX	:	User eXpereince
IxD	:	User Interaction Design
UID	:	User Interface
UM	:	User Model
VIF	:	Variance Inflation Factor
WOS	:	Web of Science
WSR	:	Wilcoxon Signed Ranks test

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

This chapter provides the background of the study, research problem and motivation which leads to the problem statement, research objectives and research questions. The scope, limitation of the research and the research methodology are also described.

1.1 Background of the study

Recommender Systems (RSs) have been an area of substantial research interest since the mid-1990s (Felfernig & Burke, 2008). In the last decade, RSs were investigated and implemented in various application domains, including social networks, e-commerce, e-learning, e-health, publications and e-resources (Verbert, Lindstaedt, & Gillet, 2010). With the increasing number of scientific publications, Scholarly Recommender Systems (SRSs), or commonly known as research paper or academic RSs, are considered an appropriate tool to facilitate and accelerate the process of information seeking for scholars by offering appropriate resources to users when they are going over a huge amount of relevant and irrelevant resources in scholarly repositories (Champiri, Shahamiri, & Salim, 2015).

SRSs normally collect data about users' activities and build user models to filter the preferences expressed either implicitly by inferring the needs from the user's item interactions (Sikka, Dhankhar, & Rana, 2012) or explicitly by a list of keywords. However, in the explicit way, an RS looks like a search engine and is not able to recommend indicators when the users do not know exactly what they need (Felfernig & Burke, 2008) (Verbert et al., 2010). In other words, the main difference between SRSs and search engines is the former's ability to predict the unknown indicators based on the limited data provided by the users (Sikka et al., 2012) (Baltrunas, Ludwig, Peer, & Ricci, 2012). Most scientists spend a lot of time on keyword-based search in order to find relevant research articles; however, their efforts yield unsatisfactory results (Mönnich &

Spiering, 2008a). A good and efficient SRS should be able to retrieve unknown papers for the scholars.

The scope of SRSs is broad, the operation of which may support scholars not only in recommending appropriate papers but also appropriate conferences, collaborators, etc. As a whole, SRSs can be helpful to researchers in multiple aspects when carrying various scholarly tasks (Champiri, Shahamiri, & Salim, 2015). However, the focus of this research is on SRSs that offer users relevant research or scholarly papers.

1.1.1 Recommending approaches

Regardless of the domain for which a recommender has been designed, so far, Classical approaches and Contextual approaches have been introduced to generate recommendations (Table 1.1). By and large, most traditional RSs adopt the classical approaches, which fall mainly into three main classes: Collaborative Filtering (CF), Content-Based Filtering (CBF), Knowledge-Based Filtering(KBF); and an additional class, known as Hybrid method, which is a combination of two or all of these three approaches (Adomavicius, Sankaranarayanan, Sen, & Tuzhilin, 2005; Pommeranz, Broekens, Wiggers, Brinkman, & Jonker, 2012).

Table 1.1: Recommending approaches (Adomavicius & Tuzhilin, 2011a)

Approach	*Classical approaches*	*Contextual approaches*
1	Collaborative Filtering (CF)	Contextual – Pre Filtering
2	Content – Based Filtering (CBF)	Contextual – Post Filtering
3	Knowledge – Based Filtering(KBF)	Contextual Modelling
4	Hybrid	

CF approaches are based on scholars' opinions and behaviours in order to find papers which are downloaded, used, read, cited, or rated, and scholars who have similar behaviours or opinions. CBF approaches are based on premises that utilise similarities between the papers and their features regardless of scholars' opinions in order to retrieve indicators with similar features. Another classical approach is Knowledge-Based Filtering (KBF), which provides recommendations based on specific knowledge or predefined (or learned) rules about users and items (R. D. Burke, Hammond, & Young, 1996) to deduce applicable links between user requirements and items that might be required to fulfil them (Resnick & Varian, 1997)(Will, Srinivasan, Im, & Wu, 2009b)(R. D. Burke et al., 1996). More details of classical approaches are discussed in Chapter 2 of this study. Generally, in classical approaches that make recommendations, RSs use a set of ratings that is either explicitly created by scholars or implicitly deduced by a system (Adomavicius & Tuzhilin, 2011a); hence, there are two types of entities, namely scholars and papers (two-dimensions) to estimate the rating function *R* (Liu, 2013):

$$R : Scholar \times Paper \to Rating \tag{1.1}$$

For each scholar s, the paper p' that maximises the scholar's utility is defined as (Adomavicius et al., 2005):

$$\forall s \in S, \mathrm{p}'_s = \arg\max R(s, p), p \in P \tag{1.2}$$

As Table 1.1 shows, the contextual approaches are classified into three approaches: pre-filtering, post-filtering, and contextual modelling (Adomavicius & Tuzhilin, 2011b; Kantor, Rokach, Ricci, & Shapira, 2011). Contextual information can be incorporated into the classical recommendation procedures in order to generate better recommendations (Baltrunas & Ricci, 2009). The preferences are estimated with the rating function of papers, users, and contexts as follows (Adomavicius et al., 2005):

$$R : \; Scholar \times Paper \times Context \to Rating \tag{1.3}$$

If the contextual information is defined with a set of contextual dimensions *D*, while two of these dimensions are *Scholar* and *Paper*, and the rest are contextual; the rating function *R* is (Adomavicius et al., 2005):

$$R = D_1 \times D_2 \times D_3 \times \ldots \times D_n \rightarrow Ratings \qquad (1.4)$$

The utility function is defined by selecting certain "what" dimensions $D_{i1},\ldots, D_{ik}\,(k<n)$ and certain "for whom" dimensions $D_{j1},\ldots,D_{jl}\,(l<n)$ that do not overlap, i.e. $\{D_{i1},\ldots, D_{ik}\}\cap\{D_{j1},\ldots,D_{jl}\}=\varnothing$, and recommending for each tuple $(d_{j1},\ldots, d_{jl})\in D_{j1}\times\ldots\times D_{jl}$ the tuple $(d_{i1},\ldots, d_{ik})\in D_{i1}\times\ldots\times D_{ik}$ that maximises rating $R(d_1,\ldots, d_n)$ (Adomavicius & Tuzhilin, 2011a). In particular:

$$\forall (d_{j1},\ldots, d_{jl})\in D_{j1}\times\ldots\times D_{jl},\ (d_{i1},\ldots, d_{ik}) = \underset{\substack{(d'_{i1},\ldots,d'_{ik})\in D_{i1}\times\ldots\times D_{ik} \\ (d'_{j1},\ldots,d'_{jl})=(d_{j1},\ldots,d_{jl})}}{\operatorname{argmax}} R(d'_1,\ldots,d'_n) \qquad (1.5)$$

For example, in recommending papers to scholars, if there are Paper (title, keywords, author, subject, publisher, year) and Scholar (name, age, degree, interests), and the contexts can be defined as *Location*, where the scholar is looking for a paper; then *L={university, home and Time*, when the scholar is looking for a paper; *T={first semester, second semester}*. Hence, the function will become (Adomavicius & Tuzhilin, 2011a):

$$R = Scholar \times Paper \times Location \times Time \qquad (1.6)$$

The term context appeared in the field of computer science in the late 1980s (Hong, Suh, & Kim, 2009), and the idea of context awareness in computing was introduced by Schilit in 1994 (Brown, Bovey, & Chen, 1997) in order to increase the richness of communication and provide more useful computational services (Dey, 2001). Since then, many studies in the field of computer science tried to define the term "context". Dey

(2001) offered the most cited definition of context from a computer science viewpoint. He expressed that context is any information can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves (Dey, 2001).

As depicted in Figure 1.1, the contextual approaches are categorised into three approaches: pre-filtering, post-filtering, and contextual modelling implicitly (Adomavicius & Jannach, 2013; Yujie & Licai, 2010) (Kobsa, 2001).

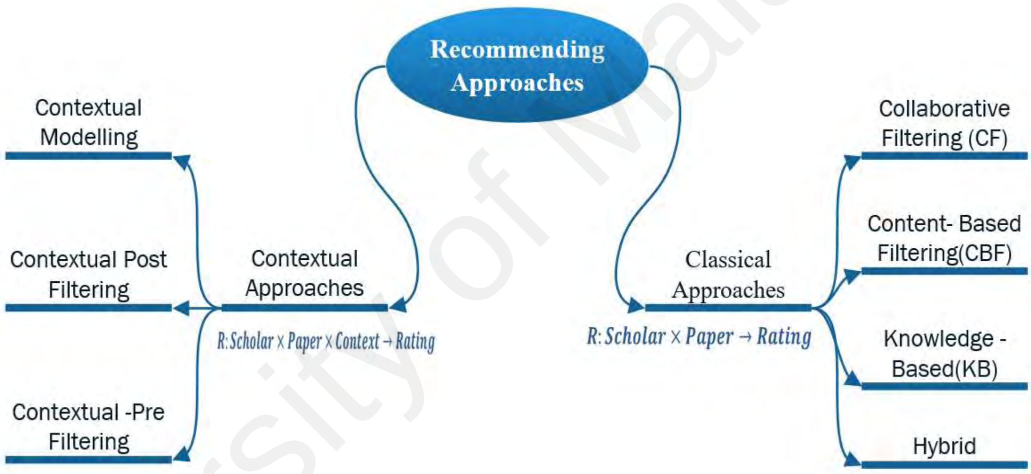


Figure 1.1: Recommending Classification

Contextual information can be incorporated into the classical recommendation procedures in order to generate better recommendations (Adomavicius & Tuzhilin, 2011b; Baltrunas, 2008). Particularly, in contextual pre-filtering, the contextual information is used before all the ranked recommendations are computed (Panniello, Gorgoglione, & Tuzhilin, 2015). The reduction-based approach is an example of pre-filtering in a way that first, all ranked recommendations are computed through classical methods like CBF; then they are adjusted or reranked for each user using contextual information (Adomavicius & Tuzhilin, 2005b). Conversely, in contextual post-filtering, after computation of all the ranked recommendations, the contextual information is used (Adomavicius & Tuzhilin, 2005b).

As discussed earlier, a recommender overpowers search engines by modelling users' preferences. In other words, regardless of the technology exploited by RSs, the high quality recommendations can be provided to users only after their preferences have been modelled, which is typically called User Model (UM) in the literature (Berkovsky et al., 2008) (Adomavicius & Jannach, 2013; Yujie & Licai, 2010) (Kobsa, 2001). In this context, quality refers to the ability of the system to produce exactly those recommendations that the user will use or would like to. To achieve this, adequate information, including contextual information, should be stored to deliver high quality recommendations; however, acquisition of sufficient data for the UM is not an easy task, especially in the initial stages of interaction with the user, when usually little information about the user is available (Berkovsky et al., 2008). In contextual modelling, the contextual information is employed directly as a main part of learning preference models (built using techniques such as decision tree, regression, and probabilistic model). To put it differently, contextual variables are added as dimensions in the recommendation function in addition to the user and item dimensions (Hariri, Mobasher, & Burke, 2014).

1.2 User eXpereince and Contextual User Modeling

In the early 1990s, cognitive scientist, Don Norman (2013) coined the term User eXperience (UX). He indicated that UX is about the user's feelings (positive and negative) about a product over time (D. Norman, 2013b). UX can be used as an umbrella term in the field of Human Computer Interaction (HCI) to focus on aspects which are beyond usability (Hassenzahl & Tractinsky, 2006), including all the feelings (positive and negative) a user is experiencing while interacting with a product, e.g. a mobile phone. There is no clear definition of UX (McCarthy & Wright, 2004) (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009); however, the current ISO (ISO 9241-110:2010 (clause 2.15) definition of UX focuses on a person's perception and the responses resulting from the use or anticipated use of a product, system, or service. In software engineering, as a matter of fact, if a product fails to meet end users' rising needs, it makes both the product and the company (creator of product) obsolete (Kraft, 2012). In other words, users will choose products with a great UX (Knijnenburg & Willemsen, 2010). Hence, UX is becoming the key competitive factor in more and more industries. Recently, researchers have acknowledged that embedding the RSs and user modelling into UX impacts dramatically on the effectiveness of recommendations for the users (Bart P. Knijnenburg et al., 2012; Joseph A Konstan & John Riedl, 2012). The UX is also affected by the users' situations, behaviours, characteristics or in a nutshell, users' contexts (Kamis & Davern, 2004; Knijnenburg & Willemsen, 2009; Knijnenburg & Willemsen, 2010). In this regard, contextual information influencing UX plays an important role in creating appropriate recommendation because it can present the status of people, places, objects and devices in the environment (Adomavicius & Tuzhilin, 2011a; Baltrunas, 2008; Baltrunas et al., 2012; J. Yuan, Sivrikaya, Marx, & Hopfgartner, 2014; Yujie & Licai, 2010) and leads to a better experience for the users, and consequently better interaction between the users and system.

1.3 Problem Statement

Since the advent of SRSs, more than 200 papers have been published (Beel, Breitinger, Langer, Lommatzsch, & Gipp, 2016), which mostly aim to create more accurate algorithms. There is a presumption that, the more accurate the algorithm, the better the predicted recommendation is for the users. Recently, the embedment of recommending methods into UX has been taken into consideration, which greatly influences the value of RSs to the users (Joseph A Konstan & John Riedl, 2012; Sean M McNee, Riedl, & Konstan, 2006a; T. Nguyen, 2016). Researchers have stated that UX is affected by the limits of human perception and the preconceptions of the individual and factors such as personal characteristics and situational characteristics, which are mostly considered as contextual information (Kamis & Davern, 2004; Knijnenburg & Willemsen, 2009; Knijnenburg & Willemsen, 2010).

Although there are a few studies on developing UMs in the field of SRSs, it has been emphasised that incorporating contextual information into user modelling and creating the recommendations based on the users' information needs, is an effective approach to enhance personalisation and consequently UX with SRSs (Beel, Breitinger, Langer, Lommatzsch, & Gipp, 2016). However, it is not clear how contexts influence the UX of SRSs and moreover it is difficult to decide which contexts must be incorporated into developing the user model (UM) (Adomavicius & Tuzhilin, 2011b; Baltrunas, 2008) because of the following reasons.

First, identification of valid contextual information for different domains is a challenge in contextual user modelling either explicitly or implicitly (Adomavicius & Tuzhilin, 2011a; Asabere, 2013). Besides, the irrelevant contextual information might have a negative impact on UX and lead to false reasoning models and irrelevant recommendations. Consequently, users lose their trust and stop using the system (Adomavicius & Tuzhilin, 2011a; Baltrunas, 2008; Baltrunas et al., 2012; J. Yuan, Sivrikaya, Marx, & Hopfgartner, 2014; Yujie & Licai, 2010). For example, the resources recommended to an undergraduate student searching for "Fuzzy method" for his class assignment may be different from those recommended to a graduate student writing a research paper on the same topic. This is due to the different requirements of the tasks they are working on and the different levels of formal education, which are considered as contextual information.

Second, exploitation of too much contextual information causes complexity and ambiguity in the system due to irrelevant, redundant, inconsistent, and noisy data

(Adomavicius & Tuzhilin, 2011a; Baltrunas, 2008; Baltrunas et al., 2012; J. Yuan, Sivrikaya, Marx, & Hopfgartner, 2014; Yujie & Licai, 2010). In fact, each piece of particular contextual information is considered an extra dimension to the utility function of the recommender. Too many dimensions might raise the problem of dimensionality (Guyon & Elisseeff, 2003) (Sarwar, Karypis, Konstan, & Riedl, 2000). Apart from the problem, it needs unbounded computational resources to discover useful knowledge patterns and imposes extra costs to the system development, especially to the data acquiring and processing.

Third, users' experiences are influenced by the users' values and expectations, which have to be considered in the design and development of RSs from the beginning. For instance, what contextual information touches a scholar's experience when an academic paper is recommended to him? Or even before and after his interaction with the system? (Sean M McNee, Riedl, et al., 2006a). It might be related to users' interests, background knowledge or novelty of the recommended paper. Depending on the situations, the experience of interacting with SRSs might be weak or convincing for the scholars. Therefore, it is significant to explore what and how contexts influence users' experiences and make evaluations over time (Champiri et al., 2015; Bart P. Knijnenburg, Willemsen, Gantner, Soncu, & Newell, 2012).

Fourth, another main concern regarding the UX of SRS is how an SRS should present the recommendations to the users. Actually, only a few studies have put efforts to develop User Interfaces (UIs) for SRSs; particularly among more than 200 studies in the field of SRSs, the number of studies which have designed UIs is less than five (Beel, Breitinger, Langer, Lommatzsch, & Gipp, 2016), and most of the researchers tried to improve the recommendations algorithms (Pu, Chen, & Hu, 2012a), while both industry practitioners and academic researchers have argued that the interface of a RS may have far larger effects on users' experience with the recommender than the recommender's algorithmic performance (McNee et al., 2006; Baudisch & Terveen, 1999; Murray & Haubl, 2008; Xiao & Benbasat, 2007; Ziegler et al., 2005; Ozok et al., 2010) (T. Nguyen, 2016). According to the RecSys09 keynote presented by Francisco Martin, up to 50% of the value of recommenders comes from a well-designed interface (Ge, Delgado-Battenfeld, & Jannach, 2010a). UI is important because it is the way users interact with the system. No matter how accurate the algorithms work, if the UI is not well designed and evaluated, it will degrade the interaction between the user and system.

Last but not least, a contextual UM should be able to infer possible cognitive process of UX, including behaviour and mind, but it is not fully tractable in practice (W. Wu, He, & Yang, 2012) since it needs deep understanding of UX in a particular domain, which is mostly uncertain. In the past decade, different Machine Learning (ML) methods had been applied to support user modelling in RSs; however, the cognitive UMs, which are based on the deep understanding of uncertain and dynamic users' contexts, is still in an early stage (Martín, Haya, & Carro, 2013; Papatheocharous, Belk, Germanakos, & Samaras, 2014) particularly in the domain of SRSs. Additionally, among the ML methods, Bayesian networks (BNs) are powerful tools used for uncertainty modelling (Pearl, 1985). However, they have rarely been applied in the field of SRSs (Hassan, 2017) (Beel, Breitinger, Langer, Lommatzsch, & Gipp, 2016).

As mentioned above, recently researchers of RSs have revealed that there are a number of issues and gaps concerning the definition and detection of contextual information influencing UX with RSs. So far, SRS studies have given little attention about the relevant contexts which influence UX with SRSs and also how these contexts should be incorporated into the UX recommending UMs (Champiri et al., 2015). The urge to conduct this research is due to addressing the following research issues, which are also the source of motivation for conducting this research.

1.3.1 User eXperience with SRSs

In SRSs, as a matter of fact, if the system fails to meet end users' rising information needs, both the system and the company (creator of product) would be considered obsolete because the users would stop working with the system. UX is becoming the key competitive factor in more and more industries (Bernhaupt, 2010). Users demand products that are not only easy to use but also joyful and fun to use. In other words, they will choose products with a great UX. There are few studies investigating the role of UX in the field of RSs and especially SRSs, and research on UX is quite new (Knijnenburg & Willemsen, 2010).

1.3.2 Exploration of how contextual information influences UX of SRSs

One of the existing issues is that there is no study in the field of SRSs to explain how contexts influence UX of SRSs (Champiri et al., 2015). To address this issue, the existing models and theories of UX, especially with RSs are reviewed, and then a conceptual framework is proposed. The framework contains in what way contexts (latent variables) influence UX of SRSs. After theoretically justifying the relationships drawn between the various latent variables leading to the proposition of relevant hypotheses, the statistical experiment of the framework is performed using the quantitative method of Partial Least Squares (PLS) Regression and Structural Equation Modelling (SEM). The proposed framework not only enriches the conceptual understanding of how contextual information influences UX of RSs but also serves as a foundation for further theoretical and empirical investigations.

1.3.3 Detection of relevant contexts influencing UX of SRSs

Due to the issues of the impact of irrelevant contexts mentioned in Section 1.2, it is necessary to estimate and analyse the impact of contextual information on UX before actually collecting and exploiting it in the recommending process (Rubens, Kaplan, & Sugiyama, 2011). Different empirical tests have been applied to assess the relevant contexts based on the guidelines of the empirical method of PLS-SEM. This consequently leads to dimension reduction and might prevent the occurrence of dimensionality problem (Sarwar, Karypis, Konstan, & Riedl, 2000). It is pertinent to mention that, the detection of contextual information does not completely solve the problem of providing a pleasing and convincing experience in using the SRSs; in fact, it is half of the way and helps to clarify the problem in order to figure out what users may experience when interacting with SRSs. The next step is how to exploit and adapt the detected contextual information in recommending UM so that the system can generate better recommendations, which is the next motivation of this research.

1.3.4 Applying BN modelling

Based on the SLR on recommending methods applying to SRSs by (Champiri et al., 2015) and review of recent studies by (Hassan, 2017) (Beel, Gipp, et al., 2016), it is concluded that, different ML methods such as Neural Networks, SVM, and Decision Trees, have been utilised in making paper recommendations by considering the CF and CBF approaches. Despite the fact that UMs play a critical role in maintaining the recommendation quality and identification of the users' needs, they are rarely used in SRSs researches (Beel, Gipp, et al., 2016). Quality of recommendations refers to the capability of the system to predict exactly those items or services the user would like or use, and to provide overall good experience for the users (Berkovsky et al., 2008; Kobsa, 2001). However, information needs are uncertain and vary among users due to different contexts, such as background knowledge, preferences and goals (Beel, Gipp, et al., 2016). Hence, it is necessary to select a method which infers dynamic context and surpasses uncertainty. Among the ML methods, BNs are powerful tools used for uncertainty modelling (Pearl, 1985) based on the probabilistic theory of Bayes' theorem, which spreads knowledge within the network (Heckerman et al., 1995; Neapolitan, 2004) and reasons complicated problems. However, contextual BNs are rarely applied for recommending scientific articles (Hassan, 2017). According to the survey conducted by (Portugal, et al., 2015), among seven RSs applied BN method, only two studies are related to book and document recommender (Ericson & Pallickara, 2013) (Lucas, Segrera, & Moreno, 2012), which are not contextual UM for paper recommending. In this research, a contextual Bayesian UX model is developed. The recommending methods used for SRSs are discussed in Chapter 2, and the reasons in selecting the BN method for modelling the users' need are provided in detail in Chapter 5.

1.3.5 rScholar: UI design

Another main concern regarding the UX of SRS is how SRS should present the recommendations to the users. Actually, only a few studies have put in efforts to develop interactive User Interfaces (UIs), and most of the researchers tried to improve the algorithms of recommendations (Pu et al., 2012a) (Beel, Gipp, et al., 2016). UI is important because it is the way users interact with the system. In other words, no matter how accurate the algorithms work, if the UI is not well designed and evaluated, it will degrade the interaction between the user and system. This research attempts to design an appropriate UI for SRSs to provide better UX.

1.4 Research Objectives

From the research issues, it is known that contextual information, which is a considerable factor in UX, has been under-utilised in the researches of SRS. This study aims to fill this research gap and deepen the understanding of UX in three ways: grasp the knowledge from the perspective of researchers who are end users of SRSs; assess the most contextual information from their point of view; and develop a UM and UI based on the identified contextual information.

The main aim of this research is to develop a Bayesian contextual UX model for SRSs. In the following section, the specific objectives for achieving the main goal of this research have been listed.

Objective 1: To propose a framework to show how contextual information influences UX with SRSs

Objective 2: To develop a contextual Bayesian UX model using the assessed relevant contexts in Objective 1

Objective 3: To design a UI using the assessed contexts related to UI and User Interaction design adequacy in Objective 1 and the inputs and outputs for Objective 2

Objective 4: To evaluate the proposed Bayesian UX model and UI

1.5 Research Questions

The research questions (RQ) posed to address the research objectives are described below. Each question along with its sub-questions are answered in one of the chapters as follows:

Objective 1: To propose a framework to show how contextual information influences UX with SRSs

Title: A framework for contextual information influencing UX of SRS

RQ1. How does contextual information conceptually influence UX with SRSs?

SRQ1.1: What models/frameworks/theories have been proposed for UX of RS/SRS in the existing studies?

SRQ1.2: What components and relationships can be applied to the framework?

SRQ1.3: What indicators can be applied to the components?

SRQ1.4: What contexts can be applied?

SRQ1.5: What is the experts' review feedback on the proposed conceptual framework?

RQ2. How does contextual information empirically influence UX with SRSs?

SRQ2.1: How can an appropriate dataset be prepared for the empirical examination?

SRQ2.2: Are indicators empirically valid?

SRQ2.3: Are constructs (components) empirically valid?

SRQ2.4: Are relationships between the constructs (components) empirically valid?

SRQ 2.5: What is the GOF of the framework?

SRQ2.6: What are the most relevant contexts influencing/contributing to UX of SRSs?

Objective 2: To develop a Bayesian UX model using the assessed relevant contexts in Objective 1

Title: A contextual Bayesian UX model

RQ1: How can a contextual Bayesian UX model be developed using the assessed relevant contexts in Objective 1?

SRQ1: How should a contextual dataset be acquired for the Bayesian model train and analysis?

SRQ2: What contexts should be incorporated into the Bayesian UX model?

SRQ3: How should the structure of BN model be built up?

SRQ4: How should the parameters of BN model be learned from the dataset?

Objective 3: To develop a UI using the assessed contexts related to UI and User Interaction design adequacy in Objective 1

Title: rScholar: the UI design and development

RQ1: How is a UI designed based on the assessed contexts related to UiD and IxD adequacy in objective 1 and the data inputs and outputs required for objective 2?

SRQ1: What are the users' requirements for the UI development?

SRQ2: What are the design-solutions based on the identified users' requirements?

SRQ3: What is the appropriate architecture for the UI development?

SRQ4. How should the IxD adequacy features (Interaction design) be designed?

SRQ5. How should the UiD adequacy features (Visual Design) be designed?

Objective 4: To evaluate the proposed Bayesian UX model and UI

Title: Evaluation

RQ1: How should the Bayesian UX model be evaluated?

SRQ1.1: Which evaluation method and metrics are appropriate to evaluate the Bayesian UX model?

SRQ1.2: What are the results of selected measures applied to evaluation of the BN Structure model?

SRQ1.3: What are the results of selected measures applied to performance of BN algorithm?

SRQ1.4: What are the results of selected measures applied to evaluation of the BN performance model?

RQ2: How should the UI be evaluated?

SRQ2.1: Which evaluation method and metrics are appropriate to evaluate the proposed UI?

SRQ2.2: What are the differences between rScholar (UiD and IxD) and Google Scholar (experts' feedback)?

SRQ2.3: What is the overall experts' evaluation on rScholar & Googlescholar?

SRQ2.4: What is the experts' evaluation of design ideas exploited in rScholar?

SRQ2.5: Is there any change in users' ratings of rScholar after three months?

SRQ2.6: Is there any differences in users' groups in evaluation of rScholar?

1.6 Research Contribution

In this section, the research contributions in different types are discussed.

1.6.1 Meta-Analysis & SLR on CASRSs

In this thesis, in order to answer a few research questions, two SLRs and meta-analysis reviews have been conducted. Table 1.2 summarises the questions that have been answered through this reviews. The depth surveys exhibit an impression about the important aspects of the SRS researches, and are the source of inspiration for formulating the research question of this thesis. The details of the above-mentioned surveys are discussed in Chapter 2 of this thesis.

Table 1.2: Exploration of SRSs by SLR and Meta-analysis

Questions that have been explored	***Publication Reference (2000-2014)***
1. What recommending methods have been used? 2. How have researchers assessed the most influencing contextual information? 3. What contextual information has been used in SRSs through a survey on related work? 4. Challenges and open issues 5. What evaluation methods have been applied? 6. What evaluation metrics have been applied?	(Champiri et al., 2015) Chapter 2

1.6.2 An empirical research to identify the contexts influencing UX

By achieving Objective 1 of this study, a conceptual framework of the UX research in RSs is provided, which gives a deeper understanding of how environment, user and system contexts influence the UX. It thereby provides a better understanding of what and how certain relevant contexts results in a better UX, which consequently helps further user-centric research and development of SRS, and serves as the backbone for understanding context before using it. Besides, the proposed conceptual framework in this thesis is examined empirically and thus the results are testable and reproducible. The results are elaborated in Chapter 4 of this thesis.

1.6.3 Decipher of context and UX concepts in SRSs research

The concepts of both context and UX are subjective, which can cause difficulties for researchers who are working with them (Champiri et al., 2015; Bart P. Knijnenburg et al., 2012). The lack of understanding of the subjective concepts not only leads to more complexity in the system but also causes the system failure from the end user's perspective. In the field of SRS, there is no comprehensive study and through the above-mentioned empirical research, the distilling of the concepts of context and UX might be the inspiration for the future studies. More details are provided in Chapter 4 of this thesis.

1.6.4 Contextual UX model

As indicated earlier, the most applied recommending approaches are CF and CBF, and there are few studies developing UMs although the main factor that influences users' satisfaction is the ability of a recommender to meet the users' information needs; and it is obvious that the users have different information needs due to different knowledge, goals, and generally different contexts which are uncertain and change consistently (Beel, Breitinger, et al., 2016). A user model tries to process the users' data in order to predict the users' preferences. In this research, the users' information needs for four levels of accurate, novel, popular and diverse papers are predicted. More details of the contextual UX model using the BNs method are discussed in Chapter 5 of this thesis.

1.6.5 Bayesian UM

BNs are better suited not only to reason with the knowledge and uncertain belief, but also the structure of knowledge representation to deal with uncertainty of context inference (Long et al., 2010) (French, 1986; Peterson, 2009). Overall, BNs are flexible models which use probability distribution to provide the predictions about a number of influential variables rather than a single variable (Zukerman & Albrecht, 2001) (Guo, 2011; Korb & Nicholson, 2003). Despite the benefits of BNs, as mentioned earlier, they have been rarely applied to model users' information needs and users' contexts in SRSs. In this research, a contextual user experience model is developed, which is discussed in Chapter 5.

1.6.6 Contextual dataset

One of the main contributions of this research is providing a new dataset. After assessment of the most influencing contextual information, the second objective of this research is to develop a Bayesian UX model based on the identified contextual data from the first objective. There is no available dataset for the SRSs containing the detected contextual variables (data) to train and test the proposed UM; hence, a web application is designed in order to acquire the data from the real users (scholars). It is worth noting that in this dataset, all indicators (papers) have been classified or tagged in four categories of novel, diverse, popular and accurate according to the relevant user's ratings. The detailed information of the dataset preparation is expatiated in Chapter 5. This dataset contains new and useful corpus and has been public-shared by this thesis's researcher for the benefit of the research community along with a few benchmark tests, if anyone wants to do some comparisons for evaluations of shared repositories or new algorithms.

1.6.7 UiD and Interaction design adequacy in SRSs

There are only a few studies which have taken into account UI and interaction adequacy in SRSs. The reasons behind this matter are discussed in detail in Chapter 2. However, it is accentuated that UI and interaction design adequacy are extremely important factors for enhancing UX of RSs (Pu & Chen, 010; Pu et al., 2012) (Bart P. Knijnenburg et al., 2012). In other words, no matter how accurate the algorithms work, if the UI and interaction design are poorly designed and evaluated, it will degrade the interaction between the user and system in a way that users might find the system intrusive, annoying or distracting, and perceive it as a factor that negatively affects their experience (Ozok et al., 2010). Additionally, without considering the above-mentioned matter, the goal of UX enhancement would be abortive since users' satisfaction is influenced by different users' feelings appertaining to UI and Ix options and elements. This research is the first attempt that investigates the impact of UI and interaction design adequacy on the UX in the field of SRSs. The proposed UI called rScholar is developed and discussed in Chapter 6 of this thesis.

1.6.8 Opening new research areas (Opinion Contributions)

There are a few open areas that may provide a step forward to extend this research. The future researchers are encouraged to investigate the contextual information that has been ignored in this research, such as reasoning methods, mood, and location for the SRSs and other relevant recommenders. Besides, making UMs more implicitly is also another open area for the interested researchers. The open research areas are discussed in detail in Chapter 8.

1.7 Research Methodology

This research mainly adopts the Design Science Research Methodology (DSRM) by (Peffers et al., 2007) and (Hevner, 2007) in order to achieve the objectives with predominantly Empirical methods (EMs) to perform each activity of DSRM. The EMs refer to the specific techniques, tools, and means by which data are collected and analysed. The research process consists of three main phases: problem & solution identification, design & development, and evaluation. In each phase, the EMs are accomplished according to the guidelines by (Wohlin et al., 2012) whenever it is required. Figure 1.2 illustrates the main phases and the activities applied in this research. The methodology is elaborated in Chapter 3.

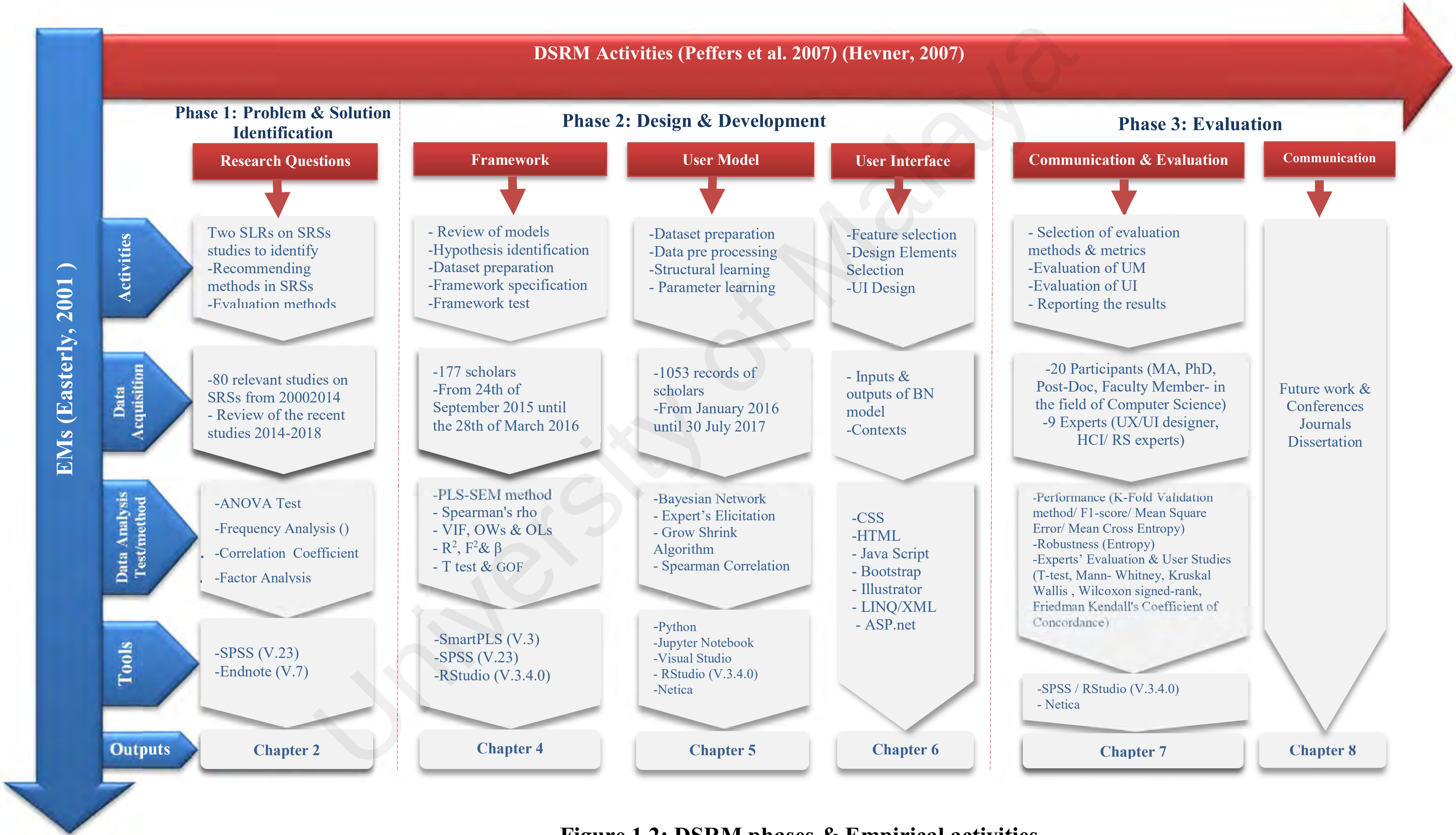


Figure 1.2: DSRM phases & Empirical activities

1.8 Scope and limitation

This research is an intersection of three main areas which include various sub-areas. As Figure 1.3 depicts, context-aware computing, users' information needs in scholarly domain, user modelling, UX, and UI are under the HCI area. RSs researches are an interdisciplinary research area, but they are mainly under the information filtering systems, and finally because the BNs method is applied to develop the UM, this research also involves partially Artificial Intelligent (AI) and ML research areas. However, the main goal of this research is to develop a UM exploiting context which influences UX of SRSs. Therefore, this research attempts to step forward in RSs research from a HCI perspective.

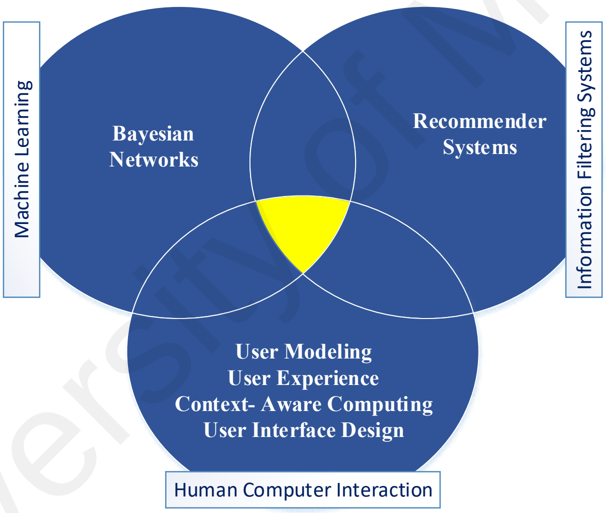


Figure 1.3: Scope of the research

It is important to emphasise that the scope of SRSs is broad, which may support scholars not only in recommending appropriate papers but also appropriate conferences, collaborators, etc. As a whole, SRSs can be helpful for researchers in multiple aspects of scholarly tasks. This research will focus on SRSs which offer research papers to the scholars, and scholars in this research are master's, PhD students, post-doc researchers and faculty members who are actively involved in doing research and looking for appropriate papers.

1.9 Structure of thesis

There are eight chapters in this thesis as explained below.

Chapter 1- Introduction: It introduces the research topic and provides an overview of the dissertation by briefly discussing the research problem, research objectives and questions, research methodology, research contribution, limitations, and methodology. It also presents the structure of the thesis.

Chapter 2- Literature Review: It gives an introduction to SRSs, recommending approaches, including classical and contextual approaches, UX with SRSs, evaluation methods and metrics, reviews of the existing related work on contextual recommending and evaluation methods to provide a comprehensive and critical overview of available SRSs by conducting two SLRs to further the knowledge on SRSs and to address a few research questions regarding this research.

Chapter 3- Research Methodology: It provides an introduction to the research methodology carried out to fulfil the objectives of this research. This research mainly adopts the DSRM, which is conducted in three phases of problem & solution identification, design & development and evaluation that predominantly uses EMs to perform each activity in the mentioned phases of DSRM.

Chapter 4- A framework to show how contexts influence UX with SRS: It explores how contexts influence UX with SRSs, and assesses what the most relevant contexts are. This chapter reviews existing models and theories of UX, especially with RSs and then proposes a conceptual framework, and examines it empirically by using quantitative method of Partial Least Squares Regression and Structural Equation modelling technique. Also, this chapter identifies the most relevant contextual information influencing or contributing to UX of SRSs.

Chapter 5- A contextual Bayesian UM exploiting the relevant identified contexts: It develops a UM applying BN method based on the most relevant contexts identified from

Objective 1 in Chapter 4. The BN model is built up in three main processes: dataset preparation and data pre-processing; BN model structure learning and; and BN model parameters learning. This chapter also discusses why BN method has been applied among the other ML methods.

Chapter 6- rScholar: A UI design considering the identified contexts: It designs a UI called rScholar mostly based on the empirical results of the most influencing contexts related to the UI and interaction design adequacy identified in Objective 1; and the data which are required for the proposed BN model developed in Objective 2. This chapter also explains how rScholar supports the required data to be exploited in the BN model, and follows the existing UI and interaction design adequacy guidelines which can be potentially applicable for enhancement of SRSs.

Chapter 7- Evaluation: It presents the evaluations performed to validate the proposed BN model and UI in Chapters 5 and 6 by applying two evaluation methods of offline and user-studies. In the offline method, the performance and robustness of the BN model have been examined. To evaluate the UI, several tests, including T-test, Mann-Whitney (MW), Kruskal Wallis (KW), Wilcoxon signed-rank, and Friedman Kendall's Coefficient of Concordance are performed. This chapter also explains the reasons for the selection of the above-mentioned evaluation methods and metrics.

Chapter 8-Conclusion & Future work: It gives a review of the research objectives and conclusion of the work. It also discusses the future work based on the results and limitation of this dissertation.

CHAPTER 2: LITERATURE REVIEW

This chapter starts with the brief overview of Scholarly Recommender Systems (SRSs), recommending approaches, context, User eXperience (UX), evaluation methods and metrics and continues by reviewing existing related work on contextual recommending and evaluation methods to provide a comprehensive and critical overview of available SRSs by conducting two Systematic Literature Reviews (SLRs). Also, this chapter aims to explain and highlight the way researchers understood and assessed relevant contextual information incorporating into the recommending process in order to provide better recommendations and enhance UX of SRSs. This review chapter ends with a discussion and critical analysis of the open issues of the existing SRSs and the research work in this thesis that aims at addressing those issues.

2.1 Scholarly Recommender Systems

In 1997, the term Recommender Systems (RSs) was posed in an article by Paul Vesnick and Hal R. Varian (Resnick & Varian, 1997). They described the RSs as a tool applying for decision making and not just for information retrieval (Ali, 2014). Apart from the information retrieval(Salton, 1989), the birth of RSs is derived from various domains such as of cognitive science(Rich, 1979), marketing(Lilien, Kotler, & Moorthy, 1992), management (Murthi & Sarkar, 2003) as well as forecasting(Armstrong, 2001) and approximation (Powell, 1981) theories (Adomavicius & Tuzhilin, 2005a).

Scholarly Recommender Systems (SRSs) sometimes called research paper or academic RSs, aim to manage information overload by filtering and personalising data according to users' needs (Champiri et al., 2015). In 1998, Giles et al. (Giles, Bollacker, & Lawrence, 1998) introduced SRSs in CiteSeer project. Since then, the dramatic data increase has necessitated the use of SRSs as an appropriate tool for facilitating and accelerating the process of information seeking for scholars (C. Porcel, Herrera-Viedma, Enrique, 2010) (Mönnich & Spiering, 2008a).

Figure 2.1 shows the structure of this chapter and how the research questions have been reached in this research.

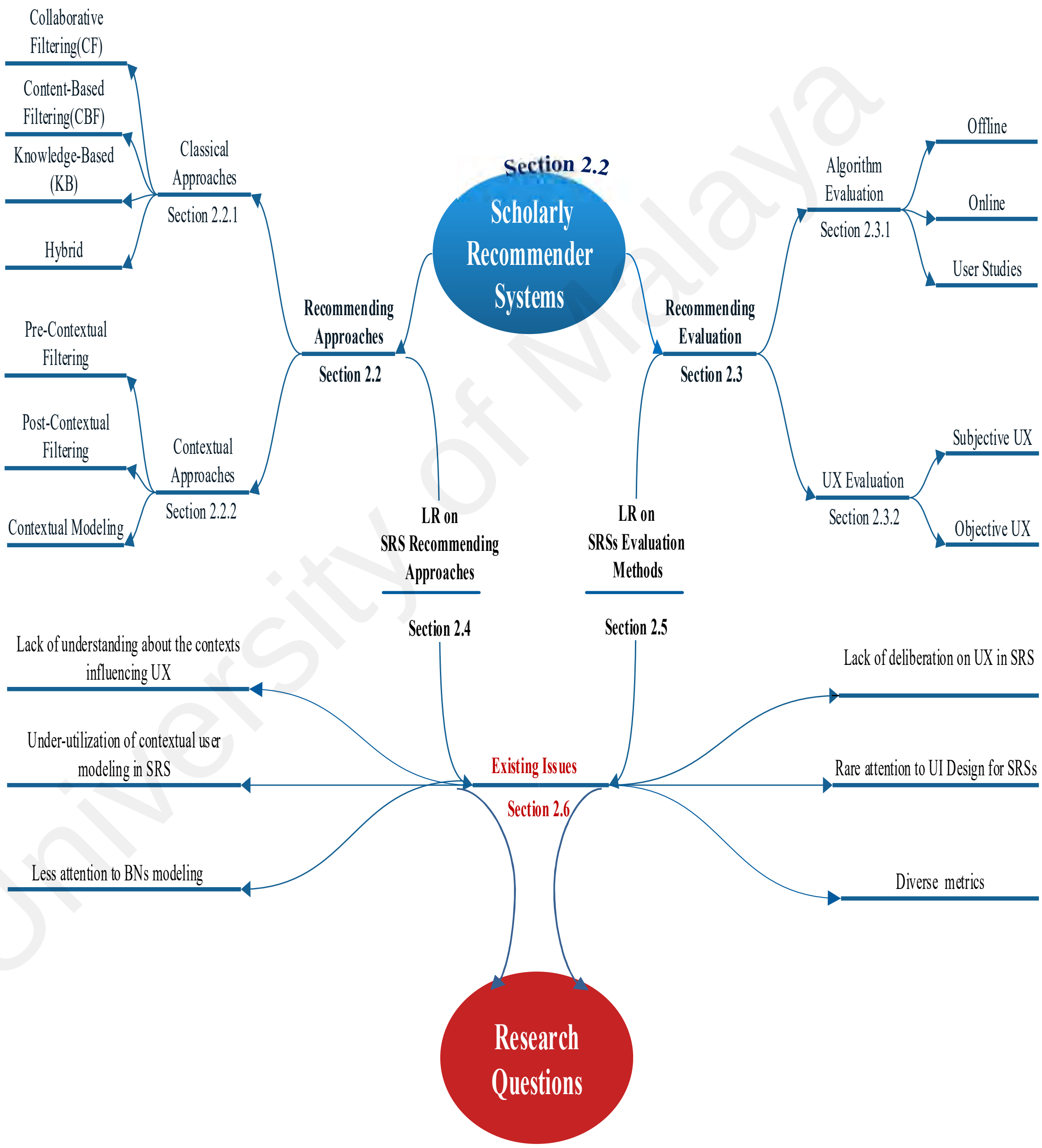


Figure 2.1: Structural of chapter 2

In the following sections, first recommending approaches including classical and contextual (Section 2.2) are discussed. Then, evaluation methods using for RSs (Section

2.3) are explained. After that, the results of literature reviews on recommending and evaluation methods are discussed (Section 2.4 & Section 2.5). At the end, the existing issues are described (Section 2.6).

2.2 Recommending approaches

As shown in Figure 2.2, recommendations mostly are provided via two approaches: 1) Classical approaches and 2) Contextual approaches. In the following figures the parts that are discussed (like section 2.2), are highlighted and the other parts (sections 2.3, 2.4) are blured.

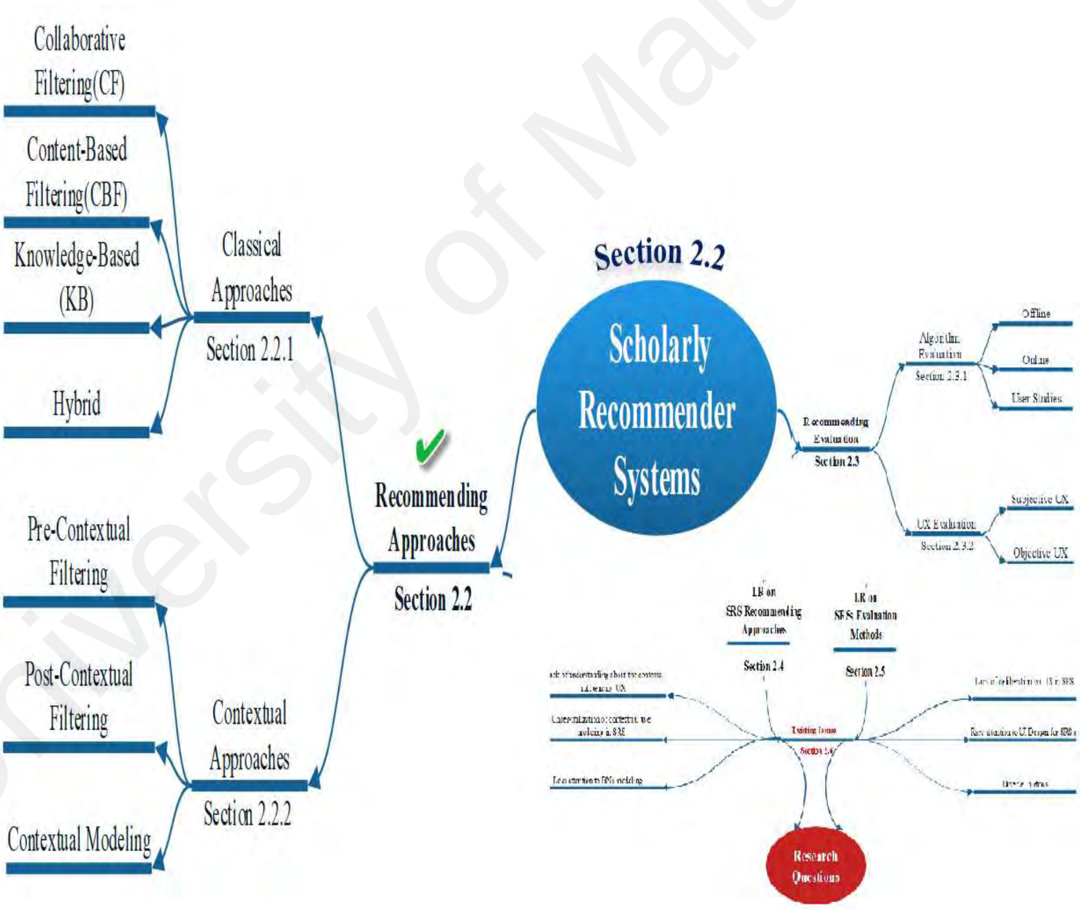


Figure 2.2: Recommending approaches (Adomavicius & Tuzhilin, 2011a)

2.2.1 Classical recommending approaches

The most common systems use classical or two dimensional (2D) approaches, which fall into three main classes: Collaborative Filtering (CF), Content- Based Filtering (CBF), and Knowledge–Based Filtering (KBF) (Adomavicius & Tuzhilin, 2005b). There is another class that is a combination of two or all of these three approaches, called Hybrid (Adomavicius et al., 2005). Classical RSs use a set of ratings that is either explicitly created by users or is implicitly deduced by a system (Adomavicius & Tuzhilin, 2011a) so that two types of entities, namely users and indicators (two-dimensions), are used to estimate the rating function R (Liu, 2013).

$$R : User \times Item \rightarrow Rating \qquad (2.1)$$

For each user u, the item i' that maximises the user's utility is defined as (Adomavicius et al., 2005):

$$\forall u \in U, i'_u = \arg\max R(u,i), i \in I \qquad (2.2)$$

This research does not attempt to discuss all available algorithms applied for classical and contextual approaches. In fact, there are a few comprehensive studies that concentrate on the methods applied for the recommenders (Adomavicius and Tuzhlin 2005; Burke 2002; Ekstrand et al. 2011; Herlocker et al. 1999, 2004;), Rather, an overview of the most important methods applied in each approach is presented.

2.2.1.1 Collaborative Filtering (CF) recommending approach

CF approaches recommend indicators to a target user based on given ratings by other users' behaviour similarities and users' functional patterns in the community (Lika, Kolomvatsos, & Hadjiefthymiades, 2013). CF is the most commonly used approach for creating recommendations based on previous users' search history. In particular, users looking for information should be able to utilise what other users have already found and evaluated (Zhang, Wang, & Li, 2008a). So this approch recommends to a target user based on the similar users' opinion on a particular item rather than the information about that indicators.

There are two main class of CF approches including memory-based (userbased) and model-based (item-based) (Ali, 2014). To make recommendation by the memory- based approch, it requires to collect and store in the momery all similar indicators rated by users, all similar users and ratings. Howevre, this approch has shorcoming of cold start problem and to cope with this shortcoming, the model-based method was developed that make recommendations by using similar indicators instead of making groups of similar users by using an offline pattern created periodically by summarizing item ratings. More infromation about the advantages and shortcomings of these two approches has been discussed in (Ali, 2014) (Lika et al., 2013).

The model-based CF methods have some advantages that make them generally smaller, faster than, and definitely as coherent as memory-based methods. First, they are predictive and can clarify the correlations in elicited data (Ekstrand, Riedl, & Konstan, 2011). Second, they need less memory space for storing data (Schafer, Frankowski, Herlocker, & Sen, 2007). Third, taking advantage of the complied model, the recommendations can be made very fast in the model-based. Indeed, they are useful in the real world where the user profiles and interests change slowly and do not require to be updated frequently (Deshpande et al. 2004; Linden et al. 2003; Sarwar et al. 2001) (Gong, 2010).

User-Item matrix is utialised to present both memory-based and model-based algorithms (Wang, et al., 2006). As dipicted in Figure 2.3, a $K \times M$ user-item matrix is showen the user's profile where X is for K number of users and M number of indicators. Each element $X_{k,m} = r$ represents the value of rating that the user K has assigned to item M. For indicators rated r ∈ {1, … , |r|} , and for unrated ones, r = ∅.

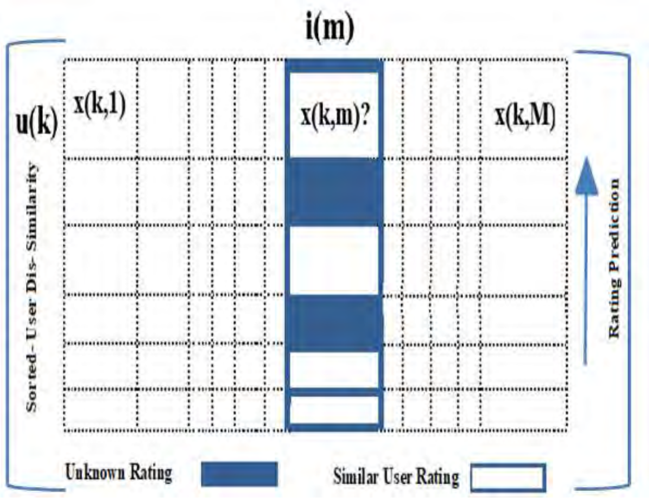


Figure 2.3: The User-Item matrix

There are two different ways to manipulate the user-item matrix considering its row vectors and column vectors. For its row vectors, each row vector u_K represents a user profile including all ratings assigned to the indicators. The memory-based CF is based on this type of representation (Ali, 2014).

$$X = [u_1, \dots, u_K], \qquad u_k = [x_{k,1}, \dots, x_{K,m}]\ k = 1, \dots, K \tag{2.3}$$

Second, it can be decomposed into column vectors: where each column vector i_m represents all ratings assigned to a specific item. This viewpoint leads to model-based CF systems(Ali, 2014).

$$X = [i_1, \dots, i_M], \qquad u_m = [x_{1,m}, \dots, x_{K,m}]\ k = 1, \dots, M \tag{2.4}$$

In the memory-based (user-based) approach, the recommender engine predicts the active user's interest for a specific item by considering the elicited information from similar users' profiles (Wang, et al., 2006). As depicted in Figure 2.4, each row vector which represents a user profile has been sorted based on its dissimilarity towards the active user's profile. Hence, the indicators rated by more similar users have better chance of being recommended to the active user.

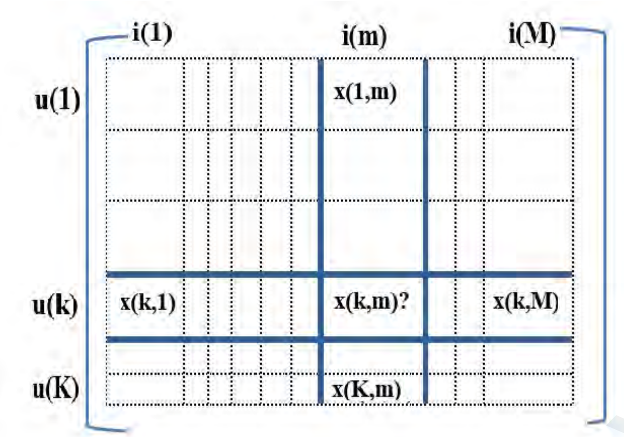


Figure 2.4: Using user similarity to predict the Ratings

A group of similar users can be generated by selecting top-N similar users $S_u(u_k)$ toward user (Wang, et al., 2006):

$$S_u(u_k) = \{u_k \mid rank\ s_u(u_k, u_a) \leq N, X_{a,m} \neq \emptyset\} \qquad (2.5)$$

$$where\ |S_u(u_k)| = N$$

In the above mentioned formula, $S_u(u_k, u_a)$ identifies the degree of similarity between users k and a. The Cosine similarity and Pearson's correlation are the most popular measures for calculation of this kind of similarity in CF (Wang, et al., 2006). In the cosine similarity measure, a and b are two users that their similarity is going to measured. The rating assigned by users a and b to item p are represented by $r_{a,p}$ and $r_{b,p}$ .The set of indicators rated by both user a and b is shown by P and finally the parameter of r_x is the average ratings which user x has been submitted. The similarity is a value between -1 and $+1$(Wang, et al., 2006).

$$Similarity\ (a,b) = \frac{\sum_{p \in P}(r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P}(r_{a,p} - \bar{r}_a)^2}\ \sqrt{\sum_{p \in P}(r_{b,p} - \bar{r}_b)^2}} \quad (2.6)$$

A few ML methods can also be used to create this ranked list of similar users (Jin, Chai, & Si, 2004). The similarity value can be calculated for all pair of users and the following formula is be applied to calculate the predicted rating $Pred\ (a,p)$ of item p by the user:

$$Pred\ (a,p) = \bar{r}_a + \frac{\sum_{b \in N}\ Similarity(a,b) \times (r_{b,p} - \bar{r}_b)}{\sum_{b \in N}\ Similarity(a,b)} \quad (2.7)$$

Where N is the set of all users who submitted a rating to item p.

Many studies have applied the memory-based CF method because of its simplicity and tangibility however as mentioned earlier, Data sparsity, Cold start, Shilling and Scalability problems are some of the shortcomings (Ekstrand, Riedl, & Konstan, 2011).

To solve some of the problems with the memory-based algorithms, the model-based CF approach, takes advantage of a model of user preferences considering the indicators instead of users to predict the ratings which includes precompiled information of indicators, users and ratings, and might be generated in several hours or days (Schafer, Frankowski, Herlocker, & Sen, 2007). As shown in Figure 2.5, the average ratings of similar indicators rated by the active user is used to make predictions (Deshpande et al. 2004; Linden et al. 2003; Sarwar et al. 2001).

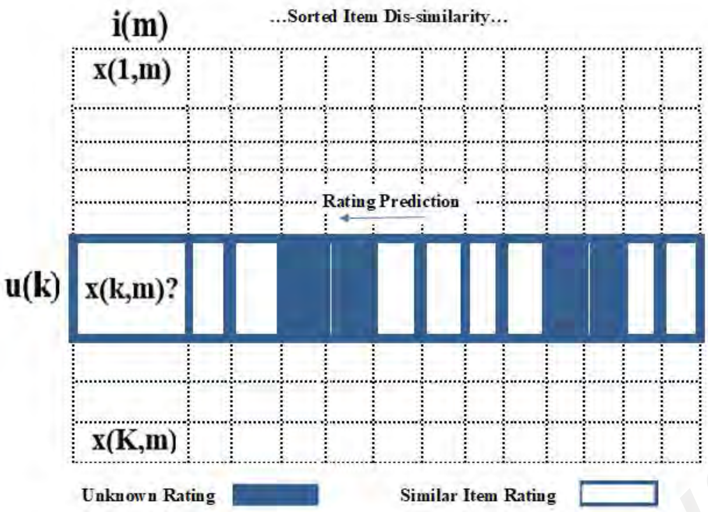


Figure 2.5: Using item similarity to predict the ratings

Like to the memory-based, here also sorting is done based on dissimilarity and the indicators (column vectors) are sorted toward the target item rather than row vectors as shown in Figure 2.5. The prediction of the most relevant indicators is based on the similarity between the indicators (Schafer et al., 2007):

$$Similarity\ (\vec{a},\vec{b}) = \frac{\sum_{u\in U}(r_{u,a}-\bar{r}_u)(r_{u,b}-\bar{r}_u)}{\sqrt{\sum_{u\in U}(r_{u,a}-\bar{r}_u)^2}\sqrt{\sum_{u\in U}(r_{u,b}-\bar{r}_u)^2}} \tag{2.8}$$

Where the set of all users who rated both item a and b is presented by U and $r_{u,a}$ and $r_{u,b}$ are the rates that have been assigned to indicators a and b respectively by user u. The prediction of the rating of user u to item p by having the similarity of indicators is calculated (Schafer et al., 2007):

$$Pred\ (u,p) = \bar{r}_a + \frac{\sum_{i\in ratedItem(u)} Similarity(i,p)\times r_{u,i}}{\sum_{i\in ratedItem(u)} Similarity(i,p)} \tag{2.9}$$

Where the degree of similarity between each member of indicators rated by user u is presented by $Similarity(i, p)$ and target item$p$, and $r_{u,i}$ donates the rating assigned by user u to item i.

(a) CF approach limitation

The model-based CF methods have some advantages that make them generally smaller, faster than, and definitely as coherent as memory-based methods. First, they are predictive and can clarify the correlations in elicited data(Ekstrand et al., 2011). Second, they need less memory space for storing data(Schafer et al., 2007). Third, taking advantage of the complied model, the recommendations can be made very fast in the model-based. Indeed, they are useful in the real world where the user profiles and interests change slowly and do not require to be updated frequently (Deshpande et al. 2004; Linden et al. 2003; Sarwar et al. 2001) (Gong, 2010). The CF approach requires user ratings however usually users are not high motivated to rate the indicators and this cause Cold start problem. To cope with this problem, implicit ratings from users' behaviors such as paper reading (C. Yang, Wei, Wu, Zhang, & Zhang, 2009), paper downloading, adding the paper to profile, editing paper details, and viewing its bibliography as positive votes (Pennock, Horvitz, Lawrence, & Giles, 2000) as well as author's citations (Sean M McNee et al., 2002) have been applied by the researchers. In the other words, users' behaviors are considered as the indicators to be liked by the user. However, considering these behaviors as the positive vote would be misguiding. For example; a user might spend long time to read a paper while it does not mean necessary that user likes the paper.

Sparsity is a general problem of in the CF approach in the field of SRSs (Beel, Gipp, et al., 2016). It means that there are typically few users but many papers, and only few users rated the same papers. Hence, finding like-minded users is often not possible. In addition, many papers are not rated by any users therefore it cannot be recommended.

Another problem with the CF is that implicit ratings is not accurate as explicit human quality assessments (Beel, Gipp, et al., 2016). In addition, using citations might also annihilate the CF's advantage of being content-independent. Typically, reliable citation data is not widely available. Therefore, access to the papers' content is required to build a citation network, but this process is even more fault-prone than word extraction in CBF. In CBF, "only" the text of the papers must be extracted, and maybe fields such as title or abstracts must be identified. For citation-based CF the text must also be extracted but in this text, the bibliography and its individual references must be identified, including their various fields (such as title and author).

2.2.1.2 Content- Based Filtering (CBF) recommending approach

On the other hand, the objectives of CBF approaches focus on finding correlations between content of indicators as opposed to correlation between users as is the case in CF approaches (Liu, 2013) (Herlocker, 2000). The root of the CBF approach can be traced back to Information Retrieval (Balabanović & Shoham, 1997). CF approach analyses the ratings for the indicators that an active user has made in the past in order to build a user model of his or her preferences (Mladenic, 1999) and then matches up the preferences with the attributes of that item. Usually, the recommendation indicators are stored in a database table.

Figure 2.6 shows a simple database of structured data. The records which have a unique identifier, or ID describe three papers and the columns represent the properties of the papers. The ML algorithms might be used to create a user profile from structured data.

ID	Title	Author(s)	Main Subject	Citation
101	Introduction to Recommender Systems	Adomavicius	Recommender System	1544
102	Contexts in recommender systems	Sankaranarayanan	Recommender System	1150
103	The next generation of recommender systems	A Tuzhilin	Recommender System	8666
104	Recommender systems handbook	F Ricci	Recommender System	2657
105	User models for enhanced personalization	S Berkovsky	Recommender System	206

Figure 2.6: A sample of structured data

Figure 2.7 shows a sample of an un-structured data which the attribute names are not with well-defined values. Furthermore, the complexity of natural language may cause problems. For example; in a text they might be words with several meaning in different contexts called polysemous words and synonyms which are different words with the similar meaning.

Generic User Modeling Systems

The paper reviews the development of generic user modeling systems over the past twenty years. It describes their purposes, their services within user-adaptive systems, and the different design requirements for research prototypes and commercially deployed servers. It discusses the architectures that have been explored so far, namely shell systems that form part of the application, central server systems that communicate with several applications, and possible future user modeling agents that physically follow the user. Several implemented research prototypes and commercial systems are briefly described.

Figure 2.7: A sample of un-structured data

There are a few methods to convert an un-structured text to a structured(Salton, 1989). One of the methods typically called stemming, formalizes the root of the words that reflects the common meaning behind the words (Porter, 1980) such as “recommend”, “recommendation”, “recommender” and “recommenders”. A value is associated with the term that represents its importance or relevance. In addition, ML methods such as Decision Trees (Kim et al., 2006), Nearest Neighbour (Yang, 1999), relevance feedback algorithms (Manning, Raghavan, & Schütze, 2008), Linear Classifiers (Zhang & Iyengar, 2002), Probabilistic and Naïve Bayes, an (Ogata, & Okuno, 2008) and clustering (Pazzani & Billsus, 1997) might be applied to create user profiles from the structured data (Pazzani & Billsus, 2007).

On the most used matures for calculation keyword weights in an un-structured text, is the Term Frequency-Inverse Document Frequency (TF-IDF) (Mangina & Kilbride, 2008). If N is defined as the total number of indicators that are eligible to be recommended to users and that keyword k_j appears in n_i of them. Furthermore, consider $f_{i,j}$ as the number of times keyword k_i appears in document d_j. Then,$TF_{i,j}$, the term frequency (or normalized frequency) of keyword k_i in item d_j, is defined as(Mangina & Kilbride, 2008):

$$TF_{i,j} = \frac{f_{i,j}}{max_{zfz,j}} \tag{2.10}$$

Where the maximum is the biggest value of the frequencies $f_{z,j}$ of all keywords k_z in item d_j. To reduce the repeated keywords in many, the Inverse Document Frequency (IDF) is used in combination with $TF_{i,j}$, which defines as(Mangina & Kilbride, 2008);

$$IDF_i = log\frac{N}{n_i} \tag{2.11}$$

Hence, for each item d_j, the TF − IDF weight for keyword k_i is identified as

$$w_{i,j} = TF_{i,j} \times IDF_i \tag{2.12}$$

Consequently, the content of item d_j is defined as:

$$Content\ (d_j) = (w_{1,j}, \dots, w_{k,j}) \tag{2.13}$$

In overall, if $ContentBasedProfile(c)$ is the profile of user c including user preferences and rates which also can be defined as a vector of weights$(w_{c,1}, \dots, w_{c,k})$, where each weight $w_{c,i}$ represents the significance of keyword k_i to user c (Cremonesi, Turrin, & Airoldi, 2011). In CBF algorithms, the utility function $u(c, s)$ is defined as:

$$u(c, s) = score\ (ContentBasedProfile(c), Conten(c)) \tag{2.14}$$

In this context, both $ContentBasedProfile(c)$ and $Conten(c)$ can be defined as TF-IDF vectors $\overrightarrow{w_c}$, $\overrightarrow{w_s}$ of keyword weights (Adomavicius & Tuzhilin, 2005), where K is the total number of keywords in system. Therefore, $u(c, s)\ is$:

$$u(c, s) = \cos(\overrightarrow{w_c}, \overrightarrow{w_s}) = \frac{\overrightarrow{w_c} \cdot \overrightarrow{w_s}}{\|\overrightarrow{w_c}\|^2 \times \|\overrightarrow{w_c}\|^2} \tag{2.15}$$

$$= \frac{\sum_{i=1}^{K} (w_{i,c}\ w_{i,s})}{\sqrt{\sum_{i=1}^{K} w_{i,c}{}^2} \sqrt{\sum_{i=1}^{K} w_{i,s}{}^2}}$$

The plain words are applied in the most approaches as features and some also have used n-grams (Ferrara, Pudota, & Tasso, 2011). It is axiom that the words in different part of a document have different discriminative powers. For example; the word in the title of a document is usually more representative of the document content rather than the word in the body-text (Nascimento, Laender, da Silva, & Gonçalves, 2011). Among those which have used the weighting scheme such as Latent Semantic Analysis (LSA), Topic

Modelling (TM) (R. Patton, T. Potok, & B. Worley, 2012), plain Term Frequency (TF), Term Frequency Inverse Document Frequency (TF-IDF), phrase depth and life span (Sugiyama & Kan, 2010), the most popular, 83% has been applied TF-IDF (Ferrara et al., 2011). Co-occurrence is a method to calculate the document relevancy by using proximity of co-citations. If the proximity of two references is high within a paper, it means that the relevancy of cited papers is more(Gipp, Beel, & Hentschel, 2009a). Co-occurrence has been used specially in two well-known recommender of bX and (Van De Sompel & Bollen, 2006) (Mönnich & Spiering, 2008b). Global relevance is an additional ranking method which measure measures overall popularity of an item (He, Pei, Kifer, Mitra, & Giles, 2010). Usually, a primary list of recommendation is generated and then the list is re-ranked using the global relevance metrics (Bethard & Jurafsky, 2010) such as PageRank, HITS, Katz, citation counts, venues' citation counts, citation counts of the authors' affiliations, authors' citation count, h-index, title length, number of co-authors, number of affiliations, and venue type(Councill, Giles, & Kan, 2008; Zarrinkalam & Kahani, 2013a).

(a) Limitation of CBF

There are various ML methods applied for CBF (Middleton, De Roure, & Shadbolt, 2002). The Vector Space Model (VSM) by using cosine measure calculates similarities between user models and recommendation candidates. The Graph networks such as Bayesian Networks (Liu, 2013)typically represent the relations between the features such as papers citations, authors, users/customers, venues, genes and proteins, and the years the papers were published(Liang, Li, & Qian, 2011). Hierarchical classification and ontology are also the two most frequently used methods. Comparing to CF, CBF systems have some advantages such as user independency in ratings, new indicators generation (Lops, De Gemmis, & Semeraro, 2011). However, the CBF approach also has its shortcomings such as limited content analysis, which requires more analysis of the content to distinguish between the users' preferences and those that the users are not interested in. A comparison between CF and CBF is mentioned in Table 2.1.

Table 2.1: CF & CBF Comparison (R. Burke, 2002; Vivacqua & Oliveira, 2009)

Feature	CF	CBF
Independency to users' fates (first- rated problem)	No	Yes
Diversification of recommendations	No	No
Recommendation to new Users	No	No
Novelty of recommendations (over- specialization problem)	No	N o
Transparency	Yes	Yes

2.2.1.3 Knowledge Based Filtering (KBF) recommending approach

Another classical or two dimensional (2D) approach is Knowledge- Based (KBF) which provide recommendations based on specific knowledge or predefined (or learned) rules about users and items (Burke, Hammond, & Young, 1996) to deduce applicable links between user requirements and items that might be required to fulfil them (Resnick & Varian, 1997). Based on the literature, KBF approaches might apply intelligent methods such as Neural Networks, Fuzzy Logic, Genetic Algorithms, Decision Trees, and Case-Base reasoning (Will, Srinivasan, Im, & Wu, 2009b). Unlike other approaches, KBF approaches do not depend on large bodies of statistical data about particular rated items or particular users, since only enough knowledge is needed to judge items as similar to each other (Will, Srinivasan, Im, & Wu, 2009b). The KBF approach is strongly complementary to other types of approaches (Burke et al., 1996).

2.2.1.4 Hybrid recommending approach

As mentioned earlier, hybrization is combination of two or three approaches of CBF, CF and KBF (Burke, 2002; Vivacqua, Oliveira, & de Souza, 2009). The Hybrid approach is utilised to reduce the limitations and improve the system efficiency. For example, CF approaches can be useful if a superabundant number of users' behaviours have been identified as well as an adequate number of rated indicators have been accounted for. As mentioned earlier, they suffer from the cold start problem (Ricci, Rokach, & Shapira, 2011). On the other hand, CBF approaches are extremely dependent on content analysis; if the content analysis does not include adequate information to differentiate a user's preferred indicators from those indicators the user does not like, no helpful recommendation can be made (Verbert et al., 2010). Hence, the hybrization of CBF, CF and KBF methods is an alternative to cope with these shortcomings (Felfernig & Burke, 2008) (Verbert et al., 2010). In a true hybrization, the combined approaches are more or less equally important (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013). TechLens recommender (Konstan, Kapoor, McNee, & Butler, 2005) is one of the most famous and influential SRSs developed by using Hybrid approach (Beel, Gipp, et al., 2016). A few researchers also have applied KB methods such as (Will et al., 2009b), (C.-S. Tsai & Chen, 2008a), (Marko A. Rodriguez, 2009) howere in the field of SRSs the number of studies applied KB, is not considerable.

Three are three more approaches in the classification of RSs which are based on context- awareness called contextual or multi- dimensional approaches. According to Bamshad, there is always a context and RSs are not "usable" without context (Mobasher, 2012). In the following section, first definition of context and contextual recommending are discussed then three contextual approaches including are Contextual pre-filtering, Contextual post- filtering and contextual modeling are elaborated.

2.2.2 Contextual recommending approaches

In the late twentieth century, the epistemological contextualisation was developed by philosophers. This theory indicates that the standards of knowledge and justification change with the context. Particularly, understanding of context is necessary for better comprehension of a situation since when the context shifts, the knowledge about the situation will shift as well (Craig, 1998). From a general point of view, the Oxford Advanced Learner's Dictionary mentions that context is “a situation in which something happens and that helps you to understand it” (Crowther, 1995). Likewise, according to the Webster’s dictionary (M. Webster, 2006), “Context is a situation in which something happens: the group of conditions that exist where and when something happens.”

Many definitions of context have been proposed in various disciplines, including computer science (primarily in artificial intelligence and ubiquitous computing), information retrieval, cognitive science, linguistics, philosophy, social science, psychology, and organisational sciences (Adomavicius & Tuzhilin, 2011a). It is beyond the scope of this research to review all of them. However, a few definitions proposed in the field of computer science are reviewed in the following.

The term context appeared in computer science in the late 1980s (Hong et al., 2009), and the idea of context awareness in computing was introduced by Schilit in 1994 (Brown et al., 1997) in order to increase the richness of communication and provide more useful computational services (Dey, 2001). Since then, many studies in the field of computer science tried to define the term “context”. Some studies present parametric definitions that stipulate context as a set of parameters such as time, temperature, lightness, and speed, while others define context generally and try to explain context and its territories. For example, Schilit and Theimer(Schilit & Theimer, 1994) defined context as location, identity, nearby people, and objects. In a similar definition by Brown (Brown et al., 1997), context consists of location, identity, nearby people and objects and season. Meanwhile, Pascoe (Pascoe, 1998) explained that context corresponds to the following questions:

1. Where are you?
2. Who are you with?
3. What resources are nearby?

One of the most cited definitions in computer science was offered by Dey and Abowd (Abowd et al., 1999). They expressed that context is any information that can be used to characterise the situation of an entity. They categorised context into four dimensions: location, identity, time, and activity (Figure 2.8). In this definition, there are two context levels: primary contexts, which are the four mentioned dimensions and secondary contexts gained from primary contexts. As an illustration, many pieces of related information such as phone numbers, addresses, email addresses, birth date, etc., can be acquired from the location of an entity. Such information acquired from primary contexts is numerated as secondary contexts (Abowd et al., 1999).

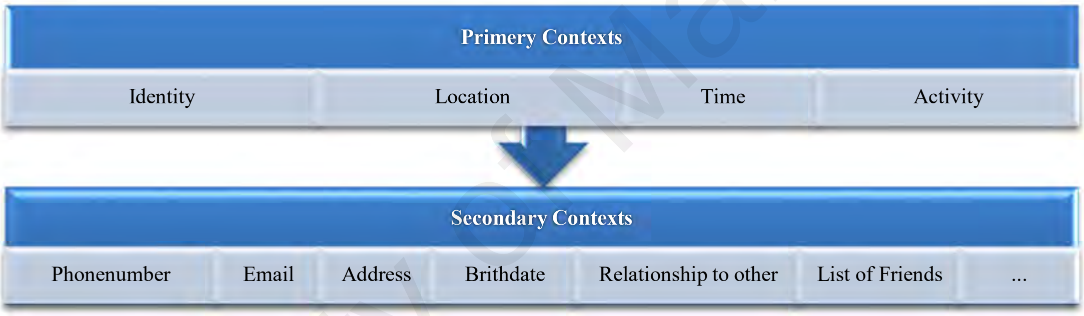


Figure 2.8: Context levels presented by Dey & Abowd (Abowd et al., 1999)

In another computer science point of view, Lieberman and Selker (Lieberman & Selker, 2000) interpreted context as "everything" that "affects the computation except the explicit input and output", including the state of user, physical environment, computational environment, and history of user-computer environmental interaction. Dourish(Dourish, 2004) expressed the context as "the features of the environment within which the activity takes place", and indicated that it is separate from the activity itself. He explains that the scope of contextual features is defined dynamically, and it is accessioned rather than static. Haseloff (Haseloff, 2005), as shown in Figure 2.9, presented a model of contextual factors based on Object Oriented (OO) concepts and

Unified Modelling Language (UML), including surroundings, state, location and reachability.

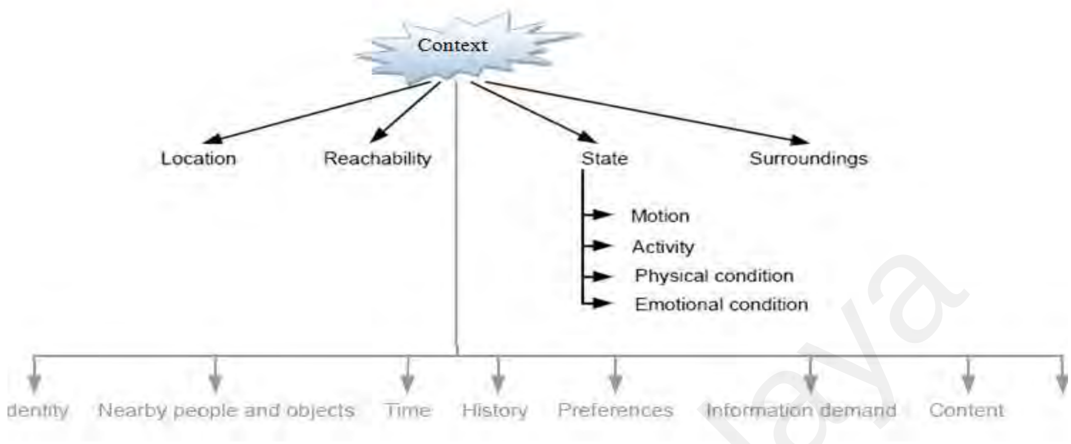


Figure 2.9: Constituent elements of context

Bazire and Brézillon (Bazire & Brézillon, 2005) analysed 150 definitions coming mainly from the web in different domains. However, they concluded that it is difficult to reach a consensus on what exactly context is. Thus, trying to reach a consensual definition for context is an ineffectual effort since the concept of ‘‘context” evokes different impressions in each reader and context may include almost everything (Kocaballı & Koçyiğit, 2007). Furthermore, it is difficult to present a definition that encompass all the aspects it refers to (Tamine-Lechani, Boughanem, & Daoud, 2010). The definition of context in RSs was investigated by Verbert (Verbert et al., 2010) and (Adomavicius & Tuzhilin, 2011a). They emphsised that contextual information is any additional information that has a direct impact on the relevance of recommendations.

In the case of scholarly recommendations, contextual information describing a scholar's environment or situation or other entities such as paper which is incorporated into the process of making recommendations to improve recommendations; these systems are called Context- Aware Scholarly RSs (CASRSs) (Adomavicius & Tuzhilin, 2011a).

The preferences are estimated with the rating function of indicators, users and context as follows c

$$R: \ Scholar \times Paper \times Context \rightarrow Rating \qquad (2.16)$$

In the classical approaches, function R is defined as a two-dimensional matrix including basic dimensions of scholar and paper. It can be also described as a multidimensional matrix by a set of features. However, since the features are not based on any accepted ontology, the basic dimensions of scholar and paper are ambiguous if they are considered separately as a multidimensional matrix. In addition, there are common features such as time that might be considered as the features describing both the users and indicators. For example, in the case of SRS, time can be considered as a feature for the paper or also might be the user feature (e.g., a user is searching for a recent published paper). To conquer this, a single multi-dimensional space of features is considered to represent the whole list of features called contextual information, where certain sets of features can be grouped into the basic dimensions of user and item(Berkovsky et al., 2008). Therefore, if the contextual information is defined with a set of contextual dimensions *D*, while two of these dimensions are *Scholar* and *Paper*, and the rest are contextual; the rating function *R* is:

$$R = D_1 \times D_2 \times D_3 \times \ \dots \ \times D_n \rightarrow Ratings \qquad (2.17)$$

The utility function is defined by selecting certain "what" dimensions $D_{i1},..., D_{ik} \left(k < n\right)$ and certain "for whom" dimensions $D_{j1},\ldots,D_{jl} \left(l < n\right)$ that do not overlap, i.e. $\left\{ D_{i1},..., D_{ik}\right\} \cap \left\{ D_{j1},\ldots,D_{jl}\right\} = \varnothing$, and recommending for each tuple $\left(d_{j1},\ldots, d_{jl}\right) \in D_{j1} \times \ldots \times D_{jl}$ the tuple $\left(d_{i1},\ldots, d_{ik}\right) \in D_{i1} \times \ldots \times D_{ik}$ that maximises rating $R\left(d_1,\ldots, d_n\right)$ (Berkovsky et al., 2008). In particular:

$$\forall \left(d_{j1},\ldots, d_{jl}\right) \in D_{j1} \times \ldots \times D_{jl}, \ \left(d_{i1},\ldots, d_{ik}\right) \qquad \mathbf{(2.18)}$$

$$= \underset{\substack{(d'_{i1},\ldots,d'_{ik}) \in D_{i1} \times \ldots \times D_{ik} \\ (d'_{j1},\ldots,d'_{jl}) = (d_{j1},\ldots,d_{jl})}}{\operatorname{argmax}} R(d'_1,\ldots,d'_n)$$

For example, in recommending papers to scholars, if the Paper (title, keywords, author, subject, publisher, year) and Scholar (name, age, degree, interests), and the contexts is also defined as *Location*, where the scholar is looking for a paper; *L={university, home} and Time*, when the scholar is looking for a paper; *T={first semester, second semester}*(Berkovsky et al., 2008). Hence, the function will become:

$$R = Scholar \times Paper \times Location \times Time \quad \textbf{(2.19)}$$

Contextual information can be incorporated into the classical recommendation procedures in order to generate better recommendations (Baltrunas & Ricci, 2009). In the following, the three categories of contextual approaches including pre-filtering, post-filtering, and contextual modelling are discussed (Adomavicius & Tuzhilin, 2011b; Kantor, Rokach, Ricci, & Shapira, 2011). Before explanation of contextual recommending approaches a brief overview of context definition for better understanding of contextual approaches is provided.

2.2.2.1 Contextual pre-filtering

In contextual pre-filtering (Figure 2.10.a), the contextual information is used before all the ranked recommendations are computed. The reduction-based approach (Adomavicius & Tuzhilin, 2005b) is an example of pre-filtering in which, first, all ranked recommendations are computed through classical methods like CFB; then they are adjusted or re-ranked for each user using contextual information (Adomavicius & Tuzhilin, 2005b).

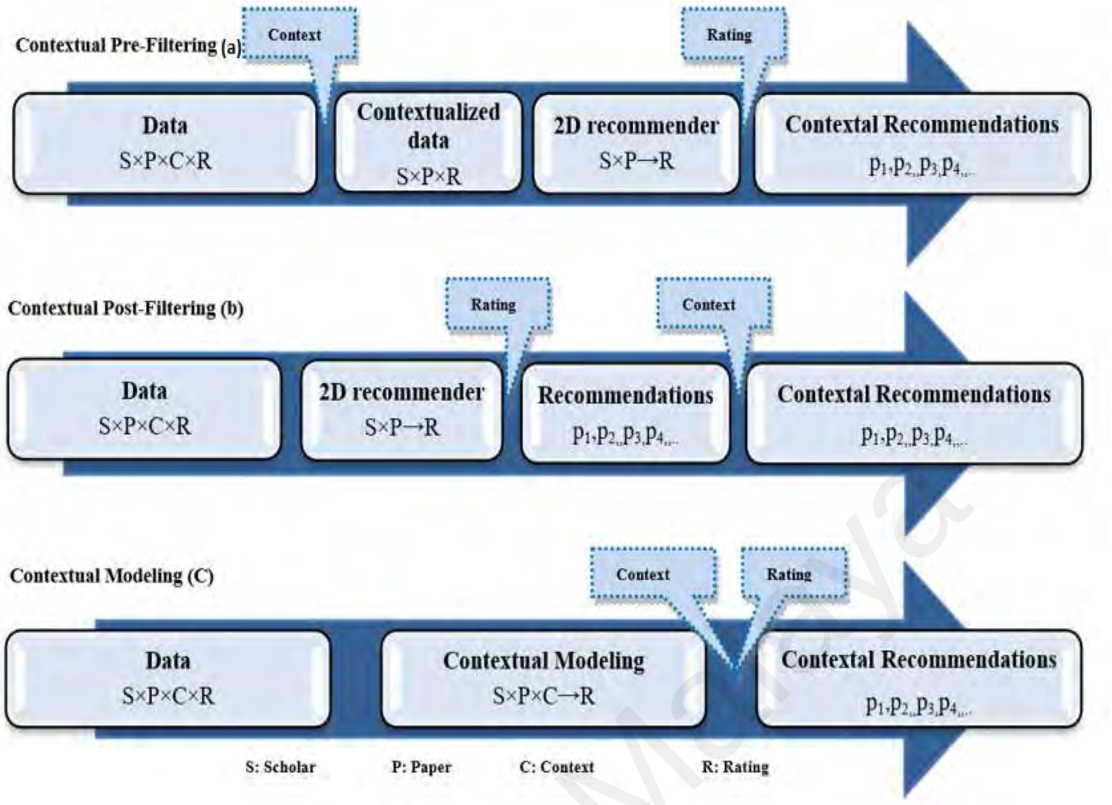


Figure 2.10: Context incorporation into recommending (Panniello et al., 2009)

2.2.2.2 Contextual post-filtering

Panniello and his colleagues (Panniello et al., 2009)presented a post-filtering strategy (Figure 2.10.b) that penalises the recommendations of indicators with few ratings in the target context (Campos, Fernández-Tobías, Cantador, & Díez, 2013) (e.g., filtering based on context similarity).Conversely, in contextual post-filtering, after computation of all the ranked recommendations, the contextual information is used (Panniello et al., 2009).

2.2.2.3 Contextual modeling

Contextual modeling approaches (Figure 2.10.c) are mostly the ML based algorithms such as decision tree, regression, and probabilistic models, which are able to alleviate the sparsity problems and produce better context-aware recommendations than the ones by the filtering-based methods. These models directly incorporate context information as parts of the predictive functions or learning preference models, and contexts are no longer used as filters in the recommendation process(Zheng, 2017). To put it differently, contextual variables are added as dimensions ($D_1, \ldots, D_n$) in the feature space or recommendation function in addition to the user (scholar) and item (paper) dimensions (Baltrunas et al., 2012). As shown in Figure 2.11, dimension “goal” has been added to the recommendation function as a contextual information.

