SEMANTIC RECOMMENDER SYSTEM FOR MALAYSIAN TOURISM INDUSTRY

TIRAD MOHAMMED AREF ALMALAHMEH

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

2014

SEMANTIC RECOMMENDER SYSTEM FOR MALAYSIAN TOURISM INDUSTRY

TIRAD MOHAMMED AREF ALMALAHMEH

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

2014

UNIVERSITI MALAYA
<u>ORIGINAL LITERARY WORK DECLARATION</u>

Name of Candidate: **TIRAD MOHAMMED AREF ALMALAHMEH** (I.C/Passport No: J098196)
Registration/Matric No: WHA080006

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):
SEMANTIC RECOMMENDER SYSTEM FOR MALAYSIAN TOURISM INDUSTRY
Field of Study**: INFORMATION RETRIEVAL AND SEMANTIC TECHNOLOGY**

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: Mar 2014

Subscribed and solemnly declared before,

Witness's Signature Date: Mar 2014

Name: Assoc. Prof. Dr. Sameem Abdul Kareem
Designation: Supervisor

To My Family

ABSTRACT

Semantic Malaysian Tourism Recommender System (SMTRS) adopts the natural language interface, recommender system and semantic technology to analyse users' query and provide answers from the Malaysian tourism domain based on the tourists' preferences. Tourists usually search for information through different search engines. However, as found by various researchers the retrieved answers have two main problems: *overloaded* and *not-related* answers. A Recommender System (RS) is one application that can provide personalized information, with the optimal goal of providing personalized information recommendation in order to customize the World Wide Web (WWW). Regular RS users query the system by choosing from a fixed set of attributes represented by option sets or dropdown lists. Menu-driven navigation and keyword search currently provided by most commercial sites have considerable limitations because they tend to overwhelm and frustrate users with lengthy, rigid, and ineffective interactions. This research proposes incorporating semantic technology with a recommender system to deliver information that is more related to the tourists' interests. At the same time a User-friendly Natural Language Interface is also included to assure convenient query access to the Semantic Web data, where the Natural Language Interfaces are perceived as the most acceptable by end-users. The approach results in a prototype with an architecture consisting of a Content-based Recommender System, Semantic Technology, ontology engineering in the Malaysian Tourism domain, and Natural Language Interface. This research found, users are satisfied with the proposed services giving it an excellent rating based on the System Usability Scale (SUS) acceptability score.

ABSTRAK

Sistem Penentu Semantik Pelancongan Malaysia (SMTRS) menerima pakai antara muka bahasa tabii, sistem penentu dan teknologi semantik untuk menganalisis pertanyaan pengguna dan menyediakan jawapan daripada domain pelancongan Malaysia berdasarkan pilihan pelancong. Para pelancong biasanya mencari maklumat melalui enjin carian yang berlainan. Namun, menurut pelbagai penyelidik, jawapan-jawapan yang didapati mempunyai dua masalah utama: jawapan sarat dan yang tidak berkaitan. Sistem Penentu (RS) adalah satu aplikasi yang boleh memberikan maklumat peribadi, dengan matlamat optimum bagi memberikan cadangan maklumat peribadi mengikut persanan Jaringan Sejagat (WWW).

Para pengguna tetap RS mengemukakan pertanyaan melalui sistem ini dengan membuat pilihan daripada satu set tetap ciri-ciri yang diwakili oleh set-set pilihan atau senarai jatuh bawah. Navigasi menggunakan menu dan carian kata kunci yang disediakan oleh kebanyakan laman web komersial mempunyai batasan yang besar kerana mereka lebih cenderung untuk membanjiri dan mengecewakan para pengguna dengan interaksi yang panjang, tegar dan tidak berkesan. Kajian ini mencadangkan penggabungan teknologi semantik dengan sistem penentu bagi menyampaikan maklumat yang lebih relevan dengan kehendak para pelancong.

Pada masa yang sama, suatu Antara Muka Bahasa Tabii yang Mesra Pengguna juga dimasukkan bagi memastikan akses pertanyaan yang mudah kepada data Web Semantik, di mana Antara Muka Bahasa Tabii dianggap sebagai paling boleh terima oleh pengguna akhir. Pendekatan ini menghasilkan satu prototaip dengan seni bina yang terdiri daripada satu Sistem Penentu Berasaskan Kandungan, Teknologi Semantik,

kejuruteraan ontologi dalam domain Pelancongan Malaysia, dan Antara Muka Bahasa Tabii.

Kajian ini mendapati bahawa para pengguna berpuas hati dengan perkhidmatan yang dicadangkan, memberikannya penarafan yang cemerlang berdasarkan skor kebolehterimaan Skala Sistem Kebolehgunaan (SUS).

ACKNOWLEDGMENTS

First and foremost, I would like to thank the Creator of the Universe, Most Gracious and Most Merciful, without whose Will, it would not have been possible for me to fulfill my wish of completing this PhD.

I would like to express my deepest and sincere gratitude to my supervisor, Assoc. Prof. Dr. Sameem Abdul Kareem for her immeasurable wisdom, guidance, patience, valuable suggestions and supervision. Also, special thanks to my friend, Dr. Mansoor Abdullateef for his helpful suggestions and constructive criticisms of this thesis. For this, I will never be grateful enough.

I have no words to express my heartfelt gratitude to my beloved parents, Mohammed and Suzan; and my brothers Rabee and Basel, and sisters Reem and Aseel for their unconditional love, continuous support and trust in me. Also this PhD would never have happened without my lovely wife Faten Maali and my two brilliant children Bashar and Malak, it is a privilege to have all of you as mine, thanks for your support and contributions to the success of this endeavour during all these years, God keep you all for me.

Finally, I would like to thank all the staff at the Faculty of Computer Science and Information Technology for their cooperation and a kind treatment throughout my study. I would also like to thank all my friends and colleagues, for their encouragement, cooperation and help. I learnt a great deal from each of you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Term	Definition
ANNIE	A Nearly-New Information Extraction System
API	Application Programming Interface
B2C	Business to Consumer
CRS	Computer Reservation System
CBF	Content-based filtering
CF	Collaborative filtering
CBR	Case-based reasoning
DAML	DARBA Agent Markup Language
DL	Description Logic
EU	European Union
GATE	General Architecture for Text Engineering
GDSs	Global Distribution Systems
GUI	Graphical User Interface
ICT	Information and Communication Technology
JAPE	Java Annotation Patterns Engine
jOWL	Plug-in JavaScript library for visualizing OWL-RDFS documents
MTO	Malaysia Tourism Ontology
NL	Natural Language
NLI	Natural Language Interface
NLIDB	Natural Language Interface to DataBase
NLP	Natural Language Processing
OIL	Ontology Inference Layer
OWL	Ontology Web Language
QA	Question Answering
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RS	Recommender System
SDD	System Design Description

Term	Definition
SMTRS	Semantic Malaysian Tourism Recommender System
SPARQL	Protocol and RDF Query Language
SQL	Structure Query Language
SRS	Software Requirements Specification
SUS	System Usability Scale
THN	Tourism Harmonisation Network
UML	Unified Modelling Language
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
WTO	World Tourism Organization
WWW	World Wide Web
XML	Extensible Markup Language

1.0 Introduction

The Tourism economy is one of the fastest growing activities in developed countries (UNWTO, 2010). As it is an information based business, Information and Communications Technology (ICT) continue to be one of the greatest influences fuelling dramatic changes in reducing uncertainty and perceived risks to enhance the quality of trips by providing renewed web services (WTOBC, 2001). Unfortunately many of such services are not very tourist-oriented. A typical scenario is where tourist agencies provide general information (In the Web) about the tourism sites without beaming in mind the tourist's interest (e.g. focus on special activities) (Marcus & Chen, 2002). For instance, some tourists visiting Malaysia might be interested in shopping others might be interested in Historical and Cultural Sites and so on. Current tourism websites lumped all kinds of activities without considering the specific interest of the user.

A Recommender System (RS) is one application that can provide personalized information especially in the field of tourism. The goal of RS is providing personalized information recommendation to customize the World Wide Web (www) environment. Tim Berners-Lee introduced the Semantic Web in 2001(Berners-Lee, Hendler, & Lassila, 2001), as "extension of the current Web in which information is given the well-defined meaning, better enabling computers and people to work in cooperation". In other words, the Semantic Web is a kind of knowledge representation which enables machine to understand terms and relations in a specific domain to support the user in his tasks(Dotsika, 2010; Janev & Vranes, 2009).

Moreover, "*The term semantic technology represents a fairly diverse family of technologies that have been in existence for a long time and seek to help derive meaning from information. Some examples of semantic technologies include natural language processing (NLP), data mining, artificial intelligence (AI), category tagging, and semantic search.*" (Bio, 2013). Semantic technology is considered as a new and emerging technology as it is only merely about 10 years old. Its services (such as searching by meanings rather than keywords) are perceived as a very promising technology to enhance the web by enabling machines to understand the information available in different domains.

The domains includes E-Learning (Rathod, Prajapati, & Singh, 2012), Graph query processing (Yıldırım, Chaoji, & Zaki, 2012), Cloud Computing (Husain, McGlothlin, Masud, Khan, & Thuraisingham, 2011; Siva & Poobalan, 2012), Information Retrieval (Duhan & Sharma, 2011) etc.

Tourism is a perfect application area for Semantic Technology, since information dissemination and exchange are the key backbones of the travel industry as mentioned by (Cardoso, 2005). Meanwhile, Ontology plays an essential role in realizing this Semantic Technology, which is a set of shared, explicit and formal concepts used to organize and classify contents. In addition, researchers have shown that the Natural Language Interface (NLI) is perceived as the most acceptable means of communication by end-users and it is used to simplify the information retrieval process (Kaufmann & Bernstein, 2007).

1.1 Motivation

Malaysia aims to be a developed nation and the Vision 2020 is one way of reaching this. Malaysia is also becoming one of the world's favourite tourism destinations. In the year 2009 Malaysia had a revenue of RM53.4 billion from 23.6 million arrivals, while the Malaysian target is 36 million arrivals with a revenue of RM168 billion as stated in the Malaysian 2020 vision. (Asean Affairs, 2010; Tourism Malaysia, 2010). Therefore we strongly believe that an ICT based tourism development package would prepare the tourism industry of Malaysia to take up this huge challenge. In addition, combining Content-based recommender systems with semantic technology will results in a new level of depth that provides seamless interoperation between systems and users for the Malaysian Tourism domain. This seamless interoperation will provide more personalized answers to the queries made by tourists. It will filter the information resources semantically according to the personalized information obtained from the users, construct the personalized information environment, and provide the information and service according to the users' interest. As a result, this will lead to a higher state of user satisfaction with the services provided (i.e. personalized answers) for the benefits of the Malaysian Tourism Industry.

1.2 Research Problem

Information and Communication Technology (ICT) has played an important role in the development of tourism (Buhalis & Licata, 2002). Traditionally, tourism information is generated and published by multiple official or unofficial tourism sources (Marchiori & Cantoni, 2011). Hence, the information required by tourists is delivered in a rather random publishing environment. Tourists usually search for information through different search engines. However, the retrieved answers have two main problems: *overloaded* and *not-related* answers (S. Choi, Lehto, & Oleary, 2007; Jones, Ravid, & Rafaeli, 2004).

These two problems arise since there are no standards for representing tourism data (Bilbao, Lejarazu, & Herrero, 2010; Mistilis & Buhalis, 2012). We elaborate the problem in three main points:

PS1. **Information Overload:** Tourists face a load of information and resources, which can lead to difficulty in the decision-making process (Park & Jang, 2013). RS is the best solution for information overload (Adomavicius & Tuzhilin, 2005). Evidence from literature about Tourism Systems reveals that the most significant research efforts have been in improving Recommender System (RS) by mobility (Gavalas & Kenteris, 2012) and context-awareness (Adomavicius, Sankaranarayanan, Sen, & Tuzhilin, 2005). Regular RS users query the system by choosing from a fixed set of attributes represented by option sets or dropdown lists. Unfortunately, this diversity of terms results in a dramatically overloaded search interface. The complexity of such overloaded interfaces is an argument in favour of query formulation in the natural language (Berger, Dittenbach, & Merkl, 2004). Menu-driven navigation and keyword search currently provided by most commercial sites have considerable limitations because they tend to overwhelm and frustrate users with lengthy, rigid, and ineffective interactions (Chai et al., 2002). For instance, in the DIETORECS system (Ricci et al., 2006) users query the system by choosing from a fixed set of attributes represented by option sets or dropdown lists as shown in Figure 1.1.

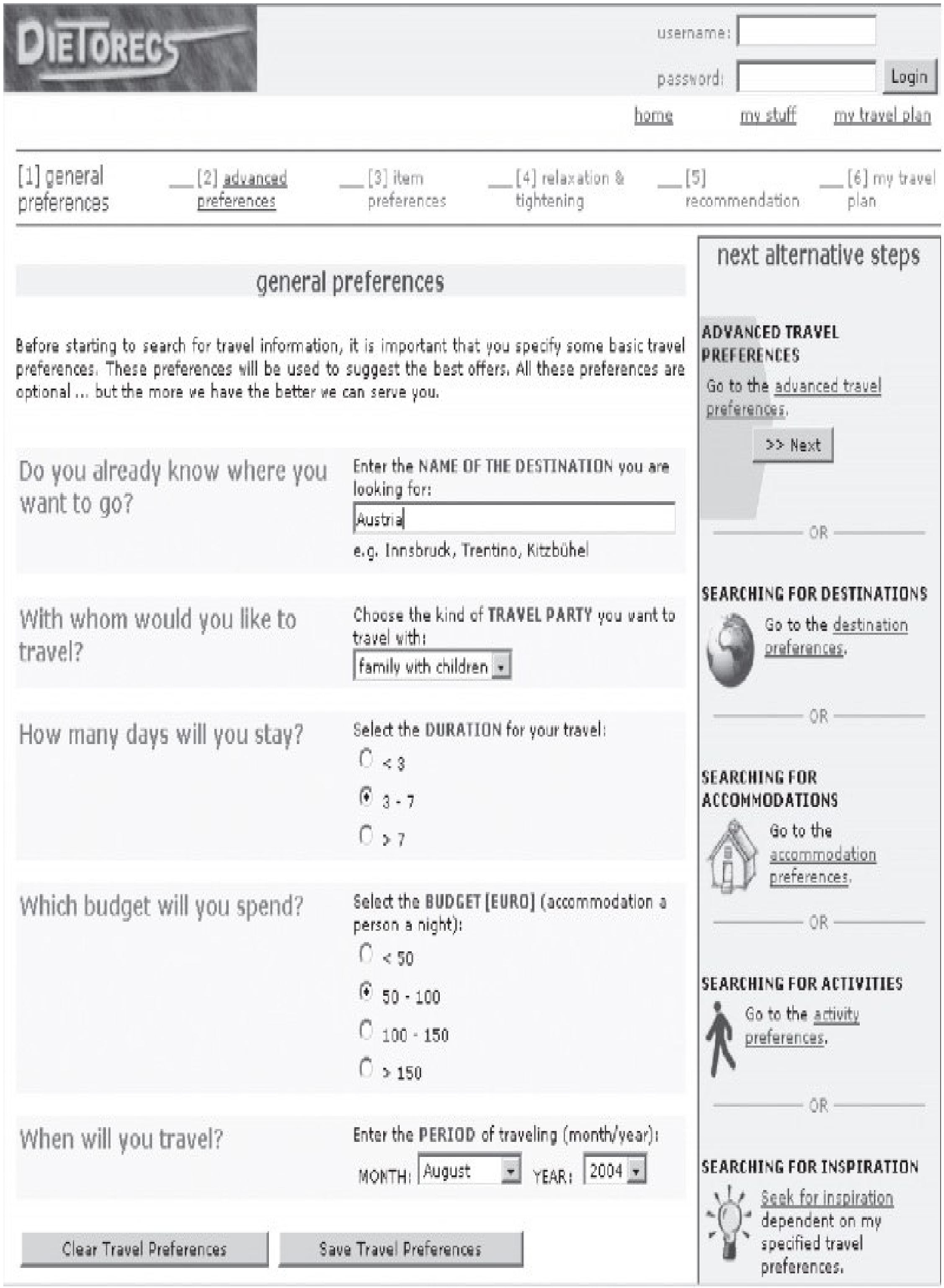


Figure 1.1 DIETORECS interface adopted from (Ricci, et al., 2006)