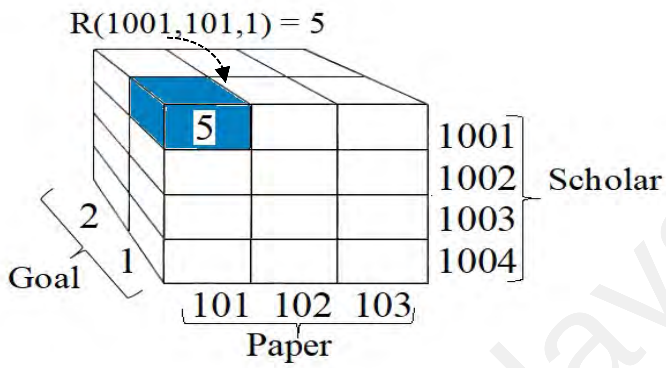


Figure 2.11: Multidimensional model (Zheng, Mobasher, & Burke, 2014)

Many different approaches in recent years have been proposed for the contextual modeling. According to(Zheng, Mobasher, & Burke, 2014), they are classified into four groups of;

1. Extension of standard CF
2. Heuristic distance-based
3. Matrix/Tensor Factorization
4. Probabilistic latent variable context models

Recent approaches have been applying regression models such as Tensor Factorization (TF) to contextual modeling in order to fit the data and to extent two-dimensional matrix factorization problem into a multidimensional version of the same problem (Karatzoglou, Amatriain, Baltrunas, & Oliver, 2010). The Multi-dimensional matrix is factored into lower-dimensional representation, where the user, the item and each contextual dimension are represented with a lower dimensional feature vector but the problem is TF might generate a huge number of model parameters that must be learned using the training data. One of the solution for solving this problem is using context-aware Matrix Factorization (MF) (Baltrunas, Ludwig, & Ricci, 2011). However, it may be difficult to interpret the models such as MF models based in order to understand why and how contexts play an important role on quality of recommendation (Zheng, 2017). Therefore, there are limited research that try to utilize the model to interpret the contextual effects in the RSs. Due to the difficulty of interpreting the contextual modeling approaches, most of the existing work focus on the interpretations by the contextual filtering methods, especially the pre-filtering approaches (Zheng, 2017).

In section 2.2, two recommending approaches: 1) Classical approaches including Collaborative Filtering (CF), Content- Based Filtering (CBF), Knowledge–Based Filtering (KBF) and, Hybrid approaches and 2) Contextual approaches including Contextual pre- filtering, Contextual post- filtering and Contextual modelling were explained along with the different methods were explained in the following section 2.3, the recommending evaluation methods are discussed.

2.3 Recommending evaluation

In this part, the recommending evaluation in two perspectives of evaluation method and UX evaluation along with the evaluation metrics applied in the literature of SRSs are discussed (Figure 2.12).

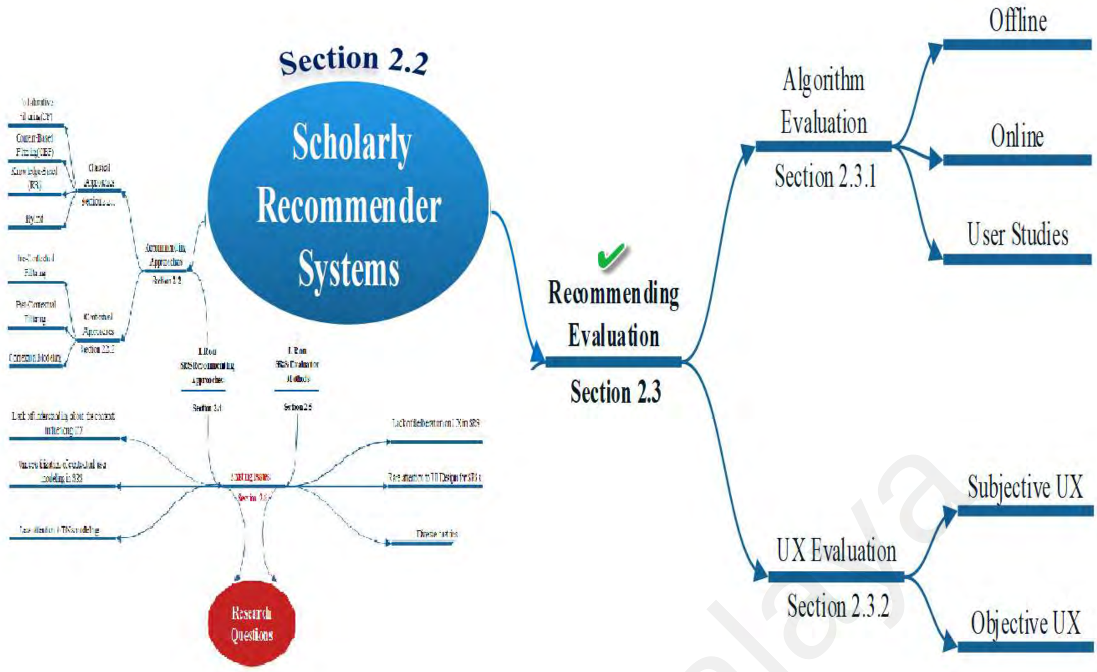


Figure 2.12: Classification of recommending evaluation methods & metrics

2.3.1 Recommending evaluation methods

As shown in Figure 2.12, evaluation methods, independent of the domain, are classified into three categories: offline, online and user studies evaluation (Beel, Genzmehr, Langer, Nürnberger, & Gipp, 2013; J. L. Herlocker, Konstan, Terveen, & Riedl, 2004; Shani & Gunawardana, 2011). In the following, a brief explanation of the above- mentioned methods is presented.

2.3.1.1 Offline evaluation method

Offline method utilises a pre-compiled offline dataset, which is divided into the test and training sets while some information is eliminated (Said, 2013). Then, the ability of an algorithm to predict the removed information (the rates in data set) is evaluated by using the ratings in the training set. The outcomes can be compared and contrasted with the real rates in the test set (de Wit, 2008). According to a recent survey of 330 papers, the research and investigation of RSs in the last five years relied very much on offline evaluation methods (Jannach, Lerche, Gedikli, & Bonnin, 2013). Three different datasets are considered in offline evaluations: Explicit ground-truths, Inferred ground-truths and Expert ground-truths (Beel, Genzmehr, et al., 2013; Beel & Langer, 2014).The ground truth is a term that refers to the data collected direct observation from the real world. To carry out an effective ground-truth dataset, random indicators from the database are discarded; and by utilising the remaining indicators, recommendations are produced.

An Explicit ground-truths is based on the explicit users' ratings about the papers. The accuracy of recommender prediction is evaluated when some ratings are removed from the dataset (ratings of papers *4*, *5* in Figure 2.13). If the recommender is able to predict the missing rating, it can say that the recommender is accurate.

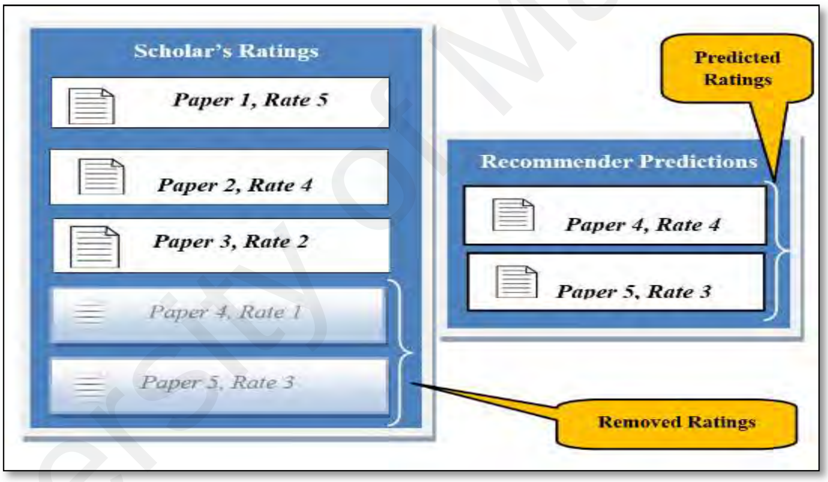


Figure 2.13: Explicit ground –truth (Beel, Genzmehr, et al., 2013)

Inferred ground-truths databases are typically based on personal users' profiles or ratings that have been implied implicitly. For instance; a list of papers that user *u* cited, or downloaded are inferred as the most relevant recommendations. As depicted in Figure 2.14, User *u* has three research papers in her profile (Paper *1*, *2*, and *4*). The recommendation approach recommends four papers (Paper *1, 2, 3* and *4*), only one of which is in *u*'s collection (Paper *3*). Only paper *3* is considered a "good" recommendation, while the others are not.

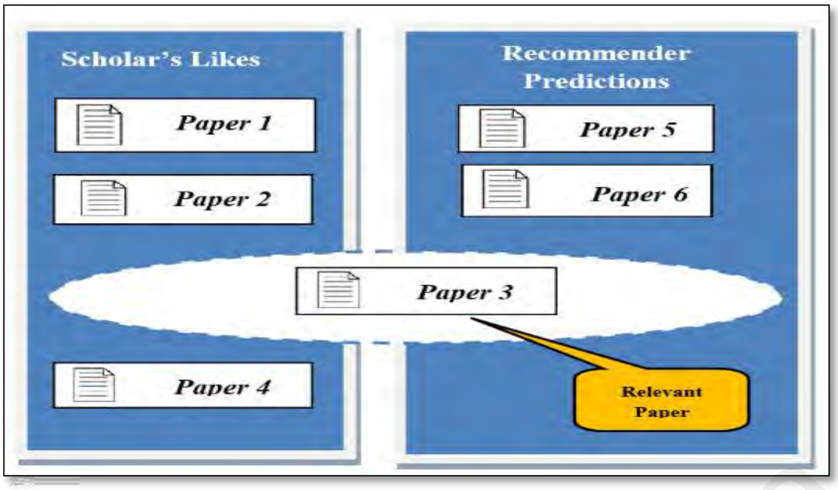


Figure 2.14: Inferred ground –truth (Beel, Genzmehr, et al., 2013)

The Expert ground-truths contain item classifications that are manually compiled by experts. Offline method is independence in interacting with real users therfore it makes this method reasonable and appealing to researchers. However, there are a number of shortcomings with offline methods. For one, they might only be applied to exmaine a very limited range of measures, such as the accuracy of a RSs (Ali, 2014). Second, the offline data sets should be as similar as possible to real data in a runtime environment. Also, it is very crucial that the distributions of selected indicators, ratings and users is not bias. To reduce the experiment cost, randomly exclusion of users or indicators with low counts which do not result in systematic bias is suggested (Mahmood & Ricci, 2007a).

2.3.1.2 Online evaluation method

Online method provides users with real recommendations from the running or online system. By implementing this method, an RS monitors users' ratings as to how they accept a recommendation given to them (Said, 2013; Zaier, Godin, & Faucher, 2008) by a few metrics such as Click Through Rate(CTR) or Cite-Through Rate(CiTR), but it is commonly measured by CTR. For example, if an RS gives 5000 papers and 220 papers have been clicked or downloaded or opened, the CRT would be 4.4 % which is considered as implicit users' satisfaction metric. Beel and Langer (2014) indicate that CRT results are not realible in all situations because a scholar might download the recommended paper but he may not be satisfied after reading it (Beel & Langer, 2014).

As in online evaluation the efficiency of the recommender is measured by real users in a real environment in terms of overall satisfaction and user retention or long-term profit, it is considered as the strongest evaluation method of a RS (Shani & Gunawardana, 2011) especially for comparison of different recommending algorithms. There are a number of essentials for running the online test. First, the user selection should be random in order to demonstrate the fair comparisons. Second, if the evaluation concentration is the UI, the underlying algorithm requires to be maintained fixed and vice versa (Kohavi et al. 2009). Third, there might be a few side effects with using online test that should be taken into consideration as well. For example, if the RS generates irrelevant indicators, it could influence users' perceptions and discourage them from using the real system ever again (Shani & Gunawardana, 2011).

2.3.1.3 User studies evaluation method

User Studies are carried out when users explicitly and implicity quantify and qualify their expectations and contentment with the real recommendations generated by various independent algorithms; and the algorithm with the highest rating is regarded as the best(Pu et al., 2012a; Shani & Gunawardana, 2011).

There are two types of "lab" and "real-world" user studies evaluations. Users are informed that they are part of the user study evaluation in lab studies; therefore, their behaviour and consequently test results might be influenced by many factors. In real-world studies, users are not apprised of the study, and real results can be observed without any intervening factors that may influence the results. It is worth telling that, user studies measures user satisfaction in an actual environment of recommending; therefore, they are not able to measure the accuracy of an RS because at the time of recommending, users are not aware of the most relevant recommendation (Said, 2013).

Among the evaluation methods, user studies is the only method that allows collection of qualitative data. Since the users are closely monitored while performing the tests, this method enables a large set of quantitative measurements to be gathered. This method is able to cover the widest range of questions.

Like other methods, user studies method also has some limitations. Indeed, the test must be conducted with the end-users in a real environment (scholars for scholarly recommender systems). To avoid bias users' responses, the goal of the experiment should not be informed the test users (Ali, 2014). It would be wise to limit the project scope to a small set of users and accordingly, a small set of tasks. It would also be preferable for each scenario to be repeated several times in order to achieve reliable results. Furthermore, in instances when applications malfunction during particular user interactions, pilot user studies should initially be executed to prevent failed experiments. User studies is primarily too costly to conduct, and accumulating a large set of test users to carry out a sufficiency large set of tasks requires a lot of effort. Therefore, this method is not popular among the RS researchers (Beel, Genzmehr, et al., 2013).

As discussed in the previous section (section 2.3.2), evaluation methods, independent of the domain, are classified into three categories: offline, online and user studies evaluation (Beel, Genzmehr, Langer, Nürnberger, & Gipp, 2013; J. L. Herlocker, Konstan, Terveen, & Riedl, 2004; Shani & Gunawardana, 2011). In the following, a brief explanation of recommending evaluation metrics is presented.

2.3.2 Recommending evaluation metrics

Many evaluation metrics for measuring various dimensions have been used to analyse the results of recommendation algorithms (Gunawardana & Shani, 2009). Mostly, offline evaluations apply variety metrics to measure the performance of the recommender (Ekstrand, 2014). Some of the metrics indicated in literature are accuracy metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)(de Wit, 2008);Top-N metrics such as Precision, Recall, ROC curves, A-measures, Mean Reciprocal Rank (MRR)(Ekstrand, 2014);and Serendipity metrics(Murakami, Mori, & Orihara, 2007) such as Unexpectedness(Adamopoulos & Tuzhilin, 2011), Relevance, Novelty, Coverage(Ge, Delgado-Battenfeld, & Jannach, 2010a),User Satisfaction, Learning Rate, and Confidence. Review and discussion on the whole metrics is beyond the scope of this thesis, for more details of different metrics, refer to studies of (Gunawardana & Shani, 2009; Murakami et al., 2007; Parra & Sahebi, 2013; Schröder, Thiele, & Lehner, 2011). In the following, a few important and most used metrics of RSs are discussed.

2.3.2.1 Prediction accuracy metrics

Prediction accuracy is one of the most important indicators that is measured in the majority of RSs (Shani & Gunawardana, 2011). The prediction accuracy is basically independent of the user interface, it can thus be used in offline experiments. A typical assumption in study of RSs is that the RS with more accurate predictions will be better from the user′s point of view. In following, three different categories of metrics for measuring the accuracy of predictions including metrics for computing the accuracy of ratings predictions, metrics for computing the accuracy of usage predictions, and finally metrics for computing the accuracy of rankings given to indicators are discussed.

(a) Measuring prediction accuracy of ratings

Sometimes the main goal of a recommender is to predict the users' ratings. Therefore, the accuracy of the predicted ratings is important in this situation. The most popular metric applied for this situation is Root Mean Squared Error (RMSE) (Shani & Gunawardana, 2011) where predicted ratings $\hat{r}_{ui}$ for a test set τ of user-item pairs (u, i) for which the true ratings r_{ui} are known. The r_{ui} are considered as known, because they are hidden in an offline experiment approach, or because they would be obtained through a user study or online experiment. The RMSE between the actual and predicted ratings is identified by:

$$RMSE = \sqrt{\frac{1}{|\tau|} \sum_{(u,i)\in\tau} (\hat{r}_{ui} - r_{ui})^2} \tag{2.20}$$

And also, there is another alternative that is called Mean Absolute Error (MAE):

$$MAE = \sqrt{\frac{1}{|\tau|} \sum_{\sum(u,i)\in\tau} |\hat{r}_{ui} - r_{ui}|} \tag{2.21}$$

Normalized RMSE (NRMSE) and Normalized MAE (NMAE) are two other versions of RMSE and MAE that have been normalized by the range of the ratings (i.e. $r_{max} - r_{min}$). As the functionality of normalized versions is the same as RMSE and MAE metrics, their resulting rankings are the same as the un-normalized metrics. Average RMSE and Average MAE are mostly used in case of unbalanced test sets. In this situation, the RMSE or MAE methods might be heavily suffered through the side effects of the error on a few very frequent indicators. To measure the prediction error on any item, it is advised to apply MAE or RMSE for each item separately and then use the average value over all indicators.

(b) Usage prediction metrics

Usage prediction metrics measures the frequency of relevant or irrelevant indicators recommender by an algorithm. It helps to know if the system recommends an appropriate item that the user will use it (Shani & Gunawardana, 2011). As mentioned previously, these metrics are not used for directly measuring the qualifications of a RS to predict ratings accurately. Precision and Recall are two metrics that mostly used for the above-mentioned context (Cleverdon, Mills, & Keen, 1966) (Shani & Gunawardana, 2011). For calculation of the precision and recall, Table 2.2 which is a 2×2 table is used.

Table 2.2: The possible conditions of recommendation to users (Rohani, 2014)

	Selected (Recommended)	**Not selected (Not Recommended)**	**Total**
Relevant (Used)	N_{rs}	N_{rn}	N_r
Irrelevant (Not used)	N_{is}	N_{in}	N_i
Total	N_s	N_n	N

Figure 2.15 also shows how to transform the rating and indicators into a binary scale if they already in a different scale. Also, it is required to separate the indicators into the set that was recommended to the user (selected/recommended), and the set that was not.

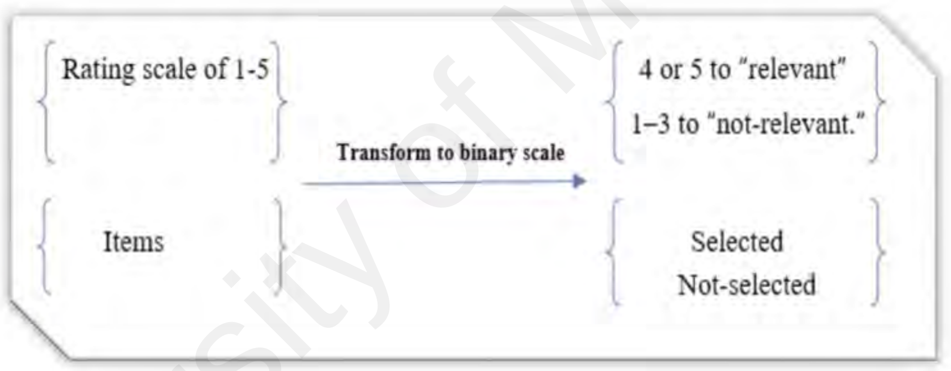


Figure 2.15: Transform to binary scale

Based on the situations as shown in Table 2.2, four conditions for recommendation of indicators to the users are possible:

N_{rs} = Relevant (Used) and Selected (Recommended) Indicators

N_{is} = Irrelevant (Not Used) and Selected (Recommended) Indicators

N_{rn} = Relevant (Used) but Not Selected (Not Recommended) Indicators

N_{in} = Irrelevant (Not Used) and Not Selected (Not Recommended) Indicators

N_s = Total number of Selected (Recommended) Indicators

N_n = Total number of Not Selected (Not Recommended) Indicators

N_r = Total number of Relevant (Used) Indicators

N_i = Total number of Irrelevant (Not Used) Indicators

Therefore, Precision or True Positive Accuracy (TPA) is calculated as the ratio of selected (recommended) indicators that are used (relevant) to the total number of selected (recommended) indicators (Herlocker, et al., 2004):

$$Precision = TPA = \frac{N_{rs}}{N_{rs} + N_{is}} \tag{2.22}$$

This is the probability that a recommended item corresponds to the user's interests and preferences. Recall or True Positive Rate (TPA) is calculated as the ratio of selected (recommended) indicators that are used (relevant) to the total number of used indicators (Herlocker, et al., 2004) which is the probability that a used (relevant) item is recommended.

$$Recall = TPA = \frac{N_{rs}}{N_{rs} + N_{rn}} \tag{2.23}$$

Precision and Recall are inversely related and depend on the separation of the concept of relevant and irrelevant indicators. There is not consensus on the definition of “relevance” and the suitable approach to calculate (Harter, 1996; Voorhees, 2000). The majority of information retrieval evaluation methods have focused on an objective viewpoint of relevance, where it is defined with respect to a query, and is independent of the user. In doing so, the documents can be compared with queries to determine which documents are relevant to which queries. However, in case of RSs, the objective relevance makes no sense and it′s not applicable.

The only person who can judge if an item is suitable, is the user. Therefore, relevance in the field of RSs is considered as a subjective concept (Herlocker, et al., 2004).

As mentioned earlier, there is a mutual dependence between Precision and Recall. Specially, in case of longer recommendation lists, the Recall is increased while the precision is decreased. To make a trade-off between these two metrics, Precision and Recall is considered in conjunction with another metric called Fall-out or False Positive Rate (FPR). Fall-out is measured as the ratio of selected (recommended) indicators that are not used (irrelevant) to the total number of not used indicators (Hernández del Olmo & Gaudioso, 2008) which is the probability that an irrelevant (not used) item is recommended to the user.

$$Fall - out = FPR = \frac{N_{is}}{N_{is} + N_{in}} \tag{2.24}$$

F-measure is a metric from combination of Recall and Fall- out, the harmonic mean of precision and recall which defines as (Rohani, 2014):

$$\boldsymbol{F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}} \tag{2.25}$$

In overall, if the number of recommended offering to the users is preordained, the most useful metric is Precision at N where N is the number of recommended indicators to the user (Shani & Gunawardana, 2011).

2.3.3 User eXpereince evaluation

As indicated earlier in Chapter 1, recently researchers have acknowledged that embedding the RSs methods into UX impacts dramatically on the effectiveness of recommendations for the users (Bart P. Knijnenburg et al., 2012; Joseph A Konstan & John Riedl, 2012). In the early 1990’s, cognitive scientist, Don Norman coined the term UX. He indicated that UX is about the user’s feeling (positive and negative) about a product over time (D. Norman, 2013b). He elucidated that the term of UX was invented because there was a need for an umbrella term beyond the human interface and usability to cover all aspects of the user's experience with a system such as industrial design, graphics, the interface, the physical interaction, and the manual.

There is no a clear definition of UX (McCarthy & Wright, 2004) (Law et al., 2009). However, the current ISO (ISO 9241-110:2010 (clause 2.15) definition on UX focuses on a person's perception and the responses resulting from the use or anticipated use of a product, system, or service. Kraft (2012) emphasized that UX is not about creating the newest and cutting-edge technologies but a great UX around them. For example; Nokia, as a leader of UX, failed to create an acceptable UX for touch screens in the mobile phone industry but Apple managed to create an excellent UX for this old technology. Therefore, it has been said that today, UX is the key battleground for all kind of products in the consumer business market (Kraft, 2012). In software engineering, as a matter of fact, if a product fails to meet rising end users' needs, it makes both the product and the company (creator of product) obsoleted. UX is becoming the key competitive factors in more and more industries. Users are demanding products that are not only easy to use but also joyful and fun to use. Users will choose the products that put a smile on their face when using the product. In other words, users will choose products with a great UX (Bernhaupt, 2010).

In the field of RSs, UX is the delivery of the recommendations to the user and the interaction of the user with those recommendations. Indeed, UX necessarily includes algorithms, often extended from their original form, but these algorithms are now embedded in the context of a certain application (Bart P. Knijnenburg et al., 2012).

2.3.3.1 UX metrics

UX is dynamic and changes in the different circumstances by users' contexts and emotional states durin, before and after an interaction with a product (Vermeeren, 2010). Considering dynamic changes of user goals and needs, it is crucial to evaluate the RSs beyond the static aspects and investigate the temporal aspects of UX to know how and why experiences evolve over time(Bart P Knijnenburg & Willemsen, 2010). Moreover, users' values and expectations influence their experiences therefore it has to be considered from the beginning of the design process (T. Nguyen, 2016).

A thorough understanding of users' experiences, their positive or negative feelings, is at the core of UX evaluation (Kraft, 2012). However, one problem with feelings and indeed the UX is that different people react differently to different situations. And the same person may get different feelings in the same situation depending on the context. The goal is, of course, to maximize the positive moments for users when they're using the system. And ideally to make the users love the product, at least some or most of the time(Hassenzahl & Tractinsky, 2006).

A collection of UX definitions and evaluation methods along with the relevant references are accessible at *http://www.allaboutux.org/ux-definitions*. The review of all methods is beyond of the scope of this research and some also are not applicable for the RSs. In the following, only the methods that extent the understanding of UX evaluation and have been applied in RSs are discussed briefly.

One of the methods for long- term evaluation of UX is the UX Curve (Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, & Sinnelä, 2011) which aims at assisting users in retrospectively reporting how and why their experience with a product has changed over time. Figure 2.16 is an example of UX Curve which positive feelings mean that the UX curve goes up, negative feelings mean that the curve goes down. If the curve goes down too much or drops repeatedly during the process, the system will most likely lose the user, or the user may end up being pretty unhappy most of the time using the system. Another very important point is that one negative user experience may need dozen good experiences to make the user satisfied again and sometimes the user does not come back and even is not willing to use other services or products from the same company (Kraft, 2012).

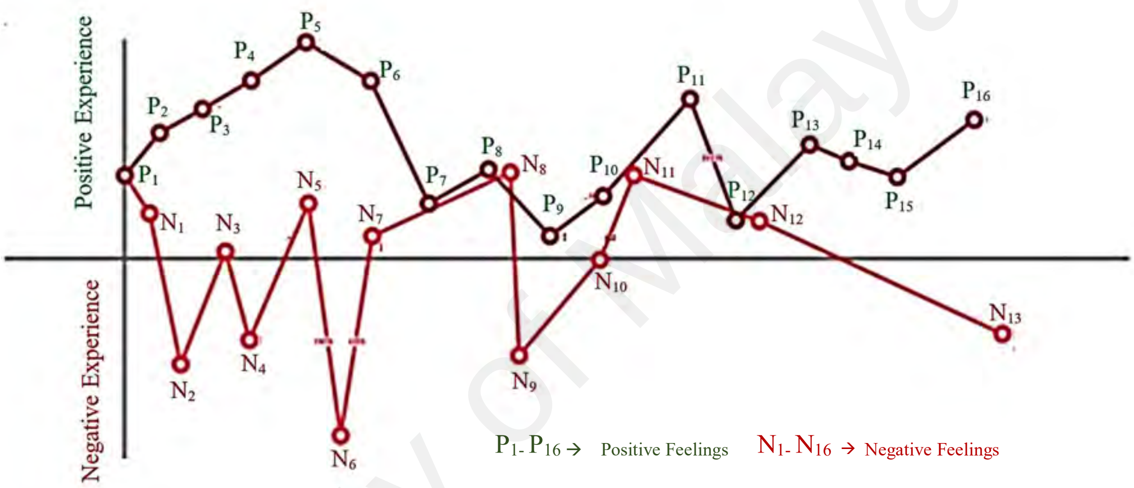


Figure 2.16: UX Curve

The UX curve helps to look beyond static aspects and to evaluate UX changes over time. And this is also the difference between usability tests and UX. Usability tests tend to focus on task performance whereas UX focuses on lived experiences(Hassenzahl et al., 2013). In the following two different perspectives in UX evaluation posted in the RSs literature are expressed briefly.

(a) Objective UX (Cognitive Load)

Objective UX measures the motivations behind the users' ratings (T. Nguyen, 2016). In other words, it is about the amount of the required memory being used by the working memory of the user to achieve his goal(Sweller, 1994).The less cognitive load is the more the positive feeling is discerned by the users while interacting with the system. Harper et al. (2005) investigated motivations behind user rating behaviours in an online movie RS. They constructed an empirical model to formalize their initial understanding of the system, and conducted survey to collect behavioral data. They found out that users perceive rating-time costs when they provided ratings. In another study, Sparling et al. (2011) the mental cost and benefits on different rating scales upon 12,847 movie and product review ratings collected from 348 users through an online survey have been investigated. Based on the Sparling et al.'s approach the rating time is applied to estimate cognitive load because user mental costs is difficult to be measured accurately (T. Nguyen, 2016).

(b) Subjective UX (Self-report)

Knijnenburg et al.'s (Bart P. Knijnenburg et al., 2012) proposed an evaluation framework for the UX of RSs. The framework aims to investigate the impact of objective system aspects such as the user interface on the users' subjective perceptions and experience. They provided a questionnaire survey of 7-likert-scale statements to measure the usefulness of the system (experience) influenced by the perceived difficulty of the system (subjective system aspect), and how each of these are affected differently by the three different interfaces, controlling for the self-reported expertise as a personal characteristic.

2.4 Systematic Literature Review on recommending approaches

The increasing number of papers on RSs is an ample evidence that in many disciplines applying contextual information has been a critical issue in the last decade (Adomavicius & Jannach, 2013) and identifying contextual information used in SRSs is effective for future studies in this field. Besides, the evaluation of the SRS' effectiveness in SRSs is not promising (Beel, Genzmehr, et al., 2013). Hence, the need to conduct two SLRs on the basis of the results from past studies exploited contexts for recommending and also reviewing the recommending evaluation methods have been identified as shown in Figure 2.17.

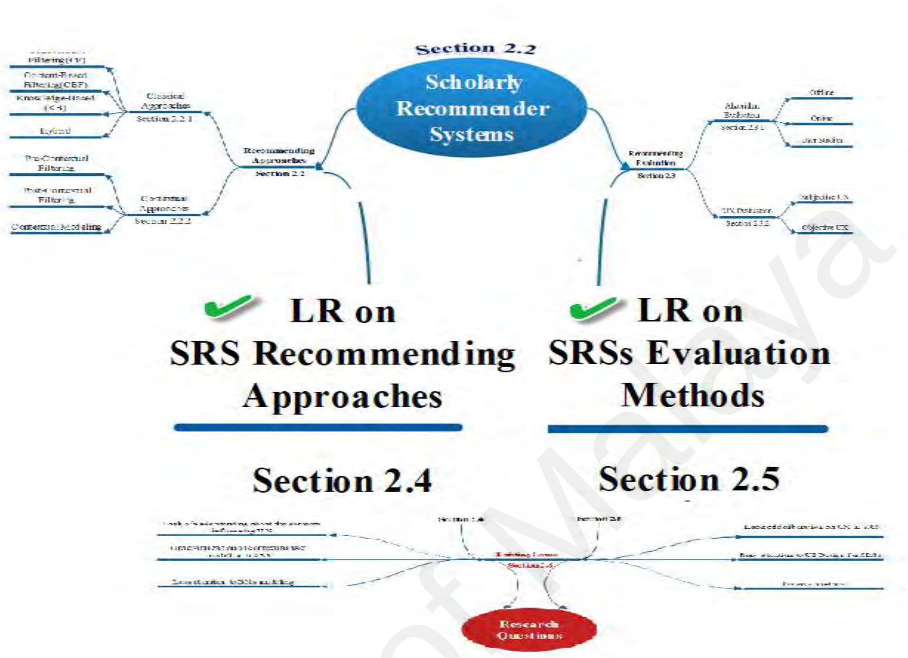


Figure 2.17: SLRs on recommending approaches & evaluation methods

The first SLR on SRSs studies was carried out in order to identify the: 1) recommending methods applied in SRSs, 2) contextual information incorporated to the recommending process, and 3) the ways that researchers have identified the contextual information. In the following section, a summary of the first SLR results is discussed.

2.4.1 Summary of results

This SLR work has been published in a journal and the article(Champiri et al., 2015) is appended in appendix A. Therefore, the following sub-section provides the summary of the SLR results.

2.4.1.1 Recommending approaches applied in SRSs

Many methods have been successfully applied to make recommendations. The majority of approaches during the years of 2000 to 2014 are typically classical approaches of CF and CBF, as shown in Figure 2.18 (The references are in the appendix A).

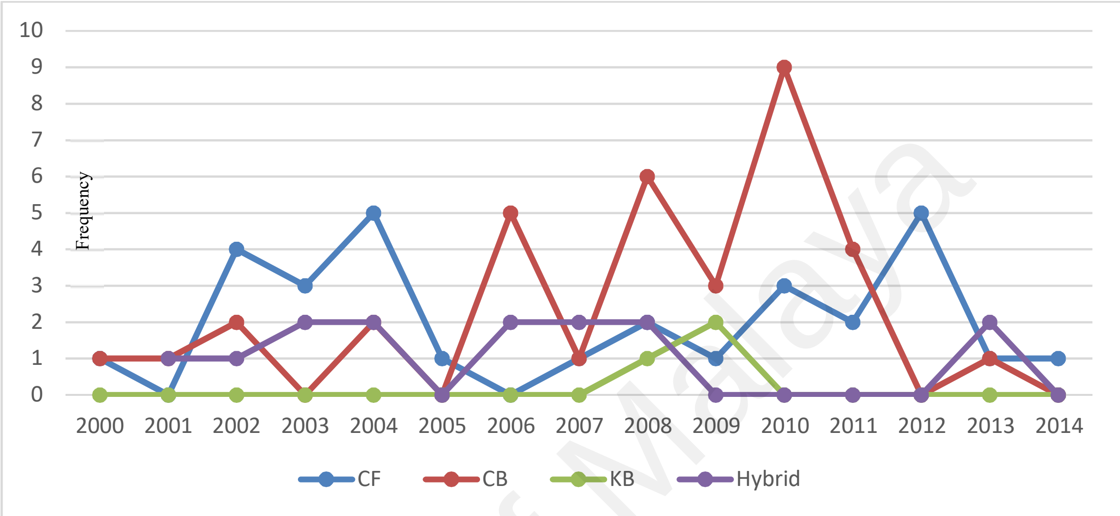


Figure 2.18: Distribution of recommending approaches

Incorporating contextual information into recommendations leads to greater data volume and a considerably more complex computation. Besides, it is even harder to estimate the important contextual factors that are relevant to users' interests (Yujie & Licai, 2010). Therefore, the majority of the existing studies have applied approaches of CF and CBF.

2.4.1.2 Contextual modeling in SRSs

Some researchers of RSs believe that regardless of the technology exploited by a RSs, the high quality recommendations can be produced only after modeling of the users preferences which is typically called User Model (UM) in the literature (Berkovsky et al., 2008; Kobsa, 2001). In this context, quality refers to the ability of the system to produce exactly those recommendations that the user will use or would like (Berkovsky et al., 2008). To achieve this, adequate information including contextual information should be stored to deliver high quality recommendations however acquisition of sufficient data for the UM, is not an easy task especially at the initial stages of interaction with the user, when usually little information about the user is available. Therefore, as a general rule, the more valid information is stored in the UM, i.e., the more knowledge the system has obtained about the user, the better the quality of the recommendations will be (Berkovsky et al., 2008).

As indicated earlier, there are little studies developing UMs in the field of SRSs while the main factor that influences users' satisfaction is the ability of a recommender to meet the users' information needs and it is obvious that the users have different information needs due to different knowledge, goals, and generally different contexts which are uncertain and change consistently (Beel, Gipp, Langer, & Breitinger, 2016). In contextual UM, the contextual information is employed directly as a main part of learning preference models (built using techniques such as decision tree, regression, and probabilistic model). To put it differently, contextual variables are added as dimensions in the recommendation function in addition to the user and item dimensions (Panniello et al., 2009). However, identification of valid contextual information for different domains are challenges in contextual user modelling either explicitly or implicitly (Yujie & Licai, 2010) (Kobsa, 2001). Moreover, a contextual UM should be able to infer possible cognitive process of user behavior and mind but it is not fully tractable in the practice (Berkovsky et al., 2008; Kobsa, 2001; J. Y. Wu & Wu, 2014). The past decade has seen research into the use of ML to support user modeling pass through a period of decline and then resurgence, however for creating the cognitive user models it is still at early stage(Martín et al., 2013; Papatheocharous et al., 2014).

2.4.1.3 The contexts incorporated into recommending in SRSs

The results taken from the SLR showed that, contextual information exploited for SRSs are categorized into three main groups including user, document, and environment contextual information. The users' contextual information implies the information explaining users' current situation such as task, information seeking. The information that characterizes the situation of a paper could be considered as document or paper context. CB approaches generally calculate the similarities between the content of documents for making recommendations. Each document has specific attributes that differ from others such as bibliographic information, citations and popularity. There are several attributes of documents (Pazzani & Billsus, 2007). For example, bibliographic information of a paper, including title, ISSN, abstracts, keywords are key factors in generation of ratings. Like bibliographic information, co-citations are semantic similarities ratings for documents that present the frequency of two documents cited together by other documents (Franke, Geyer-Schulz, & Neumann, 2008). Some studies proposed approaches where they compute the similarity of citations between scientific papers to recommend appropriate papers (Sean M McNee et al., 2002; Sean M McNee, Kapoor, & Konstan, 2006; J. Webster et al., 2004). Besides, the modelling of researcher's past works as well as papers that cite the work are effective to be used for the purpose of formulating scholarly papers recommendations. This model was implemented and tested by Sugiyama, K. and M.-Y. Kan (Sugiyama & Kan, 2010). Based on the results obtained from the users' feedback, they proved that filtering these sources of information has a significant impact on accuracy of recommendation. Environment contextual information presents a set of information to formalise the situation of users, especially when the situation of users is dynamic and changes frequently. Based on the results of SLRs, examples of the environment contextual information exploited for scholarly recommendations include location, time, service type and surrounding conditions. It seems that environment contextual information is mostly employed in mobile recommender systems, which are characterised by dynamic changes in the environment. The contextual information along with its conditions based on the analysis of the past studies are sorted and categorised in the paper attached to the Appendix A.

2.4.1.4 The methods of contextual information identification

Another question of the above mentioned review addressed the way in which researchers have understood the relevancy of contextual information that is expressed in below. The importance of this question is intensified when researchers on the field of context awareness are unanimous that context is an ill-defended concept (Bazire & Brézillon, 2005) (Adomavicius & Jannach, 2013). Besides, exploitation of all or too much contextual information might cause computational complexity and ambiguity in the system due to irrelevant, redundant, inconsistent, and noisy data (Baltrunas et al., 2012; J. Yuan et al., 2014; Yujie & Licai, 2010). Moreover, it is discussed later that each particular contextual information is considered as an extra dimension to the utility function of recommender. Hence, understanding of how researchers assess the relevant contextual information employed in SRSs, is crucial. The results showed that researchers in the field of SRSs, mostly relied on the past studies or have not discussed how they find out the relevant contextual information which they have exploited in the recommending process.

2.5 Systematic Literature Review on recommending evaluation methods

A separate SLR have been conducted to review comprehensively recommending evaluation methods applied in the field of SRSs. In the following sections, the SLR methodology, data acquisition, data analysis and results are elaborated respectively. To accomplish the SLR, the meta-analysis method has been applied. Systematic review is also a part of the meta-analysis method (Ferrier et al., 1995; Tenenhaus et al., 2004), for the systematic review, the Kitchenham's guidelines are inspired, which is a rigorous and well-defined guideline for reviewing the sources in the field of software engineering(Barbara Kitchenham, 2007). The process consists of the following steps and activities (Figure 2.19) which are discussed in the next sections.

Figure 2.19: SLR activities

2.5.1 Data Acquisition

The evaluation of the SRS' effectiveness in SRSs is not promising (Beel, Genzmehr, et al., 2013) and there is not consensus between researchers on the best evaluation method; hence, the need to review this area has been identified (DA1). Figure 2.20 depicts the research questions that the SLR aim to investigate and respond (DA2).

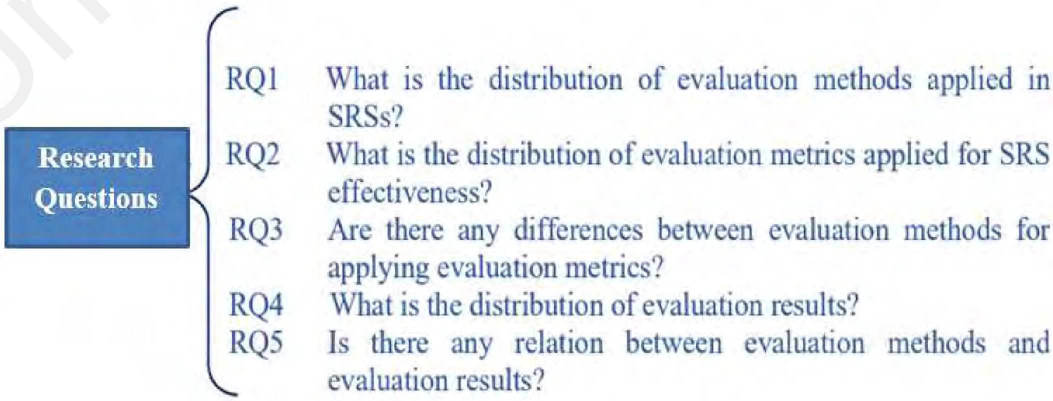


Figure 2.20: SLR research questions on recommending evaluation

Several terms have been used differently in publications to identify the area of CARSs(Baltrunas & Ricci, 2009). Besides, a few researchers had used contextual data such as user profile information for creating scholarly recommendations, but they did not indicate directly the terms “contextual” or “context aware” in their titles or abstracts. Likewise, different terms and synonyms have been applied for Scholarly RSs, including university and scientific systems (C. Porcel, Herrera-Viedma, Enrique, 2010). Thus, in order to find the maximum number of related papers, the searches have been carried out in 14 bibliographic databases in two steps of narrow and broad searches (DA3). In the first step as shown in Figure 2.21, the searches specified the retrieval of papers directly discussing CASRSs by using a “Boolean strategy” to locate title, abstract and keywords; a total of 12 papers were retrieved through this step. For those studies that did not have words that matched the related terms, a broader strategy as contained in the second-step search was used.

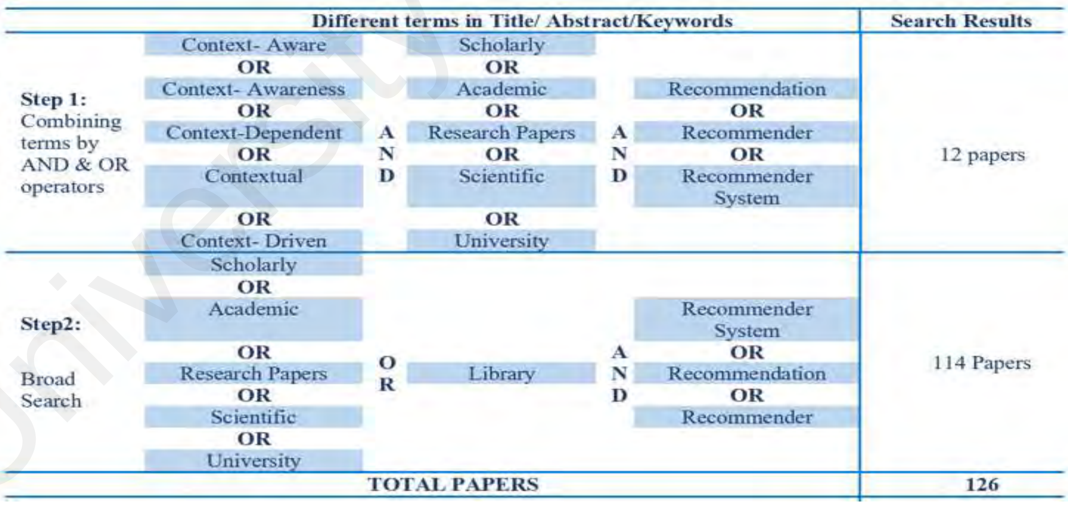


Figure 2.21: Search strategies

To retrieve the relevant studies because After conducting the second step search, an additional 114 papers that discussed RSs for recommending books and papers were identified; these papers were published from the years 2000 to 2014. All of them have been added to the database. Thus, the total number of papers retrieved through the two search steps is 126, all written in English. The EndNote software (Reuters, 2013) is used to store the retrieved papers because with this software it is easy to keep record of references (DA4). Figure 2.22 shows data acquisition and analysis process. Based on the exclusion and inclusion criteria (DA5), 67 papers were selected to be considered for purposes of final review and Meta-Analysis to provide answers to the research questions.

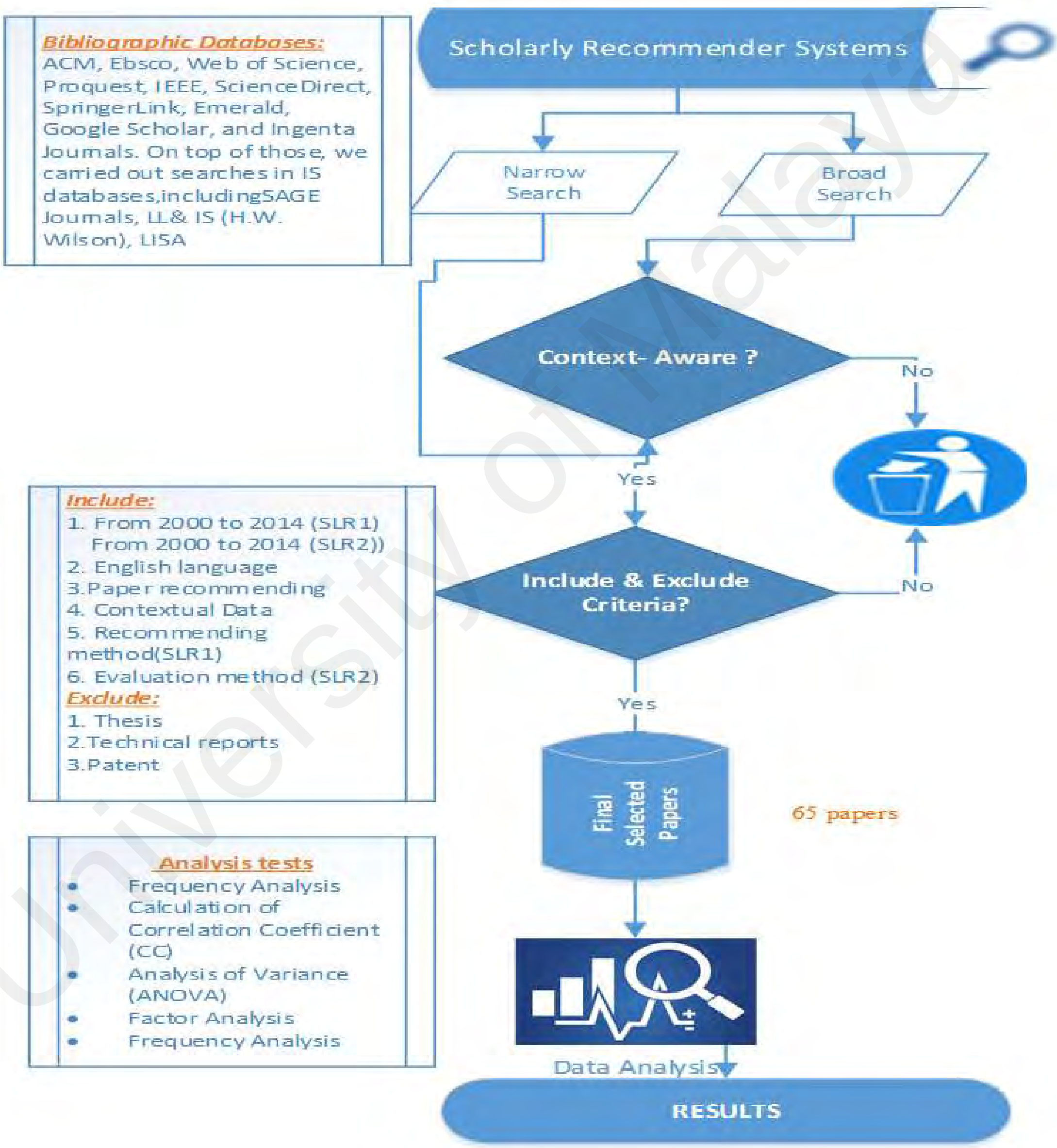


Figure 2.22: Data acquisition and analysis process

The date of January 2000 was fixed as the starting point for this review and November 2014 as the last date. Next, the papers were scrutinised for finding the values for variables such as recommending method, contextual information (DA6). Then a table was prepared when the rows are papers and each column is a determined variable, e.g. offline, CF, CB. The table is filled out by 0 and 1; for example if a paper used offline evaluation method, the value of 1 should be assigned for that in offline column (Figures 2.23).

	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE
1	EV-Methods																			EV-Metrics	
2						AI-based metrics							Text-based Metrics			Implicit user' prefrences Metrics					Ac
3	Offline	online	user - studies	Others-qualitative	Not-mentioned	precision (P@n)	Recall	F-Measure	Coverage	effective rank	NDCG	Hit-rate	Co-cited probability	CCIDF	TFIDF measure	distance	MRR	CTR (set)	LTR	CiTR	WS
4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
9	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2.23: A sample of data preparation

For the validity control of data collection, randomly 20 % of total papers for each SLR have been chosen and the second author manually reviewed them again. The review was based on each paper's abstract; introduction and conclusion, and the results were used to certify the accuracy of the final selected papers. Due to the implementation of the additional control, there was no new discovery of relevant papers that met the aforementioned criteria (DA7). An additional control was established so as to limit the maximum number of relevant papers. All the references of the selected papers were reviewed; the titles of the references were double-checked with the database. Due to the implementation of the additional control, there was no new discovery of relevant papers that met the aforementioned criteria. This served as a confirmation that within the scope of this study, the maximum number of relevant papers had been reviewed.

2.5.2 Data Analysis

The Frequency Analysis (FA), Correlation Coefficient (CC), Analysis of Variance (ANOVA), factor Analysis and Frequency Analysis to respond the identified questions. All of these analyses are conducted by IBM SPSS Statistics v.23. In order to ensure validity controls of data analysis tests, the five essential assumptions (Harman, 1960) in factor analysis were duly checked. Specific assumptions are made in carrying out the ANOVA test, which include those in the following list: 1) each sample is an independent sample; 2) the normal distribution (the Kolmogorov-Smirnov test is conducted to verify the normal distribution of the sample; 3) at the group level, the population variances are equal in responses(DA8).

2.5.3 Summary of results

Following are the key findings from the second SLR conducted on recommending evaluation methods in SRSs (DA9). In the rest of this section, the questions indicated earlier are responded separately.

2.5.3.1 Distribution of evaluation methods

As shown in Figure 2.24, 38 of the 67 papers used offline approach; 11 papers applied online approach and 28 papers applied user studies evaluations. A few studies (G. Geisler, McArthur, & Giersch, 2001), (De Giusti, Villarreal, Vosou, & Martínez, 2010) did not discuss the evaluation method, but it seems that they had used structured interviews in order to evaluate users' views (Torres, McNee, Abel, Konstan, & Riedl, 2004). Figure 2.25 depicts the trend of evaluation methods from the years of 2000 to 2014. It illustrates that between the years of 2009 to 2011, the offline method has been the most used method compare to other evaluation methods. The evaluation methods with the reference have been listed in Appendix B.

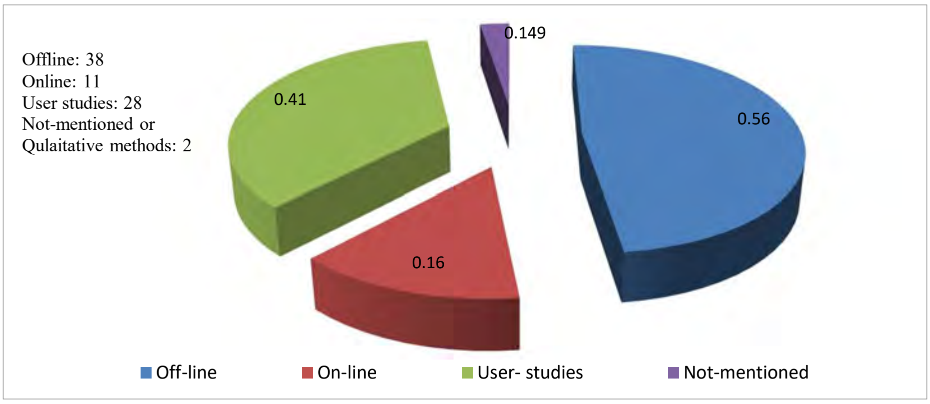


Figure 2.24: Distribution of recommending evaluation methods

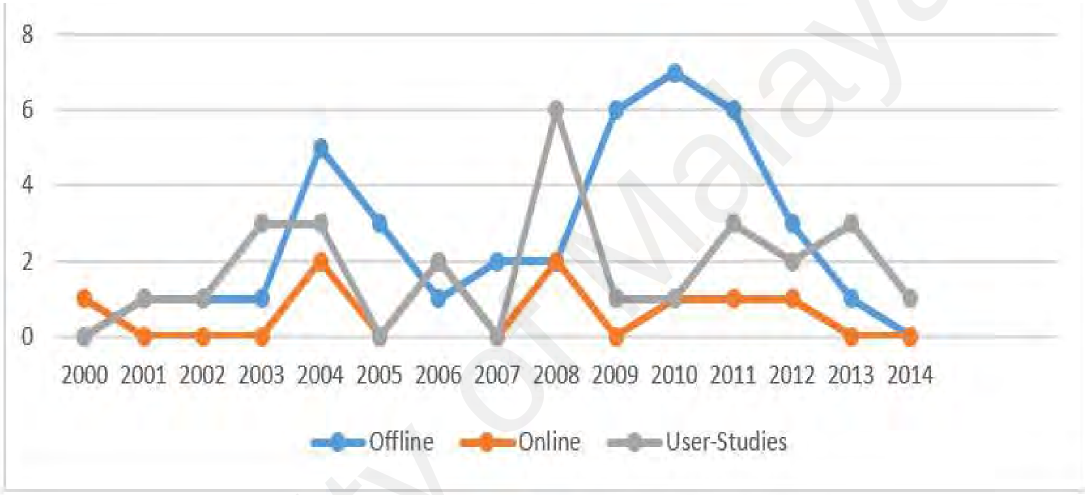


Figure 2.25: Trend of recommending evaluation methods

2.5.3.2 Distribution of evaluation metrics

A variety of metrics have been applied to measure the different dimensions of SRSs. Based on the results, 28 different metrics have been used listed in Appendix C. In this review, first the metrics are classify and then the metrics distribution is calculated. Through the principle component analysis, EFA is conducted to reduce the observed variables. The high membership degree indicates the strong membership of a metric under a factor (group). The maximum degree to place a metric under a proper factor is considered. The factors which do not have any maximum score of metrics are deleted. Then the final classifications of metrics under 8 groups are concluded (Figure 2.26). The grouping of metrics based on the maximum number of membership degree is as below:

Metrics	Factors (Groups)							
	1	2	3	4	5	6	7	8
precision	-.057	.801	-.387	.077	-.065	.126	.013	.086
Recall	-.077	.861	.048	.190	.168	.016	-.017	-.062
FMeasure	-.048	.619	-.273	.115	.202	-.076	-.053	-.149
Coverage	-.046	.237	-.195	-.059	.589	.275	.062	.128
effective	-.023	-.130	-.111	-.130	.583	.090	.024	.091
NDCG	-.044	.188	.851	.201	.094	.128	-.018	.001
Hitrate	-.047	.161	.663	.174	.131	-.130	.013	.055
Cocitedpro	-.029	.203	.665	.190	.146	.039	.029	.121
CCIDF	-.026	-.290	-.193	.907	-.033	.197	.000	-.008
TFIDF	-.026	-.290	-.193	.907	-.033	.197	.000	-.008
semanticd	-.017	-.078	.022	-.059	-.134	-.005	.600	-.146
MRR	-.022	-.006	.414	.057	-.039	.166	-.059	-.123
CTRset	-.067	-.344	-.079	-.274	.386	.542	-.037	-.038
LTR	-.032	-.078	-.092	-.089	.068	.195	.183	.652
CiTR	-.021	-.159	.000	-.156	.101	.436	-.228	-.589
WSM	-.017	-.078	.022	-.059	-.134	-.005	.600	-.146
MAE	.443	.442	-.187	.003	-.363	.224	-.031	-.043
RMS	.011	.195	-.127	-.036	-.430	.295	-.010	.052
Quetionares	.291	-.068	-.094	-.004	.152	-.612	-.107	-.205
Reliability	.997	-.011	.017	.012	.034	.017	.006	.014
Accessibility	.997	-.011	.017	.012	.034	.017	.006	.014
Assistance	.997	-.011	.017	.012	.034	.017	.006	.014
Usability	.997	-.011	.017	.012	.034	.017	.006	.014
Applicability	.997	-.011	.017	.012	.034	.017	.006	.014
Performance	.997	-.011	.017	.012	.034	.017	.006	.014
Feasibility	.997	-.011	.017	.012	.034	.017	.006	.014
Labstudies	-.099	-.126	-.310	.274	.356	-.421	-.036	.045
statistical	-.040	-.189	.053	-.146	-.348	-.014	-.445	.300

Figure 2.26: Metric classification groups based on membership degree

Based on Factor Analysis results, Metrics were classified into eight groups. However, because of the diversity in metrics used for SRSs' evaluations, a general concept or purpose to classify eight groups under them was not founded. For example, metrics in Group 6 including Precision, Recall and F-Measure can be classified under the label of "accuracy" metrics or CTR (Click-Through Rate), CiTR (Cite-Through Rate) can be considered under "users' feedback" metrics but general labels do not occur for all groups. Considering the above-mentioned classification and the results taken from this study, evaluation metrics used for SRSs have been listed and categorised into eight categories as shown in Figure 2.27. Based on the results, lab studies (user's behaviour observation), precision, and recall are the most commonly used metrics for evaluation of SRSs (Figure 2.28).

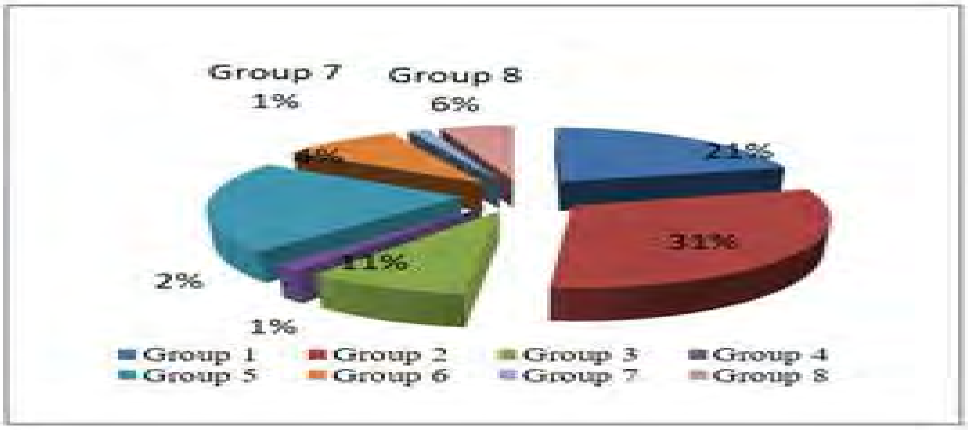


Figure 2.27: Distribution of Evaluation metrics used for SRSs

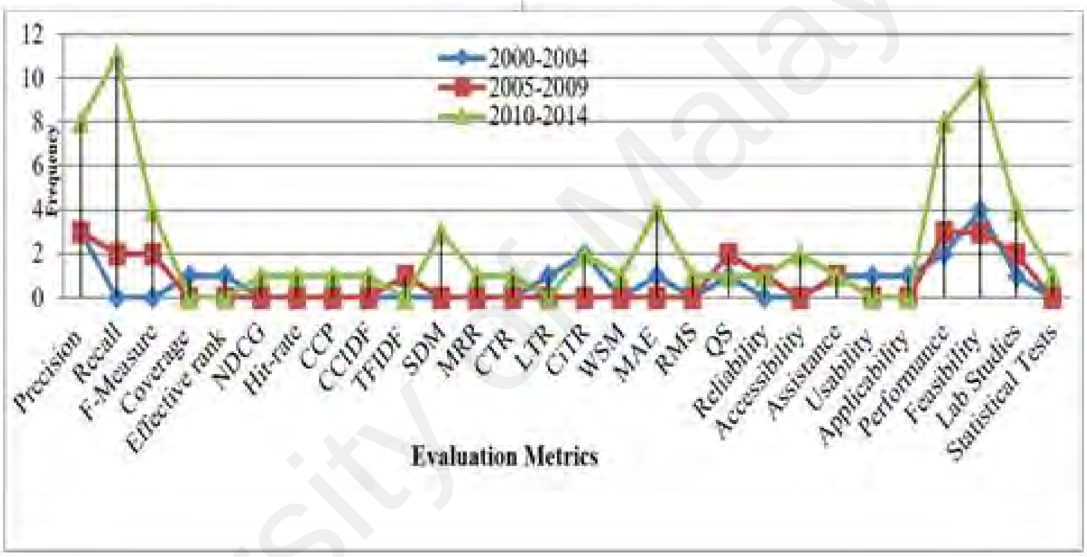


Figure 2.28: Trend of detected evaluation metrics

The results of frequency analysis of evaluation metrics show the recent (2010-2014) evaluation metrics in the following order of popularity: Precision, Recall, Questionnaire studies, Lab Studies, Mean Absolute Error (MAR) (Figure. 2.27). The more details of the above mentioned metrics are provided by (Parra & Sahebi, 2013), (Said, 2013; Schröder et al., 2011).

As the SLRs have conducted the studies until 2014, the next section reviews the new studies until the end of this research.

2.5.4 Review of the recent related work

During the years of 2014 to 2018, there have been a few researches in the field of SRSs. Wesley-Smith & West have applied citation analysis in order to identify papers that are similar to an input paper, they used the databases such as CiteSeerX3(Wesley-Smith & West, 2016). They used TF-IDF and LSA methods to discover groups of words that are equivalent in their meaning. Shahin Mohammadi (2016) has proposed a new method to integrate structural and contextual information for building a context specific network for similar PubMed articles (Shahin Mohammadi, 2016). Beel et al (2016) has been indicated that CF methods in SRSs are not effective because there is not balance between the number of papers and number of users. In other words, a huge number of papers compared with the number of users, and only few users rated the similar papers(Beel, Gipp, et al., 2016). Therefore, they suggested that user mind modeling is more effective than CFs.

The use of deep neural networks for Natural Language Processing (NLP) has recently received much attention; it provides high quality semantic word representations. Deep neural network models have been applied to tasks ranging from machine translation to question answering, but not much attention is paid to the RSs area. For instance, in (Mueller & Thyagarajan, 2016), the authors showed that LSTM can be used to build a language model and assess semantic similarity between sentences. These models are usually trained on large amounts of data. To the best of the research's knowledge, there have been no work done before for recommending scientific articles based on contextual Bayesian networks (Hassan, 2017). Also, the use of machine learning algorithms in RSs has been reviewed and analyzed by (Portugal, 2015). According to this survey, among seven studies that used Bayesian method (Table 2.3), only two studies are related to book and document recommender(Ericson & Pallickara, 2013) (Lucas, Segrera, & Moreno, 2012) which they are not considered contextual information.

Table 2.3: BN in RSs (Portugal, Alencar, & Cowan, 2015)

Domain	*Reference*
Documents	(Ericson & Pallickara, 2013)
E-shop	(Felden & Chamoni, 2007)
E-mail	(Gorodetsky, Samoylov, & Serebryakov, 2010)
Books	(Lucas et al., 2012)
Movie	(Marović, Mihoković, Mikša, Pribil, & Tus, 2011)
Traffic	(Šerić, Jukić, & Braović, 2013)
Tourism	(Y. Wang, Chan, & Ngai, 2012)

The issues derived from the SLRs results and review of recent are discussed in the following section.

2.6 Discussion: Existing issues

As shown in Figure 2.29, a few crucial issues that built the fundamental of this thesis to formulate the research questions are discussed here. The mapping of the research issues and objectives of this research are described in Chapter 3.

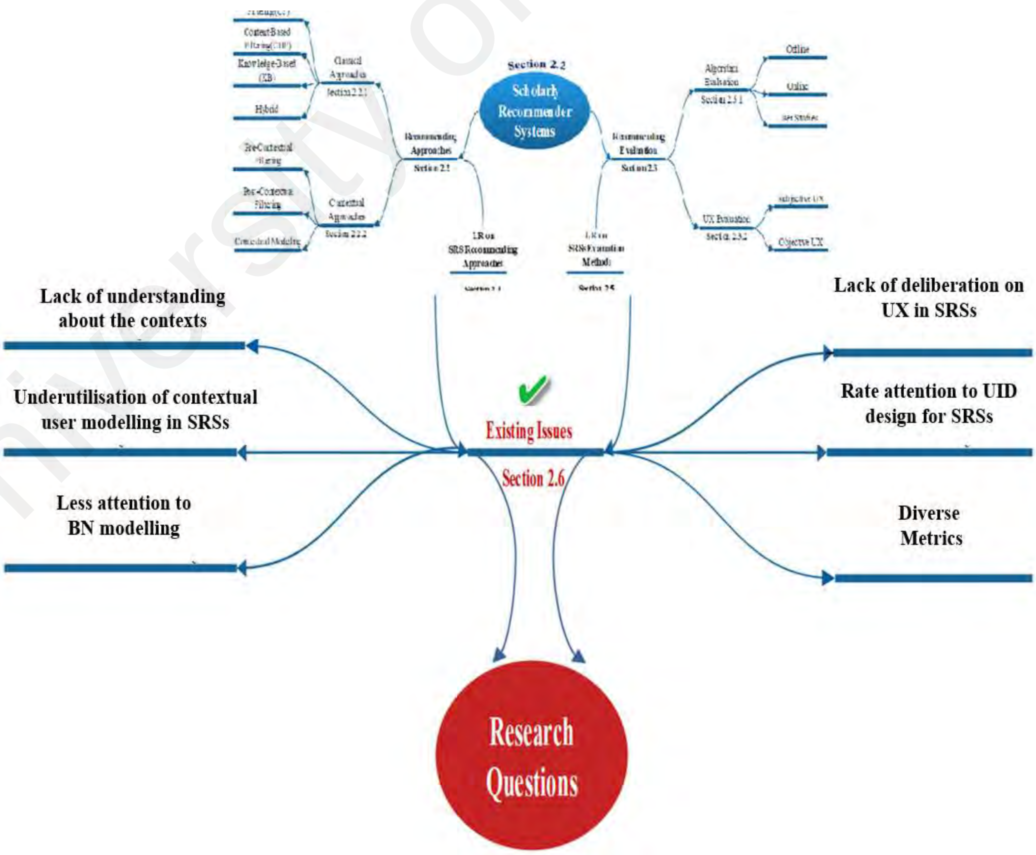


Figure 2.29: Existing issues in SRSs

2.6.1 Lack of understanding about the contexts influencing UX

Despite the fact that there is a strong relationship between context and UX(Bart P. Knijnenburg et al., 2012), and accentuation of the RS researchers on using UX in the recommendation process, there is a lack of understanding about the contexts influencing UX in the RS literature (Adomavicius & Tuzhilin, 2011b; Bart P. Knijnenburg et al., 2012). To make inferences about the users' experience, it is required to move beyond measuring their behavior and measure users' subjective valuations as well. Moreover, as users' interaction with RSs is highly context-dependent, personal and situational characteristics also need to be taken into account. Unfortunately, even studies that consider aspects other than accuracy look at a limited set of variables that influence each other (e.g., how satisfaction changes due to diversification, or how choices become more accurate with the inclusion of a recommender engine) without integrating these variables into a model of overall UX (McNee et al., 2006) (Pu & Chen, 010; Pu et al., 2012). However no study have discussed in detail how contexts impact UX. One of the reasons behind that might be the subjectivity of UX, which makes it extremely difficult to measure without explicitly asking a user how good a recommendation is in their eyes in a long period(Tintarev & Masthoff, 2007). Rather than the subjectivity of UX, following problems also lead to the aforementioned issue.

2.6.1.1 Indeterminate contexts

The results of the SLR on CASRSs showed that contextual information can be categorized into three groups of user, document or paper (system) and environment contexts however the contexts are not indeterminate completely. According to Dey's (2010) illustration, context can be anything that characterizes the situation of an entity which helps better interaction between the user and the system. In SRSs, if a paper is considered as an entity so the context would be anything that characterizes the situation of a paper but before detecting the relevant contexts to describe the situation of an entity, the more important question is that what is the situation of a paper? And how a paper's situation (e.g. paper's accuracy) makes a better interaction between the user and system? Many recent works have raised the issue that beyond accuracy other characteristic such as popularity, diversity and novelty of a paper also influence the quality of paper recommendations (Ge, Delgado-Battenfeld, & Jannach, 2010b) (Adamopoulos & Tuzhilin, 2011; Ge et al., 2010a; Sean M McNee et al., 2002) This initiative has opened up a new perspective regarding evaluating and improving recommendation techniques but some challenges are still to be faced. For example, diversity of recommendations has been mentioned in only in a few studies. Vellino et al (2010) measured diversity as the number of different journals from which articles were recommended (André Vellino, 2010a) meaning that if recommendations were all from the same journals, diversity was zero. Despite this fact that novelty introduced one of the ways that improves users' satisfaction in RSs (Adamopoulos & Tuzhilin, 2011; Ge et al., 2010a; Sean M McNee et al., 2002), The SLR results show that SRSs researchers have not taken into account providing novel recommendations. Past studies relied on users' rating to find out the preferences for diversity, popularity, novelty and accuracy while the users' ratings cannot easily show the users' preferences (Beel & Dinesh, 2017). The above explanation demonstrates that the concept of context in SRSs is a crude concept and creating recommendations for users in an academic domain to cater to their needs and tasks needs more analysis of contextual information affecting decision-making in this domain.

2.6.1.2 Difficulties on detection of relevant contexts

It is difficult to understand and exploit the relevant contexts which influence UX for all applications (Adomavicius et al., 2005) due to the various situational parameters that might influence users 'decisions in an intuitive way (Hariri, Mobasher, & Burke, 2014). Besides, contribution of irrelevant contextual information in the process of recommending might leads to false reasoning models and worse recommendations so that users experience negative feeling with the system and consequently lose their trust and stop using the system (Baltrunas et al., 2010)(Panniello, Gorgoglione, & Tuzhilin, 2015) (Adomavicius & Jannach, 2013).

2.6.1.3 Method unanimity in assessment of relevant contexts

As mentioned before, using contextual information has been considered as the main factor for creating better recommendations and enhancing UX (Adomavicius & Tuzhilin, 2011a; Baltrunas, 2008). Researchers emphasize applying contextual approaches in order to recommend indicators to users based on certain circumstances (Baltrunas & Ricci, 2009; Kaminskas & Ricci, 2011). However, the variety of application scenarios and user requirements cause difficulties in presenting a unanimous method for detection of contextual information for all RSs. The review of past studies accomplished on SRSs have been captured the relevancy of contextual information on recommendation through five ways shown in Appendix C. The majority of studies employed contextual information based on the past studies were used these contextual information while few studies performed separate investigation such as interview with users to explore the relevant contextual information from the users' point of views. In addition, the prediction of relevant contexts in the recent studies mostly is assessed by using statistical methods. The studies and detection methods along with the references have been summarized in Table 2.3.

Table 2.4: Methods of relevant contexts detection in RSs

Domain	*method*	*Relevant detected contexts*	*Reference*
Movie	t-test	Movie genre, Time of week, Accompany type	(Adomavicius et al., 2005)
Travel	χ2 test	Travel accompany with type(Sole, with accompany)	(Liu, Lecue, Mehandjiev, & Xu, 2010)
Tourist	t-test	Distance, Season, Temperature, Day time, Distance, Mood, Travel goal, Mood, Distance, Mood, Time available	(Baltrunas et al., 2012)
Restaurant	A simple feature selection approach	Service model :23 attributes such as latitude, longitude, address, city ,state ,country, days, hours User model : 21 attributes such as interests, personality, religion, occupation, budget Environment model: 2 attributes of time, weather	(Vargas-Govea, González-Serna, & Ponce-Medellın, 2011)
Movie	Pearson's χ2 test Freeman–Halton test	Time, location, social, end emotion, dominant emotion, mood interaction	(Odic, Tkalcic, Tasic, & Košir, 2012)
Movie	User's opinions	Day type, location, end emotion, dominant emotion, Mood, physical, decision, interactiont ime, location, social, end emotion, dominant emotion, mood interaction	(Odić, Tkalčič, Tasič, & Košir, 2013)

2.6.2 Under-utilization of contextual user modeling

As concluded, the most applied SRSs recommending approaches are CF and CBF. This is surprising that there is little interest (about 20%) in user modeling specially based on contextual information in SRSs while user modeling is one of the most important parts of a RS and Amazon, Google's business model and Netflix are heavily dependent on user modeling. User modeling analyzes the users' indicators or actions and consequently infers information (Beel, Breitinger, et al., 2016) which is one of the main differences between a RS and search engine as well (Berkovsky et al., 2008; Kobsa, 2001). The aspect called "concept drift" in user modeling meaning that the automatic information inferring should detect the current relevant indicators (meaningful data) for the user-modeling process. However in SRSs researches, concept drift is widely ignored. Besides, user model development requires to gain access to a stream of user actions which can be gained through the enhancements of user interface by establishment a link between user actions and system events (Bart P. Knijnenburg et al., 2012) while in SRSs studies this matter has not taken much attention. The user-model size is another important aspect about the user modeling. While in search, user models (i.e. search queries) typically consist of a few words, user models in RSs may consist of hundreds or even thousands of words. Of the reviewed approaches, 91% did not report the user-model size, which leads us to the assumption that they simply used all features. Those few that reported on the user-model size usually stored fewer than 100 terms. For instance, Giles et al. utilized the top 20 words of the papers (Beel, Breitinger, et al., 2016). The following problems lead to the aforementioned issue.

2.6.2.1 Less attention to the users' information needs

The meeting of user's information need is the main contribution of a good scholarly recommender. Users have different information needs due to different knowledge, preferences and goals, and contexts. One user might look for novel papers in a particular area while another user might be interested in the most popular papers. Indicators that meet the information needs are "relevant" to the user (Beel, Breitinger, et al., 2016). The more a SRS meet the users' information needs the better is the SRS. So far, this task is reflected by measuring the accuracy of recommender: the more relevant, and the less irrelevant indicators it recommends, the more accurate it is. The problem is that identification of users' information is not an easy and needs better understanding about the users' information seeking behavior and whether the methods such as log analysis are adequate enough to recognize users' information needs.

2.6.2.2 Less attention to BNs modeling

Based on the SLR on recommending methods applying for SRSs by (Champiri et al., 2015) and review of recent studies by (Hassan, 2017) (Beel, Breitinger, et al., 2016), it is concluded that, different ML methods such as Neural Networks, SVM, Decision Trees have been utilized in making paper recommendations considering the CF and CBF approaches. Despite that UMs play a critical role on recommendation quality and identification of the users' needs, they are rarely have been used in SRSs researches(Beel, Gipp, et al., 2016). Quality of recommendations refers to the capability of the system to predict exactly those items or services that make the user would like or use, overall to provide good experience for the users (Berkovsky et al., 2008; Kobsa, 2001). However; information needs are uncertain and vary among users due to different contexts such as background knowledge, preferences and goals (Beel, Gipp, et al., 2016). Hence, it is required to select a method which infers dynamic context and surpass uncertainty of them. Among the ML methods BNs are powerful tools used for uncertainty modeling (Pearl, 1985) based on the probabilistic theory of Bayes' theorem which spreads knowledge within the network (Heckerman et al., 1995; Neapolitan, 2004) and reason complicated problems. However, As mentioned in Chapter 1, contextual BNs have been rarely applied for recommending scientific articles (Hassan, 2017).

2.6.2.3 Lack of real databases

Once contexts are exploited in recommenders especially for ML methods, it is important that there is a dataset containing relevant parameters. For example; if the algorithm is using users' information needs, it requires data of users' behaviors which reflects users' information needs. If the required data is not available, the researcher have to collect the data before applying the method. In some cases, the process of dataset preparation is costly and time consuming therefore the researcher prefer to exploit the existing datasets and not get involve in preparing a new dataset. Indeed, the preparation of new dataset is difficult and sometime seems to be impossible since it needs serious dedication of users but the problem is that using contextual information in the field of SRS needs the relevant real datasets and it cannot be tested by datasets in other domains which is mismatch with the users' needs later on in the real world (Beel, Gipp, et al., 2016).

2.6.3 Rare attention to UI

There is a similar lack of attention concerning the content of what and how SRS should present the recommendations to the users. Indeed, only a few studies have developed UI and tested the UI usability, (Ozok et al., 2010) (Hiesel, Wörndl, Braunhofer, & Herzog, 2016). Recommendation agents are almost never stand-alone applications but are usually one of several components of an e-commerce website. SRSs are sometimes part of other systems such as digital libraries or bibliographic databases and one of the reasons little attention has been given to UI in this area. However, it does not mean that UI must be ignored in SRSs(Calero Valdez, Ziefle, & Verbert, 2016) (di Sciascio, 2017; Pu et al., 2012a).UI is important because it is the way users interact with the system. No matter how accurate the algorithms work, if the UI is not well designed and evaluated, it will degrade the interaction between the user and system. For example, the study of Middleton et al. (Middleton et al., 2004), the recommendations presented through website received much more click-through rate than the similar recommendations delivered to users via email. As mentioned before, among the reviewed SRSs, the majority of recommendations are delivered through other websites (like library) and only (Beel, Gipp, Langer, & Genzmehr, 2011) and Mendeley (Zaugg, West, Tateishi, & Randall, 2011)provide recommendations via a desktop software however it is still unknown which method surpasses the others(Beel, Genzmehr, et al., 2013).

It shows that there are not many studies that focus on user interface design guidelines. Hence, there is a need for guidelines that can potentially help designers to create more effective and satisfying UIs for SRSs. A well-design UI influences the users' perceptions, for example a user might better perceive a higher degree of diversity and novelty, if in the interface there is transparency or explanation about the indicators (Ge, 2010). Following issues explain more the above mentioned conclusion.

2.6.3.1 Lack of deliberation on UX of SRS

As mentioned earlier, part of the users' experience is formed by interacting the UI. However as researchers have indicated user centric evaluation of RSs which is beyond the accurate prediction and needs UI, has not been the main focus of RSs researchers(Joseph A Konstan & John Riedl, 2012) (Bart P. Knijnenburg et al., 2012). This is actually a paradigm shift in RSs research field since before that all the researches have been trying to develop more accurate algorithms algorithms (McNee et al., 2006; Cosley et al., 2003; Murray & Haubl, 2008; Murray & Haubl, 2009; Ozok et al., 2010; Pu et al.,2012; Konstan & Riedl, 2012). As discussed before, the results of SLRs on exiting studies have drawn this conclusion that UX has rarely received attention in the field of SRSs consequently the UI desing of the SRSs have not been the main focus of researchers in this filed (Beel, Gipp, et al., 2016). Although there are many published articles, it seems that SRS studies rarely have been implemented completely and used by the end users.

2.6.3.2 Diverse and un-reproducible metrics

Past studies on SRSs have considered 28 diverse metrics in order to evaluate the algorithms effectiveness, which means that researchers are not unanimous in determining attributes for effectiveness SRSs. However, metrics such as Precision, Recall and F-Measure have been widely used to evaluate the accuracy of recommendations. Also, a few researchers believe the variety of algorithms and variations in the implementations make it difficult to reproduce the results due to the absence of standards to guide RSs researchers as to how they should document their proposed algorithms and evaluation methods. The problem of limited reproducibility in RSs has been highlighted recently

(Beel, Breitinger, et al., 2016; Ekstrand, 2014; Konstan, 2004). Beel, et al (2016) criticizes that for 67% of the CBF approaches no information was given on the fields the terms were extracted from (e.g. title or abstract). Which might cause problems in replicating evaluations, and reproducing research results(Beel, Gipp, et al., 2016). In addition, researchers do not normally provide clear definitions of metrics and reason or justification for their choice of a specific metric and the crux of the matter is that some of the mentioned metrics such as performance, applicability, usability can be defined differently depending on the evaluator's viewpoint (Powers, 2007; Schröder et al., 2011). Based on a few significant guidelines done by Gunawardana & Shani, 2009; Murakami et al., 2007; Parra & Sahebi, 2013; Schröder, Thiele, & Lehner, 2011), the first step of the evaluation process is to state its goal with a clear definition.

2.7 Summary

This chapter started with the brief overview of SRSs, recommending approaches, context, UX, evaluation methods and metrics and continued by reviewing existing related work on contextual recommending and evaluation methods to provide a comprehensive and critical overview of available SRSs by conducting two SLRs and also reviewing the recent works on the files of SRSs. Also, this chapter aimed to explain and highlight the way researchers understood and assessed relevant contextual information in the recommending process in order to provide better recommendations for the users and enhance UX of SRSs. This review chapter ended with a discussion and critical analysis of the open issues of the existing SRSs and the research work in this thesis that aims at addressing those issues. In a nutshell, lack of understanding about the contexts influencing UX, difficulties on detection of relevant contexts, under-utilization of contextual user modeling, less attention to BNs modeling, lack of real databases and rare attention to UI are the issues of existing studies which formulate the foundation of research questions of this research.

CHAPTER 3: RESEARCH METHODOLOGY

In this chapter, the research methodology is described and the activities that have been performed for this research are listed. As discussed in Chapter 1, this research mainly adopts the Design Science Research Methodology (DSRM) process; however, the activities are conducted by using Empirical Methods (EMs) to mainly acquire and analyse the quantitative data. In addition, the description of the selected approaches and methods used to achieve the defined objectives is provided. It also describes the overall operational framework for conducting the research and tools utilised to carry out the research. The data acquisition and analysis methods for each process, along with the selection justification, are discussed in separate sections. Finally, the mapping of research objectives, research methods and deliverables is provided.

3.1 Introduction to Research Methodology

Prior to discussing the research methodology, clarification of the terms 'methodology' and 'method' must be essentially provided. Runeson et al. (2009) argue that the term 'methodology' refers to the principles and procedures of orderly through, or certain processes applied to a specific discipline of science (Runeson & Höst, 2009). Collis & Hussey (2003) described methodology as the overall approach of the research process from theoretical underpinning to the processes of data collection and analysis. Methodology provides a starting point for selecting suitable make-up of theories, concepts, ideas and definitions of the topics (Collis & Hussey, 2003). Considering this, all types of research follow a distinct methodology varying from study to study. In contrast, 'method' refers to the specific techniques, tools and means by which data are collected and analysed (Runeson and Skitmore, 1999; Hussey and Hussey, 1997).

Many suggestions are proposed to conduct an appropriate process in DSRM. Peffers et al. (2007) reviewed and evaluated the processes of conducting design science research in software engineering and information systems, and concluded that some studies have presented design research processes (Vaishnavi & Kuechler, 2004), some have not proposed a process (Cole, Purao, Rossi, & Sein, 2005), and some have proposed research frameworks that do not clearly state a process (Von Alan, March, Park, & Ram, 2004). However, in the prominent work of Hevner (Hevner, 2007), a set of guidelines was proposed in light of structuring the appropriate process or method in DSRM. The guidelines are summarised in Table 3.1.

Table 3.1: Design Science guidelines (Hevner, 2007)

No	*Guidelines*	*Descriptions*
G1	Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
G2	Problem Relevance	The objective of design-science research is to develop technology based solutions to important and relevant business problems.
G3	Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
G4	Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
G5	Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
G6	Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
G7	Communication of Research	Design-science research must be presented effectively both to technology oriented as well as management-oriented audiences.

Based on the research process proposed by (Hevner, 2007), the second and fifth guidelines, which are highlighted more than the other existing guidelines, focus on the design and action, and consist of three main phases of problem and solution identification, development and evaluation. Each phase is divided into steps that work iteratively.

Considering the above-mentioned explanation, this study utilises DSRM by (Hevner, 2007) and (Peffers et al., 2007) to guide the whole research process in order to fulfil the defined research objectives and address the hypotheses (Research Questions). In each process of answering a specific research problem, if any experiment is required, which involves conducting data acquisition and analysis, the guidelines of Empirical Methods (EMs) by (Easterly & Levine, 2001) are applied.

In fact, the roots of DSRM are in the engineering and the sciences of the artificial, and primarily are considered as a problem-solving paradigm (Peffers et al., 2007). This paradigm attempts to create innovative artefacts and solutions, define ideas, practices, and design an acquisition method of knowledge. This methodology uses a systematic approach to develop and evaluate an artefact to solve a particular problem. The results can be theoretical, practical, or both, based on the problem targeted in the research. Therefore, DSRM is used in this research with emphasis on the extensive review of the literature to identify the problem (Phase 1), develop an artefact (UM and UI in this research: Phase 2), demonstrate its use, evaluate and communicate the findings with the researchers and relevant audiences (Phase 3) (Hevner, 2007).

In the present study, as mentioned, for each process based on DSRM, some empirical activities have been done. EM is a research method by using empirical observations to collect and analyse the data (Easterly & Levine, 2001). Figure 3.1 depicts the combination of DSRM (X-axis) and a set of data acquisition and analysis activities carried out based on EM (Y-axis). Accordingly, each of the design artefacts (UM and UI) is be used to further discuss the following process.

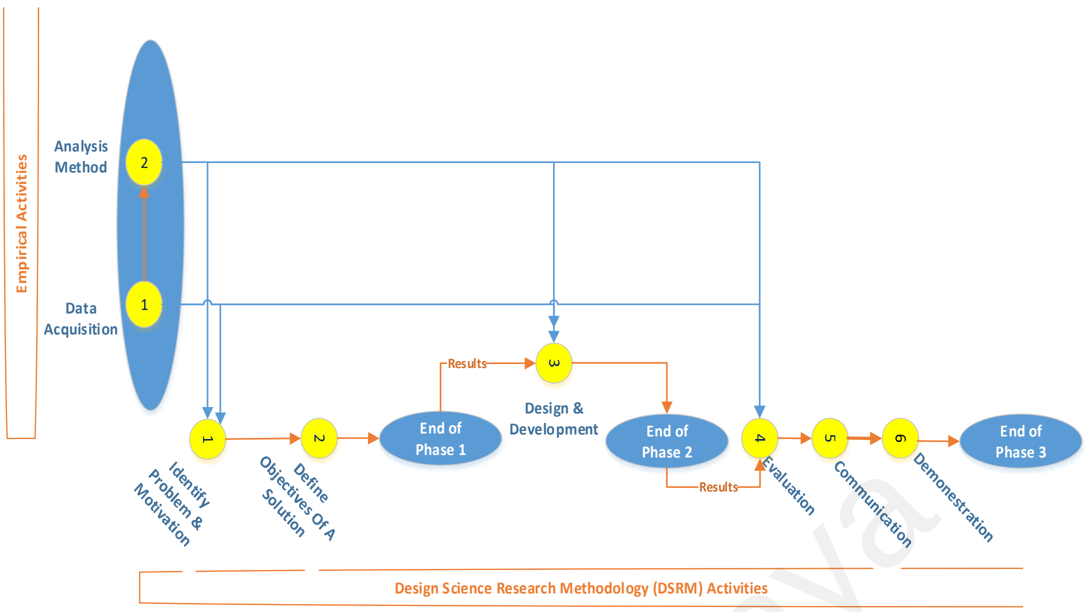
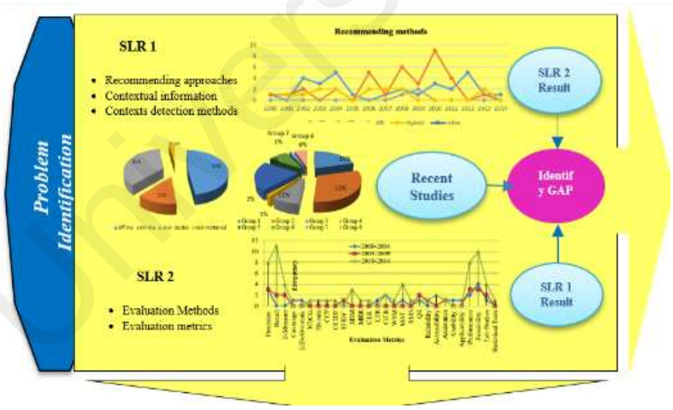


Figure 3.1: Combination of DSRM & EM

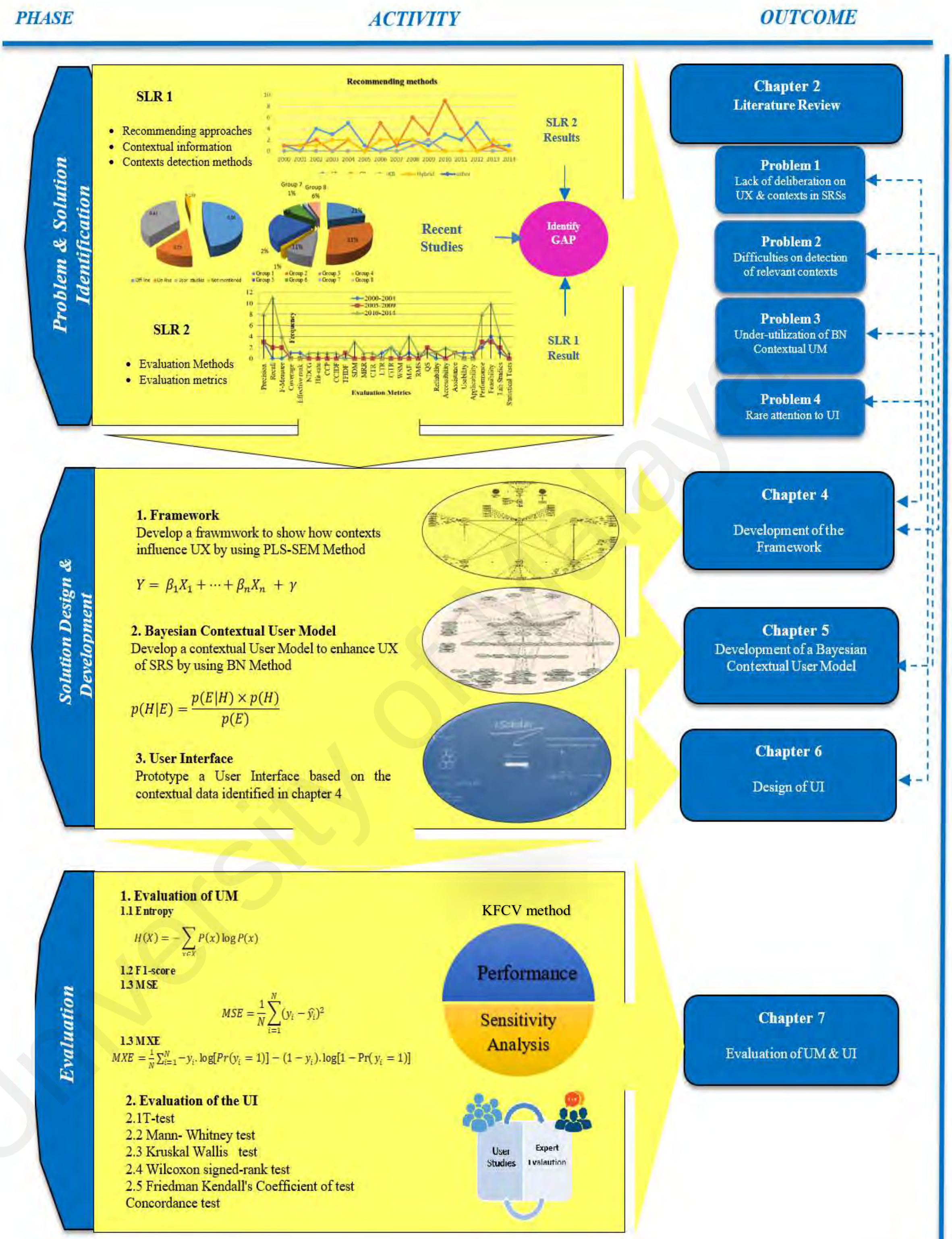
3.2 Research methodology (DSRM) process

Considering Hevner's guidelines (Hevner, 2007), this research is conducted in three phases as shown in Figure 3.2. In the following section, the three phases and activities done to address the research questions for each phase are elaborated.