According to Staab (2002) Natural Language Interface is one of the requirements of future systems. According to the interface evaluation conducted in Kaufmann and Bernstein (2007), systems developed to support Natural Language Interfaces (NLI) are perceived as the most acceptable by end-users. Familiarity with the natural language used in these systems is a key to simplify the information retrieval processes. Hence, to provide an efficient solution for information access, the Natural Language Interface is required.

PS2. **Poor Knowledge Representation:** Tourism information is freely available in the Web but most of these sources are isolated from each other. Huge detailed pieces of information are available in the internet but this information is difficult to connect with the offers available. For instance, the information provided for a "hotel query" in a specific Hotel Website will link the activities happening nearby the hotel. However, Tourism websites do not present the available data in this linked format. Thus, tourists are required to link these data to have a complete picture of the information (Lam & McKercher, 2013).

PS3. **Absence of Personalization and User Profile Utilization:** With the huge amount of information available on the internet the tourists are confronted with the difficult decisions about how to select products more suited to their needs. This is due to the absence of personalization and the lack of information about user profiles. For instance, a system that knows that the user is Muslim, would be able to provide him/her with the more appropriate restaurant information that serves Halal food. Although these problems are discussed under different sub-topics, they are essentially inter-related, as information overload and poor knowledge representation are the result of the absence of personalized information. It is also noted that semantic technology, ontology and NL are also inter-related as the term “semantics” simply means “meaning” and an ontology is merely a way of capturing this meaning (Uschold, 2010).

NL technology can be utilized in any application where there is a large amount of unstructured information, particularly if the underlying information is related and structured stored in conventional databases (Gonzalez, 2013). NLP can be used to extract the structured data in the existing databases. These data can then be linked through Semantic technologies to pre-existing data located in other databases and elsewhere, thus bridging the gap between the unstructured and the structured data.

As a conclusion, the critical challenge in tourism domain is information explosion, particularly with respect to the amount of information available on the web. This often leads to the phenomenon of information overload where people get too much irrelevant information, which can lead to difficulty in the decision-making process.

Semantic technology, Natural Language Interface and recommender systems are discussed in chapter 2.

1.3 Research Objectives

In this research I would like to take advantage of Semantic Technology, Natural Language Interface and Recommender System in order:

RO1. To integrate the Natural Language Interface and Content-based Recommender System and incorporate Semantic Technology for a Malaysian tourism web service.

RO2. To develop a Malaysian Tourism Ontology to:

- Formalize a conceptual tourism knowledge representation (content-based)
- Capable to interact using a Natural Language Interface.

RO3. To build a semantic personalized information retrieval architecture using a filtering facility, natural language processing and querying ontology components.

RO4. To validate the proposed architecture: by developing a Semantic Malaysian Tourism Recommender System SMTRS (prototype) with the capability to answer users' question and recommending the best answer based on users' interest.

In order to link between the problem statements and the research objective:

RO 1 addresses the issues of the first problem statement, namely, information overload. RS is considered as the best solution in order to overcome the problems associated with information overload (Costa & Macedo, 2013). At the same time, NLI is perceived as ideal to tackle the problems relating to searches which result in information overload. Integrating RS with NLI provide personalization from RS with the ease of use interface from the NLI. Semantic technology are algorithms and solutions that seek to help derive structure and meaning from information. Thus, using semantic technology will ultimately reduce the issues with respect to information overload (Bio, 2013).

RO 2 An ontology formally represents knowledge as a set of concepts within a domain, using a shared vocabulary to denote the types, properties and interrelationships of those concepts. The importance of the ontology does not rest on the vocabulary per se but on the conceptualizations the terms in the vocabulary capture. Identifying such vocabulary and the underlying conceptualizations generally requires careful analysis of the kinds of objects and relations that can exist in the domain. Thus, an ontology is used to refer to a body of knowledge describing the domain, namely the tourism knowledge domain, using a representation vocabulary. The ontology captures the entities, ideas, and events, along with their properties, intrinsic conceptual structure and relations of the domain. The ontology can also represent goals, beliefs and predictions about the domain which in this research is the Malaysian tourism. There is no single correct ontology for any domain. Ontology design is a creative process and no two ontologies designed by different people would be the same. The potential applications of the ontology and the designer's understanding and view of the domain will affect ontology design choices. The quality of the ontology can only be assessed by using it in applications for which it was designed (Chandrasekaran, Josephson, & Benjamins, 1999; Noy & McGuinness, 2001). Hence, Objective 2 is related to addressing the issues in the second problem statement of knowledge representation. Objective 2 also addresses the issues of the third problem statement since the content-based recommender system is generally used to provide the preference that user would give to a particular item in the tourism domain they had not yet been considered, using a model built from the characteristics of an item (content-based approaches) (Lops, Gemmis, & Semeraro, 2011). With the rapid progress of technologies in the areas of computers and communications, the future computing environments will support seamless interactions from computers, networks and the web. That is, users could access the computer at anytime and anywhere requiring users to interact with computers through more natural and comfortable interfaces (Rhodes & Maes, 2000).

RO 3 Due to the explosion of the volumes of information available to users to deal with, there is a need to retrieve only the appropriate data suitable to the users' preferences or profiles and to present the retrieved information appropriately based on the users' special interest or special object. Only the most relevant information to the user is retrieved from the system and this is achieved through a personalized information retrieval architecture. Thus, Objective 3 addresses the issue in relation to PS3 (Hong, Park, Lee, Shin, & Woo, 2005)

1.4 Context and Scope

Tourism is an information based domain(Garzotto et al., 2004). The researchers introduce tourism as a hybrid industry since: it is information-based services while the core product is mainly physical services. Hence, this combination requires integration between information and physical services (Werthner & Klein, 1999). This information-based domain adopted the World Wide Web in order to improve the process of retrieving the tourism related information. Therefore, the WWW is the main source of information. Information is considered as one of the biggest needs for tourists. If they have the appropriate information, it will help them in making their choices about (*what to do*, *where to stay*, and *how to get there)* the trip (Siricharoen, 2008). However, this type of information are often isolated or intended with predefined programmes which simply broadcasts the tourism information, and provide the same information to users regardless of their interests.

On the other hand, information-overload is another issue, since providers seemed to focus on delivering as much information as possible and fail to take into account specific users' needs (Hinze & Buchanan, 2005). Furthermore, information searching mechanism depends on users' effort which machine agents are still unable to perform, as the information is not machine readable. Meanwhile, moving from the current broadcasting information to intelligent machine readable information is not an easy task as satisfying users' requirements should be considered during the designing and implementation of these intelligent services.

Recommender systems are known to have made contributions to the general success of personalized Websites by providing answers specific to the user's interest. In this context, Semantic Technology is perceived as a very promising technology. It enhances the web by enabling machines to understand the information available. Semantic technology enables data representation in a machine-readable form. Such representation facilitates the integration of tourist resources and data exchange among systems, which may include semantic descriptions of users and products provided(C. Choi et al., 2009; Damljanović & Devedžic, 2008). Meanwhile, Ontology plays an essential role in realizing the Semantic Web, which is a set of shared, explicit and formal concepts used to organize and classify contents of the domain (which in this case is the tourism domain). The task of inferring new knowledge from facts and rules is expressed in an ontology language so it can be used to reason about most important concepts of that domain, their attributes and relations between concepts. Several ontology query languages have been developed for extracting this knowledge from ontologies such as SPARQL, but for the casual end-users it is highly impossible to learn and use one of these query languages. Also, users need to understand the contents of the ontology in order to build a query, where users actually prefer to query the ontology using their natural language such as (English). Supporting the system with natural language interface is more user-friendly and will bring the advantages of this knowledge closer to the casual users (Kaufmann & Bernstein, 2007).

According to the interface evaluation conducted in Kaufmann and Bernstein [2007], systems developed to support Natural Language Interfaces (NLI) are perceived as the most acceptable by end-users. Familiarity with the natural language used in these systems is a key to simplifying the information retrieval processes. Natural language interfaces have the possibility to answer tourist questions about tourism related items. Tourists may find many answers provided by the Natural languages interface for each question. These answers are not personalized information with a particular interest to each individual user. A Recommender System (RS) is one of the applications that provide personalized information. The goal of RS is providing customized information in the World Wide Web environment. It actively constructs the personalized information environment, provides services according to the users' interest and filters the information resources according to the users' profile (Ricci, Rokach, & Shapira, 2011). Tourists have individual preferences so the user profile plays a vital role in the personalization process. The tourism User Profile is a structured representation of the tourists' preferences. An accurate User Profile enhances the information customization tremendously. In general, user profiles distinguish the needs of different users. It assists in providing customized answers to a particular tourist's query based on their needs.

Therefore, we summarise the scope of our research in adopting the natural language interface, recommender system and semantic technology in tourism to understand users' query and provide answers from the Malaysian tourism domain based on the tourists' preferences.

1.5 Methodology

In setting out to achieve the stated scope of this thesis, we carried out the process based on the methodology designed by (Mohammad Abooyee Ardakan, 2009; V. Vaishnavi & Kuechler, 2007; V. a. K. Vaishnavi, W, 2004). Figure 1.2 demonstrates the methodology used in the study.

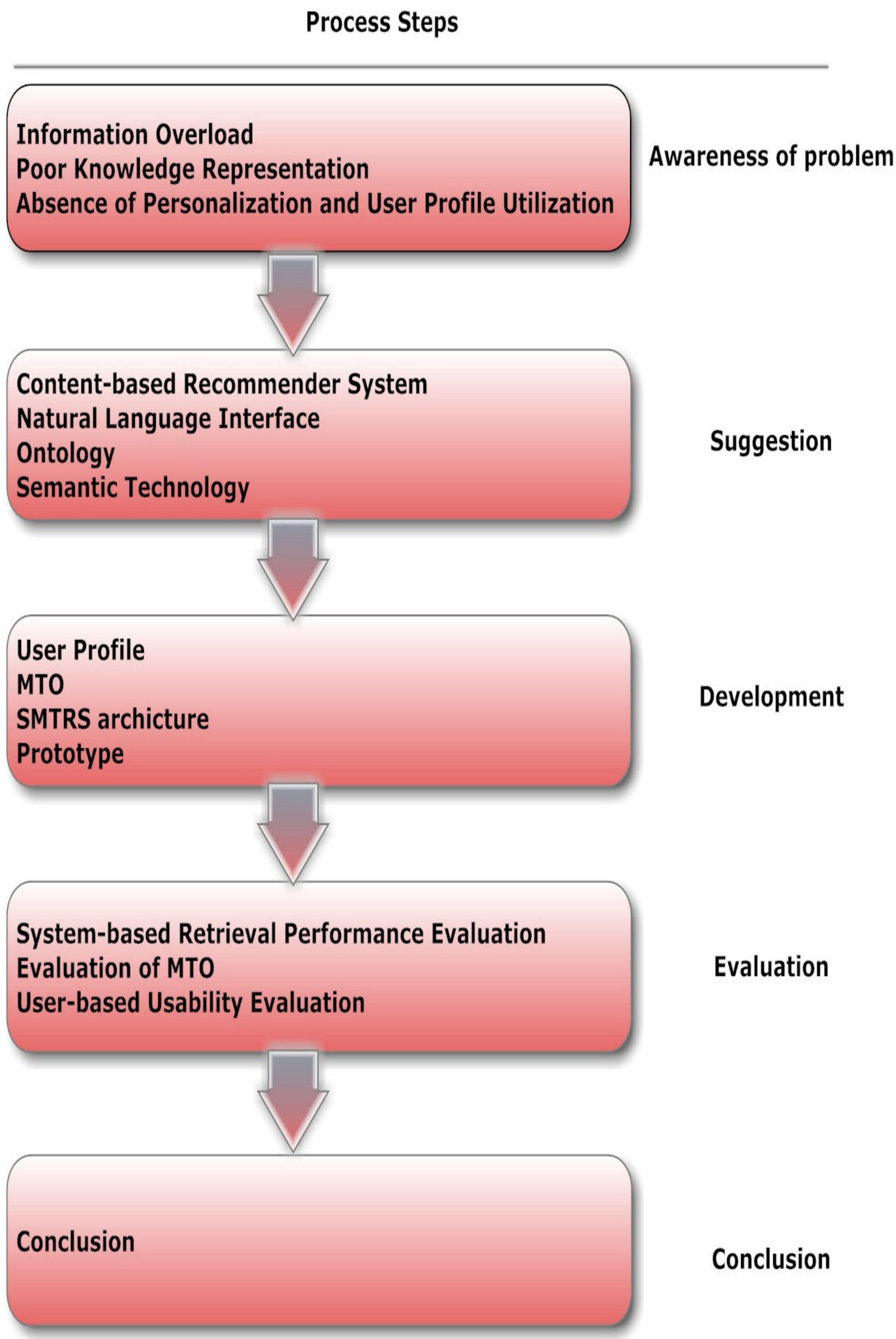


Figure 1.2 Research Methodology

1. Awareness of problem

    I. Investigate the literature on the application of ICT in the Tourism domain, the investigation includes its role in the Tourism development, the values of information in the Tourism domain and challenges facing ICT

    II. Investigate the literature on the different type of information technologies used in the tourism domain, also the type of information that tourists need

    III. Identify the challenges facing tourists while they search for tourism related information on the web, and how to improve the information search process

    IV. Explore and analyze technologies that can be employed to handle (overcome) the challenges in the Malaysian tourism domain

2. Suggestion

    I. Study possible solutions to process and answer tourist questions

    II. Study possible solutions to improve the quality of tourism recommendation with the use of a Malaysian Tourism Ontology

3. Development

    I. Prepare a set of questions to be used solely for testing purposes

    II. Create a semantic personalized information retrieval architecture

    III. Develop the SMTRS prototype to show how the proposed architecture will work

4. Evaluation

    I. Evaluate and validate the prototype

5. Conclusion

    I. Findings and Contributions

1.6 Research Questions

Q1.What do we understand about the challenges of the current information systems in tourism domain? What kind of technologies is nominated to overcome these challenges?

Q2. What is the appropriate recommender system for the tourism domain? Is there any need for improvement? What type of improvement need to be considered?

Q3. How can we use semantic technology to allow users the freedom to build natural language questions for Tourism information enquiry?