3.2.1 Phase 1- Problem & Solution Identification

This phase is performed through two activities of identification of the problems and finding the solution. The identification and establishment of a specific research problem is the preliminary activity which needs the justifiable value of the solution to the problem. This activity seeks to motivate the researcher and audiences to strive for a solution, agree on the results, and to understand the reasoning associated with the problem.



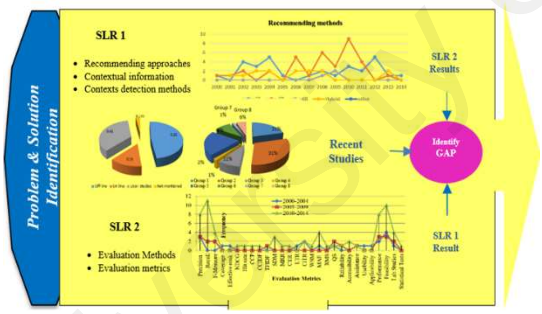
Note:
SLR = Systematic Literature Review **UX** = User Experience **UI**= User Interface **UM**= User Model **SRS** = Scholarly Recommender Systems **PLS-SEM** = Partial Least Square regression - Structural Equation Modeling **BN**= Bayesian Network **C**= Contextual **KFCV**= K-Folds Cross Validation **MSE**= Mean Square Error **MXE** =Mean Cross- Entropy

Figure 3.2: Research Methodology

The purpose of the first activity is to identify what the problem is (Peffers et al., 2007). The problem can be identified from the literature review or it can come from own or colleagues' expertise and experience (Lassenius, Soininen, & Vanhanen, 2001). For identification of the research problems in this thesis, as elaborated in Chapter 2, extensive literature reviews on recommending methods of SRSs, contextual information, and evaluation methods are conducted. It is pertinent to note that until the last months of submitting this thesis, the review of the relevant studies have been ongoing to keep the related work up-to-date.

In the following section, a brief discussion on the activities performed for the identification of problems and solutions is provided, and then the research objectives are mapped to the research problems derived from the existing issues.

3.2.1.1 Data acquisition and analysis process in phase 1



The data acquisition process and analysis of literature for identification of the research problems have been elaborated in Chapter 2. This phase includes the development of a systematic review protocol, conducting of review according to the defined protocol, analysis, reporting and visualisation of results, and discussion on findings. Figure 3.2 summarises the existing issues, and highlights the fact that there is a gap of how contexts influence UX of SRSs. There are difficulties in detecting the most influencing context. Besides, user modelling is generally mostly ignored in SRSs researches, but this is recognised recently as one of the best solutions to enhance the UX of RS. Finally, the UI has been rarely taken into consideration by the researchers of SRSs.

3.2.1.2 Solution identification and objectives

The first activity is concerned with the identification of the problem (gap), whereas the purpose of the second activity is to determine how the problem should be solved. This activity seeks to define specific objectives and solution requirements to determine how the problem should be solved. To carry out this activity, essential requirements of the solution are identified after the research problem is established and the current state-of-the-art solutions are reviewed. Four primary research objectives are intended to be achieved by this study as discussed in Chapter 1.

3.2.1.3 Mapping of research problem and research objectives

Figure 3.3 shows how research objectives are mapped to the research problems derived from the existing issues. Considering the focus of this research, Chapters 4, 5, 6 and 7 describe the objectives of this thesis in terms of how they can support the designing of SRSs to enhance UX.

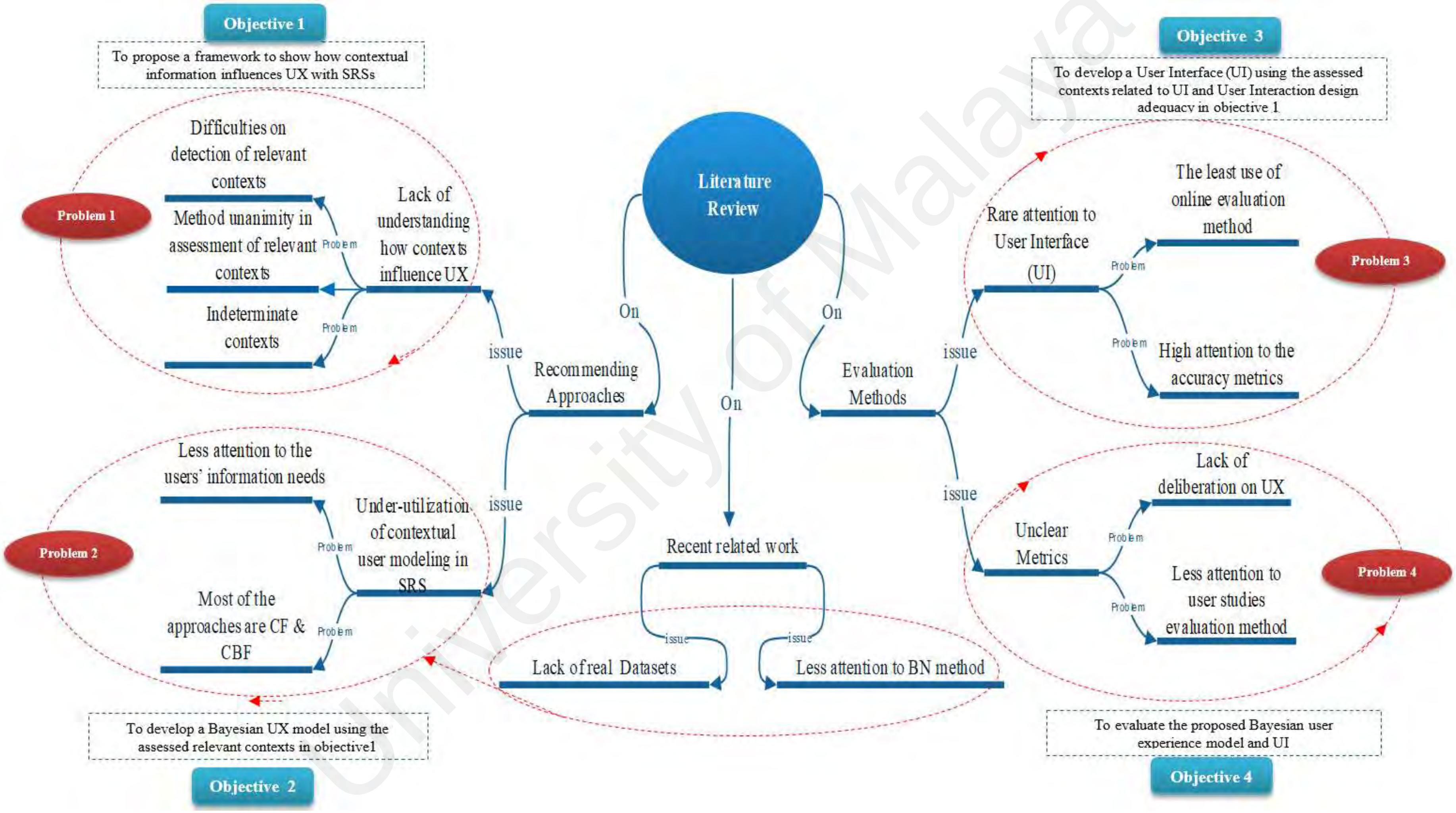
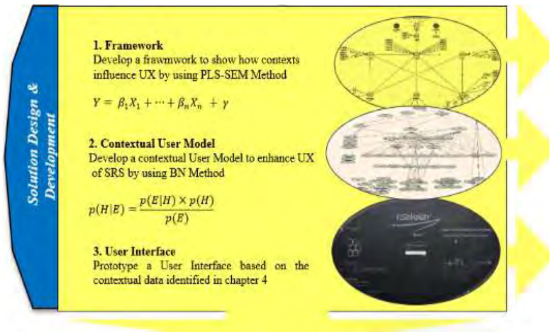


Figure 3.3 : Mapping research problems & objectives

3.2.2 Phase 2- Design & Development



This activity refers to the creation or development of the artefact or solution such as models, methods, constructs, or instantiations (Peffers et al., 2007), (Hevner, 2007). This activity also specifies the desired functionality of the underlying concepts and of the architecture. In Phase 2, the objectives from the previous activity are used as a basis for the development of a solution, which is the answer to the mentioned problems.

Three artefacts are proposed in order to solve the mentioned problems. As the current research focuses on contextual UM, these two steps are taken: firstly, investigate how contextual information influences UX of SRSs; secondly, detect the most influencing contextual information to be considered in a Bayesian UM. Since UX is dependent on UI, and some of the contextual information characterises the situation of UI, a UI is designed in this thesis. Therefore, the design and development phase includes three primary sections, which are also the first three objectives of this research. They are discussed separately in the following section.

3.2.2.1 Framework Development: Achieving Objective 1

The objective explores how contexts influence UX with SRSs and to assess what the most relevant contexts are. The framework has been formed in two stages: firstly, the concept of the framework is composed; and secondly, it is empirically examined. Therefore, the existing models and theories of UX, especially with RSs, are reviewed first, and then a conceptual framework is proposed. The framework shows in what way contexts influence UX with SRSs. The proposed framework not only enriches the conceptual understanding of how contextual information influences UX of RSs, but also serves as a foundation for further theoretical and empirical investigations. Moreover, a better understanding of relevant contextual information can also help researchers to design an effective SRS and reduce the data complexity. The set of structurally related contexts assessed in this research can be embedded into both back-end (algorithms) and front-end (user interface) in order to enhance the UX of SRSs. Chapter 4 discusses the theoretical basis, adoption of main component and identification of variables of the framework. The relationships drawn between the various latent variables and concepts leading to the proposition of relevant hypotheses are theoretically justified. Finally, the proposed framework is examined empirically. The activities performed to achieve this objective are shown in Figure 3.4.

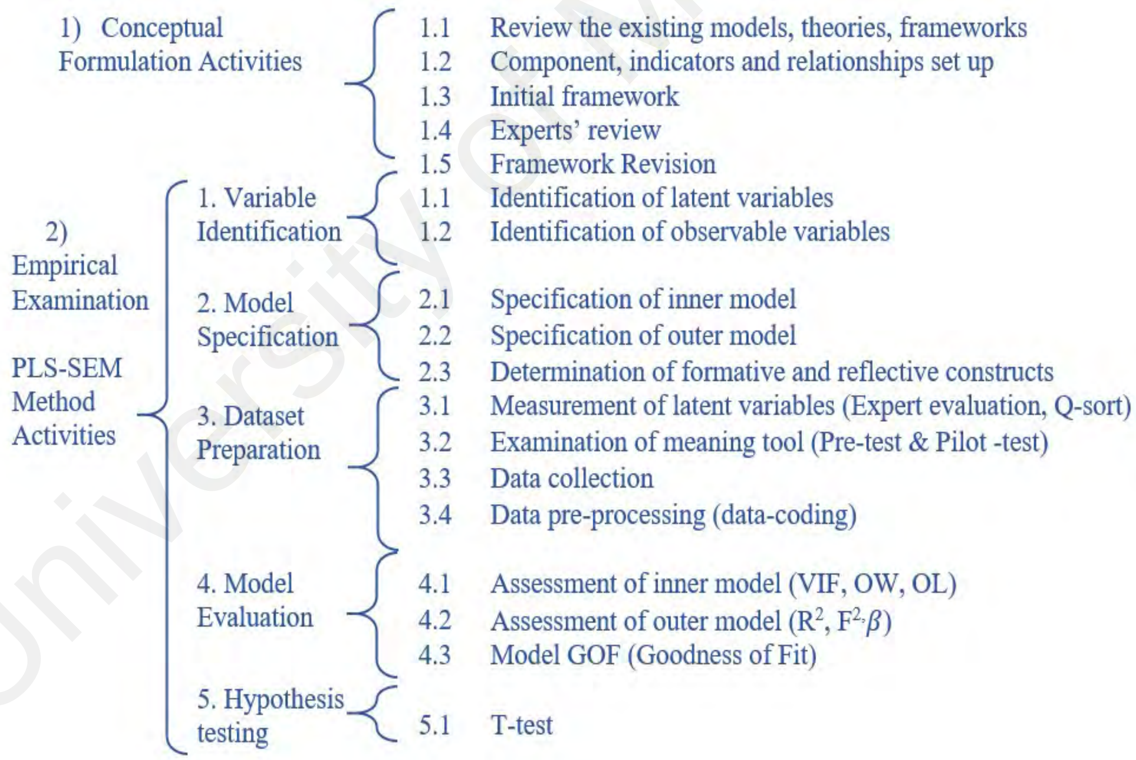


Figure 3.4: Activities performed for Objective 1

(a) Data acquisition for Objective 1

As mentioned, the conceptual foundation of the proposed framework is based on the relevant existing theories and experts' feedback, but it is empirically examined by the data collected through the end users of SRSs. For collecting the required data, this study targets the population of computer science scholars in three different countries, which are Malaysia, Iran and Canada. The questionnaire was available on the *"www.rscholar.com/quest"* server from 24 September 2015 until the 28 March 2016 (almost 6 months). During the time period of data gathering, a total of 177 useful responses were received and used in the data analysis. Chapter 4 describes in detail the demographic information of the participants of this study and the process of data collection.

(b) Data analysis method and justifications for Objective 1

The examination of the framework and detection of most relevant contexts are performed using the empirically activated quantitative method of PLS-SEM. This method is a predictive technique (Pratley, van Voorthuysen, & Chan, 2014) and useful for exploring research objectives (Hulland, 1999); it is particularly relevant in studies using less developed theories (Henseler et al., 2014) and the phenomena of which is relatively new (Sharma & Kim, 2012). This method helps to estimate the values of latent variables of relationship models that are mostly subjective and not directly measurable. Additionally, a few pre-processing assumptions such as large sample size and normality have been avoided in this method (Hair et al., 2014; Urbach & Ahlemann, 2010). Besides, in comparison with the methods that can analyse only one layer of linkages between variables at a time, the PLS-SEM method has the advantage of answering a set of interconnected enquires in a single, systematic, and comprehensive analysis (Hair et al., 2011). In other words, by using PLS-SEM, a single run of analysis algorithm can simultaneously calculate both the variable values of measurement model (the correlation between the measurement indicators and their related construct) and the structural model (the conceptualised linkages between the various constructs in the research model)

(Pratley, van Voorthuysen, & Chan, 2014). Therefore, this method is useful for data or dimension reduction to eliminate or mitigate the complexity and ambiguity in the system.

3.2.2.2 Development of Bayesian UM: Achieving Objective 2

To achieve this objective, the relevant contexts identified in Objective 1 have been exploited to design a Bayesian UM in order to diagnose the users' information needs in searching for scholarly papers at four levels: accurate, novel, diverse and popular papers. The proposed UM can be embedded in the recommending process to improve the UX. The activities performed to achieve Objective 2 are shown in Figure 3.5. Details of user modelling process are discussed in Chapter 5 and evaluation of UM is discussed in Chapter 7.

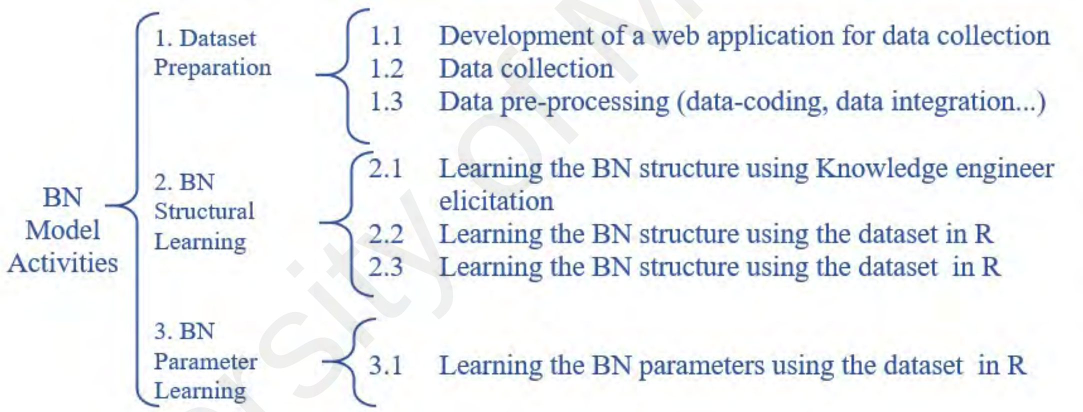


Figure 3.5: BN model activities

(a) Data acquisition for Objective 2

In order to gather the required data for the BN modelling, a web application was prepared and the link of the web application *"www.rscholar.com/app"* was shared by an email invitation to computer science scholars, including master's, PhD, post-doc students, and faculty members. The data collection started from 10 January 2016 until 30 July 2017. During the time period of data gathering, a total of 1121 records were received, of which 1053 records have been used in the process of BN development. After the dataset preparation, some of the activities shown in Figure 3.6, such as data cleaning and missing data, are performed. The detailed information about the data acquisition, dataset preparation, and data pre-processing activities are provided in Chapter 5.

Figure 3.6: Data cleaning process

(b) Data analysis method and justifications for Objective 2

As discussed in Chapters 1 and 2, the BNs are powerful methods most commonly used for uncertainty modelling of behavioural data, especially, but not exclusively, in capturing perceptions (R. Rim, M. Amin, & M. Adel, 2013a). Their initial appearance was in the field of medical decision systems in the late 1970s (Pearl, 1985). The inference of BNs is based on a probabilistic theory called Bayes' theorem to spread knowledge within the network (Heckerman, Mamdani, & Wellman, 1995; Neapolitan, 2004).

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E)} \tag{3.1}$$

In Formula 3.1, the prior probability of A is $p(H)$, $p(E|H)$ is the likelihood function of H, and $p(E)$ is the prior probability of E, which is called marginal probability. Thus, $p(H|E)$ is a posterior probability of H given E (R. Rim, M. Amin, & M. Adel, 2013a). In a general form, $p(H)$ and $p(E)\geq 0$, and $p(H_i)$ consists of mutually exclusive events within the universe S; the Bayes' formula would be (Bolstad & Curran, 2016) as below:

$$(H_i|E) = \frac{p(E \cap H_i)}{p(E)} = \frac{p(E|H_i) \times p(H_i)}{\sum_{j=1}^{n} p(E|H_i) \times p(H_i)} \tag{3.2}$$

More precisely, BNs are a class of graphical models that allow a concise representation of the probabilistic dependencies between a given set of random variables (nodes)$X = \{X_1, X_2, X_3, \dots X_n\}$ as a directed acyclic graph (DAG) *G=(V,A).* Each node $v_i \in V$ corresponds to a random variable X_i, which might represent a causal link from a parent node to its children (R. Rim, M. Amin, & M. Adel, 2013a). Each node is associated with a conditional probability distribution which assigns a probability to each possible value of this node for each combination of the values of its parent nodes (Zukerman & Albrecht, 2001). More information about BN method and justification is provided in Chapter 5. Briefly, the BN is chosen for the UM in this thesis because of the advantages over other ML methods:

1. It is able to overcome uncertainty of contextual data
2. It is able to represent the cause-effect relationships between the contexts
3. It is well adapted for development of user modelling
4. It is appropriate for diagnosing users' information needs of scholarly papers through the established theory of probability
5. It is well adapted for other recommending approaches.

3.2.2.3 Design of UI: Achieving Objective 3

As discussed in Chapter 2, the UI has a strong impact on the user perception of the recommendations' quality. The SRSs studies have mostly neglected this impact, which might lead to biased conclusions. A UI named rScholar is designed in this research. Figure 3.7 briefly describes the activities performed in order to design the rScholar.

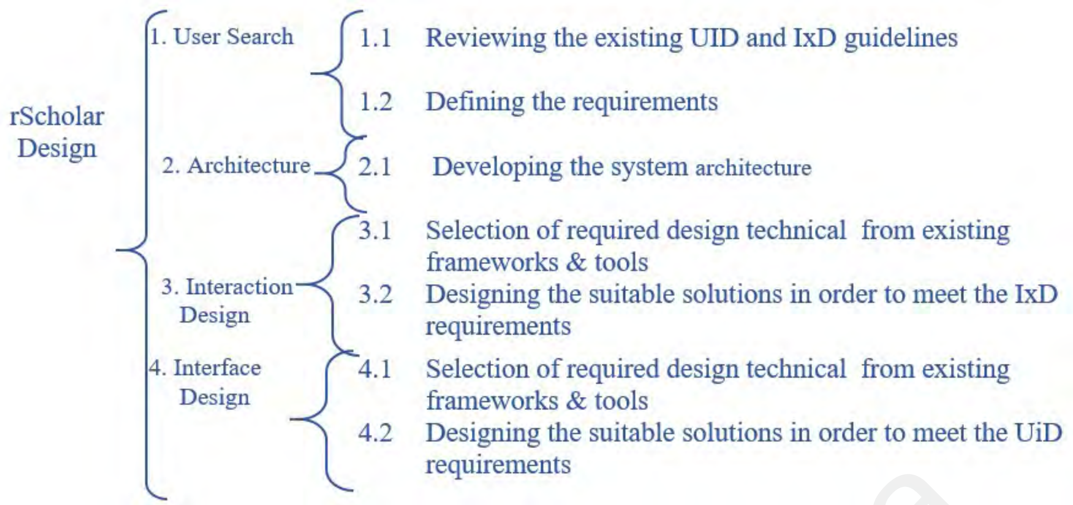


Figure 3.7: UI design activities

(a) Data acquisition for Objective 3

As mentioned earlier in Chapter 1, the rScholar (Figure 3.8) (available at *www.rscholar.com)* has been designed based on the contextual features (UI and interaction design adequacy) identified from Objective 1, and considering the functionality of the UM for interaction between the user and SRS (Data inputs and outputs of UM).

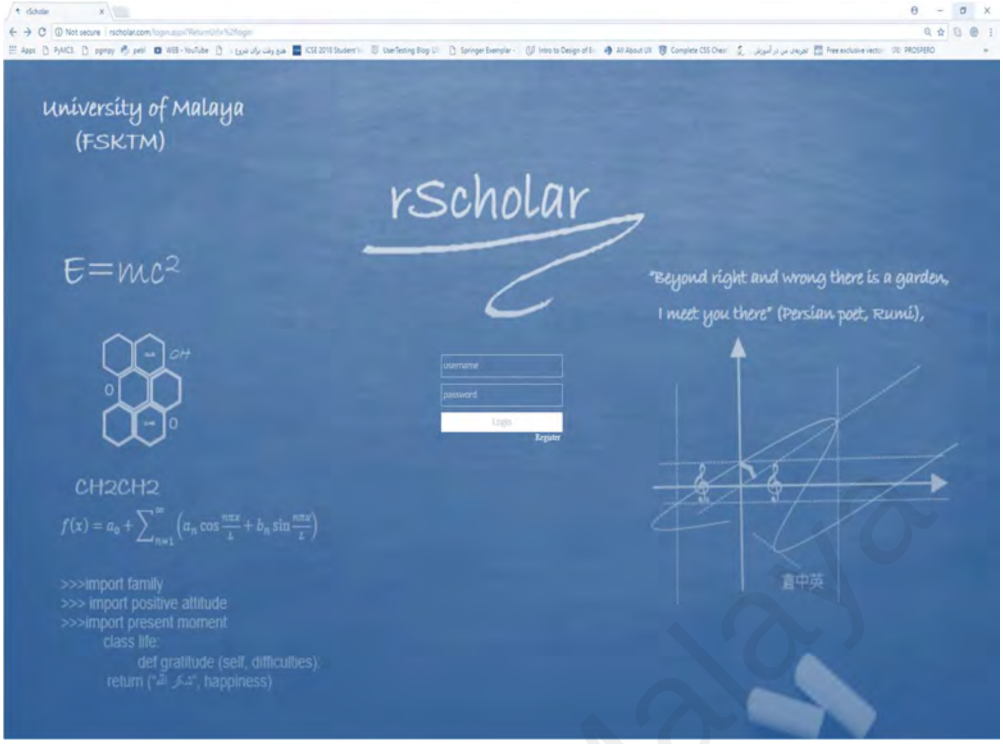
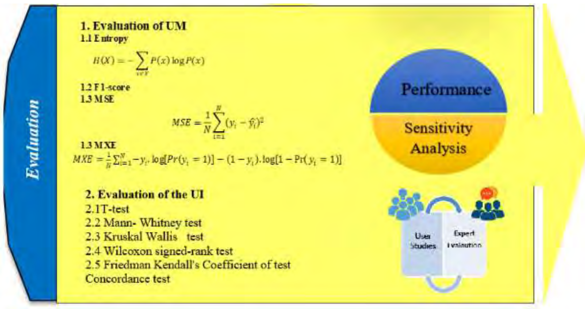


Figure 3.8: rScholar - login page

(b) Design methods and justifications for Objective 3

In order to design rScholar, Bootstrap - CSS, JS, and HTML frameworks are used for building responsive web pages. In addition, Illustrator, a vector graphics editor, is used for the primary design of web pages along with logos. ASP.net, a fully supported and free web application framework for building standards connected web solutions, is also applied to develop the UI. LINQ to XML is a LINQ-enabled, in-memory XML programming interface that enables it to work with XML from within the .NET Framework programming languages. The rScholar aims to provide users with a better experience when interacting with SRSs. One of the advantages of the proposed UI (rScholar) is that it has been designed based on the UI and interaction design adequacy of SRSs identified from Objective 1. For example, it offers two different page layouts, and the pages are responsive. All features are discussed in detail in Chapter 6 and the UI evaluation is described in Chapter 7.

3.2.3 Phase 3- Evaluation



The purpose of this phase is to prove that the artefacts developed in the previous phase work well by solving one or more problems (Peffers et al., 2007) (Hevner, 2007). The evaluation consists of two parts: the first part discusses the UM evaluations; and the second part explains the UI evaluation. Figure 3.9 depicts the activities performed for the evaluations of the mentioned artefacts.

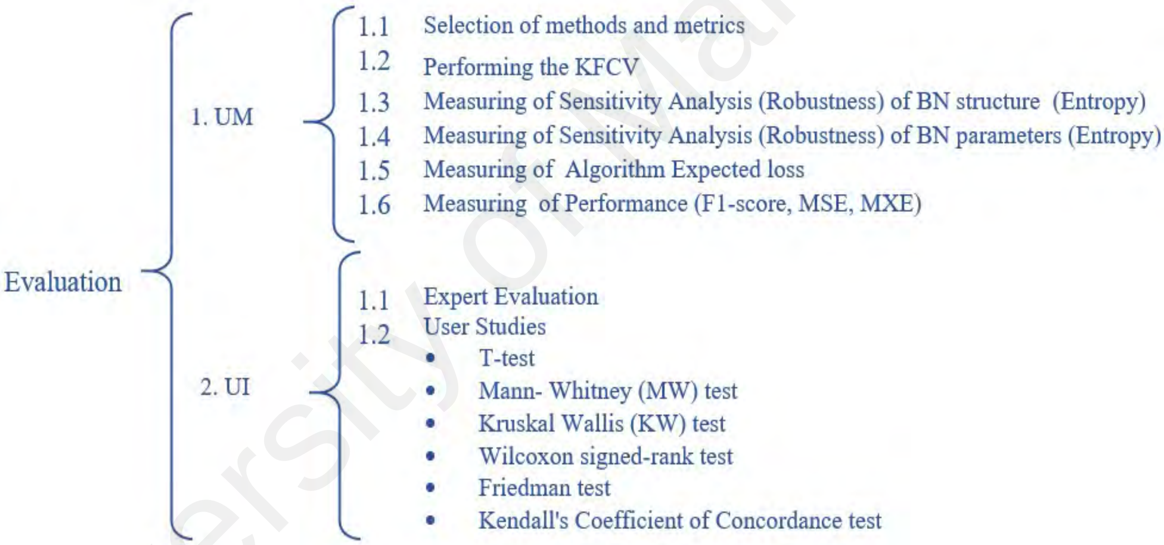


Figure 3.9: Evaluation activities

3.2.3.1 Data acquisition for Objective 4

For evaluations of the UM, the KFCV method is applied to split the dataset provided in Chapter 5 (for the UM development). For the UI evaluation, the data is collected by obtaining the feedback of nine experts and twenty participants (end user). The details of the data preparation for the evaluation of both UM and UI are presented in Chapter 7.

3.2.3.2 Evaluation methods and justifications for Objective 4

The offline evaluation is used for the evaluation of the BN model. Offline method has some shortcomings, but it is the most used method in the evaluations of SRSs (Beel, Genzmehr, et al., 2013; Beel & Langer, 2014; Beel, Langer, Genzmehr, et al., 2013b). As shown in Figure 3.9, the Entropy metric is applied to conduct the sensitivity analysis in evaluating the robustness of BN structure and parameters against the changes. For evaluation of the BN algorithm, the expected loss method is used. Finally, to evaluate the performance of BN, three metrics of F1-score, MSE and, MXE have been calculated (Korb & Nicholson, 2003; Kuenzer, Schlick, Ohmann, Schmidt, & Luczak, 2001) (Flores et al., 2011; Margaritis, 2003). The above-mentioned metrics are selected mostly based on the past studies in the field of BN modelling (Seixas, Zadrozny, Laks, Conci, & Saade, 2014) (G.Marcot, 2012). The evaluation of the user interface is performed by two methods: user studies and expert evaluation. As depicted in Figure 3.9, the expert review and user studies method is applied for the evaluation of the proposed UI. After data collection, six different tests are applied in order to analyse the data. The details of the evaluation are presented in Chapter 7.

3.2.3.3 Communication

The research by Peffers et al. (2007) and Hevner, March, Park, and Ram (2004) have emphasised the importance of communication as a part of research. The construction of the artefact (framework) as a part of this research is a PhD research project; therefore, communication was carried out through conferences and journals related to the area of this research. Furthermore, this dissertation is also a single comprehensive piece of communication.

3.3 Summary

This chapter describes the overall research strategy and roadmap of the current research in order to achieve the defined objectives. This chapter also lists the specific quantitative methods used in different stages of this research, along with the outcomes or deliverables. Justification for certain methods has also been discussed in this chapter. The next chapter focuses on devising a framework to demonstrate in what way contexts influence UX with SRSs, introduced as Objective 1.

CHAPTER 4: FRAMEWORK DEVELOPMENT

The primary goals of this chapter are to explore how contexts influence UX and to assess the most relevant contexts incorporating in the UX with SRSs. This chapter reviews existing models and theories of UX especially in the field of RSs and then proposes a conceptual framework. The framework manifests in what way contexts (latent variables) influence UX with SRSs. After theoretically justifying the relationships drawn between the various latent variables leading to the proposition of relevant hypotheses, the experiment of the conceptual framework and detection of most relevant contexts are performed using the quantitative method of Partial Least Squares (PLS) Regression and Structural Equation Modeling (SEM). In this research based on the instruction of PLS-SEM method, various tests of VIF, OW, OL, β, R^2, F^2 and, SRMR have been applied in order to empirically examine the framework. The proposed framework not only enriches the conceptual understanding of how contextual information influences UX of SRSs but also serves as a foundation for further theoretical and empirical investigations. Moreover, a better understanding of relevant contextual information can also help researchers to design effective SRS and reduce the complexity of data. This chapter provides a conspectus of the whole contexts influencing user's perceptions.

4.1 Conceptual formulation of the framework

For conceptual construction of the framework, five steps are performed as shown in Figure 4.1. Prior to describing the steps, the RQ1 and its sub research questions are reviewed and then the steps to achieve them are elaborated respectively

RQ1: How does contextual information conceptually influence UX with SRSs?

- ✓ SRQ1: What models/frameworks/theories have been proposed for UX of RS/SRS in the existing studies?
- ✓ SRQ2: What components and relationships can be applied for the framework?
- ✓ SRQ3: What indicators can be applied for the components?

✓ SRQ4: What contexts can be applied?

✓ SRQ5: What are the experts' review feedbacks on the proposed conceptual framework?

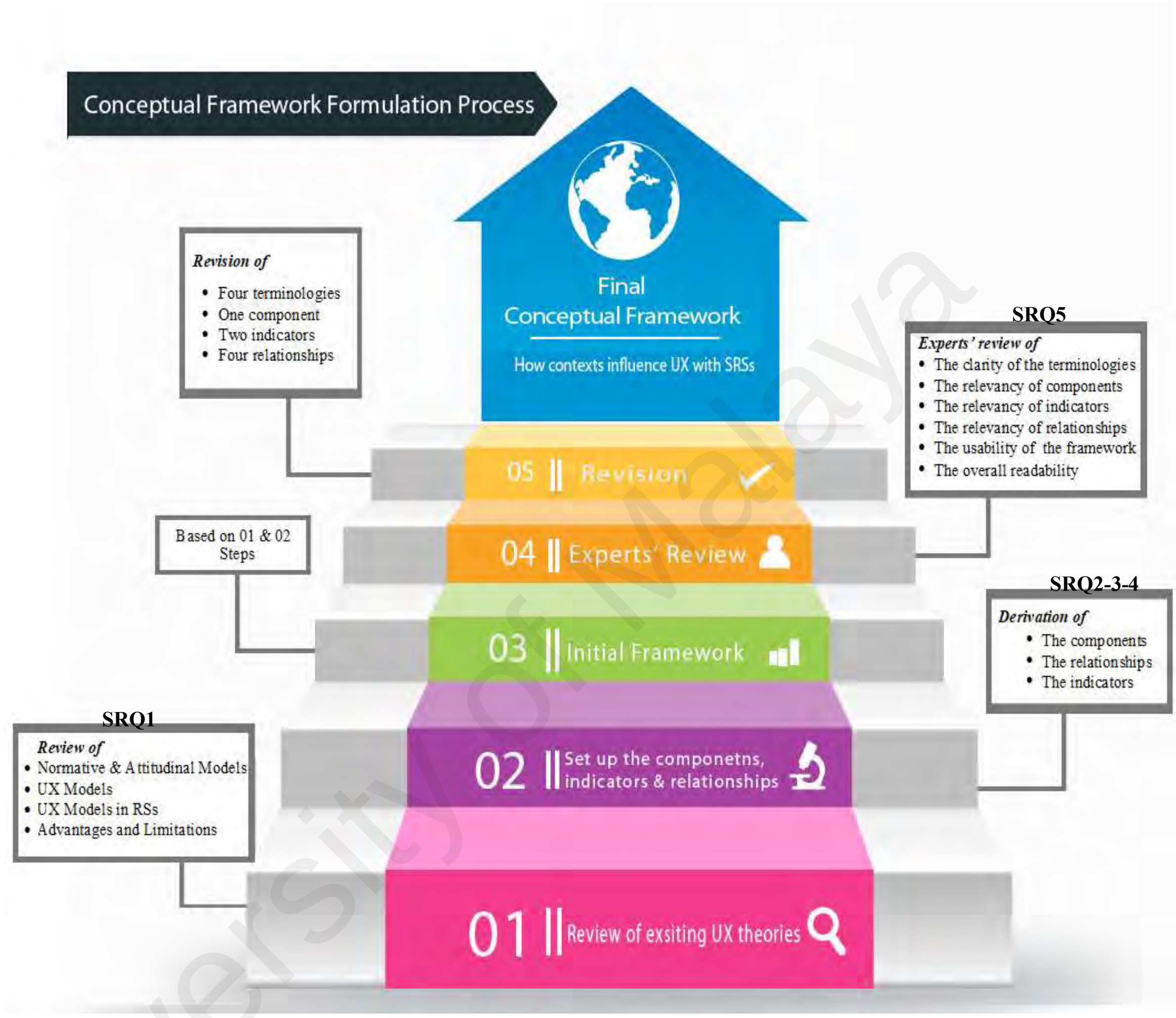


Figure 4.1: Conceptual formulation of the framework

4.1.1 Review of existing studies

The main goal of proposed framework is to present a set of structurally relevant contexts influencing UX of SRSs, which can be embedded into both back- end (algorithms) and front- end (user interface) of SRS development in order to enhance the UX of SRS. In the following, as a first step, several theories, models and frameworks that utilized as a basis for establishment of proposed framework are conversed.

4.1.1.1 Normative and attitudinal models

The Theory of Reasoned Action (TRA) proposed by Fishbein and Ajzen (Fishbein, 1975) is the fundamental of many human behavior models. This theory posits that person's behavioral intention is influenced by attitudinal and normative factors therefore for understanding of person's behavior intention, it is important to find out the attitude toward that behavior (Figure 4.2).

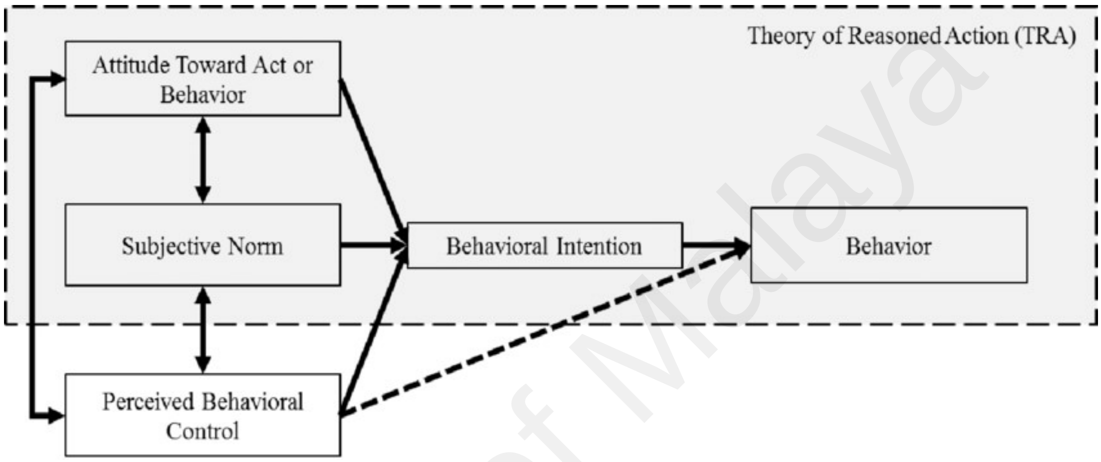


Figure 4.2: Theory of TRA proposed by Fishbein and Ajzen (Fishbein, 1975)

Later the TRA was adopted as a part of another theory called Technology Acceptance Model (TAM) developed by Davis, Bagozzi and Warshaw (Davis, Bagozzi, & Warshaw, 1989). TAM theory claims that the attitude towards using a technology is influenced by two perceptions of usefulness and ease of use of the system. In 2003, a modified version of TAM theory was formulated by (Venkatesh, 2003) named the Unified Theory of Acceptance and Use of Technology (UTAUT). The UTAUT inspires the normative part of TRA and aims to bold the role of personal and situational characteristics on behavioral intention.

4.1.1.2 UX Models

Researchers indicated that UX is a subjective phenomenon however its impacts might be reflected by the users' observable behaviors(Bart P. Knijnenburg et al., 2012). Since the 1980s, several models, theories have been developed to illuminate the interactive experience of using digital technologies (Hart, 2015). Once researchers explored that usability does not account for subjective emotions, UX emerged to explain the subjective experience when user is interacting with a product or a system and led to a shift from designing for users to designing with users (Visser, Stappers, Van der Lugt, & Sanders, 2005). This shift mostly brought concepts such as fun(Monk, Hassenzahl, Blythe, & Reed, 2002), pleasure(Green & Jordan, 2003), aesthetics(Tractinsky, Katz, & Ikar, 2000) and hedonic qualities (Hassenzahl & Tractinsky, 2006) up to understanding of UX. Some of the models have considered UX as a cognitive process that can be modelled and used to measure or evaluate changes in perception and judgement over time. "Sander's Experience Model" (Sanders, 2002; Visser et al., 2005) as shown in Figure 4.3, postulates that experience is an intersection of memories of the past, current experience, and future dreams which is felt individually. Sander believes that UX can be involved in the process of design once we have access to people's experiences (past, present and potential) (Sanders, 2002).

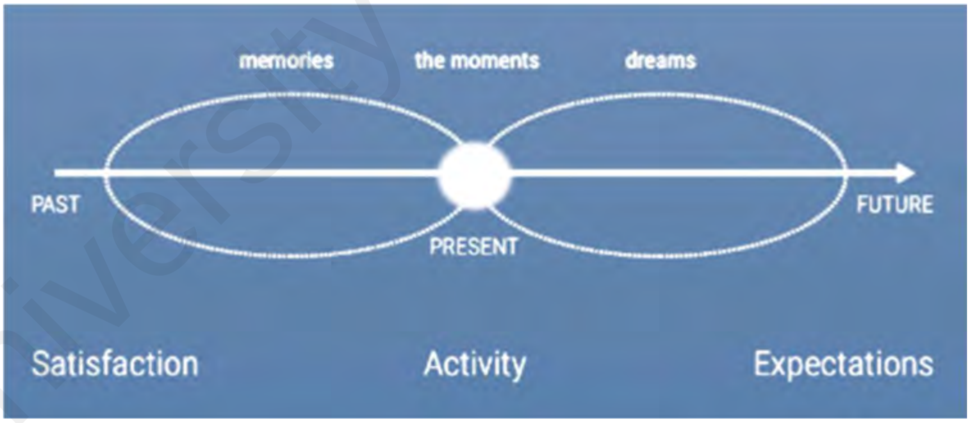


Figure 4.3: Sander's Experience Model

According to Sander's model, by accessing what people say, think, do, use, know, feel and dream, four sources of knowledge including explicit knowledge, observed experience, tacit knowledge and latent needs are achievable (Shown in Table 4.1). Consequently, the resonance with people is established by having those knowledge.

Table 4.1: Different Knowledge's taken from users

Knowledge Type	*How to achieve the knowledge*
Explicit knowledge	Listening to what people say (what they are able to express in words)
Observable information	Watching what people do/ use
Tacit knowledge	Discovering what people think and know (provides us with their perceptions of experience)
	Understanding how people feel (gives us the ability to empathize with them)
Latent needs	Seeing and appreciating what people dream (how their future could change for the better)

The three levels of design theory also is considered cognitive process of UX over the time. This theory was proposed by Don Norman(D. Norman, 2013a), who is most well-known for advocacy of user-center design. He discussed that there are three different levels of experience and that these experiences can be triggered by three different levels of design including visceral, behavioral and reflective. The visceral reaction is immediate and often beyond our control, is the one precipitated by the initial sensory scan of the experience. The behavior experience is when the user is using the product and finally there is an experience beyond the initial experience of using a product. It's the experience of association and familiarity when the user is not holding the product however has feeling for it and is able to put values on the product in retrospect (D. Norman, 2013a).

Mahlke & Thüring (Thüring & Mahlke, 2007) proposed the CUE (Components of User Experience) model which consists of three UX components, instrumental, non-instrumental and the emotional reactions of the user (Figure 4.4). In CUE model, system properties, user-characteristics, and task/context influence interaction characteristics; in turn, interaction characteristics influence perceptions of instrumental qualities and perceptions of non-instrumental qualities, both of which lead to emotional reactions; all three are antecedents of appraisal of the system.

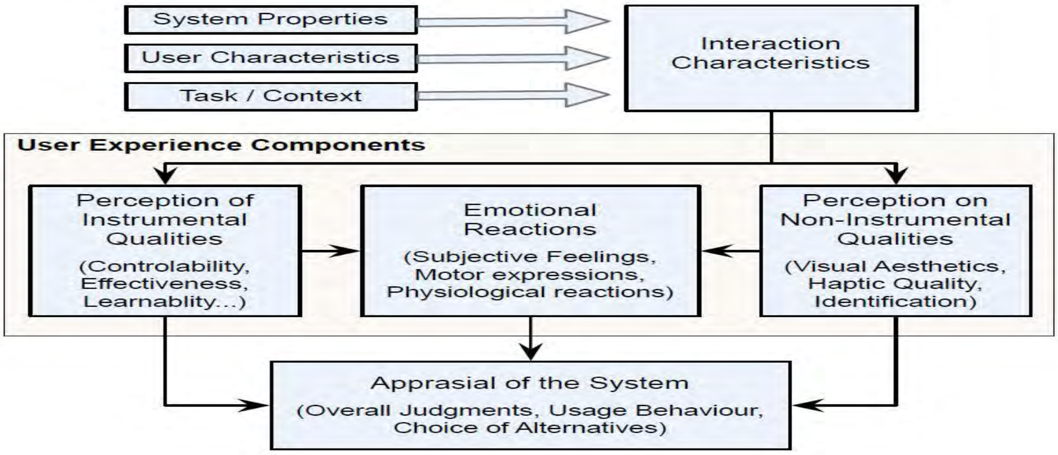


Figure 4.4: The integrated CUE model proposed by Mahlke & Thüring, (2007)

Hassenzahl foreground the impact of hedonic attributes like pleasure in Hassenzahl et al.'s UX model (Hassenzahl & Tractinsky, 2006), while all TRA-related theories only discuss pragmatic attributes. He indicates that the product evaluations in terms of appeal, pleasure and satisfaction is affected by perceptions of product characteristics including pragmatic quality and hedonic quality(Hassenzahl et al., 2013). There is a user interface quality assessment model developed by Hartmann et al.'s (Hartmann, Sutcliffe, & Angeli, 2008). This model presents that users' background, goals and task impact the system assessment as well as decision-making criteria (usability, aesthetics…).

4.1.1.3 RSs Models

The overall system evaluation is also influenced by decision- UX models for RSs. Zins and Bauernfield (Bauernfeind & Zins, 2005)conducted a survey among users of two travel RSs and an application for finding digital cameras then based on the survey results, they constructed a model for the UX of RS. The model displays how user's satisfaction is influenced by trust, flow, and browsing behavior, and how these in turn is effected by the personal characteristics. However, this model does not explain how objective system aspects such as quality of recommendation impact on UX which is a prominent limitation of their model.

McNee et al. (Sean Michael Mcnee, 2006; Sean M McNee, Riedl, & Konstan, 2006b) have proposed an analytic model called Human-Recommender Interaction (HRI). He accentuated the significance role of the context such as users' tasks on the UX of RSs. In addition, they emphasize on the appropriate RS dialogue through the analysis of users' needs. McNee believes that if recommender is going to help users, it has to be designed based on the real-world information seeking tasks. Not considering users' information needs and even though their background knowledge puts the recommender's designers in to pitfalls(Sean M McNee, Kapoor, et al., 2006).

Xiao and Benbasat (Xiao & Benbasat, 2007) conducted a vast and considerable literature review of the business and marketing oriented research on RSs and then proposed a framework showing how certain characteristics of RSs such as type and process, and output design influence users' trust and satisfaction. They considered personal and situational characteristics as the moderator variables, which influence users' evaluation. Ozok et al.'s (Ozok et al., 2010) studies are mainly on RS usability to yield design guidelines based on the users' study survey. The guidelines mostly describe the impacts of specific system aspects on the usability of RSs. However, their approach is mostly descriptive which depends on the users' opinions about RSs rather than experimental results of a specific system. Pu and Chen's framework (Pu et al., 2011) consists of four main dimensions as shown in Figure 4.5. This framework relies on a user centric approach to RS evaluation and links user's perception of quality to user's beliefs while user's beliefs are antecedents of the users' attitudes, which are antecedents of the behavioral intentions (inspired by TRA). This framework brought a new approach to user's perception of recommendation quality that signalized this approach compared to the exiting frameworks. However, the framework does not explain what contexts influence user's perceptions.

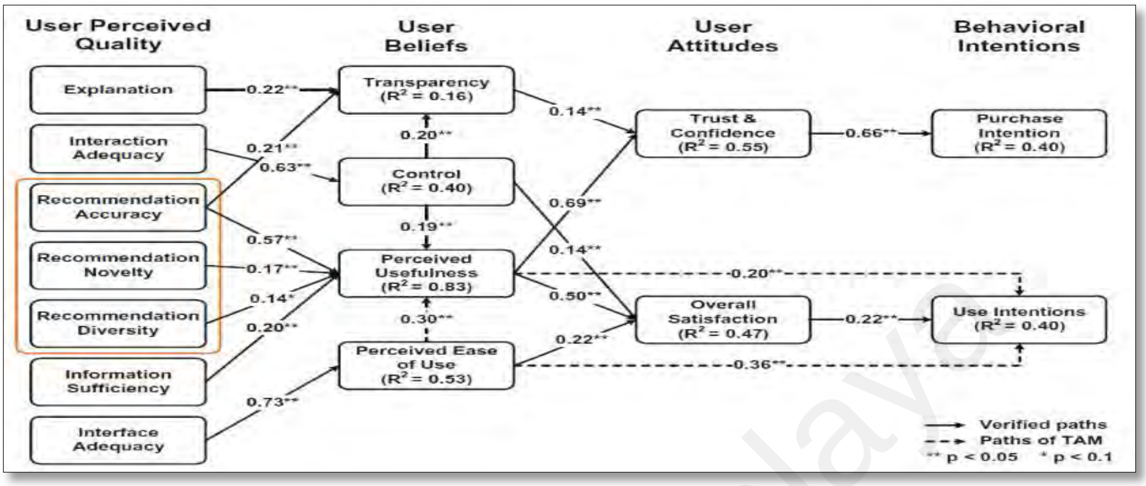


Figure 4.5: Pu and Chen's Framework of perceived qualities of recommenders (Pu et al., 2011)

(Bart P. Knijnenburg et al., 2012) disputed that for analyzing the UX of SRs, the accuracy of recommendations, consideration of other aspects also is essential since measuring accuracy is an insufficient method. They advocated an evaluation framework that examines the influence of subjective system aspects such as recommendation and interaction into objective user behaviors such as purchase and use as personal. They found that the subjective aspects such as perception of quality, interaction usability and appeal, have strong correlation with users' behaviors shown in Figure 4.6. In addition, their experiments showed that why and how subjective system aspects bring about the user experience of RSs. However, they have not elaborated the subjective system aspects and the relationships with situational characteristics and personal characteristics.

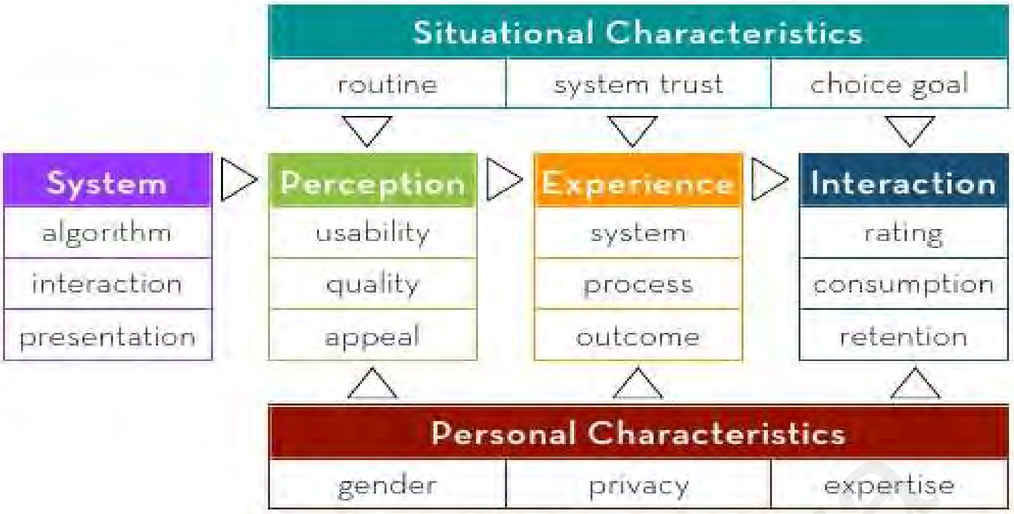


Figure 4.6: UX evaluation framework (Bart P. Knijnenburg et al., 2012)

4.1.1.4 Advantage and limitation of the existing models

As the main goal of this research is enhancing UX of SRS therefore in the following Table 4.2, the limitation and advantage of the relevant existing models have been discussed.

The key limitation of existing models is that they have not discussed how contextual data impacts on the UX. Exploiting contextual information in the process of recommending has been recognised as the most appropriate method to make tailored recommendations for the users and enhance users' experiences (Adomavicius & Jannach, 2014; Zheng, 2017). Recently, researchers emphasize on user centric evaluation of RSs (Beel, Breitinger, et al., 2016). This is actually a paradigm shift in RSs research field since before that all the researches have been trying to develop more accurate algorithms not to enhance UX (Joseph A Konstan & John Riedl, 2012) (Bart P. Knijnenburg et al., 2012). However, it is still not clear that how contexts and what contexts impact on users' evaluation of the RSs. Besides, it have not proved empirically. Moreover, in the field of SRS, there is no study that has investigated the most influencing contexts on the users' experience of scholarly recommenders and clarified that what are exactly those contexts. Another key limitation of existing models is that they have not paid attention into long term user experiences. UX should not only be seen as something evaluable after interacting with an object, but also before and during the interaction. While it is relevant to evaluate short-term experiences, given dynamic changes of user goals and needs related to contextual factors, it is also important to know how (and why) experiences evolve over time.

Besides, in the models that have been proposed specifically for RSs, the users' perceptions or beliefs have not elaborated completely. For example; in Pu and Chen's framework, only four indicators of “control”, “ease of use”, “usefulness” and “transparency” have been considered while the users' perceptions which users perceives are more than the indicators mentioned in the past models. The perceptions such as visual aesthetics, personalization and fun that are also very important in users' centric evaluations have been ignored to consider in the existing models in the field of RSs.

Finally, among the existing models and theories, only three models are related to the recommenders and among those three also the HRI Model proposed by (Sean M McNee, Riedl, et al., 2006b) was examined for the domain of research paper recommenders. However, the main focus of this model is not the impact of the contexts on the UX which is a prominent limitation of this model.

Table 4.2: Advantage and limitation of existing theories/ models/ frameworks

Model/ Theory	*Advantage*	*Limitation*
TRA, TAM, UTAUT theories	- Discuss understanding of person's behaviour intention - Discuss attitude towards using a technology - Discuss the factors influence acceptance and Use of Technology	- Not discuss the impact of contexts - Not discuss UX - Not discuss UX of RSs - Not examine the results empirically
Sander's Experience Model	- Discuss UX - Discuss explicit, tacit knowledge and, observed experience influencing UX	
CUE Model	- Discuss UX - Discuss context - Discuss instrumental, non-instrumental and the emotional reactions	- Not discuss the impact of the whole contexts - Not discuss UX of RSs - Not discuss long term impact of UX - Not examine the results empirically
Hassenzahl UX model	- Discuss UX - Discuss pragmatic quality and hedonic quality	- Not discuss the impact of contexts - Not discuss UX of RSs - Not discuss long term impact of UX - Not discuss how contexts influence UX
HRI Model	- Discuss UX of SRSs and users' needs - Discuss users' seeking behaviors	
Pu & Chen's framework	- Discuss RS evaluation from a user centric view point - Results were examined empirically	- Not discuss the impact of contexts - Not discuss long term impact of UX
Bart P. Knijnenburg et al.'s framework	- Discuss UX of RSs - Results were examined empirically	- Not elaborated the subjective system aspects - Not elaborated the relationships with situational characteristics and personal characteristics - Not discuss long term impact of UX

4.1.2 Components, indicators & relationships set-up

Based on the limitation of existing models particularly in the field of RSs, the main components of the framework has been initiated. In addition, based on the existing LRs on RSs, the indicators and relationships are set up which are discussed in the following section separately.

4.1.2.1 Derivation of components & relationships

As shown in Figure 4.7, the framework in this research has five main components including context, perception, attitude, feeling and appraisal.

Figure 4.7: The components involving the proposed framework

Five main components were chosen that are mostly based on taking the advantages and overcoming the limitations of existing models which are discussed in the following.

First, the contexts are taken as a starting point because in most current models, factors such as users' background, characteristics, goals, task which considered as contexts are the starting point of proposed frameworks which impact on users' perceptions (Pu et al., 2011) (Thüring & Mahlke, 2007) (Hartmann et al., 2008) (Bart P. Knijnenburg et al., 2012). However; as discussed there is a lack of elaboration on how and what contexts impact on the UX of RSs in the existing models. Besides, as McNee (Sean M McNee, Riedl, et al., 2006b) and a few studies in the field of RSs have indicated; incorporating contexts into recommending process can enhance UX (Adomavicius & Tuzhilin, 2011b; Baltrunas et al., 2012; Panniello & Gorgoglione, 2011). Therefore, this research devotes a considerable attention to conceptualization of contexts including user, system and environment contexts and investigation of their impact on the users' perceptions in the process of UX of SRS. Second, based on the (Bart P. Knijnenburg et al., 2012) and CEU models, contexts affect users' perceptions (second component). Also, Pu and Chen's (Pu et al., 2011) framework conceptualizes that the user's perception of quality is initiated by recommendation diversity, novelty, accuracy, interaction adequacy, interface adequacy, information sufficiency and transparency. However, Pu and Chen's (Pu et al., 2011) has not discuss context influencing UX. From this point of view, this research bears much similarity to Pu and Chen's framework, but goes one step back to explore how these quality perceptions come out and one step beyond it to detect the whole contextual information. Furthermore, as mentioned before, a key shortcoming of current models is that they have considered only a few perceptions such as "control", "ease of use", "usefulness" and "transparency" while in the literature there are more perception such as perception of fun, personalization and visual aesthetics that users might perceived. Also, Hassenzahl's UX model emphases on considering visual aesthetics however they have been ignored in the most current models particularly in RSs models. This research takes the mentioned perceptions into consideration and aims to experiment them in the UX process. Third, the TRA, TAM, UTAUT theories have demonstrated that the person's behavioral intention is influenced by attitudinal and normative factors. Based on the aforementioned theories, it has been differentiated between attitudes and behavioral intention that created by attitudes towards this framework for SRSs. Forth, CUE model, perceptions and attitudes influence emotional reactions or feelings; hence, in the initial proposed framework, the feeling is presented after attitudes(Green & Jordan, 2003; Hassenzahl et al., 2013). Users might feel pleasure during the experience of interaction with the RS. For example, when user revives a good and unexpected recommendation that can meet his/her information need, he/she might feel pleasure. Hassenzahl emphasizes on pleasure moment and design for the happiness which embraces both features of a product; functionality and aesthetics (Hassenzahl et al., 2013). Like

Hassenzahl's model, the pleasure with the addition of trust are considered in this framework. Fifth, according to the TAM, UTAUT, CUE models, the people' behaviors or reactions is a product of emotional reactions which is called appraisal in CUE model. Knijnenburg et al (Bart P. Knijnenburg et al., 2012) also revealed that what users' feel about the system impact on the subjective system aspects such as user satisfaction in this way the appraisal component is derived to end up the flow of the proposed framework.

Finally, a major shortcoming of existing UX research in the field of RSs is that UX happens in a long term interval and the long term aspect of UX is mostly ignored in the existing models. This framework is also inspired by Norman and Sander's theories whereas the influence of long term variable is conceptualized as a moderator variable on of users' feeling and appraisal. Table 4.3 summarized the derivation of the components, relationships as well as references.

Table 4.3: Derivation of the components and relationships

Component	Reference	Relationships	Reference
Context	CUE Model (Thüring & Mahlke, 2007), (HRI) (Sean Michael Mcnee, 2006; Sean M McNee, Riedl, et al., 2006b), (Bart P. Knijnenburg et al., 2012)	*Context → Perception*	(Pu et al., 2011) (Thüring & Mahlke, 2007) (Hartmann et al., 2008) (Bart P. Knijnenburg et al., 2012)
Perception	TRA, TAM, UTAUT theories, CUE Model (Thüring & Mahlke, 2007), Pu and Chen's framework (Pu et al., 2011), (Bart P. Knijnenburg et al., 2012)	*Perception → Attitude*	TRA, TAM, UTAUT theories, Pu and Chen's framework (Pu et al., 2011)
Attitude	Pu and Chen's framework (Pu et al., 2011)	*Attitude → Feeling*	UX model (Hassenzahl & Tractinsky, 2006)
Feeling	CUE Model (Thüring & Mahlke, 2007), UX model (Hassenzahl & Tractinsky, 2006), Pu and Chen's framework (Pu et al., 2011)	*Feeling → Appraisal*	CUE Model (Thüring & Mahlke, 2007), UX model (Hassenzahl & Tractinsky, 2006)
Appraisal	TRA, TAM, UTAUT theories, CUE Model (Thüring & Mahlke, 2007), UX model (Hassenzahl & Tractinsky, 2006), Hartmann et al.'s (Hartmann et al., 2008)	*Over time → Feeling-Appraisal*	Norman & Sander's Experience Model

4.1.2.2 Derivation of the indicators

Each component consists of a few determinants or indicators taken from the literature of RSs. In the following, the derivation components' indicators have been described. Also, each indicator is discussed accordingly. Table 4.4 shows an overview of the whole indicators. The indicators mostly have been identified using a Systematic Literature Review (SLR) discussed in Chapter 2. The published paper from the results of SLR is also attached to the Appendices. In addition, all indicators' references are provided in Appendices D, E, and F.

Table 4.4: Derivation of the indicators

	Component	Indicators	Component	Indicators
Context	*User Situation*	Profile	*Interface Design Adequacy*	Visualization
		Task		Gamification
		Pre-Knowledge		Consistency
		Scholarly network		Info. sufficiency
		Search logs		Display
		Learning style	*Perception*	Fun
		Mood		Transparency
		Personality trait		Personalization
		Search Status		Usefulness
	Environment Situation	Time		Visual authentic
		Location		Dominance
System Situation	*Paper Quality*	Accuracy	*Attitude*	Trust
		Novelty	*Feeling*	Pleasure
		Popularity		Surprise
		Diversity	*Over time*	
		Serendipity	*Appraisal*	Usage
	Interaction Design Adequacy	Preference elicitation		Overall satisfaction
		Preference refinement		Expectation
		Explanation	*The references are provided in Appendices*	
		Privacy consideration		

As mentioned earlier, the contexts are taken as a starting point for the UX. As discussed in chapter 2, the SLR results Champiri,et al(Champiri et al., 2015) has revealed that contextual information applied in SRSs has been categorized into three class of user situation, environment and resource[1] (paper) contexts however considering the provided

[1] The resource is a scientific paper throughout this research

discussion in chapter 2 and Dey's definition of context where the context is any information that can be used to characterized the situation of an entity (Dey, 2001), there are other contextual information such as interaction and interface contexts influencing user experience with SRSs which have not been considered in existing SRSs. Furthermore, most studies have considered accuracy as the main part of resource context or quality while accuracy partially constitutes UX of SRSs and researchers have recommended to apply other features diversity, novelty and popularity to make a better list of recommendation and improve UX of RSs(Kotkov, Veijalainen, & Wang, 2016; Kotkov, Wang, & Veijalainen, 2016; McCay-Peet & Toms, 2011). The major contexts of user, environment and system situation are considered as the contextual information in this research.

4.1.2.3 Derivation of the indicators for user situation-context

User situation characterise users' current situation interacting with the SRSs or even before and after it. In this research, only the assumed features that have an impact on UX are considered. As Figure 4.8 depicts, based on the literature eight features have been identified as the user context.

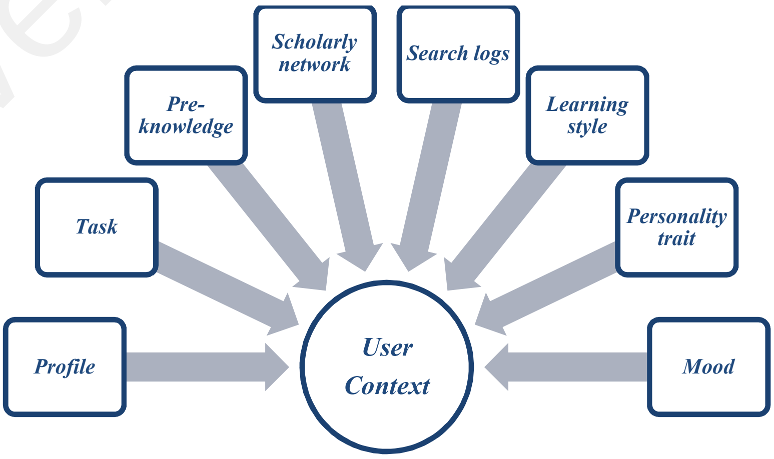


Figure 4.8: User situation contexts- indicators

In the following sub-sections, each user context feature is discussed. The conditions of each context are also provided in Appendix D.

(a) Profile

The users' profile information, including personal or demographic information, general interests and research areas is used to make recommendations. The fixed information such as identity, name, age and gender is considered as the long-term profile information. Meanwhile, information like research interests, which change consistently, is considered as the short-term information. Additional profile information include degree (undergraduate, graduate), relation between the majors and reading materials (e.g. computer science students read the electronic books), primary article publications, project descriptions, roles and memberships.

(b) Task

Users look for papers for various purposes such as to maintain awareness, explore research area and find relevant sources. To make better recommendations, RSs must take information-seeking purposes into consideration (Kuo & Zhang, 2012). For example, it may be useful and satisfactory to send the last updated papers in a specific domain if the user is continually searching to obtain more information about a certain research area. Different scenarios of the users' tasks have been presented to support the scholarly communication process in RSs. Matching the users to their specific tasks leads to increased user satisfaction, efficiency, and usefulness of the recommender system (Sean M McNee, Kapoor, et al., 2006). Marko A. Rodriguez. et al. (Marko A. Rodriguez, 2009) extrapolated four kinds of purposes from the investigations of the scholars' information seeking behaviours. These purposes usually are: 1) An article related to another article of interest, 2) A potential collaborator for a funding opportunity,3) An optimal venue to submit their article and 4) Referees to review an article in the role of an editor or other tasks such as completing an assignment, preparing a paper or proposal and writing a thesis, which have been indicated in a few studies (Dehghani, Afshar, Jamali, & Nematbakhsh, 2011; Konstan et al., 2005; R. Patton et al., 2012).

(c) Pre-knowledge

Pre-knowledge or background knowledge has been defined as the primary specific knowledge about a topic an individual has learnt formally or informally (via experience). In the academic context, it is considered as the content knowledge, academic language and vocabulary necessary for understanding content information (Strangman & Hall, 2004). According to the reviewed studies, two types of pre-knowledge, including pre-knowledge in searching or information literacy and pre-knowledge in research area, were incorporated into useful recommendations for users seeking information. The RSs should identify users in terms of pre- knowledge and support them by offering relevant recommendations (Amini, Ibrahim, Othman, & Rastegari, 2011) (Sean M McNee et al., 2002)distributed a questionnaire to students at the University of Minnesota to survey the users' information seeking behaviours. They classified students according to their pre-knowledge in research area into four categories of novice users, experienced users in this field, experienced users in a related field and experienced users in a non-related field.

(d) Scholarly network

The collaboration with other users, work team and co-author relationships traceable by social networking is a relatively rich source of users' networks for creating relevant recommendations(Serrano-Guerrero, Herrera-Viedma, Olivas, Cerezo, & Romero, 2011). For example, the users' research network of social relationships, expertise, user similarities in research areas, published papers, and preferable journals are factors exploited by RSs (W.-S. Yang & Lin, 2013). Sinha and Swearingen (2001) interestingly realized that recommendations from friends are more helpful for users than those from system.

(e) Search logs

The ways individuals interact with systems in order to find and utilise information are described as the information behaviour such as seeking, reading, saving, downloading and printing information (Geyer-Schulz, Neumann, & Thede, 2003b). SRSs provide models based on users' information behaviours to calculate the users' preferences. For example, when a user downloads a paper (De Giusti et al., 2010) or reads a paper (Cheng Li 2008), it would be rated as the user' preferences. The effect of information taken from information seeking behaviour such as the browsing logs, search logs, saving logs and past work referenced papers on relevant recommendations was examined and presented in a few studies (Dehghani et al., 2011; S.-Y. H. Hwang, Wen-Chiang; Yang Wan-Shiou, 2003; Jung, Harris, Webster, & Herlocker, 2004; Middleton, Shadbolt, & De Roure, 2004; Tsuji et al., 2012; Yoshikane & Itsumura, 2013).

(f) Learning style

People learn differently since anyone has a unique learning style. Three different learning styles of visual, auditory and tactile have been indicated in the literature (Dunn, Beaudry, & Klavas, 2002). Visual leaners are those who understand the things from charts, pictures, diagrams, films, and written directions. These students will value to-do lists, assignment logs, and written notes. Many of these techniques, however, also benefit tactile learners. For the auditory learners the directions should be read aloud, speeches are required, or information is presented and requested verbally. They learn mostly from traditional teaching techniques such as lecture-style forum and presentation by regulating voice tone, inflection, and body language. Tactile learners understand the concepts through doing and touching. They are more successful when totally engaged with the learning activity. In this research, this is assumed that the learning style might have influence the way individual evaluate a paper as a relevant paper. For example if a paper includes lots of figures and charts, it might be more relevant to a visual learners. Also, they might consider interaction design as a key factor for working with the SRSs.

(g) Personality trait

Like the above mentioned indicator, this indicator also was added after the expert's review of the initial framework however because of the consistency of the component explanation, this is discussed right after of the other user situation contexts. Another feature that can influence user's preferences is user personality. Prior work in user personality have identified five personality traits of openness to new experiences, conscientiousness, extraversion (or introversion), neuroticism and agreeableness (S. D. Gosling, Rentfrow, & Swann, 2003; McCrae & Costa, 1987). Prior work also showed that there are significant connections between these traits and people's tastes, interests and preferences(S. Gosling, 2009) (Kraaykamp & Van Eijck, 2005; Rentfrow & Gosling, 2003). In 2016, Tien (T. Nguyen, 2016) conducted a research for a movie recommender to investigate the relationships between user personality and user satisfaction with the levels of recommendation diversity, popularity, and serendipity. The results showed that the integration of users' personality into recommending process helps to generate recommendations with the preferred levels of diversity, popularity, and serendipity to individual users and enrich the UX.

(h) Mood

Users go through different mood states during the accomplishment of scholarly activities (Jamali et al., 2011). These mood states can play an important role in their interaction with the system. For example, when a user cannot find appropriate papers due to the lack of IT literacy s/he might be in desperation status. It is important to mention that, among the user context, user's mood is an emotional state that the UXs in a particular situation. Moods express both positive and negative emotions. Based on the literature, the moods that may happen in a scholarly domain are such as dissatisfaction from not founding useful resources, anxiety of not aware of the subject and the appropriate keywords, confusion about the information needs (Jamali et al., 2011)(Will et al., 2009). At the first glance, it seems that user's mood does not have an impact on recommending personalised resources however a few researchers has pointed out the influence of users' moods (J. Herlocker et al., 2012), (Jung et al., 2004).

(i) Search Status

The UX is not only when users are interacting with the system but also it is before and after of users interactions (D. Norman, 2013b). Therefore it seems that for providing a better UX with the SRSs, it is helpful if examine if the users is interacting with the system or it is a time before or after it.

4.1.2.4 Derivation of the indicators for Environment context

As shown in Figure 4.9, environmental situation or context determines the physical environment that the users are in including time, location, service type, connecting mode and network bandwidth (J. Luo, Dong, Cao, & Song, 2010) and surrounding conditions. It seems that environmental contextual information is mostly employed in mobile RSs which are characterized by dynamic changes of environment (Biancalana, Gasparetti, Micarelli, & Sansonetti, 2013; C.-M. Chen & Yang, 2010). Environment contextual information might incorporate into recommendations to ensure that users receive fast, secure and relevant services in a dynamic and adaptive environment (Gómez, Zervas, Sampson, & Fabregat, 2014). Such context information is used to predict the user preferences for a location or specific time (Cheng Li 2008). In this research, only the impact of time and location are examined. The list of references is provided in Appendix E.

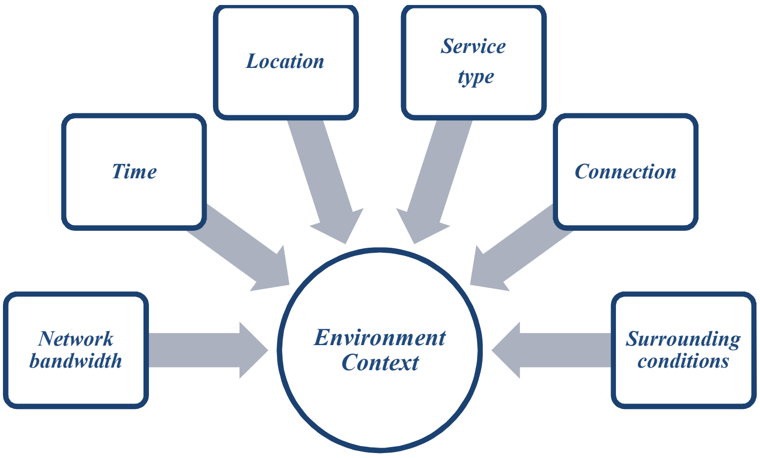


Figure 4.9: Environment context

4.1.2.5 Derivation of the indicators for system context

As depicted in Figure 4.10, the system situation or context includes paper quality, Interaction Design (IxD) and Interface design (UiD) adequacy which are discussed in respectively. In the following, first the paper quality is explained.

Paper quality refers to the attributes that each paper has to be matched with a specific scholar. Researchers argued that apart from the accuracy, other qualities such as diversity, novelty and popularity of the recommended papers are also important for the users (Adamopoulos & Tuzhilin, 2011; Hurley, 2011; Sridharan, 2014). For example, the aim of SRSs is not simply provide papers with similar keywords. To diversify the recommendation, the papers from different disciplines but with several shared citations, papers on a different topic, but adopting a similar methodology might be a good recommendation or papers on similar topics from journals that the user has never accessed before (Hurley, 2011).

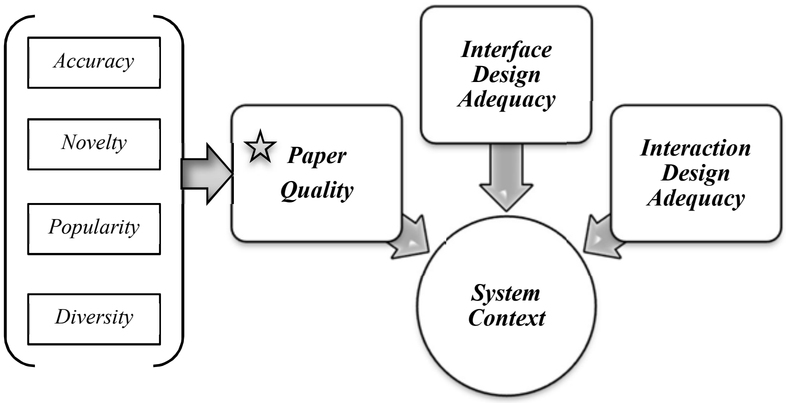


Figure 4.10: System context-paper quality

In chapter 2, it was discussed why diversity, popularity, and novelty matter in RSs (Shown in Figure 4.10). Here, a brief introduction of them is presented. The list of reference is provided in Appendix F.

(a) Accuracy

To achieve high accuracy, RSs tend to suggest indicators similar to a user profile. It means that what really is similar to the users' preferences and past ratings(T. Nguyen, 2016) (Felfernig, Burke, & Pu, 2012). So far most of the SRSs aimed to predict more accurate recommendations (Bart P Knijnenburg & Willemsen, 2010; Joseph A Konstan & John Riedl, 2012; Said, 2013). Depending on a particular application scenario, different actions that a user performs indicate his/her interest.

(b) Novelty

In the RSs literature, novelty is define as a recently added item that users have not yet rated or used or forgotten item. A user might forget that she/he consumed the item some time ago. Also is could be an unknown item that the user has never consumed (Rana, 2013). In other words, the indicators that users have not previously experienced or seen before. Therefore, it might be unpopularity and dissimilarity to a user profile (T. Nguyen, 2016) (Felfernig, Burke, & Pu, 2012). Adamopoulos and Tuzhilin (Adamopoulos & Tuzhilin, 2015) presented an approach to generate unexpected recommendations by maximizing a utility function which combines an item's relevance and its distance from a set of expected indicators. The set of expected indicators includes all indicators rated (i.e., seen) by the user and indicators to the seen similar ones. While user experience unexpected and fortuitous indicators from the McNee et al's perspective (Sean M McNee, Riedl, et al., 2006b) is called serendipitous indicators. However, Herlocker et al. (2004) asserted that novelty and serendipity are different concepts, since novelty only covers the new indicators while serendipity covers new and surprising indicators.

(c) Popularity

Popular item is widely recognized among a community. For example, a paper which has the high citation in a field might be considered as a popular item(Jannach et al., 2013). The user may have heard about the indicators from others or through the social networks(T. Nguyen, 2016). One of the ways to approximate popularity of recommendation is to filter the recommended item by frequent users.

(d) Diversity

Diversification is defined the variety in a list of recommendations. Recently, diversity has been the main focus of a few studies in RSs and Information Retrieval (IR) to improve user satisfaction (Adomavicius & Kwon, 2011; Jannach et al., 2013; C.-H. Tsai, 2017). The average of all pairwise distances (representing the differences) of any two recommendations is an indicator for measuring the diversity. Typical approaches replace indicators in the derived recommendation lists to minimize similarity between all indicators or remove "obvious" indicators from them (Adomavicius & Kwon, 2011).

4.1.2.6 Indicators Derivation - Interaction Design (IxD) Adequacy

IxD is a part of UX that focuses to understand how a person interact with an entity (website, application, car, and microwave) to meet a certain goal and how to tailor (design) the process of interaction (user's experience) to help user to meet that goal. Visual elements (user interface design) such as page layout, buttons are utilized to build up the process of interaction in a best way. The selection of visual elements (user interface) is determined by how helpful the element will be for user to achieve the goal. Sometimes in an interaction, both person and entity have certain goals to achieve. For example, a person interacts with a SRS to receive appropriate papers and SRS wants to get users' preference at the same time. During this interaction, the person's goal and system's goal are different while the ultimate goal is to make a better UX therefore the interaction should be tailored in a way that both system and user achieve their goals. In latter example, imagine that there are two different actions of rating and gaming available on the screen for the user in order to transfer his preference to the system however one of the action of rating is the primary action and for providing a better service, the system needs to know users' rating rather than gaming results. So in this case,

- IxD decides to offer two different options of rating and gaming to elicit the user preference whereas one of the options is the primary option.
- Interface design decides how design these two options while shows to the user the one which is more important.

HCI research has considerably focused on the use of interactive features to enhance UX within the intelligent systems (Pu et al., 2011) however in the field of RSs it is almost new(Jannach, Nunes, & Jugovac, 2017). Pu and Chen (Pu et al., 2011) have investigated the impact of appropriateness of interaction including elicitation of users' preference, revision of users' preference, and explanation of recommendations on the users' perception and concluded that besides recommendation quality, it is also important to design a suitable interaction between the users and the system because it influence a lot on the users' perception of overall satisfaction.

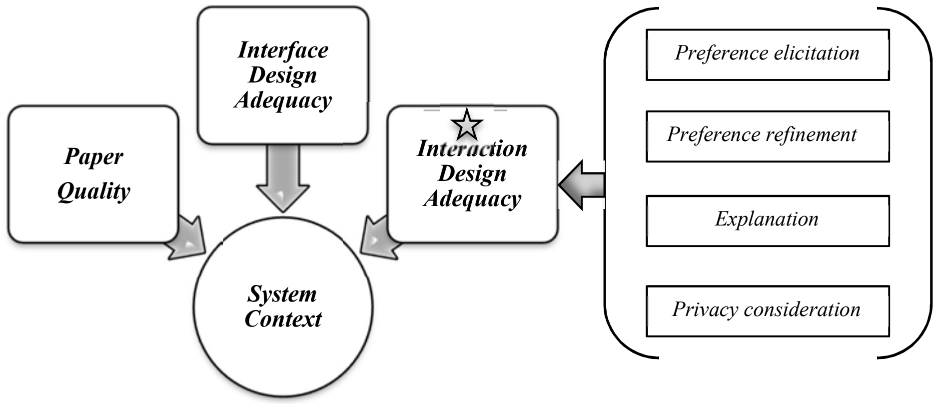


Figure 4.11: System context-interaction design adequacy

In this research as shown in Figure 4.11, the preference elicitation, preference refinement, explanation, privacy consideration are considered as the IxD indicators taken from literature. In the following paragraphs, they are explained briefly.

(a) Preference elicitation

Preference elicitation refers to the implicit and explicit methods to obtain user preferences. In implicit methods such as users' searching behaviours, users might not be aware or actively engage in the preference elicitation but users might expect system to give appropriate feedbacks to their searching behaviours. When users need to input their interests or rate the recommendations directly, it considered as the explicit elicitation method (Chen and Pu 2004; Peintner et al. 2008).The preference elicitation should be designed in a way to take out preferences from the user's mental representation and transfer the preferences into the system to reason with (Pommeranz et al., 2012; Pu et al., 2012a). McNeee founded through a user survey on a movie recommender that if the users are allowed to rate the indicators when they are sign up in the system, this influence positively users' loyalty to the system (Sean Michael Mcnee, 2006).

(b) Preference refinement

By applying preference refinements methods, users are allowed to refine their preferences (Pu et al., 2012a). Since the initial recommendations might not be accurate enough, the preference refinement helps users who seeks more accurate indicators(Swearingen & Sinha, 2002). Preference refinement sometimes is called relevance feedback mostly in the information retrieval literature(Kelly & Fu, 2006). It is a useful method to improve the recommendation after presenting to the users however it is mostly ignored in SRSs researches(Beel, Langer, Genzmehr, et al., 2013a). In fact, relevance feedback and profile feedback are two types of preference refinement methods. Middleton et al. (Middleton et al., 2004) showed that profile feedback (refinement of user models) is more effective than relevance feedback however There are not much researches on which method surpass the other one in RSs.

(c) Explanation

Another key part of a RS is the possibility to explain the inner logic of recommended indicators to the user (Felfernig & Burke, 2008). In other words, a RS should be able to explain why a set of specific recommendations are being recommended to a target user (J. L. Herlocker et al., 2004). Pu and Chen have indicated that explanation interfaces could effectively help build users' trust in the system. Other researchers also contend that explanation interfaces can cultivate user trust (Sinha and Swearingen 2002). Chen and Pu (Pu et al., 2012a) conducted a study on 54 users and concluded that recommendation explanations are more effective than annotations or unorganized lists. Besides, they indicated that explanations increase the system's acceptance and users' trust.

Labeling can be considered as a method for explanation of recommendations. For example, a study investigated the effect of labeling on paper recommendations. In this study similar papers recommended to users by two different labels. The showed papers labeled as ‘sponsored recommendation’ have an negative impact on users compare to those papers that have a label of ‘organic,’ though the recommended papers were similar(Beel, Langer, Nürnberger, & Genzmehr, 2013).

(d) Privacy consideration

Privacy consideration aims to protect users’ privacy while acquiring information about the users to generate recommendations(Resnick & Varian, 1997).The privacy concern is essential once the user’s information is more personal and users intend to keep the information confidential(M. S. Ackerman & S. D. Mainwaring, 2005). Some studies have shown that users overlook privacy concerns once they receive benefits or perceived usefulness from the system(Bart P Knijnenburg & Willemsen, 2010; Tinschert, Natt, Mautsch, Augthun, & Spiekermann, 2001). As indicated in chapter 2, SRSs mostly are part of other systems like digital libraries and the user’s privacy is very critical in libraries; hence, the process of making recommendations should not intrude the users’ privacy (M. Ackerman & S. Mainwaring, 2005). RSs should make recommendations based on the desired privacy level that users have predetermined in their user profile (Bollen & Van de Sompel, 2006) (Geyer-Schulz, Neumann, & Thede, 2003a).

4.1.2.7 Indicators Derivation - UI Design (UiD) Adequacy

One of the most important factors for users’ satisfaction with RSs is how well the recommendation presented to the users because most of the time, people are influenced by what they see(Pu et al., 2011; Svensson, Höök, & Cöster, 2005). As discussed before, Interaction and Interface design work together to meet the users’ goals. Several studies have investigated interface design issues of RS and provided some guidelines (Hiesel et al., 2016; Hu, 2010; Jannach et al., 2017; Ozok et al., 2010; Viriyakattiyaporn & Murphy, 2009). For example, a detailed set of design guidelines were proposed in (Cremonesi, Elahi, & Garzotto, 2017). As shown in Figure 4.12, in the literature, page layout and navigation are being emphasized rather than other features which influence on overall perceived ease of use and usefulness of system (Ozok et al., 2010; Swearingen & Sinha, 2001).

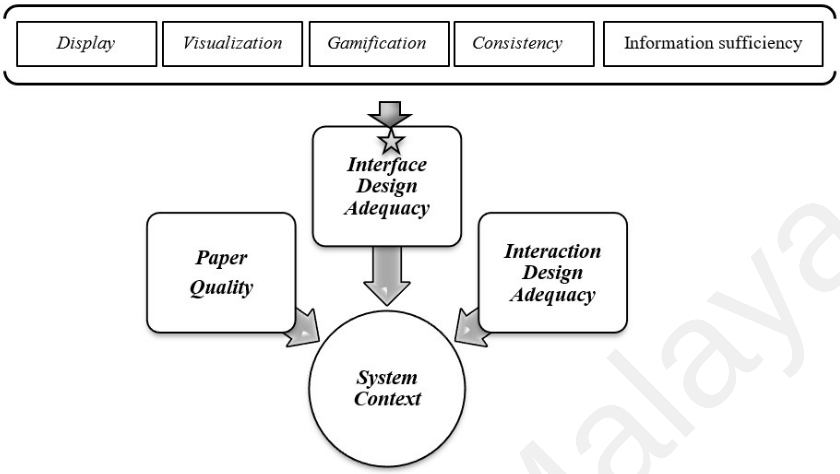


Figure 4.12: System context-interface design adequacy

Based on the best knowledge, this study is the first study which investigates the IxD and UiD features in the field of SRSs. Each of the UiD features are described in the following.

(a) Display

Display include a set of elements such as layout, input controls, navigational components, informational components and containers(Garrett, 2010). As mentioned earlier, page layout and navigation are the most important elements influence ease of use and usefulness of RSs (R. Rim, M. M. Amin, & M. Adel, 2013b; Swearingen & Sinha, 2001).

Number of recommendations or the set size (how many indicators to recommend) is one of the features that have been discussed in a few studies (Bart P. Knijnenburg et al., 2012) (L. Chen & Tsoi, 2011). Set composition of a few recommendations is a mixture of different kinds of recommendations such as top-ranked, relevant and diverse.

Researchers mostly indicated that showing more than five recommendation might make choice overloaded(L. Chen & Tsoi, 2011).

(e) Visualisation

Visualization leverages visual and graphical representations to facilitate user's perception(Hiesel et al., 2016). This is contrast to many textual recommendations, which precisely explain the recommendation. Trumper and Dollner (Murphy-Hill & Murphy, 2014) provide an overview of visualization techniques for RSs in software engineering. They emphasize visualization helps to make more transparency and interactivity in RSs. Also, another recommender called SmallWorlds visualizes the relationships between recommended indicators and similar friends' profile therefore users understand why certain indicators are being recommended to them and consequently it raises transparency and ultimately increases the chance of finding indicators. Context visualization also has been done by (Hiesel et al., 2016) to denote the user's current context (time, location, weather). Context visualization makes more apparent to users how the recommender works and reduce the cognitive effort which is required to uptake the meaning of the recommendation when it is presented (di Sciascio, 2017).