Q4. How to evaluate the SMTRS efficiency and the Malaysia Tourism Ontology (MTO)?

1.7 Thesis Overview

In Chapter 1.0, an introduction to the thesis is presented, the motivation, the main problems, the overall objectives, the methodology and the scope of this research.

Chapter 2.0, presents related literature review by exploring the importance of tourism, the role of ICT in Tourism Development, the challenges of the ICT in Tourism domain from an information based business perspective. Subsequently, this chapter discusses Semantic Technology and Ontology, with some existing Ontologies from the domain of Tourism. The chapter ends with a review of Natural Language Interface and Recommender systems from the Tourism Domain.

Chapter 3.0 discusses the steps followed to develop the prototype, including the implementation of the MTO using Protégé 4.0. This chapter also presents the justification of categorizing the collected tourism data.

Chapter 4.0 discusses the evaluation results achieved in detail, showing the usability and performance of SMTRS.

Chapter 5.0 summarizes the major contributions made in this thesis, followed by future work.

2.0 Literature Review

2.1 ICT and the Tourism Industry

2.1.1 Tourism Industry Features

Travel & Tourism is an important economic activity in most countries around the world (Chi-Ok, 2005; Kim, Chen, & Jang, 2006). It encompasses transportation, catering, accommodations, shops, entertainment, activity facilities, and other hospitality services for travellers. Actually tourism is one of the world's largest industries and the largest generator of jobs and it ranks fourth after fuels, chemicals and automotive products. For many developing countries it is one of the main sources of foreign exchange income and the number one export category, creating much needed employment and opportunities for development (Cernat & Gourdon, 2012). According to World Tourism Organization (UNWTO, 2011) tourists' arrivals estimated receipt reached US$ 1,030 billion worldwide for 2011. In Malaysia, for 2011 alone as shown in Table 2.1 a total of RM 58.3 billion of receipts and 24.7 million arrivals (Tourism Malaysia, 2010).

Table 2.1 Malaysia Tourist Facts and Figures (Tourism Malaysia, 2010)

Year	Arrivals	Receipts (RM)
2011	24.7 Million	58.3 Billion
2010	24.6 Million	56.5 Billion
2009	23.6 Million	53.4 Billion
2008	22.0 Million	49.6 Billion
2007	20.9 Million	46.1 Billion

Generally, tourism is defined as people travelling for a particular place to perform: specific tasks, for leisure or business purposes. The World Tourism Organization's definition for tourism is "Tourism comprises the activities of persons travelling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business, and other purposes."

(Goeldner & Ritchie, 2003) defined Tourism as the processes, activities, and outcomes arising from the relationships and the interactions among tourists, tourism suppliers, host governments, host communities, and surrounding environments that are involved in the attracting and hosting of visitors.

Tourism is defined as a short-term movement of people to places some distance from their normal place of residence to indulge in pleasurable activities. It may also involve travel for business purposes. (Horner & Swarbrooke, 1996)

The difference between these definitions comes from the different factors such as the main purpose of the trip, how far you have to travel and how many nights you have to stay away from home to be classified as a tourist. Our concern here, are the activities which are serviced by the Tourism industries such as hospitality and entertainment, therefore for the sake of this we select the World Tourism Organization definition.

Over the past six decades, tourism has expanded to become one of the largest and fastest growing economic sectors in the world. This growth and deepening diversification is linked to development and encompasses a growing number of new destination. Tourism has become one of the major player in international trade, and has become a valuable source of income even for developing countries

Some information provided by The World Tourism Organization (UNWTO); specialized agency of the United Nations and the leading international organization in the field of tourism are as follows:

- The contribution of tourism to economic activity worldwide is estimated at some 5%. Tourism's contribution to employment tends to be slightly higher relatively and is estimated in the order of 6-7% of the overall number of jobs worldwide (direct and indirect).
- From 1950 to 2010, international tourism arrivals expanded at an annual rate of 6.2%, growing from 25 million to 940 million.
- The income generated by these arrivals grew at an even stronger rate reaching around US$ 919 billion (€ 693 billion) in 2010.
- While in 1950 the top 15 destinations absorbed 88% of international arrivals, in 1970 the proportion was 75% and 55% in 2010, reflecting the emergence of new destinations, many of them in developing countries.
- As growth has been particularly fast in the world's emerging regions, the share in international tourist arrivals received by emerging and developing countries has steadily risen, from 32% in 1990 to 47% in 2010.

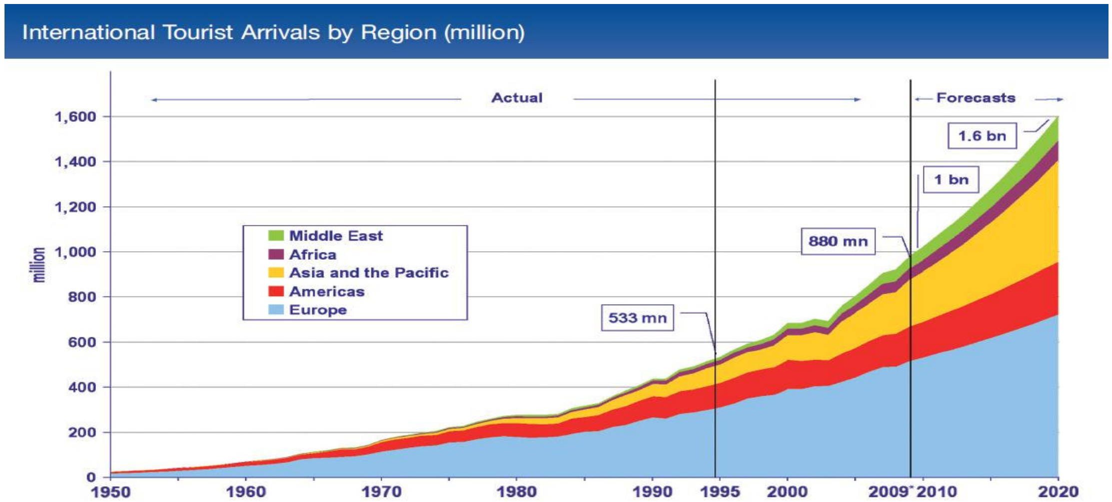


Figure 2.1: UNWTO Highlights 2010

According to Figure 2.1, international arrivals are expected to reach nearly 1.6 billion by the year 2020. The total tourist arrivals by region shows that, by 2020, the top three receiving regions will be Europe, East Asia and the Americas, followed by Africa, the Middle East and South Asia.

2.1.2 The role of ICT in Tourism Development

Information and Communication Technology (ICT) has played an important role in the development of tourism (Buhalis & Licata, 2002). Travel & Tourism is an important economic activity in most countries around the world (Chi-Ok, 2005; Kim, et al., 2006). This impact is accelerated by technological developments, as tourism is an information based business(Garzotto, et al., 2004). The impact of ICT is reflected on tourism industry market. This is through the distributed benefits to the market's members such as: tour operators, travel agencies, hotel enterprises, car rental companies, and cruise companies (Hojeghan & Esfangareh, 2011). ICT changes this business in terms of opening various channels for organizations to market their products and gain more relationships with customers(Poon, 1993).

The fast development of ICT and the increase of Internet users have reshaped the tourism Industry around the world, ICT has been the backbone of many process innovations, and it has attracted a great deal of research interest with its own agendas and institutions (Buhalis & Law, 2008). ICTs have been user in tourism since the establishment of the Computer Reservation System (CRS) for airlines in the 1970s and in the transformation to Global Distribution Systems (GDSs) in the late 1980s. This is followed by the development of the Internet in the late 1990s, which improved interconnectivity, interoperability, and altered the business operational practices in the tourism industry. The Internet has a major influence on the way people plan for and consume travelling (Buhalis & Law, 2008).