(f) Gamification

The term gamification consists in the use of game design elements in a non-game context to motivate and involve users in an activity, environment or any task for a long time that requires user engagement such as item ratings in RSs. In overall, games have this amazing ability to make relationships and trust between people, and develop their creative potentials (de CA Ziesemer, Müller, & Silveira, 2014; Feil, Kretzer, Werder, & Maedche, 2016). Gamification in SRs rewards users with the true recommendations generated by game feedbacks results. Moreover, at the same time, makes fun moments for the users (Hussain et al., 2014). Though the academic and scholarly tasks seem involving or relating to serious activities with less fun, this research investigates the role of using games in SRSs.

(g) Consistency

Consistency is using familiar icons, colors, menu hierarchy, call-to-actions, and user flows when designing similar situations and sequence of actions. It makes users to perceive the system easy to use and stable (Nielsen, 1995) (Shneiderman, 2004).There is not any trackback assumption in past studies discussing the relationships between consistency and UX of SRSs however this factor in this research is investigated as this factor is one of the most important UI design factors(Nielsen, 1995).

(h) Information sufficiency

Information sufficiency indicates the content of the recommendation (a set of features) should be sufficient for users to make confident decisions while saving time and effort (Ozok et al., 2010). This may evoke users' interest. In the example of SRSs it might be the information about the paper including title, abstract, keywords, author names and so on. Overall, information sufficiency refers to a piece of information about the recommendations that users would like to see once the recommendations are presenting to them. Based on the analysis taken from the CiteSeer recommender, Farooq et al. (Farooq, Ganoe, Carroll, Councill, & Giles, 2008) pointed out that that the information to display varies on the type of recommendation. For example in a RS which recommends paper, the bibliographic data such as title, author, citations, etc. are the most important data that a user likes to see before using the paper. Also, they found out that users would like to see both bibliographic data and abstract while the recommended papers are similar to the users' publication papers. Beel et al. (Beel, Gipp, et al., 2016) also indicated the Farooq et al.'s findings are interesting because the majority of SRSs display only the title and do not include the abstract in the presented information.

4.1.2.8 Indicators Derivation - User's perception

When people actively select, organize and interpret the information received to their brain by the sense, this process is called perception. In this essential process, the brain processes, rationalizes or makes sense the received information. It is noted that the perceptions are subjectively intuited by the users therefore they might change time by time(Bart P. Knijnenburg et al., 2012). Mahlke & Thüring (Thüring & Mahlke, 2007), the perceptions include instrumental (like transparency, dominance) and non-instrumental (like visual aesthetic) qualities however, in this study they have not divided specifically in two aforementioned categories.

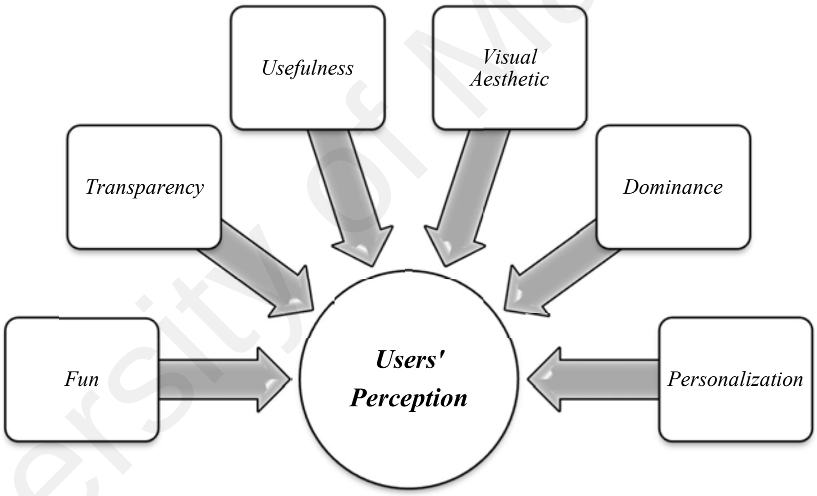


Figure 4.13: Users' perception indicators

As Figure 4.13 shows, six users' perception indicators are identified which are introduced here.

(a) Fun

Perception of fun is not vastly investigated in the field of RSs. Shneiderman has indicated(Shneiderman, 2004)that if the experiences are playful and liberating, users perceived fun-filled experiences which make users smile. There are two kinds of fun; fun-in-doing and fun-in-not-doing. The first one might be perceived by physical activities such as sports or mental challenges such as solving problems while fun-in-not-doing is not tied to action or goals such as relaxing. In creating technologies, mostly the focus is on fun-in-doing and the ways in which technology can be designed to make more fun for users. The topic of fun-in-doing traces back to early studies of games, such as Tom Malone's study on educational games(Malone, 1981).

(b) Transparency

Transparency is related to rationale. It clarifies to users why a particular recommendation is given to them (Murphy-Hill & Murphy, 2014). If the recommender explain its inner logic to users through the user interface, the users most likely perceive transparency(Pu et al., 2011). Also, it might influence the perception of usefulness(Swearingen & Sinha, 2002).

(c) Usefulness

Usefulness is the extent to which a user believes that using a certain product would improve his job performance (Davis et al., 1989) (Pu et al., 2012a). Totally, the objective of a SRS is to provide useful papers for the scholars and overcome the overwhelming information overload with which scholars are struggling.

(d) Visual Aesthetic

Visual design refers to the balance, emotional appeal, or aesthetic of a user interface and it might be expressed through colors, shapes, font type, music or animation (Cyr, Head, & Ivanov, 2006; Deaton, 2003; Hekkert, 2006). Visual aesthetic of RS has been examined in the study of of (Ozok et al., 2010). The users indicated that the color of pages are not important for them however the study of (Karvonen, 2000) has shown a the influence of aesthetic beauty on the users' trust.

(e) Dominance

User control of SRSs or dominance is about letting users decide what to do. For example, users have this authority to rate the indicators that they like to rate(Pu et al., 2012a). This approach will likely enables users to control their preferences also motivates them to rate more indicators in the future (di Sciascio, 2017; Joseph A Konstan & John Riedl, 2012). McNee et al.'s experimented an interface of RS and the results revealed that a slower initial rating interface that gave users more control (at the cost of more effort) led to higher user retention even though it did not improve actual prediction quality (Sean M McNee, Riedl, et al., 2006a).

(f) Personalization

SRSs aim to help users in a way that a resource is adapted to each user characteristic and preferences. This type of delivering tailored service is referred to in the literature as personalized recommendations(J. Luo, Dong, Cao, & Song, 2009; Neves, Carvalho, & Ralha, 2014).One of the promising ways to achieve personalized recommendations is by exploiting contextual information defined as any information which describes users 'situations such as location, time, and task(Joonseok Lee, 2013). Personalization can be perceived not only the quality of recommendation but also from the interaction and interface.

4.1.2.9 Indicators Derivation -User's attitude

As discussed before perceptions affect attitude. Based on the literature, trust is considered as the indicator of attitude.

(a) Trust

Trust is an important factor for making a strong and long term relationship (Bitner, 1995), it helps to maintain the relationship and resolve the conflicts (Roberts, 2001). In this research, trust is a product of user's perceptions from the interaction with system.

Trust influence overall user satisfaction and users' intention to use the system(Montaner Rigall, 2003; Panniello et al., 2015). According to the past studies, recommendation quality and explanation impact on users' of RSs. However, it is sometimes not easy to measure trust after a short term interaction of user with the system and it must be measured over time (Lim, 2012) (Viriyakattiyaporn & Murphy, 2009).

4.1.2.10 Indicators Derivation- User's feeling

In a few studies, feeling and attitude can be used interchangeably however in the initial framework, they have been considered as two separate components.

(a) Pleasure

Since the beginning of time, human has sought pleasure(Jordan, 1998). Pleasure is one of the major feelings from the impression occurred. Pleasure is also considered as the emotional, hedonic and practical benefits associated with a product(Green & Jordan, 2003; Hassenzahl et al., 2013). Users might feel pleasure during the experience of interaction with the RS. For example, when the user is playing a game and receiving a list of good recommendations, the user may feel pleasure. Like Hassenzahl's model(Hassenzahl et al., 2013), the pleasure with the addition of trust are considered in this framework.

(b) Surprise

Surprise refers to the feeling of receiving useful and unexpected recammendation (Kaminskas & Bridge, 2014b) (Reisenzein, 2000). The term of serendipity coined in the 18th century (Hu, 2010) also has been used in RSs literature to define the feeling of surprise(Kaminskas & Bridge, 2014a). The surprise is considered as an indicator to measure the impact of overall feeling.

4.1.2.11 Indicators Derivation - User's appraisal

Researchers in cognitive studies have conducted many studies to show how feelings mostly influence judgment and determination of the product's quality and also govern the user experience with the product (Spillers, 2004). The antecedents of user's feelings in producing the user's appraisal of SRS which might happen in a long term are examined. So, the variable of long term is considered as a moderator variable to show that the user's appraisal may build in a long term period. Like Sander's model, the user's appraisal is an intersection of three dimensions of user expectation, user behaviour, and overall satisfaction.

(a) Behavioral intention: Usage

Behavioral intentions are the users' feedbacks such as usage, rating, reading towards a SRS. It represent the ability of RS to influence users' behaviors and the degree that the users have accepted the system(Venkatesh et al., 2003). Once a scholar receives a few paper recommendations, the intentions might be one of the following behaviors;

- Reading the paper
- Recommending the paper to the interested friends
- Citing the paper in his academic works
- Saving in his library such as End note

(b) Expectation

When users use apps, websites, or software, they have different expectations about the product itself and its associated usage, which is reflected in the dialog design, user guidance, and achievement of goals. Users have very different expectations. This includes the click of a button, the need for information, or an aesthetically consistent design that can also be accessed on mobile devices.

(c) Overall Satisfaction

In a relationship, once people evaluate their experiences over the time and the positive status outcomes, it is so-called satisfaction (Anderson & Narus, 1990; Storbacka, Strandvik, & Grönroos, 1994). In a relationship or interaction with the RS, overall satisfaction measures the overall users' feeling and opinion in a direct way (Pu et al., 2011). User satisfaction is measured through two dimensions: internal user satisfaction (Torkzadeh & Doll, 1999)and overall satisfaction(Seddon & Kiew, 1996). Based on the literature, overall satisfaction can be measured by a single item(Seddon & Kiew, 1996) (Keiningham, Perkins-Munn, & Evans, 2003). In this study, the overall satisfaction is measured as a single item.

4.1.3 Initial framework: How contexts influence UX of SRSs

After connsidering the limitation of existing models particularly in the field of RSs and seting up the component, indicators, relationsips, a framework draft has been initiated shown in Figure 4.14 which shows the component, indicators and relatiosnhips. This framework is expected to lead developing and designing new SRSs in which UX has been centralized.

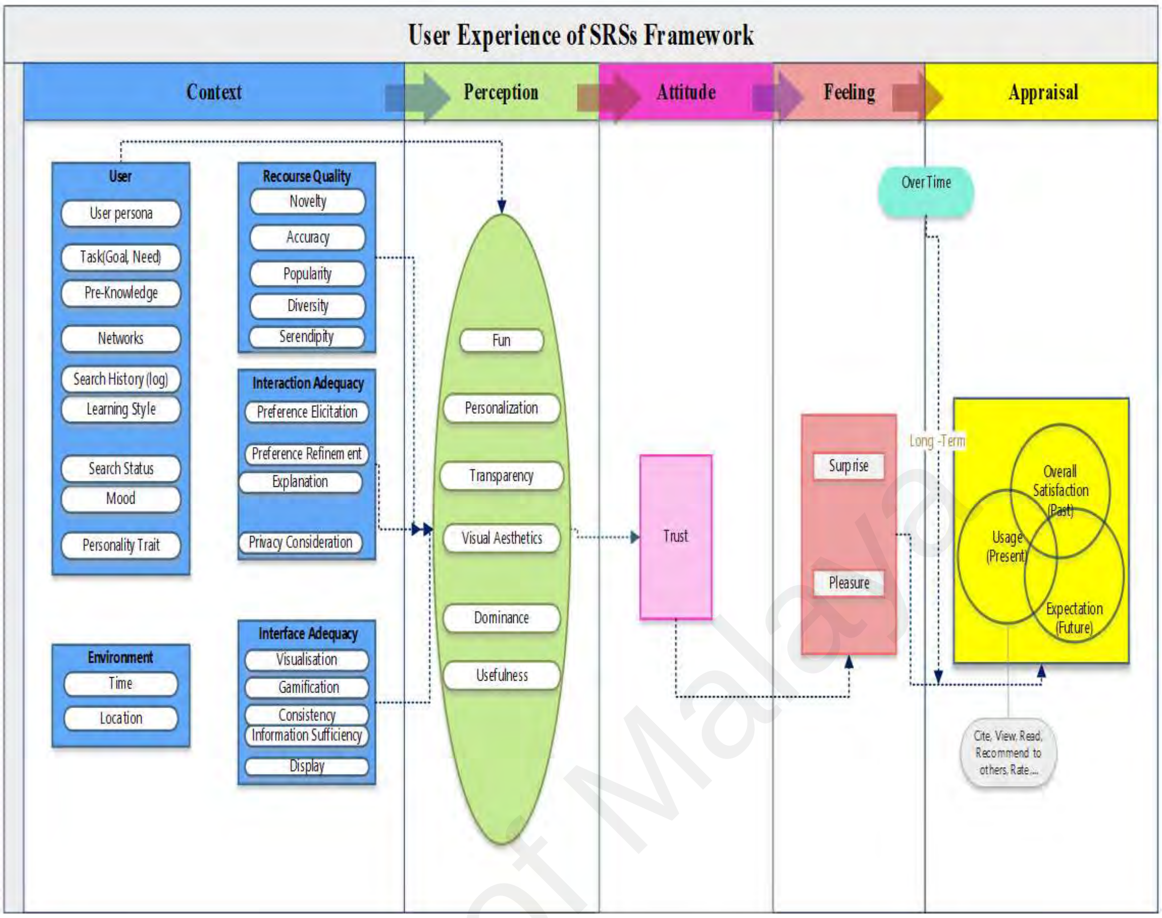


Figure 4.14: Initial conceptual framework

The initial conceptual framework is also validated through the expert review technique explained in the next section.

4.1.4 Expert Review

Before the empirically examination, the proposed conceptual framework is also validated through expert review technique. Five experts attended the review sessions and demographic profiles of the experts are as displayed in Table 4.5.

Table 4.5: Demographic Profiles of Experts

No.	*Domain of RS*	*Field of Expertise*	*Experience (Year)*	*Country*
1	Movie	RS , HCI	5	Germany
2	Digital Library	RS, UM, HCI	10	Iran
3	Movie	RS, ML, NLP	8	Italy
4	Tourism	RS, HCI	6	Malaysia
5	Movie	RS, ML, UM	5	Malaysia

RS: Recommender System, HCI: Human- Computer Interaction, UM: User Modelling, ML: Machine Learning, NLP: Natural Language Processing

The 13 experts have been selected through the ResearchGate, LinkedIn and Google Scholar. From 13 invitations, only five experts accepted to attend the review sessions. The selection of expert reviewers are based on the following criteria:

- ✓ Have the PhD degree in Computer Science , Software Engendering or related fields
- ✓ Have research background in RSs or SRSs
- ✓ Have knowledge of framework construction in the above mentioned field
- ✓ Have at least five research publications in RSs or SRSs

4.1.4.1 Instrument and Procedures

The aim of this expert review is to validate the proposed conceptual framework, its components, indicators and relationships. An initial figure of the framework along with a questionnaire which examines the following features have been provided for the experts:

1. The clarity of the terminologies used in the conceptual framework
2. The logic relevancy of proposed components (5 components)
3. The logic relevancy of proposed indicators/indicators in each component (36 indicators)
4. The connections and flows (12 relationships)

5. The usability of the framework to show how contexts influence UX of SRSs
6. The overall readability of the conceptual framework

The interviews were conducted online through Skype while face-to-face interviews were conducted with the participants from Malaysia. The experts were asked to validate if the above mentioned component and indicators along with the relationships are logically relevant of not relevant in the process of UX of SRSs by giving “yes” if they agree with and “no” if vice versa. Lastly, the experts were also asked to add their further comments and any suggestions for the parts with which they do not agree in the provided questionnaire. At the beginning of the interview sessions, the background of study along with the proposed framework explained and introduced briefly by the researcher. The review session involved two-way interactions, where experts may ask questions and give their opinions of the focused matter. Afterwards, the experts were required to answer all the questions. The expert review findings are discussed in the next section.

4.1.4.2 Findings

In this section, only the results that need to be revised are explained and visualized.

(a) The clarity of the terminologies used in the conceptual framework

The majority of the experts agreed that the most selected terminologies are clear enough. They only revised two terminologies that listed in Table 4.6.

Table 4.6: Experts recommended terminologies

Component	***Unclear -terminology***	***Experts recommendation***
User Context	Search history(logs)	Info Seeking Behaviour
Feeling	Feeling	Overall Feeling

(b) The logic relevancy of the proposed components (5 components)

As shown in Figure 4.8, the majority of experts were unanimous that two components of "Attitude" and "Feeling" should be merged under one component. Therefor the "overall feeling" was chosen for the final revision. Consequently, the all indicators of "trust", "pleasure" and "surprise" were moved under the "overall feeling" component. After the experts reviewing, Users' intentions are highly influence by users' feeling (Y. Chen & Pu, 2011). Regarding the other components, the majority of the experts agreed that the proposed conceptual components are well relevant in the proposed framework.

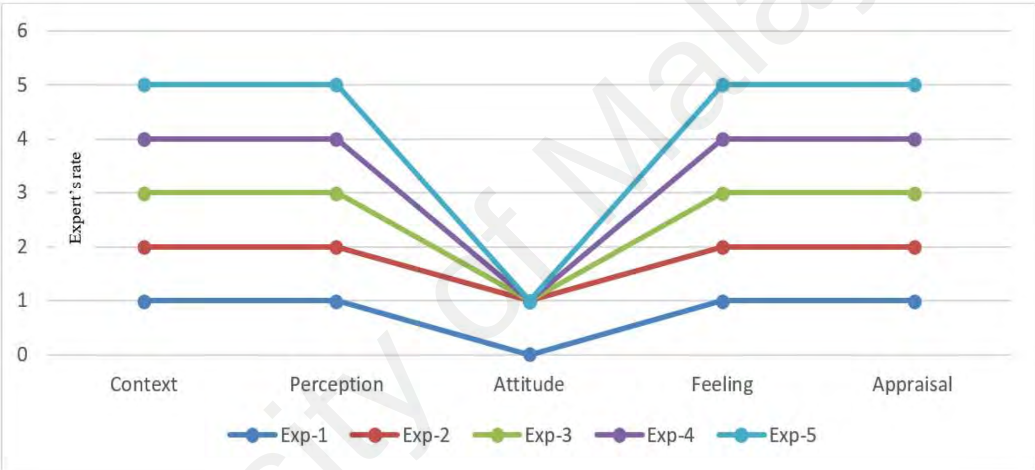


Figure 4.15: Relevancy of components

(c) The logic relevancy of indicators in each component

For the relevancy of indicators, the majority of the experts agreed that the most indicators are relevant however as shown in Figure 4.9, two indicators of "Information Sufficiency" and "Serendipity" under the "System Situation Contexts" are not relevant. According to the reviewers' recommendations the indicator of "Serendipity" should be removed because the indicator of "Novelty" is more clear and appropriate to explain the paper quality. Also, four experts recommended that the indicator of "Information Sufficiency " is better to be moved under the "IxD" component because in overall, "Information Sufficiency" refers to a piece of information about the recommendations that users would like to see once the recommendations are presenting to them and it recommended to be categorized under the “IxD ”. The above mentioned indicators were revised in the final version of framework.

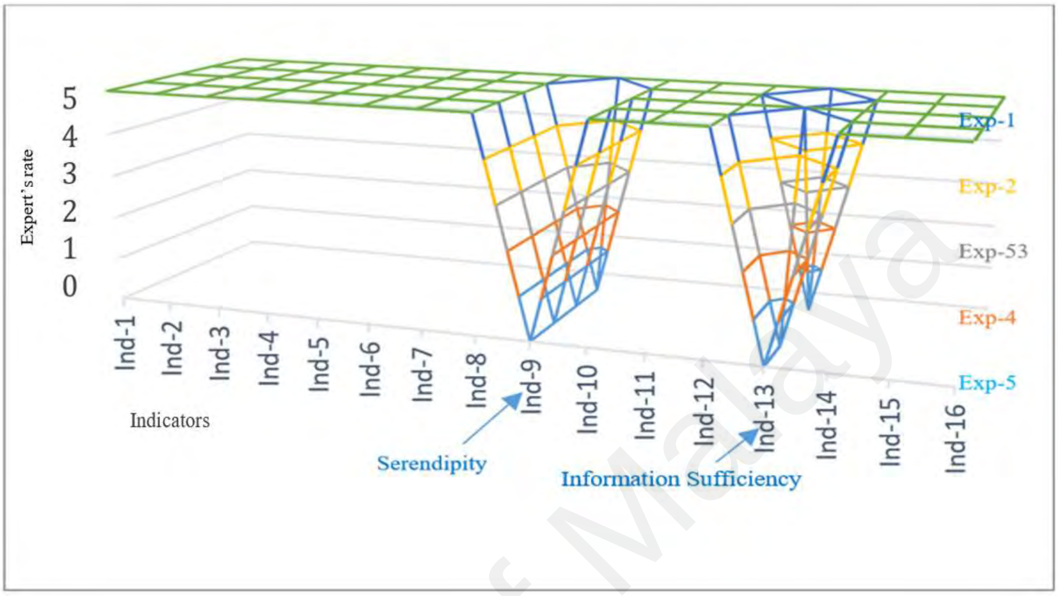


Figure 4.16: Relevancy of system context indicators

Apart from the above-mentioned indicators, the experts also recommended a few indicators to be add into the components. These indicators have not trace back in the literature of RSs however, they have been recommended by experts to be included in the initial framework for further empirical examination. The examination of the recommended indicators in the field of RSs is quite new. The added indicators based on the experts’ point of views, are briefly listed and explained in Table 4.7.

Table 4.7: New indicators recommended by the experts

Component	***New indicator***	***Description***	***References recommend by experts for more info.***
User situation	Reasoning methods	Each person might have particular reasoning method to apply the exiting knowledge for making judgements, predictions, and explanations or drawing conclusions. Three methods of reasoning are the deductive, inductive, and abductive	(Shneiderman, 2004). (D. Norman, 2013b) (Sweller, 1994)
IxD adequacy	Dialog	The aim of dialogue design is to yield closure in order to prevent users to think much or guess what to do or what is the next action. The dialogue make the system fool-proof as possible by for example using messaging box, flags, and icons	
UiD adequacy	Signifier	Signifiers are communication signs or signals that tell users what this object for. Signifiers help users to understand what to do and where to do it, what is happening and what is the alternative therefore they convey the individual to the possible action. The signifier should be clear for the user who wants to use the product so it is easy to understand	
Perception	Interactivity	It is about building systems and platforms that allows interaction between product/service and its users. It aims to build meaningful relationships between people and the product/services they use	
	Affordance	When users interacting with a product, if they perceive affordance, it means that the UI design's clues or identifiers are visually clear enough to guide users what to do. Some of the studies in the field of RS have used term of "Ease of Use" which is defined as "the degree to which a person believes that using a certain system would be free of effort". The affordances are based on real-world experiences or standard UI conventions.	
	Cognitive barrier	When users interacting with SRSs, if something temporary put a stop to their actions required to complete in order to gain their goal, it called cognitive barrier. For example, when the user is not able to find the desired information and the system has not a good help option. The less the user perceives cognitive barriers interacting with RSs, the more they have a better experience	
	Cognitive load	It is about the amount of the required memory being used by the working memory of the user to achieve his goal. The less cognitive load and cognitive barrier are the more the positive felling such as pleasure is discerned by the users while interacting with the system	

*** It is pointed out that the description of the new indicators are based on the experts' explanations and the references recommended for further information by the experts. As mentioned, these indicators have not trace back in the literature influencing UX of RSs.

(d) The connections and flows (relationships)

In the initial framework, a total of 8 relationships were defined between the components and sub-components. According to the results of experts review, the whole relationships are well-linked but the relationships between the "Attitude" and "Feeling" should be removed because of the integration of these two components. Three experts pointed out that there are interrelationships between the contexts therefore investigation of following impacts is recommended by the experts and added to the initial framework.

1. The impact of the environment context on the user context
2. The impact of the user context on the paper quality
3. The impact of the user context on the interface adequacy
4. The impact of the user context on the interaction adequacy

They might be a few relationships between some of the indicators, for example, signifier might influence, affordance however such relationships are not examined in this research.

(e) The usability of the framework to show how contexts influence UX of SRSs

From the findings depicted in Figure 4.17, it can be concluded that whole experts agreed that the proposed framework is overall useable in terms of showing how contexts influence the UX of SRSs however two of the experts remarked on the addition of the above-mentioned relationships to make the whole process more useable.

Figure 4.17: Usability and Readability of the framework

(f) The readability of the conceptual framework

Like usability, all the overall readability, which is the ease of framework understanding, was acceptable by the experts (Figure 4.17).

4.1.5 Revised conceptual framework

As discussed in the previous section, before the empirical examination, the framework was reviewed by five experts in this field and after applying the comments and recommendation of experts, the framework was revised as depicted in Figure 4.18.This is important to add, the time (context) characterises different situations. For example, it might represent the time a user spends to receive recommendations which influence UX, in this case time is considered as a systemic context not environmental context (Beel & Langer, 2014). Moreover, there are other factors such as accessibility of recommendations. For examples, in most systems, recommendations are free however some SRSs such as Mendeley provides paper recommendations only as a premium service (Beel & Langer, 2014) which are not free and unappealing for the users. Also, the impact of the accessibility of recommendations was excluded from this research. As indicated earlier, this is assumed that the perceptions mediate the impact of contexts on users' feelings. In this framework, context influence user's perceptions and there are interrelationships between the contexts.

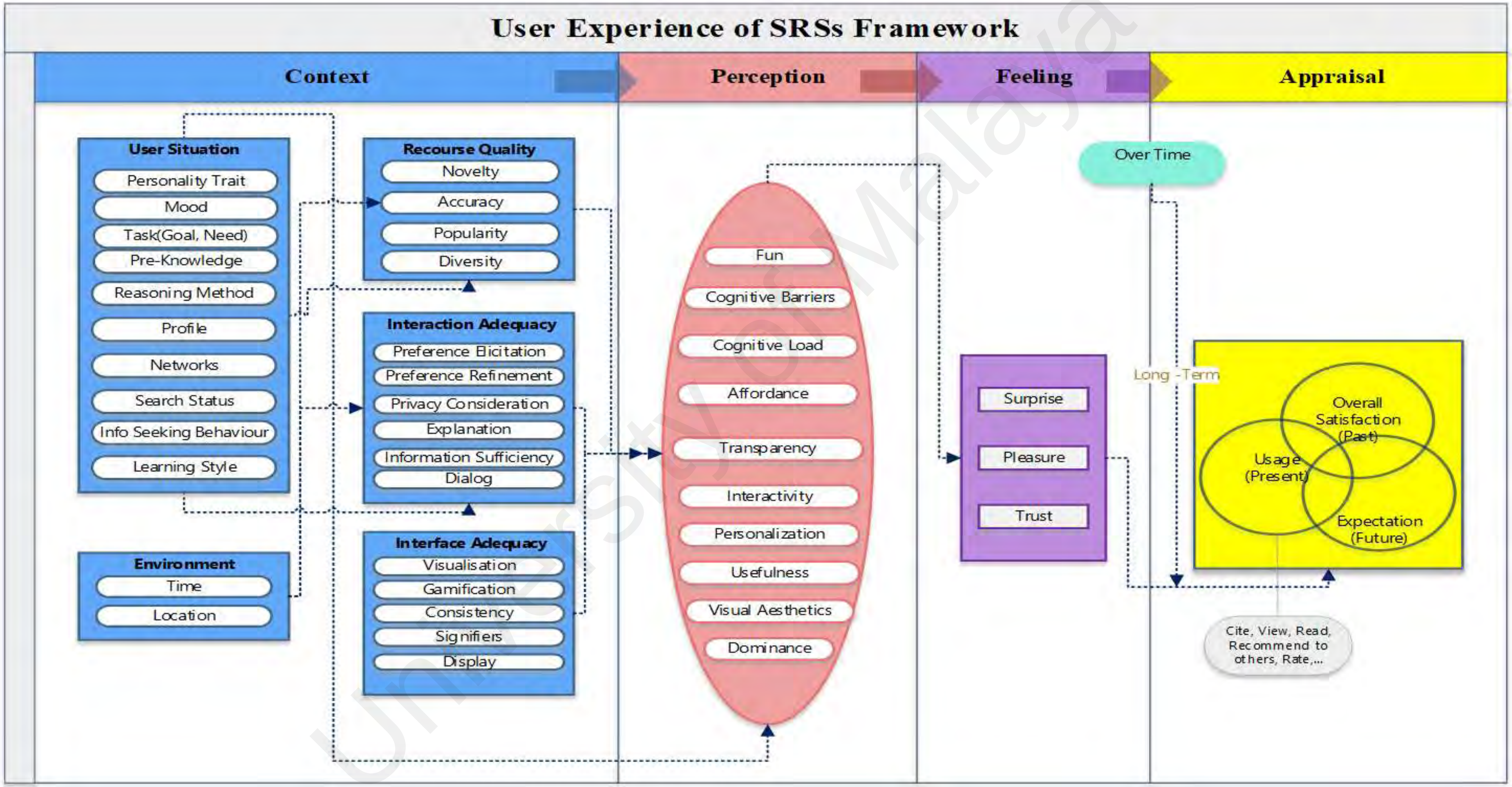


Figure 4.18: Revised conceptual framework

After the framework has been generated and explained theoretically, it is the time to empirically validate it and the hypothesized relationships between its components. In the following, the steps that test the framework has been discussed. Table 4.8 shows the terminology used in this section. This terminology helps to understand the model specification and data analysis using PLS-SEM method. Additionally, different terms have been used for a single concept. For example, indicator, item, measure and scale have the same meaning.

Table 4.8: Terminology used in the empirical section

Term	*Definition*
Observable variable (indicator/ item/ measure/scale)	A variable that can be directly observed and measured even with a single indicator
Latent variable (construct)	A theatrical concept that cannot be directly observable or measured but can only be inferred from observable variable
Mediator variable	This variable explains the relation between two variables and why there is a relation
Moderator variable	This variable affects the direction and/or strength of the relation between two variables
Formative construct	The latent variable which is formed or induced by its indicators and convinced as a composite of these indicators. Changes in the indicators influence variable
Formative indicator	The indicator of formative construct
Reflective construct	The latent variable which is reflectively measured and changes in the construct are reflected manifested by its indicators. Variation in the leads to variation in to its indicators
Reflective indicator	The indicator of reflective construct
First- order construct (uni-dimensional)	The construct is first-order where the it is assessed directly using a number of indicators
Second-order construct (multi-dimensional)	The construct is second-order where the it is represented by a number of dimensions or sub-constructs which are assessed by a number of indicators

4.2 Empirical test of framework

This section aims to respond to RQ2 which indicates how does contextual information empirically influence UX with SRSs? To do so, the following sub-questions of are addressed:

SRQ2.1: Are indicators empirically valid?

SRQ2.2: Are constructs (components) empirically valid?

SRQ2.3: Are relationships between the constructs (components) empirically valid?

SRQ 2.4: What is the framework GOF?

SRQ2.4: What are the most relevant contexts influencing/contributing into UX of SRSs?

The quantitative method of PLS-SEM is used to empirically test the framework. This method is a predictive technique (Pratley et al., 2014) and useful for exploratory research objectives (Hulland, 1999)where theory is less developed (Henseler et al., 2014)a and studying a phenomena is relatively new(Sharma & Kim, 2012). In addition, it helps to estimate relationship models with latent variables which are mostly subjective and not directly measurable. Also, a few pre- processing assumptions such as large sample size and normality have been avoided in this method (Hair et al., 2014; Urbach & Ahlemann, 2010). This method also can be considered as one of the useful methods for data or dimension reduction to reduce the complexity and ambiguity in the system. Besides, in comparison to the methods that can analyses only one layer of linkages between variables at a time, the PLS-SEM method has this advantage to answer a set of interconnected enquires in a single, systematic and comprehensive analysis (Hair et al., 2011). In other words, using PLS-SEM, a single run of analysis algorithm can calculate simultaneously both the measurement model (the correlation between the measurement indicators and their related construct) and the structural model (the conceptualized linkages between the various constructs in the research model) (Pratley et al., 2014). Therefore, this method is useful for data or dimension reduction to reduce the complexity and ambiguity in the system. As illustrated in Figure 4.19, five multi-stage processes have been applied.

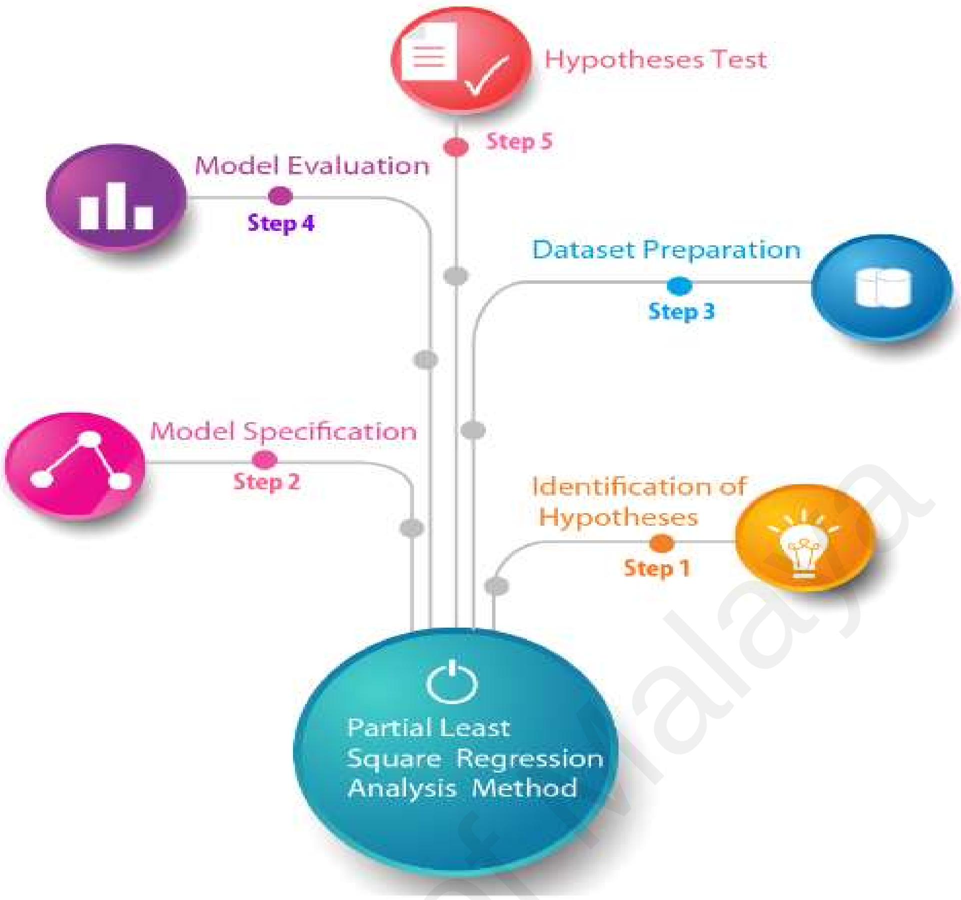


Figure 4.19: PLS-SEM procedures

In the following, the PLS-SEM process along with the activities accomplished in each process are explained.

4.2.1 Identification of hypotheses

In the previous section, the theoretical basis, adoption of main component and identification of variables of the framework was discussed and theoretically justified the relationships drawn between the various latent variables and concepts leading to the proposition of relevant hypotheses. In Table 4.9, the hypothesized relationships are summarised.

Table 4.9: Hypotheses of this research based on the literature

No	*Description*
H_1	The user's context influences resource's context toward the user experience with SRSs
H_2	The user's context influences interaction's context toward the user experience with SRSs
H_3	The user's context influences interface's context toward the user experience with SRSs
H_4	The user's context influences user's perception toward the user experience with SRSs
H_5	The environment's context influences resource's context toward the user experience with SRSs
H_6	The environment's context influences interface's context toward the user experience with SRSs
H_7	The environment's context influences interaction's perception toward the user experience with SRSs
H_8	The environment's context influences user's perception toward the user experience with SRSs
H_9	The resource's context influences user's perception toward the user experience with SRSs
H_{10}	The interface's context influences user's perception toward the user experience with SRSs
H_{11}	The interaction's context influences user's perception toward the user experience with SRSs
H_{12}	The user's perception influences user's feeling toward the user experience with SRSs
H_{13}	The user's feeling influences user's appraisal toward the user experience with SRSs
H_{14}	The long term moderates the relationships between feeling and appraisal toward the user experience with SRSs

The hypotheses examine the impacts of contexts on users' perception (H_1-H_{11}), perception on users' feelings (H_{12}), feeling on users' appraisal (H_{12}) and finally the impact of long- term variable on the users' appraisal (H_{14}).

4.2.2 Model Specification

Based on the PLS-SEM method guidelines (Hair et al., 2014), the second step is modulation or model specification made of two sub-models of structural model (inner model) and measurement model (outer model).

4.2.2.1 Specification of inner and outer models

The structural model displays the relationships (paths) between the constructs or latent variables. The measurement models display the relationships between the latent variables and indicators (measurable variables) (Vinzi, Chin, Henseler, & Wang, 2010). The inner and outer models are set up based on the inspiration from the literature as they were elaborated previously. Empirically, the inner and outer model relationships are examined by using PLS. As Figure 4.20 depicts, user, resource, environment, interface and interaction contexts along with the perception, feeling and appraisal compose the latent constructs that build up the structural or inner model (Grey circles and relationships between them). Each of aforementioned constructs along with its indicators (Green rectangular) establishes measurement or outer model (Grey circle and green rectangular). The pink circles show the impact of moderator variable of "over time" on the relationships between "feeling" and "appraisal" variables. The hypothesised relationships also have been shown.

4.2.2.2 Determination of formative and reflective constructs

Latent variables can be measured by two different ways of reflective and formative measurement. The main goal in the examination of construct conceptualization is to specify whether the construct should be conceptualized as a formative or a reflective construct. Based on the instructions mentioned by (Hair et al., 2013; Henseler, Ringle, & Sinkovics, 2009), if the indicators cause the constructs meaning that causality direction or arrows point is from measure (indicator) to construct(latent variable) then the relationships are formative. In the formative constructs, the indicator(s) form the construct therefore they are conceptually not-correlated and independent components of a construct. After reviewing the whole construct and indicators as well as consulting statistical advisor, the conceptualizing of the latent variables in the modulation are multi-dimensional formative constructs.

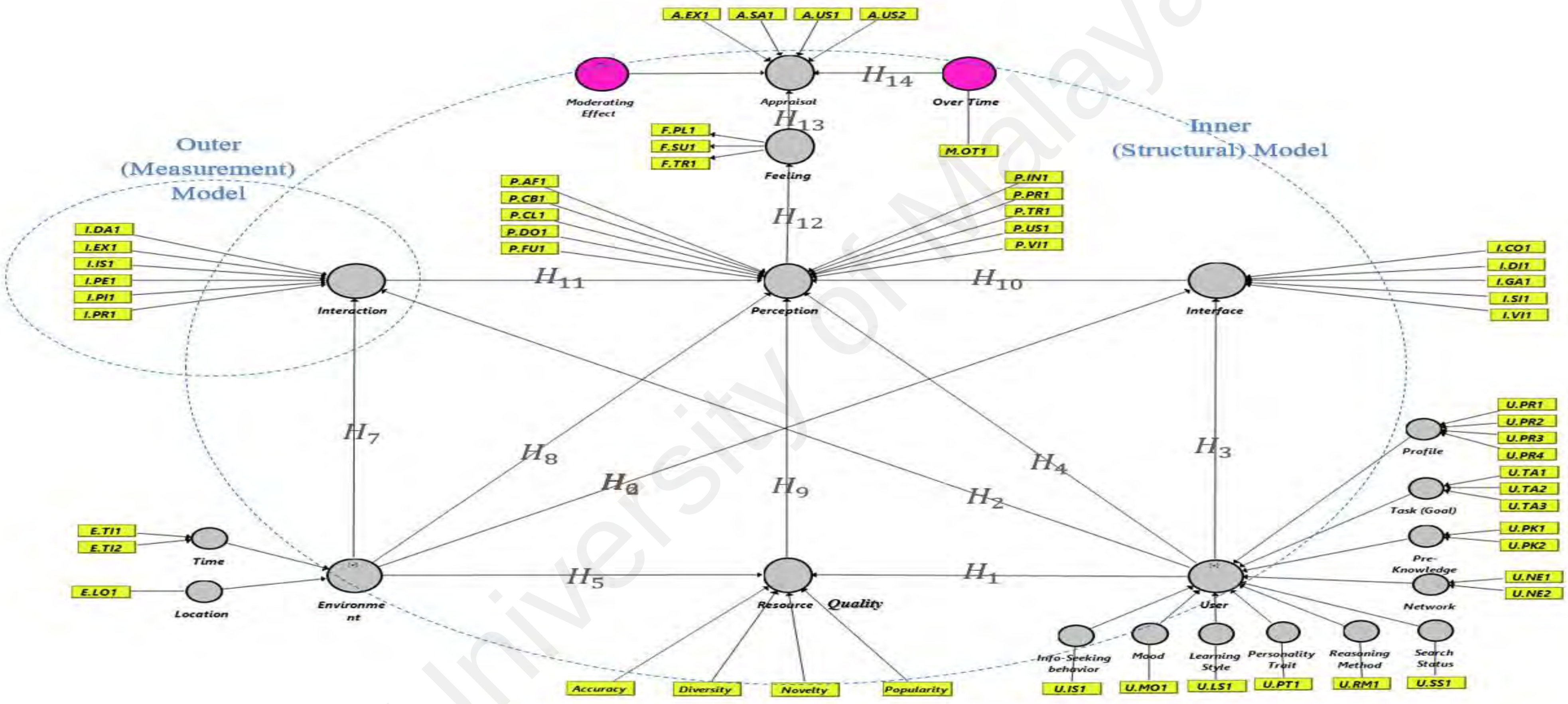


Figure 4.20: Model Specification (Inner and outer model)

Therefore, based on the PLS regression equation, this is;

$$Y = \beta_1 X_1 + \cdots + \beta_n X_n + \gamma \qquad (4.1)$$

Where Y = the formative construct being estimated , $\beta_i =$ $Beta\ wights\ for\ indicators$, $X_i = indicator\ scors(observation), \gamma =$ $a\ disturbance\ term$

4.2.3 Dataset preparation

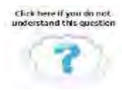
Before evaluating the validity of proposed model and examining the research hypotheses, the preparation of the adequate dataset is required. A few activities discussed in the following sections have been performed to prepare the dataset.

4.2.3.1 Measurement of latent variables

Generally, for meaning the conceptual variable (latent variable), there are two methods. One is to use existing measurement scales from the literature if there are and the other is to create new measures. The use of exiting scales is recommended and has the advantages to save time and cost of study since other studies have already created reliable and validated scales. However, if there are not reliable variables scales, the creation of new scales is equally recommend because the effort for creation of new scales add an extra contribution to the research field that can be applied in the future studies(Straub, Boudreau, & Gefen, 2004).

To decide the best scale for each of our latent variables, a thorough literature review of existing measurement scales was conducted. The past studies of Pu and Chen (Pu et al., 2012a) and Knijnenburg et al (Bart P. Knijnenburg et al., 2012) provided extensive questionnaires to test several concepts related to UX of RSs. The scales which are similar to this research concepts are utilized. However, because the scales are for a scholarly RS domain and UX is highly contingent upon the purpose of the system (Bart P. Knijnenburg et al., 2012), the necessary wording modifications were made to the original scales to suit them to the measurement needs of this study without affecting the original conceptual bases of such scales. In addition, past studies mostly do not include latent variables of contexts therefore in this study, these scales were developed.

In this research, most of the second-ordered constructs such as learning style, reasoning method are measured by single scales. The reason of measuring them by a single scale are as follows;

1. Too many questions make participants to quit the survey before completing all questions.
2. The length of survey should be short enough that participant pay full attention to respond to all questions throughout the survey. Too many questions distract the participant’s attention and makes biased results.
3. Some of the constructs can be measured by an overall or general item such as satisfaction.
4. For all the single scales whenever it is required, the question is clarified by presenting a few examples or specification. In other words, instead of creation more questions to measure an item, the aspects of an item is exemplified and specified in order to transfer a valid and consistent concept to the participants.
5. Also, more clarification has been made by a question mark icon 

for each question on the online web application. The more information is presented if the participant presses this icon.

4.2.3.2 Examination of measuring tool

Before the actual analysis, it is needed to make sure that the test instrument or measuring tool is valid and consistent. For examining the validity, Pre-test and for examining the consistency, Pilot-test have been performed (Figure 4.21). Obviously, it is important to perform the validity examination before the actual analysis since in the analysis phase it is too late to change the experiment in order to obtain better validity.

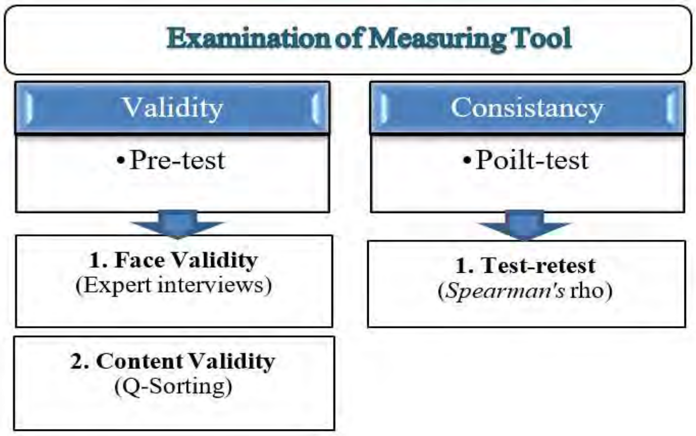


Figure 4.21: Measuring tool examination

4.2.3.3 Pre-test: indicator screening and validation

Pre-test examine the validity of the test instrument or measuring tool. After preparing the list of the indicators list which already confirmed theoretically, the validity of indicators empirically were examined and also, a few scales were developed that need to be validated. Two validities have been performed; 1. Face validity that can be examined by the expert interviews and 2. Content validity by using user card sorting exercises (also called Q-Sorting).

(a) Face validity control

According to Lewis et al. (Lewis, Templeton, & Byrd, 2005), "the early and prudent use of experts in the design of philosophical elements can expedite scientific progress and make construct development projects more efficient". Therefore, it was important that experts in the field get involved before any empirical validation with the potential survey. The aim of this experts review is to check the validity control of the selected indicators. All indicators were reviewed by three experts from the Human- Computer Interaction and UX fields to check the indicators linguistically and conceptually. Table 4.10 describe the demographic information of experts involved in the face validity control.

Table 4.10: Experts' profile

No	*Experts' Research interests/ profession*	*Field*	*Country*
1	Recommender Systems, Multimedia Retrieval, Active Learning, Human Computer Interaction	Academia	Italy
2	Human Computer Interaction, Recommender Systems	Academia	Malaysia
3	Recommender Systems, Human Computer Interaction, User Experience Research, Personalization Technology, User Modeling	Academia	Switzerland

The experts were asked to score (from 1 to 5) to some extent they think the indicator can measure the construct. Figure 4.22 shows that the content of indicators are valid and transparent from the experts' point of views.

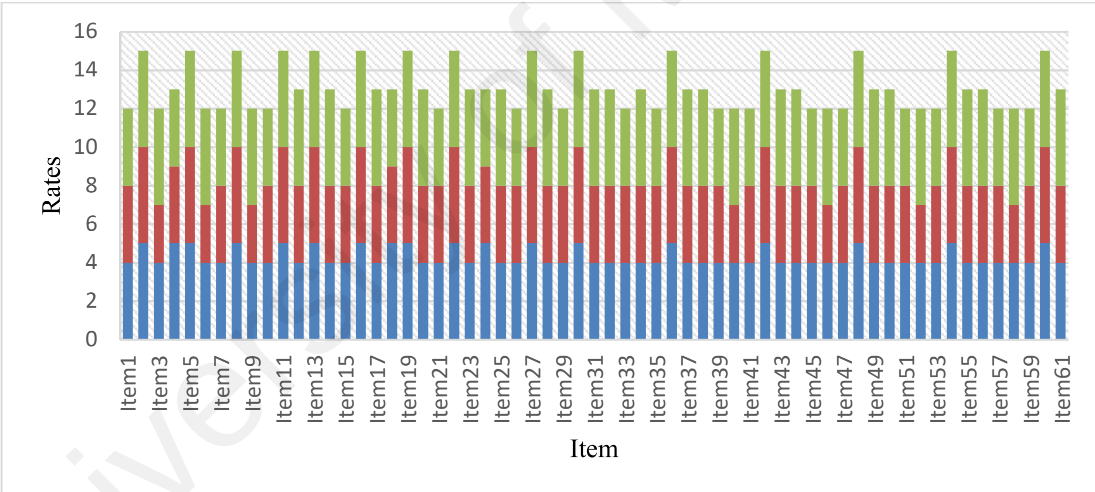


Figure 4.22: Face validity of indicators from the experts' views

From this review, all indicators found to be relevant and valid based on the conceptual framework and the design of the study have been shown in Figure 4.22. Some indicators or statements, however, needed some language improvements. Such improvements were made and all indicators were taken to the next stage of the scale development process. Also the indicators which entirely developed in this study were separated and ask two more experts to assure the face validity.

(b) Content validity control

The method used to check empirically if the indicators can be classified in a certain group or class is Card Sorting or Q-sorting (Moore & Benbasat, 1991). The Q-sorting was initially developed by William Stephenson (Stephenson, 1953) in order to inspect peoples' views about a target topic.

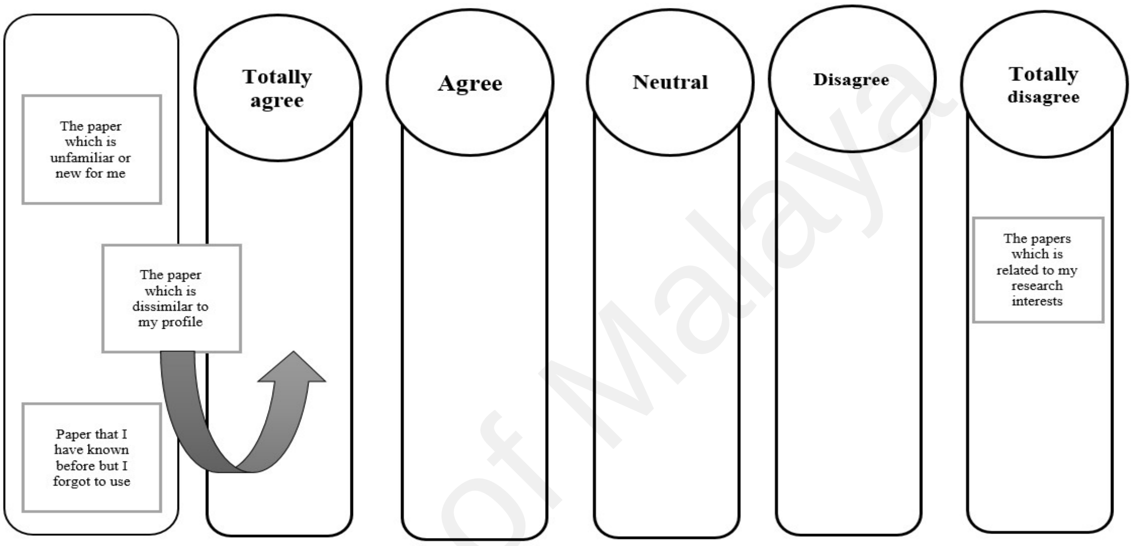


Figure 4.23: Sorting for the novelty construct

The Q- sorting is used to double-check the validity of the initial conceptual classification of the pre- prepared statements using participants' feedback and also to discover any wording or language issues that made a misunderstanding for participants. The cards were provided and selected 15 researchers (5 master/ 5 PhD students/ 5 lectures) from the Faculty of Computer Science and Information Technology (FSKTM) at the University of Malaya at random in order to participate in this survey. At the beginning of each Q-sort, a summary of the research and the how to perform the experiment was explained to the participants. Then, they were asked to read the statements and sort a per-determined number of statements (Q-sort desk which composed by cards) into a set of boxes (close sort). The sorting test was designed in the way that only a fixed number of cards can be sorted in a box; for example, in the box labeled

"*Novelty"* (Figure 4.24) only four cards can be selected. This guarantees that participants give attention to choose carefully those four cards they completely agree with. It was made clear to the participants that each card had to be related to one, and only one box. If a card can fall into more than one group, the participant had to decide about which group the card fits best within. The participants were asked to express their opinion in a Likert scale that to how extent they agree with the statement for a particular constructs (columns in Figure 4.7). The participant was also requested to ask if they have any question during the sorting. Once the participant was done classifying all the statements, the data was analyzed quantitatively by carefully reviewing and reading the contents and participants notes if there was any. The insights gained from the Q-sort analysis led to revising and removing a few indicators such as that were ambiguous and multi-interpretable to participants. Their feedback and comments were thoughtfully taken into consideration and enhancements were made to the instrument.

4.2.3.4 Pilot-test

The aim of the Pilot-test is to examine the consistency of the test instrument. The test-retest reliability was performed by running a same questionnaire twice over a period of two weeks by the 15 scholars who already accepted to participate two times. Table 4.11 describes the profile of 15 scholars.

Table 4.11: Scholars who accepted to participate in the Pilot-test

No	*Scholar's role*	*Number*	*Country*
1	PhD Students	4	Iran, Malaysia, Canada
2	Master Students	5	Iran, Malaysia, Canada
3	Post-doc	4	Iran, Malaysia,
4	Faculty	3	Iran, Canada

To examine the consistency of indicators, the Spearman correlation coefficient (Spearman's rho) was applied and tested the reliability of all indicators. Table 4.12 summarizes the results of Spearman's rho test.

Table 4.12: Correlation between test and re-test survey for all indicators

Indicator	r	Indicator	r	Indicator	r	Indicator	r
Q-AC1	0.783	U-SS1	0.938	U-PR1	0.950	I-CO1	0.919
Q-AC2	0.811	U-MO1	0.890	U-PR2	0.883	I-VI1	0.955
Q-PO1	0.826	U-PT1	0.944	U-PR3	0.930	I-GA1	0.902
Q-PO2	0.793	U-RM1	0.985	U-PR4	0.725	I-SI1	0.954
Q-PO3	0.813	U-LS1	0.970	U-PK1	0.822	I-DI1	0.950
Q-PO4	0.883	U-IS1	0.951	U-PK2	0.902	P-CB1	0.906
Q-NO1	0.843	E-TI1	0.977	U-TA1	0.865	P-AF1	0.922
Q-NO2	0.874	E-TI2	0.980	U-TA2	0.888	P-CL1	0.909
Q-NO3	0.871	E-LO1	0.700	U-TA3	0.964	P-PR1	0.943
Q-NO4	0.726	I-PE1	0.712	P-FU1	0.895	F-PL1	0.862
Q-DI1	0.870	I-PR1	0.777	P-IN1	0.853	F-TR1	0.859
Q-DI2	0.926	I-EX1	0.948	P-US1	0.896	A-SA1	0.867
Q-DI3	0.911	I-IS1	0.912	P-VI1	0.849	A-EX1	0.897
U-NE1	0.841	I-PI1	0.909	P-TR1	0.884	A-US1	0.983
U-NE2	0.879	I-DA1	0.938	P-DO1	0.854	A-US2	0.986

According to the results all correlation coefficients were above 0.7 which indicates that all indicators can be reliable and there is no confusing or correlated item. Figure 4.25 also visualizes the correlation coefficients values.

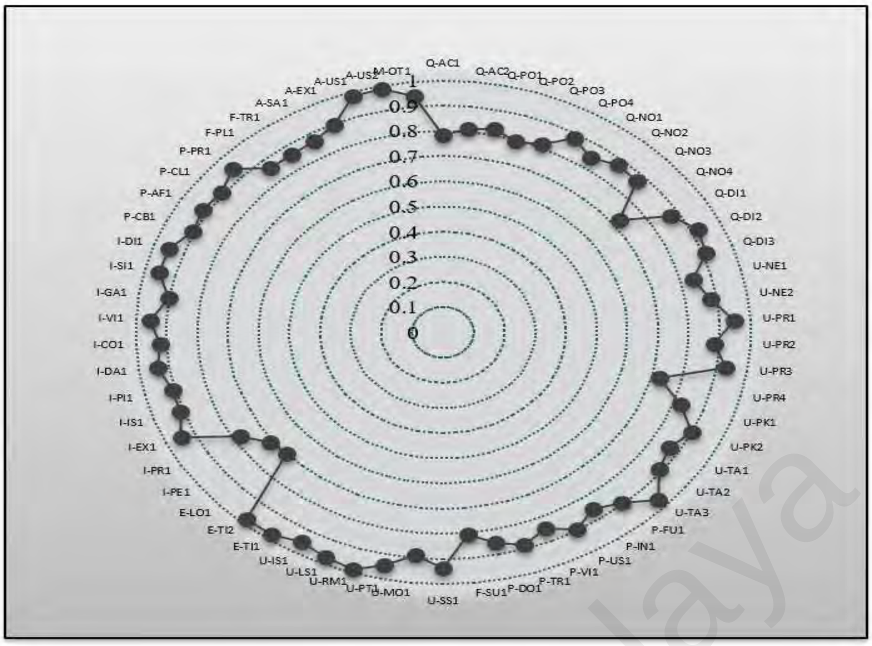


Figure 4.24: The result of correlation between test and re-test

After examining the validity and consistency of the indicators, a web application was developed which enables the participants to express their feedbacks in 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). To assess the appropriateness of the whole process of data gathering by a highly controlled sample and knowledgeable about the construct under investigation, the link of application was sent to one of the experts who participated to the face validity control and asked to check the clarity, terminology and understanding of indicators in the web application. Hence, it was made sure that the final instrument is ready for use in the data collection stage. The final version of the questionnaire contains the Plain Language Statement (PLS) that describes to participants the research objectives, the questionnaire format and content, approximate completion time, operating instructions for the online system, confidentiality provisions for information collected, and contact information. This questionnaire along with the web application are presented in Appendices G and H.

4.2.3.5 Population

This study targets the population of computer science scholars in three different countries of Malaysia, Iran and Canada. Table 4.13 describes the demographic information of final participants of this study. On average, the participants spent about 30 min to respond the questions.

Table 4.13: Demographic profile of participants

	Indicators	*Frequency*	*Percentage*
Gender	Female	85	.483
	Male	73	.412
	Not- mentioned	19	.102
Group	PhD	53	.293
	Master	69	.385
	Post-doc	31	.175
	Faculty	24	.136
University	UM	49	.277
	UTM	24	.138
	UBC	31	.179
	SFU	28	.152
	UOI	26	.144
	UOT	19	.102
Country	Malaysia	73	.413
	Canada	59	.334
	Iran	45	.253

The bachelor students in this examination were excluded because it was assumed that under-graduate students are not seriously involved in research and scholarly tasks such as finding appropriate papers. Besides, for this study that the academic people are needed who have the experience of working with Bibliographic Databases (BDB) such as science direct, web of knowledge and also know what a SRS is and what it is for. The computer science community was selected because of two main reasons; first the availability of this community in the locations where this research has been conducted and second the necessity of well-acquainted participants with bibliographic databases.

4.2.3.6 Data gathering & data pre- processing

The Questionnaire went live on the "www.rscholar.com\quest" server from the 24th of September 2015 until the 28th of March 2016(Almost 6 months). To encourage participation, an incentive was provided in the form of $ 5.00 USD Starbucks Card eGift to spend on the favorite drink. Because the Starbucks are not available in Iran, for the Iranian participants, this amount has paid via the Iran's bank eGift cards. During the time range of data gathering, a total of 177 useful responses were received and used in the data analysis. As Hair et al. indicated the PLS method is not restricted by the sample size and normality distribution (Hair et al., 2014). However, based on suggestion of minimum sample size table in Appendix I by Hair et al., the maximum sample size for the minimum R^2 in the field of Engineering is 189. Therefore, the sample size of 177, is an adequate sample size for this study. As the online web-based application was used to collect the data, the controls of not responded questions in terms of missing values already have been checked. Preparation for analysis also involved checking for errors in the data and correcting them before any further analysis is accomplished. Using the SPSS software, descriptive statistics were calculated for all variables in the questionnaire to detect out-of-range or erroneous data entries. As a result of this check, no data entry errors were found.

4.2.4 Empirical results

Once the inner and outer models were specified and the dataset was prepared, in the next step, the Partial Least Square algorithm was run using the SmartPLS3 tool (*www.smartpls.com*) to estimate the research model and test the hypotheses. The empirical analysis of the proposed framework consists of four experiments of 1) The assessment of the measurement model, 2) The assessment of the structural model, 3) The examination of hypotheses and 4) The examination of model goodness of fit.

4.2.4.1 Assessment of outer model

In this section SRQ2.1 is addressed which investigates the validity of the indicators. In the PLS-SEM method, it is performed by assessment of measurement model. This model is a formative construct model and based on the guidelines by (Hair et al., 2014), the measurement model was assessed by examining the Multicollinearity and Significance and relevance of indicator weights (outer weight). The flow of how an item/indicator is retained or removed is shown in Figure 4.25.

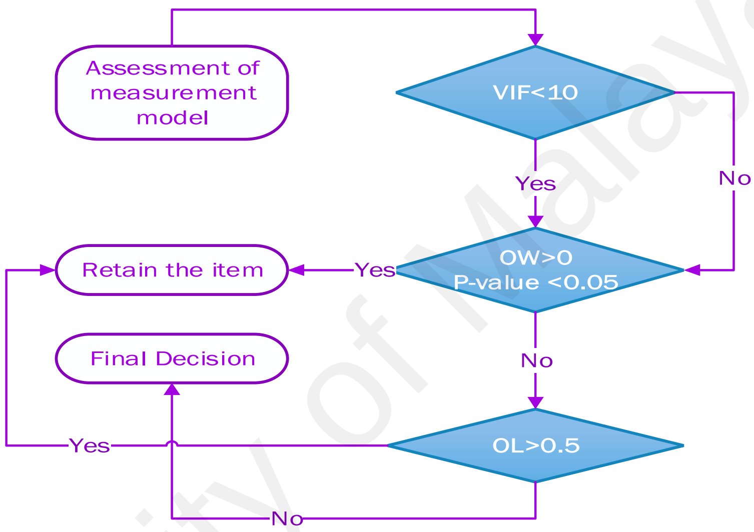


Figure 4.25: Assessment of measurement model

(a) Multicollinearity

Multicollinearity indicates the extent to which an independent variable varies with other independent variables; excessively high multicollinearity would challenge the statistical assumption that the independent variables are truly independent of one another(Hair et al., 2014). The Variance Inflation Factor (VIF) is a metric for multicollinearity (Hair et al., 2014). The term VIF is derived from the fact that its square root is the degree to which the standard error has been increased due to multicollinearity.

Amount of the R^2 for each item is equal to the square of the load factor between the Construct and the Item. As a rule of thumb, the VIF should not exceed a value of 10 (Hair et al., 2014).

$$Foreach\ item \quad VIF = \frac{1}{1 - R^2} \tag{4.2}$$

If the indicators do not pass the significance level for VIFs, the item is not removed from the model because the absolute contribution of formative indicators is assessed outer loading, which is always provided along with the item weights. Therefore, after measuring the item weight, the decision to remove the indicator or keep the item is made as shown in Figure 4.27. As Table 4.14 indicates, the resulting of the VIF for the constructs after running the PLS algorithm, are all lower than 10 which indicates the absence of multicollinearity in the indicators.

Table 4.14: Overview of VIFs of outer model (formative indicators)

Indicator	VIF	Indicator	VIF	Indicator	VIF	Indicator	VIF
Q-AC1	1.374	U-PR1	1.070	U-SS1	1.027	I-CO1	1.082
Q-AC2	1.595	U-PR2	1.170	U-MO1	1.216	I-VI1	1.153
Q-PO1	2.234	U-PR3	1.301	U-PT1	1.435	I-GA1	1.184
Q-PO2	2.113	U-PR4	1.148	U-RM1	1.086	I-SI1	1.075
Q-PO3	1.142	U-PK1	1.029	U-LS1	1.589	I-DI1	1.146
Q-PO4	1.251	U-PK2	1.368	U-IS1	1.197	P-VI1	1.297
Q-NO1	1.606	U-TA1	1.051	E-TI1	1.072	P-TR1	1.055
Q-NO2	1.537	U-TA2	1.041	E-TI2	1.453	P-DO1	1.176
Q-NO3	1.319	U-TA3	1.021	E-LO1	1.403	F-PL1	1.845
Q-NO4	1.267	P-AF1	1.122	I-PE1	1.074	F-TR1	1.846
Q-DI1	1.346	P-CL1	1.180	I-PR1	1.064	F-SU1	1.198
Q-DI2	1.744	P-CB1	1.643	I-EX1	1.334	A-SA1	1.165
Q-DI3	1.756	P-PR1	1.945	I-IS1	1.074	A-EX1	1.192
U-NE1	1.343	P-FU1	1.267	I-PI1	1.742	A-US1	1.023
U-NE2	1.148	P-IN1	1.228	I-DA1	1.241	A-US2	1.011
						M-OT1	1.000

(b) Significance and relevance of indicators

The Outer Weights (OWs) are checked to examine the significance and relevance of indicators using bootstrapping of 3000 sample data (Figure 4.28). The OWs should be different from zero (p-value < 0.05; T-values > 1.96). However, if the OWs is different from zero but p-value ≥ 0.05, in such cases, as suggested by (Hair Jr et al., 2014; Hair et al., 2011; Hair, Ringle, & Sarstedt, 2013), the Outer Loadings(OLs) should be checked for the particular indicators to see if they pass a minimum threshold of 0.5 (OL > 0.5). If they pass, the indicators should be retained in the analysis otherwise the item is retained but it is interpreted as absolutely important and not as relatively important. If the item is not significant neither the OW nor OL, the researcher should decide whether to retain or remove the indicator by examining its theoretical relevance and potential content overlap with other indicators of the same construct (Hair et al., 2011). Table 4.15 shows the results of significance relevance of indicators by representing of OWs and OLs which meet the above mentioned thresholds (Significant level: 95%).

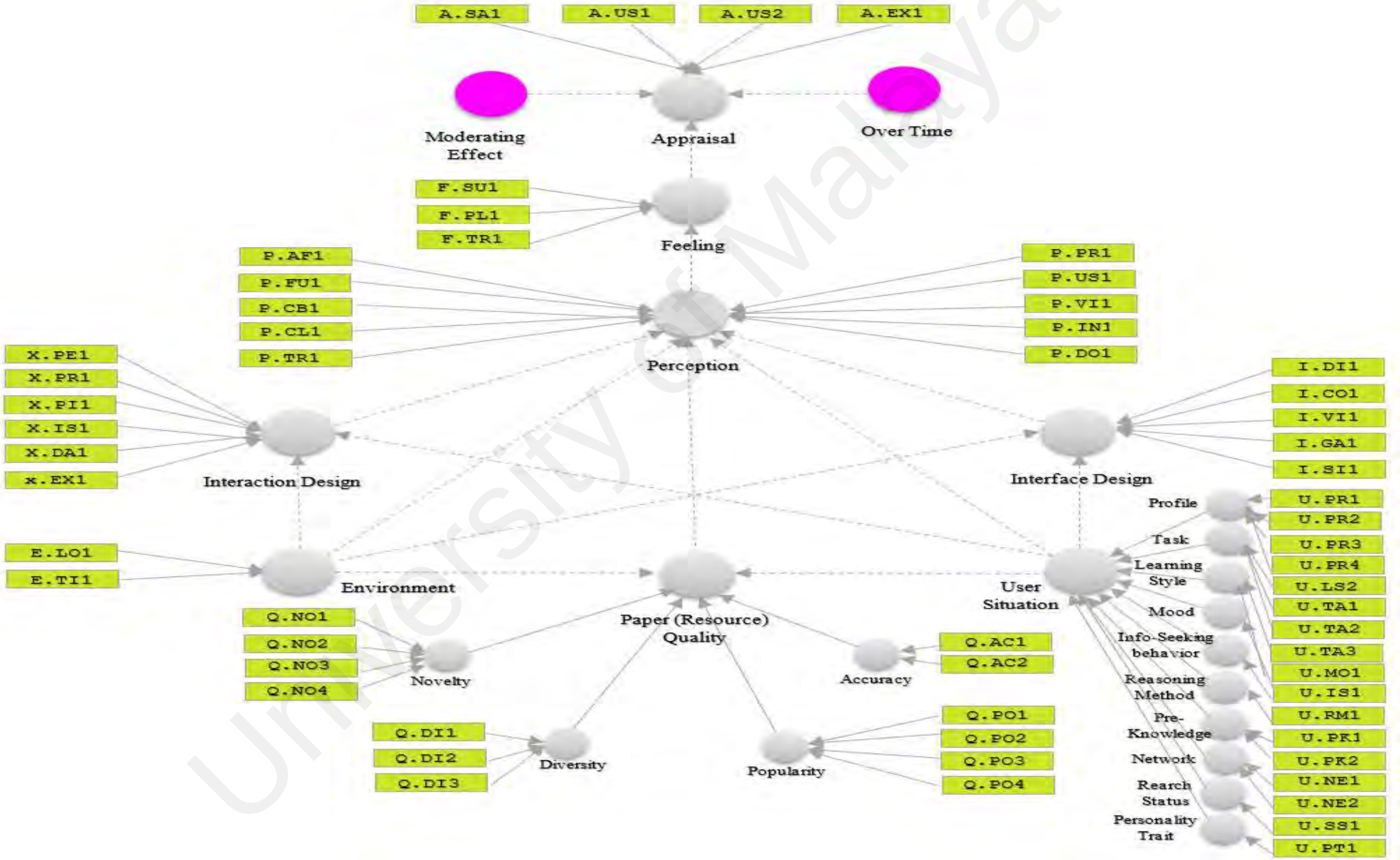


Figure 4.26: Bootstrap model

Table 4.15: Significance and relevance assessment of indicators

Indicator	OW	T - Values	P - Values	OL	T- Values	P- Values
U.IS1	0.674	3.034	0.015	0.420	5.365	0.036
U.LS2	0.654	3.388	0.014	0.293	2.989	0.001
LS_Mean	0.664					
U.MO1	0.067	2.381	0.013	0.468	2.972	0.000
U.NE1	0.334	3.125	0.012	0.912	3.862	0.000
U.NE2	0.312	2.255	0.011	0.181	2.666	0.048
NE_Mean	0.323					
U.PK1	0.567	2.493	0.012	0.395	2.771	0.000
U.PK2	0.556	2.762	0.009	0.529	3.006	0.000
PK_Mean	0.561					
U.PR1	0.893	3.718	0.008	0.346	2.347	0.010
U.PR2	0.873	2.587	0.007	0.823	2.484	0.009
U.PR3	0.894	2.013	0.006	0.518	2.467	0.015
U.PR4	0.947	2.093	0.004	0.279	2.415	0.008
PR_Mean	0.901					
U.PT1	0.092	3.678	0.003	0.282	3.144	0.001
U.RM1	0.054	2.592	0.002	0.614	2.294	0.003
U.SS1	0.412	3.306	0.001	0.396	4.302	0.002
U.TA1	0.712	2.668	0.001	0.429	2.583	0.005
U.TA2	0.721	2.100	0.001	0.591	2.088	0.006
U.TA3	0.625	5.930	0.002	0.643	4.555	0.009
TA_Mean	0.685					
U_Mean	0.417					
Q.AC1	0.376	3.577	0.003	0.113	3.594	0.007
Q.AC2	0.401	3.808	0.004	0.588	3.899	0.000
AC_Mean	0.388					
Q.DI1	0.537	2.154	0.005	0.134	2.248	0.005
Q.DI2	0.614	2.998	0.006	0.722	2.219	0.007
Q.DI3	0.538	2.053	0.007	0.166	2.770	0.006
DI_Mean	0.563					
Q.NO1	0.532	2.731	0.008	0.425	3.878	0.004
Q.NO2	0.630	3.166	0.009	0.192	3.743	0.004
Q.NO3	0.558	2.870	0.010	0.524	3.965	0.003
Q.NO4	0.578	2.198	0.011	0.170	2.045	0.000
NO_Mean	0.574					
Q.PO1	0.286	2.067	0.002	0.304	2.569	0.010
Q.PO2	0.339	2.507	0.013	0.201	2.943	0.000
Q.PO3	0.366	2.684	0.004	0.608	3.844	0.004
Q.PO4	0.357	2.766	0.005	0.165	2.068	0.007
PO_Mean	0.336					
Q_Mean	0.406					
E.LO1	0.052	3.237	0.045	0.846	3.087	0.001
E.TI1	0.367	2.548	0.042	0.491	3.070	0.001

Table 4.15: continued

Indicator	*OW*	*T - Values*	*P - Values*	*OL*	*T- Values*	*P- Values*
E.TI2	0.494	2.210	0.033	0.848	3.958	0.001
TI_Mean	0.430					
E_mean	0.241					
I.DI1	0.396	3.403	0.008	0.279	3.143	0.001
I.CO1	0.387	2.211	0.014	0.488	5.862	0.001
I.VI1	0.124	3.635	0.001	0.843	3.347	0.001
I.GA1	0.374	2.658	0.011	0.564	2.865	0.001
I.SI1	0.226	5.216	0.021	0.509	4.742	0.001
I_Mean	0.301					
X.PE1	0.526	2.395	0.031	0.741	3.652	0.001
X.PR1	0.529	3.655	0.002	0.420	3.025	0.000
X.PI1	0.553	2.731	0.003	0.710	2.756	0.001
X.IS1	0.460	2.195	0.014	0.464	5.933	0.000
X.DA1	0.132	2.893	0.029	0.403	4.344	0.000
x.EX1	0.298	4.681	0.005	0.585	4.161	0.001
X_Mean	0.416					
P.AF1	0.398	2.884	0.031	0.004	4.036	0.485
P.FU1	0.385	2.778	0.038	0.494	4.776	0.000
P.CB1	0.361	2.356	0.009	0.012	4.105	0.458
P.CL1	0.290	3.242	0.035	0.568	3.322	0.000
P.TR1	0.235	3.091	0.012	0.580	3.361	0.000
P.DO1	0.218	2.114	0.057	0.374	3.415	0.000
P.PR1	0.214	2.180	0.045	0.607	2.562	0.000
P.US1	0.198	2.791	0.001	0.424	4.355	0.000
P.VI1	0.196	3.094	0.001	0.105	2.892	0.186
P.IN1	0.184	2.136	0.016	0.215	3.925	0.027
P_Mean	0.267					
F.SU1	0.572	3.263	0.001	0.882	3.680	0.001
F.PL1	0.412	2.275	0.005	0.675	2.015	0.001
F.TR1	0.367	2.905	0.002	0.814	2.184	0.001
F_Mean	0.450					
A.EX1	0.037	2.442	0.029	0.402	2.049	0.071
A.SA1	0.479	3.639	0.031	0.999	3.615	0.065
A.US1	0.363	3.045	0.002	0.076	2.794	0.055
A.US2	0.326	2.353	0.002	0.024	2.240	0.005
A_Mean	0.286					

A high indicator weight suggests that the indicator is making a significant contribution to the formative latent variable (Tate, 2010). The results of significance relevance of indicators by representing of OWs and OLs which meet the above mentioned thresholds (Significant level: 95%) are shown in Table 4.15.

4.2.4.2 Assessment of inner model

In this section SRQ2.2 is addressed which investigates the validity of the constructs (components). In the PLS-SEM method, it is performed by assessment of structural model.The relationships between the latent variables (constructs) represent the structural model. Two criteria were applied to assess the structural model including Coefficient of determination (R^2) and Effect size (F^2) (Hair et al., 2011). For the easier reference, the constructs are coded as listed in Table 4.16.

Table 4.16: Constructs coding

Construct	*Code*	*Construct*	*Code*	*Construct*	*Code*
Environment$_{context}$	*E*	*IxD $_{adequacy}$*	*X*	*User's Feeling*	*F*
User$_{context}$	*U*	*UiD $_{adequacy}$*	*I*	*User's overall Appraisal*	*A*
Resource $_{quality}$	*Q*	*User's Perception*	*P*	*Over time*	*O*

(a) Coefficient of determination (R^2)

The coefficient of determination (R^2) is the second key criterion for evaluating the structural model and is used determining the degree of linear-correlation of variables in regression analysis. It measures the proportion of the variance of a dependent variable that is explained by independent variables. It shows the model's ability to explain and predict the dependent latent variables. In other words, R^2 explains how much of the variability of a factor (latent or dependent variable) can be influenced by its relationship to another factor (independent variable).

The latent variable in this research is UX. Therefore, R^2 shows how much of the variability of UX can be influenced by independent variables mentioned in Table 4.17.

An R-squared equal to zero means that the dependent variable cannot be predicted using the independent variable. Conversely, if it equals one, it means that the dependent of variable is always predicted by the independent variable.

Table 4.17: R-Squares of dependent (latent) variables

IV	R^2	R^2 *effect*	R^2 *adjusted*
A	0.554	Moderate	0.546
E	0.596	Moderate	0.556
F	0.664	Moderate	0.660
X	0.762	Moderate	0.758
I	0.716	Moderate	0.711
P	0.519	Moderate	0.505
Q	0.735	Moderate	0.726
U	0.992	Substantial	0.992

A coefficient of determination that falls within this range measures the extent that the dependent variable is predicted by the independent variable. For example, an R-squared of 0.99, means that 99% of the dependent variable (UX) is predicted by the independent variable (user $_{context}$). Overall, R^2 values of 0.75, 0.50, or 0.25 for dependent variables are viewed as substantial, moderate, or weak (Hair et al., 2013). As depicted in Figure 4.27, the R^2s of the dependent variables indicate that the variance of UX is explained substantially only by the effect of User $_{context}$ variable.

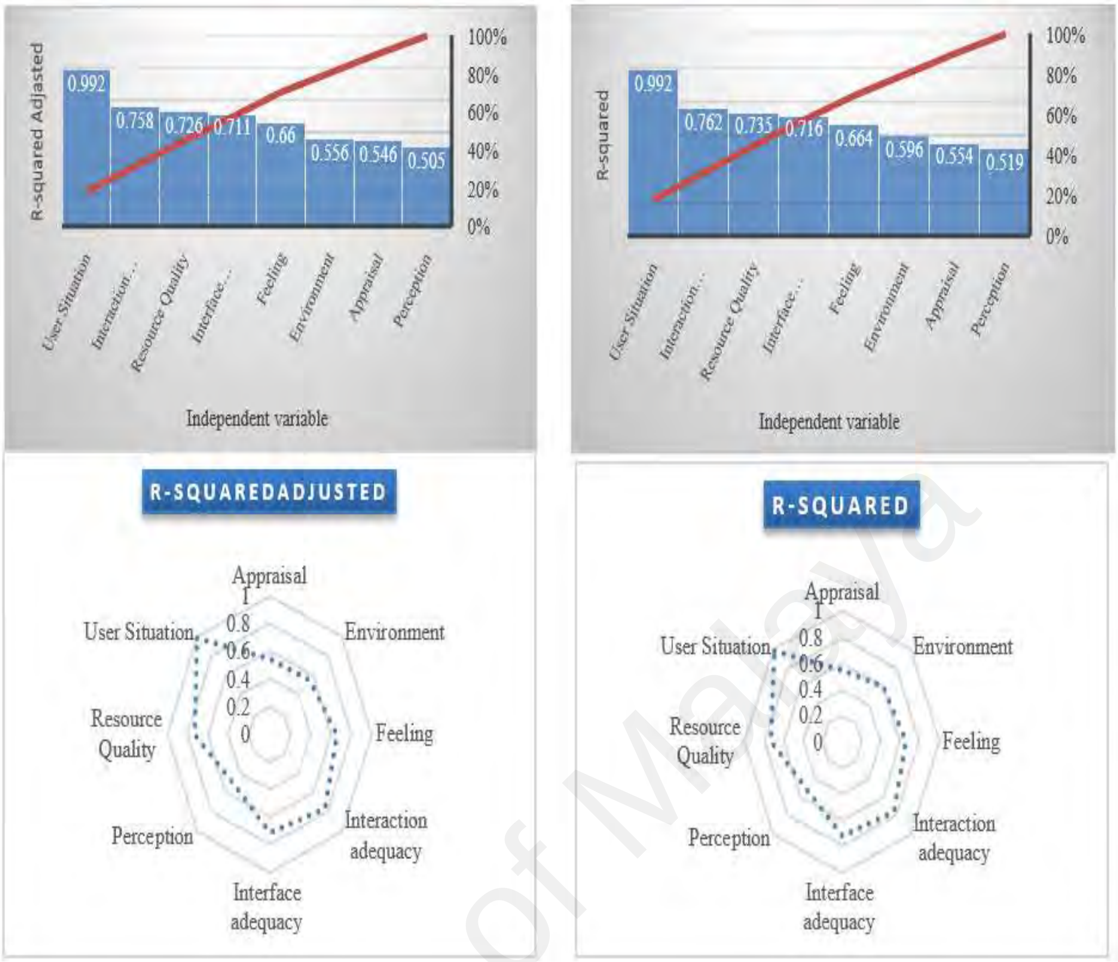


Figure 4.27: R^2s and adjusted R^2s of the independent variables

(b) Effect size (f^2)

Effect size (f^2) is an important tool in reporting and interpreting the impact of a construct on another one. It simply shows the correlation between two variables (constructs). The f^2 is calculated by Cohen's formula (Cohen, 1992). According to Cohen, the values of 0.35, 0.15, and 0.02 signify respectively large, medium, and small effects. f^2 values of less than 0.02 indicate that there is no effect.

The direct and in-direct relationships between the constructs, defined in the proposed model, have been illustrated in Table 4.18. The correction between two variables of X and Y is shown by $\rho\ (X, Y)$. Based on the f^2 values, the smallest effects are between the environment context construct and other constructs. It means that environment changes such as location, time might not be influence on the interaction adequacy, interface adequacy and user's perception however the environment changes may impact on resource quality which includes novelty, accuracy, diversity and popularity of a paper in this research. For example, the academic time can impact on the recommended paper. Also there is a small correlation between the overtime construct (moderating variable) and appraisal. As mentioned before, f^2 does not necessarily mean that the change in one variable is the cause of the change in the values of the other variable.

Table 4.18: f^2 values

ρ (X,Y)	*f^2*	*Effect*	*ρ (X,Y)*	*f^2*	*Effect*
ρ (E, Q)	0.161	Medium	*ρ (Q, P,F)*	0.282	Medium
ρ (E, X)	0.003	Small	*ρ (Q, P,F,A)*	0.282	Medium
ρ (E, I)	0.004	Small	*ρ (I,P)*	0.174	Medium
ρ (E, P)	0.002	Small	*ρ (I, P,F)*	0.162	Medium
ρ (E, P,F)	0.011	Small	*ρ (I, P,F,A)*	0.280	Medium
ρ (E, P,F,A)	0.006	Small	*ρ (X,P)*	0.102	Medium
ρ (U, Q)	0.376	Large	*ρ (X, P,F)*	0.148	Medium
ρ (U, I)	0.351	Large	*ρ (X, P,F,A)*	0.102	Medium
ρ (U, X)	0.356	Large	*ρ (P,F)*	0.354	Large
ρ (U, P,F)	0.271	Medium	*ρ (P,F,A)*	0.371	Large
ρ (U,P,F, A)	0.282	Medium	*ρ (F,A)*	0.374	Large
ρ (Q, P)	0.375	Large	*ρ (O,A)*	0.001	Small

The following constructs have obtained the large effect and the correlations between the user's context and resource quality is the largest effect.

```
Largest effects = [(ρ (U, Q),   0.376), (ρ (U, I),0.351), (ρ
(U, X),0.356), (ρ (Q, P), 0.375),(ρ (P,F), 0.354)), (ρ (P,F,A),
0.371)),(ρ (F, A), 0.374)]
```

4.2.4.3 Hypothesis testing (β test)

To examine the hypothesized relationships between the constructs, the path coefficients (β) are assessed which have standardized values of linear regression weights between -1 and + 1 (Hair et al., 2011). Estimated path coefficients close to + 1 represent strong positive relationships (and vice versa for negative values) that are almost always statistically significant (p-value < 0.05; T-values > 1.96) (Hair et al., 2011). The closer the estimated coefficients are to 0, the weaker the relationships are (Normal data distribution). In other words, very low values close to 0 are usually none-significant. Figure 4.28 shows the Normal data distribution.

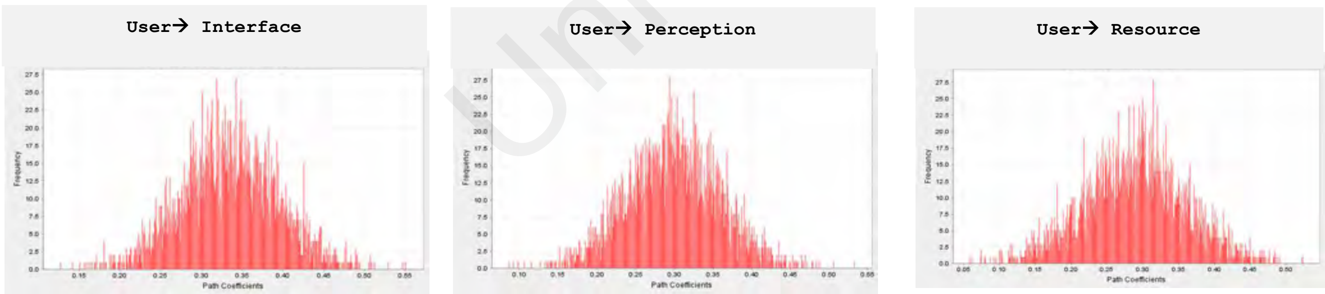
As shown in Table 4.19, the obtained path coefficients (p-value < 0.05; T-values > 1.96) examined the relationships between the constructs. Among the hypothesized relationships, all relationships were found statistically significant other than the relationship of the *Moderating Effect (over time)* → *Appraisal* (p-value = 0.201) which is discussed in the next section. The results also revealed that the strongest relationships are between the constructs listed below;

$User_{context}$ → $Resource_{quality}$ 0.790
$User_{context}$ → $IxD_{adequacy}$ 0.788
$User_{context}$ → $UiD_{adequacy}$ 0.786
$Resource_{Quality}$ → Perception 0.760

In PLS-SEM method, all direct and in-direct coefficients paths between the variables are assessed by using bootstrapping test. Since the focus of this research is on the contextual data and the strongest relationships were predicted between (U, Q), the direct and in-direct coefficients paths of variables were assessed and illustrated in the Appendix K and J. The relationships between the contexts are the source of inspiration for the construction of BN model and UI for SRSs which are discussed respectively in Chapters 5 and 6.

Table 4.19: Significance of the path coefficients (β)

Constructs	*β*	*T-value*	*P-value*	+→	*Hypo. Testing*
E → *X*	0.305	3.656	0.000	Weak	Supported
E → *I*	0.350	3.020	0.000	Weak	Supported
E → *P*	0.387	2.012	0.022	Weak	Supported
E → *Q*	0.449	1.787	0.037	Moderate	Supported
F → *A*	0.586	2.035	0.000	Moderate	Supported
X → *P*	0.428	2.846	0.002	Moderate	Supported
I → *P*	0.456	2.590	0.005	Moderate	Supported
O* → *A	**0.056**	**0.837**	**0.201**	**-**	**N-Supported**
P → *F*	0.605	3.076	0.000	Strong	Supported
U → *X*	0.788	2.609	0.000	Strong	Supported
U → *I*	0.786	2.118	0.000	Strong	Supported
U → *P*	0.649	2.376	0.354	Moderate	Supported
U → *Q*	0.790	2.942	0.000	Strong	Supported
Q → *P*	0.760	3.942	0.003	Strong	Supported



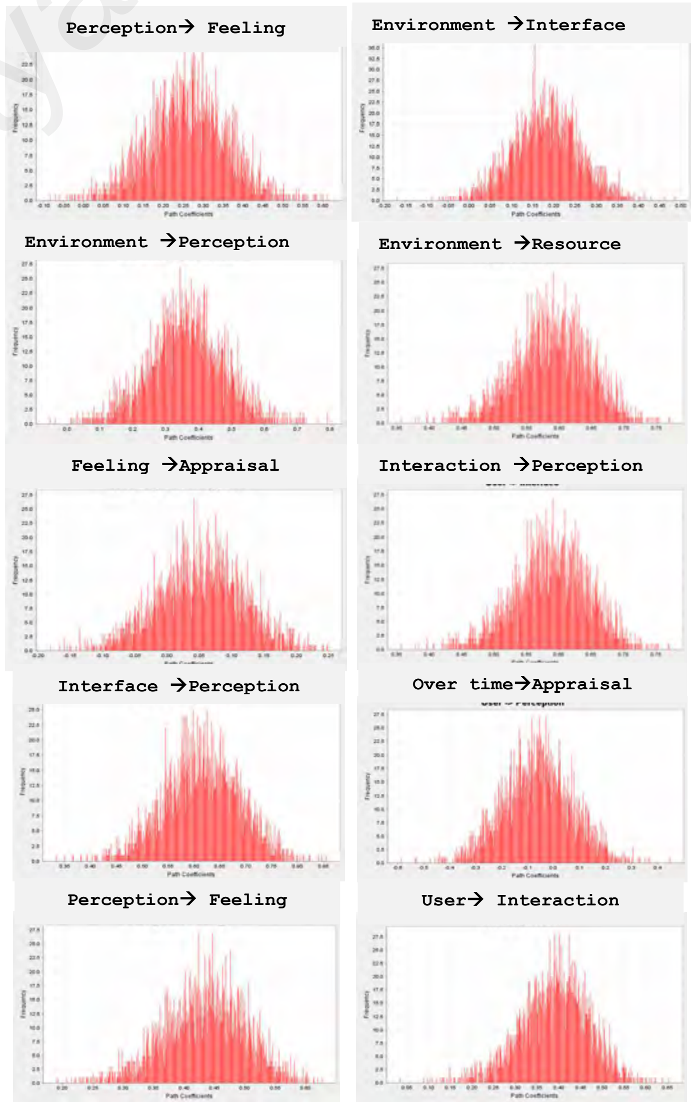


Figure 4.28: Normal distribution(s)

(a) Moderating effect impact

As mentioned above, the results of path coefficients showed that the moderating effect indicator (over time) does not surpass the minimum threshold of p-value ($p\text{-value}_{overtime} = 0.201$) which means that there is not statistically significant relationship between this construct and overall appraisal (Figure 4.29).

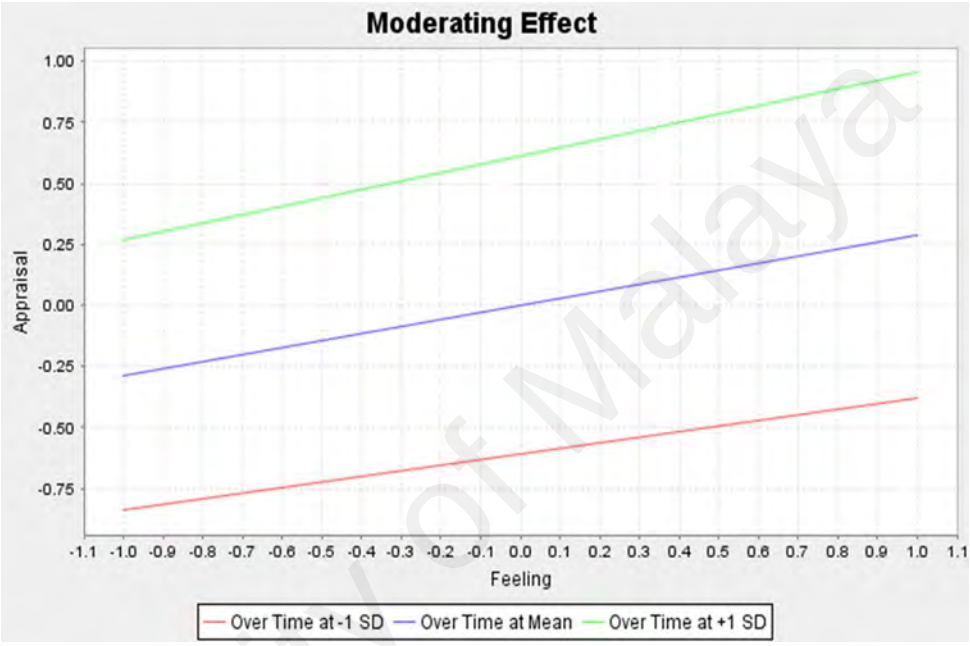


Figure 4.29: The impact of moderator variable

As discussed earlier, researchers have emphasized that the UX is not built one night. The users' appraisal of SRSs might be changed over the time and it is not stable. Therefore, from the conceptual view point, there must be a relationship between these two constructs. However, the empirical results revealed no relationship. One main reason might be that the ratings collected in this research, were provided by the users not while experiencing the paper recommendations but by imagining the situation and providing judgments. Hence, the impact of "over time" construct has not been assessed significant.

4.2.4.4 Framework revision

Based on the PLS-SEM method guidelines, different tests have been applied to empirically validate the framework (shown in Table 4.20). For the Significance and relevance of indicators OWs and VIFs have been applied. The results reveled that all defined indicators are empirically valid therefore no indicator is removed in the revised conceptual framework. Also, the R^2 and F^2 tests examined the validity of the constructs (components) and the results revealed that all predefined components empirically contribute into construction of UX of SRSs. In addition, the results of path coefficients (β) and F^2 showed that all the relationships are valid in the conceptual framework however, the moderating effect indicator (over time) does not surpass the minimum threshold of p-value ($p\text{-value}_{overtime} = 0.201$) which means that there is not statistically significant relationship between this construct and overall appraisal. Therefore, this relationship is removed in the conceptual framework. As discussed earlier, some of the relationships are weak but they remain in the framework as long as they are statistically significant.

Table 4.20: Tests applied for the empirical examination

Step	*Examination*	*Test*
Dataset	Validity of measuring tool (Pre-test)	-Face validity (Expert's interview) -Content Validity (Q-soring)
	Consistency of measuring tool (Pilot-test)	-Test- retest (Spearman's rho)
Measurement (outer) model	Multicollinearity	-Variance Inflation Factor (VIF)
	Significance and relevance of indicators	-Outer weight (OW) -Outer load (OL)
Structural (inner) model	The constructs' variabilities	-Coefficient of determination (R^2)
	The constructs' effects	-Effect size (F^2)
Hypo. Testing	Variables' relationships	-Significance of the path coefficients (β)
GOF	The model performance in explaining different datasets	-The Standardized Root Mean Square Residual (SRMR)

.

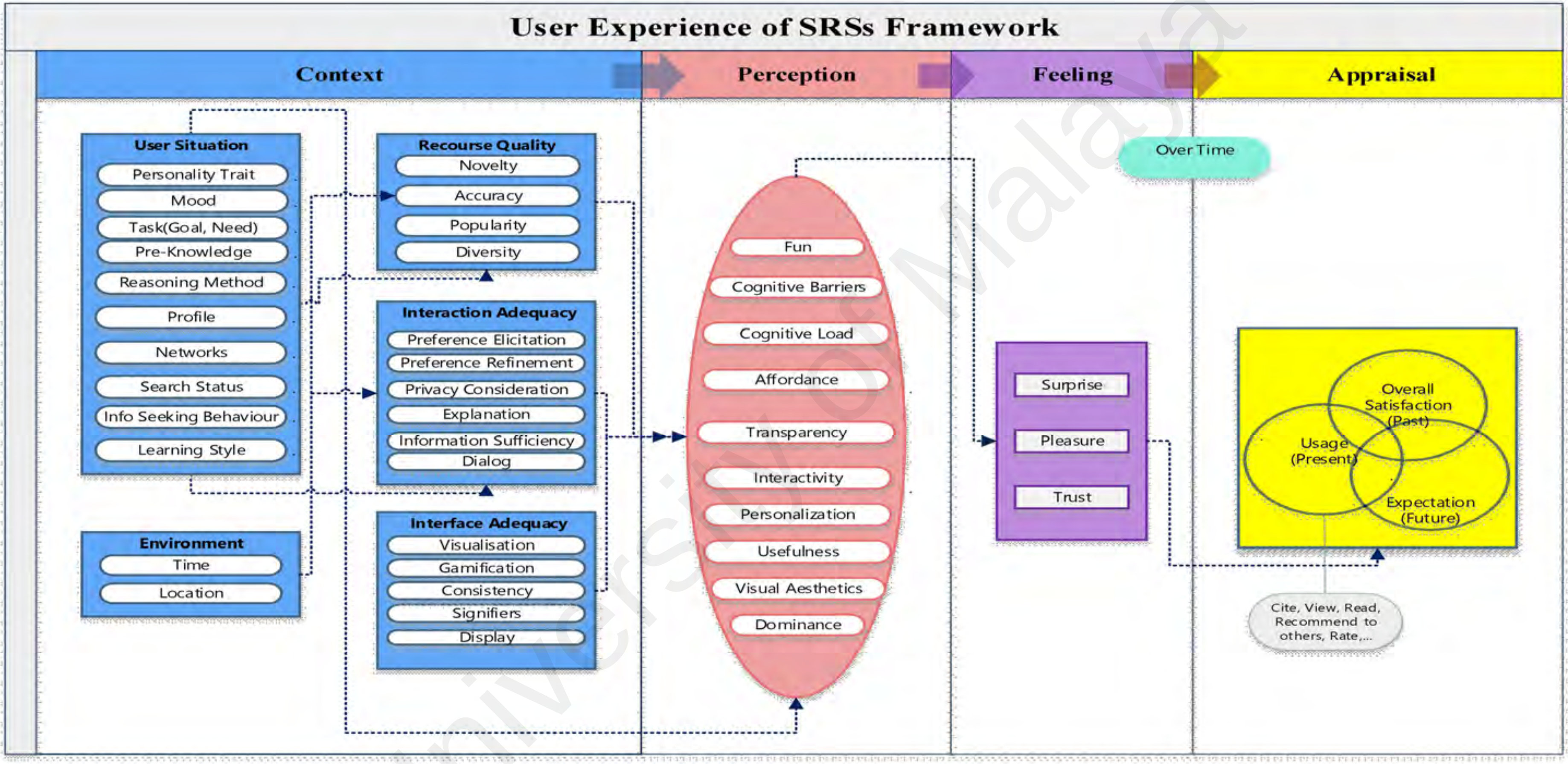


Figure 4.30: Revised framework after the empirical experiment

4.2.4.5 Goodness of Fit (GoF)

There is no overall fit index in PLS path modeling. The GoF can be useful to assess how well a PLS path model can explain different sets of data (Henseler et al., 2014). A measure of the Goodness of Fit (GoF) index has been proposed by (Tenenhaus, Amato, & Esposito Vinzi, 2004). This criterion takes into account the model performance in both the measurement and the structural model while provides a single measure for the overall prediction performance of the model and is obtained as the geometric mean of the Average Variance Extracted (AVE) and the average R^2 value shown by the following formulas (Tenenhaus, Amato, & Esposito Vinzi, 2004):

$$\boldsymbol{GOF = \sqrt{\overline{AVE} \times \overline{R^2}}} \tag{4.3}$$

$$\boldsymbol{For\ n\ item:}\quad \boldsymbol{\mu_{AVE=\frac{1}{n}}\sum_{i=1}^{n} x_i \quad and \quad \mu_{R^2=\frac{1}{n}}\sum_{i=1}^{n} x_i} \tag{4.4}$$

$$\boldsymbol{Relative\ GOF = \sqrt{\overline{\mu_{AVE}} \times \overline{\mu_{R^2}}}} \tag{4.5}$$

The Standardized Root Mean Square Residual (SRMR) is another measure introduced by Henseler et al. (2014) to evaluate the GoF. The SRMR is defined as the difference between the observed correlation and the model implied correlation matrix which allows assessing the average magnitude of the discrepancies between observed and expected correlations as an absolute measure of (model) fit criterion. The Smart-PLS calculates SRMR by bootstrapping and for the proposed model in this research is 0.09. A value of less than 0.10 (Hu and Bentler, 1999) is considered a good fit and allows to conclude that the model performs well compared to the baseline values. However, in the recent years, a few researchers such as Henseler and Sarstedt (2012) criticized that the GoF by Tenenhaus et al. (2004) does not represent a fit measure and should not be used as such.

Researchers showed this measure is unsuitable for identifying mis-specified models and have advised not to use it (Henseler et al., 2014).

4.2.4.6 Detection of the most relevant contexts

The R^2s of the dependent variables indicated that the variance of UX of SRSs is explained substantially only by the effect of User $_{context}$ variable. In addition, based on the outputs of Outer Weights (OWs) (Table. 4.15; Section: 4.2.4.1), among the indicators of one construct, the indicator which have more weight is relevant for the construction of the formative index demonstrates a sufficient level of validity. Among the user context construct, respectively profile `(PR_Mean, 0.901)`, task `(TA_Mean, 0.685)`, learning style `(LS_Mean, 0.664)`, pre-knowledge `(PK_Mean, 0.561)` and, information seeking behavior have obtained the highest weights which confirm that they have a significant contribution to the formation of user contexts for a SRS.

`U_weight = [(PR_Mean, 0.901), (TA_Mean, 0.685), (LS_Mean, 0.664), (PK_Mean, 0.561), (IS, 0.460)]`

Also, between time `(TI_Mean, 0.430)` and location `(E.LO1, 0.052)` for the environment context, time is more relevant for the construction of the formative the environment context and demonstrates a sufficient level of validity. Although the weight of location is low but the value of `0.52` still surpasses the threshold. Therefore, it is not removed. Among the paper quality context constructs, novelty (`NO_Mean, 0.574)` and diversity (`DI_Mean, 0.563)` have more weights.

`Q_weight = [(AC_Mean, 0.388), (DI_Mean, 0.563), (NO_Mean, 0.574), (PO_Mean, 0.336)]`

Also, display (`I. DI1, 0.396`), consistency (`I. CO1, 0.387`) and gamification (`I.GA1, 0.374`) have the highest weights among the UiD's indicators. And for the IxD construct, as Table 4.15 explains, the preference elicitation, refinement and privacy consideration have received the highest weights in contribution to the IxD the formative latent variable;

IxD_{weight} = [(X.PE1, 0.526), (X.PR1, 0.529), (X.PI1, 0.553), (X.IS1, 0.460), (X.DA1, 0.132), (X.EX1, 0.298)]

Between the perception's indicators, the three highest weight respectively belong to affordance (P.AF1, 0.398), fun (P.FU1, 0.385) and, cognitive barrier (P.CB1, 0.361).

P_{weight} = [(P.AF1, 0.398), (P.FU1,0.385),(P.CB1,0.361),(P.CL1, 0.290),(P.TR1, 0.235),(P.DO1, 0.218),(P.PR1, 0.214),(P.US1, 0.198),(P.VI1, 0.196), (P.IN1,0.184)]

Among the perception indicators, affordance is the indicator which has contributed mostly in UX of SRS. Interestingly, transparency has not been the most contributed indicator.

Between the users' feeling and appraisal indicators, surprise and satisfaction have the highest weight (F.SU1, 0.572), (A.SA1, 0.479).

F_{weight} = [(F.SU1, 0.572), (F.PL1, 0.412), (F.TR1, 0.367)]
A_{weight} = [(A.EX1, 0.037), (A.SA1, 0.479), (A.US1, 0.363), (A.US2, 0.326)]

The results path coefficients (β) (section 4.2.4.2) revealed that the strongest relationships are between the constructs listed below;

$User_{context} \rightarrow Resource_{quality}$	0.790
$User_{context} \rightarrow IxD_{adequacy}$	0.788
$User_{context} \rightarrow UiD_{adequacy}$	0.786
$Resource_{Quality} \rightarrow Perception$	0.760

Additionally, the results of f^2 test showed that the following constructs have obtained the large effect and among them, the correlations between the user's context and resource quality is the largest effect.

```
Largest effects = [(ρ (U, Q),   0.376), (ρ (U, I),0.351), (ρ
(U, X),0.356), (ρ (Q, P), 0.375),(ρ (P,F), 0.354)), (ρ (P,F,A),
0.371)),(ρ (F, A), 0.374)]
```

Considering the above mentioned data analysis results emphasizing on the strong relation between the $\text{User}_{\text{context}} \rightarrow \text{Resource}_{\text{quality}}$, the relationships between the relevant contexts in this two variables are the source of inspiration for the construction of BN model. Hence, for more assessment of the relationships between these two variable and their indicators, the path coefficients (β) of specific indirect effects were assessed and listed in Appendices J.

In summary, as Figure 4.31 depicts, the predicted relevant contexts are applied in both development of both UM and UI in this research. More details are provided in chapter 5 and 6. The focus of this research is not on the designing the gamification and visualization solutions for the SRSs.

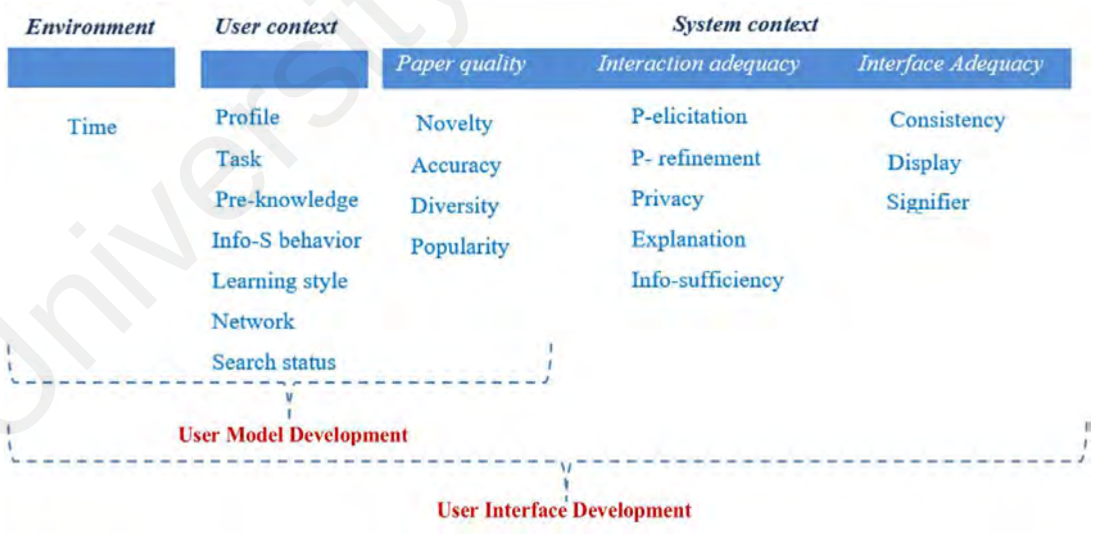


Figure 4.31: Contribution of the relevant context into UX of SRS

4.3 Discussion

This chapter continues existing lines of research that aim to better understand the UX of SRSs and does not stop here. However, the implication of this study is multi-fold. First, it proposes a theory-based conceptual framework for understanding how contexts influence UX of SRSs. It attempts to unify various conceptualizations in the literature relating UX and RSs to give them a theoretical foundation with the help of relevant experts. It also makes the first attempt to empirically examine the impact of contexts on UX and to detect the most relevant contexts and their conditions influencing UX in the field of SRSs. Moreover, the proposed framework bridges the user's contexts with the system contexts and provides finer insights into both back-end and front-end of RSs development from a UX design perspective. Additionally, the relationships between the contexts are the source of inspiration for the construction of a BN user model and UI for SRSs which are discussed respectively in Chapters 5 and 6. This research makes important contributions for both academics and the SRS providers, however, some limitations are acknowledged in the following.

First, the participants of this study had a certain background in Computer Science. While they are part of scholarly community, there may be aspects of UX of RSs that are different in other majors. Second, this research focus to detect contexts influencing research paper recommending. The empirical results showed that location does not have a significant impact on UX of SRSs while for offering a book, location might be one of the most influencing contexts. Another possible limitation is that, the ratings collected in this research, were provided by the users not while experiencing the paper recommendations but by imagining the situation. Finally, there are possible relationships between contexts and user's perceptions. However, analysis of those relationships requires huge data collection which creates a complex predictive model. Hence, the analysis of their direct impacts were ignored and left for the future studies.

4.4 Summary

This chapter proposes a framework that the primary goals are to explore how contexts influence UX with SRSs. Furthermore, it identifies what are the relevant contexts incorporated in the UX. To achieve this goal, first the existing models and theories of UX especially with RSs are reviewed and then several gaps in existing works through reviewing the current studies and experts feedbacks in this field are identified. The proposed framework consists four main components including context, perception, feeling and appraisal and manifests in what way contexts (latent variables) influence UX with SRSs. Additionally, it not only enriches our conceptual understanding of how contextual information influences UX of scholarly recommender systems but also serves as a foundation for further theoretical and empirical investigations. The proposed framework is evaluated empirically by using PLS-SEM method. Besides, the relevant contexts influencing UX of SRSs are assessed and discussed in this chapter.

CHAPTER 5: CONTEXTUAL BAYESIAN USER EXPERIENCE MODEL

In this chapter, the relevant contexts identified in objective 1 are exploited to design a Bayesian UM for assisting the diagnosis of scholars' information needs in terms of accurate, novel, diverse and popular papers. Prior to explaining the UM development, a few essential definitions are briefly presented, and then the reasons behind applying the BNs technique are discussed. Thenceforth, three phases of BN modeling including dataset preparation, network structure learning, and parameter learning and inference are elaborated. The proposed user model can be embedded in the process of recommending to identify the users' information needs and help recommenders retrieve more appropriate recommendations which consequently leads to the enhancement of the UX of SRSs.

5.1 Essential definitions

Bayesian networks (BNs) are powerful tools used for uncertainty modeling. Their first appearance was in the field of medical decision systems in the late 1970s (Pearl, 1985). For better understanding of BN modeling, in the following, a few crucial definitions are discussed succinctly.

5.1.1 Bayes' theorem

The inference in BNs is based on a probabilistic theory called Bayes' theorem or Bayes' law which spreads knowledge within the network (Heckerman et al., 1995; Neapolitan, 2004). Bayes' theorem is a widely accepted and uncontroversial formula and has been around for hundreds of years(Korb & Nicholson, 2003). It describes the probability of an event based on prior knowledge of the conditions that might be related to that event. For instance, for hypothesis H and evidence E, Bayes theorem states that the relationship between the probability of the hypothesis before getting the evidence $p(H)$ and the probability of hypothesis after getting the evidence $p(E|H)$ which defines as follows (Jameson, 1995):

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E)} \tag{5.1}$$

Where the prior probability is $p(H)$, $the\ p(E|H)$is the likelihood function of H, and $p(E)$is the prior probability of E which is called marginal probability. Thus, $p(H|E)$is a posterior probability of H given E (Rim et al., 2013a).

In a general form, where $p(H)$and $p(E)\geq 0$ and $p(H_i)$consist of mutually exclusive events within the universe S, the Bayes' formula would be (Bolstad & Curran, 2016);

$$p(H_i|E) = \frac{p(E \cap H_i)}{p(E)} = \frac{p(E|H_i) \times p(H_i)}{\sum_{j=1}^{n} p(E|H_i) \times p(H_i)} \tag{5.2}$$

BNs are better suited not only to reason with the knowledge and belief uncertain, but also the structure of knowledge representation (Darwiche, 2009; Mahjoub & Kalti, 2011).

5.1.2 BN Graph structure

More precisely, BNs are a class of graphical models that allow a concise representation of the probabilistic dependencies between a given set of random variables (nodes) $X = \{X_1, X_2, X_3, \dots X_n\}$ as a Directed Acyclic Graph (DAG) *G=(V,A)*. Each node $v_i \in V$ corresponds to a random variable X_i which might represent a causal link from parent node to their children (Rim et al., 2013a). Each node is associated with a conditional probability distribution which assigns a probability to each possible value of this node for each combination of the values of its parent nodes (Zukerman & Albrecht, 2001).

Graph *G* is an independency map (I-map) of the probabilistic dependence structure *P* of *X* if there is a one-to-one correspondence between the random variables in *X* and the nodes *V* of *G*, such that for all disjoint subsets *A*, *B*, *C* there is;

$$\boldsymbol{A} \perp \boldsymbol{p}\ (\boldsymbol{B}|\boldsymbol{C}) \Leftarrow \boldsymbol{A} \perp \boldsymbol{G}\ (\boldsymbol{B}|\boldsymbol{C}) \tag{5.3}$$

Similarly, G is a *dependency map* (D-map) of P if X , there is

$$\boldsymbol{A} \perp \boldsymbol{p}\ (\boldsymbol{B}|\boldsymbol{C}) \Rightarrow \boldsymbol{A} \perp \boldsymbol{G}\ (\boldsymbol{B}|\boldsymbol{C}) \tag{5.4}$$

G is said to be a *perfect map* of P if it is both a D-map and an I-map,

$$\boldsymbol{A} \perp \boldsymbol{p}\ (\boldsymbol{B}|\boldsymbol{C}) \Leftrightarrow \boldsymbol{A} \perp \boldsymbol{G}\ (\boldsymbol{B}|\boldsymbol{C}) \tag{5.5}$$

and in this case, P is said to be *isomorphic* or *faithful* to G.

Overall, BNs consist of both qualitative and quantitative parts. With regard to the qualitative part, it is the structure of the network: a directed acyclic graph where nodes correspond to variables and arcs representing influences between variables. The quantitative part, however, provides the conditional probability tables that make up the network settings (Zukerman & Albrecht, 2001) (Korb & Nicholson, 2003). The learning applies both the network structure and parameters that can be obtained from complete or incomplete data(Mahjoub & Kalti, 2011). The correspondence between the structure of the DAG G and the conditional independence relationships is elucidated by the directed separation criterion (Pearl, 1988), or d-separation, as discussed below.

5.1.3 D-separation

If A, B, and C are three disjoint subsets of nodes in a DAG G, then C is said to *d-separate* A from B, denoted $A \perp G\ (B|C)$, if along every sequence of arcs (path) between a node in A and a node in B, there is a node v satisfying one of the following two conditions(Zukerman & Albrecht, 2001):

1) v has converging arcs (i.e., there are two arcs pointing to v from the adjacent nodes in the path) and none of v or its descendants (i.e., the nodes that can be reached from v) are in C.

2) v is in C and does not have converging arcs.

5.1.4 Markov property of BNs

The Markov property of BNs, which follows directly from d-separation, enables the representation of the joint probability distribution of the random variables in X (the global distribution) as a product of conditional probability distributions (the local distributions associated with each variable X_i) (Nagarajan, Scutari, & Lèbre, 2013) (Ono, Kurokawa, Motomura, & Asoh, 2007). This is a direct application of the chain rule (Korb and Nicholson, 2010). In the case of discrete random variables, the factorization of the joint probability distribution $\boldsymbol{P}_X$ is given by

$$\mathrm{P}_X(\mathrm{X}) = \prod_{i=1}^{p} \mathrm{P}_{X_i}(X_i | \prod X_i) \tag{5.6}$$

where $\prod X_i$ is the set of the parents of X_i; in the case of continuous random variables, the factorization of the joint density function f_X is given by

$$f_X(\mathrm{X}) = \prod_{i=1}^{p} f_{X_i}(X_i | \prod X_i) \tag{5.7}$$

Bayes' theory is an old and tested math rule which helps to make good decisions when there is uncertainty (W. S. Geisler, 2008). It can be applied not only in developing complex systems but also in making an individual a better thinker in its daily life. It asserts that this theory is the single most important tool for representing appropriate strengths of belief to understand human opinion which is constrained by ignorance and uncertainty(Korb & Nicholson, 2003). In other words, it is the closest technique to the human reasoning or rationality which is so-called Bayesian thinking or Bayesianism philosophy (Korb & Nicholson, 2003) (Easwaran, 2011). As shown in Figure 5.1, the Bayes' rule applies the past data (prior probability) and present data (likelihood of the evidence) to predict the future (posterior). Therefore, it updates the degree of belief in order to come up with a new or updated strength of belief (posterior) (Korb & Nicholson, 2003)

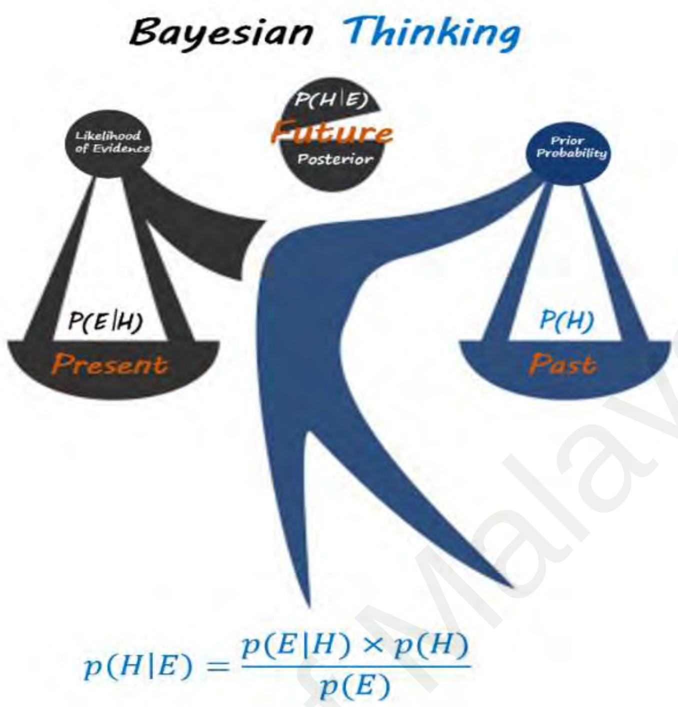


Figure 5.1: Bayesian thinking

To summarize, Bayesian thinking can be used to avoid common mistakes and fallacies in thinking since having a strong opinion/belief about an issue can make it hard to take in new information about it or to consider other options when they are presented. It can also be used to reach decisions in those circumstances when very few observations or pieces of evidence are available.

In the next section, it is given a substance why the BNs method is an appropriate method for contextual users modeling for SRSs in this research.

5.2 The reasons for selection of BNs method

This research aims to develop a UM which exploits the relevant contexts (Objective 1) in order to identify the users' information needs for a SRS. The BN method is a suitable method to achieve objective 2 in this research. The reasons for the choice of BNs are varied and are discussed as what follows.

5.2.1 Suitable to deal with uncertain and dynamic contexts

As discussed in chapter 2, contexts are dynamic and change over the time. Therefore, their conditions are very complex and uncertain (Hariri, Mobasher, & Burke, 2014). Moreover, the user's need of scholarly papers is uncertain and changes due to the different contexts such as background knowledge, preferences, and goals. The BN method provides an effective approach for constructing and manipulating probabilistic models for handling uncertainty in context awareness applications (Rim et al., 2013a). Besides, this method is appropriate for representing complex relations between the variables which their states change over the time (Codina Busquet & Ceccaroni, 2014; Ono et al., 2007).

5.2.2 Well adapted for UMs developing

A fundamental goal of HCI research is to make systems more usable and more useful, and to provide users with better experiences which fit their situations (Fischer, 2001). A UM strives to accomplish this goal by discovering the laws of nature and inferring unobservable information about a user from observable information to analyze perceptual and cognitive processes and characterize individual differences. BNs are well adapted to the problem of user modeling because they can represent the uncertainty related to the modeling of users' preferences (Guo, 2011; Korb & Nicholson, 2003). In addition, BN method is a natural and scientific method to cope with the variability and the complexity of users' situations by probability distribution of over the whole related variables and explicitly represents causal relations (Long et al., 2010) (French, 1986; Peterson, 2009). Apart from this, as discussed in chapter 2, several ML methods can be applied for UMs developing such as BNs, fuzzy logic, and neural networks (Ono et al., 2008). The use of ML algorithms in RSs has been reviewed by Portugal et al.(2015) and the results revealed that no study has applied BNs for SRSs. Interestingly, among the existing SRSs studies, there is no work on contextual BNs models for CASRSs (Hassan, 2017).

5.2.3 Appropriate for diagnose of user's information needs

Quality of recommendations (papers) refers to the capability of the system to predict exactly those papers that make the user would like or use (Berkovsky et al., 2008; Kobsa, 2001). In other words, those papers which are well matched with the user's information needs. The information needs vary among users owning to different contexts such as background knowledge, preferences, and goals. The results of path coefficients (β) from objective 1 unveiled that among the contexts, the strongest relationships are between `[(User`$_{context}$ → `Resource`$_{quality}$`), 0.790]`. Additionally, the Analysis of OWs, in chapter 4, attested that four indicators of novelty, diversity, popularity, and accuracy form the quality of a paper (`Q`$_{weight}$ `= [(AC_Mean, 0.388), (DI_Mean, 0.563), (NO_Mean, 0.574), (PO_Mean, 0.336)])`. Therefore, as Figure 5.2 shows, a user model should match the recommended papers on the basis of four identified levels of accurate, novel, popular, and diverse with users' information needs or users' contexts.

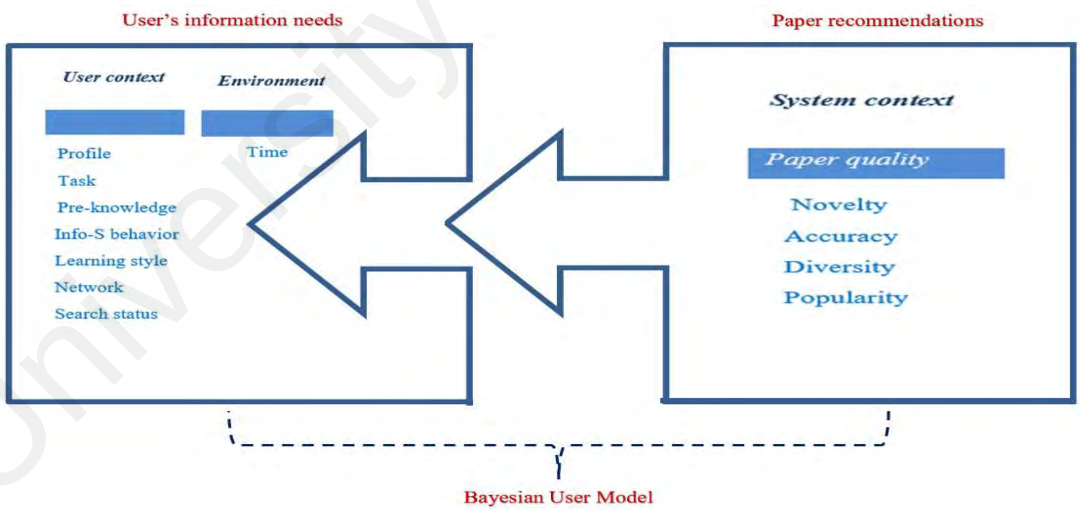


Figure 5.2: Paper quality matching with user's information needs

As mentoned ealier, BNs are flexible models not only to reason with the knowledge and belief uncertain(Sean M McNee, Riedl, et al., 2006a), but also the structure of knowledge representation (Darwiche, 2009; Mahjoub & Kalti, 2011) (Park, Yoo, & Cho, 2006). They make the system dynamic and evolving in order to infer users' needs and react depending on them (Guo, 2011). Also, BNs provide the predictions to be made about a number of influential variables rather than a single variable by using probability distribution (Zukerman & Albrecht, 2001). Hence, BNs are relevant for diagnosing users' information needs.

5.2.4 Appropriate for representation of casualty relationships

The results of objective 1 showed that there are cause-effect (casualty) relationships between the contexts (presented in appendix J). Among the methods that can be used to represent uncertain domains such as decision trees, neural networks, mixtures of basic functions, and Markov networks, etc., researchers have claimed that BNs are better suited for representation and learning of the directed causal relationships (Darwiche, 2009; Mahjoub & Kalti, 2011). In addition, BNs aim to facilitate the description of a collection of beliefs by making explicit the relationship of causality and conditional independence among these beliefs and to provide a more efficient way to update the strength of beliefs when new evidence is observed(Mahjoub & Kalti, 2011).

5.2.5 Well adapted to other recommending approaches

An important feature of BNs is that they are able to support hybrid recommending approaches of the CF and CBF. The CF might be used to obtain the conditional probability tables and the initial beliefs of a BN (Bart P. Knijnenburg et al., 2012). These beliefs can then be updated in a CBF manner when the network is accessed by a user. This mode of operation enables a predictive model to overcome the data collection problem of the CBF (which requires large amounts of data to be gathered from a single user), while at the same time enables the tailoring of aspects of a collaboratively-learned model to a single user.

5.3 Framework & tools applied for Bayesian UM development

As Table 5.1 explains, different frameworks, libraries, and tools as well as languages such as R, Python, and ASP. NET have been utilized in the Bayesian UM development.

Table 5.1: Frameworks & Tools for Bayesian UM development

Phase	*Programming Language*	*Frameworks & Libraries*	*Tools*
Web-based application & Dataset preparation	ASP.NET JavaScript (JS) HTML CSS	.NET; Bootstrap; Language Integrated Query(linq)	Visual studio.net
Data preprocessing	Python	Pandas; Numpy; openpyxl; xlsxwriter; nltk; sklearn.metrics.pairwise; tabulate; TfidfVectorizer; operator; SV;consine_similarity;	Ubuntu Server 14.04 LTS; Amazon EC2; Jupyter; SPSS
Model structural learning	R	bnlearn	R Studio Netica MSBNx
Model parameter learning			

The existing R packages (libraries) for BN modeling developing are less restricting and well-tested in terms of operations compared to other existing libraries such as PyMC3 library in Python. Some of the R packages only deal with structure learning while others only deal with parameter learning and inference. Table 5.2 illustrates the features of different packages in R for BN modelling. Within the existing BN packages in R, "bnlearn" offers a wide variety of structure learning algorithms, parameter learning approaches (maximum likelihood for discrete and continuous data, Bayesian estimation for discrete data), and inference techniques (cross-validation, bootstrap, conditional probability queries, and prediction) (Albert, 2009). It is also the only package that keeps a clear separation between the structure of a network and the associated probability distribution, which are implemented as two different classes of R objects (Albert, 2009). Hence, "bnlearn" package is applied for developing the BN model in this research.

Table 5.2: Features of BN packages in R (Albert, 2009)

Feature	bnlearn	catnet	deal	pclag	gRbase	gRain
Discrete data	1	1	1	1	1	1
Continuous data	1	0	0	1	1	0
Mixed data	0	0	0	1	0	0
Constrained- based learning	1	0	0	0	1	0
Scored-based learning	1	1	1	1	0	0
Hybrid learning	1	0	0	0	0	0
Structural manipulation	1	1	1	1	0	0
Parameter estimation	1	1	1	1	1	0
Prediction	1	1	1	0	0	1
Approximate inference	1	0	0	0	0	1

References: Packages bnlearn (Scutari, 2010, 2012), deal (Bøttcher and Dethlefsen, 2003), pcalg (Kalisch et al., 2012), and catnet (Balov and Salzman, 2012), gRbase (Højsgaard et al., 2010) and gRain (Højsgaard, 2010) (1→ Yes;) 0→No)

To implement model construction and parameter learning, the R Studio 3.4.0 has been used in this study.

5.4 Bayesian Network algorithms

Generally, the BN modeling development task involves some major activities including structural, parameter learning, and inference (Darwiche, 2009) (Margaritis, 2003) (Korb & Nicholson, 2003). The structure and parameters learning can be yielded either by experts' knowledge or by automatic learning from a dataset, provided that the dataset is complete and unbiased (Tibshirani et al., 2013) (Flores et al., 2011). Despite the variety of theoretical backgrounds and terminology, the methods for automated learning of structure and parameter are categorized into three methods: constrained-based, scored-based or metric/search –based (Albert, 2009)(Amirkhani, Rahmati, Lucas, & Hommersom, 2016; Margaritis, 2003), and hybrid methods (Korb & Nicholson, 2003). Figure 5.3 depicts a classification of BN algorithms exploited in the existing studies along with an algorithm example.

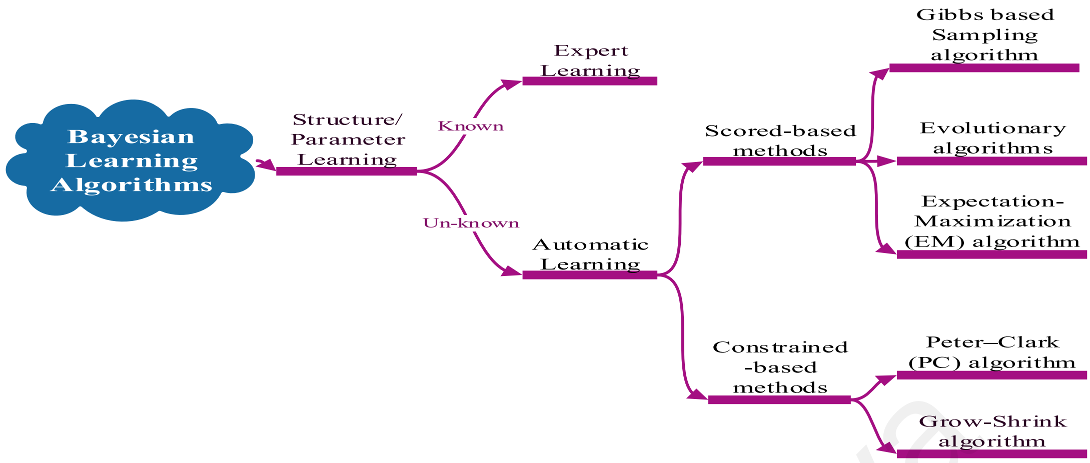


Figure 5.3: Algorithms applied for Bayesian Network learning

Constraint-based methods apply information about conditional independencies gained by performing statistical significance tests on the data(Albert, 2009) (Guo, 2011).Score-based methods search for a BN to minimize or maximize a score (Flores et al., 2011; Korb & Nicholson, 2003) (Albert, 2009) (Guo, 2011). In fact, scored and search based methods have two major components: a scoring metric that measures every candidate BN using a score function with respect to a dataset and a search procedure to move through a solution space composed by possible BNs (Guo, 2011). The Bayesian inference algorithms are also used to calculate marginal probabilities, given an evidence set (Scutari & Denis, 2014). The Grow-Shrink algorithm (constrained based algorithm) in the "bnlearn" package is one of the most recommended algorithms for the BN structure, parameter, and computation of the posterior probability distribution (Albert, 2009) which is applied in this research. The performance of the Grow-Shrink algorithm is compared to a hybrid BN algorithm by applying the expected loss metric which is discussed in Chapter 7.

It is noted that the BN parameter learning can be conducted by the domain experts and also knowledge engineering method but the most basic problem here is finding suitable experts who have the time and interest to assist with the parameter learning modelling process.

Another difficulty is that humans including expert humans, almost always display various kinds of bias in estimating probabilities. One of the biases in estimating probabilities is the tendency to attribute higher than justifiable probabilities to events that have a probability sufficiently greater than 0.5 (Flores et al., 2011; Korb & Nicholson, 2003). This bias estimating is so- called overconfidence. Thus, an event which objectively has a probability of 0.9 is usually attributed a probability that is somewhat higher. Availability is another difficulty of parametrising by using exerts' knowledge. Availability is considered as assessing an event as more probable than is justifiable since it is easily remembered or more salient. There is an extensive and problematic literature on assessing these biases and proposals to eliminate the bias of human probability estimates. Therefore, the BN model in this research is semi-known and parameters (conditional probabilities) are unknown and are computed by the data.

5.5 Bayesian UM development: Addressing RQ2

The process of BN model development in this research is accomplished in three phases of dataset preparation, structure, and parameter learning. Figure 5.4 depicts the phases and major activates performed in this research for BN model development. The evaluation of proposed UM is discussed in Chapter 7.

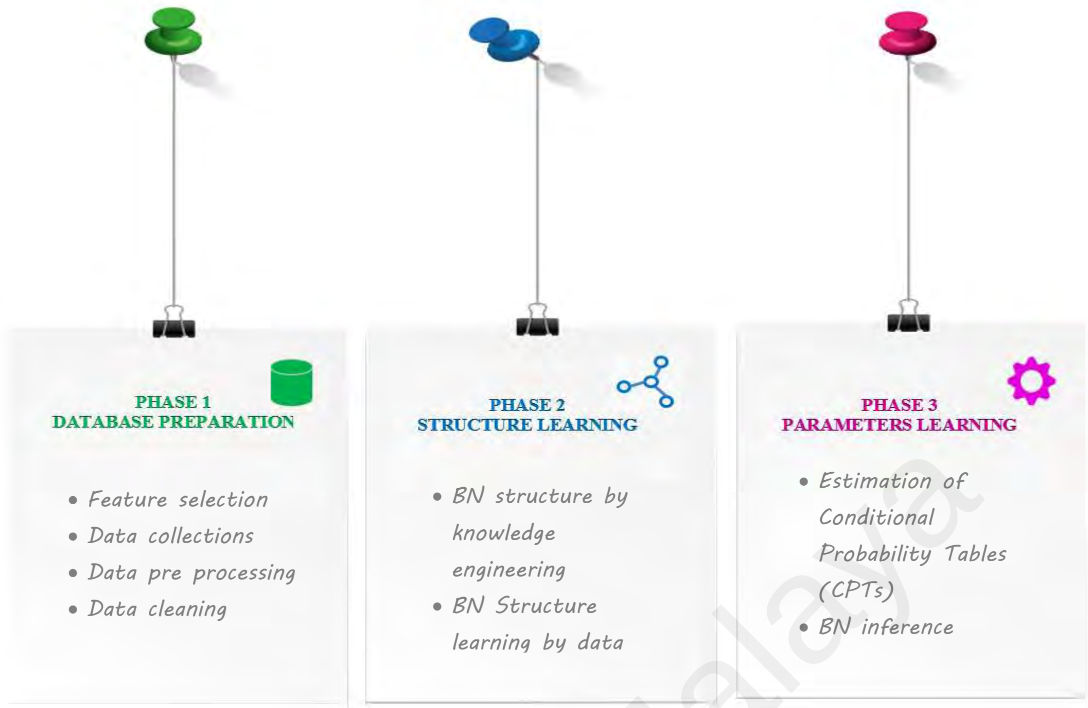


Figure 5.4: BN development phases

5.5.1 Dataset preparation

In practice, obtaining the sufficient data for user modeling to deliver high quality recommendations is difficult (Mobasher, 2013) (Berkovsky et al., 2008). Among the current datasets using for SRSs, there is no dataset which includes the required contexts as well as indexes or labels for novel, diverse, popular, and accurate papers. Therefore, there is an urgent need to prepare an appropriate dataset for the BN modeling. In the following, the activities undertaken for the dataset preparation are discussed.

5.5.1.1 Feature selection

The experts unanimously agree that the feature selection (engineering) is a key and vital step in the success of applied ML which leads to improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and a better understanding of the underlying process (Guyon & Elisseeff, 2003). Besides, it overcomes the problem of dimensionality curse (Sarwar, Karypis, Konstan, & Riedl, 2000) which mostly happens in processing of too many features in a ML problem. Apart from the problem, the feature selection reduces the unbounded computational processing to discover useful knowledge patterns. Consequently, it reduces the extra costs for system development, especially in the data acquisition(Guyon & Elisseeff, 2003). In the contextual modeling, using too many contexts leads to computational complexity. Therefore, identification of the relevant contexts is considered as the feature selection for the data or dimension reduction to reduce the complexity and ambiguity in the system (Mobasher, 2013) (Berkovsky et al., 2008).

As discussed in chapter 2, researchers have applied different methods such as correlation analysis, regression, and factor analysis to find the relevant contexts and there is no technique that is widely accepted as the best. In this research, PLS-SEM method, which is based on the partial least square regression, elaborated in chapter 4, has been applied to find the relevant contexts influencing UX of SRSs. Accordingly, the users' contexts in addition to four levels of novelty, diversity, accuracy and popularity are exploited to contribute in the UM development. Among the environment contexts, time is used in this research as well.

5.5.1.2 Data acquisition (web-based app)

The goal of Bayesian UM is to diagnose the users' information need in terms of four levels of accurate, novel, diverse, and popular papers based on the contextual data. As mentioned earlier, among the current SRSs datasets, there is not an appropriate dataset coupled with the goal of this research. Therefore, one of the research contributions of this study is to prepare the required dataset, which can be used for the future studies in this field.

To acquire the data, a web-based application was developed. The data acquisition procedure was composed of a large-scale questionnaire survey. The questionnaire's fields (items/ features of each context) were designed based on the comprehensive studies that have been performed in chapter 4 on the contexts influencing UX of SRSs. The details of identification and validation of all items/ features of each context have been elaborated in Chapter 4 and further information is provided in Appendix G.

This study targets the population of computer science scholars including master, PhD students, post-doc researchers as well as faculty members. For the privacy, more details of participants are kept confidential.

Requirements for participating in this survey are listed as below;

- Participant must be a master or PhD student or Postdoc or Faculty member in computer fields.
- Participant must have the experience of working with WoS bibliographic database.
- Participants must have minimum 15 minutes time to conduct and finish the survey.

For this survey, the scholars were asked to conduct searches for suitable papers for their current work in a naturalistic setting by using Web of Science (WoS) bibliographic database. This data collection was performed in two steps; in step 1 as shown in Figure 5.5, the participants are asked to submit their current contexts such as task, pre-knowledge, etc. If the participants have no idea of how to respond and fill out the form, more information is provided by clicking the blue question mark buttons provided for each field. Also, the clear button enables the participants to remove the whole information in the form if it is required.

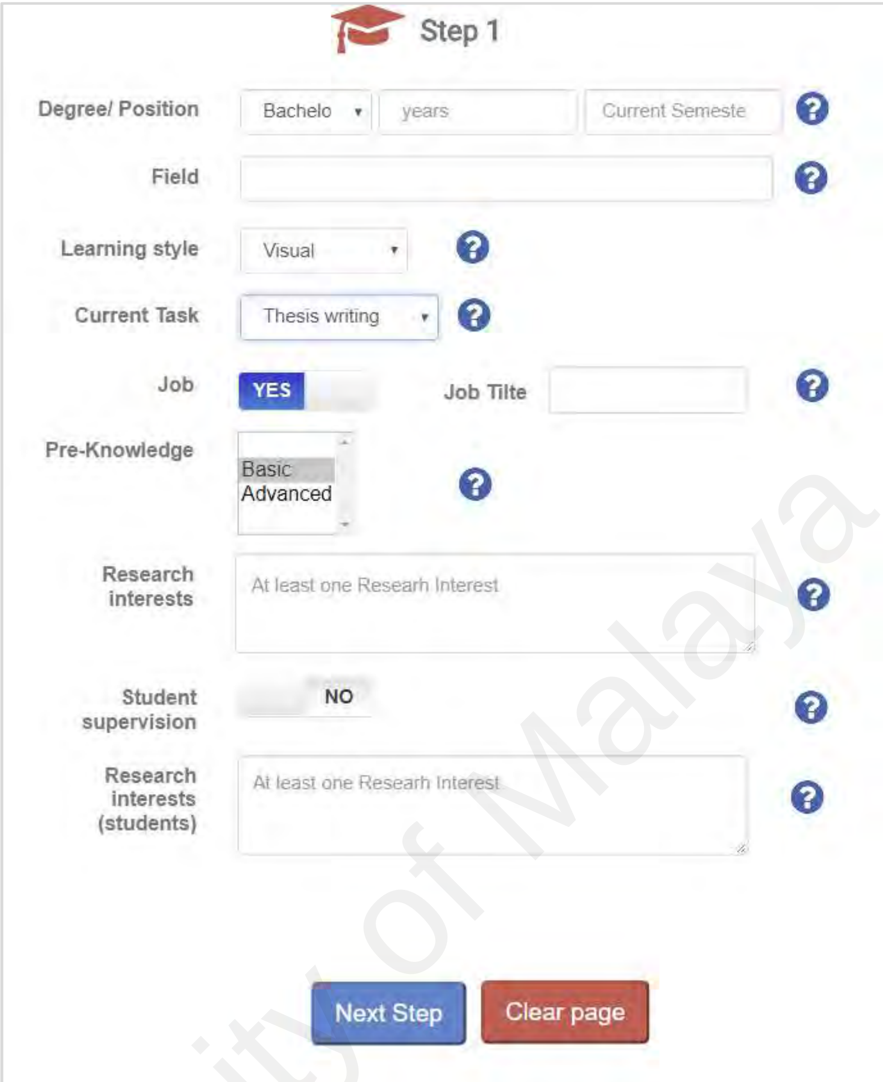


Figure 5.5: Acquisition of Bayesian data-Step 1

In Step 2, the scholars were asked to select the most appropriate paper relevant to their current needs and to rate the paper in a 5 Likert scale in terms of novelty, accuracy, popularity, and diversity as shown in Figure 5.6. Finally, participants were required to submit the paper ID (identification paper produced in WOS) or upload the paper in the PDF format.

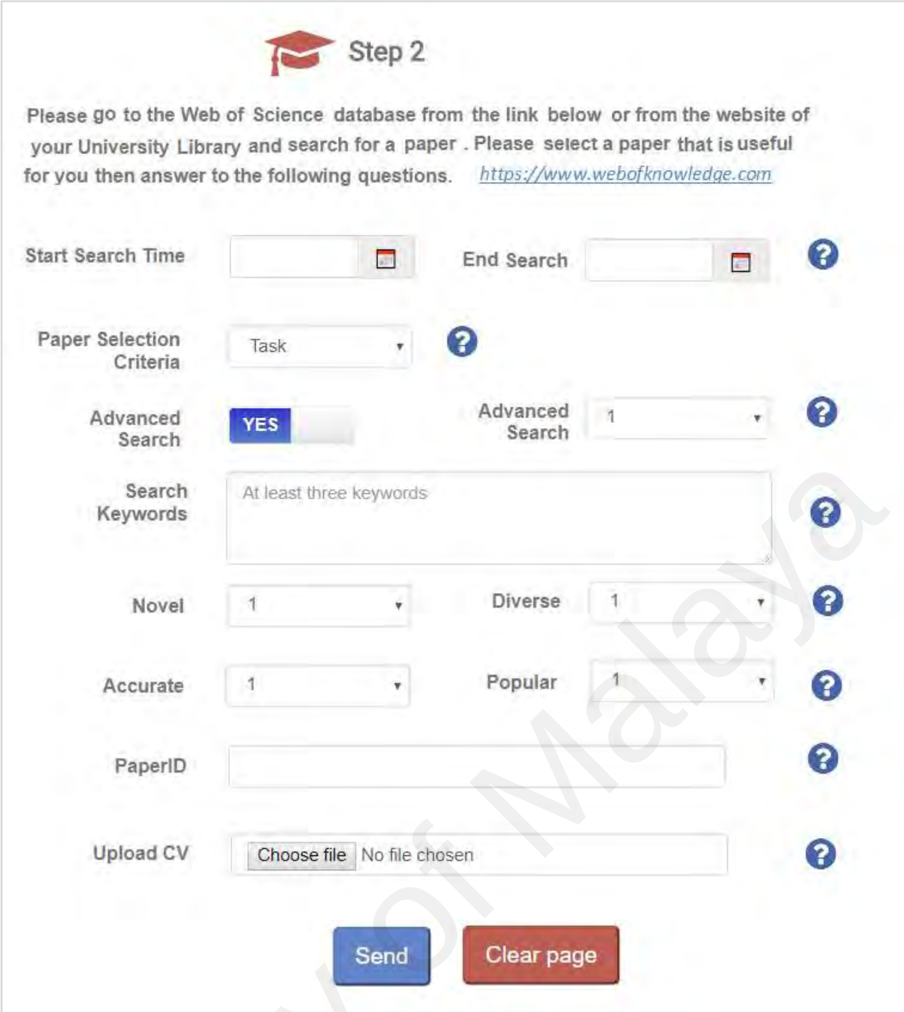


Figure 5.6: Acquisition of Bayesian data-Step 2

The data acquisition survey was conducted from January 2016 until 30 July 2017. In the primary assessments, 1121 records have been registered during the one and half a year data collection period. After dataset preparation, a few tasks were accomplished to prepare the final dataset for the BN model learning. In the following, the pre-processing tasks are discussed.

5.5.1.3 Data pre- processing

A crucial step in ML is pre-processing (Kotsiantis, Kanellopoulos, & Pintelas, 2006). Depending on the data and the selected methods for the problem which is going to be solved (research purpose), there are many pre-processing tasks that can be undertaken (Flores, Nicholson, Brunskill, Korb, & Mascaro, 2011). In this research, a few activities have been performed in order to prepare the final dataset for the BN modeling. As shown in Figure 5.7, after users' and paper's data collection through the web application, the papers' data such as title, authors, and citations were retrieved from the WOS. Then, the CSV data files of users' data and paper's data were merged to make the final dataset.

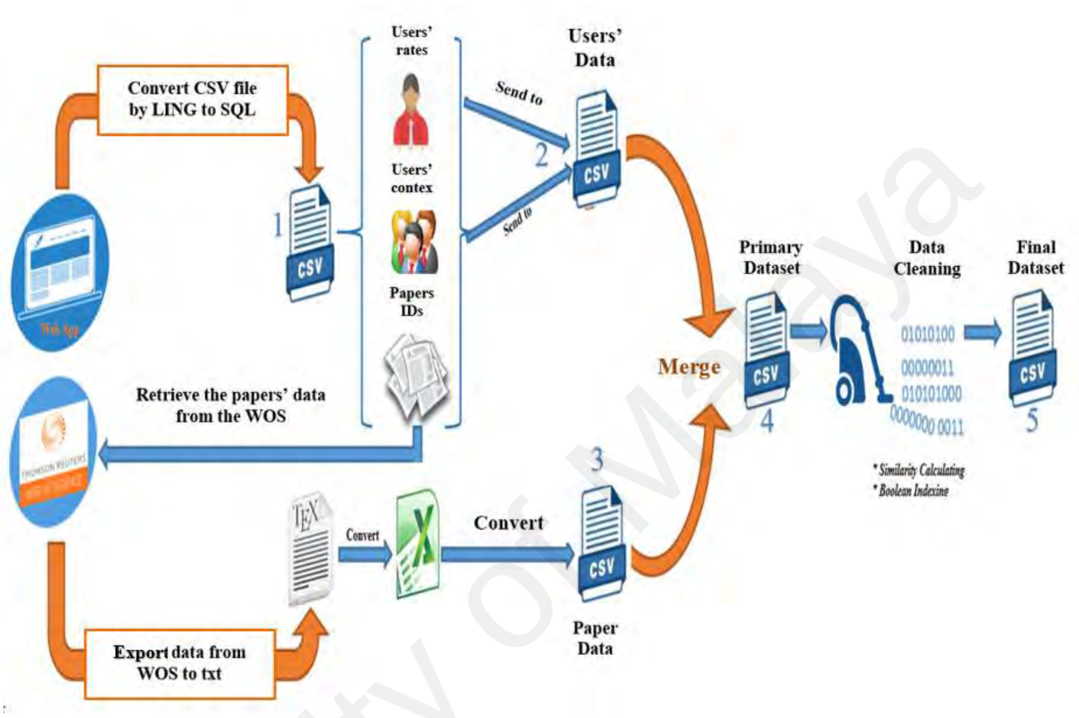


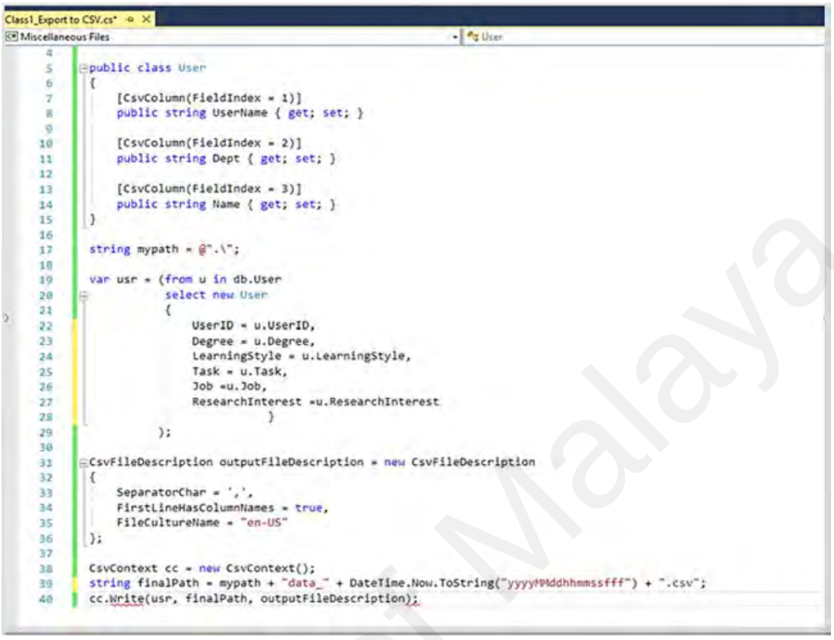
Figure 5.7: Final dataset preparation process

For the data pre-processing and data cleaning tasks, mostly the libraries in Python such as numpy and pandas were chosen. The hand out libraries in Python have a great strength among all existing options with supreme functionality for analyzing and cleaning the data (Bird, Klein, & Loper, 2009). The details of the tasks are discussed in the following sections.

(a) Exporting data to CSV file by LINQ to SQL query

The collected data by the web application is required to be exported to a CSV file for the further analysis. It is possible to export the data from the web application to a CSV file by querying to database through LINQ to SQL query in the Visual studio.net (Figure 5.8).

This original CSV file (CSV file 1 in Figure 5.7) contains users' contexts, users' ratings and papers' IDs. The users' data was transformed into a separate CSV file (CSV file 2 in Figure 5.7).




```
public class User
{
    [CsvColumn(FieldIndex = 1)]
    public string UserName { get; set; }

    [CsvColumn(FieldIndex = 2)]
    public string Dept { get; set; }

    [CsvColumn(FieldIndex = 3)]
    public string Name { get; set; }
}

string mypath = @".\";

var usr = (from u in db.User
           select new User
           {
               UserID = u.UserID,
               Degree = u.Degree,
               LearningStyle = u.LearningStyle,
               Task = u.Task,
               Job =u.Job,
               ResearchInterest =u.ResearchInterest
                       }
         );

CsvFileDescription outputFileDescription = new CsvFileDescription
{
    SeparatorChar = ',',
    FirstLineHasColumnNames = true,
    FileCultureName = "en-US"
};

CsvContext cc = new CsvContext();
string finalPath = mypath + "data_" + DateTime.Now.ToString("yyyyMMddhhmmssfff") + ".csv";
cc.Write(usr, finalPath, outputFileDescription);
```

Figure 5.8: Data export by using LINQ to SQL query

The primary dataset contains 1121 participants' records. Among the collected records, 675 participants have entered the paper's IDs and 446 participants have uploaded the PDF paper files. However, 10 records of users' data were invalid and 58 files of paper data were useless and irrelevant files. Therefore, the whole 68 invalid records have been eliminated from the dataset and the final dataset size is 1053 records.

(b) Importing data from text to CSV

As four levels of accuracy, diversity, popularity, and novelty of the papers have been considered in the BN modelling, more bibliographic information about the papers such as title, authors, and keywords are also required. The papers' bibliographic information is provided by using the WOS exporting option. Figure 5.9 shows the procedure of data exporting the data from WOS and then emporting the plain text to a CSV file (CSV file 3 in Figure 5.7) by using convert option in the Microsoft EXCEL.

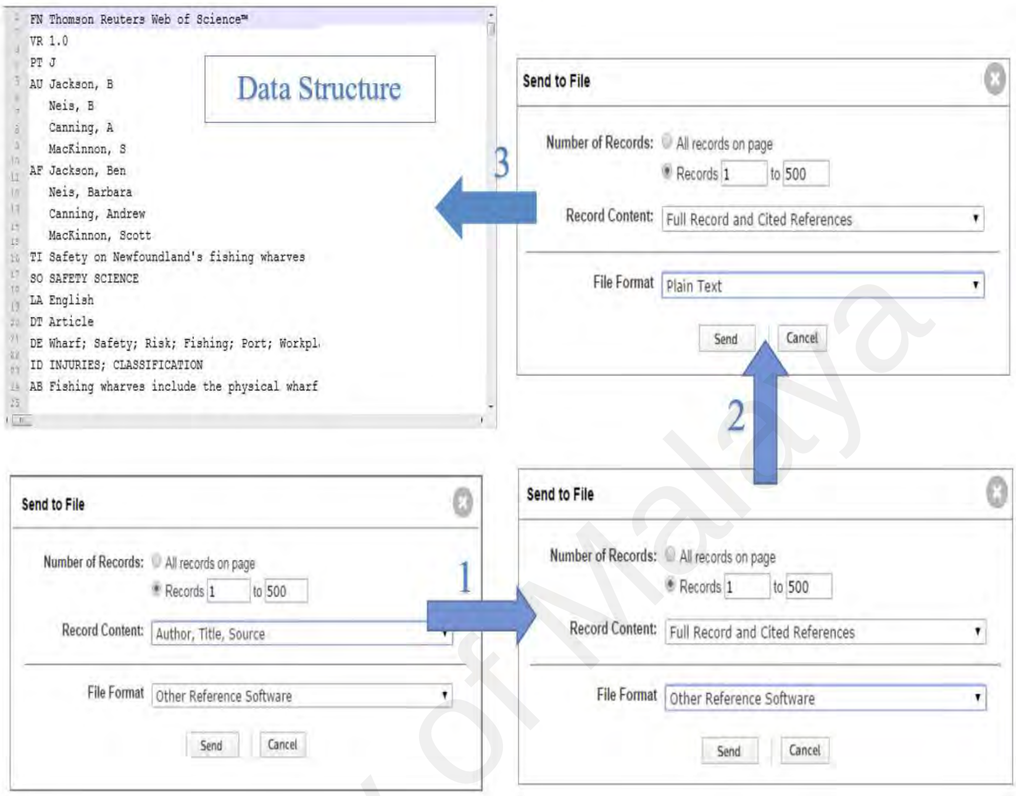


Figure 5.9: Paper data import from the WOS

The WOS approximately retrieves 73 bibliographic features for a paper. The list of the whole paper's information is provided in the Appendix K.

(c) Combining datasets

By taking on the above-mentioned tasks, two CSV files were produced. The CSV file 2 (Figure 5.7) includes the users' data and CSV 3 (Figure 5.7) includes the papers' data imported from the WoS. For the further data analysis and data cleaning, the datasets were merged by using the `pd.merge ()` function in numpy library as follows in the following codes shown in Figure 5.10.

```
import pandas as pd
import numpy as np
db1= pd.read_csv('user.csv')
db2= pd.read_csv('paper.csv')
Merged= pd.merge(db1.db2,how= 'left', on ='num')
Merged.to_csv('finaldb.csv', index =False)
```

Figure 5.10: Dataset combining

5.5.1.4 Data cleaning

Handling the missing data and incorrect (invalid) values are the most common task in data cleaning. In this research, the users' data were collected through a web application. Therefore, most of the input data were validated runtime during the data collection using JavaScript and HTML 5 by throwing validity exceptions in terms of data type validation, data range validation, and constraint validation. The participants are able to submit the data while the inputs are valid and non-null. Figure 5.11 shows a sample of data validity controls which allows users to upload only PDF files for the paper data in step 2 of data collection.

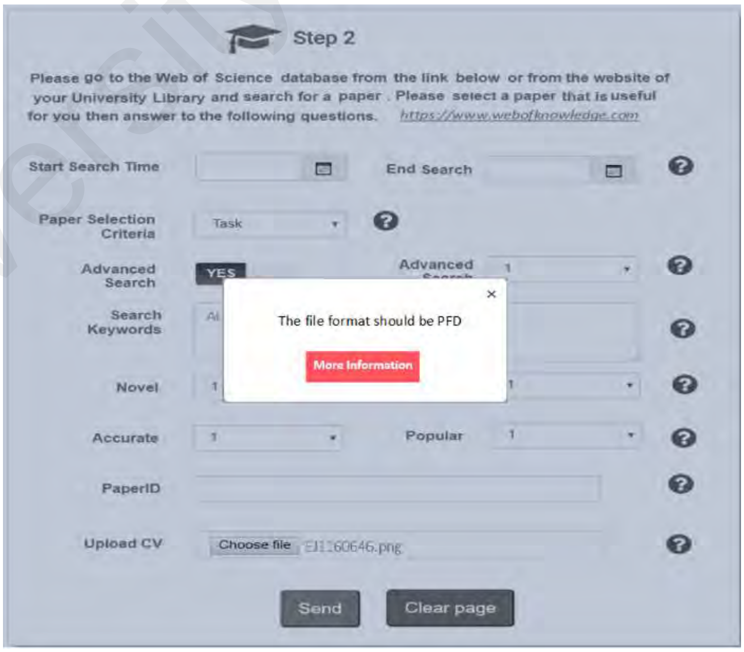


Figure 5.11: Throwing validity exceptions in data collection

Additionally, to ensure the data accuracy, a sample of 200 records were selected randomly and examined manually in order to check the quality of the data. Among the records in the selected sample, no invalid data was detected.

As mentioned earlier, the WOS approximately provides 73 bibliographic features for a paper (Appendix K). In this research, only 15 features have been exploited and the rest such as PI: Publisher City, PA: Publisher Address, SN: International Standard Serial Number (ISSN), BN: International Standard Book Number (ISBN), DI: Digital Object Identifier (DOI) were omitted from the papers' dataset. The empirical results in chapter 4 revealed that the valid indicators contributed to the formation of novel, diverse, popular, and accurate papers (Appendix G); therefore, the irrelevant identified bibliographic features were not considered in the UM development.

5.5.1.5 Numerical data discretisation

Based on guidelines provided by Korb & Nicholson (2003), the nodes that take mutually exclusive and exhaustive discrete values are recommended for the BN modeling. It means that the variable should take on exactly one of the values at a time. For example, a user might have two scholarly tasks at the same time but in the BN modelling, one of the tasks should be considered at a time. The Boolean nodes (binary values of true & false), ordered values (low, medium, high), and integral values (values from 1 to 120 for the users' age) are three common types discrete nodes in BN modeling (Korb & Nicholson, 2003).

In this research, most of the variables have different states such as `LS = [visual, verbal, physical]; TA = [Thesis writing, Paper writing, Course taking, Topic finding, Course teaching]`. Therefore, based on the guidelines, the best modelling solution is to have Boolean variables for each state. In other words, for the different states of the same variable, separate variables are created and they must be mutually exclusive and it is solved by adding an extra arc between the nodes and a deterministic CPT that enforces the mutual exclusion (Korb & Nicholson, 2003) (Darwiche, 2009). When considering the variables, it is also required to decide what states or values the variable can take. For example, when modeling the user's contexts, the pre- knowledge (PK) variable could take the two values of advanced and basic which are represented by the values of 0 and 1 (binary values) in the final dataset. Moreover, some of the variables might represent integral values. For instance, search time might have possible values of more than 20 minutes or less than 20 minutes. The following code (Figure 5.13) is a sample of the variable definition and of assigning the relevant values using R programming to prepare the dataset for the BN modeling development and evaluation.

```
> bndata <- read.table ("bntrain.csv", header=TRUE, sep=",")

Pre-knowledge = c (0, 1),
labels = c ("Basic", "Advanced")
```

Figure 5.12: Assigning variables' values

The knowledge of engineer method might be very helpful to examine the final variables and also the suitable values (Korb & Nicholson, 2003) which is discussed more in the BN structure learning phase.

Most of the variables in this research are already discrete or transferred to Boolean/binary scales. There are only two continuous variables (time and paper impact factor). To cope with the continuous variables, the selected BN learning algorithm is appropriate for continuous variables to automatically convert them to discrete data types. Thus, by using an objective function and a search algorithm, the discretization algorithm estimates the cut-off points for numerical attributes, splits them into well-defined numerical ranges covering the whole numerical domain. More details on the BN learning algorithm is provided in BN structure and parameter learning phases.

5.5.1.6 Converting text data to numeric

In the data collection phase, the participants were asked to identify their research interests and input search keywords which are text data and should be converted to the numeric data type for the BN learning (Korb & Nicholson, 2003). For the diagnosis of the problem in this research, the similarities between users' input search keywords and paper's keywords, title, and abstract were calculated using Cosine Similarity (CS). CS is a measure, which calculates the similarity between two texts or vectors (on the vector space) by calculation of the cosine of angle between two vectors (Huang, 2008). As;

Cos 0° = 1 → two vectors are similar

Cos 90° = 0→ two vectors are not similar

The cosine of two vectors (non-zero) is calculated by the Euclidean dot product formula, therefore the cosine similarity, cos(θ), is defined as similarity between two vectors a and b (Huang, 2008):

$$\boldsymbol{a} \,.\, \boldsymbol{b} = \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \, \cos \boldsymbol{\theta} \tag{5.8}$$

$$\textbf{Similarity} = \cos \boldsymbol{\theta} = \frac{\boldsymbol{a} \,.\, \boldsymbol{b}}{\|\boldsymbol{a}\| \, \|\boldsymbol{b}\|}$$

To calculate the similarity between the above –mentioned text variables, first, the target columns were selected from the data () and tabulated separately. Then, by using two functions of dep_product and getSimilarity, the similarity was calculated. To calculate the cosine similarity, NLTK, sklearn.metrics.pairwise libraries as well as sklearn.feature_extraction.text have been utilized (Bird, 2009) (appendix L). The CS function returns a value (between 0 and 1) which is the similarity between two columns (e.g. users' search keywords (UKW) and paper's keywords (PKW)). Figure 5.13 shows the CSs between the text data for the first five records of dataset. Therefore, the text data (columns) of users' search keywords, papers' keywords and abstract have been removed and the CSs were added to the dataset.

User search key words	Paper keywords	Abstract	Title	UKW-PKW	UKW-PKW	UKW-PTI
UKW	PKW	PAB	PTI	CS_1	CS_1	CS_3
posture recognition human factors	Sign language Gesture recognition Feature selection	Hand gestures are an intuitive way for humans to interact with computers…	Multi-objective optimization for hand posture recognition	0.204124	0.150946	0.377964
Neural network metrics and methods	Siamese neural networks Metric learning Human action recognition	This paper focuses on metric learning with Siamese Neural Networks (SNN)…	Class-balanced siamese neural networks	0.00	0.035399	0.010313
Sentiment analysis attentive behavior and learning style	Ambient intelligent ,Machine learning ,Learning activities Attentiveness Learning styles	Learning styles are strongly connected with learning and when …	Characterizing attentive behavior in intelligent environments	0.00	0.445399	0.138975
User interface design in electrooculography	Electrooculography (EOG) Eye gesture Unit saccadic signals Common spatial pattern (CSP) Joint approximate diagonalization Support vector machine (SVM)	People with motor diseases have suffered from deprivation of both verbal …	Design and implementation of an eye gesture perception system based on electrooculography	0.003496	0.203490	0.129099
Interaction design theory system design user centered design	Nonantropocentric design; posthumanism; sm art city; anthropocene; meshwork; dichotomies; urban planning; theory	While the smart city agenda is critiqued for its focus on technology and business led solutions, …	Nonanthropocentric design and smart cities in the anthropocene	0.3223291	0.028171	0.283473

Figure 5.13: Calculation of cosine similarity for the text data

The similarities between the users' interests and the uploaded papers were assessed through the "criteria" filed. This field determines the relevance of the uploaded papers to the users' tasks, research interests, etc. If the paper is selected based on the research interests; then, the similarity between these two columns is considered as 1 otherwise it is considered as 0. In this research, the CSs have been calculated for the titles, keywords, and abstracts of the papers. The CS of a paper body was not calculated because, firstly, an accurate CS calculation requires to be compared with classification systems such as the ACM computing classification system (ACM- CCS). The ACM- CCS is a poly-hierarchical ontology based on semantic vocabularies, which reflects the state of the art of the computing disciplines, concepts, and categories. Secondly, text classification is out of the scope of this research.

5.5.1.7 Final dataset

The objective of the Bayesian UM in this study is to diagnose users' information needs for four levels of novel, accurate, diverse, and popular papers based on users' contexts. Figure 15.14 shows, the dataset preparation, BN modeling activities, and evaluation. The BN evaluation is discussed in Chapter 7.

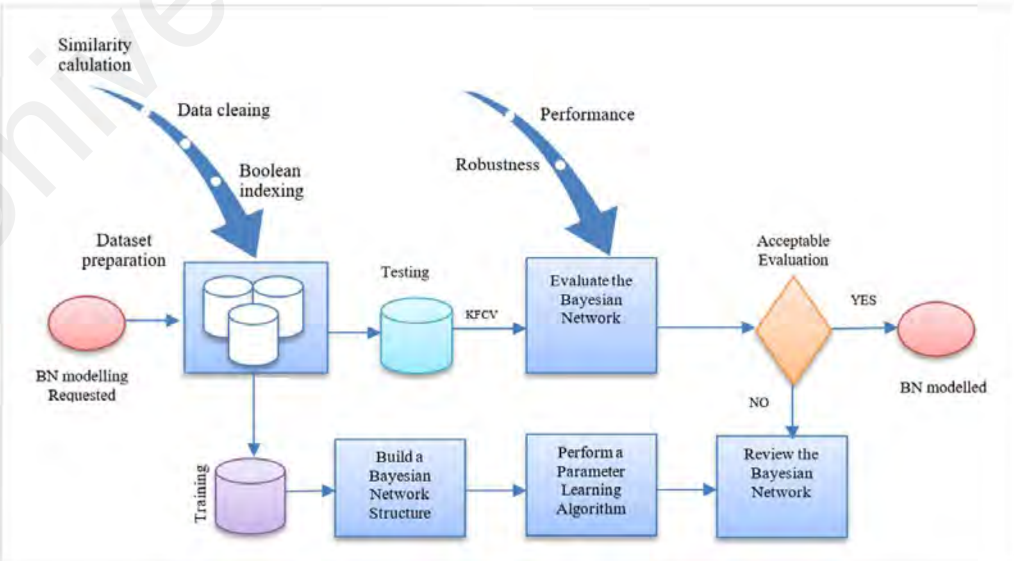


Figure 5.14: BN modeling process

By performing the above pre-processing tasks, the final dataset was produced. As proposed in ML guidelines, when the dataset is not so large, the whole dataset is applied for the training phase and K-Fold Cross Validation (KFCV) method is recommend (Russell & Norvig, 2016) (Tibshirani, James, Witten, & Hastie, 2013).

5.5.1.8 Preparing data codebook

The codebook serves as a reference and ensures that the data is understood and interpreted properly. The data codebook in this research has been created by uploading the final CSV data file in `SPSS Analysis-> Report-> Codebook` (Figure 5.15). It contains data, values, and assigned variable codes for user and paper data files. The data code book also represents more information about the data such as measure (e.g. nominal, ordinal, and scale) and type (i.e. numeric, string; how many characters wide it is; how many decimal places it has) which are omitted for brevity.

Variable	Code	State	Values
Role	RO1	PhD	0-1
	RO2	Master	0-1
	RO3	Post-Doc	0-1
	RO4	Faculty Member	0-1
Pre-knowledge	PK_1	Basic	0-1
	PK_2	Advanced	0-1
Learning Style	LS_1	Visual	0-1
	LS_2	Verbal	0-1
	LS_3	Physical	0-1
Task	TA_1	Thesis writing	0-1
	TA_2	Paper writing	0-1
	TA_3	Course taking	0-1
	TA_4	Topic finding	0-1
	TA_5	Course teaching	0-1

Variable	Code	State	Values
Time	SE_1	First Semester	0-1
	SE_2	Second Semester	0-1
	SE_3	Third Semester	0-1
	SE_4	Fourth Semester	0-1
	SE_5	Semester Above	0-1
Search Time	ST1	More than 5 times	0-1
		Less than 5 times	
Number of Search Modification	SM1	More than 5 times	0-1
		Less than 5 times	
Advanced Search	AS1	Yes	0-1
		NO	
Paper quality	NO1	Novelty	0-1
	PO1	Popularity	0-1
	AC1	Accuracy	0-1
	DI1	Diversity	0-1

Variable	Code	State	Values
Criteria	SI1	UKW-PTI Similarity	0-1
	SI2	UKW-PAB Similarity	0-1
	SI3	UKW-PKW Similarity	0-1
	SI4	URI-PAP Similarity	0-1
	SI5	UJO-PAP Similarity	0-1
	PY1	Publication in current Year	0-1
	SU1	Seen by User	0-1
	RU1	Read by User	0-1
	CN1	Citation Number	0-1
	AR1	Author Reputation	0-1
	JR1	Journal Reputation	0-1
	PR1	Publisher Reputation	0-1
	DN1	Download Number	0-1
	PT1	Conference	0-1
	PT2	Journal	0-1
	PT3	Review	0-1

Figure 5.15: Data codebook

5.5.2 BN structure learning

The BN structure or topology learning is defined a set of relevant variables and their possible values built by connecting the variables into a DAG to represent a network which best describes the observed data (Flores et al., 2011) (Guo, 2011). In particular, two nodes should be directly connected if one affects or causes the other with the arc indicating the direction of the effect (Amirkhani et al., 2016) (Korb & Nicholson, 2003).

The BN structure can be built either manually by knowledge engineer (expert elicitation) or by automated learning from the data. Applying the hybrid method leads to the better analytical and predicting ability of the BN model (Ono et al., 2007) (Flores et al., 2011). In the coming sections, the BN structure using the experts' knowledge and automatic learning are discussed respectively.

5.5.2.1 BN structure by the knowledge engineer

The engineering discipline which involves integrating the high level of human expertise into computer systems in order to solve complex problems is called knowledge engineering (Korb & Nicholson, 2003). The knowledge engineering aims to build a model for an expert system which represents realistically the problem features and is able to reason and respond as close as an expert human (Tibshirani et al., 2013) (Korb & Nicholson, 2003).

To create a BN structure by the knowledge engineers, four sessions have been conducted. In the first three sessions, the draft of the BN structure has been built and in the last session, the structure has been reviewed. Overall, the whole sessions took nine hours. Since the experts were located far from the knowledge engineer (researcher of this study), the most communications were electronic and have been performed by Zoom video conferencing software. The experts who already agreed to contribute in this research are in the fields of scholarly communication and recommending system. The sessions are discussed in the following subsections.

(a) Session 1: variable checking

The knowledge engineer and experts reviewed the variables (presented already in data codebook) that could influence the users' needs of scholarly papers based on four levels of accurate, novel, diverse, and popular. Afterwards, the clarity and consistently examination have been performed. In the clarity test, the whole identified variables were checked to have a clear operational meaning and clear agreements on the following statements.

- Are all the relevant variables included? Are they named usefully?
- Are all states (values) appropriate? Exhaustive and exclusive?
- Are all state values useful, or can some be combined?
- Are state values appropriately named?
- Where variables have been discretized, are the ranges appropriate?

Also, the consistency of the state spaces across different variables was examined. For example, the states or values of a parent and its child must be consistent and do make sense without causing any misunderstanding.

(b) Session 2: relationships

According to the guidelines indicated in the BN modelling literature (Tibshirani et al., 2013) (Korb & Nicholson, 2003) (Flores et al., 2011), it is appropriate to create direct questions about the causes or effects to elicit and identify the relationships between the nodes. Hence, with the help of the experts, the first several questions have been posted and then, based on these questions, the arcs (relations) were added from those causal variables to the affected variables. For the brevity, only the questions conducted for the novel paper node is presented in the following. In this research, the users' perceptions were not considered in the BN model because of the difficulties faced in the data collection of users' perceptions.

Q: "What are contexts that would cause users to need a novel paper?"
A: "topic finding and course teaching"
Modelling: suggests arcs from those nodes to the novel paper node.

Q: "Is there any context which prevents need for a novel paper?"
A: "if the user has already knows and have advanced pre -knowledge in the."
Modelling: suggests an arc from pre- knowledge to the novel paper.

Q: "What can cause a paper to be considered as a novel paper?"
A: "publication data, users' awareness"
Modelling: suggests the arcs from those nodes to a novel paper.

Q: "What are the effects of looking for a novel paper?"
A: "searching behaviour"
Modelling: suggests the arcs from novel paper to different searching behaviour.

Q: "Is there any context that might interfere with the users' need for a novel paper?"
A: "Yes, the users' perceptions"

Q: "what are the possible states and values for the variables incorporating in need for a novel paper?
A: "definition of values and states as indicated in the BN structure model"

(c) Session 3: pairwise relations by the expert elicitation

In the BN structure, the main focus was on capturing expert understanding of the relationships between variables performed by pairwise elicitation method. The expert was provided with a cross-table of 42 rows (R) and 42 columns (C). For each cell in *R* and *C*, the expert indicated the relations providing the below signs.

→	R directly causes C
≺	R occurs before C
-	R and C are directly related
~	R and C are correlated

To reduce the complexity of BN structure and the possibilities of inconsistent information as well as elicitation burden on the expert, which consequently cause mistakes, the reverse direction relationships were not considered. The result of the expert elicitation was the set of priors for pairwise relations shown in Figure 5.16. The 934 cells out of 1764 (56.26%) are signed by × which represents no relations. As mentioned, the sign of – or ~ also shows the non-directional relationship. The total of 105 relations were indicated as → (directly causes), 7 ≺ (R occurs before C) and 718 ~, - (correlations & relations) have specified by the knowledge engineering method which provided 830 of the 1764 possible pairwise relationships. In addition, there are no ⇒ relationships, as the experts did not consider non-direct causality a natural relationship to specify.

In some studies, the Delphi method (Hsu & Sandford, 2007) or focus group method (Kontio, Lehtola, & Bragge, 2004) is recommended to validate expert opinion and to elicit information independently from a group of experts to reach a consensus (Korb & Nicholson, 2003). In this study, only one expert has contributed into the BN structure by practical considerations because among the three relevant experts identified in this field, only one individual agreed to dedicate this research. Moreover, this study also applies automated BN construct learning using the dataset, which avoids adding the additional complexity to the BN structure process.

(d) Session 4: BN structure review

At the final session, the graph structure was reviewed. It is about looking at the implications of the d-separation dependencies and independencies and at whether the structure violates any prior knowledge about time and causality considering the pairwise relations. The BN structure for four paper levels (Figure 5.17) elicited by the experts' knowledge was drawn by MSBNx. MSBNx is a component-centric toolkit for BN modelling designed at Microsoft company (Kadie, Hovel, & Horvitz, 2001).