Werthner and Ricci (2004) presented in their famous paper "E-COMMERCE AND TOURISM" that despite severe economic problems that cause fewer passengers overall, online transactions in the tourism industry are continuously increasing, which puts the tourism industry on (business-to-consumer) B2C as leading applications. More recently, web2.0 became popular following the first O'Reilly Media Web 2.0 conference in 2004, in which there was no update to any technical specifications, but to changes in the ways software developers and end-users use the web. The term facilitates new generation of web-based communities and hosted services to enable collective intelligence on the internet. Users of Web 2.0 has become the information consumer, player and provider (Nicholas, Huntington, Jamali, & Dobrowolski, 2007), and Web 2.0 makes websites more interactive and user friendly.

As a result, incorporation ICTs in tourism businesses enrich the field with greater productivity, decreased costs, increased revenues and improved customer service.

2.1.3 Tourism as an Information Based Business

Information is one of the biggest needs for tourists. If they have the appropriate information it will help them in making their choices about what to do on the trip, where to stay, and how to get there (Siricharoen, 2008). This information is the first step of the vacation decision-making process. The necessity of tourism information is for choosing a destination and for on-site decisions such as selecting accommodations, travel mode, location activities, and tours (Chen & Gursoy, 2000; Gursoy & Terry Umbreit, 2004; Snepenger, Meged, Snelling, & Worrall, 1990).

According to KIM (2000) "*the Internet offers the potential to make information and booking facilities available to large numbers of tourists at relatively low costs. It also provides a tool for communication between tourism suppliers, intermediaries, as well as end-consumers*". Therefore, information search is one of the most widely researched issues by tourism researchers (Chen & Gursoy, 2000; Hyde, 2008; Schul & Crompton, 1983). Information search is defined as "*the motivated activation of knowledge stored in memory or acquisition of information from the environment*"(Engel, Blackwell, & Miniard, 1995). As the definition suggests, information sources are classified into two types(Fodness & Murray, 1997);

1. **Internal source:** include past experience of the destination, either with a specific or a similar destination, and based on the retrieval of knowledge from memory.

2. **External source:** which include advice from friends and relatives, market dominated (magazine, newspaper), destination-specific literature (travel guidebooks, government publications), travel advisor (travel agents, travel clubs) and more recently, the Internet. (Crotts, 2000; Hyde, 2008)

In order to support tourists with the information they need, it is extremely important to understand how tourists acquire information so you can get your message to them. It is a key issue for new development in supporting tourists in their search and decision process. There are four types of questions tourist may ask to get the required information (Cardoso, 2005; Cardoso & Sheth, 2006b; Siricharoen, 2008). These questions involve the “WH” type questions What, Where, When and How as described below:

1. **What**. What can a tourist see, visit and what can he do while staying at a tourism destination?
2. **Where**. Where are the interesting places located to see and visit? Where can a tourist carry out activity X?
3. **When**. When can the tourist visit a particular place? This includes exact time and weather. Consider that some activities cannot be undertaken under a raining climate.
4. **How**. How can the tourist get to X destination to see or do an activity?

In travel and tourism, studies indicate that users’ questions tend to be short, usually consisting of less than four keywords. Most users also do not go beyond those results on the second page. As a result, only a relatively small number of results are visible to the user though millions of potential web pages were found (Xiang, Gretzel, & Fesenmaier, 2009).

The Literature reveals several tourism activities that may be classified into four main categories of goods and services(Schmidt-Belz & Poslad, 2003):

- **Accommodation:** Accommodation classifies all facilities like hotels, guest houses and apartments.
- **Transportation**: Comprises all travels related to the vacation, including travel to/from airports, return flights, and all travels at the destination.
- **Activity:** Activities include the visitation of specific locations for recreational purposes and may be generally divided into attractions (museums, visitor centers, botanical gardens, etc.), entertainment (cinema, bar, shopping, etc.), and sport activities (diving, jet boating, golf, etc.).
- **Food:** Food includes dining at several kinds of restaurants (such as local cuisines, high quality restaurants and familiar chain restaurants), also purchasing local food products and consuming local beverages.

Searching for information on a desired spot for vacation is usually very time-consuming as will be discussed in the next section. For example, tourists are aware of several criteria that should be followed when searching for accommodation such as the distance from the shopping centre, sandy beach, a possibility to rent a car, etc., as well as of some individual constraints such as prices, departure times, etc. (Damljanović & Devedžic, 2008). When retrieving information from the internet, the search engines are only concerned with the information syntax but not the semantics of the information. The above shortcoming affects general search engines as discussed in (Horrocks, 2007) and tourism search engines as discussed in (Xiang, et al., 2009). It is overcome with the evolution of the next generation web 'Semantic Web'. Semantic Web maintains the web in a structured form and makes web accessible data more amenable to machine processing (Janev & Vranes, 2009). Such a representation, enables easier integration of tourist resources and data exchange, which may include semantic descriptions of users and products(Mistilis & Buhalis, 2012). The details of this representation will be elaborated in section 2.2.

2.1.4 Challenges of ICT in Tourism

- **Information Overload:** Information overload is the state of an individual (or system) in which not all communication inputs can be processed and utilized, leading to breakdown (Jones, et al., 2004) . Since the beginning of the World Wide Web, countless tourism businesses and organizations offer their products and services to their customers through the Web (Buhalis & Law, 2008). Information is generated and published by multiple official or unofficial tourism sources. This open and distributed nature of the Web makes it difficult for Web search engines to find information related to user needs due to the massive amount of information published (Aldebert, Dang, & Longhi, 2011). A typical scenario is where people get too much irrelevant information alongside relevant ones as a response to queries posed on the web. For instance, a Google (www.google.com) search for "hotels in Kuala Lumpur" brings up over 73,000,000 different links including advice needed in deciding where to stay in Kuala Lumpur for a search that was performed on the 27th of September 2013 as shown in the Figure 2.2**.**