C

R

VAR	RO1	RO2	RO3	RO4	PK1	PK2	LS1	LS2	LS3	TA1	TA2	TA3	TA4	TA5	SE1	SE2	SE3	SE4	SE5	ST1	SM1	AS1	SI1	SI2	SI3	SI4	SI5	PY1	SU1	RU1	CN1	AR1	JR1	PR1	DN1	PT1	PT2	PT3	NO1	PO1	AC1	DI1
RO1		×	×	×	–	–	×	×	×	~	~	~	~	~	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	×	×	×	×	×	×	–	–	–	~	~	~	~
RO2	×		×	×	–	–	×	×	×	~	~	~	~	~	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	×	×	×	×	×	×	–	–	–	~	~	~	~
RO3	×	×		×	–	–	×	×	×	~	~	~	~	~	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	×	×	×	×	×	×	–	–	–	~	~	~	~
RO4	×	×	×		–	–	×	×	×	~	~	~	~	~	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	×	×	×	×	×	×	–	–	–	~	~	~	~
PK1	~	~	~	~		–	×	×	×	~	~	×	×	×	×	×	×	×	×	~	~	~	×	×	×	×	×	×	~	~	×	×	×	×	×	–	–	–	~	~	~	~
PK2	~	~	~	~	≺		×	×	×	~	~	×	×	×	×	×	×	×	×	~	~	~	×	×	×	×	×	×	~	~	×	×	×	×	×	–	–	–	~	~	~	~
LS1	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	–	–	–	~	×	×	×	~	~	~	~
LS2	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	–	–	–	~	×	×	×	~	~	~	~
LS3	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	–	–	–	~	×	×	×	~	~	~	~
TA1	~	~	~	~	→	→	–	×	×		×	×	×	×	×	×	×	×	~	×	×	×	–	–	–	–	–	–	×	×	×	×	×	×	×	–	–	–	~	~	~	~
TA2	~	~	~	~	→	→	×	×	×	×		×	×	×	×	×	×	×	~	×	×	×	–	–	–	–	–	–	×	×	×	×	×	×	×	–	–	–	~	~	~	~
TA3	~	~	~	~	→	→	×	×	×	×	×		×	×	~	~	~	×	~	×	×	×	–	–	–	–	–	–	×	×	×	×	×	×	×	–	–	–	~	~	~	~
TA4	~	~	~	~	→	→	×	×	×	×	×	×		~	~	~	~	×	×	×	×	×	–	–	–	–	–	–	×	×	×	×	×	×	×	–	–	–	~	~	~	~
TA5	~	~	~	~	→	→	×	×	×	×	×	×	×		~	~	~	~	~	×	×	×	–	–	–	–	–	–	×	×	×	×	×	×	×	–	–	–	~	~	~	~
SE1	~	~	~	~	~	~	×	×	×	×	×	~	~	~		×	×	×	×	×	×	×	×	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	~	~	~	~
SE2	~	~	~	~	~	~	×	×	×	×	×	~	~	~	×		×	×	×	×	×	×	×	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	~	~	~	~
SE3	~	~	~	~	~	~	×	×	×	×	~	~	~	~	×	×		×	×	×	×	×	×	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	~	~	~	~
SE4	~	~	~	~	~	~	×	×	×	~	~	~	×	~	×	×	×		×	×	×	×	×	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	~	~	~	~
SE5	~	~	~	~	~	~	×	×	×	~	~	~	×	~	×	×	×	×		×	×	×	×	×	×	×	×	×	~	~	×	×	×	×	×	×	×	×	~	~	~	~
ST1	~	~	~	→	→	≺	≺	~	~	×	×	×	×	×	×	×	×	×	×		×	×	~	~	~	~	~	×	×	×	×	×	×	×	×	×	×	×	~	~	~	~
SM1	~	~	~	→	→	≺	≺	~	~	×	×	×	×	×	×	×	×	×	×	×		×	~	~	~	~	~	×	×	×	×	×	×	×	×	×	×	×	~	~	~	~
AS1	~	~	~	→	→	≺	≺	~	~	×	×	×	×	×	×	×	×	×	×	×	×		~	~	~	~	~	×	×	×	×	×	×	×	×	–	–	–	~	~	~	~

Figure 5.16: Pairwise relations through the expert elicitation process

C →

R ↓

VAR	RO1	RO2	RO3	RO4	PK1	PK2	LS1	LS2	LS3	TA1	TA2	TA3	TA4	TA5	SE1	SE2	SE3	SE4	SE5	ST1	SM1	AS1	SI1	SI2	SI3	SI4	SI5	PY1	SU1	RU1	CN1	AR1	JR1	PR1	DN1	PT1	PT2	PT3	NO1	PO1	AC1	DI1
SI1	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	—	—	—	×
SI2	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	—	—	—	×
SI3	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	—	—	—	×
SI4	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	—	—	—	×
SI5	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	—	—	—	—
PY1	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	—	—	—	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	—	×	×	×
SU1	—	—	—	—	—	—	~	~	~	~	~	~	~	~	×	×	×	×	×	—	—	—	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	—	×	×	×
RU1	—	—	—	—	—	—	~	~	~	~	~	~	~	~	×	×	×	×	×	—	—	—	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	—	×	×	×
CN1	×	×	×	×	×	×	~	~	~	~	~	~	~	~	×	×	×	×	×	—	—	—	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	—	×	×
AR1	×	×	×	×	×	×	~	~	~	~	~	~	~	~	×	×	×	×	×	—	—	—	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	—	×	×
JR1	×	×	×	×	×	×	~	~	~	~	~	~	~	~	×	×	×	×	×	~	~	~	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	—	×	×
PR1	×	×	×	×	×	×	~	~	~	~	~	~	~	~	×	×	×	×	×	~	~	~	×	×	×	×	×	×	×	×	×	×			×	×	×	×	×	—	×	×
DN1	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	—	—	×
PT1	×	×	×	×	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	×	×	×	×	×	×	×	×		×	×	—	—	—	—
PT2	×	×	×	×	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	×	×	×	×	×	×	×	×	—		—	—	×	—	—
PT3	×	×	×	×	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	×	×	×	×	×	×	×	×	×	×		—	—	—	—
NO1	~	~	~	~	—	→	—	—	→	→	→	→	→	→	—	—	→	→	→	→	→	→	→	→	→	→	→	→	→	→	×	×	×	×	×	→	→	→		×	×	×
PO1	~	~	~	~	→	—	→	—	—	→	→	—	→	→	→	→	—	→	—	→	→	→	→	→	→	→	×	×	×	×	→	→	→	→	→	→	→	→	×		×	×
AC1	~	~	~	~	→	→	→	→	→	→	→	→	→	→	→	→	—	→	—	→	→	→	→	→	→	→	×	×	×	×	×	×	×	×	×	→	→	→	×	×		×
DI1	~	~	~	~	—	→	—	—	—	—	→	—	—	—	—	→	→	—	→	→	→	→	→	→	→	→	×	×	×	×	×	×	×	×	×	→	→	→	×	×	×	

Figure 5.16: Continoued

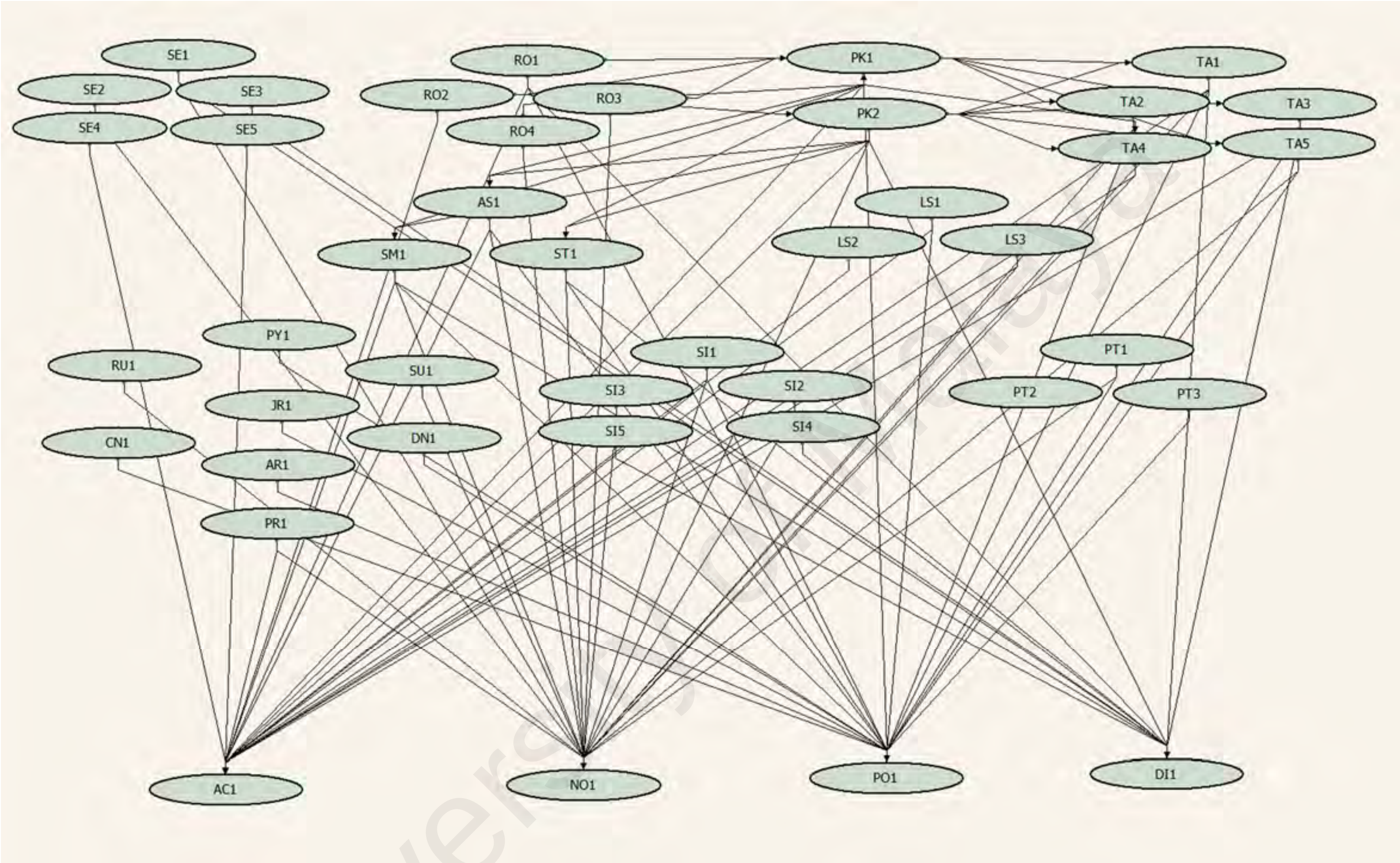


Figure 5.17: BN structure by expert's elicitation

5.5.2.2 Automated BN structure by data (GS algorithm)

The purpose of this section is to illustrate BN structure by data computations using "bnlean" library based on the guidelines given by Albert (2009). The latest version of "bnlearn" and its dependencies are available at CRAN (https: //cran. r-project.org,) (Gentleman et al. 2010) (Scutari & Denis, 2014). As mentioned earlier, the GS algorithm was implemented in this study to learn the BN structure by the data. Also, the Pearson's Linear Correlation (Cronbach's α: 0.05) was used as the conditional independence test.

After loading the data, an empty network with the nodes (variables) was made using the `empty.graph` function. `> ug = empty.graph(names(bn train))`. In addition, by `Complete.subset` function, the missing data was checked once again and the BN object was built accordingly using the `bn` class, which provides description of the network structure.

The structure of the BN associated with the dataset was learned with the GS algorithm, implemented in the gs function, and stored in an object of class bn. As Figure 5.18 shows, a node is a parent of a child, if there is an arc from the former to the latter (TA_4 & PK_1 are parents of ST_1). Therefore, if each node is defined as X, then BN structure is defined as;

$$\boldsymbol{P}(\boldsymbol{x_1}, \boldsymbol{x_1}, \dots, \boldsymbol{x_n}) = \prod_{\boldsymbol{i}} \boldsymbol{P}\ (\boldsymbol{x_i} | \boldsymbol{Parents}\ (\boldsymbol{x_i})) \tag{5.9}$$

$$\text{Provided } \boldsymbol{Parents}\ (\boldsymbol{x_i}) \subseteq \{\boldsymbol{x_1}, \boldsymbol{x_1}, \dots, \boldsymbol{x_{i-1}}\}$$

In this directed chain of nodes, TA_4 & PK_1 are ancestors of NO_1 & AC_1 since they appear earlier in the chai and AC_1 is a descendant of TA_4 node because it comes later in the chain.

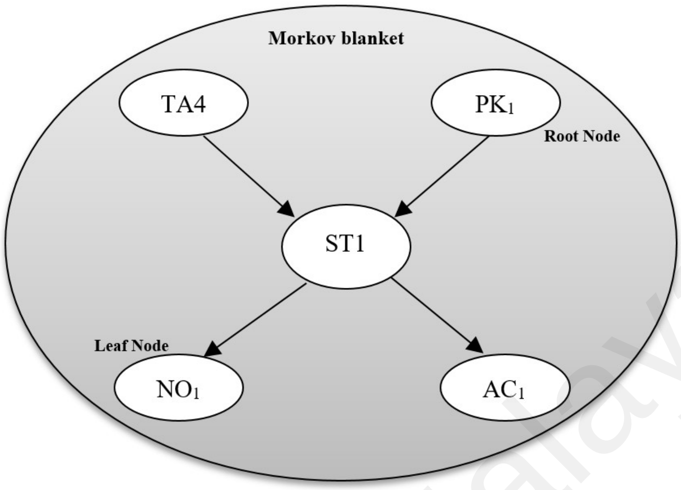


Figure 5.18: Markov blanket sample

The Morkov blanket node PK is composed of the node's parents, its children, and its children's parents and all the other nodes sharing a child with PK. Although BNs are considered as the graphs rather than trees but there are terminologies that represent any node without parents as a root node (TA_4 & PK_1), any node without children as a leaf node (NO_1 & AC_1) and a non-leaf and non-root node as intermediate node. Therefore, root nodes (TA_4 & PK_1) represent original causes and the leaf nodes (NO_1 & AC_1) represent final effects, which make the casual BN structure.

In the most constrained – based algorithms such as GS algorithm for building the BN structure, the Markov blanket is computed separately. Also, each neighborhood is a subset of the corresponding Markov blanket and therefore, can be learned independently from the others(Korb & Nicholson, 2003). Figure 5.19 shows the Morkov blankets and neighborhoods of PK1and NO1 nodes learned by the data using GS algorithm.

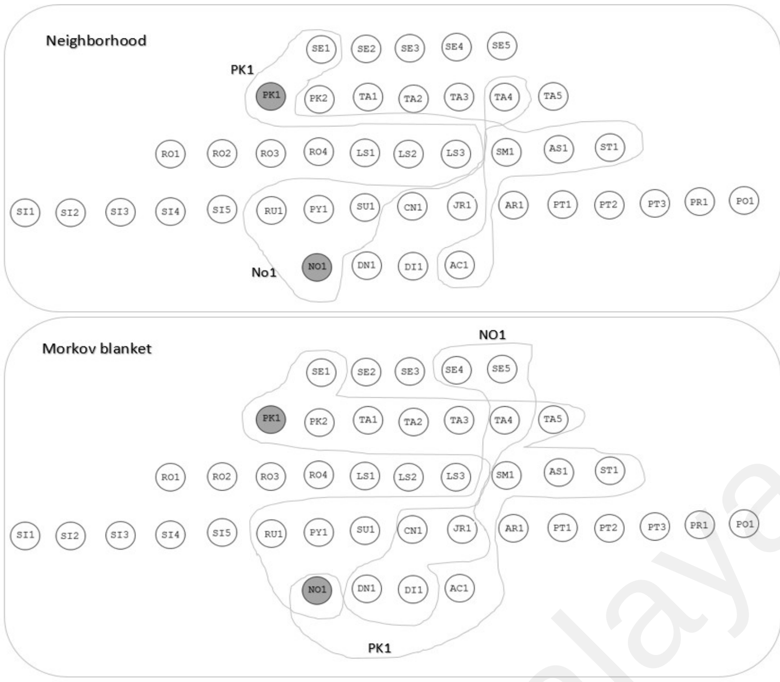


Figure 5.19: Learning Markov blanket & neighbourhood

According to the BN modeling guidelines, before learning the neighborhoods, the consistency of all Markov blankets should be checked to examine their symmetric differences(Albert, 2009) (Gentleman et al. 2010). Therefore, all pairs of nodes were checked and were removed from each other's Markov blanket if they did not appear in both of them. Figure 5.20 shows the automated BN structure derived from data analysis using GS algorithm.

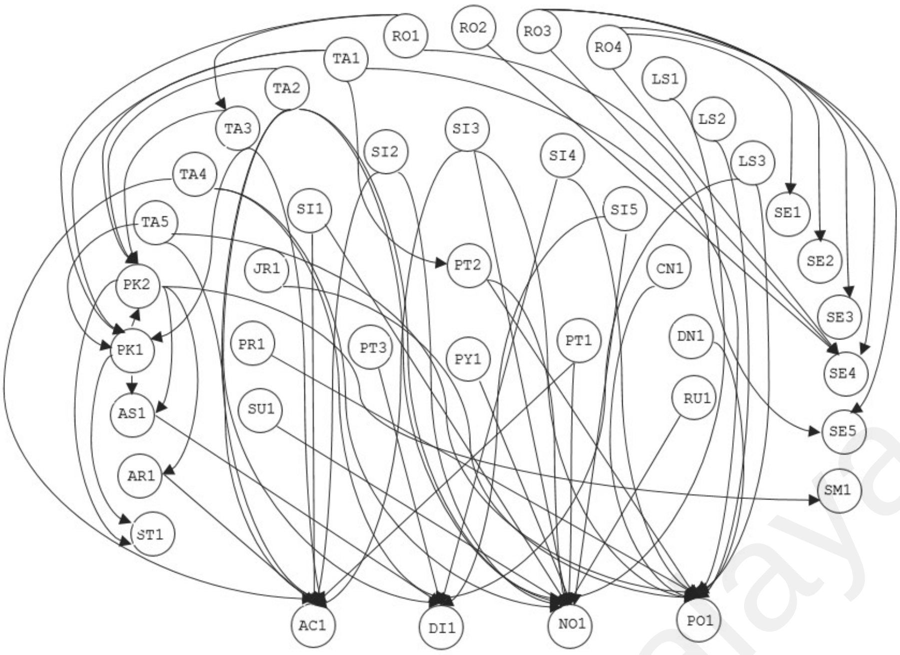


Figure 5.20: BN structure derived from data

After learning the Markov belanket and neighbourhoods, the correlation between the variables was computed to create a correlation matrix using the dataset. Then, the nodes and relationships were compared to the results and consequently, the BN structure was built. As shown in the following code, by applying the Pearson function in R, the Pearson's Correlation was applied to investigate the dependence between multiple variables at the same time. The function rcorr() in Hmisc package also was applied to calculate the significance levels for Pearson correlations which return the correlation coefficients for all possible pairs of variables. It is a value in the range of (-1, +1). The values of -1 and +1 stand respectively a strong negative correlation and strong positive correlation. The value of 0 also represents zero correlation or no correlation.

```
> cor(x, method = "pearson", use = "complete.obs")
> df <- read.csv(bndata())
> data("bndata")
> df <- mtcars[]
> res <- cor(df)
> cor(df,use= "complete.obs")
>rcorr(x, type = c("pearson","spearman"))
>install.packages("Hmisc")
>library("Hmisc") res2 <- rcorr(as.matrix(df)) res2
> head(8)
```

	RO1	RO2	RO3	RO4	PK1	PK2	LS1	LS2
RO1	1.0000	0.0000	0.0000	0.0000	0.0315	0.0201	0.0000	0.0000
RO2	0.0000	1.0000	0.0000	0.0000	0.0613	0.0400	0.0000	0.0000
RO3	0.0000	0.0000	1.0000	0.0000	0.0331	0.0000	0.0000	0.0000
RO4	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
PK1	0.0315	0.0613	0.0331	0.0000	1.0000	0.0000	0.0000	0.0000
PK2	0.0201	0.0400	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
LS1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
LS2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

The colored correlation matrix plot, as depicted in Figure 5.21, shows relatively strong and statistically significant relationships. As mentioned earlier, the values of -1 and +1 stand respectively a strong negative correlation and strong positive correlation between the 42 variables. Highest positive correlations that are shown in navy blue (Figure 5.21) also represented the value near to +1.

In the following, it is discussed how the results taken from the knowledge engineer and automated learning as well as coloration matrix are compared to build the final BN structure.

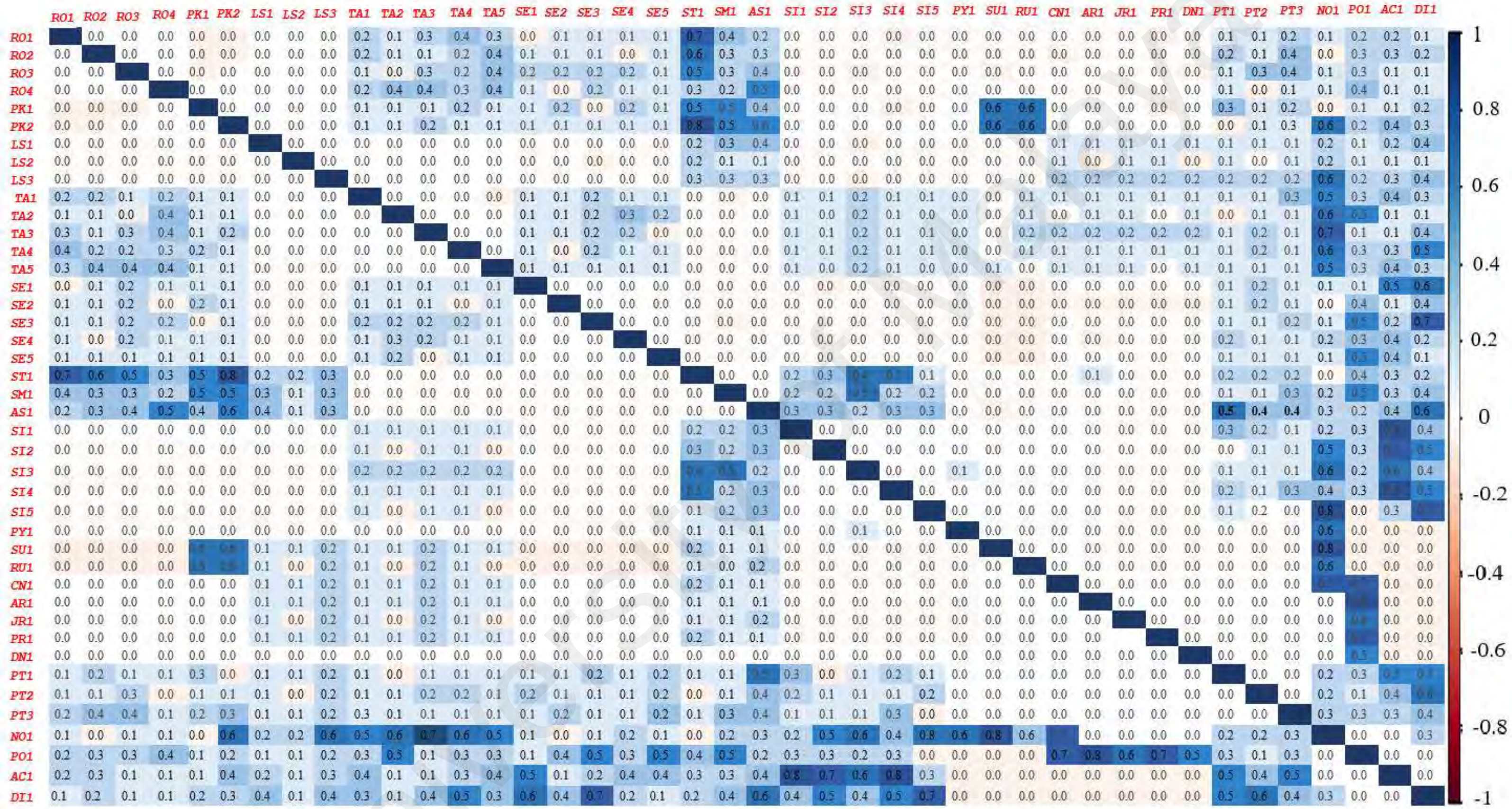


Figure 5.21: Correlation matrix

After comparison of the BN structure derived from the knowledge engineer and automated learning, it was noticed that there are a few nodes that do not have same correlations based on the results of automated BN learning, therefore, by consulting with the knowledge engineer and expert, the final relationships are decided for four paper levels as depicted in Figure 5.22.

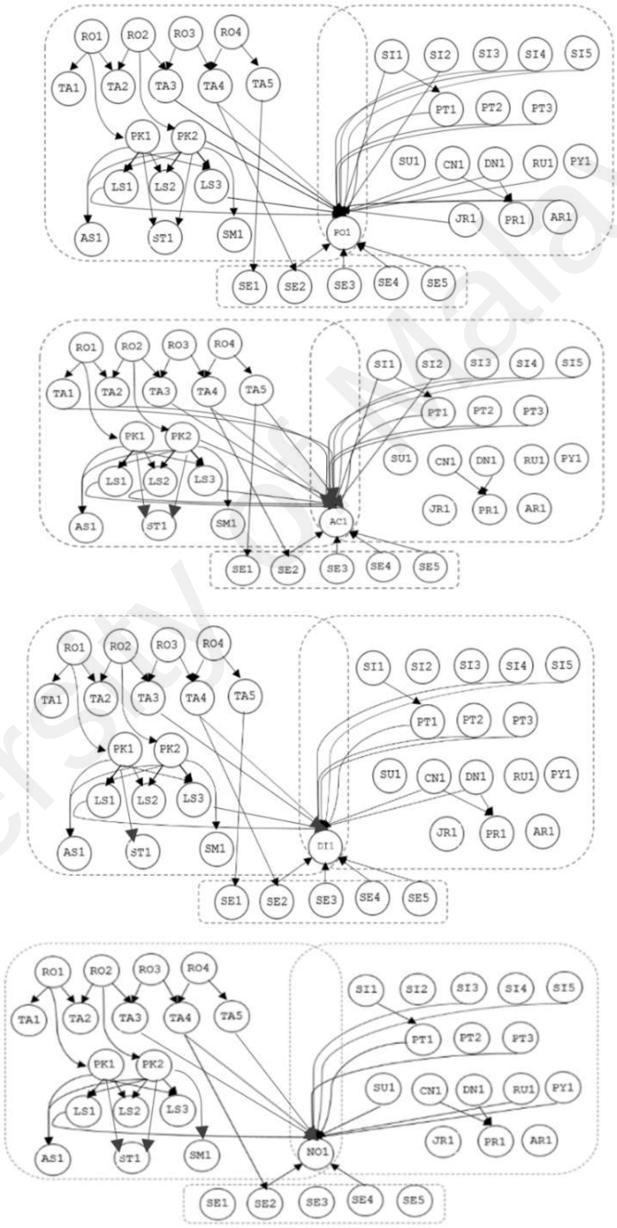


Figure 5.22: BN structure four paper’s levels

5.5.3 BN parameter learning and inference

After BN structure modelling, the next step is the specification of conditional probability distribution or the Conditional Probabilities Tables (CPTs) entailed by a network which is called parameter learning(Korb & Nicholson, 2003) (Guo, 2011).

In the BN modelling literature, Expectation-Maximization (EM) algorithm is normally applied for the numerical parameter learning when there are missing values or biased data (Babas, 2014). However, for the complete and fairly unbiased datasets, the maximum likelihood is recommend for the calculation of CPTs (Margaritis, 2003) (Ono et al., 2007) (Gentleman et al. 2010). Hence, in this research, the parameters were calculated applying the bn.fit function using bnlearn package which utilises the network data to estimate their maximum likelihood(Albert, 2009) (Scutari & Denis, 2014).

For each variable in a Boolean network which has n parents, the size of CPT would be 2^{n+1} probabilities(Albert, 2009). So, in this research, for four nodes of the novel, accurate, diverse, and popular papers, the sizes of the CPTs are respectively 2^{10}, 2^{13}, 2^{12}, and 2^{8} probabilities.

For the brevity, in the next, only the probability distributions of a single node are provided. Figure 5.23 shows part of the BN network along with the joint probability distributions.

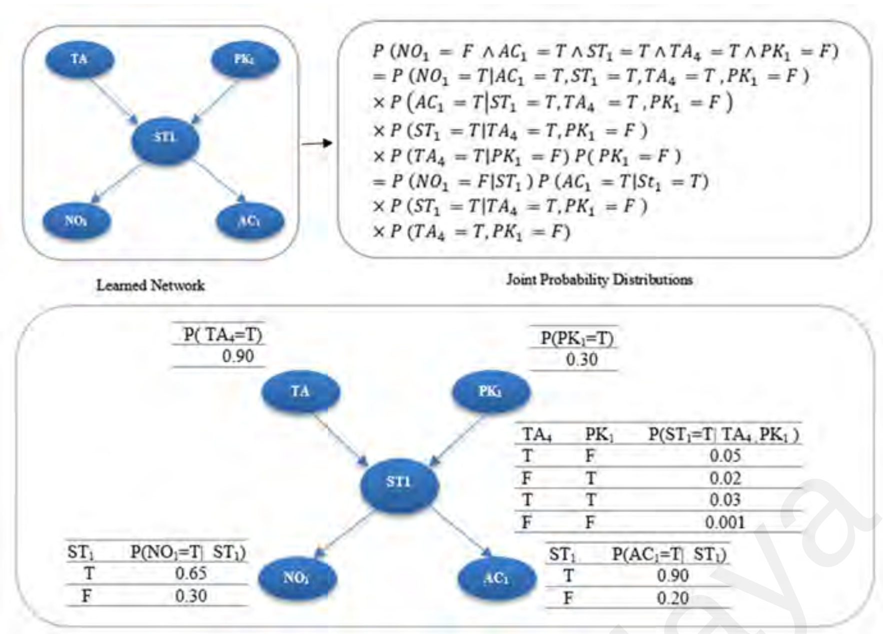


Figure 5.23: BN network along with the distributions

In addition, the following code shows the learning of parameter estimations using bn.fit function for the nodes of LS1, LS2, andLS3 that represent the user's learning style in the BN structure. For the other nodes, the parameter learning is in the same way.

```
> fit = bn.fit (res2, dat)
> fit

  Bayesian network parameter

Parameters of node LS3 (Gaussian distribution)

Conditional density: LS3
Coefficients:
(Intercept)
  0.2560051
Standard deviation of the residuals: 0.4365626

  Parameters of node LS2 (Gaussian distribution)
Conditional density: LS2 | LS3
Coefficients:
(Intercept)     PHYSICAL
  0.5497026   -0.5497026
Standard deviation of the residuals: 0.4294111

  Parameters of node VISUAL (Gaussian distribution)

Conditional density: LS2 | LS1
Coefficients:
(Intercept)       VERBAL
  0.5668449   -0.5668449
Standard deviation of the residuals: 0.3811809
```

The single BN structure along with the possible sates and values were drawn (Figure 5.24) by the Netica tool (Norsys, 1998) which is a powerful, easy-to-use, complete program for working with BNs and diagrams.

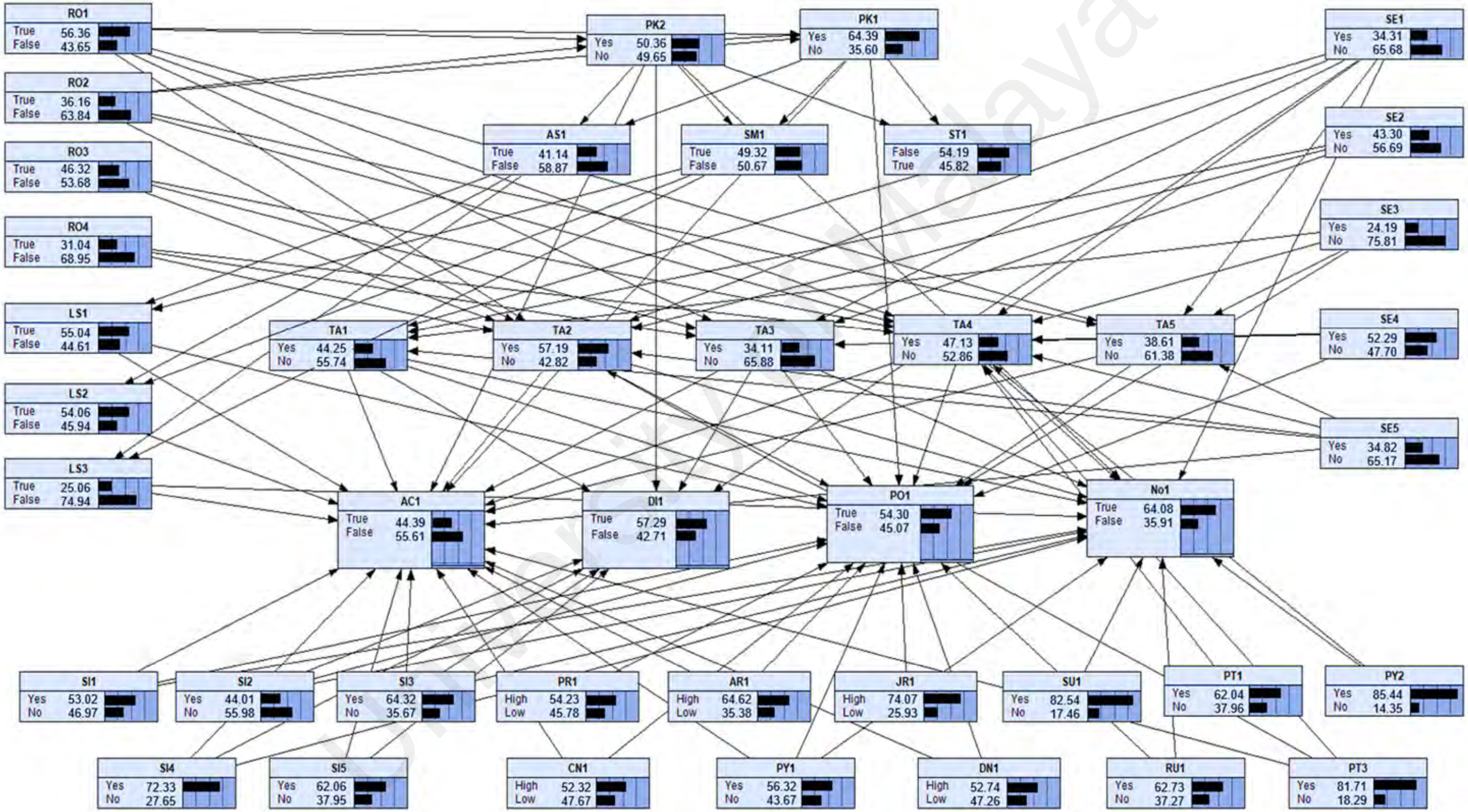


Figure 5.24: BN network and parameters (Netica)

After parameter learning, it is time to investigate the estimated probability distributions of the model(Albert, 2009) (Scutari & Denis, 2014). The BN inference is carried out to estimate the posterior probability distribution or suitable values for the tuning parameters(Korb & Nicholson, 2003) (Margaritis, 2003). There are various methods for performing the inference on a Bayesian model such as bootstrapping, conditional probability queries, and cross validation. Based on the guidelines, cross validation is the most used and appropriate method to validate BN models algorithms and parameters(Albert, 2009). Sensitivity analysis and predictive performance assessment are key elements of the modelling process (Chen and Pollino, 2012;Fienen and Plant, 2015). The validation of BN structure and parameters are discussed in detail in chapter 7.

5.6 Discussion

The more a SRS meets the users' information needs, the better is the SRS(Beel, Breitinger, et al., 2016) (Shirude & Kolhe, 2018). Prior works on RSs argued that users would be more satisfied by a list of the papers that not only are accurate but also are novel and diverse (Kotkov, Wang, et al., 2016) (Kotkov, Veijalainen, et al., 2016). Also, a few studies have emphasized on the diversification of recommendations (Ziegler, McNee, Konstan, & Lausen, 2005). Moreover, today the user might look for novel papers in a particular area while tomorrow she/he might be interested in the most popular papers. Therefore, the users' information needs change due to different knowledge, preferences, goals, and contexts. (Beel, Breitinger, et al., 2016). In addition, the results of regression analysis in Chapter 4 in this research revealed that there are significant relationships between users' contexts and paper quality (diversity, popularity, novelty, and accuracy of a paper).The proposed Bayesian UM in this research aimed to diagnose the users' information needs. This model could be embedded in the recommending process to generate more appropriate paper recommendations related to their contexts. The development of the BN model was both expert oriented and data-oriented. However, a few contexts such as users' reasoning method and the mood have been left out for the future researches because they had small effect on the UX of SRSs based on the examinations derived from chapter 4. The identification of users' information needs is not an easy task and needs better understanding about the users' perceptions via long term studies and observations.

5.7 Summary

This chapter proposed a contextual Bayesian UM for supporting the diagnosis of scholars' information needs in terms of four levels of diverse, novel, accurate, and popular papers exploiting the contexts identified in objective 1. Such diagnoses are important because of high prevalent emphasizes on providing paper recommendations which are relevant to the users' context and appropriate to their information needs. The proposed decision UM in this research applied a BN probabilistic approach where the contexts and four paper levels are all connected together to build a realistic graphical model. The model could be used to build more intelligent SRSs, which diagnose scholars' information needs and provide better experiences for them. The UM development involved three major phases of dataset preparation, structural network learning, and parameter and inference learning. The BN structure modeling was built by the domain expert elicitation method and automated learning by the data leading to a more robust and reliable UM. This UM is able to deal with partial observations and uncertainty, which make the model suitable for SRSs where the scholars' information needs are changing consistently. Additionally, a brief description on the BN modelling and essential definitions as well as the reasons for selection of BNs method for this research were discussed at the beginning of this chapter. The UM evaluation is explained in chapter 7.

CHAPTER 6: USER INTERFACE DESIGN

In this chapter, the proposed User Interface (UI) called rScholar is discussed and presented. The rScholar design is mostly based on the empirical results of the most influencing elements of UiD and IxD adequacy identified in objective 1 and the inputs and outputs for the proposed Bayesian UM developed in objective 2. To embed Bayesian UM into SRSs, it is crucial to gain access to a stream of user actions data. The UI is a critical tool to support gaining the required data. First, the importance of UI and IxD in RSs and SRSs is briefly discussed. Meanwhile, the existing UI guidelines have also been reviewed and those which can be potentially applicable for enhancement of SRSs have been utilized. To design the proposed UI, five steps are performed; the first four steps are discussed in this chapter and the details of the last step, which is the evaluation, is explained in chapter 7.

6.1 The importance of UI and IxD in the RSs and SRSs

As already mentioned in Chapter 2, RSs development can be divided into two parts: the back end that decides what to recommend, and the front end that delivers the recommendation (Murphy-Hill & Murphy, 2014). Both industry practitioners and academic researchers argue that the interface of a RS may have profound effects on users' experience with the recommender than the recommender's algorithmic performance (McNee et al., 2006; Baudisch & Terveen, 1999; Murray & Haubl, 2008; Xiao & Benbasat, 2007; Ziegler et al., 2005; Ozok et al., 2010) (T. T. Nguyen et al., 2013). According to the RecSys09 keynote presented by Francisco Martin, up to 50% of the value of recommenders comes from a well-designed interface (Ge et al., 2010b). This hypothesis is also empirically supported by the results of the framework presented in Chapter 4 of this study which indicates that the adequacy of UI and IxD can have a critical as well as decisive effect on users' perceptions and consequently on the experience of

SRSs. In other words, no matter how accurate the algorithms might work, if the UiD and IxD are poorly designed and evaluated, it can degradethe interaction between the users and system in a way that users might find the system intrusive, annoying or distracting, and they might perceive it as a factor that negatively affects their experience (Ozok et al., 2010) (Pu and Chen 2007; Pu et al. 2012). As an example: if the recommending algorithm performs very well and retrieves an appropriate set of novel and accurate recommendations for a target user but the delivery of recommendation set is not well enough; hence, the user might not able to touch upon the usefulness of recommendations(Ge et al., 2010b). In this case, the users might be confused or misunderstood. That is why such certain recommendation set is provided for them because they might not able to find any link between their real needs and recommendations. This misleading perception might make users stop using the system forever (Callahan and Koenemann, 2000) (Ozok et al., 2010). Exploiting the UI and IxD features help users visually understand the logic behind the recommendations (algorithm functionality) and consequently, help them perceive the usefulness of these recommendations. In addition, selecting the interface and items which are carefully presented to the users for rating can have significant effects on dealing with the “new user problem” in RSs (Sean M McNee, Riedl, et al., 2006a).

The little attention given to UiD and IxD of RSs derives from the fact that they have been mostly implemented as a part of other systems (Abdrabo & Wörndl, 2016). SRSs are mostly part of digital libraries, reference management tools, and bibliographic databases such as Science Direct recommender, Mendeley and Docear. Despite the given fact, RSs are usually considered as one of the critical components of e-commerce websites such as Amazon, e-Bay (Calero Valdez et al., 2016). None of the researches surveyed so far has considered the impact of the distribution of UI of SRSs on the user's experience. This is where this study provides one of its main contributions. In this research, the

Bayesian UM is embedded into recommending process in order to predict the appropriate recommendations in four levels of accuracy, novelty, popularity and diversity; hence, it is essential to gain access to a stream of user actions and preferences to retrieve the tailored recommendations for the users. Unfortunately, as discussed in Chapter 2, SRSs have not been developed with an eye on user modelling. Thus, a censorious problem in developing probabilistic enhancements is to design effective interface and interaction which establish a link between user actions and system events and synchronise the interchange information in a way that users experience positive feelings when they interact with the system (McNee et al., 2006; Murray & Haubl, 2008; Xiao & Benbasat, 2007).

6.2 The UI development process

Figure 6.1 depicts the steps accomplished for designing the rScholar. The aim is to apply the UiD and IxD features identified in objective 1 and to make a link between those features and the inputs and outputs required for the proposed Bayesian UM developed in objective 2. To achieve this, five steps including user search, information architecture, interaction design, interface design and evaluation were carried out. The UI development process was mostly inspired from the design principle introduced by Staffer (2010) which distinguishes between the IxD and UiD and emphasizes IxD adequacy by improving the interactions between or among humans and the UI affordances. According to Staffer (2010), the UiD is the visible part (physical expression) of IxD. IxD is that part of a product or service which is usually invisible. UiD usually controls the manipulation of the features and functionality that makes up the IxD. The first four steps are discussed in this chapter and the details of step 5 are explained in chapter 7. In the following section, each step along with the activities is discussed separately.

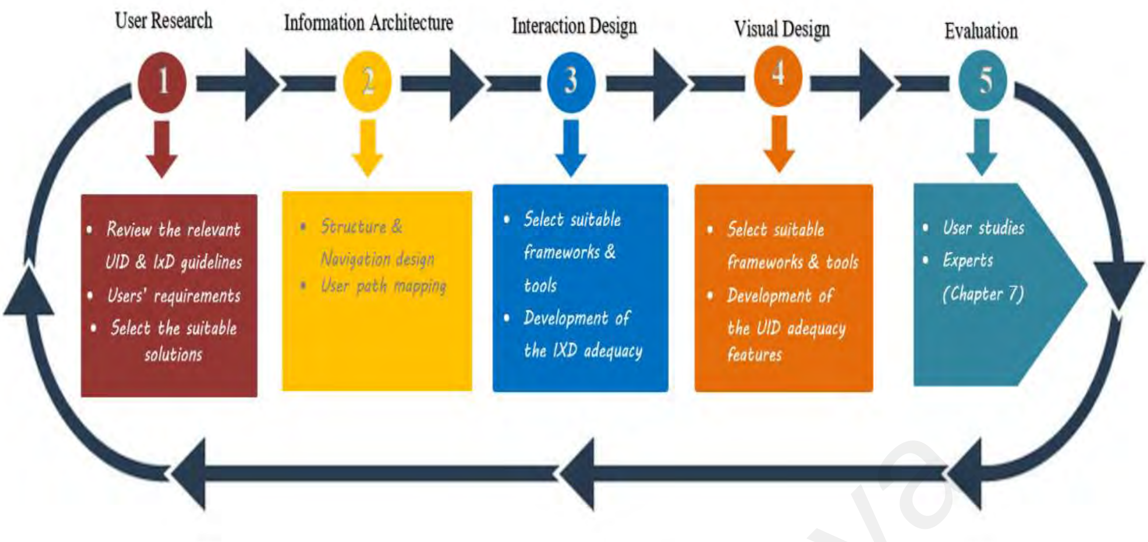


Figure 6.1: UI development steps

6.2.1 User research

User research phase investigates the user requirements for designing an effective interface for SRSs regarding the results of objective 1 and 2 of this research. Therefore, first the suitable and general UiD and IxD guidelines are reviewed. Then, the user requirements are identified and finally, the UiD and IxD features are selected and discussed in the upcoming sections.

6.2.1.1 Review of the general UiD and IxD guidelines

Table 6.1 describes the most important rules of UiD and IxD which are general as proposed and highlighted by a number of studies such as (Galitz, 1985) (Nielsen, 1999; Shneiderman, 2010).

Table 6.1: Design rules & guidelines (Sarif, 2011)

	Design Rules	***Descriptions***
1	Maintain consistency	Identical terminology, menus and help screens; and consistent commands
2	Reduce information load	Information should appear in a natural and logical order
3	Create an aesthetic & minimalist interface design	The interface/display should be pleasing and in many cases the layout should be balanced and proportional.
4	Provide informative feedback	Appropriate, human-readable feedback within a reasonable amount of time
5	Design interactions to create closure	Sequences of actions, which involve beginning, middle and end
6	Provide error recovery	A simple, comprehensible mechanisms for handling the error
7	Support internal control of users	Undo previous action, which provide sense of relief and also encourages exploration of unfamiliar control
8	Provide "shortcuts" for frequent users	Abbreviations, special keys, hidden commands, and automation facilities Shorter response times and faster display rates are also important for frequent users
9	Provide "help" capabilities	This feature helps users learn and use the system

While it is recommended that general webpage designing guidelines can be applied to RSs, it should be noted that the RS is a component of the page that is heavily related to the item (paper in this research) that is delivered to the user rather than anything else on the Web page (Svensson et al., 2005). Therefore, among the general guidelines those which support a pleasant experience of paper delivery to the scholars have been considered. It is noted that there are also some overlaps between the general and RSs guidelines.

6.2.1.2 Review of the existing UiD and IxD guidelines for RSs

The impacts of recommender UIs on user behavior have already been discussed in a few studies and have resulted in a few design guidelines and rules, but the applicability of these guidelines on the specific RS component have rarely been tested empirically (Felfernig et al., 2012) (Bart P. Knijnenburg et al., 2012). Ozok et al (Ozok et al., 2010) and Pu and Chen (2011) provide an in-depth analysis of common design pitfalls when developing UIs for preference elicitation, preference revision, page layout and explanation. Besides, Felfernig et al. (2006) analyze the impact of different recommender UI functionalities such as explanations, product comparison pages, and repair actions on factors such as perceived increase of domain knowledge, increase of usability and trust. In another study, Felfernig et al. (Ozok et al., 2010) also analyze different preference elicitation interfaces through three user studies. The results are four design guidelines for preference elicitation interfaces: (1) users willing to spend more effort in preference elicitation should be able to do so, (2) affective feedback interfaces (e.g., in terms of so-called affect buttons) should be considered as a means of detailed preference feedback, (3) design interfaces should be organized in an explorative fashion where the consequences of preference shifts are easily visible, and (4) preference elicitation processes should not start from scratch but rather rely on initial system preference suggestions. Table 6.2 represents an overview of features emphasized on the design rules and guidelines indicated in several RSs researches.

Table 6.2: Design rules and guidelines for the RSs

Study title	***Design features***												***Reference***
	PE	PR	IS	SC	DN	EX	RL	RD	PC	SI	PL	NA	
Beyond algorithms: An HCI perspective on recommender systems				✓							✓	✓	(Swearingen & Sinha, 2001)
Evaluating recommender systems from the user's perspective: survey of the state of the art	✓	✓	✓		✓	✓	✓	✓			✓		(Pu, Chen, & Hu, 2012a); (Felfernig, Burke, & Pu, 2012)
Design guidelines for effective recommender system interfaces based on a usability criteria conceptual model: results from a college student population			✓	✓	✓	✓	✓	✓			✓	✓	(Ozok et al., 2010)
Advanced user interfaces & hybrid recommendations for exploratory search						✓						✓	(di Sciascio, 2017)
Adaptive user interface and user experience based authoring tool for recommendation systems	✓												(Hussain et al., 2014)
The effect of preference elicitation methods on the user experience of a recommender system				✓		✓			✓		✓		(Knijnenburg & Willemsen, 2010)
Interaction design for recommender systems	✓	✓	✓			✓					✓	✓	(Swearingen & Sinha, 2002)
Interfaces for eliciting new user preferences in recommender systems						✓	✓						(S. McNee, Lam, Konstan, & Riedl, 2003)
Is seeing believing?: how recommender system interfaces affect users' opinions	✓							✓					(Cosley, Lam, Albert, Konstan, & Riedl, 2003)
Recommendation delivery	✓										✓	✓	(Murphy-Hill & Murphy, 2014)
Recommender systems: from algorithms to user experience		✓					✓				✓	✓	(Konstan & Riedl, 2012)
User interface patterns in recommendation-empowered content intensive multimedia applications				✓		✓					✓	✓	(Cremonesi, Elahi, & Garzotto, 2017)

Preference Elicitation= PE; Preference Refinement= PR; Information Sufficiently= IS; Size & Composition of recommendation sets=SC; Dialog & Natural Language= DN; Explanation= EX; Recommendation Label= RL; Recommendation Display= RD; Privacy Consideration= PC; Signifier = SI; Page Layout= PL ; Navigation= NA

6.2.1.3 Defining the user requirements

The output of the Bayesian UM is detection of the users' information needs in four levels of accurate, novel, diverse and popular papers. To be able to detect the users' information needs, the Bayesian UM requires the data of users' actions and users' preferences. Based on objectives 1 and 2, the contextual indicators that have the great contribution are Research Interests, Learning Style, Task, and Pre-knowledge already discussed in details in Chapters 4 and 5. Also, the UiD indicators that have a major contribution to the enhancement of SRSs are Display, Consistency, and Signifier and the IxD indicators are Pre-elicitation, Pre- refinement, Privacy, Explanation, and Info-sufficiency. Therefore, the requirements are defined as:

1. An efficient mechanism to obtain user's data (user context)
2. An efficient mechanism to obtain the environment's data (environment context)
3. An efficient mechanism to show the Bayesian user model outputs (to present recommended papers) (system context)
4. An efficient mechanism to meet the UiD adequacy (system context)
5. An efficient mechanism to meet IxD adequacy (system context)

The final goal of the aforementioned requirements is to make users feel a good experience while using rScholar.

6.2.1.4 Selecting the design solution

Considering the abovementioned requirements, the user persona design is applied as a suitable solution to obtain the required information from the scholars. Based on Cooper et al's (2014) point of view, the personas are very effective in helping designers to know more about the users' contexts since they reflect the users' voice and they are a key element in critiques (Cooper, Reimann, Cronin, & Noessel, 2014). The user persona is the representation of a user .It is typically based on user's research and incorporates user's goals, needs, and interests, which are realistic. The user persona page can serve as a personal web page/profile for the scholars; therefore, they do not need to have their own web pages. Using the user persona design not only makes the rScholar serve as a scholarly search engine and a recommender but also as a scholarly personal web page which allows the scholars to share their scholarly profile and upload their CVs. Moreover, in the studies of Pu, Chen, & Hu (2012b), guideline #9 emphasizes on setting specific goals for users to achieve for motivation of users to contribute. The user persona design also helps users identify themselves within scholarly communities and they might contribute more. All elements of a user persona are carefully selected through consultation with two UX/UI designers for their relevance to the design of rScholar as a recommending system and for meeting the defined requirements. It is noteworthy to mention here that the researcher of this study has also almost four years job experience in the field of UX/UI design.

Table 6.3 represents the final selected design elements applied in order to meet the abovementioned requirements. Approximately, eighteen design elements have been exploited and codded as shown in the Table 6.3.

Table 6.3: Design element selection

Requirements/ Features		Influential in Perception										Design Solution: User Persona	Selected Design Element(s)																	
		AF	FU	CB	CL	TR	IN	PR	US	AV	DO		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
UI & Ix adequacy Features	PE	✓	✓	✓	✓						✓		✓	✓			✓		✓	✓		✓	✓	✓	✓		✓			
	PR	✓	✓	✓	✓						✓		✓	✓			✓		✓	✓		✓	✓	✓	✓		✓			
	IS			✓	✓						✓			✓			✓	✓	✓			✓	✓		✓		✓		✓	
	EX/ RL			✓	✓	✓			✓				✓	✓	✓	✓						✓	✓		✓		✓		✓	
	PC	✓						✓			✓		✓	✓				✓							✓	✓				
	RD	✓	✓					✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	SC			✓	✓				✓	✓									✓	✓										
	SI	✓												✓	✓	✓	✓	✓	✓					✓	✓	✓	✓	✓	✓	✓
	CO	✓		✓	✓		✓		✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obtain users / environment data	PR						✓	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
	TA						✓	✓	✓									✓												
	PK						✓	✓	✓									✓												
	LS						✓	✓	✓									✓												
	IS						✓	✓	✓				✓	✓							✓	✓	✓						✓	✓
	TI						✓	✓	✓									✓												
Bayesian Model outputs	NO						✓	✓	✓				✓	✓			✓		✓	✓		✓	✓	✓	✓		✓	✓	✓	
	DI						✓	✓	✓				✓	✓			✓		✓	✓		✓	✓	✓	✓		✓	✓	✓	
	PO						✓	✓	✓				✓	✓			✓		✓	✓		✓	✓	✓	✓		✓	✓	✓	
	AC						✓	✓	✓				✓	✓			✓		✓	✓		✓	✓	✓	✓		✓	✓	✓	

Features: Preference Elicitation= PE; Preference Refinement= PR; Information Sufficiently= IS; Explanation= EX; Recommendation Label= RL; Privacy Consideration= PC; Size & Composition of recommendation sets=SC; Recommendation Display= RD; Signifier = SI; Consistency= CO
User/ Environment contexts: Profile = PR; Task= TA; Pre-knowledge= PK; Learning style= LS; Info-seeking behavior= IS; Time= TI;
Paper context: Novelty= NO; Diversity= DI; Popularity= PO; Accuracy= AC
Perception: Affordance =AF; Fun= FU; Cognitive barrier= CB; Cognitive Load= CL; Transparency=TR; Interactivity=IN; Personalization=PR; Usefulness=US; Visual authentic=VA Dominance= DO
Design Elements: Sliders =1; Icons =2; Dropdown lists =3; List boxes = 4; Buttons = 5; Dropdown Button= 6; Toggles (Switches)= 7; Text fields= 8; Search Field= 9; Pagination= 10; Tags= 11; Notifications= 12; Tool Tips = 13; Message Boxes =14; Modal Window (pop-up) =15; Accordion =16; Menu Bar= 17; Tabs= 18

6.2.2 rScholar architecture

In this section, the rScholar architecture is explained which is a Multilayer (n-tier) architecture. Multilayer architecture is a common and successful software architecture in software engineering(Richards, 2015). In particular, it is a very suitable architecture for web applications and is applied for development of rScholar in this research.

As Figure 6.2 depicts, layering in layered architecture pattern is some sort of logical grouping of subsystems according to their functionality. Layering divides the process of software development into some logical sections. Each layer of the layered architecture pattern functions independently and has a specific role and responsibility within the application. Therefore, if a technology used in one of the layers needs some changes, it would not require any changes in the other layer since the relation between the layers is limited to the service they give to each other. The separation of concerns among components is one of the distinctive features of this architecture which means that components within a specific layer deal only with logic that pertains to that layer (Richards, 2015). For instance, components in the reasoning layer have limited relation to other components. As shown in Figure 6.2, three layers of data acquisition, reasoning and interface have been developed.

In the following sections, it is discussed which tools have been selected to develop both interaction and visual design.

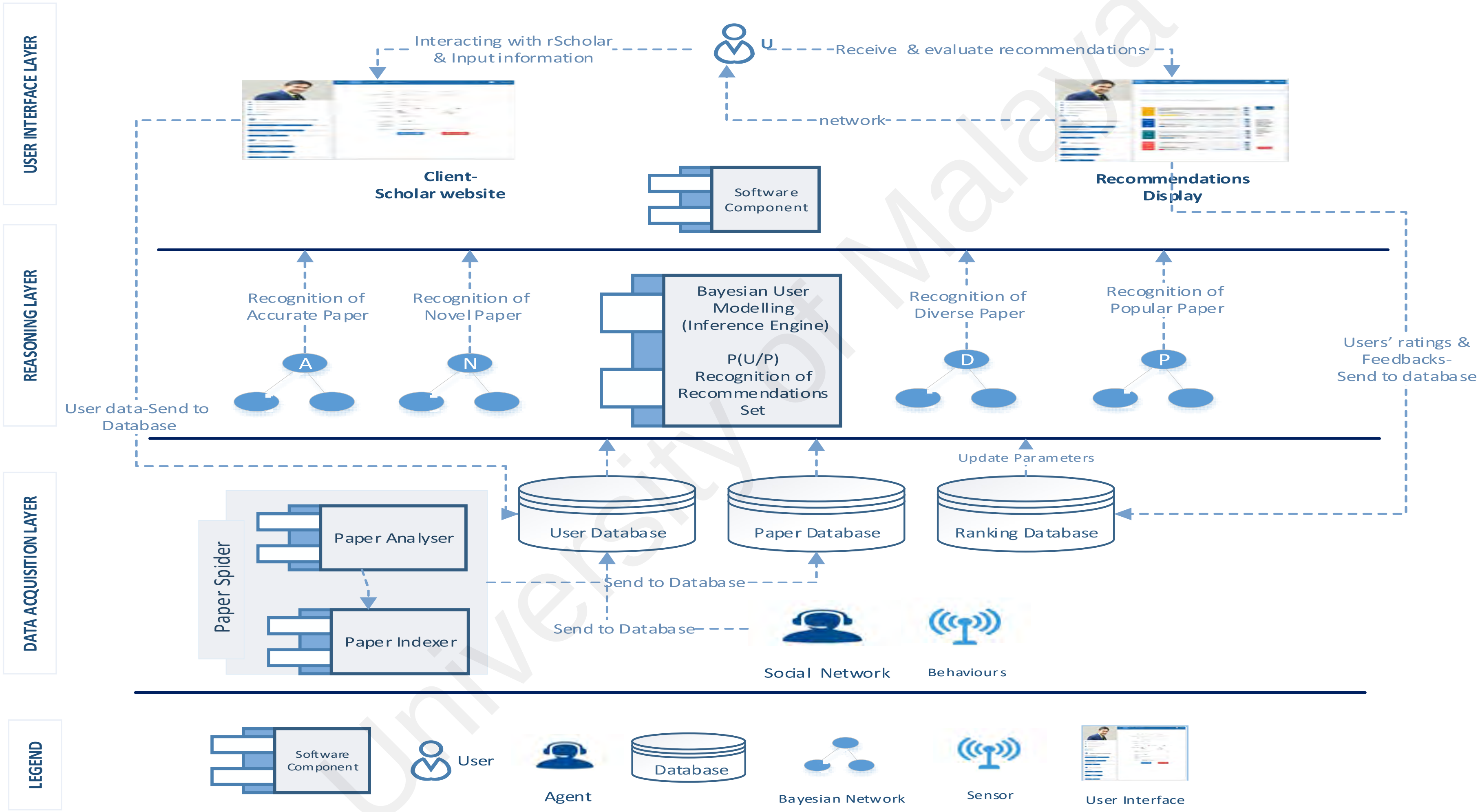


Figure 6.2: rScholar Architecture

6.2.2.1 Selection of design framework/tools

This sub-section presents the technical frameworks and tools used for the development of the UI. As Table 6.4 shows, Bootstrap - CSS, JS, and HTML frameworks are applied for building responsive web pages. In addition, the Illustrator, a vector graphics editor, is used for the primary design of web pages along with logos. ASP.net, a fully supported and free web application framework for building standards connected web solutions, is also applied to develop the UI. LINQ to XML is a LINQ-enabled, in-memory XML programming interface that enables to work with XML from within the .NET Framework programming languages.

Table 6.4: UI development framework & and tools

Part	*Framework & Tools*
Front-end Frameworks	CSS3
	JavaScript (JS)
	HTML5
	Bootstrap
	ASP.net
Database	LINQ/XML
Design tool	Illustrator

6.2.3 rScholar interaction & visual design development

In this section, it is described how the interaction and visual design are designed to meet the defined requirements in the previous section. In the following sub-sections, the responses to the SRQ4 and SRQ5 are explained.

6.2.3.1 Obtaining the required contextual data

Figure 6.3 shows different screens of the rScholar. In each screen, it is aimed to follow the guidelines and rules. The alphabetic letter in each circle corresponds to one of the UI screens. The arrows between screens show that the users can navigate from the current screen represented by the tail of an arrow to the next screen which is represented by the head of an arrow. The description of each screen is provided in separate sub-section respectively..

Figure 6.3: rScholar screens

6.2.3.2 Screen A – Login page

As shown in Figure 6.4, the login page is the first screen visited when rScholar is run. This screen represents the procedure used to get access to rScholar (Oval 1). The user requires to enter the username and password created before by using Register link (Oval 2). After the user's authentication, the user has access to screen C- Home page which is unique and personalised for each user.

The background screen of login and register pages is a board with some notations taken from different academic majors such as mathematics, music, literature, computer and etc. which recalls the academic environment as the end users of rScholar are academic people.

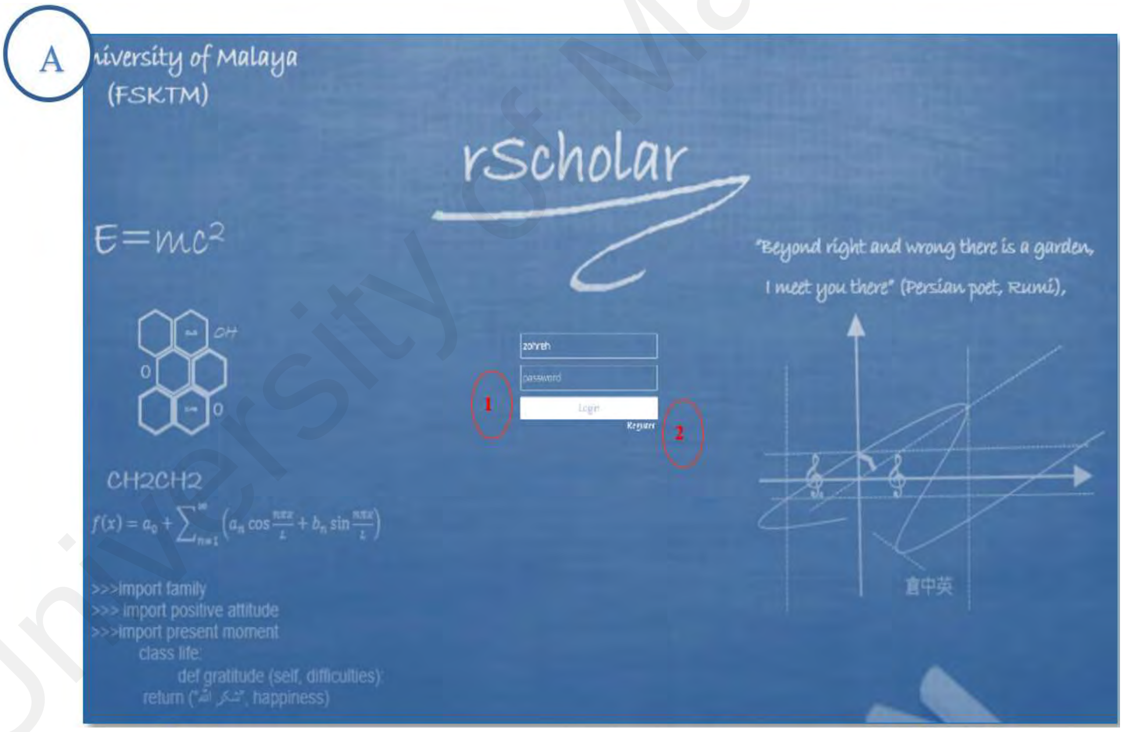


Figure 6.4: Login page

6.2.3.3 Screen B – Register page

If the user is not yet registered s/he needs to register first by entering email address as the username and a combination of letters and/or numbers as the password. Once the user is registered, s/ he receives a confirmation email. Figure 6.5 shows the data which is needed for the registration (Oval 3) including username, email, password and confirmed password. The registration and receiving the services in the rScholar are all free.

Figure 6.5: Register page

6.2.3.4 Screen C- Home page

Once the user logged in, Screen C- Home page (Figure 6.6) which has been designed based on the user persona design solution and has been personalised for each user is presented. In this page:

1. User is able to submit his academic information such as name, major, contact number etc. as well as social media accounts and upload his/her photo (Oval 1:Figure 6.6)

2. User is able to submit his/her research interests and pre-knowledge. The maximum allowed inputs are ten per user. The sliders can be applied to rate the degree of his/her knowledge from a range of 0 to 100 % (Oval 2: Figure 6.6).

3. User is able to input persona information such as learning style, task and current academic semester (Oval 3: Figure 6.6).

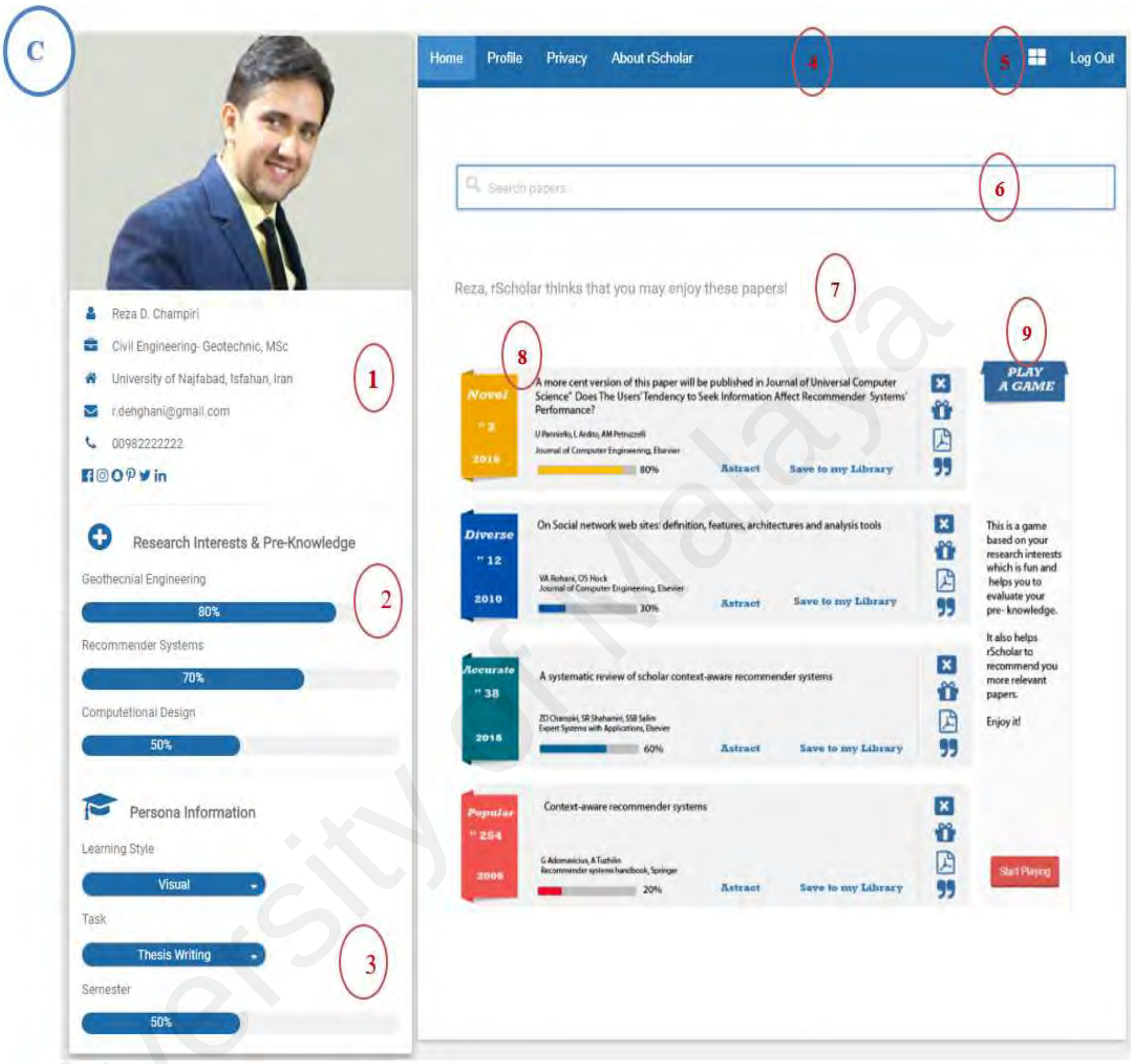


Figure 6.6: Home page- user persona

Figure 6.7 shows the dropdown lists, which represent available options for learning style, task and academic semester. As mentioned before, these pieces of information are needed for the Bayesian UM, therefore, the Bayesian UM exploits these pieces of information and once the user updates his/her information, the Bayesian UM is also updated.

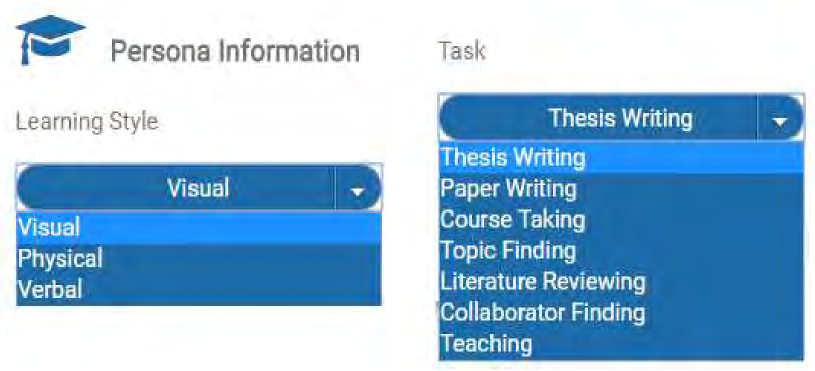


Figure 6.7: Available options for learning style & task (dropdown lists)

4 User is able to navigate through different pages via the menu bar including homepage, profile, privacy setting and about rScholar respectively on screens D, E, F and G. These pages are discussed separately later. In the menu bar, there is also an option to log out from the system (Oval 4: Figure 6.6).

5. There is another option in the menu bar, which allows the user to change the page layout (Oval 5: Figure 6.6). Two different page layouts have been designed including list and pie layouts as shown in Figure 6.8. By using this option, the user is able to personalise the page layouts based on his/her preferences.

6. User is able to search in the datasets (Oval 6: Figure 6.6). Screen F in Figure 6.9 depicts how rScholar presents the search results. User is also able to view the abstract of the paper and save the papers to his/her library. This library is a self-created library by the user.

In the future, it can also be connected to reference management tools. However, this option has been excluded from the present research. There are a few icons as tooltips represent, the user can remove the paper which is not relevant, send the paper as a gift to someone else, see the full text and cite the paper (Oval 2-screen F: Figure 6.9). These options can also be considered as the options for the preference elicitation and refinement, which are discussed in the next section.

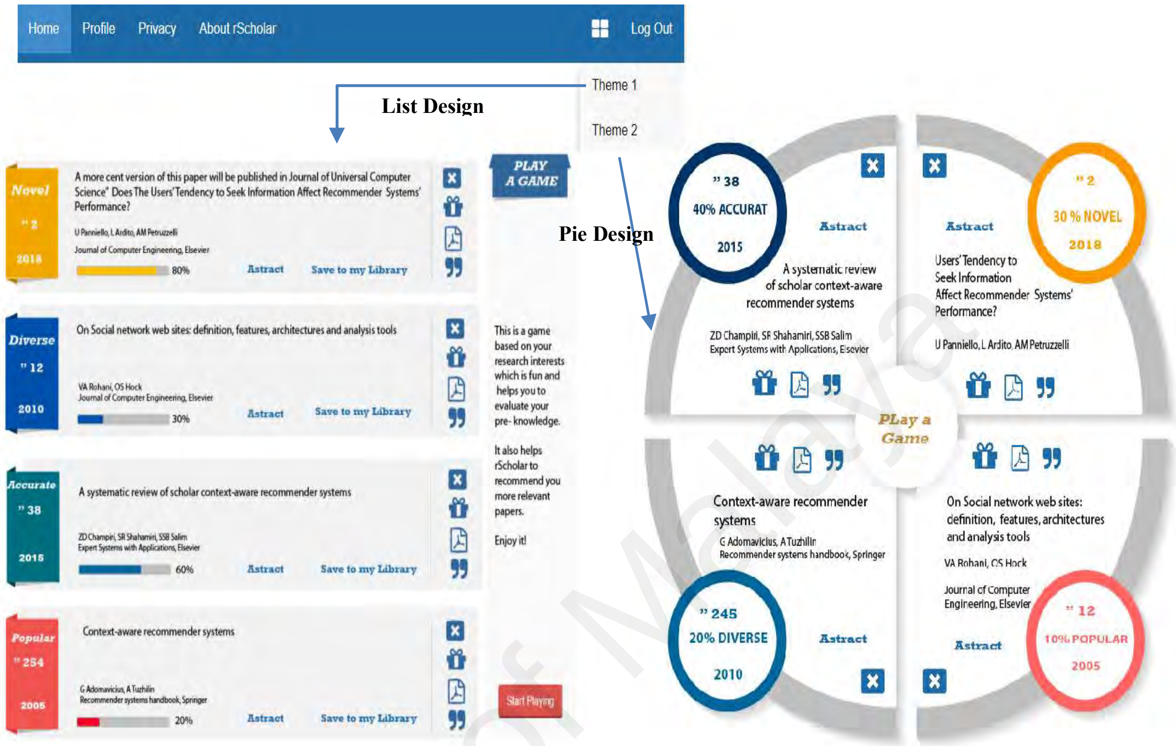


Figure 6.8: Page layout options

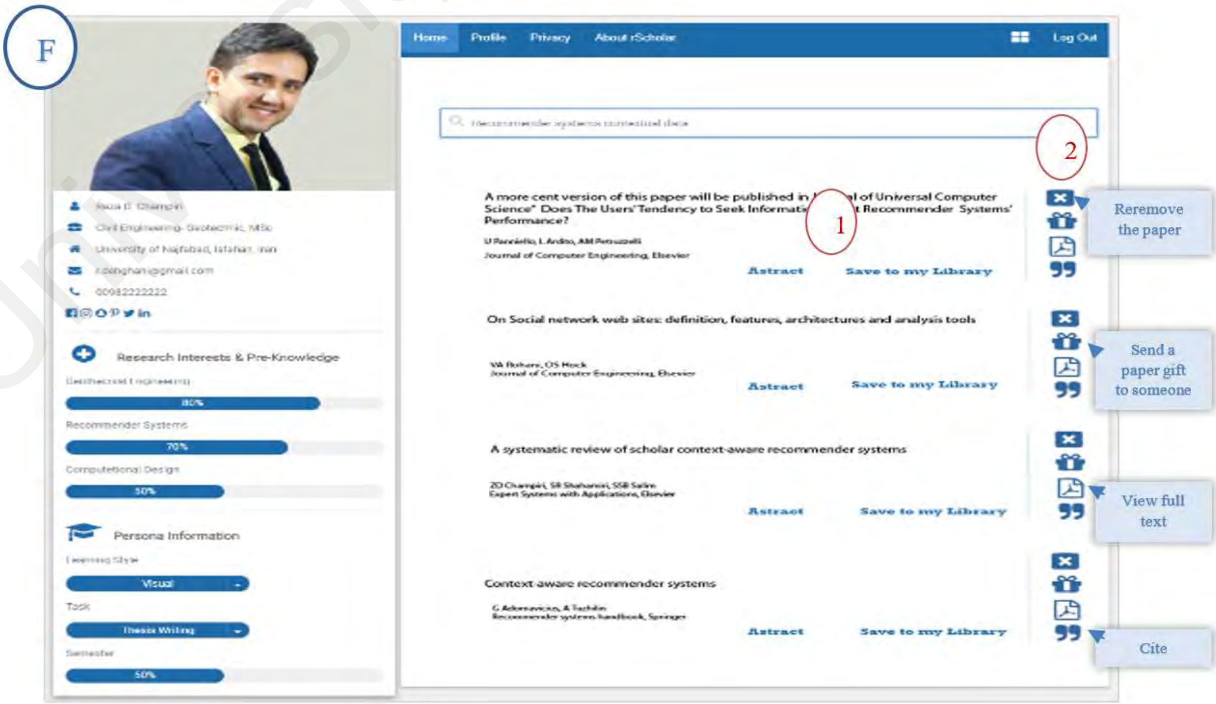


Figure 6.9: Search results

Oval 7, 8 and 9 (Figure 6.6) show the recommendation label, recommendation presentation and gamification respectively which are discussed in the next section in detail.

6.2.3.5 Screen D- Profile

In profile menu, the user is able to input information about him/herself and edit the data. The user can manage his/her recommendation delivery time interval by choosing three options of daily, weekly and monthly (Oval1- Screen D). In addition, the user is able to set if s/he is willing to get mail notifications or not (Oval 2- Screen D). If the user is a faculty member and has a few students to supervise, s/he might input the students' research interests or topics so that s/he may receive some recommendations related to his/her students' research interests (Oval 3- Screen D). The user's CV can also be uploaded by using the last option in the profile page (Oval 4- Screen D). As mentioned earlier, one of the advantages of the proposed UI is that the user persona page can serve as a personal web page/profile for the scholars.

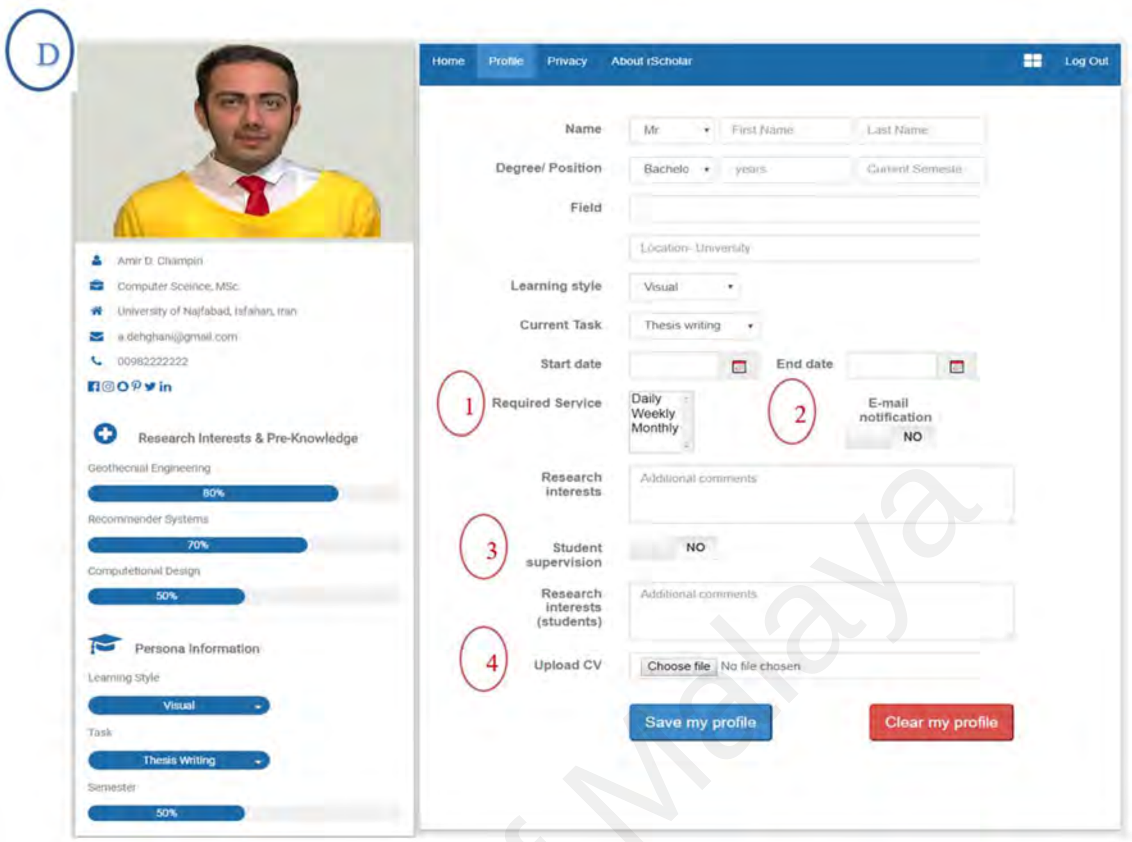


Figure 6.10: User profile

All the fields in homepage are editable through two ways on the homepage and profile menu as shown in Screen D. Guideline #9 in Pu et al.'s study, (2012b) emphasizes on setting specific goals for users to achieve as users' motivation to contribute. The user persona selection also helps users identify themselves within scholarly communities, thus, they might want to contribute more. Besides, Beenen et al. (2004) also discovered that most users are social loafers rather than contributors and they work harder when their effort is important to the group's performance and when their contributions to the group can be identified. As mentioned before, the aim of the proposed UI is to develop an efficient mechanism to obtain user's data (user context), the environment's data (environment context) and also an efficient mechanism to show the Bayesian UM outputs (to present recommended papers) (system context). In the following sections, it is discussed how an efficient mechanism is applied to meet the UiD and IxD adequacy for the reposed UI.

6.2.4 Meeting the UiD adequacy requirement

Based on the results of objective 1; display, consistency and, signifier are major contributors to the recommendation perceptions. In online RSs, interface issues such as page layout and navigation are the most important factors relating to the overall ease of use and perceived usefulness of RSs (Swearingen and Sinha, 2001) (Ozok et al., 2010). Gamification and visualization have not been designed seriously in this research since they require other studies in terms of design and data analysis. However, they are briefly discussed in order to give insights for the future UI design.

6.2.4.1 Recommendation display

Page layout, color, icons, number of recommendations, recommendation notification, recommendation time, and navigation design patterns all can address the recommendation display. Guideline # 14 in Pu et al.'s study, (2012b) gives a special importance to more attractive layout design and effective labels in order to enhance users' perceived accuracy , and explains how the systems compute the recommendations (recommendation explanation: IxD adequacy indicators). Doing so can increase users' perception of the system's effectiveness, their overall satisfaction of the system, their readiness to accept the recommended items, and their trust in the system. In one of the earlier studies examining the HCI aspects of such systems, Swearingen et al. (2001) found that some of the interface issues including graphics and color are not strongly correlated to the ease of use and perceived usefulness of RSs. In the survey conducted by Ozok et al. (2010), participants overwhelmingly (85.5%) preferred to see RSs as part of the regular web page content. Chen (2011) has indicated that most of current RSs follow the list structure, where recommended items are sited one after another. The grid layout, a two-dimensional display with multiple rows and columns, has also been applied in one movie recommender sites to display the items. As the third alternative design, pie layout has been rarely used in RSs. Howevre, it has been proven as an effective menu design for accelerating users' selection process and has significantly enhanced the users' decision confidence, enjoyability, perceived recommender competence, and usage intention via user evaluation(L. Chen & Tsoi, 2011). Therefore, as mentioned before and shown in Figure 6.8, two page layouts, list and pie, have been designed in this research. The effectiveness evaluation of the designed page layouts for SRSs is discussed in Chapter 7.

The number of recommendations, which is also called set composition or recommendation set size, is another concern of RSs researchers. Bart P. Knijnenburg et al., (2012) and Ozok et al.'s (2010) studies indicate that while showing one item is too few, showing more than five items increases users' choice difficulty. Hence, it seems that the adequate recommendation number is three (Ozok et al., 2010). They also imply that different kinds of items should be mixed and balanced to make up the final set (Knijnenburg et al. 2012) (McNee et al. 2006a; Ziegler et al. 2005). Besides, presentation is a crucial factor in persuading users to accept the recommended items; therefore, each RS must carefully employ special strategies that are sensible to users' information needs as well as the business goals of the RS. For making an adequate recommendation set in this research, a set of four paper recommendations consisting of novel, popular, accurate and diverse is offered (Figure 6.8). The user persona solution makes it easier to design a good and smooth navigation structure for the users. As depicted in Figure 6.3, the user is able to navigate easily through the screens.

There are two ways of reactive (manual) in which the recommendations are sent to users when they ask for it and proactive initiation (automatic) in which the recommendations are sent as they are planned at a scheduled time to deliver the recommendation (Murphy-Hill & Murphy, 2014). Several user interface techniques have been proposed to help balance the need for timely recommendations with the need to avoid distracting users which informs the user that a recommendation is available without forcing the user to acknowledge it immediately. It seems that providing both ways and a combination of them is preferable by users. In other perspective, Ho & Tam (2005) indicate that in the early stage of the decision process, the users are more likely to accept and review the recommendations; therefore, exactly at that time, it is better to present some recommendations to them. In rScholar, users receive recommendation by the time they explicitly ask for it at a set time or after searching and interacting with the system. As discussed earlier and shown in Figure 6.10 (oval 1 & 2), rScholar has two methods of recommendation delivery in three time intervals of daily, weekly and monthly including recommendations that are delivered via email notification and those presented in rScholar homepage. The user is able to set his /her preferable method. If the email method is set, the recommendations are sent to the users' mail inbox at a set time as Figure 6.11 shows. The user is also able to view more detailed information about the recommendations. The deactivation of receiving recommendations is provided by the link at the bottom of email notification.

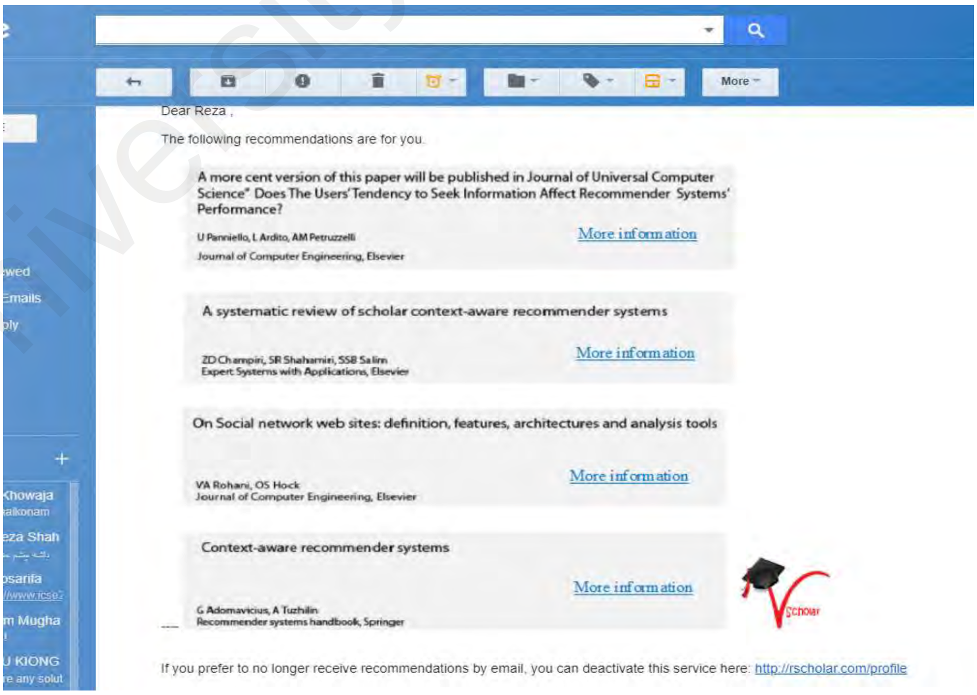


Figure 6.11: Sample of recommendation delivery by email notification

Visualization techniques are used in RSs to support transparency, acceptance and controllability of the recommendation process. As pointed out earlier, visualization has not been designed seriously in this research. PeerChooser and SmallWorlds are good examples of visual interactive recommenders that show relationships between users and recommended items (Murphy-Hill & Murphy, 2014). Recently, a few studies have utilized visualization techniques specially to denote the user's current context such as time, location and weather along (Hiesel et al., 2016) and also multidimensional visualization to promote the diversity of recommended items through an interpretable and interactive interface (C.-H. Tsai, 2017). Like the dashboard on a car, RS dashboards typically integrate recommendations of different types from different sources allowing the user to glance at recommendations frequently and with low commitment (Murphy-Hill & Murphy, 2014). In the design of rScholar, not any specific game and visualization method have been desinged since they have not been the objectives of this research. In addition, adequate design of a game and visualization for SRSs are not easy tasks and require separate studies. However, the game icon is considered in UI design in order to emphasise using the game in scholarly contexts for consideration in the future studies and achieving the users' feedback at the User Studies evaluation which is discussed more in chapter 7.

6.2.5 Meeting the IxD adequacy requirements

In the following, it is discussed how the features influencing the IxD adequacy have been considered in the design of rScholar.

6.2.5.1 Preference elicitation & refinement

The preference elicitation method is the way in which the RS discovers what the user likes and/or dislikes (Chen and Pu 2004; Peintner et al. 2008). In rScholar, there are a few methods for discovering the users' preferences implicitly and explicitly. In the explicit mode, users are asked to rate items or specify preferences on the features of the products such as user profile, item rating (and using filtering methods) and Questionnaire. In the implicit mode, the system makes predictions of the users' preferences by observing users' browsing, searching, selecting, purchasing, and rating behaviors. As a matter of fact, feature-based preference elicitation has been more popularly applied to high-risk and high-involvement products, such as cars, computers, houses, cameras, with which users rarely have experience (so it is difficult to obtain their ratings), while it is feasible to ask them to identify criteria on specific features (Herlocker et al. 2004). Figure 6.12 shows the implicit and explicit preference elicitation applied in the design of rScholar. Based on the interface design guidelines for preference elicitation indicated in Pommeranz et al. (2012), the profile/interest selection serves as an easy (i.e. reduced effort) starting point for showing default preferences that can subsequently be adapted by the users. Game is also an implicit method for getting the users' preferences.

Amazon employs a simple preference refinement method. It asks users to rate some specific items under the box that says "Improve Your Recommendation". This facility may convince users that their work leads to more accurate recommendations and encourage them to put forth more effort. Guideline #11 in Pu et al.'s research (2011) indicates that refinement facilities, such as critiquing helps the system increase recommendation accuracy and the users' sense of control.

Figure 6.12: Implicit and explicit preference elicitation

In rScholar design, the users' feedbacks are used to develop and update the Bayesian UM. Table 6.5 illustrates the preference refinement methods used in rScholar. In other words, users take these actions towards the recommended papers in order to refine their preferences.

Table 6.5: Preference elicitation/refinement methods

Method	*rScholar*
Deleting	✓
Viewing the full-text	✓
Saving to the library	✓
Citing	✓
Sending as a gift to a friend	✓
Changing the rate of paper	✓
Updating the user profile	✓

This preference elicitation likely helps establish a more accurate user model for finding the users' information needs. According to Swearingen and Sinha (2002), this aspect is highly correlated to the users' trust in the system.

6.2.5.2 Recommendation label

The recommendation label identifies the area on the screen where the recommended items are displayed. "Recommendation for you" (by Amazon), "Movies you'll like" (by Netflix), or simply "Suggestion" (by YouTube) are samples of recommendation labels (Murphy-Hill & Murphy, 2014). In rScholar, the statement of "'User's name, rScholar thinks that you may enjoy these papers" is applied as the recommendation label (Figure 6.13).

Figure 6.13: Recommendation labelling

6.2.5.3 Explanation

In addition to labels, recent recommenders employ explanation techniques to help users understand the recommender's logic and thus, augment the transparency of the user interaction (Murphy-Hill & Murphy, 2014) which is one of the IxD indicators that plays a highly important role in the success of the recommenders (Ozok et al., 2010) (Felfernig et al., 2012). To date, many researchers have demonstrated that providing good explanations for recommendations could help inspire users' trust and satisfaction, increase users' involvement and educate users on the internal logic of the system (Herlocker et al. 2000 (Joseph A Konstan & John Riedl, 2012); Sinha and Swearingen 2001, 2002; Simonson 2005; Tintarev and Masthoff 2007a,b). Table 6.6 summarizes the explanation type and methods for recommendations stated in various studies.

Table 6.6: Methods for the recommendation explanation

Type	*Method/Technique*	*References*
Textual	Comments, feedback, chat box, social texture	(Cosley et al., 2003); (Kim et al., 2004); (Svensson et al., 2005)
Graph	Statistical presentations, images, graphs, Star ratings	(Åberg and Shahmehri, 2000) (Konstan, 2012); (Herlocker et al. 2000; Tintarev and Masthoff 2008; Vig et al. 2009)
Cascaded	Displays only the category of information sources	(Pu and Chen 2007)
Labelling	"Customers who bought/ viewed this item also bought/ viewed:"	(Vig et al. 2009)
Tabular	Tables	(Ozok et al., 2010), (Pu et al. 2011)

As shown in Figure 6.8, in rScholar, considering the outputs of Bayesian UM, the papers recommended to the users include four different categories of novel, accurate, popular and, diverse. The recommendations of these four categories are generated by considering the users' situation and preferences, which have been completely explained earlier. However, it should be noted that, in rScholar, the logic of recommendations is based on more than one criteria. In this space, describing the rationale for the recommendation is not easy to simply provide some logic such as the same research interest. Textual descriptions may be an appropriate method in this situation (Murphy-Hill & Murphy, 2014) which is considered for the future enhancement.

6.2.5.4 Information sufficiency

Information sufficiency refers to the content of the recommendation and specification that should be sufficient for users to make confident using decisions while saving time and effort (Ozok et al., 2010). In rScholar, as Figure 6.14 depicted, paper title, paper author(s), abstract, full text (if accessible), citations, journal publication date, and publisher are presented to the users and the detailed information such as abstract are provided by additional link.

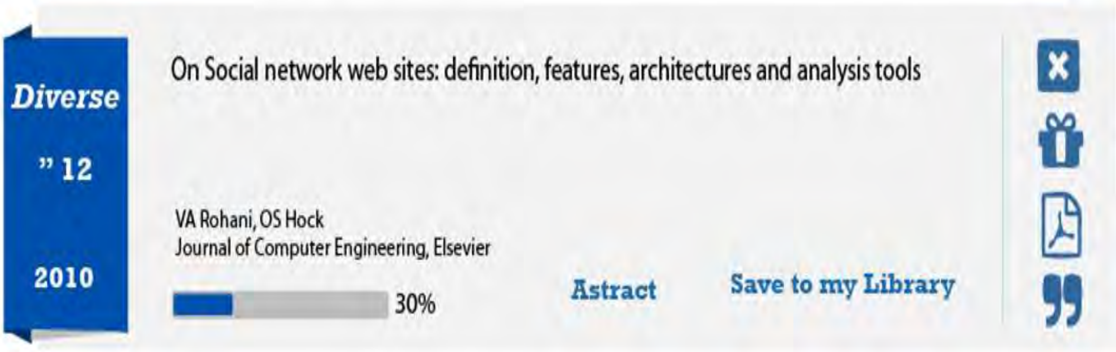


Figure 6.14: Information sufficiency in rScholar

Based on the study of (Ozok et al., 2010), participants mostly want short and concise information.

6.2.5.5 Privacy consideration

Privacy is a critical issue for RSs, regardless of whether the adopted user modeling method is explicit or implicit (Resnick andVarian 1997; Riedl 2001). To know users well enough and make effective recommendations, a recommender must acquire sufficient information (e.g., demographic information, preference information, personality information etc.) about the users. The privacy concern becomes more important when the required information is more personal and the users want to keep it confidential. Finding the optimal balance between privacy protection and personalization remains a challenging task(Knijnenburg & Berkovsky, 2017). There is a general assumption that people are sensitive to privacy issues. However, a mismatch exists between people's privacy preferences and their actual behavior. Spiekermann et al. (2001) compared self-reported privacy preferences of 171 participants with their actual disclosing behavior during an online shopping episode. In their study, most individuals stated that privacy is important to them, with concern centering on the disclosure of different aspects of personal information. However, regardless of their specific privacy concerns, most participants did not adhere to their self-reported privacy preferences. The results suggest that people overlook privacy concerns once they are highly involved in the system (Ramakrishnan, Keller, Mirza, Grama, & Karypis, 2001).

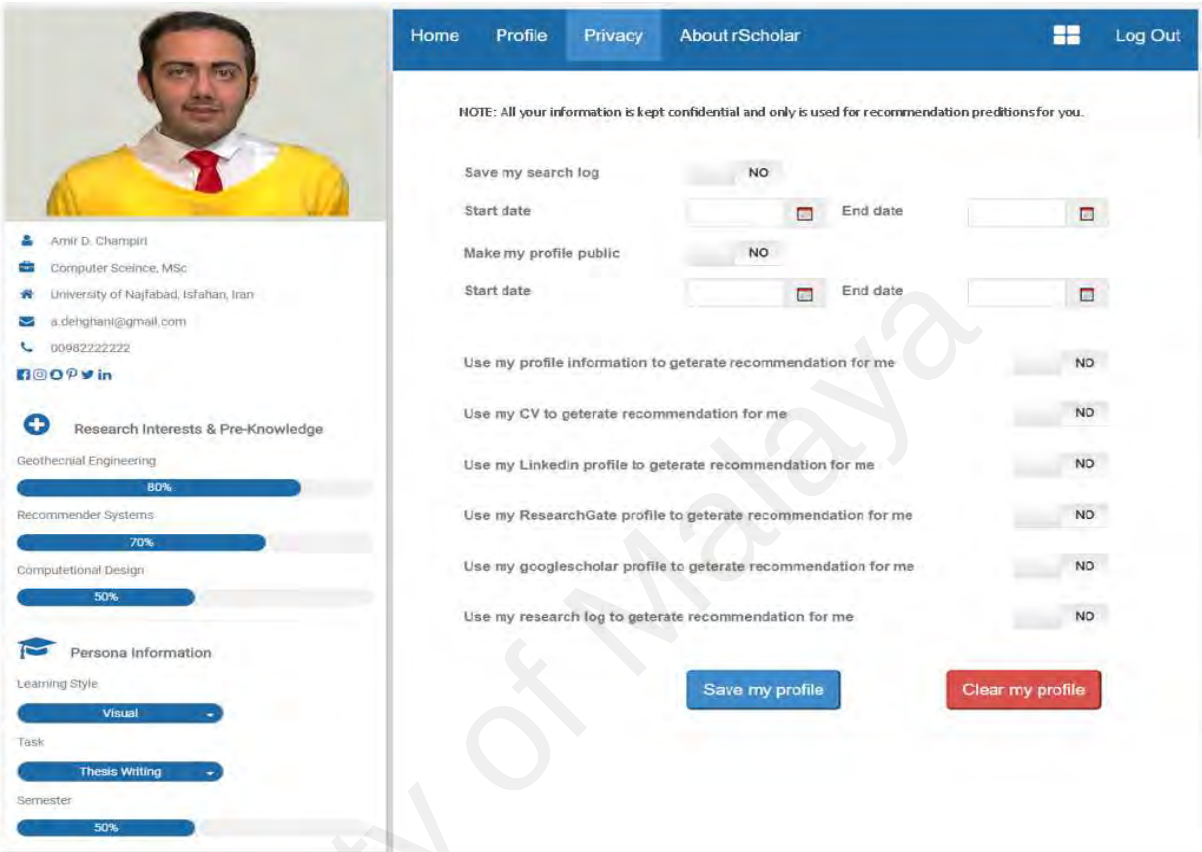


Figure 6.15: Privacy consideration in rScholar

In addition, when users decide whether to provide personal information or not, they would like to know who is able to access the information and its specific purpose (Kobsa and Schreck 2003) (Brodie et al. 2004). Figure 6.15 shows that users are asked to give permission to the rScholar to access their social media information and search logs while apprising them that their information is kept confidential and only is used for making more personalized recommendations for them. Although this method protects users' information appropriately (Lam et al. 2006) and might make users trust the system, it is still a long way to fully deal with the privacy issues in RSs.

6.3 Discussion

As mentioned in chapter 2, there has been given little attention to UI design for RSs and particularly SRSs. rScholar is one of the first serious attempts in the field of UI design for the SRSs considering the UiD and IxD adequacy factors and data required for the proposed Bayesian UM. However, this attempt does not stop here and requires more and more investigations especially to make a trade-off between user contribution and data acquisition (e.g. ratings) for UMs to make more effective recommendations. Making this trade-off is not always easy since first, the users perceive rating-time costs which means they perceive the benefits outweigh the costs they pay (T. Nguyen, 2016). Second, users are often more satisfied when they are given control over how the recommender functions on their behalf, even when that control increases the effort that is required of them, and when the resulting recommendations are objectively less accurate. Therefore, a recommender should provide balance between users' control which they desire and effective recommendation service(Joseph A Konstan & John Riedl, 2012). McNee et al. (2003) compared three interface strategies for eliciting movie ratings by 192 users. It was found that a higher level of user control over the interaction gives rise to a more accurate preference UM. The experiment also revealed that although the user-controlled group spent more time, they did not perceive any additional effort.

Besides, design of adaptive RSs, which tailors the user interaction effort to his / her individual characteristics and motivates users to expend this effort, might be a good solution to provide more trade-off and ultimately a pleasant experience for the users. Therefore, this research does not end here and encourages further researches on UI design for RSs and SRSs. Recommendations for further studies on this area are discussed in Chapter 8.

6.4 Summary

In this chapter, the proposed UI called rScholar was discussed. rScholar is mostly designed based on the empirical results (identified in objective 1) of the most influencing elements of UID adequacy such as consistency, signifier and, display as well as IxD adequacy including preference elicitation, preference refinement, privacy consideration and, explanation. rScholar also was designed to perform effectively for the data acquisition for the proposed Bayesian UM developed in objective 2 and to present four categories of novel, diverse, accurate and, popular paper recommendations. . In this chapter also the importance of UI and IxD in RSs and SRSs along with the UI guidelines were briefly discussed. To design the proposed UI, five steps are performed, first four steps was explained. However, the details of step 5 (evaluation) is discussed in chapter 7.

CHAPTER 7: EVALUATION

This chapter provides details of the methods applied in order to evaluate the UM and UI developed respectively in Chapters 5 and 6. Since the proposed products in this research are based on the contexts influencing UX of SRSs, multi evaluation methods of offline and user studies have been chosen to ensure that the products are effective for SRSs (G.Marcot, 2012)(Aggarwal, 2016; Bart P. Knijnenburg et al., 2012; Pu et al., 2012a). In the offline method, the performance and robustness of the Bayesian UM have been examined. The experts and users have assessed the UI, which is a significant way to improve the quality of a developed software (Wiegers, 2002). The objective is to gather feedback on the rScholar and to compare it with the UI of Google Scholar to improve based on the feedbacks from the experts and users. Several tests including T-test, Mann-Whitney (MW), Kruskal Wallis (KW), and Wilcoxon signed-rank as well as Friedman Kendall's Coefficient of Concordance were applied to analyze the experts and users' evaluation data. In this chapter, first, it is discussed how the appropriate methods and metrics for evaluation of UM and UI have been selected and then, each of the evaluation task is separately discussed and accordingly, the results are reported.

7.1 Bayesian UM evaluation

In the following section, the selected method and metrics for the evaluation of Bayesian UM are discussed, and then, the results are reported.

7.2 Methods and metrics for BN model evaluation

As discussed in chapter 2, there are three different methods for the evaluation of the RSs including offline, online and user studies (J. L. Herlocker et al., 2004). However, each method has its shortcomings and strengths (Shani, 2011). In this research, the offline method, the highly used method in the evaluations of SRSs (Beel, Genzmehr, et al., 2013; Beel & Langer, 2014; Beel, Langer, Genzmehr, et al., 2013b), is applied to evaluate the

BN model. Based on the literature, the majority of studies have applied the offline method and have assessed the performance and robustness of the BN models (Seixas, Zadrozny, Laks, Conci, & Saade, 2014) (Korb & Nicholson, 2003; Kuenzer, Schlick, Ohmann, Schmidt, & Luczak, 2001) by using the metrics such as Entropy, F-measure (G.Marcot, 2012), Mean Square Error (MSE), and Mean Cross Entropy (MXE)(Flores et al., 2011; Margaritis, 2003). Additionally, Albert (2009) and Long et al. (2015) have also recommended to evaluate the performance of the applied BN learning algorithm by employing the measure of expected loss (Long et al., 2015) (Albert, 2009).

Elicitation review is another method for the BN models evaluation (Tibshirani et al., 2013) (Korb & Nicholson, 2003). The aim of elicitation review is to check the clarity and consistency mostly in the structure of the BN model by knowledge engineers and domain experts' involvement. As explained in Chapter 5, rather than the automated structure learning, in this research, the BN structure model was built by using the knowledge engineer who examined the clarity and consistency as well. Therefore, the elicitation review method has been ignored in the evaluation part.

As Figure 7.1 shows, the entropy metric is applied to measure the robustness of BN model structure and parameters. The expected loss is also gauged in order to evaluate the performance of the applied learning algorithm (GS algorithm). Three predictive performance metrics of F-measure, MSE, and MXE have been applied to evaluate the performance of the BN model. Moreover, the cross validation technique has been used to perform the above mentioned assessment. Before reporting the results, the cross validation technique is briefly discussed here.

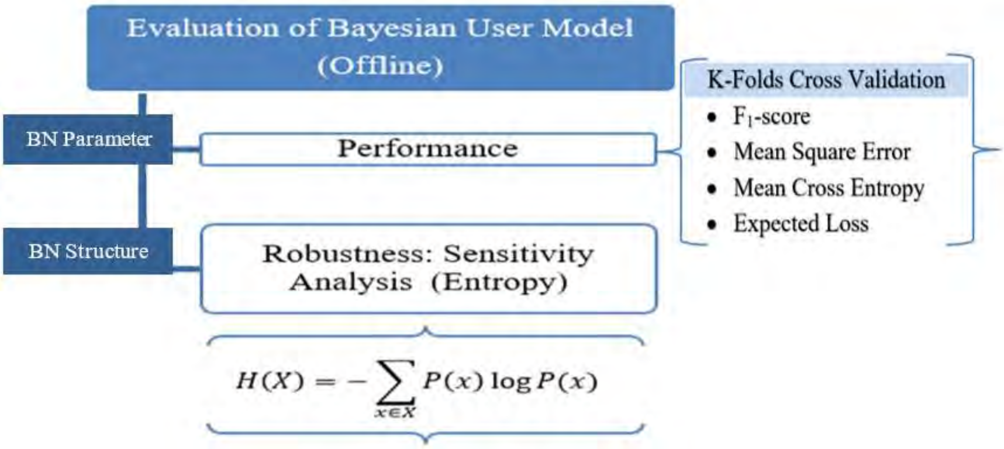


Figure 7.1: BN model evaluation measures

7.2.1 K-Folds Cross Validation

According to Albert(2009), cross-validation is the most widely used technique to validate statistical models such as BNs and to select suitable values for their tuning parameters. It can be applied to evaluate any combination of structure learning algorithms, parameter learning methods, and the respective tuning parameters. There are three different techniques for the cross validation including Holdout, K-Folds, and Leave-one-out (Hastie et al., 2009). For examining the BN model predictive performance, the K-Folds Cross Validation (KFCV) was used. KFCV is an evaluation technique model, which iteratively selects K different learning sets and test sets, and then, based on these sets, learns the networks and evaluates the performance. Compared to the other cross validation techniques mentioned above, in KFCV technique, most of the data is used for the model fitting and validation; hence, the data bias is reduced significantly (G.Marcot, 2012). Besides, the training and test sets are interchanged and this adds to the effectiveness of this technique.

In the KFCV, the original dataset is split into equal K subsets/folds and is repeated K times. In each time, one of the K subsets is applied as the test or validation set and the remaining subsets (K-1) are utilized as the training set.

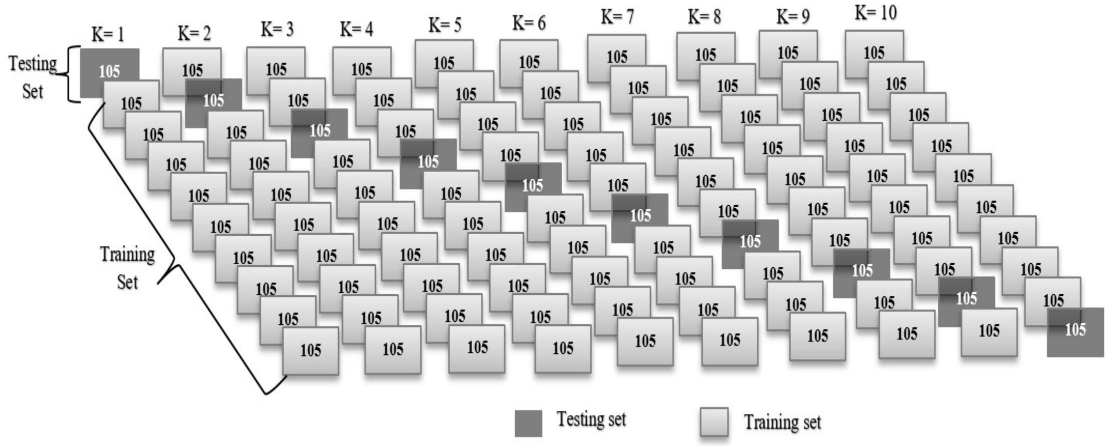


Figure 7.2: Visualisation of K- Fold Cross Validation (K=10)

There is no rigorous rule for the number of K (G.Marcot, 2012), but in BN performance validation, K=10 is mostly preferable (Scutari & Denis, 2014). Thus, in this research, K= 10 is also considered which means for the total dataset of 1053 cases, each fold contains around 105 randomly drawn cases for the test (leave out for the test set), and the remaining 964 cases as the training set (shown in Figure 7.2). Prior to the explanation of results, it is explained how the dataset is randomly split using the KFCV technique.

7.2.1.1 Dataset randomly split into K-Fold

As mentioned before, the golden standard is 10 runs of 10-fold cross validation (Scutari & Denis, 2014) in particular for small datasets. Cross validation can be applied by using bn.cv() method in "bnlearn" package (R programming) as shown in the following code.

```
> library(bnlearn)
> data(bndata)
> bn.cv(bndata, loss = "pred", k = 10, loss.args = list(target="T"), debug
= TRUE)
# setting the debug option to TRUE displays the results for the loss
function per fold
```

7.2.2 Bayesian UM evaluation results

The results of the structure and parameters of the BN model using the aforementioned metrics are reported in the following sections separately.

7.2.2.1 Robustness- Sensitivity analysis

For examining the BN structure robustness, the sensitivity analysis is carried out which analyses how sensitive the network outputs or parameters are against the inputs (observations) (Tibshirani et al., 2013)(Korb & Nicholson, 2003). This analysis assures the BN structure is correct or highlight errors(Chan & Darwiche, 2004) (Hansson & Sjökvist, 2013).The Entropy is the widely used metric in sensitivity analysis to examine the uncertainty in a probability distribution or BN modeling (Tibshirani et al., 2013)(Korb & Nicholson, 2003). In the proposed UM, the inputs are the users' contexts and the outputs are the users' information needs for novel, accurate, diverse, and popular papers. Therefore, it is examined how the model diagnoses for novel, accurate, diverse and, popular papers are sensitive against the users' contexts changes. Therefore, the Entropy *H(X)* in the target node *X*, where for example; *X* represents the probability of the user needs a novel paper (positive for diagnosis of novel paper) is calculated as (Korb & Nicholson, 2003);

$$H(X) = -\sum_{x \in X} P(x) \log P(x) \tag{7.1}$$

Thus far, there is only limited support in current BN software tools for sensitivity analysis (Tibshirani et al., 2013); In this research, the "sensitivity to findings" function in the Netica's software was used to measure the Entropy of four nodes of novel, accurate, diverse, andpopular papers against the users' contexts.

Since the calculation of Entropy is performed by using GUI in the Netica software (Figure 7.3), it gives this benefit to change the nodes which have high Entropy in the GUI and then, calculate the new Entropy for the target node easily. The advantage of using

Netica software for calculating the Entropy is its GUI. As shown in Figure 7.3, the target node *H(X)* is chosen and, "*the sensitivity to findings"* function from the network menu (Network → Sensitivity to Findings) is run, then, the sensitivity report for the target node is displayed. Entropy is calculated for the nodes of novel, accurate, diverse, and popular papers after performing the Sensitivity to Findings function using each piece of evidence (contextual nodes) separately.

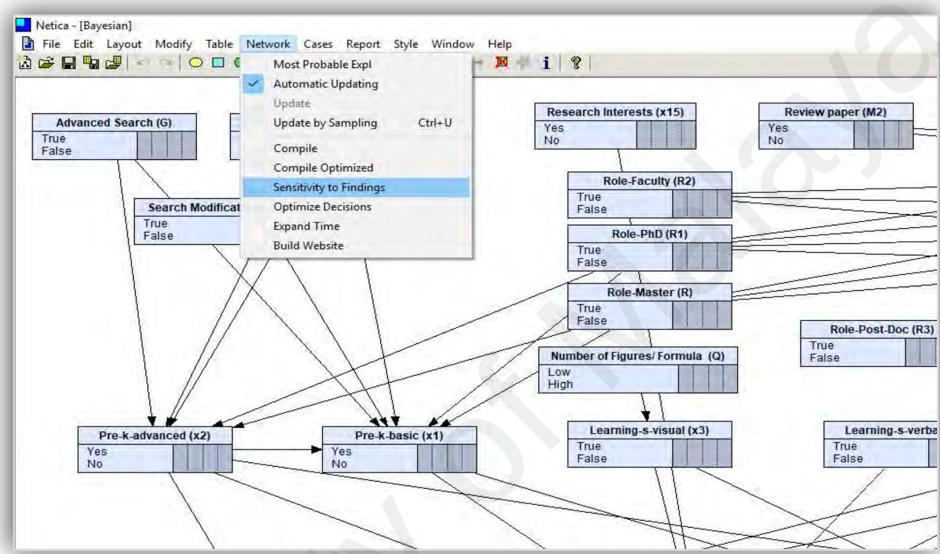


Figure 7.3: Sensitivity analysis using Netica software

The sensitivity analysis or entropy values for diagnosis of $H(X_1), H(X_2).., H(X_3), H(X_4)$ where the contextual nodes are evidences (C), are defined as follows:

$$H(X_1) \rightarrow pr(X_1 = \text{ Positive}|C) > 0.5 \ \ (Novel\ paper)$$

$$H(X_2) \rightarrow pr(X_2 = \text{ Positive}|C) > 0.5\ (Accurate\ paper)$$

$$H(X_3) \rightarrow pr(X_3 = \text{ Positive}|C) > 0.5\ (Popular\ paper) \qquad (7.2)$$

$$H(X_4) \rightarrow pr(X_4 = \text{ Positive}|C) > 0.5\ (Diverse\ paper)$$

So, the entropy value helps better understand which node (evidence (C)) is more relevant to corresponding positive or negative diagnosis.

```
if Pr ((X = Positive|C))>0.5:
   print "C is in a positive diagnosis group"
else:
   print "C is in a negative diagnosis group"
```

For the binary variables, the entropy ranges from zero to one ($H(X)$: (0,1)). Zero indicates the minimum uncertainty and one represents the maximum uncertainty for a target node against the evidence. Therefore; $H(X) = 0$ represents $\rightarrow \min_{uncertanity}$ and $H(X) = 1$ represents $\rightarrow \max_{uncertanity}$. The sensitivity analysis results were ranked by entropy value in Table 7.1. These results only consider observations one at a time. Additionally, each variable or node was associated to a diagnosis node, which has certain states.

Table 7.1: Sensitivity results

Node	X_1 Diagnosis $H(X_1)$	Node	X_2 Diagnosis $H(X_2)$	Node	X_3 Diagnosis $H(X_3)$	Node	X_4 Diagnosis $H(X_4)$
RO_1	0.37	RO_1	0.45	RO_1	0.10	RO_1	0.22
RO_2	0.43	RO_2	0.44	RO_2	0.13	RO_2	0.45
RO_3	0.27	RO_3	0.20	RO_3	0.36	RO_3	0.51
RO_4	0.22	RO_4	0.25	RO_4	0.37	RO_4	0.29
PK_1	0.45	PK_1	0.45	PK_1	0.29	PK_1	0.51
PK_2	0.51	PK_2	0.51	PK_2	0.45	PK_2	0.54
LS_1	0.29	LS_1	0.34	LS_1	0.44	LS_1	0.52
LS_2	0.51	LS_2	0.36	LS_2	0.42	LS_2	0.28
LS_3	0.24	LS_3	0.45	LS_3	0.46	LS_3	0.26
TA_1	0.22	TA_1	0.54	TA_1	0.56	TA_1	0.34
TA_2	0.54	TA_2	0.37	TA_2	0.53	TA_2	0.36
TA_3	0.32	TA_3	0.56	TA_3	0.44	TA_3	0.40
TA_4	0.35	TA_4	0.55	TA_4	0.57	TA_4	0.42
TA_5	0.44	TA_5	0.40	TA_5	0.40	TA_5	0.43
SE1	0.01	SE1	0.08	SE1	0.00	SE1	0.04
SE2	0.00	SE2	0.04	SE2	0.03	SE2	0.03
SE3	0.02	SE3	0.12	SE3	0.00	SE3	0.05
SE4	0.00	SE4	0.01	SE4	0.07	SE4	0.01
SE5	0.00	SE5	0.02	SE5	0.02	SE5	0.04
ST1	0.04	ST1	0.03	ST1	0.00	ST1	0.02
SM1	0.54	SM1	0.55	SM1	0.64	SM1	0.45
AS1	0.14	AS1	0.73	AS1	0.35	AS1	0.44
SI1	0.63	SI1	0.67	SI1	0.74	SI1	0.43
SI2	0.52	SI2	0.82	SI2	0.78	SI2	0.42
SI3	0.42	SI3	0.56	SI3	0.54	SI3	0.55
SI4	0.56	SI4	0.64	SI4	0.42	SI4	0.24
SI5	0.00	SI5	0.06	SI5	0.00	SI5	0.18
PY1	0.62	PY1	0.24	PY1	0.30	PY1	0.24
SU1	0.65	SU1	0.45	SU1	0.40	SU1	0.27
RU1	0.71	RU1	0.56	RU1	0.48	RU1	0.30
CN1	0.35	CN1	0.42	CN1	0.56	CN1	0.55
AR1	0.42	AR1	0.25	AR1	0.59	AR1	0.57
JR1	0.36	JR1	0.40	JR1	0.62	JR1	0.59
PR1	0.27	PR1	0.23	PR1	0.64	PR1	0.30
DN1	0.32	DN1	0.39	DN1	0.54	DN1	0.44
PT1	0.21	PT1	0.45	PT1	0.36	PT1	0.43
PT2	0.40	PT2	0.32	PT2	0.56	PT2	0.42
PT3	0.11	PT3	0.30	PT3	0.42	PT3	0.40

As Figure 7.4 shows, the results of sensitivity to diagnoses nodes revealed that, for all four nodes of novel, popular, diverse, and accurate, the entropy reduction values were approximately equal to zero for the nodes of SI5, SE1, SE2, SE3, SE4, SE5 and ST1. Further examination of the CPTs of the mentioned nodes confirmed that SI5, SE1, SE2, SE3, SE4, SE5 have no impact on the rest of the network. For this reason, these nodes were removed from the network. However, since the ST1 node has an impact on the PK node, this node remained in the network.

Interestingly, the results of sensitivity analysis also confirms the empirical outputs revealed in chapter 4 which indicates that the impact of environment context is not significant compared to other contexts. Thus, entropy values substantiate that BN model's outputs are not sensitive against the SE1, SE2, SE3, SE4 variables that represent the time (environment context). Figure 7.5 shows the maximum uncertainties for $H(X_1), H(X_2)., H(X_3), and\ H(X_4)$. The nodes of pre-knowledge and task are in a positive group. The nodes that are in a positive group for a novel, accurate, diverse, and popular papers are as follows:

$$\begin{aligned} H(X_1) &\to pr(X_1 = \text{ Positive}|PK, TA, RU, SU, SI) > 0.5 \\ H(X_2) &\to pr(X_2 = \text{ Positive}|SI, AS,\ TA, PK) > 0.5 \\ H(X_3) &\to pr(X_3 = \text{ Positive}|SI, PR, JR, Pk, TA) > 0.5 \\ H(X_4) &\to pr(X_4 = \text{ Positive}|SI, JR, PK, TA\ ) > 0.5 \end{aligned} \tag{7.3}$$

Based on the entropy values, a moderate uncertainty level has been obtained for the diagnosis nodes against the users' learning style nodes (SL1, SL2, and SL3) which divulges the moderate significance of this context compared to other contexts such as TA and PK for the accuracy of the BN model.

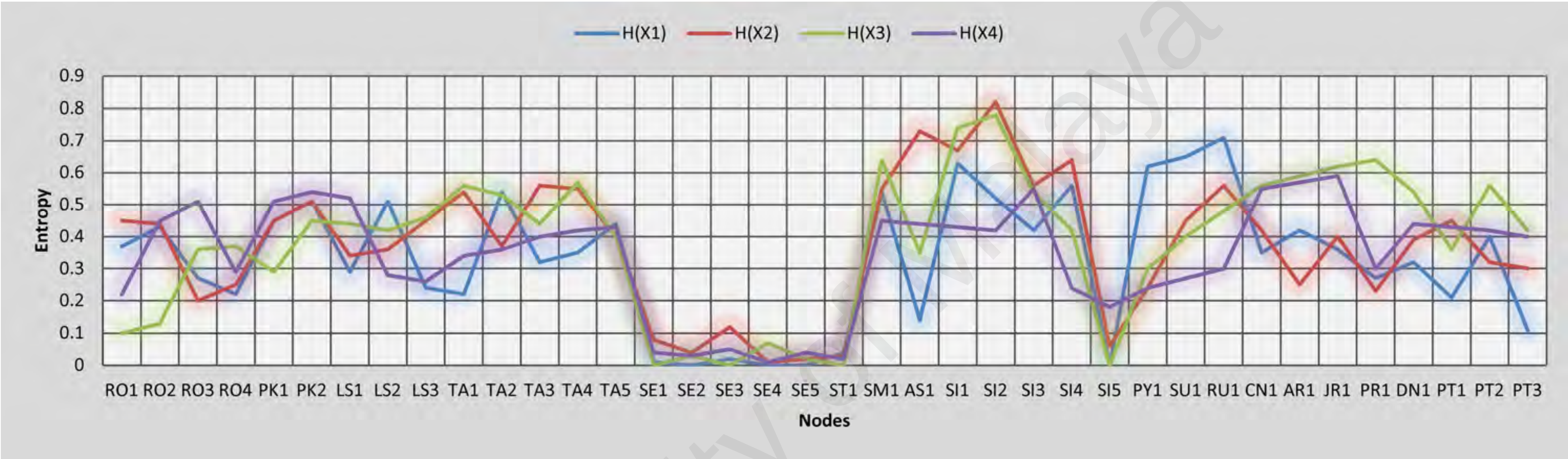


Figure 7.4: Entropy values of BN nodes

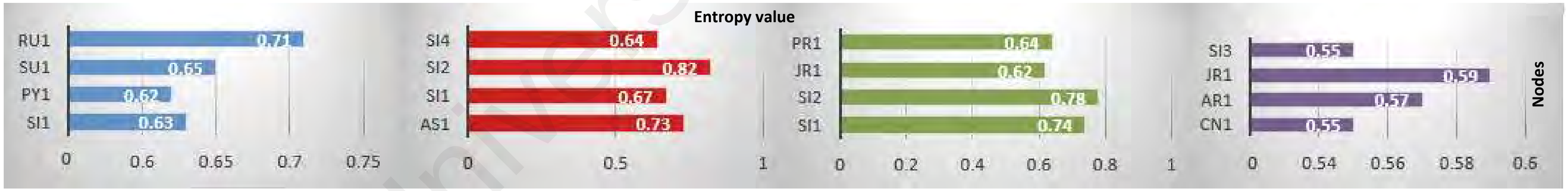


Figure 7.5: Maximum entropies

7.2.2.2 Comparison of BN algorithm- Expected loss

As mentioned in chapter 5, for learning the BN structure, a constraint-based learning algorithm called Grow- Shrink (GS) algorithm has been applied based on the guidelines indicated by (Albert, 2009). To evaluate the performance of the GS algorithm, the Expected Loss $\boldsymbol{\rho}(\mathbf{a})$ of the applied algorithm is calculated and compared to the $\boldsymbol{\rho}(\mathbf{a})$ of Max-Min Hill-Climbing (MMHC) algorithm which is a combination of constraint-based and search-and-score techniques (Hybrid algorithm) (Tibshirani et al., 2013) using the same dataset. The Bayesian expected loss is defined as the expected loss under the predictive distribution (G.Marcot, 2012):

$$\boldsymbol{\rho}(\mathbf{a}) = \int \boldsymbol{l}(\mathbf{a},\mathbf{y})\mathbf{P}(\mathbf{y}|\mathbf{x},\mathbf{D})\boldsymbol{dy} \tag{7.4}$$

As shown in Figure 7.4, if training examples are drawn independently at random according to unknown distribution P(x,y) and the learning algorithm analyzes the training examples and produces a function f, given a new point <x,y> drawn from P, the function is given x and predicts ŷ = f(x) therefore the loss L(ŷ,y) is measured (Schain & Schain, 2015). The goal of a Bayesian learning algorithm is to find the f that yields the the lowest expected loss $E_{P(x,y)}[L(f(x),y)]$. In other words, loss function is used to determine which Bayesian learning algorithm is better suited for a certain problem and dataset.

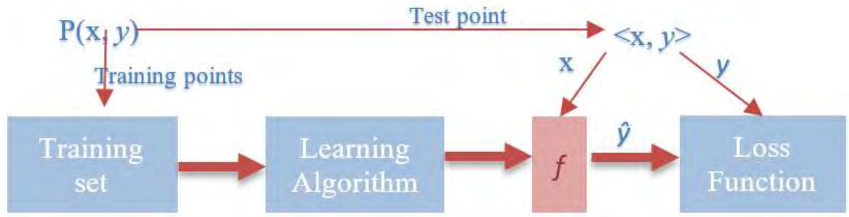


Figure 7.6: Loss function for a supervised ML algorithm

To calculate the $\boldsymbol{\rho}(\mathbf{a})$, the bn.cv method in "bnlearn" package was used. The less is the expected loss, the better is the algorithm in terms of the performance (Albert, 2009).

```
> bn.cv(bndata, bn = "mmhc", algorithm.args = list(score = "bde",
iss = 1))

target learning algorithm:            Max-Min Hill-Climbing
number of folds:                      10
loss function:                        Log-Likelihood Loss (disc.)
expected loss:                        2.350423
```

```
> bn.cv(bndata, bn = "gs", algorithm.args = list(score = "bde", iss
= 1))

target learning algorithm:            Grow- Shrink
number of folds:                      10
loss function:                        Log-Likelihood Loss (disc.)
expected loss:                        0.341486
```

As Figure 7.5 represents, the results disclosed that the EL of GS algorithm is ($\rho(a)$ = 0.341486) and MMHC is ($\rho(a)$ = 2.350423) applying KFCV method.

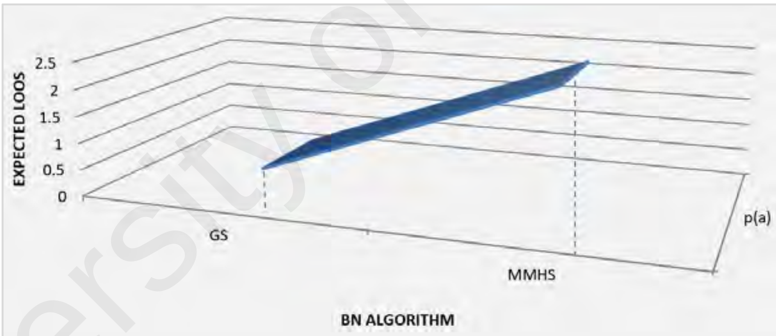


Figure 7.7: Loss function

It means that the GS algorithm produces a BN model which fits the data of this study better than the MMHC algorithm. Therefore, GS performs better than the MMHC algorithm and it is more fitted to the dataset of this study for the BN modeling.

7.2.2.3 Predictive performance assessment

To assess the predictive performance of the BN model, three metrics of F_1-score, Mean Square Error (MSE), and Mean Cross- Entropy (MXE) have been estimated using CFCV method (K=10) as recommend in the literature (Seixas, Zadrozny, Laks, Conci, & Saade,

2014) (G.Marcot, 2012) (Albert, 2009) (Olson & Delen, 2008). In the following sections, first, a brief description of each measure is presented and accordingly, the results of the predictive performance assessment are reported.

(a) F_1-score

F_1-score or F-measure is a very common measure and is obtained by the harmonic mean of precision and recall (Olson & Delen, 2008) (Seixas, Zadrozny, Laks, Conci, & Saade, 2014). F-score is a composite measure which benefits algorithms with higher sensitivity and challenges algorithms with higher specificity(Sokolova, Japkowicz, & Szpakowicz, 2006).

(b) Mean Square Error (MSE)

MSE is a metric that can be applied to examine how well a statistical model fits the data (Witten, Frank, Hall, & Pal, 2016) by calculating the mean of the squared difference between the estimated parameters by the model and known parameters. MSE can be used to minimize the errors of a given model. For dataset D_N which contains $i\ th$ samples, the MSE prediction error is calculated by Equation 7.5 (G.Marcot, 2012) (Witten, Frank, Hall, & Pal, 2016) as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad (7.5)$$

(c) Mean Cross- Entropy (MXE)

MXE is used to measure in average how close all predicted probabilities are to the true probabilities. It can be shown that minimizing the cross entropy gives rise the maximum likelihood hypothesis. Skalak and Caruana (2007) showed that for the probabilistic output models with small available data, MXE is the most robust metric (Skalak, Niculescu-Mizil, & Caruana, 2007) in comparison to other accuracy metrics. The MXE for N cases is calculated by Equation 7.6 (Witten, Frank, Hall, & Pal, 2016) (Huang, Ling, Zhang, & Matwin, 2008).

$$MXE = \frac{1}{N}\sum_{i=1}^{N} -y_i . \log[Pr(y_i = 1)] - (1 - y_i) . \log[1 - \Pr(y_i = 1)] \qquad (7.6)$$

As mentioned earlier, 10-fold cross validation has been applied for the experiment which is particularly recommended for the small datasets and assures a comparative data distribution in all folds (Scutari & Denis, 2014). The training was iterated 10 times using 90% data and testing using 10% data on the corresponding fold. Table 7.2 represents F1 score, MSE, and MXE of the BN model (overall and four levels) for each of the folds. In addition, the mean of the measures has been calculated. The best score for F1 score, MSE, and MXE are respectively 1, 0, and 0 which indicate that the results surpass the determined thresholds. Figure 7.8 shows the average for means calculated for 10 folds. The performance results in this research are promising, however, better results might be achieved by using a large dataset with more instances.

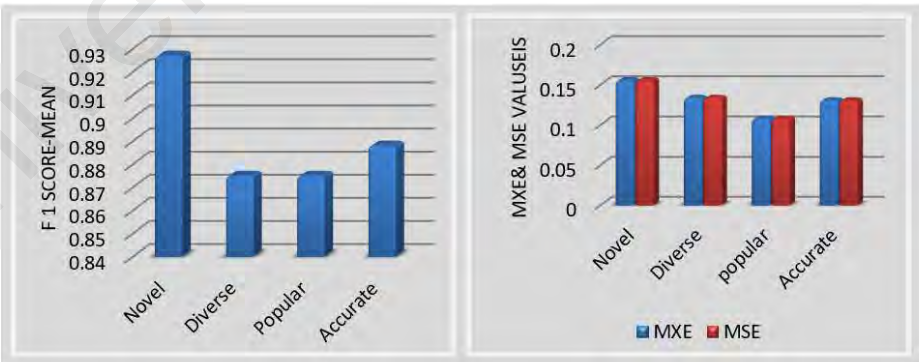
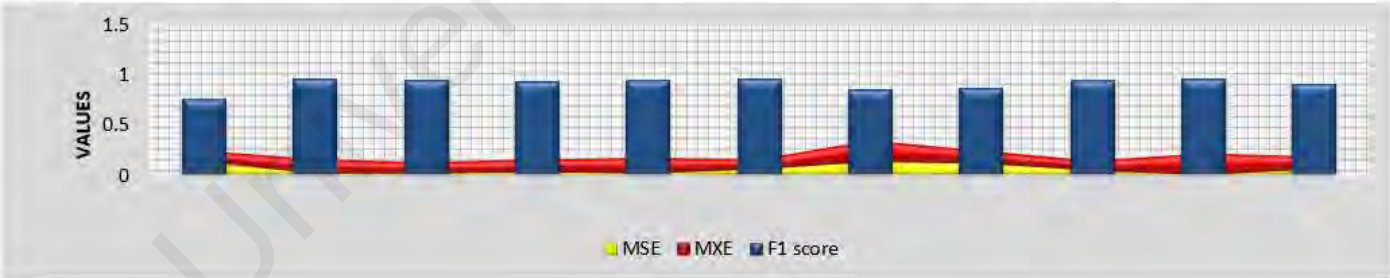


Figure 7.8: F1score, MXE & MSE means of 10 folds

Table 7.2: BN model predictive performance results

Dataset	*Folds* / *Metrics*	*Fold-1*	*Fold-2*	*Fold-3*	*Fold-4*	*Fold-5*	*Fold-6*	*Fold-7*	*Fold-8*	*Fold-9*	*Fold-10*	*Mean*
Novel	F_1	0.89	0.91	0.98	0.86	0.92	0.93	0.96	0.88	0.96	0.98	0.92
	MSE	0.12	0.11	0.02	0.15	0.14	0.03	0.09	0.05	0.01	0.01	0.07
	MXE	0.22	0.16	0.17	0.20	0.14	0.15	0.18	0.12	0.11	0.11	0.15
Diverse	F_1	0.92	0.98	0.93	0.86	0.85	0.87	0.85	0.68	0.86	0.95	0.87
	MSE	0.02	0.03	0.05	0.14	0.12	0.07	0.02	0.15	0.11	0.03	0.07
	MXE	0.10	0.13	0.10	0.13	0.15	0.16	0.15	0.10	0.20	0.12	0.13
Popular	F_1	0.92	0.98	0.93	0.86	0.85	0.87	0.85	0.68	0.86	0.95	0.87
	MSE	0.05	0.05	0.03	0.12	0.12	0.05	0.1	0.02	0.12	0.05	0.07
	MXE	0.04	0.03	0.12	0.12	0.05	0.10	0.20	0.12	0.10	0.20	0.10
Accurate	F_1	0.87	0.88	0.96	0.96	0.95	0.94	0.76	0.88	0.75	0.93	0.88
	MSE	0.09	0.12	0.02	0.08	0.05	0.03	0.18	0.12	0.11	0.01	0.08
	MXE	0.11	0.13	0.12	0.10	0.13	0.12	0.12	0.15	0.11	0.22	0.13
Overall	F_1	0.76	0.96	0.95	0.94	0.95	0.96	0.86	0.87	0.95	0.97	0.91
	MSE	0.14	0.03	0.02	0.04	0.02	0.06	0.14	0.12	0.05	0.01	0.06
	MXE	0.10	0.13	0.12	0.12	0.15	0.10	0.20	0.12	0.10	0.20	0.13
Acronyms	MXE→Mean Cross Entropy [0, ∞] Best score = 0; MSE→Mean Square [0, 1] Best score = 0; Error F_1→ F_1Score [0, 1] Best score = 1											



7.3 UI evaluation

In the following sections, the selected method and metrics for the evaluation of UI are discussed, and then, the results are reported.

7.3.1 Methods and metrics for evaluation of the UI

The results of empirical framework in Chapter 4 revealed that the UiD and IxD adequacy features influence UX of SRSs. In chapter 6, rScholar (proposed UI) was designed with regard to the identified UiD and IxD adequacy factors. In this section, the evaluation of rScholar is performed and in turn, the results are discussed.

In the study conducted by Vermeeren et al(2010) 96 methods for UX evaluation were collected and analyzed. In addition, a list of all UX evaluation methods have been provided on *http://www.allaboutux.org/all-methods*. UX Curve method, Evaluating Long-Term User Experience, EAX are a few samples, but none of them has been applied in evaluation of RSs so far. The results of SLR conducted on evaluation methods and metrics of SRSs (discussed in Chapter 2) showed that UX evaluation has only started recently and is still in its infancy. Besides, very little attention has been paid to the UI design in the RSs studies (Beel, 2011). Hence, there are not many certain metrics for measuring UI adequacy. However, Knijnenburg et al. (2012) and Pu & Hu (2012) have proposed frameworks and guidelines for evaluation of RSs from a user- centric perspective. Several guidelines by the above-mentioned studies emphasize UI and IxD adequacy already discussed in details in Chapter 4 and Chapter 6. For the rScholar evaluation, it was concluded that combining and inspiring survey questionnaires (UI & Ix aspects) of the above-mentioned studies with the questionnaire (UI & Ix aspects) in this study (Chapter 4) might provide a comprehensive method about the UI evaluation of Scholarly recommending systems. The questionnaire has 34 questions along with the data codes presented in the Appendix O. As mentioned earlier, the expert evaluation and user studies have been performed in order to evaluate the UI. In users' studies, users' experiences were examined over time (at three months intervals). Figure 7.9 depicts the tests applied in order to analyse the users' and experts' feedbacks on the rScholar evaluation.

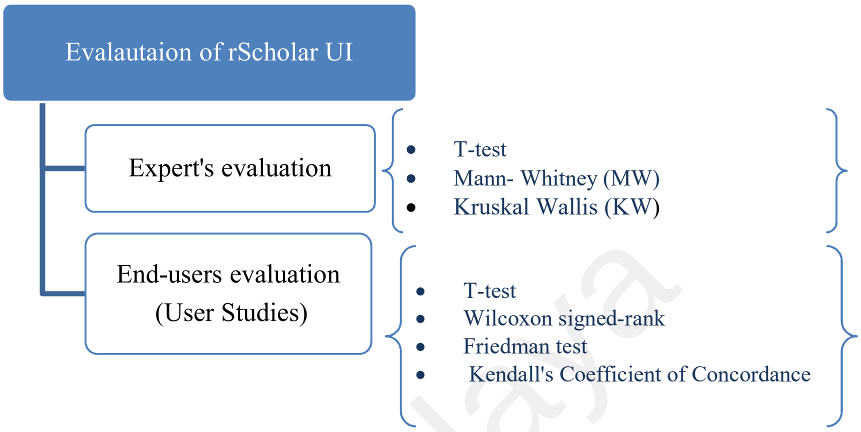


Figure 7.9: Evaluation method and metric for the UI

7.3.2 Expert Evaluation

Expert evaluation has been accepted as a significant way to improve the quality of a developed software and as a complement for testing other products (Wiegers, 2002). Thus, this study adopted experts in the evaluation of rScholar. The objective is to gather feedback on the rScholar and to compare it with the UI of Google Scholar to improve according to the suggestions and opinions of the experts. The Goal, Question, Metric (GQM) statement to perform this validation is shown in Figure 7.10.

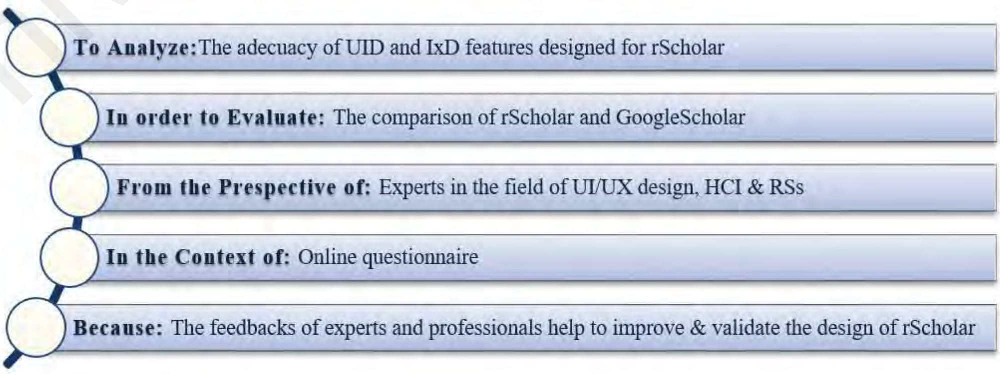


Figure 7.10: GQM statements to perform validation with experts

7.3.2.1 Evaluation instrument and procedure

The experts were selected based on their professional backgrounds, research interests, and their publications as well as job experience after pursuing their due willingness to participate. The invitations, along with the details of the validation study, were sent to 15 experts within the field. However, only nine experts were willing to participate. The research interest and country-wise distribution of the participants are shown in Table 7.3. The rest of the details of the experts are kept hidden in order to maintain the privacy of personal data.

Table 7.3: Expertise of the participants

No	*Experts' Research interests/ profession*	*Field*	*Country*
1	RSs, HCI & Cognition	Academia	Italy
2	UI/UX Design	Industry	Canada
3	UI Design	Academia	Malaysia
4	HCI, Interaction Design, Computer Supported Cooperative Work	Academia	Canada
5	HCI	Academia	Malaysia
6	RSs, HCI, UX Research, Personalization Technology, User Modeling	Academia	Switzerland
7	UX Design	Industry	Iran
8	UX/UI Design	Industry	Malaysia
9	UX Design	Industry	Iran

Both the rScholar and the questionnaire were accessible through *www.rscholar.com.* First, the participants were asked to work with Googlescholar for five minutes without providing any information or guidance to how to use the system. Then, they were asked to use the rScholar and rate the features indicated in questionnaire in a 5- Likert scale for both Googlescholar and rScholar separately. The rates were recorded in an excel file automatically and the validity and missing data were checked at runtime. Figure 7.11 depicts the screenshots of paper recommending in Googlescholar.

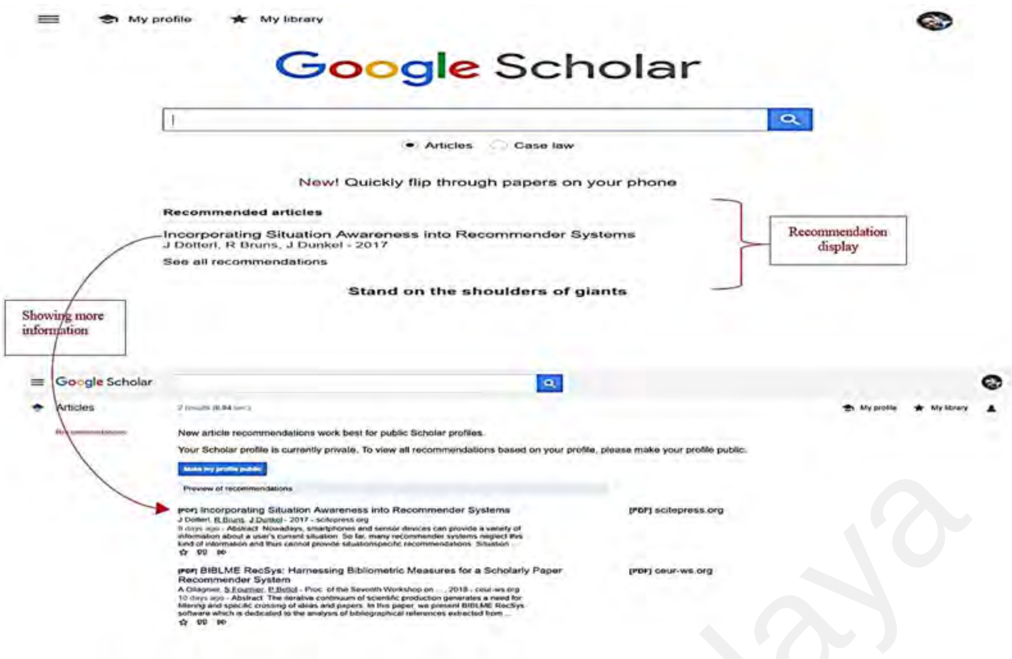


Figure 7.11: Googlescholar recommending system

The experts also were highly welcomed to express their additional feedbacks through a Skype talk or leave written comments via the available text field in the questionnaire. Among eight experts, only one expert left some comments and one expert was willing to make a Skype talk. After agreement time set, the Skype talk was made with the expert and recorded.

7.3.2.2 Differences of rScholar & Googlescholar

Before examining the UiD & IxD differences of rScholar and Googlescholar from the experts' viewpoints, the Normality test has been performed in order to choose the suitable parametric or non-parametric tests for the features indicated in the questionnaire. Figure 7.12 shows a sample of Normality test for the PE feature. The complete list of features and codes is provided in Appendix O.

As shown in Table 7.4, for comparison of two independent samples of rScholar and Googlescholar, for CSM[1] and RDM[2] features (sig. > 0.05), the parametric test of T-test and for comparing other features (sig. < 0.05), non-paramedic test of Mann- Whitney (MW) should be applied. In addition, for the features that are ordinal measures, Kruskal Wallis (KW) is applied.

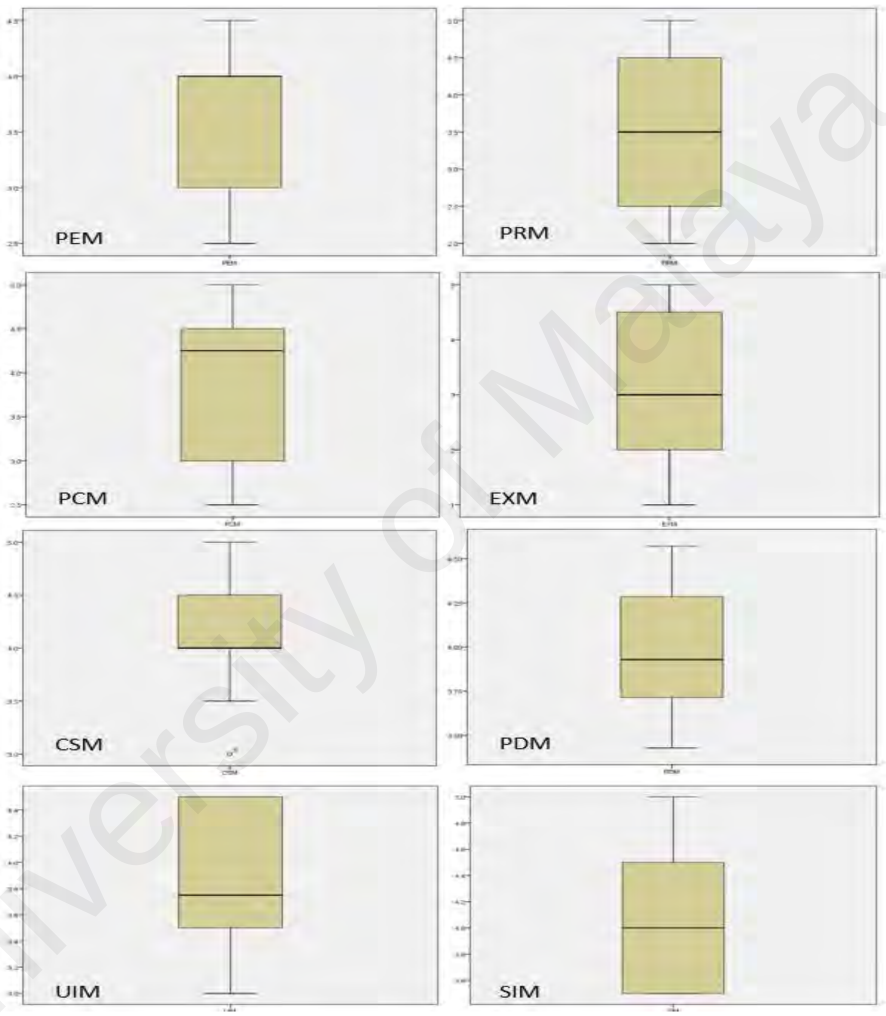


Figure 7.12: Normality test results

---

[1] CS is abbreviation of consistency and CSM presents the average of CS rates

[2] RD is abbreviation of recommendation display and RDM presents the average RD rates

Table 7.4: Results of Normality test-experts' data

	Kolmogorov-Smirnov[a]			*Shapiro-Wilk*			***Test type***	***Test***
	Statistic	df	Sig.	Statistic	df	Sig.		
PEM	.366	18	.000	.759	18	.000	Non- parametric	MW
PRM	.210	18	.035	.880	18	.026	Non- parametric	MW
PCM	.234	18	.010	.864	18	.014	Non- parametric	MW
EXM	.223	18	.019	.868	18	.017	Non- parametric	MW
CSM	.208	18	.038	.922	18	.138	Parametric	T-test
RDM	.170	18	.182	.957	18	.541	Parametric	T-test
UIM	.211	18	.034	.842	18	.006	Non- parametric	MW
SIM	.205	18	.043	.871	18	.018	Non- parametric	MW
Ordinal measures							Non- parametric	KW

a. Lilliefors Significance Correction

In the following sections, the parametric and non-parametric test results are discussed.

(a) T-test results

As mentioned, t-test is applied to examine if there is any significant differences between the CSM (consistency) and RDM (recommendation display) features in comparison of the two independent systems of rScholar and Google scholar. As the data analysis results in Table 7.5 shows, there is not significant differences (sig. = .622 >.05) between these two features in two systems, however, by comparing the means (r_CSM = 4.00 < g_CSM = 4.27; r_RDM =4.22 > g_RDM=3.77) (Table 7.6), it is concluded that experts have indicated more rates to the Google scholar's consistency than rScholar and also expressed more rates to rScholar recommendation display than Googlescholar. The reason of giving more consistency rates to Googlescholar might be the familiarity of experts with Google products and their design.

Table 7.5: Independent samples test results (T-test)

	Levene's Test for Equality of Variances		*t-test for Equality of Means*						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
CSM (Equal variances assumed)	.253	.622	-1.048	16	.310	-.27778	.26498	-.83952	.28396
CSM (Equal variances not assumed)			-1.048	13.938	.312	-.27778	.26498	-.84635	.29079
RDM (Equal variances assumed)	.253	.622	-1.048	16	.310	-.27778	.26498	-.83952	.28396
RDM (Equal variances not assumed)			-1.048	13.938	.312	-.27778	.26498	-.84635	.29079

Table 7.6: Group Statistics for CSM & RDM

	C6	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
CSM (C6)	r_CSM	9	4.0000	.66144	.22048
	g_CSM	9	4.2778	.44096	.14699
RDM(C8)	r_RDM	9	4.2222	.23810	.07937
	g_RDM	9	3.7778	.23810	.07937
		r→ rScholar		g→Googlescholar	

(b) Mann Whitney & Kruskal Wallis tests results

As mentioned before, for the scaled measures, Mann Whitney tests and for the discrete measures Kruskal Wallis test have been applied in order to examine the defined features of rScholar and Googlescholar.

As Table 7.7 shows, among the defined features, the distribution of PEM (preference elicitation), PRM (preference refinement), and EXM (explanation) are not the same across categories of rScholar and Googlescholar (Sig. < .05), therefore, the null hypothesis is rejected. The distribution of the rest of features is the same across categories of rScholar and Google scholar which means from the experts' point of view statistically there is not significant difference between two systems among those features; hence, the null hypothesis is retained.

So, based on the test results, among the examined UiD and IxD features, statistically, there is not significant difference (sig. >.05) in CSM, RDM, PCM, IS and SIM features. However, there is significant difference (sig. <.05) in PEM, PRM, EXM features between rScholar and Googlescholar.

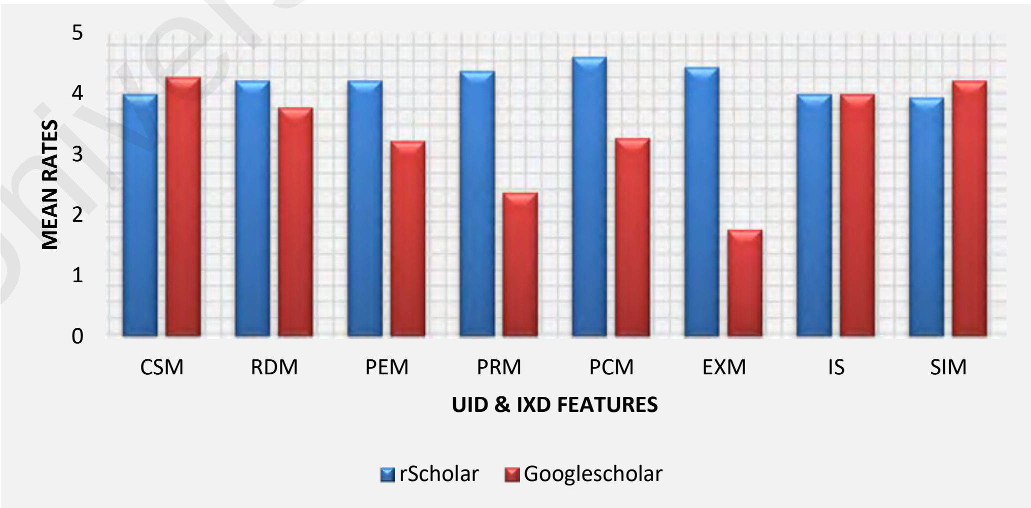


Figure 7.13: UiD & IxD differences in rScholar & Googlescholar

Figure 7.13 shows the experts' rates means. The means comparison reveals that experts have indicated more rates to rScholar for the features of Preference Elicitation (PEM) (r_CSM = 4.22 > g_CSM = 3.22), Preference Refinement (PRM) (r_CSM = 4.38 > g_CSM = 2.38), Privacy Consideration (r_PCM = 4.61 > g_PCM = 3.27) as well as Explanation of recommendation (EXM) (r_EXM = 4.44 > g_EXM =1.77) features in rScholar than Googlescholar.

7.3.2.3 Overall evaluation

The overall perception, feeling as well as appraisal of experts' feedback in terms of UiD and IxD on both UIs of rScholar and Googlescholar have been examined by a few questions (indicated in appendix O). The results of the experts' rates are illustrated in Table 7.7 and Figure 7.14.

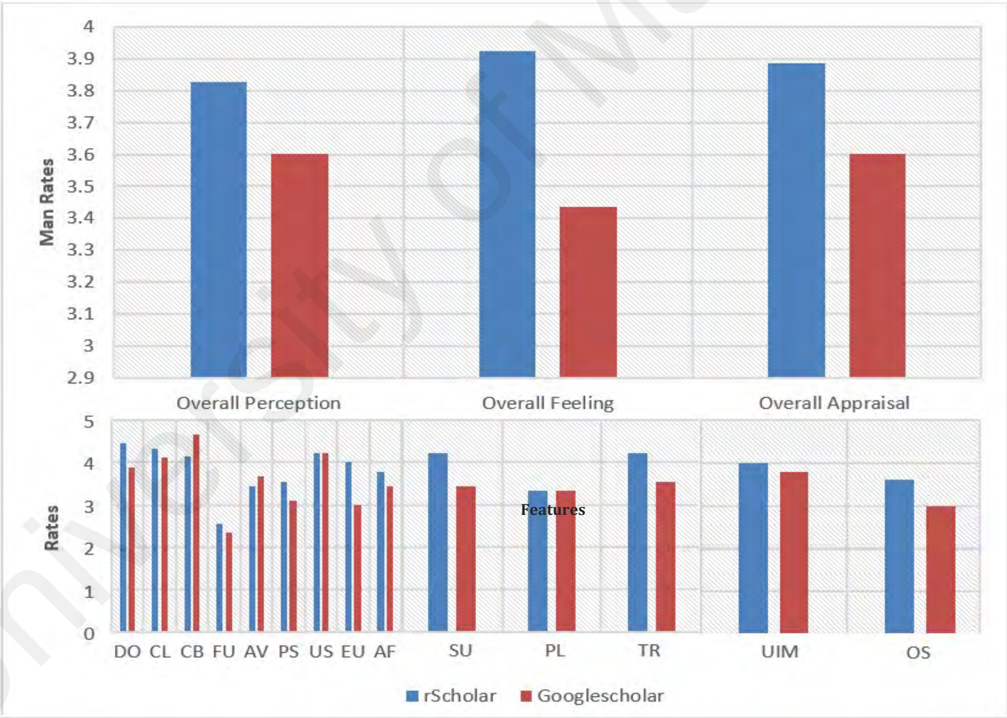


Figure 7.14: Experts' Overall evaluation

Table 7.7: Independent samples test results (Mann Whitney & Kruskal Wallis)

*Decision Null Hypo.		Sig.	Feature	Mean	Decision Null Hypo.		Sig.	Feature	Mean	Decision Null Hypo.		Sig.	Feature	Mean
PEM (C1)	×	.006	r-PEM	4.22	CL1(C10)	✓	.667	r-CL	4.33	AF1 (C17)	✓	.608	r-AF	3.77
			g-PEM	3.22				g-CL	4.11				g-AF	3.44
PRM (C2)	×	.000	r-PRM	4.38	CB1 C11)	✓	.445	r-CB	4.15	SU1 (C18)	✓	.169	r-SU	4.22
			g-PRM	2.38				g-CB	4.66				g-SU	3.43
PCM(C3)	×	.000	r-PCM	4.61	FU1(C12)	✓	.713	r-FU	2.55	PL1 (C19)	✓	.879	r-PL	3.33
			g-PCM	3.27				g-FU	2.33				g-PL	3.33
EXM (C4)	×	.000	r-EXM	4.44	AV1(C13)	✓	.396	r-AV	3.44	TR1(C20)	✓	.95	r-TR	4.22
			g-EXM	1.77				g-AV	3.66				g-TR	3.55
IS1 (C5)	✓	1.00	r-IS	4.00	PS1(C14)	✓	.268	r-PS	3.55	UIM (C21)	✓	.190	r-UIM	4.00
			g-IS	4.00				g-PS	3.11				g-UIM	3.61
SIM (C7)	✓	.190	r-SIM	3.94	US1(C15)	✓	1.00	r-US	4.22	OS (C22)	✓	.136	r-OS	3.77
			g-SIM	4.22				g-US	4.22				g-OS	3.00
DO1 (C9)	✓	.171	r-DO	4.44	EU1(C16)	✓	.077	r-EU	4.00	r→ rScholar			g→Googlescholar	
			g-DO	3.88				g-EU	3.00					

* Rejected null hypothesis (×): The distribution of feature is not the same across categories of rScholar and Google scholar

* Retain null hypothesis (✓): The distribution of feature is the same across categories of rScholar and Google scholar

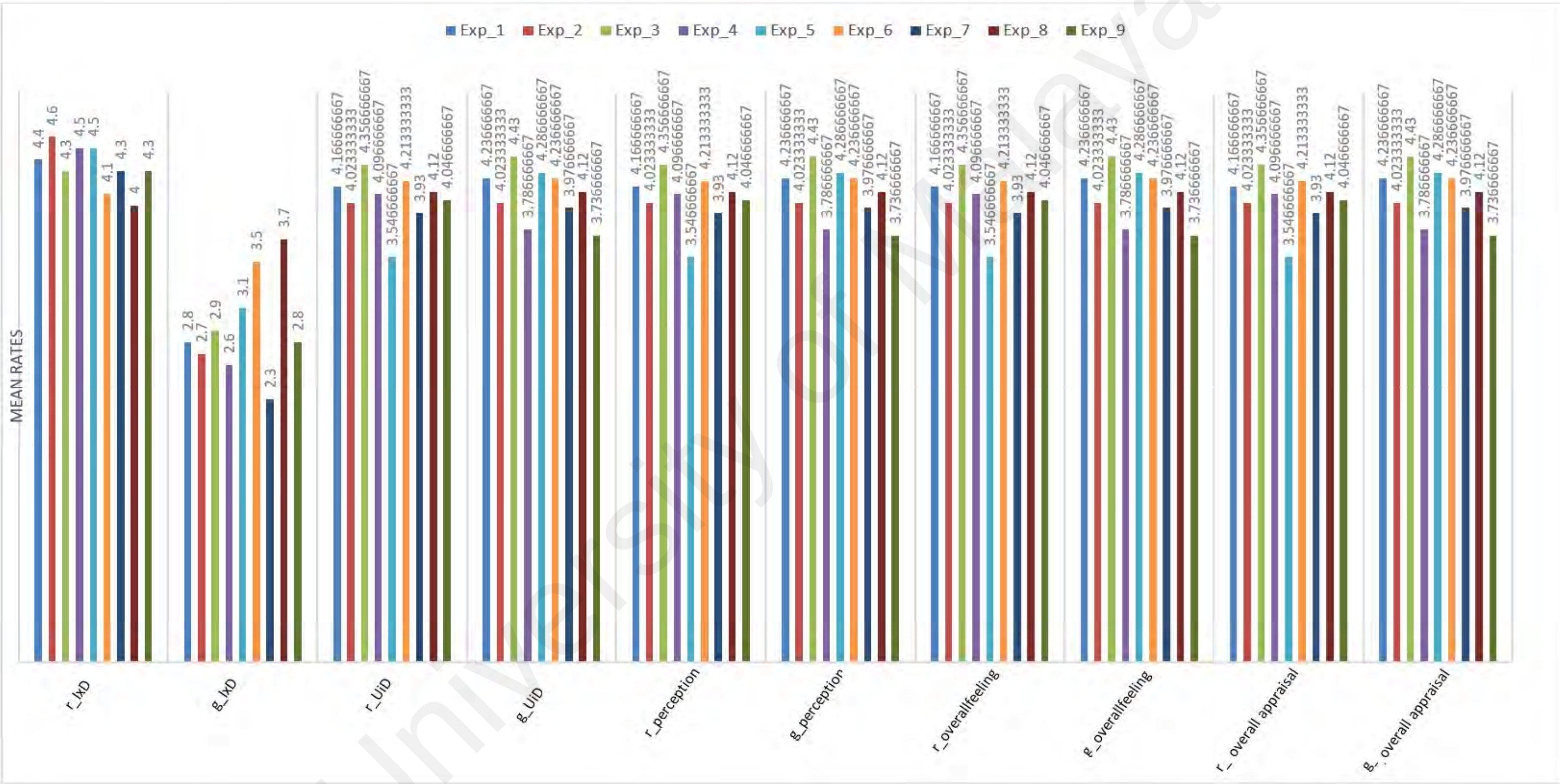


Figure 7.15: Mean experts' rates in rScholar & Google Scholar

As Figure 7.14, 7.15 demonstrate among the users' perception features, it is clear that the feature of FU (perceived fun) is the lowest rate for both UIs of rScholar and Googlescholar. This result might suggest more attentions on using gamification and visualisation in the field of SRSs. There is not any difference between two UIs among other features, except CB (perceived cognitive barrier). In this examination, the experts were asked to rate the CB item by considering the steps that they have performed in order to do what they have wanted. Based on the results, for this item, Googlescholar surpass the rScholar which means the Googlescholar design does create less cognitive barriers. Based on the experts' feedback, between the UiD & IxD in both UIs, IxD features including preference elicitation, refinement, privacy consideration, explanation, and information sufficiency have the lowest rates compared to the UiD features.

7.3.2.4 Evaluation of design ideas

As Figure 7.16 depicts, six different design ideas have been exploited in rScholar design including paper gift, logo, pie design, list design, logging page and background as well as user persona. They have been elaborated in Chapter 6.

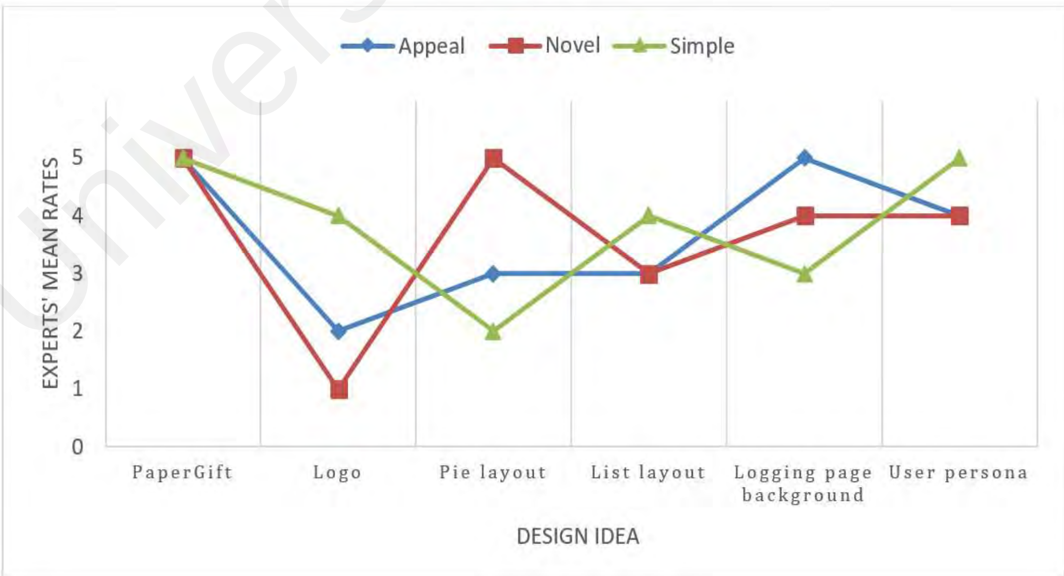


Figure 7.16: Experts' rates on design ideas in rScholar

Since the aforementioned design ideas have not been applied in other existing RSs, the experts' feedback on them were separately examined in three dimensions of "appeal", "novel" and "simple" by using scale of 1-5. As mentioned in Chapter 6, the pie layout design of recommendations has been a user- friendly design for the users of a movie recommender (Chen & Pu, 2014). In rScholar, the pie layout design is fairly appealing and extremely novel from the experts' viewpoints. However, the list layout design is simpler than the pie layout design. In other words, the experts have found the Pie layout not a simple design which might cause cognitive load or barrier for the end-users. Athough the rScholar logo is simple (Mean rate~4), among the all design ideas, the experts have given the logo the lowest rate.

A paper gift icon was designed which enables users to send a paper as a gift to their friends or colleagues. Experts have indicated the highest rate (Mean rate~5), of three dimensions of "appeal", "novel" and "simple" to this idea. Therefore, this idea has fully been supported by the experts of this research. The idea of user persona also seems the simple and engaging one. Finally and interestingly the idea of logging page (background) design has not been simple to understand for the experts. In the following section, the rScholar evaluation by the users is discussed.

7.3.3 Users studies evaluation

The objective of this evaluation by end-users' feedback is to examine the UiD and IxD features influencing UX two times in a three month interval by applying user studies method. The Goal, Question, Metric (GQM) statement to perform this validation is shown in Figure 7.17.

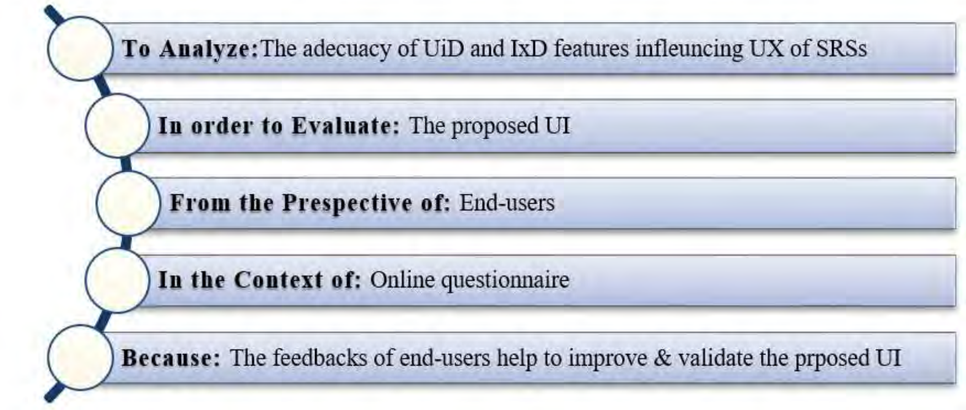


Figure 7.17: GQM statements to perform validation with end-users

7.3.3.1 Evaluation instrument and procedure

The announcement for the user studies evaluation was advertised through the bulletin of FSKTM at UM. The people, who accepted to participate in the evaluation phase in two interval times, received a gift-card of 20RM after the second session. Four groups of Master, PhD students, Post-doc researchers, and Faculty members have participated in this research. Each group consists five participants, which means the total of participants are twenty. Like the expert evaluation, both the rScholar and the questionnaire were accessible through *www.rscholar.com.* First, the participants were asked to work with rScholar for five minutes without providing any information or guidance to how to use the system. Then, they were asked to provide their rates on the features indicated in the questionnaire in a 5- Likert scale. The rest of the evaluation is like the experts evaluation. The data was collected in two times with three months time interval. The first data collection is a pre-test and the second is a post-test (after three months).

7.3.3.2 Differences in users' ratings after three months

First, the Normality test has been applied to select the appropriate statistical test in order to understand the differences between users' rating in pre-test and post-test (after three months) and the differences between the users in groups of PhD, Master, Faculty member and post-docs in evaluating the rScholar.

Table 7.8: Test of Normality for differences of pre & post tests

	Kolmogorov-Smirnov[a]			***Shapiro-Wilk***		
	Statistic	***df***	***Sig.***	***Statistic***	***df***	***Sig.***
Dif_PEM	.520	20	.000	.354	20	.00
Dif_PRM	.527	20	.000	.351	20	.00
Dif_PCM	.413	20	.000	.732	20	.00
Dif_EXM	.414	20	.000	.686	20	.00
Dif_CSM	.441	20	.000	.624	20	.00
Dif_SIM	.509	20	.000	.433	20	.00
Dif_RDM	.414	20	.000	.689	20	.00
Dif_UIM	.527	20	.000	.351	20	.00
a. Lilliefors Significance Correction						

As Table 7.8 and Figure 7.18 indicate, the results of Normality test on both the difference of scaled pre-test and post-test data showed that non-parametric tests should be applied to analyze the scalded data (Sig. = .00 < 0.05). Besides, the samples of pre-tests and post-tests are related samples. Therefore, the Wilcoxon signed-rank test is applied to examine the scalded data, which is the non-parametric test equivalent to the dependent t-test.

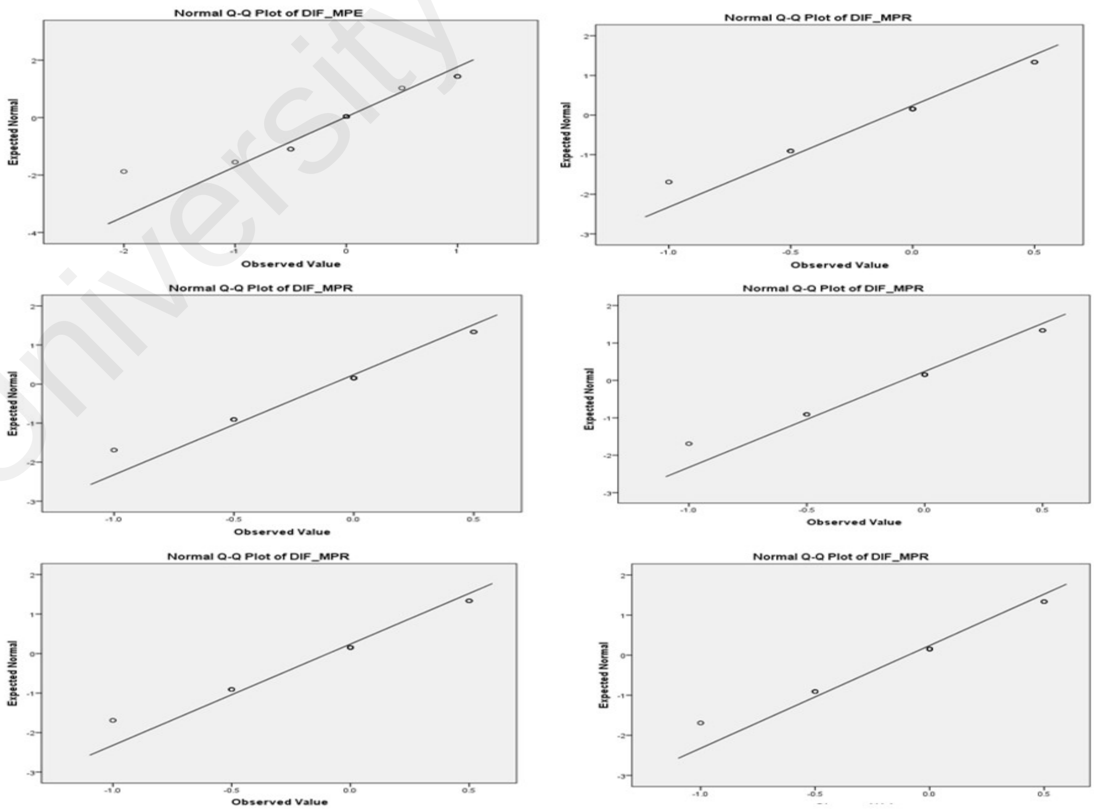


Figure 7.18: Test of Normality

(a) Wilcoxon Signed Ranks tests

To measure the changes resulting from experimental environment, pretest-posttest designs are mostly used to compare the changes in different groups and times. High changes in pre-test and post-test scores assert low reliability (Dimitrov & Rumrill Jr, 2003).

The Wilcoxon signed-rank test is applied for the comparison of two sets (2-ralated samples) of scores that come from the same participants to investigate any changes in scores from one time point to another, or when individuals are subjected to more than one condition and for the discrete data, also, the Wilcoxon signed ranks test is used (Woolson, 2007).

In user studies evaluation, the Wilcoxon signed ranks test is used to determine if the time (three months) as an intervention has a significant effect on the users' scores on UiD and IxD accuracy features and on users' overall evaluation.

As the results in Tables 7.9 and 7.10 reveal, the median of differences between Pre-test and Post-test equals zero (0) for all features which statistically means there is no difference between the scores of participants. In other words, there is no change in scores from the first time to the second time after three months, therefore, it can be discerned that the users' rates are reliable.

Table 7.9: Related samples test results-scale data (Wilcoxon Signed Ranks)

*Decision Null Hypo.		Sig.	Decision Null Hypo.		Sig.	*Decision Null Hypo.		Sig.	*Decision Null Hypo.		Sig.
Pre_PEM Post_PEM	✓	.180	Pre_PCM Post_PCEM	✓	.060	Pre_CSM Post_CSM	✓	.581	Pre_RDM Post_RDM	✓	.216
Pre_PRM Post_PRM	✓	.157	Pre_EXM Post_EXM	✓	.396	Pre_SIM Post_SIM	✓	.083	Pre_UIM Post_UIM	✓	.157

* Retain null hypothesis (✓): The median of differences between Pre-test and Post-test equals zero (0)
* Rejected null hypothesis (×):The median of differences between Pre-test and Post-test does not equal zero (0)

Table 7.10: Related samples test results- discrete data (Wilcoxon Signed Ranks)

Feature	*Pos_IS1 / Pre_IS1*	*Pos_DO1/Pre_DO1*	*Pos_CL1/Pre_CL1*	*Pos_CB1 /Pre_CB1*	*Pos_FU1 /Pre_FU1*	*Pos_AV1/ Pre_AV1*	*Pos_PS1/ Pre_PS1*
Z	-1.000[b]	.000[c]	.000[c]	.000[c]	-2.152[d]	.000[c]	-1.000[d]
Asymp. Sig. (2-tailed)	.317	1.000	1.000	1.000	.131	1.000	.317
Feature	*Pos_US1/Pre_US1*	*Pos_EU1/Pre_EU1*	*Pos_AF1/Pre_AF1*	*Pos_SU1/Pre_SU1*	*Pos_PL1 /Pre_PL1*	*Pos_TR1 / Pre_TR1*	*Pos_OS1/ Pre_OS1*
Z	.000[c]	.000[c]	-1.000[d]	.000[c]	-1.000[d]	.000[c]	.000[c]
Asymp. Sig. (2-tailed)	1.000	1.000	.317	1.000	.317	1.000	1.000

[b]. Based on positive ranks. [c]. The sum of negative ranks equals the sum of positive ranks. [d]. Based on negative ranks.

7.3.3.3 Differences in users' groups

In this section, it is examined if statistically there is any difference between the means scores of users' groups (Master, PhD student, Post-doc researcher, and Faculty member) on UiD and IxD features, and users' overall evaluation.

(a) Analysis of Variance (ANOVA)

To examine the above-mentioned difference, the Friedman Two- way Analysis of Variance (ANOVA) and Kendall's Coefficient of Concordance tests have been applied. Table 7.11 shows that the pretest-posttest data differences are not normal.

Table 7.11: Tests of Normality: Pretest-posttest differences

	Kolmogorov-Smirnov[a]			**Shapiro-Wilk**		
	Statistic	df	Sig.	Statistic	df	Sig.
Dif_IxD	.210	20	.022	.937	20	.206
Dif_UiD	.265	20	.001	.825	20	.002
Dif_PER	.492	20	.000	.425	20	.000
Dif_OFE	.538	20	.000	.236	20	.000
Dif_OAP	.527	20	.000	.351	20	.000
a. Lilliefors Significance Correction						

The analysis of two above-mentioned tests shown in Table 7.12 revealed that there is not statistically significant differences between the different groups of users in ratings to rScholar evaluation and that the distribution of all features in different groups are the same (Sig. = .00 < 0.05).

Table 7.12: Differences between users' groups

*Decision Null Hypo.		*Sig.1	*Sig.2	*Decision Null Hypo.		*Sig.1	*Sig.2
IXD_Mas IXD_PhD IXD_ Pos IXD_Fac	✓	.597	.567	UiD_Mas UiD_PhD UiD_ Pos UiD_Fac	✓	.351	.351
PEM_Mas PEM _PhD PEM _ Pos PEM _Fac	✓	.300	.300	OFE_Mas OFE _PhD OFE _ Pos OFE _Fac	✓	.392	.392
OAP_Mas OAP _PhD OAP _ Pos OAP_Fac	✓	.572	.572	*Retain null hypothesis (✓): The distribution of k_1,k_2 ,k_3 and k_4 are the same (k= feature) * Rejected null hypothesis (×):The distribution of k1,k2 ,k3 & k4 are not the same *Sig.1: Friedman Two- way Analysis of Variance (ANOVA) *Sig.2: Kendall's Coefficient of Concordance			
Mas: Master student; PhD: PhD student; Pos: Post-doc researcher; Fac: Faculty member IxD: Interaction Design adequacy;				UiD: User interface Design adequacy; PEM: User Perception; OFE: User Overall feeling; OAP: User Overall appraisal			

Therefore, the rScholar (UiD & IxD) features have the same effects on different user groups. In addition, there is not significant difference in overall evaluation of users' groups (Figure 2.19). However, between the UiD and IxD features, the UiD features including consistency, signifier and recommendation display represented lower scores (Figure 7.20).

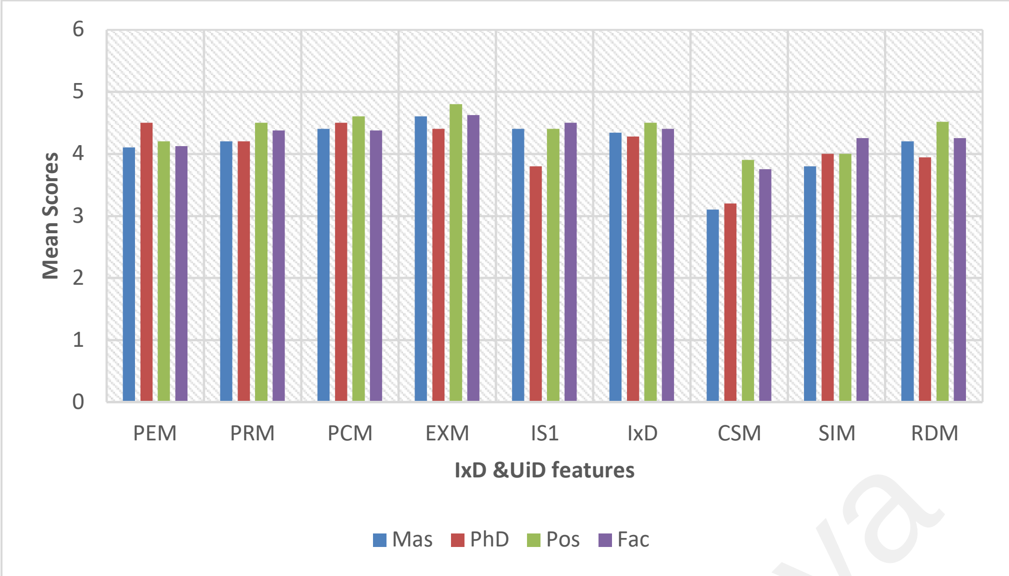


Figure 7.19: IxD &UiD features scores for different groups

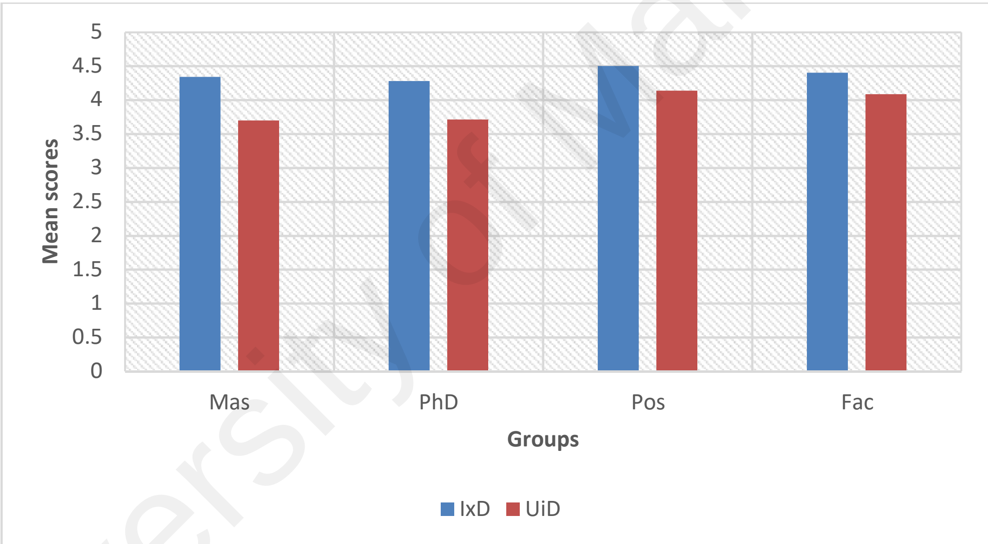


Figure 7.20: IxD &UiD scores for different groups

As discussed in Chapter 6, consistency is using familiar icons, colors, menu hierarchy, call-to-actions, and user flows when designing similar situations and sequence of actions. It makes users perceive the system as easy to use and stable (Nielsen, 1995) and it also leads less congestive load while using the system (Shneiderman, 2004). Based on the evaluation results, it was concluded that both experts and users have not rated the rScholar's consistency high compared to other features.