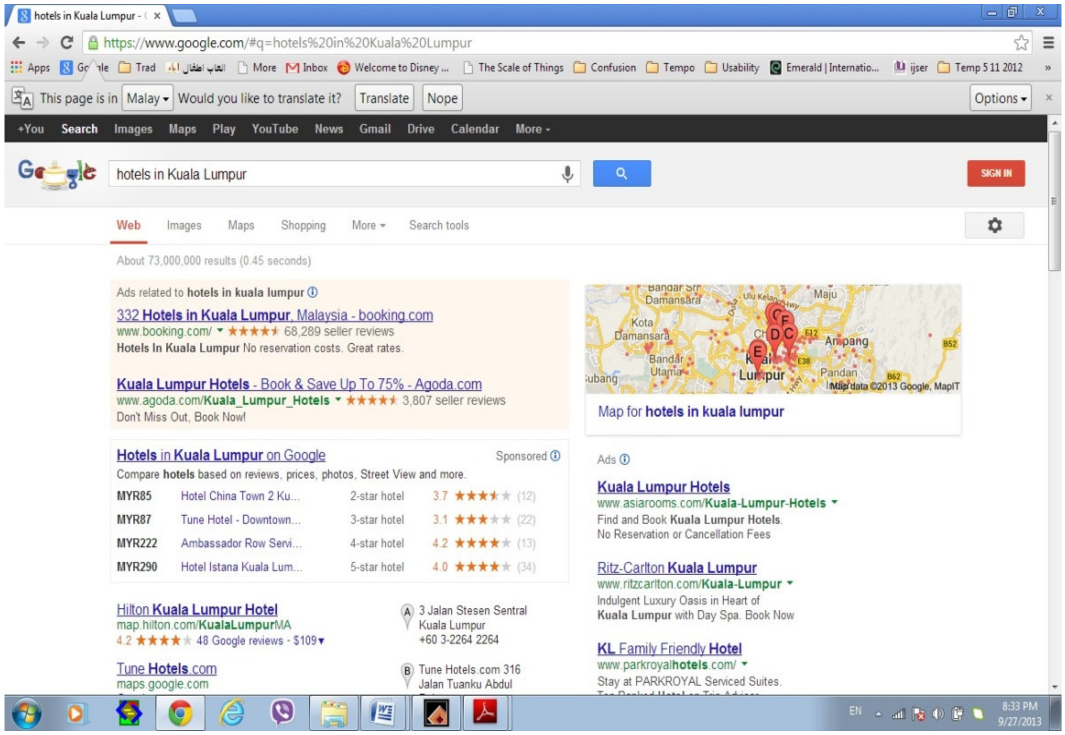


Figure 2.2 Google search for Hotels in Kuala Lumpur

Damljanovic and Devedzic (2008) mentioned that searching for information on a desired spot for vacation is usually very time-consuming. Moreover, tourists need to be aware of several criteria that should be followed when searching for accommodation such as the distance from the shopping centre, sandy beach, a possibility to rent a car, etc., as well as of some individual constraints such as prices, departure times, etc. Tourists feel overwhelmed before finding the intended information where the individual cannot process all the input communication. Tourists are often overloaded by heterogeneous information and resources, often leading to individual breakdown (Inversini & Buhalis, 2009; Jones, et al., 2004). RS is the best solution for information overload (Adomavicius & Tuzhilin, 2005). Regular RS users query the system by choosing from a fixed set of attributes represented by option sets or dropdown lists. Unfortunately, this diversity of terms results in a dramatically overloaded search interface. The complexity of such overloaded interfaces is an argument in favour of query formulation in the natural language (Berger, et al., 2004). Menu-driven navigation and keyword search currently provided by most commercial sites have considerable limitations because they tend to overwhelm and frustrate users with lengthy, rigid, and ineffective interactions (Chai, et al., 2002). For instance, in the DIETORECS system (Ricci, et al., 2006) users query the system by choosing from a fixed set of attributes represented by option sets or dropdown lists.

According to Staab (2002) Natural Language Interface is one of the requirements of future systems. According to the interface evaluation conducted in Kaufmann and Bernstein (2007), systems developed to support Natural Language Interfaces (NLI) are perceived as the most acceptable by end-users. Familiarity with the natural language used in these systems is a key to simplify the information retrieval processes. Hence, to provide an efficient solution for information access, the Natural Language Interface is required.

- **Poor Knowledge Representation:** Data providers represent the information in their own vocabulary. Pan and Fesenmaier (2006) found that the "language of tourism" (Dann, 1996) is extremely rich; further, their study indicated that the vocabularies used on destination marketing organization websites differ substantially from those of potential users. As such, they concluded that the richness in language and the differences in perspectives make it very difficult for Internet users to have a satisfying online search experience. The challenge in the tourism domain is that tourists fall victims to poor communication and lack of a common understanding (Cardoso, 2005). Massive repositories in the Web contain huge volumes of distributed heterogeneous data. This situation is especially true for the tourism industry where a broad spectrum of tourism information is already distributed over various web sites and stored using heterogeneous formats (for instance Malaysian food websites as shown in Table 3.1). It is obvious that this situation is very undesirable since the tourist is burdened with finding and visiting various web sites in order to gather all the desired tourism information and products (Haller, Pröll, Retschitzegger, Tjoa, & Wagner, 2000). Moreover, enabling machine to machine interaction makes data interoperable and capable of manipulation by computer programs, bringing 'structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users' (Berners-Lee, et al., 2001; Mistilis & Buhalis, 2012). Since current tourist web information are not machine readable (Mistilis & Buhalis, 2012), the idea is to enable the information to be used in a way that computer can understand (i.e. hotel has "internet access" and hotel has "Wifi" is same and easily understood by the human but for the machine this is not so). Semantic technology allows data representation in a machine-readable form. Such a representation, enables easier integration of tourist resources and data exchange (i.e. "where to stay" as a search term will mean "Accommodation"), which includes semantic descriptions of users and products. (Damljanovic, 2009; Damljanović & Devedžic, 2008). This will support interoperability and integration between systems and applications. Enabling machines to exchange and automate processing may became a reality by adapting semantic technology. The idea is to provide the information in a way that computers can understand, so it will allow software agents to analyze the web on our behalf, and produce results more relevant to users' needs. Semantic Technology provide the aforementioned promises with the assist of Metadata, Resource Description Framework (RDF) (RDF, 2002), and the Web Ontology Language (OWL)(Cardoso, Sheth, & Sheth, 2006a; OWL, 2004).

- **Absence of Personalization and User Profile Utilization:** Leading search engines suffer from low success rates when it comes to delivering relevant results to the average searcher. The need become more urgent to develop a new adaptive and evolving system that can help the information retrieval community on the internet obtain more individualized services (Smyth, Coyle, & Briggs, 2011). The tourism sector is one of the most striking that suffer from essentially the same problem, the "one-size-fits-all" nature of mainstream web search. There are huge detailed pieces of information available on the internet but tourists are confronted with the difficult decisions about how to select products more suited to their needs. Tan et al. (2007) caution that "many of such services are not tourist-oriented". A typical scenario is where tourist agencies provide general information in the Web about the tourism sites without bearing in mind the tourist interest (e.g. a tourist who would be interested in mountain climbing should be provided with hotels near the mountain or a tourist who wants to go scuba diving should only be provided with hotels near the sea). Service providers seemed focused on delivering as much information as possible and fail to take into account specific needs or interests of the users (Hinze & Buchanan, 2005). However, other research has highlighted how tourists expect individualised information and services, taking into account their own personal interests (Barta, Feilmayr, Pröll, Grün, & Werthner, 2009; Poslad et al., 2001).

2.2 Recommender System in Tourism Domain

2.2.1 Recommender System

A recommender System (RS) is one of the applications that provides personalized information. The goal of RS is to provide personalized information recommendation to customize the World Wide Web environment. It actively filters the information resources according to the personalized information obtained from users, constructs the personalized information environment, and provides the information and service according to the users' interest (Ricci, et al., 2011). Moreover, Recommender Systems represent a class of systems designed to help individuals and communities deal with information overload and incomplete information to make evaluative decisions (Adomavicius & Tuzhilin, 2005).

According to Ricci, et al. (2011) "*Item is the general term used to denote what the system recommends to users. A RS normally focuses on a specific type of item (e.g., CDs, or news) and accordingly its design, its graphical user interface, and the core recommendation technique used to generate the recommendations are all customized to provide useful and effective suggestions for that specific type of item*". Since the items (in our research, Tourism items; i.e. Accommodation, Activity and Food) affects the type of system recommendation, the RS selection will thus be based on the tourism domain. Furthermore, RS help individuals by predicting the required information automatically on behalf of the users according to their user profile. The user profile contains information about users' tastes, preferences, and needs.

2.2.2 Personalizing Recommendations for Tourists

RS for tourism have attracted a lot of research energy and interest (Kabassi, 2010). The main goal of these systems is that they can personalize their recommendations to each user interacting with the system. Personalization involves the design of enabling systems to capture the needs of each person and then to satisfy those needs in a known context by collecting user information. The objective of collecting user information is to create a structured representation that describes user preferences. In general, user profiles distinguish between different users. It assists in providing customized information to the users. The customized information matches user's requirements. Moreover, user profile is considered as a prerequisite for information retrieval and filtering. An accurate User Profile enhances the information customization efficiently.(Eirinaki & Vazirgiannis, 2003). There are different techniques to build a user profile.