7.4 Summary

In this chapter, the details of the methods applied for evaluating the proposed UM and UI developed in Chapters 5 and 6 have been discussed. In the offline method, the robustness of the Bayesian UM was examined by sensitivity analysis. Entropy is the highly used metric for the sensitivity analysis. Aftercomputation of entropy reduction values for four diagnoses of novel, accurate, popular, and diverse paper; the nodes which showed minimum entropy to corresponding diagnoses were removed from the BN network. To evaluate the performance of the GS algorithm applied for BN modeming, the expected loss$\boldsymbol{\rho}(\mathbf{a})$ of the GS algorithm is calculated and compared to the $\boldsymbol{\rho}(\mathbf{a})$ of Max-Min Hill-Climbing (MMHC) algorithm. The results show that the applied algorithm performs better and is more fitted to the dataset of this study for the BN modeling. Finally, to assess the predictive performance of the BN model, three metrics of F_1-score, Mean Square Error (MSE), and Mean Cross- Entropy (MXE) have been estimated using CFCV method (K=10) which surpassed the thresholds and revealed the promising results. The proposed UI called rScholar has been evaluated by nine experts and twenty end-users including master, PhD students, post doc researchers, and faculty members. Several tests including T-test, Mann- Whitney (MW), Kruskal Wallis (KW), Wilcoxon signed-rank as well as Friedman Kendall's Coefficient of Concordance were applied to analyse the data. The results divulged that between the UID and IxD features, consistency has obtained the lowest rates not only from the end-users' evaluation but also from the experts' evaluation comparing with Googlescholar. Additionally, the results showed that there is not statistically significant difference between the different groups of users in rScholar evaluation and the examined features.

CHAPTER 8: CONCLUSION AND FUTURE WORK

This chapter discusses the overall work that has been carried out throughout this research. First, it presents the findings in line with the research objectives and research questions. Second, it presents the future work that can be conducted using this research.

8.1 Research objectives revisited

The following discussion revisits all of the research objectives related to this research.

8.1.1 Research objective 1

The first objective is to construct a framework which explores how contextual information influence UX and to assess the most relevant contexts incorporating in the UX of SRSs. To this end, initially, the conceptual framework was proposed and then, it was empirically examined. To develop a conceptual framework, the following activities have been completed;

- ✓ Review of relevant existing models, frameworks, and theories
- ✓ Review of the literature to identify the components and relationships
- ✓ Review of the literature to identify the indicators
- ✓ Systematic review of the literature to identify the contextual information
- ✓ Validation of the proposed conceptual framework by experts' review

The main goal of the proposed framework is to present a set of structurally relevant contexts influencing UX of SRSs. The framework provides insights into development of SRSs both back- end (algorithms) and front- end (user interface) in order to enhance the UX of SRS. After the experts' revisions, four main components of context, perception, feeling and appraisal were confirmed along with the relationships and relevant indicators for each component were also confirmed. After theoretically justifying the relationships drawn between the various latent variables leading to the proposition of relevant hypotheses, the experiment of the conceptual framework and detection of most relevant contexts are performed using the quantitative method of Partial Least Squares (PLS) Regression and Structural Equation Modelling (SEM). Based on the PLS-SEM method, the following activities and relevant tests have been done to empirically validate the framework;

- ✓ Dataset preparation--- Face validity; Content validity; Spearman's rho tests
- ✓ Examination of identified indicators (assessment of measurement model in PLS-SEM terminology)--- VIF and OWs tests
- ✓ Examination of identified components (assessment of structural model in PLS-SEM terminology)--- R^2, F^2 tests
- ✓ Examination of relationships (assessment of hypotheses)—β tests
- ✓ Examination of Goodness of Fit
- ✓ Assessment of relevant context --- OWs tests

The empirical results reveled that all defined indicators are empirically valid, therefore no indicator was removed in the revised conceptual framework. Also, the R^2 and F^2 tests examined the validity of the constructs (components) and the results revealed that all predefined components empirically contributed into the construction of UX of SRSs. In addition, the results of path coefficients (β) and F^2 showed that all the relationships are valid in the conceptual framework, however, the moderating effect indicator (over time) does not surpass the minimum threshold of p-value ($p\text{-value}_{overtime} = 0.201$) which means that there is not statistically significant relationship between this construct and overall appraisal. Therefore, this relationship is removed in the conceptual framework. Based on the β tests results, some of the relationships are weak, but they remain in the framework as long as they are statistically significant. The R^2s of the dependent variables indicated that the variance of UX of SRSs is explained substantially only by the effect of User $_{context}$ variable. In addition, according to the outputs of OWs, among the user context, respectively profile (PR_Mean, 0.901), task (TA_Mean, 0.685), learning style (LS_Mean, 0.664), pre-knowledge (PK_Mean, 0.561), and information seeking behavior have obtained the highest weights which confirm that they have a significant contribution to the formation of user contexts for a SRS. Moreover, between time (TI_Mean, 0.430) and location (E.LO1, 0.052) for the environment context, time was more relevant for the construction of the environment context and demonstrates a sufficient level of validity. Although the weight of location was low, the value of 0.52 still surpasses the threshold. Therefore, it was not removed in the framework. The empirical results also attested that four indicators of novelty, diversity, popularity, and accuracy form the quality of a paper and novelty (NO_Mean, 0.574) and diversity (DI_Mean, 0.563) have more weights. Additionally, the results of β test demonstrated that among the contexts, the strongest relationships are between [(User$_{context}$ → Resource$_{quality}$), 0.790]. A user model should match the recommended papers for four identified levels of accurate, novel, popular, and diverse with users' information needs which change by the users' contexts. Thus, the relationships between the relevant contexts in these two variables are the source of inspiration for the construction of BN model.

Besides, the results of f^2 test showed that IxD and UiD have large effect on the UX of SRSs. Also, display (I. DI1, 0.396), consistency (I. CO1, 0.387) and gamification (I.GA1, 0.374) have the highest weights among the UiD's indicators and for the IxD construct, the preference elicitation, refinement, and privacy consideration have received the highest weights in contribution to the IxD adecuacy (formative latent variable). The IxD and UiD indicators, which have most contribution to the UX of SRSs, are the source of inspiration for the UI development in this research.

8.1.2 Research objective 2

The second objective entails developing a contextual Bayesian UM using the relevant contexts identified in objective 1 that can be embedded in the process of recommending to diagnose the accurate, novel, diverse, and popular papers by considering the users' contexts. BN method has the ability to deal with partial observations and uncertainty, which makes the model suitable for a context-aware SRS. Based on the BN modeling guidelines, three phases including dataset preparation, BN structure learning, and parameter learning have been relatively accomplished.

A web-based application was developed to gather the required data. The data acquisition procedure was composed of a large-scale questionnaire survey. The questionnaire's items were designed based on the comprehensive studies performed in objective 1. After the data collection procedure, the following activities have been performed in order to prepare the required dataset for the BN structure and parameter learning.

- ✓ Importing data to CSV file by LINQ to SQL query
- ✓ Data retrieving from WOS and importing to CSV
- ✓ Datasets combining
- ✓ Data cleaning

After dataset preparation, the Bayesian UM structure was built by applying a hybrid method, which applies both the knowledge engineer (expert elicitation), and automated learning estimated by a learning algorithm from data. The hybrid method leads to the better analytical and predicting ability of the BN model and makes the proposed UM more robust and reliable. For the consistency and clarity of the contributed variables in the Bayesian UM, the knowledge engineer with the help of an expert checked the variables (node) that could diagnose the users' needs of scholarly papers for levels of accurate, novel, diverse, and popular paper. In the clarity test, the whole identified variables were checked to have a clear operational meaning. The structure of BN model includes the relationships between the contextual indicators of user profile, task, learning style, pre-knowledge, and information seeking behavior and indicators of accurate, novel, diverse, and popular papers. In the BN structure, the focus was on capturing the expert understanding of the relationships between variables performed by pairwise elicitation method. The Grow-Shrink algorithm, a constrained- based algorithm, was implemented in this research to learn the BN structure by the data. Also, the Pearson's Linear Correlation (alpha threshold: 0.05) was used as the conditional independence test. In addition, the Pearson's Linear Correlation was utilized to investigate the dependence between multiple variables at the same time. After learning the BN topology or structure, and specification of the relationships between the connected nodes, the conditional probability distribution for each node or parameter learning was performed by using the Grow-Shrink algorithm. The Conditional Probabilities Tables (CPTs) entailed by the BN network for each node overcome the contexts' changes and uncertainty that might happen due to the users' contexts such as task, background knowledge, and research interests.

8.1.3 Research objective 3

The third objective is to design a UI called rScholar. The rScholar was designed mostly based on the most influencing indicators of UiD and IxD adequacy identified in objective 1 and the inputs and outputs of proposed Bayesian UM developed in objective 2. The UI development process consists of thesephases;

- ✓ User research
- ✓ Information architecture
- ✓ Interaction design
- ✓ Visual design

In the user research phase, the user requirements for designing an effective UI for SRSs with regard to the results of objective 1 and 2 of this research were analyzed. The inputs are users' contexts and preferences and the outputs are four papers in different levels of accurate, novel, diverse, and popular. In addition, the general UiD and IxD guidelines were reviewed. Then, the proper design elements were identified. In the second phase, a multilayer (n-tier) architecture was designed and applied for the development of rScholar. The separation of concerns among components is one of the distinctive features of this architecture, which means that components within a specific layer deal only with the logic that pertains to that layer. Three layers of data acquisition, reasoning, and UI application were developed. This phase also involved finding the technical frameworks and tools, which are suitable for the UI development considering the requirements and system architecture.

The third phase of UI development was interaction design. Based on the results of objective 1, preference elicitation, preference refinement, privacy consideration, explanation, and info-sufficiency are indicators of IxD adequacy and contributed mostly into the UX of SRSs. Therefore, in rScholar development, different methods have been designed separately in order to meet the IxD adequacy. For example, for the preference elicitation; various options of paper citing, saving to the library, paper rating, user profile updating, full-text presenting and, paper gift giving were designed. Paper gift is an icon that enables the user to offer a paper to others as a gift.

In the visual phase, the recommendations display were designed which are often related to UiD indicators. Nine different screens (pages) have been designed regarding to the identified requirements in phase 1. The whole pages were designed based on the user persona solution, which enables the scholars to use the rScholar as an academic profile for presenting their skills, academic information, etc. Additionally, for presentation of the recommendations, two different layouts of pie and list have been designed.

rScholar is the first serious attempt for development of UI in the field of SRSs. In the design of rScholar, the identified requirements and existing UI/UX standards /guidelines, adapted for the SRSs, were considered. In addition, the design solutions were mainly created based on the researcher's job experience in system analysis and UI/ UX design.

Gamification and visualization have shown significant impact on the UX of SRSs, but they have taken less attention in the rScholar design since game development and data visualization for SRSs require separate extensive investigations in both design solutions and data analysis. Although it seems that SRSs do not require providing fun for the scholars interacting with them; interestingly, gamification has a significant effect on making the users' experiences better in the SRSs and perceiving fun. Hence, in this research gamification and visualization were briefly discussed in order to offer insights into UI design for SRSs in the future studies.

8.1.4 Research objective 4

The last objective is to conduct an experimental evaluation for the Bayesian UM and UI proposed for the scholarly recommending system. For this reason, the suitable methods and metrics were selected based on the results of meta-analysis carried out to identify the methods and metrics applying in the existing SRSs and past studies in the field of BN modeling.

For the Bayesian UM evaluation, the offline method was conducted to measure the robustness (sensitivity analysis) and performance of the UM. In addition, the performance of the applied algorithms was examined by the measure of expected loss evaluation. Besides, the task of dataset split into test and train sets have been performed using the KFCV method. The following metrics have been applied in order to examine the robustness and performance of the UM.

- ✓ Robustness of BN structure --- Sensitivity analysis --- Entropy metric
- ✓ Algorithm performance--- Expected loss
- ✓ Performance of BN model--- F1-score, MSE and, MXE metrics

In the entropy assessment of the BN structure, only three variables' values did not pass the threshold of minimum uncertainty (Pr $(\mathrm{X} = \text{ Positive}|E)$ >0.5). All the entropy values of parameter passed the minimum uncertainty. Also, the results of expected loss have shown an acceptable performance for the Grow shrink algorithm applied for the BN model learning. The results of predictive performance metrics including MXE→Mean Cross Entropy ([0, ∞] Best score = 0), MSE→Mean Square ([0, 1] Best score = 0), Error F1→ F1Score ([0, 1] Best score = 1) also revealed that the proposed Bayesian UM achieved acceptable predict performance rates.

The evaluation of the UI was carried out by two methods of user studies and expert evaluation. Two groups including nine experts and twenty users (master, PhD students, post-doc researchers, and faculty members) have separately participated in the UI experimental evaluation. After collecting data, the data normality tests were performed and six suitable statistical tests were selected in order to analyze the data based on the defined goals. The following tests:

- ✓ UiD and IxD differences in rScholar & google Scholar experts' feedbacks --- T-test, KW & MW test
- ✓ Experts' evaluation of design ideas in rScholar--- T-test, KW & MW test
- ✓ Differences in users' ratings of rScholar after three months --- Wilcoxon signed-rank test
- ✓ Differences in users' groups in evaluation of rScholar --- ANOVA & KCC

The results of T-test, KW & MW tests showed that there is not statistical significant difference (sig. = .622 >.05) between these IxD and UiD features in rScholar & Googlscholar. However, by comparison of the features' means, it is concluded that the experts have indicated more rates to the Google scholar's consistency than rScholar and expressed more rates to rScholar recommendation display than Googlescholar. In addition, the evaluation of the design ideas including paper gift, logo, pie design, list design, logging page and background as well as user persona showed that the pie layout design in rScholar is fairly appealing and extremely novel, but not simple compared to the list design from the experts' viewpoints. In addition, amongst all design ideas applied in rScholar, the experts have given the lowest rate to the logo and highest rate (Mean rate$\sim$5) to the paper gift icon. Besides, the idea of user persona was rated simple and attracting. Finally and interestingly enough, the idea of logging page (background) design has not been perceived simple for the experts. The median of differences between Pre-test and Post-test was equal to zero (0) which means there is no significant differences between the users' ratings after three months and proves that the users' rates are reliable. Thus, the rScholar (UiD & IxD) features have same effects on the different user groups. Besides, there was not significant difference in overall evaluation of users' groups. However, between the UiD and IxD features, the UiD features including consistency, signifier and recommendation display presented lower scores among the users.

8.2 Future work

During the course of this research, several potentially interesting and relevant subjects presented, but to keep focused on the objectives of this study, the topics were abandoned. In the following sections, some of the works that may provide steps forward to extend this research are presented.

8.2.1 Future work related to Objective 1

The following research gaps are related to the contextual information influencing UX of SRSs.

8.2.1.1 Exploiting more contextual information in recommending

Among the identified contextual information, this research has not incorporated contexts such as users' reasoning method, mood, academic social network as well as personality traits. Therefore, future studies are encouraged to consider the above-mentioned contexts in developing systems for recommending papers.

8.2.1.2 Extension of the proposed framework

The framework proposed in this research is a specialised framework as it allows SRSs designers and researchers to target contexts associated with users' experience while they are working with the SRSs. The details of the framework components can further be enriched so that it can be utilised and evaluated for various domains. The details may include investigation of all the users' perceptions related to the identified contexts. For example, how contexts can be influential in providing fun for the users of SRSs.

8.2.1.3 Identification of users' needs in long term

Meeting of the user's information need is the main contribution of a good scholarly recommender. Users have different information needs due to different knowledge, preferences, goals, and contexts. In this research, the user model aimed to diagnose the users' information needs, however, the problem is that the identification of users' information is not an easy task and needs better understanding of the users' information needs not only when they are working with the system but also before and after that. Therefore, it requires monitoring the users' needs in a long term in order to provide good service and recommend adequate items.

8.2.2 Future work related to Objective 2

The following research gaps are related to the UX evaluation of SRSs.

8.2.2.1 Using the user model for existing recommenders

As a future work, another goal would be to deploy the BN model in a real environment and evaluate acceptance and feedback reported by the end users.

8.2.2.2 Using the user model for existing methods' optimization

The contextual user experience model using Bayesian networks is able to diagnose the users' information needs in four levels of accurate, diverse, novel, and popular papers. Identification of the user information needs is very vital in retrieving the best papers for them; therefore, embedding the proposed model in the scholarly recommending systems can be fruitful in recommending process to analyze the users' needs in long-term intervals.

8.2.2.3 Exploiting identified contexts in other methods

This research applied the Bayesian network, however, exploiting the identified contextual information, especially user and system contexts in other ML methods, which are appropriate for the decision-making such as Random forests, can be interesting topics for the future work.

8.2.3 Future work related to Objective 3

The following research gaps are related to the UI design of SRSs.

8.2.3.1 Development of new UIs

Both industry practitioners and academic researchers have argued that the interface of a RS may have far larger effects on users' experience with the recommender than the recommender's algorithmic performance (McNee et al., 2006; Baudisch & Terveen, 1999; Murray & Haubl, 2008; Xiao & Benbasat, 2007; Ziegler et al., 2005; Ozok et al., 2010). Therefore, designing new UIs and comparing them with the current study might be an amazing research topic for the future studies.The interested researchers might design new UIs for SRSs considering the UiD and IxD adequacy features assessed in this research.

8.2.3.2 Gamification in SRS

Gamification is one of the novel and open research topics in the field of RSs and particularly SRSs. The term gamification consists of the use of game design elements in a non-game context in order to motivate and involve users in an activity, environment or any tasks for a long time that requires the users engagement such as item ratings in RSs (de CA Ziesemer et al., 2014; Feil et al., 2016). Gamification in SRs rewards users with the true recommendations generated by the results of game feedbacks. Moreover, at the same time, it makes fun moments for the users (Hussain et al., 2014). In RSs, games are often being used in e-commerce in order to predict users' preferences and suggest them new products(Azam & Yao, 2014). Though the academic and scholarly tasks often pertain to some serious activities without considering fun, this research has investigated the role of using games in SRSs. Gamification and visualization have not been designed seriously in this research, however, they are open to the future researchers.

8.2.3.3 Data visualization and dashboard design for SRS

Visualization leverages visual and graphical representations to facilitate user's perception(Hiesel et al., 2016) and helps to make more transparency and interactivity in RSs. Also, context visualization makes more apparent to the users how the recommender works and reduces the cognitive effort which is required to uptake the meaning of the recommendation when it is presented (di Sciascio, 2017). Visualizations are only produced from data, but dashboards are regularly updated based on the new data. Like the dashboard on a car, RS dashboards typically integrate recommendations of different types or from different sources allowing the user to glance at the recommendations frequently and with low commitment(Murphy-Hill & Murphy, 2014). In the design of rScholar, not any specific game and visualtion method have been used since they have not been the objective of this study. In addition, adequate design of a game and recommendation visualsation are not easy tasks and require sperate studies; however, these great topics are presented to the interested researchers for more investigation.

8.2.3.4 Adaptive dialog UI for SRS

The aim of dialogue design is to yield closure and to reduce cognitive load. The dialogue makes the system as fool-proof as possible by for example using messaging box, flags, and icons(Shneiderman, 2004). Moreover, the medium that presents the recommendations also influences the user's satisfaction of RS (Beel, Genzmehr, et al., 2013). Designing adaptive dialog UIs for SRS is also an open research area.

8.2.4 Future work related to Objective 4

The following research gaps are related to the evaluation of SRSs.

8.2.4.1 UX methods & metrics

UX of RSs means the delivery of the recommendations to the user and the interaction of the user with those recommendations. UX evaluations have only started recently and are still in their infancy (Pu et al., 2012; Konstan & Riedl, 2012). Additionaly, there are not many certain metrics for measuring the UX of RSs; hence, working on the UX methods and metrics is very valuable and considered as an existing issue in the field of RSs.

8.2.4.2 Online evaluation metrics

Researchers seldom apply online tests to evaluate their proposed algorithms. The main reason is that online evaluations need a fully implemented system and a community of end-users(de Wit, 2008); therefore, they are very costly, time-consuming and entail many efforts. Despite the difficulties with online evaluation, as mentioned before, online method is the only way that can measure users' experience truly where the users are in a real interaction with the system and can leave real feelings and feedback (Cremonesi, Garzotto, & Turrin, 2013; Bart P. Knijnenburg et al., 2012). Thus, working on designing suitable and reproducible metrics is a research gap for the interested investigators.

8.2.5 General future work

The scope of scholarly recommending systems is broad. One of their applications is in paper recommending, which is commonly known as research paper or academic RSs. However, they can be applied as an appropriate tool to facilitate and accelerate the process of doing scholarly tasks such as finding appropriate conferences, research collaborators, journals to help scholars when they are going-over huge amount of relevant and irrelevant data (Champiri et al., 2015). Most of the current SRSs studies have been done to present research papers, there are a few studies on conference recommendations; however, offering appropriate research collaborators or journals to scholars still needs more investigation.

REFERENCES

Abdrabo, W., & Wörndl, W. (2016). *DiRec: A Distributed User Interface Video Recommender.* Paper presented at the IntRS@ RecSys.

Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. (1999). *Towards a better understanding of context and context-awareness.* Paper presented at the Handheld and ubiquitous computing.

Ackerman, M., & Mainwaring, S. (2005). Privacy issues in human-computer interaction. *Computer, 27*(5), 19-26.

Ackerman, M. S., & Mainwaring, S. D. (2005). Privacy issues and human-computer interaction. *Comput er, 27*(5), 19-26.

Adamopoulos, P., & Tuzhilin, A. (2011). *On unexpectedness in recommender systems: Or how to expect the unexpected.* Paper presented at the Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011), at the 5th ACM International Conference on Recommender Systems (RecSys' 11).

Adamopoulos, P., & Tuzhilin, A. (2015). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST), 5*(4), 54.

Adomavicius, G., & Jannach, D. (2013). Preface to the special issue on context-aware recommender systems. *User Modeling and User-Adapted Interaction*, 1-5.

Adomavicius, G., & Jannach, D. (2014). Preface to the special issue on context-aware recommender systems. *User Model. User-Adapt. Interact., 24*(1-2), 1-5.

Adomavicius, G., & Kwon, Y. (2011). *Maximizing aggregate recommendation diversity: A graph-theoretic approach.* Paper presented at the Proc. of the 1st International Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011).

Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems (TOIS), 23*(1), 103-145.

Adomavicius, G., & Tuzhilin, A. (2005a). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering, 17*(6), 734-749.

Adomavicius, G., & Tuzhilin, A. (2005b). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on, 17*(6), 734-749.

Adomavicius, G., & Tuzhilin, A. (2011a). Context-aware recommender systems *Recommender systems handbook* (pp. 217-253): Springer.

Adomavicius, G., & Tuzhilin, A. (2011b). Context-Aware Recommender Systems. 217-253. doi: 10.1007/978-0-387-85820-3_7

Aggarwal, C. C. (2016). Evaluating Recommender Systems *Recommender Systems* (pp. 225-254): Springer.

Aittola, M., Ryhänen, T., & Ojala, T. (2003). SmartLibrary–location-aware mobile library service *Human-computer interaction with mobile devices and services* (pp. 411-416): Springer.

Al-Hamad, A., Yaacob, N., & Al-Zoubi, A. (2008). Integrating 'learning style'information into personalized e-learning system. *IEEE Multidisciplinary Engineering Education Magazine, 3*(1), 2-6.

Albert, J. (2009). *Bayesian computation with R*: Springer Science & Business Media.

Ali, R. V. (2014). Content-based recommender system for an academic social network/Vala Ali Rohani,*University of Malaya*

Amini, B., Ibrahim, R., Othman, M. S., & Rastegari, H. (2011). *Incorporating scholar's background knowledge into recommender system for digital libraries.* Paper presented at the Software Engineering (MySEC), 2011 5th Malaysian Conference in.

Amirkhani, H., Rahmati, M., Lucas, P., & Hommersom, A. (2016). Exploiting Experts' Knowledge for Structure Learning of Bayesian Networks. *IEEE transactions on pattern analysis and machine intelligence*.

Anderson, J. C., & Narus, J. A. (1990). A model of distributor firm and manufacturer firm working partnerships. *the Journal of Marketing*, 42-58.

Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners* (Vol. 30): Springer Science & Business Media.

Asabere, N. Y. (2013). Towards a Viewpoint of Context-Aware Recommender Systems (CARS) and Services. *International Journal of Computer Science and Telecommunications, 4(1), 10–29.*

Azam, N., & Yao, J. (2014). Game-theoretic rough sets for recommender systems. *Knowledge-Based Systems, 72*, 96-107.

Balabanović, M., & Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM, 40*(3), 66-72.

Baltrunas, L. (2008). *Exploiting contextual information in recommender systems.* Paper presented at the Proceedings of the 2008 ACM conference on Recommender systems.

Baltrunas, L., Ludwig, B., Peer, S., & Ricci, F. (2012). Context relevance assessment and exploitation in mobile recommender systems. *Personal and Ubiquitous Computing, 16*(5), 507-526.

Baltrunas, L., Ludwig, B., & Ricci, F. (2011). *Matrix factorization techniques for context aware recommendation.* Paper presented at the Proceedings of the fifth ACM conference on Recommender systems.

Baltrunas, L., & Ricci, F. (2009). *Context-based splitting of item ratings in collaborative filtering.* Paper presented at the Proceedings of the third ACM conference on Recommender systems.

Barbara Kitchenham, u. S. C. K. U. a. D. U. J. R. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering.

Basu, C., Cohen, W. W., Hirsh, H., & Nevill-Manning, C. (2011). Technical paper recommendation: A study in combining multiple information sources. *Journal of Artificial Intelligence Research, 14*, 241-262.

Bauernfeind, U., & Zins, A. H. (2005). The perception of exploratory browsing and trust with recommender websites. *Information Technology & Tourism, 8*(2), 121-136.

Bazire, M., & Brézillon, P. (2005). Understanding context before using it *Modeling and using context* (pp. 29-40): Springer.

Beel, J., Breitinger, C., Langer, S., Lommatzsch, A., & Gipp, B. (2016). Towards reproducibility in recommender-systems research. *User Modeling and User-Adapted Interaction, 26*(1), 69-101.

Beel, J., & Dinesh, S. (2017). Real-World Recommender Systems for Academia: The Pain and Gain in Building, Operating, and Researching them [Long Version]. *arXiv preprint arXiv:1704.00156*.

Beel, J., Genzmehr, M., Langer, S., Nürnberger, A., & Gipp, B. (2013). *A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation.* Paper presented at the Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation.

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). paper recommender systems: a literature survey. *International Journal on Digital Libraries, 17*(4), 305-338.

Beel, J., Gipp, B., Langer, S., & Genzmehr, M. (2011). *Docear: An academic literature suite for searching, organizing and creating academic literature.* Paper presented at the Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries.

Beel, J., & Langer, S. (2014). A Comparison of Offline Evaluations, Online Evaluations, and User Studies in the Context of Research Paper Recommender Systems. *Under Review. Pre-print available at http://www. docear. org/publications*.

Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breitinger, C., & Nürnberger, A. (2013a). Research paper recommender system evaluation. 15-22. doi: 10.1145/2532508.2532512

Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breitinger, C., & Nürnberger, A. (2013b). *Research paper recommender system evaluation: a quantitative literature survey.* Paper presented at the Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation.

Beel, J., Langer, S., Nürnberger, A., & Genzmehr, M. (2013). The impact of demographics (age and gender) and other user-characteristics on evaluating recommender systems *Research and Advanced Technology for Digital Libraries* (pp. 396-400): Springer.

Berkovsky, S., Kuflik, T., & Ricci, F. (2008). Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction, 18*(3), 245-286.

Bernhaupt, R. (2010). *Evaluating User Experience in Games: Concepts and Methods*: Springer Publishing Company, Incorporated.

Bethard, S., & Jurafsky, D. (2010). *Who should I cite: learning literature search models from citation behavior.* Paper presented at the Proceedings of the 19th ACM international conference on Information and knowledge management.

Biancalana, C., Gasparetti, F., Micarelli, A., & Sansonetti, G. (2013). An approach to social recommendation for context-aware mobile services. *ACM Transactions on Intelligent Systems and Technology, 4*(1), 1-31. doi: 10.1145/2414425.2414435

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*: " O'Reilly Media, Inc.".

Bitner, M. J. (1995). Building service relationships: it's all about promises. *Journal of the Academy of marketing science, 23*(4), 246-251.

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*.

Bollen, J., & Van de Sompel, H. (2006). *An architecture for the aggregation and analysis of scholarly usage data.* Paper presented at the Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries.

Bolstad, W. M., & Curran, J. M. (2016). *Introduction to Bayesian statistics*: John Wiley & Sons.

Brown, P. J., Bovey, J. D., & Chen, X. (1997). Context-aware applications: from the laboratory to the marketplace. *Personal Communications, IEEE, 4*(5), 58-64.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction, 12*(4), 331-370.

Burke, R. D., Hammond, K. J., & Young, B. C. (1996). *Knowledge-based navigation of complex information spaces.* Paper presented at the Proceedings of the national conference on artificial intelligence.

Calero Valdez, A., Ziefle, M., & Verbert, K. (2016). *HCI for recommender systems: the past, the present and the future.* Paper presented at the Proceedings of the 10th ACM Conference on Recommender Systems.

Campos, P. G., Fernández-Tobías, I., Cantador, I., & Díez, F. (2013). Context-Aware Movie Recommendations: An Empirical Comparison of Pre-filtering, Post-filtering and Contextual Modeling Approaches *E-Commerce and Web Technologies* (pp. 137-149): Springer.

Champiri, Z. D., Shahamiri, S. R., & Salim, S. S. B. (2015). A systematic review of scholar context-aware recommender systems. *Expert Systems with Applications, 42*(3), 1743-1758.

Chandrasekaran, K., Gauch, S., Lakkaraju, P., & Luong, H. P. (2008). *Concept-based document recommendations for citeseer authors.* Paper presented at the Adaptive Hypermedia and Adaptive Web-Based Systems.

Chen, C.-M., & Yang, Y.-C. (2010). An intelligent mobile location-aware book recommendation system with map-based guidance that enhances problem-based learning in libraries *Advances in Neural Network Research and Applications* (pp. 853-860): Springer.

Chen, L., & Tsoi, H. K. (2011). *Users' decision behavior in recommender interfaces: Impact of layout design.* Paper presented at the RecSys' 11 Workshop on Human Decision Making in Recommender Systems.

Chen, R.-S., Tsai, Y.-S., Yeh, K., Yu, D., & Bak-Sau, Y. (2008). Using data mining to provide recommendation service. *WSEAS Transactions on Information Science and Applications, 5*(4), 459-474.

Chen, Y. (2011). *Interface and interaction design for group and social recommender systems.* Paper presented at the Proceedings of the fifth ACM conference on Recommender systems.

Chen, Y., & Pu, P. (2011). *Do You Feel How I Feel? An Affective Interface in Social Group Recommender Systems.* Paper presented at the Proceedings of the 3rd Workshop on Recommender Systems and the Social Web.

Cheng Li , Q. D., Zhen-Hua; Li, Tuo. (2008). Research of Information Recommendation system based on Reading Behaviour. *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming,.*

Codina Busquet, V., & Ceccaroni, L. (2014). Exploiting distributional semantics for content-based and context-aware recommendation.

Cole, R., Purao, S., Rossi, M., & Sein, M. (2005). Being proactive: where action research meets design research. *ICIS 2005 Proceedings*, 27.

Collis, J., & Hussey, R. (2003). Business Research (ed.): New York: Palgrave Macmillan.

Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). *About face: the essentials of interaction design*: John Wiley & Sons.

Councill, I. G., Giles, C. L., & Kan, M.-Y. (2008). *ParsCit: an Open-source CRF Reference String Parsing Package.* Paper presented at the LREC.

Craig, E. (1998). *Routledge Encyclopedia of Philosophy: questions to sociobioligy* (Vol. 8): Taylor & Francis.

Cremonesi, P., Elahi, M., & Garzotto, F. (2017). User interface patterns in recommendation-empowered content intensive multimedia applications. *Multimedia Tools Appl., 76*(4), 5275-5309. doi: 10.1007/s11042-016-3946-5

Cremonesi, P., Garzotto, F., & Turrin, R. (2013). User-centric vs. System-centric Evaluation of Recommender Systems *Human-Computer Interaction–INTERACT 2013* (pp. 334-351): Springer.

Crowther, J. (1995). Oxford Advanced Learner's Dictionary. *Oxford Advanced Learner's Dictionary, 5th edn, Oxford University Press, Oxford, UK.*

Cyr, D., Head, M., & Ivanov, A. (2006). Design aesthetics leading to m-loyalty in mobile commerce. *Information & Management, 43*(8), 950-963.

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*: Cambridge University Press.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science, 35*(8), 982-1003.

de CA Ziesemer, A., Müller, L., & Silveira, M. S. (2014). *Just rate it! gamification as part of recommendation.* Paper presented at the International Conference on Human-Computer Interaction.

De Giusti, M. R., Villarreal, G. L., Vosou, A., & Martínez, J. P. (2010). An ontology-based context aware system for selective dissemination of information in a digital library. *arXiv preprint arXiv:1005.4008*.

De Nart, D., Ferrara, F., & Tasso, C. (2013). Personalized access to scientific publications: from recommendation to explanation *User Modeling, Adaptation, and Personalization* (pp. 296-301): Springer.

de Wit, J. (2008). Evaluating recommender systems. *An Evaluation Framework to Predict User Satisfaction for Recommender Systems in an Electronic Program Guide Context*.

Deaton, M. (2003). The elements of user experience: user-centered design for the Web. *interactions, 10*(5), 49-51.

Dehghani, Z., Afshar, E., Jamali, H. R., & Nematbakhsh, M. A. (2011). A multi-layer contextual model for recommender systems in digital libraries. *Aslib Proceedings, 63*(6), 555-569

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.

Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing, 5*(1), 4-7.

di Sciascio, C. (2017). *Advanced User Interfaces and Hybrid Recommendations for Exploratory Search.* Paper presented at the Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion.

Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing, 8*(1), 19-30.

Dunn, R., Beaudry, J. S., & Klavas, A. (2002). Survey of research on learning styles. *California Journal of Science Education, 2*(2), 75-98.

Easterly, W., & Levine, R. (2001). What have we learned from a decade of empirical research on growth? It's Not Factor Accumulation: Stylized Facts and Growth Models. *the world bank economic review, 15*(2), 177-219.

Ekstrand, M. D. (2014). *Towards Recommender Engineering Tools and Experiments for Identifying Recommender Differences.* UNIVERSITY OF MINNESOTA.

Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. *Foundations and Trends® in Human–Computer Interaction, 4*(2), 81-173.

Ericson, K., & Pallickara, S. (2013). On the performance of high dimensional data clustering and classification algorithms. *Future Generation Computer Systems, 29*(4), 1024-1034.

Hair, J. F., Sarstedt, M., Hopkins, L., & G. Kuppelwieser, V. (2014). Partial least squares structural equation modeling (PLS-SEM) An emerging tool in business research. *European Business Review, 26*(2), 106-121.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). PLS applications in strategic management: Partial least squares modeling in strategy research. Long Range Planning, 46(1-2), 1-194.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). The use of partial least squares (PLS) to address marketing management topics. Journal of Marketing Theory and Practice, 19(2), 135-138.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). A primer on partial least squares structural equation modeling (PLS-SEM). Sage Publications.

Farooq, U., Ganoe, C. H., Carroll, J. M., Councill, I. G., & Giles, C. L. (2008). Design and evaluation of awareness mechanisms in CiteSeer. *Information processing & management, 44*(2), 596-612.

Feil, S., Kretzer, M., Werder, K., & Maedche, A. (2016). *Using gamification to tackle the cold-start problem in recommender systems.* Paper presented at the Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion.

Felden, C., & Chamoni, P. (2007). *Recommender systems based on an active data warehouse with text documents.* Paper presented at the System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on.

Felfernig, A., & Burke, R. (2008). *Constraint-based recommender systems: technologies and research issues.* Paper presented at the Proceedings of the 10th international conference on Electronic commerce.

Felfernig, A., Burke, R., & Pu, P. (2012). Preface to the special issue on user interfaces for recommender systems. *User Modeling and User-Adapted Interaction, 22*(4), 313-316.

Ferrara, F., Pudota, N., & Tasso, C. (2011). *A Keyphrase-Based Paper Recommender System.* Paper presented at the IRCDL.

Ferrier, L., Shane, H., Ballard, H., Carpenter, T., & Benoit, A. (1995). Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. *Augmentative and Alternative Communication, 11*(3), 165-175. doi: doi:10.1080/07434619512331277289

Figueroa, C., Vagliano, I., Rodríguez Rocha, O., & Morisio, M. (2015). A systematic literature review of Linked Data-based recommender systems. *Concurrency and Computation: Practice and Experience*.

Fischer, G. (2001). User modeling in human–computer interaction. *User Modeling and User-Adapted Interaction, 11*(1-2), 65-86.

Fishbein, M. leek Ajzen (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*, 181-202.

Flores, M. J., Nicholson, A. E., Brunskill, A., Korb, K. B., & Mascaro, S. (2011). Incorporating expert knowledge when learning Bayesian network structure: a medical case study. *Artificial intelligence in medicine, 53*(3), 181-204.

Franke, M., Geyer-Schulz, A., & Neumann, A. W. (2008). Recommender services in scientific digital libraries *Multimedia Services in Intelligent Environments* (pp. 377-417): Springer.

G.Marcot, B. (2012). Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling, 230*, 50-62.

Galitz, W. O. (1985). *Handbook of screen format design*: North-Holland Amsterdam.

Gantner, Z., Rendle, S., Freudenthaler, C., & Schmidt-Thieme, L. (2011). *MyMediaLite: A free recommender system library.* Paper presented at the Proceedings of the fifth ACM conference on Recommender systems.

Garrett, J. J. (2010). *Elements of user experience, the: user-centered design for the web and beyond*: Pearson Education.

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010a). *Beyond accuracy: evaluating recommender systems by coverage and serendipity.* Paper presented at the Proceedings of the fourth ACM conference on Recommender systems.

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010b). User-perceived recommendation quality-factoring in the user interface.

Geisler, G., McArthur, D., & Giersch, S. (2001). *Developing recommendation services for a digital library with uncertain and changing data.* Paper presented at the Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries.

Geisler, G., McArthur, D., and Giersch, S. (2001). Developing recommendation services for a digital library with uncertain and changing data. *In Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries (Roanoke, Virginia, United States,JCDL '01. ACM Press,New York, NY*, New York, NY.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol., 59*, 167-192.

Geyer-Schulz, A., Neumann, A., & Thede, A. (2003a). An architecture for behavior-based library recommender systems. *Information Technology and Libraries, 22*(4), 165-174.

Geyer-Schulz, A., Neumann, A., & Thede, A. (2003b). Others also use: A robust recommender system for scientific libraries *Research and Advanced Technology for Digital Libraries* (pp. 113-125): Springer.

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). *CiteSeer: An automatic citation indexing system.* Paper presented at the Proceedings of the third ACM conference on Digital libraries.

Gipp, B., Beel, J., & Hentschel, C. (2009a). *Scienstein: A research paper recommender system.* Paper presented at the International Conference on Emerging Trends in Computing.

Gipp, B., Beel, J., & Hentschel, C. (2009b). *Scienstein: A research paper recommender system.* Paper presented at the Proceedings of the international conference on emerging trends in computing (ICETiC'09).

Gómez, S., Zervas, P., Sampson, D. G., & Fabregat, R. (2014). Context-aware adaptive and personalized mobile learning delivery supported by UoLmP. *Journal of King Saud*

University - Computer and Information Sciences, 26(1), 47-61. doi: 10.1016/j.jksuci.2013.10.008

Gorodetsky, V., Samoylov, V., & Serebryakov, S. (2010). *Ontology–based context–dependent personalization technology.* Paper presented at the Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on.

Gosling, S. (2009). *Snoop: What your stuff says about you*: Hachette UK.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality, 37*(6), 504-528.

Green, W. S., & Jordan, P. W. (2003). *Pleasure with products: Beyond usability*: CRC Press.

Gunawardana, A., & Shani, G. (2009). A survey of accuracy evaluation metrics of recommendation tasks. *The Journal of Machine Learning Research, 10*, 2935-2962.

Guo, S. (2011). *Bayesian Recommender Systems: Models and Algorithms*: Australian National University.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res., 3*, 1157-1182.

Hahn, J. (2011). Location-based recommendation services in library book stacks. *Reference Services Review, 39*(4), 654-674.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). The use of partial least squares (PLS) to address marketing management topics. *Journal of Marketing Theory and Practice, 19*(2), 135-138.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Editorial-partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long Range Planning, 46*(1-2), 1-12.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2013). *A primer on partial least squares structural equation modeling (PLS-SEM)*: Sage Publications.

Hanson, E. M. (2014). A Beginner's Guide to Creating Library Linked Data: Lessons from NCSU's Organization Name Linked Data Project. *Serials Review, 40*(4), 251-258.

Hariri, N., Mobasher, B., & Burke, R. (2014). *Context adaptation in interactive recommender systems.* Paper presented at the Proceedings of the 8th ACM Conference on Recommender systems.

Harman, H. H. (1960). Modern factor analysis.

Harper, F. M., Li, X., Chen, Y., & Konstan, J. A. (2005). An economic model of user rating in an online recommender system. *Lecture notes in computer science, 3538*, 307.

Hart, J. (2015). Investigating User Experience and User Engagement for Design.

Hartmann, J., Sutcliffe, A., & Angeli, A. D. (2008). Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction (TOCHI), 15*(4), 15.

Haseloff, S. (2005). *Context awareness in information logistics.* Universitätsbibliothek.

Hassan, H. A. M. (2017). *Personalized Research Paper Recommendation using Deep Learning.* Paper presented at the Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization.

Hassenzahl, M., Eckoldt, K., Diefenbach, S., Laschke, M., Len, E., & Kim, J. (2013). Designing moments of meaning and pleasure. Experience design and happiness. *International Journal of Design, 7*(3).

Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & Information Technology, 25*(2), 91-97.

Hayes, C., & Cunningham, P. (2002). An on-line evaluation framework for recommender systems: Trinity College Dublin, Department of Computer Science.

He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, L. (2010). *Context-aware citation recommendation.* Paper presented at the Proceedings of the 19th international conference on World wide web.

Heckerman, D., Mamdani, A., & Wellman, M. P. (1995). Real-world applications of Bayesian networks. *Communications of the ACM, 38*(3), 24-26.

Hekkert, P. (2006). Design aesthetics: principles of pleasure in design. *Psychology science, 48*(2), 157.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., . . . Calantone, R. J. (2014). Common beliefs and reality about PLS comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, 1094428114526928.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing (AIM), 20*, 277-320.

Herlocker, J., Jung, S., & Webster, J. G. (2012). Collaborative filtering for digital libraries.

Herlocker, J. L., & Konstan, J. (2001). Content-independent task-focused recommendation. *Internet Computing, IEEE, 5*(6), 40-47.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5-53.

Herrera-Viedma, E., Porcel, C., López-Herrera, A. G., & Alonso, S. (2008). A fuzzy linguistic recommender system to advice research resources in university digital libraries *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models* (pp. 567-585): Springer.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems, 19*(2), 4.

Hiesel, P., Wörndl, W., Braunhofer, M., & Herzog, D. (2016). A User Interface Concept for Context-Aware Recommender Systems. *Mensch und Computer 2016-Tagungsband*.

Hong, J.-y., Suh, E.-h., & Kim, S.-J. (2009). Context-aware systems: A literature review and classification. *Expert Systems with Applications, 36*(4), 8509-8522. doi: 10.1016/j.eswa.2008.10.071

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Practical assessment, research & evaluation, 12*(10), 1-8.

Hu, R. (2010). *Design and user issues in personality-based recommender systems.* Paper presented at the Proceedings of the fourth ACM conference on Recommender systems.

Huang, A. (2008). *Similarity measures for text document clustering.* Paper presented at the Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal, 20*(2), 195-204.

Hurley, N. J. (2011). *Robustness of recommender systems.* Paper presented at the Proceedings of the fifth ACM conference on Recommender systems.

Hussain, J., Khan, W. A., Afzal, M., Hussain, M., Kang, B. H., & Lee, S. (2014). *Adaptive user interface and user experience based authoring tool for recommendation systems.* Paper presented at the International Conference on Ubiquitous Computing and Ambient Intelligence.

Hwang, S.-Y., Hsiung, W.-C., & Yang, W.-S. (2003). A prototype WWW literature recommendation system for digital libraries. *Online Information Review, 27*(3), 169-182.

Hwang, S.-Y., Wei, C.-P., & Liao, Y.-F. (2010). Coauthorship networks and academic literature recommendation. *Electronic Commerce Research and Applications, 9*(4), 323-334.

Hwang, S.-Y. H., Wen-Chiang; Yang Wan-Shiou. (2003). A Prototype WWW Literature Recommendation System for Digital Libraries. *Online Information Review, 27, No. 3*, 169–182.

Jahanian, A., Liu, J., Lin, Q., Tretter, D., O'Brien-Strain, E., Lee, S. C., . . . Allebach, J. (2013). *Recommendation system for automatic design of magazine covers.* Paper presented at the Proceedings of the 2013 international conference on Intelligent user interfaces.

Jamali, H. R., Dehghani, Z., Afshar, E., Jamali, H. R., & Nematbakhsh, M. A. (2011). *A multi-layer contextual model for recommender systems in digital libraries.* Paper presented at the Aslib Proceedings.

Jannach, D., Lerche, L., Gedikli, F., & Bonnin, G. (2013). What recommenders recommend–an analysis of accuracy, popularity, and sales diversity effects *User Modeling, Adaptation, and Personalization* (pp. 25-37): Springer.

Jannach, D., Nunes, I., & Jugovac, M. (2017). *Interacting with Recommender Systems.* Paper presented at the Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion.

Joonseok Lee, K. L., and Jennifer G Kim. . (2013). Personalized academic research paper recommendation system. . *arXiv preprint arXiv:1304.5457*.

Jordan, P. W. (1998). Human factors for pleasure in product use. *Applied ergonomics, 29*(1), 25-33.

Jung, S., Harris, K., Webster, J., & Herlocker, J. L. (2004). *SERF: integrating human recommendations with search.* Paper presented at the Proceedings of the thirteenth ACM international conference on Information and knowledge management.

Kadie, C. M., Hovel, D., & Horvitz, E. (2001). MSBNx: A component-centric toolkit for modeling and inference with Bayesian networks. *Microsoft Research, Richmond, WA, Technical Report MSR-TR-2001-67, 28*.

Kaminskas, M., & Bridge, D. (2014a). *Measuring surprise in recommender systems.* Paper presented at the Proceedings of the Workshop on Recommender Systems Evaluation: Dimensions and Design (Workshop Programme of the 8th ACM Conference on Recommender Systems).

Kaminskas, M., & Bridge, D. (2014b). *Measuring surprise in recommender systems/ serendipity.* Paper presented at the Proceedings of the Workshop on Recommender Systems Evaluation: Dimensions and Design (Workshop Programme of the 8th ACM Conference on Recommender Systems).

Kaminskas, M., & Ricci, F. (2011). Location-adapted music recommendation using tags *User Modeling, Adaption and Personalization* (pp. 183-194): Springer.

Kang, J., & Choi, J. (2011). *An ontology-based recommendation system using long-term and short-term preferences.* Paper presented at the Information Science and Applications (ICISA), 2011 International Conference on.

Kantor, P. B., Rokach, L., Ricci, F., & Shapira, B. (2011). *Recommender systems handbook*: Springer.

Kapoor, N., Chen, J., Butler, J. T., Fouty, G. C., Stemper, J. A., Riedl, J., & Konstan, J. A. (2007). *Techlens: a researcher's desktop.* Paper presented at the Proceedings of the 2007 ACM conference on Recommender systems.

Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010). *Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering.* Paper presented at the Proceedings of the fourth ACM conference on Recommender systems.

Karvonen, K. (2000). *The beauty of simplicity.* Paper presented at the Proceedings on the 2000 conference on Universal Usability.

Keiningham, T. L., Perkins-Munn, T., & Evans, H. (2003). The impact of customer satisfaction on share-of-wallet in a business-to-business environment. *Journal of Service Research, 6*(1), 37-50.

Kelly, D., & Fu, X. (2006). *Elicitation of term relevance feedback: an investigation of term source and context.* Paper presented at the Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval.

Kitchenham, B. A., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *EBSE Technical Report EBSE*, pp. 1-57.

Knijnenburg, B. P., & Willemsen, M. C. (2010). *The effect of preference elicitation methods on the user experience of a recommender system.* Paper presented at the CHI'10 Extended Abstracts on Human Factors in Computing Systems.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction, 22*(4-5), 441-504. doi: 10.1007/s11257-011-9118-4

Kobsa, A. (2001). Generic user modeling systems. *User Modeling and User-Adapted Interaction, 11*(1), 49-63.

Kocaballı, A. B., & Koçyiğit, A. (2007). Granular best match algorithm for context-aware computing systems. *Journal of Systems and Software, 80*(12), 2015-2024.

Konstan, J. A. (2004). Introduction to recommender systems: Algorithms and evaluation. *ACM Transactions on Information Systems (TOIS), 22*(1), 1-4.

Konstan, J. A., Kapoor, N., McNee, S. M., & Butler, J. T. (2005). *Techlens: Exploring the use of recommenders to support users of digital libraries.* Paper presented at the CNI fall task force meeting project briefing. Coalition for networked information. Phoenix, AZ.

Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction, 22*(1-2), 101-123. doi: 10.1007/s11257-011-9112-x

Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction, 22*(1), 101-123.

Kontio, J., Lehtola, L., & Bragge, J. (2004). *Using the focus group method in software engineering: obtaining practitioner and user experiences.* Paper presented at the Empirical Software Engineering, 2004. ISESE'04. Proceedings. 2004 International Symposium on.

Korb, K. B., & Nicholson, A. E. (2003). Bayesian artificial intelligence. . *Florida: Chapman & Hall/CRC.*

Kotkov, D., Veijalainen, J., & Wang, S. (2016). *Challenges of Serendipity in Recommender Systems.* Paper presented at the WEBIST 2016: Proceedings of the 12th International conference on web information systems and technologies. Volume 2, ISBN 978-989-758-186-1.

Kotkov, D., Wang, S., & Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems, 111*, 180-192.

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Data preprocessing for supervised leaning. *International Journal of Computer Science, 1*(2), 111-117.

Kraaykamp, G., & Van Eijck, K. (2005). Personality, media preferences, and cultural participation. *Personality and individual differences, 38*(7), 1675-1688.

Krafft, D. B., Birkland, A., & Cramer, E. J. (2008). *Ncore: architecture and implementation of a flexible, collaborative digital library.* Paper presented at the Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries.

Kraft, C. (2012). *User experience innovation: User centered design that works*: Apress.

Kuenzer, A., Schlick, C., Ohmann, F., Schmidt, L., & Luczak, H. (2001). *An empirical study of dynamic bayesian networks for user modeling.* Paper presented at the Proc. of the UM'2001 Workshop on Machine Learning for User Modeling.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers, 23*(5), 473-483.

Kuo, J.-J., & Zhang, Y.-J. (2012). A library recommender system using interest change over time and matrix clustering *The Outreach of Digital Libraries: A Globalized Resource Network* (pp. 259-268): Springer.

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009). *Understanding, scoping and defining user experience: a survey approach.* Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems.

Levy, M. (2013). *Offline evaluation of recommender systems: all pain and no gain?* Paper presented at the Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation.

Lewis, B. R., Templeton, G. F., & Byrd, T. A. (2005). A methodology for construct development in MIS research. *European Journal of Information Systems, 14*(4), 388-400.

Li, Q.-C., Dong, Z.-H., & Li, T. (2008). *Research of information recommendation system based on reading behavior.* Paper presented at the Machine Learning and Cybernetics, 2008 International Conference on.

Liang, Y., Li, Q., & Qian, T. (2011). Finding relevant papers based on citation relations. *Web-age information management*, 403-414.

Liao, I.-E., Hsu, W.-C., Cheng, M.-S., & Chen, L.-P. (2010a). A library recommender system based on a personal ontology model and collaborative filtering technique for English collections. *Electronic Library, The, 28*(3), 386-400.

Liao, I.-E., Hsu, W.-C., Cheng, M.-S., & Chen, L.-P. (2010b). A library recommender system based on a personal ontology model and collaborative filtering technique for English collections. *The Electronic Library, 28*(3), 386-400.

Liao, S.-C., Kao, K.-F., Liao, I.-E., Chen, H.-L., & Huang, S.-O. (2009). PORE: a personal ontology recommender system for digital libraries. *Electronic Library, The, 27*(3), 496-508.

Lieberman, H., & Selker, T. (2000). Out of context: Computer systems that adapt to, and learn from, context. *IBM Systems Journal, 39*(3.4), 617-632.

Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2013). Facing the cold start problem in recommender systems. *Expert Systems with Applications, 41*(4), 2065-2073.

Lilien, G. L., Kotler, P., & Moorthy, K. S. (1992). *Marketing models*: Prentice Hall.

Lim, B. Y. (2012). *Improving Understanding and Trust with Intelligibility in Context-Aware Applications.* Oregon State University.

Liu, L. (2013). *The implication of context and criteria information in recommender systems as applied to the service domain.* University of Manchester.

Liu, L., Lecue, F., Mehandjiev, N., & Xu, L. (2010). *Using context similarity for service recommendation.* Paper presented at the Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on.

Lopes, G. R., Souto, M. A. M., Wives, L. K., & de Oliveira, J. P. M. (2008). *A personalized recommender system for digital libraries.* Paper presented at the Proceedings of the 14th Brazilian Symposium on Multimedia and the Web.

Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends *Recommender systems handbook* (pp. 73-105): Springer.

Lucas, J. P., Segrera, S., & Moreno, M. N. (2012). Making use of associative classifiers in order to alleviate typical drawbacks in recommender systems. *Expert Systems with Applications, 39*(1), 1273-1283.

Luo, J., Dong, F., Cao, J., & Song, A. (2009). A context-aware personalized resource recommendation for pervasive learning. *Cluster Computing, 13*(2), 213-239. doi: 10.1007/s10586-009-0113-z

Luo, J., Dong, F., Cao, J., & Song, A. (2010). A context-aware personalized resource recommendation for pervasive learning. *Cluster Computing, 13*(2), 213-239.

Luo, Y., Le, J., & Chen, H. (2009). *A privacy-preserving book recommendation model based on multi-agent.* Paper presented at the Computer Science and Engineering, 2009. WCSE'09. Second International Workshop on.

Mahjoub, M. A., & Kalti, K. (2011). Software comparison dealing with bayesian networks *Advances in Neural Networks–ISNN 2011* (pp. 168-177): Springer.

Malone, T. W. (1981). What makes things fun to learn? A study of intrinsically motivating computer games. *Pipeline, 6*(2), 50.

Mangina, E., & Kilbride, J. (2008). Evaluation of keyphrase extraction algorithm and tiling process for a document/resource recommender within e-learning environments. *Computers & Education, 50*(3), 807-820.

Margaritis, D. (2003). Learning Bayesian network model structure from data: CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE.

Marko A. Rodriguez, D. W. A., Joshua Shinavier, Gary Ebersole. (2009). A Recommender System to Support the Scholarly Communication Process. *CoRR abs/0905.1594*

Marović, M., Mihoković, M., Mikša, M., Pribil, S., & Tus, A. (2011). *Automatic movie ratings prediction using machine learning.* Paper presented at the MIPRO, 2011 Proceedings of the 34th International Convention.

Martín, E., Haya, P. A., & Carro, R. M. (2013). *User Modeling and Adaptation for Daily Routines: Providing Assistance to People with Special Needs*: Springer Science & Business Media.

Martínez, J. P., Vosou, A., Villarreal, G., & De Giusti, M. R. (2010). An ontology-based context aware system for Selective Dissemination of Information in a digital library. *Journal of Computing, 2*.

McCarthy, J., & Wright, P. (2004). Technology as experience. *interactions, 11*(5), 42-43.

McCay-Peet, L., & Toms, E. (2011). Measuring the dimensions of serendipity in digital environments. *Information Research: An International Electronic Journal, 16*(3), n3.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology, 52*(1), 81.

Mcnee, S. M. (2006). *Meeting user information needs in recommender systems*: Proquest.

McNee, S. M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S. K., Rashid, A. M., . . . Riedl, J. (2002). *On the recommending of citations for research papers.* Paper presented at the Proceedings of the 2002 ACM conference on Computer supported cooperative work.

McNee, S. M., Kapoor, N., & Konstan, J. A. (2006). *Don't look stupid: avoiding pitfalls when recommending research papers.* Paper presented at the Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006a). *Being accurate is not enough: how accuracy metrics have hurt recommender systems.* Paper presented at the CHI'06 extended abstracts on Human factors in computing systems.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006b). *Making recommendations better: an analytic model for human-recommender interaction.* Paper presented at the CHI'06 extended abstracts on Human factors in computing systems.

Middleton, S. E., Alani, H., & De Roure, D. C. (2002). Exploiting synergy between ontologies and recommender systems. *arXiv preprint cs/0204012*.

Middleton, S. E., De Roure, D. C., & Shadbolt, N. R. (2002). Foxtrot recommender system: User profiling, ontologies and the World Wide Web. *Poster, Proc. 11th Int. World Wide Web Conference (WWW'2002), Hawaii, USA, May 2002.*

Middleton, S. E., Shadbolt, N. R., & De Roure, D. C. (2003). *Capturing interest through inference and visualization: Ontological user profiling in recommender systems.* Paper presented at the Proceedings of the 2nd international conference on Knowledge capture.

Middleton, S. E., Shadbolt, N. R., & De Roure, D. C. (2004). Ontological user profiling in recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 54-88.

Mikawa, M., Izumi, S., & Tanaka, K. (2011). *Book Recommendation Signage System Using Silhouette-Based Gait Classification.* Paper presented at the Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on.

Mobasher, B. (2012). Context Aware Recommendation. *School of Computing DePaul University, Chicago*.

Mobasher, B. (2013). Context-Aware User Modeling for Recommendation. *UMAP 2013, Rome, Italy*.

Mobasher, B., Dai, H., Luo, T., & Nakagawa, M. (2001). *Effective personalization based on association rule discovery from web usage data.* Paper presented at the Proceedings of the 3rd international workshop on Web information and data management.

Monk, A., Hassenzahl, M., Blythe, M., & Reed, D. (2002). *Funology: designing enjoyment.* Paper presented at the CHI'02 Extended Abstracts on Human Factors in Computing Systems.

Mönnich, M., & Spiering, M. (2008a). Adding value to the library catalog by implementing a recommendation system. *D-Lib Magazine, 14*(5), 4.

Mönnich, M., & Spiering, M. (2008b). Adding value to the library catalog by implementing a recommendation system. *D-Lib Magazine, 14*(5/6), 1082-9873.

Montaner Rigall, M. (2003). *Collaborative recommender agents based on case-based reasoning and trust*: Universitat de Girona.

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192-222.

Morales-del-Castillo, J. M., Peis, E., & Herrera-Viedma, E. (2009a). *A filtering and recommender system prototype for scholarly users of digital libraries*: Springer.

Morales-del-Castillo, J. M., Peis, E., & Herrera-Viedma, E. (2009b). A filtering and recommender system prototype for scholarly users of digital libraries *Visioning and Engineering the Knowledge Society. A Web Science Perspective* (pp. 108-117): Springer.

Mueller, J., & Thyagarajan, A. (2016). *Siamese Recurrent Architectures for Learning Sentence Similarity.* Paper presented at the AAAI.

Murakami, T., Mori, K., & Orihara, R. (2007). *Metrics for evaluating the serendipity of recommendation lists.* Paper presented at the Annual Conference of the Japanese Society for Artificial Intelligence.

Murphy-Hill, E., & Murphy, G. C. (2014). Recommendation delivery *Recommendation Systems in Software Engineering* (pp. 223-242): Springer.

Murthi, B., & Sarkar, S. (2003). The role of the management sciences in research on personalization. *Management science, 49*(10), 1344-1362.

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). Bayesian networks in R. *Springer, 122*, 125-127.

Nakagawa, A., & Ito, T. (2002). *An implementation of a knowledge recommendation system based on similarity among users' profiles.* Paper presented at the SICE 2002. Proceedings of the 41st SICE Annual Conference.

Nascimento, C., Laender, A. H., da Silva, A. S., & Gonçalves, M. A. (2011). *A source independent framework for research paper recommendation.* Paper presented at the Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries.

Neapolitan, R. E. (2004). *Learning bayesian networks* (Vol. 38): Pearson Prentice Hall Upper Saddle River, NJ.

Neves, A. R. d. M., Carvalho, Á. M. G., & Ralha, C. G. (2014). Agent-based architecture for context-aware and personalized event recommendation. *Expert Systems with Applications, 41*(2), 563-573. doi: 10.1016/j.eswa.2013.07.081

Nguyen, T. (2016). Enhancing User Experience With Recommender Systems Beyond Prediction Accuracies. *PhD Dissertation,The University of Minnesota*.

Nguyen, T. T., Kluver, D., Wang, T.-Y., Hui, P.-M., Ekstrand, M. D., Willemsen, M. C., & Riedl, J. (2013). *Rating support interfaces to improve user experience and recommender accuracy*. Paper presented at the Proceedings of the 7th ACM conference on Recommender systems, Hong Kong, China.

Nielsen, J. (1995). 10 usability heuristics for user interface design. *Fremont: Nielsen Norman Group.[Consult. 20 maio 2014]. Disponível na Internet*.

Nielsen, J. (1999). *Designing web usability: The practice of simplicity*: New Riders Publishing.

Norman, D. (2013a). The design of everyday things: Revised and expanded edition. *Basic Books*.

Norman, D. (2013b). *The design of everyday things: Revised and expanded edition*: Basic Books (AZ).

Norman, D. A. (1999). Affordance, conventions, and design. *interactions, 6*(3), 38-43.

Odic, A., Tkalcic, M., Tasic, J. F., & Košir, A. (2012). Relevant context in a movie recommender system: Users' opinion vs. statistical detection. *ACM RecSys, 12*.

Odić, A., Tkalčič, M., Tasič, J. F., & Košir, A. (2013). Predicting and detecting the relevant contextual information in a movie-recommender system. *Interacting with Computers*, iws003.

Ono, C., Kurokawa, M., Motomura, Y., & Asoh, H. (2007). A context-aware movie preference model using a Bayesian network for recommendation and promotion. *User Modeling 2007*, 247-257.

Ono, C., Takishima, Y., Motomura, Y., Asoh, H., Shinagawa, Y., Imai, M., & Anzai, Y. (2008). Context-aware users' preference models by integrating real and supposed situation data. *IEICE TRANSACTIONS on Information and Systems, 91*(11), 2552-2559.

Ozok, A. A., Fan, Q., & Norcio, A. F. (2010). Design guidelines for effective recommender system interfaces based on a usability criteria conceptual model: results from a college student population. *Behaviour & Information Technology, 29*(1), 57-83.

Pagonis, J., & Clark, A. F. (2010a). *Engene: A genetic algorithm classifier for content-based recommender systems that does not require continuous user feedback.* Paper presented at the 2010 UK Workshop on Computational Intelligence (UKCI).

Pagonis, J., & Clark, A. F. (2010b). *Engene: A genetic algorithm classifier for content-based recommender systems that does not require continuous user feedback.* Paper presented at the Computational Intelligence (UKCI), 2010 UK Workshop on.

Panniello, U., & Gorgoglione, M. (2011). *Context-Aware Recommender Systems: A Comparison Of Three Approaches.* Paper presented at the DART@ AI* IA.

Panniello, U., Gorgoglione, M., & Tuzhilin, A. (2015). In CARSWe Trust: How Context-Aware Recommendations Affect Customers' Trust And Other Business Performance Measures Of Recommender Systems.

Panniello, U., Tuzhilin, A., Gorgoglione, M., Palmisano, C., & Pedone, A. (2009). *Experimental comparison of pre-vs. post-filtering approaches in context-aware recommender systems.* Paper presented at the Proceedings of the third ACM conference on Recommender systems.

Papatheocharous, E., Belk, M., Germanakos, P., & Samaras, G. (2014). Towards Implicit User Modeling Based on Artificial Intelligence, Cognitive Styles and Web Interaction Data. *International Journal on Artificial Intelligence Tools, 23*(02).

Park, H.-S., Yoo, J.-O., & Cho, S.-B. (2006). *A context-aware music recommendation system using fuzzy bayesian networks with utility theory.* Paper presented at the International Conference on Fuzzy Systems and Knowledge Discovery.

Parra, D., & Sahebi, S. (2013). Recommender systems: Sources of knowledge and evaluation metrics *Advanced Techniques in Web Intelligence-2* (pp. 149-175): Springer.

Pascoe, J. (1998). *Adding generic contextual capabilities to wearable computers.* Paper presented at the Wearable Computers, 1998. Digest of Papers. Second International Symposium on.

Patton, R., Potok, T., & Worley, B. (2012). *Discovery & Refinement of Scientific Information via a Recommender System.* Paper presented at the INFOCOMP 2012, The Second International Conference on Advanced Communications and Computation.

Patton, R. M., Potok, T. E., & Worley, B. A. (2012). *Discovery & refinement of scientific information via a recommender system.* Paper presented at the The Second International Conference on Advanced Communications and Computation.

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems *The adaptive web* (pp. 325-341): Springer.

Pearl, J. (1985). *Bayesian networks: A model of self-activated memory for evidential reasoning.* Paper presented at the Proceedings of the 7th Conference of the Cognitive Science Society, 1985.

Pearl, J. (2009). *Causality*: Cambridge university press.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems, 24*(3), 45-77.

Pennock, D. M., Horvitz, E., Lawrence, S., & Giles, C. L. (2000). *Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach.* Paper presented at the Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence.

Pham, M. C., Cao, Y., Klamma, R., & Jarke, M. (2011). A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis. *J. UCS, 17*(4), 583-604.

Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.-P., & Jonker, C. M. (2012). Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. *User Modeling and User-Adapted Interaction, 22*(4-5), 357-397.

Popa, H.-E., Negru, V., Pop, D., & Muscalagiu, I. (2008). *DL-AgentRecom-A multi-agent based recommendation system for scientific documents.* Paper presented at the Symbolic and Numeric Algorithms for Scientific Computing, 2008. SYNASC'08. 10th International Symposium on.

Porcel, C., del Castillo, J. M., Cobo, M., RuÌz-RodrÌguez, A., & Herrera-Viedma, E. (2010). An improved recommender system to avoid the persistent information overload in a university digital library. *Control and Cybernetics, 39*(4), 899-924.

Porcel, C., & Herrera-Viedma, E. (2010). Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries. *Knowledge-Based Systems, 23*(1), 32-39.

Porcel, C., Herrera-Viedma, Enrique. (2010). Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries. *Knowledge-Based Systems, 23*(1), 32-39.

Porcel, C., López-Herrera, A. G., & Herrera-Viedma, E. (2009). A recommender system for research resources based on fuzzy linguistic modeling. *Expert Systems with Applications, 36*(3), 5173-5183.

Porcel, C., Moreno, J. M., & Herrera-Viedma, E. (2009). A multi-disciplinar recommender system to advice research resources in university digital libraries. *Expert Systems with Applications, 36*(10), 12520-12528.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130-137.

Portugal, I., Alencar, P., & Cowan, D. (2015). The use of machine learning algorithms in recommender systems: a systematic review. *arXiv preprint arXiv:1511.05263*.

Powell, M. J. D. (1981). *Approximation theory and methods*: Cambridge university press.

Powers, D. (2007). Evaluation: From precision, recall and F-factor to ROC, informedness, markedness & correlation (Tech. Rep.). *Adelaide, Australia*.

Pratley, A., van Voorthuysen, E., & Chan, R. (2014). A step-by-step approach for modelling complex systems with Partial Least Squares. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 0954405414534430.

Pu, P., Chen, L., & Hu, R. (2011). *A user-centric evaluation framework for recommender systems.* Paper presented at the Proceedings of the fifth ACM conference on Recommender systems.

Pu, P., Chen, L., & Hu, R. (2012a). Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction, 22*(4-5), 317-355.

Pu, P., Chen, L., & Hu, R. (2012b). Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction, 22*(4), 317-355.

Raamkumar, A. S., Foo, S., & Pang, N. (2016). A Framework for Scientific Paper Retrieval and Recommender Systems. *arXiv preprint arXiv:1609.01415*.

Rana, C. (2013). New dimensions of temporal serendipity and temporal novelty in recommender system. *Advances in Applied Science Research, 4*(1), 151-157.

Rao, K. N., & Talwar, V. (2011). Content-based document recommender system for aerospace grey literature: System design. *DESIDOC Journal of Library & Information Technology, 31*(3).

Saffer, D. (2010). Designing for interaction: creating innovative applications and devices. New Riders.

Reisenzein, R. (2000). The subjective experience of surprise. *The message within: The role of subjective experience in social cognition and behavior*, 262-279.

Renda, M. E., & Straccia, U. (2002). A personalized collaborative digital library environment *Digital Libraries: People, Knowledge, and Technology* (pp. 262-274): Springer.

Renda, M. E., & Straccia, U. (2005). A personalized collaborative digital library environment: a model and an application. *Information processing & management, 41*(1), 5-21.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology, 84*(6), 1236.

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM, 40*(3), 56-58.

Reuters, T. (2013). EndNote X7. *Thomson Reuters: Philadelphia, PA, USA*.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook *Recommender systems handbook* (pp. 1-35): Springer.

Rich, E. (1979). User modeling via stereotypes. *Cognitive science, 3*(4), 329-354.

Rim, R., Amin, M., & Adel, M. (2013a). *Bayesian networks for user modeling: Predicting the user's preferences.* Paper presented at the Hybrid Intelligent Systems (HIS), 2013 13th International Conference on.

Rim, R., Amin, M. M., & Adel, M. (2013b). *Bayesian networks for user modeling: Predicting the user's preferences.* Paper presented at the Hybrid Intelligent Systems (HIS), 2013 13th International Conference on.

Roberts, J. (2001). Trust and control in Anglo-American systems of corporate governance: The individualizing and socializing effects of processes of accountability. *Human relations, 54*(12), 1547-1572.

Rocha, L. M. (2001). TalkMine: a soft computing approach to adaptive knowledge recommendation *Soft computing agents* (pp. 89-116): Springer.

Rodriguez, M. A., Allen, D. W., Shinavier, J., & Ebersole, G. (2009). A recommender system to support the scholarly communication process. *arXiv preprint arXiv:0905.1594*.

Rubens, N., Kaplan, D., & Sugiyama, M. (2011). Active learning in recommender systems *Recommender Systems Handbook* (pp. 735-767): Springer.

Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering, 14*(2), 131.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*: Malaysia; Pearson Education Limited.

Said, A. (2013). Evaluating the Accuracy and Utility of Recommender Systems;*Doctoral dissertation,Universitätsbibliothek der Technischen Universität Berlin*

Salton, G. (1989). Automatic text processing: The transformation, analysis, and retrieval of. *Reading: Addison-Wesley*.

Sanders, E. B.-N. (2002). From user-centered to participatory design approaches. *Design and the social sciences: Making connections, 1*(8).

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Application of dimensionality reduction in recommender system-a case study: Minnesota Univ Minneapolis Dept of Computer Science.

Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems *The adaptive web* (pp. 291-324): Springer.

Schilit, B. N., & Theimer, M. M. (1994). Disseminating active map information to mobile hosts. *Network, IEEE, 8*(5), 22-32.

Schröder, G., Thiele, M., & Lehner, W. (2011). *Setting Goals and Choosing Metrics for Recommender System Evaluations.* Paper presented at the UCERSTI2 Workshop at the 5th ACM Conference on Recommender Systems, Chicago, USA.

Seddon, P., & Kiew, M.-Y. (1996). A partial test and development of DeLone and McLean's model of IS success. *Australasian Journal of Information Systems, 4*(1).

Seixas, F. L., Zadrozny, B., Laks, J., Conci, A., & Saade, D. C. M. (2014). A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer′ s disease and mild cognitive impairment. *Computers in biology and medicine, 51*, 140-158.

Šerić, L., Jukić, M., & Braović, M. (2013). *Intelligent traffic recommender system.* Paper presented at the Information & Communication Technology Electronics & Microelectronics (MIPRO), 2013 36th International Convention on.

Serrano-Guerrero, J., Herrera-Viedma, E., Olivas, J. A., Cerezo, A., & Romero, F. P. (2011). A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0. *Information Sciences, 181*(9), 1503-1516.

Shahamiri, Z. D. C. S. S. B. S. S. R. (2015). The Role of Context for Recommendations in Digital Libraries. *International Journal of Social Science and Humanity, 5*(11).

Shahin Mohammadi, S. K., Giorgos Kollias, and Ananth Grama. (2016). Context-Specifc Recommendation System for Predicting Similar PubMed Articles. *In Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on. IEEE, 1007–1014*.

Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems *Recommender systems handbook* (pp. 257-297): Springer.

Sharma, P. N., & Kim, K. H. (2012). Model Selection in Information Systems Research Using Partial Least Squares Based Structural Equation Modeling.

Shneiderman, B. (2004). Designing for fun: how can we design user interfaces to be more fun? *interactions, 11*(5), 48-50.

Shneiderman, B. (2010). *Designing the user interface: strategies for effective human-computer interaction*: Pearson Education India.

Sikka, R., Dhankhar, A., & Rana, C. (2012). A Survey Paper on E-Learning Recommender System. *International Journal of Computer Applications, 47*(9), 27-30.

Sinha, R., & Swearingen, K. (2002). *The role of transparency in recommender systems.* Paper presented at the CHI'02 extended abstracts on Human factors in computing systems.

Sparling, E. I., & Sen, S. (2011). *Rating: how difficult is it?* Paper presented at the Proceedings of the fifth ACM conference on Recommender systems.

Spillers, F. (2004). Emotion as a cognitive artifact and the design implications for products that are perceived as pleasurable. *Experience Dynamics*.

Sridharan, S. (2014). Introducing serendipity in recommender systems through collaborative methods.

Steinberg, R. M., Zwies, R., Yates, C., Stave, C., Pouliot, Y., & Heilemann, H. A. (2010). SmartSearch: automated recommendations using librarian expertise and the National Center for Biotechnology Information's Entrez Programming Utilities. *Journal of the Medical Library Association: JMLA, 98*(2), 171.

Stephenson, W. (1953). The study of behavior; Q-technique and its methodology.

Storbacka, K., Strandvik, T., & Grönroos, C. (1994). Managing customer relationships for profit: the dynamics of relationship quality. *International journal of service industry management, 5*(5), 21-38.

Strangman, N., & Hall, T. (2004). Background knowledge. *Wakefield, MA: National Center on Assessing the General Curriculum. Retrieved September, 26*, 2005.

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *The Communications of the Association for Information Systems, 13*(1), 63.

Sugiyama, K., & Kan, M.-Y. Towards Higher Relevance and Serendipity in Scholarly Paper Recommendation.

Sugiyama, K., & Kan, M.-Y. (2010). *Scholarly paper recommendation via user's recent research interests.* Paper presented at the Proceedings of the 10th annual joint conference on Digital libraries.

Sun, Y., Ni, W., & Men, R. (2009). *A Personalized Paper Recommendation Approach Based on Web Paper Mining and Reviewer's Interest Modeling.* Paper presented at the Research Challenges in Computer Science, 2009. ICRCCS'09. International Conference on.

Svensson, M., Höök, K., & Cöster, R. (2005). Designing and evaluating kalas: A social navigation system for food recipes. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(3), 374-400.

Swearingen, K., & Sinha, R. (2001). *Beyond algorithms: An HCI perspective on recommender systems.* Paper presented at the ACM SIGIR 2001 Workshop on Recommender Systems.

Swearingen, K., & Sinha, R. (2002). *Interaction design for recommender systems.* Paper presented at the Designing Interactive Systems.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction, 4*(4), 295-312.

Tamine-Lechani, L., Boughanem, M., & Daoud, M. (2010). Evaluation of contextual information retrieval effectiveness: overview of issues and research. *Knowledge and Information Systems, 24*(1), 1-34.

Tejeda-Lorente, Á., Porcel, C., Peis, E., Sanz, R., & Herrera-Viedma, E. (2014). A quality based recommender system to disseminate information in a university digital library. *Information Sciences, 261*, 52-69.

Tenenhaus, M., Amato, S., & Esposito Vinzi, V. (2004). *A global goodness-of-fit index for PLS structural equation modelling.* Paper presented at the Proceedings of the XLII SIS scientific meeting.

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology, 42*(4), 253-264.

Tibshirani, R., James, G., Witten, D., & Hastie, T. (2013). An introduction to statistical learning-with applications in R: New York, NY: Springer.

Tinschert, J., Natt, G., Mautsch, W., Augthun, M., & Spiekermann, H. (2001). Fracture Resistance of Lithium Disilicate--, Alumina-, and Zirconia-Based Three-Unit Fixed Partial Dentures: A Laboratory Study. *International Journal of Prosthodontics, 14*(3).

Tintarev, N., & Masthoff, J. (2007). *A survey of explanations in recommender systems.* Paper presented at the Data Engineering Workshop, 2007 IEEE 23rd International Conference on.

Torkzadeh, G., & Doll, W. J. (1999). The development of a tool for measuring the perceived impact of information technology on work. *Omega, 27*(3), 327-339.

Torres, R., McNee, S. M., Abel, M., Konstan, J. A., & Riedl, J. (2004). *Enhancing digital libraries with TechLens+.* Paper presented at the Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*(2), 127-145.

Trewin, S. (2000). Knowledge-based recommender systems. *Encyclopedia of library and information science, 69*(Supplement 32), 180.

Trujillo, M., Millan, M., & Ortiz, E. (2007). A recommender system based on multi-features *Computational Science and Its Applications–ICCSA 2007* (pp. 370-382): Springer.

Tsai, C.-H. (2017). *An Interactive and Interpretable Interface for Diversity in Recommender Systems*. Paper presented at the Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion, Limassol, Cyprus.

Tsai, C.-S., & Chen, M.-Y. (2008a). Using adaptive resonance theory and data-mining techniques for materials recommendation based on the e-library environment. *Electronic Library, The, 26*(3), 287-302.

Tsai, C.-S., & Chen, M.-Y. (2008b). Using adaptive resonance theory and data-mining techniques for materials recommendation based on the e-library environment. *The Electronic Library, 26*(3), 287-302.

Tsuji, K., Kuroo, E., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., & Itsumura, H. (2012). *Use of library loan records for book recommendation.* Paper presented at the Advanced Applied Informatics (IIAIAAI), 2012 IIAI International Conference on.

Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using partial least squares. *Journal of Information Technology Theory and Application, 11*(2), 5-40.

Vaishnavi, V., & Kuechler, W. (2004). Design research in information systems.

Van De Sompel, H., & Bollen, J. (2006). *An architecture for the aggregation and analysis of scholarly usage data.* Paper presented at the Digital Libraries, 2006. JCDL'06. Proceedings of the 6th ACM/IEEE-CS Joint Conference on.

Vargas-Govea, B., González-Serna, G., & Ponce-Medellın, R. (2011). Effects of relevant contextual features in the performance of a restaurant recommender system. *ACM RecSys, 11*.

Vellino, A. (2010a). A comparison between usage-based and citation-based methods for recommending scholarly research articles. *Proceedings of the Association for Information Science and Technology, 47*(1), 1-2.

Vellino, A. (2010b). A comparison between usage-based and citation-based methods for recommending scholarly research articles. *Proceedings of the American Society for Information Science and Technology, 47*(1), 1-2.

Vellino, A., & Zeber, D. (2007). *A hybrid, multi-dimensional recommender for journal articles in a scientific digital library.* Paper presented at the Proceedings of the 2007 IEEE/WIC/ACM international conference on web intelligence and international conference on intelligent agent technology.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Verbert, K., Lindstaedt, S. N., & Gillet, D. (2010). Context-aware Recommender Systems J. UCS Special Issue. *Journal of Universal Computer Science, 16*(16), 2175-2178.

Vermeeren, A. P., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). *User experience evaluation methods: current state and development needs.* Paper presented at the Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries.

Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (2010). *Handbook of partial least squares*: Springer.

Viriyakattiyaporn, P., & Murphy, G. C. (2009). *Challenges in the user interface design of an IDE tool recommender.* Paper presented at the Proceedings of the 2009 ICSE Workshop on Cooperative and Human Aspects on Software Engineering.

Visser, F. S., Stappers, P. J., Van der Lugt, R., & Sanders, E. B. (2005). Contextmapping: experiences from practice. *CoDesign, 1*(2), 119-149.

Vivacqua, A. S., Oliveira, J., & de Souza, J. M. (2009). i-ProSE: inferring user profiles in a scientific context. *The Computer Journal, 52*(7), 789-798.

Von Alan, R. H., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly, 28*(1), 75-105.

Wakeling, S. (2012). *The user-centered design of a recommender system for a universal library catalogue.* Paper presented at the Proceedings of the sixth ACM conference on Recommender systems.

Wang, C.-Y., Wei, F.-H., Chao, P.-Y., & Chen, G.-D. (2004). *Extending e-books with contextual knowledge recommenders by analyzing personal portfolio and annotation to help learners solve problems in time.* Paper presented at the Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on.

Wang, F.-H., & Shao, H.-M. (2004). Effective personalized recommendation based on time-framed navigation clustering and association mining. *Expert Systems with Applications, 27*(3), 365-377.

Wang, Y., Chan, S. C.-F., & Ngai, G. (2012). *Applicability of demographic recommender system to tourist attractions: A case study on trip advisor.* Paper presented at the Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03.

Webster, J., Jung, S., & Herlocker, J. (2004). Collaborative Filtering: a new approach to searching digital libraries. *New review of information networking, 10*(2), 177-191.

Webster, M. (2006). Merriam-Webster online dictionary.

Wesley-Smith, I., & West, J. D. (2016). *Babel: A Platform for Facilitating Research in Scholarly Article Discovery*. Paper presented at the Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, Québec, Canada.

Whitney, C., & Schiff, L. R. (2006a). The Melvyl recommender project: Developing library recommendation services. *California Digital Library*.

Whitney, C., & Schiff, L. R. (2006b). The Melvyl recommender project: Developing library recommendation services.

Wilensky, R. (2002). *Re-inventing scholarly information dissemination and use.* Paper presented at the III Jornadas de Bibliotecas Digitales:(JBIDI'02): El Escorial (Madrid) 18-19 de Noviembre de 2002.

Will, T., Srinivasan, A., Im, I., & Wu, Y.-F. B. (2009a). Search personalization: Knowledge-based recommendation in digital libraries. *AMCIS 2009 Proceedings*, 728.

Will, T., Srinivasan, A., Im, I., & Wu, Y.-F. B. (2009b). Search Personalization: Knowledge-Based Recommendation in Digital Libraries. *in AMCIS 2009 Proceedings*.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*: Springer Science & Business Media.

Wu, D., Yuan, Z., Yu, K., & Pan, H. (2012). Temporal social tagging based collaborative filtering recommender for digital library *The Outreach of Digital Libraries: A Globalized Resource Network* (pp. 199-208): Springer.

Wu, J. Y., & Wu, C. L. (2014). *The Study of User Model of Personalized Recommendation System Based on Linked Course Data.* Paper presented at the Applied Mechanics and Materials.

Wu, W., He, L., & Yang, J. (2012). *Evaluating recommender systems.* Paper presented at the Digital Information Management (ICDIM), 2012 Seventh International Conference on.

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: use, characteristics, and impact. *MIS quarterly, 31*(1), 137-209.

Yang, C., Wei, B., Wu, J., Zhang, Y., & Zhang, L. (2009). *CARES: a ranking-oriented CADAL recommender system.* Paper presented at the Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries.

Yang, Q., Zhang, S., & Feng, B. (2007). *Research on personalized recommendation system of scientific and technological periodical based on automatic summarization.* Paper presented at the Information Technologies and Applications in Education, 2007. ISITAE'07. First IEEE International Symposium on.

Yang, S.-Y. (2010). Developing an ontology-supported information integration and recommendation system for scholars. *Expert Systems with Applications, 37*(10), 7065-7079.

Yang, W.-S., & Lin, Y.-R. (2013). A task-focused literature recommender system for digital libraries. *Online Information Review, 37*(4), 581-601. doi: 10.1108/oir-10-2011-0172

Yang, X., Zeng, H., & Huang, Y. (2009). *Artmap-based data mining approach and its application to library book recommendation.* Paper presented at the Intelligent Ubiquitous Computing and Education, 2009 International Symposium on.

Yang, Y., & Yun, L. (2010). *Literature recommendation based on reference graph.* Paper presented at the Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on.

Yoshikane, F., & Itsumura, H. (2013). Book Recommendation based on Library Loan Records and Bibliographic Information.

Yuan, J., Sivrikaya, F., Marx, S., & Hopfgartner, F. (2014). *When to Recommend What? A Study on the Role of Contextual Factors in IP-based TV Services.* Paper presented at the MindTheGap@ iConference.

Yuan, Z., Yu, T., & Zhang, J. (2011). A social tagging based collaborative filtering recommendation algorithm for digital library *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation* (pp. 192-201): Springer.

Yujie, Z., & Licai, W. (2010). *Some challenges for context-aware recommender systems.* Paper presented at the Computer Science and Education (ICCSE), 2010 5th International Conference on.

Zaier, Z., Godin, R., & Faucher, L. (2008). *Evaluating recommender systems.* Paper presented at the Automated solutions for Cross Media Content and Multi-channel Distribution, 2008. AXMEDIS'08. International Conference on.

Zarrinkalam, F., & Kahani, M. (2013a). SemCiR: A citation recommendation system based on a novel semantic distance measure. *Program, 47*(1), 92-112.

Zarrinkalam, F., & Kahani, M. (2013b). SemCiR: A citation recommendation system based on a novel semantic distance measure. *Program: electronic library and information systems, 47*(1), 92-112. doi: 10.1108/00330331311296320

Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. *TechTrends, 55*(1), 32.

Zhang, M., Wang, W., & Li, X. (2008a). A Paper Recommender for Scientific Literatures Based on Semantic Concept Similarity. *Digital Libraries: Universal and Ubiquitous Access to Information*, 359-362.

Zhang, M., Wang, W., & Li, X. (2008b). A paper recommender for scientific literatures based on semantic concept similarity *Digital libraries: Universal and ubiquitous access to information* (pp. 359-362): Springer.

Zheng, Y. (2017). Interpreting Contextual Effects By Contextual Modeling In Recommender Systems. *arXiv preprint arXiv:1710.08516*.

Zheng, Y., Mobasher, B., & Burke, R. (2014). *CSLIM: Contextual SLIM recommendation algorithms.* Paper presented at the Proceedings of the 8th ACM Conference on Recommender Systems.

Zhu, Z., & Wang, J.-y. (2007). *Book Recommendation Service by Improved Association Rule Mining Algorithm.* Paper presented at the Machine Learning and Cybernetics, 2007 International Conference on.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). *Improving recommendation lists through topic diversification.* Paper presented at the Proceedings of the 14th international conference on World Wide Web.

Zukerman, I., & Albrecht, D. W. (2001). Predictive statistical models for user modeling. *User Modeling and User-Adapted Interaction, 11*(1-2), 5-18.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Champiri, Z. D., Shahamiri, S. R., & Salim, S. S. B. (2015). A systematic review of scholar context-aware recommender systems. Expert Systems with Applications, 42(3), 1743-1758.

Champiri, Z. D., Shahamiri, S. R., & Salim, S. S. B. (2015). The role of context for recommender systems in digital libraries, International Journal of Social Science and Humanity (IJSSH), V.5, N.10, 2014 (SCOPUS-Cited Publication)

Champiri, Z. D., Asemi, A., & Salim, S. S. B. (2019). A meta-analysis of evaluation methods, metrics in context-aware scholarly recommender systems, Knowledge and Information Systems journal-۱-۳۲

Champiri, Z. D., Salim, S. S. B., Loo, C.K., (2018). A Bayesian user model for scholarly recommender systems, (Under review in the journal of Expert Systems with Applications)

Champiri, Z. D., & Salim, S. S. B., B., Loo, C.K. (2018). How contextual data influences user experience with scholarly recommender systems: A conceptual Framework, (Under review in the journal of Behaviour & Information Technology (BIT))

Champiri, Z. D. & Salim, S. S. B. (2018). Interaction design guidelines for recommender systems: a case study of a scholarly recommender system (under review in: big data recommender systems: recent trends and advances, book, the Institution of Engineering and Technology (IET))

Note. The status of “Under review” is based on the time of printing this thesis.