The most common techniques are:

- **Explicit profiling**: each user indicate his/her preferences by filling in a form when visiting the web site; this method has the advantage of letting users specify directly their interests(Middleton, Shadbolt, & Roure, 2004).
- **Implicit profiling**: user preferences inferred automatically from his/her transactional behaviour over time. This technique is generally transparent to the user. Behaviour information is generally stored in a log file (Sugiyama, Hatano, & Yoshikawa, 2004).
- **Legacy data**: they provide a rich source of profile information for known users (Buono, Costabile, Guida, & Piccinno, 2002).

Explicit profiling is when each user indicate his/her preferences by filling in a form when visiting the web site; this method has the advantage of letting users specify directly their interests(Middleton, et al., 2004). Therefore, our selection is based on the common and direct technique to facilitate the information gathering, which is explicit profiling. Personalized tourism services aim at helping the users finding what they are looking for, easily without spending time and effort. Therefore, a variety of approaches have been used to perform recommendations in these domains, including content-based, collaborative, demographic, knowledge-based or hybrid approaches and many others (Montaner, López, & de la Rosa, 2003).

2.2.3 Recommender System Techniques

Various approaches for recommender systems have been developed. (Burke, 2002) showed that most existing recommender systems adopt three main types of techniques: namely, Content-based filtering, Collaborative filtering, and knowledge-based filtering. Burke (2002) also showed that these techniques have complementary advantages and disadvantages. Thus, Burke (2002) claims this fact has provided incentive for research in hybrid recommender systems that combine techniques for improved performance.

I. Content-based Filtering

Content-based filtering is a development of information filtering research and is based on the idea of recommendation as classification (Belkin & Croft, 1992). Content-based recommendation systems analyze item descriptions to identify items that are of particular interest to the user (Pazzani & Billsus, 2007). The details of recommendation systems differ based on the representation of items. Item is the general term used to denote what the system recommends to users. A RS normally focuses on a specific type of item (e.g., tourism, or movies) and accordingly its design, its graphical user interface, and the core recommendation technique used to generate the recommendations are all customized to provide useful and effective suggestions for that specific type of item (Ricci, et al., 2011).

The User Profile is a structured representation of user interests, adopted to recommend interesting items, in Content-based filtering a menu interface can be created to allow a user to create a profile, and after using the system for a period of time depending on the domain, items that recommended to the users are often stored, and machine learning algorithms may be used to create a dynamic user profile. The recommendation process basically consists of matching up the attributes of the user profile against the attributes of a content object. The result is a relevance judgment that represents the user's level of interest in that object. If a profile accurately reflects the user's preferences, it is of tremendous advantage for the effectiveness of an information access process. For instance, it could be used to filter search results by deciding whether a user is interested in a specific Web page or not and, in the negative case, preventing it from being displayed. Systems that use Content-based filtering in Tourism domain are shown in Table 2.2.

II. Collaborative Filtering (CF)

Collaborative filtering predicts the utility of items for a particular user based on the items previously rated by other users(Adomavicius & Tuzhilin, 2005). A typical user profile in a collaborative system consists of a matrix of items and their ratings. The greatest strength of Collaborative techniques is that they are completely independent of any machine-readable format of the registered items that will be recommended, and work well for complex objects such as music and movies where variations in taste are responsible for much of the variation in preferences. (Burke, 2002; Orgun, Thornton, Bohnert, & Zukerman, 2007)

Collaborative Filtering is more appropriate for areas where the universe of items is small or static and the density of user interests is relatively high. If the universe of items is static; rating from previous users will help to make more accurate prediction for the new user. Nevertheless, small set of items will increase the probability of overlap in the matrix of the items and their ratings.

Collaborative recommender systems depend on the overlap in ratings across users. It has difficulty when the space of ratings is sparse(Burke, 2005); mainly, there is a high number of items-of-interest and few user votes on items, therefore user's interests are missing. Another problem is the recommendation of a new item, as very few users have rated this item (Berka & Plößnig, 2004). Commercially, the most well known usage of collaborative-style explanations are the ones used by Amazon.com: "Customers Who Bought This Item Also Bought". This explanation assumes that the user is viewing an item which they are already interested in. It implies that the system finds similar users (who bought this item), and retrieves and recommends items that these similar users bought. Systems used Collaborative filtering in Tourism domain are shown in Table 2.2.

III. Knowledge-based Filtering

Knowledge-based Filtering uses the query to make recommendations based on inferences about a user's needs and preferences. Knowledge-based approaches predict recommendations based on the functional and domain knowledge: the system have the capability to know about how a particular item meets a particular user's need, and can therefore reason about the relationship between a need and a possible recommendation(Burke, 2002).

Knowledge-based filtering suffers from two major shortcomings:

i. **Knowledge Acquisition**: a well-known bottleneck for many artificial intelligence applications, which denotes the communication overhead between domain experts and knowledge engineers in the phase of constructing knowledge bases. Knowledge base filtering requires description of the specific knowledge by the domain expert who may know the rules of relationship governing the body of knowledge but may not know how to translate these into a knowledge product, but on the other hand, knowledge engineers who know how to create these programs may have a limited knowledge about the product domain.

ii. **Structured Domain Knowledge**: in order to build knowledge-based recommendation systems, the domain knowledge should be readily available in some structured machine-readable form, e.g. ontology. (Adomavicius & Tuzhilin, 2005).

2.2.4 Selection of Recommender System Techniques

Recommender Systems represent the main area where principles and techniques of Information Filtering are applied. Nowadays many web sites embody recommender systems as a way of personalizing their content for users (Resnick & Varian, 1997).

Table 2.2 summarize the tourism recommendation system with respect to the recommendation techniques

Table 2.2 Tourism recommender system with respect to the recommendation techniques. adapted from (Kabassi, 2010)

Content-based filtering	Collaborative	Hybrid/knowledge based
Triplehop's TripMatcher (Ricci et al., 2003)	**Speta** (García-Crespo et al., 2009)	**PTA** (Coyle & Cunningham, 2003)
VacationCoach (Staab, et al., 2002)	(Maw & Naing, 2006)	**Entreé** (Burke, 1999)
CAPA (Tung & Soo, 2004)	**UMT** (Yang & Marques, 2005)	(Huang & Bian, 2009)
Cyberguide (Abowd et al., 1997)	(Soe, Naing, & Ni Lar, 2006)	**Traveller** (Schiaffino & Amandi, 2009)
GUIDE system (Cheverst, Davies, Mitchell, Friday, & Efstratiou, 2000)		**Travel Planner** (Chin & Porage, 2001)
WebGuide (Fink & Kobsa, 2002)		**PTS** (Srivihok & Sukonmanee, 2005)
Sightseeing4U (Scherp & Boll, 2004)		(Hinze & Voisard, 2003)
MastroCARonte (Console, Torre, Lombardi, Gioria, & Surano, 2003)		
CATIS (Pashtan, Blattler, Heusser, & Scheuermann, 2003)		
MAIS Project (Corallo, Lorenzo, & Solazzo, 2006)		
INTRIGUE (Ardissono, Goy, Petrone, Segnan, & Torasso, 2003)		
Gulliver's Genie (O'Grady & O'Hare, 2004)		
MobiDENK (Krösche, Baldzer, & Boll, 2004)		
PinPoint (Roth, 2002)		
m-ToGuide prototype (Kamar, 2003)		
Method for personalising route planning (Niaraki & Kim, 2009)		
PRSET (Srisuwan & Srivihok, 2008)		
personalized recommendation in car navigation system (Iwasaki, Mizuno, Hara, & Motomura, 2007)		
ATA (Linden, Hanks, & Lesh, 1997)		
ITR (Ricci, et al., 2003)		

From Table 2.2 we conclude that there are two approaches that are more popular Content-based filtering and Collaborative filtering. Each type of filtering method has its own weaknesses and strengths as discussed in Table 2.3.

Table 2.3 Strength & Weakness of filtering methods adopted from (Adomavicius & Tuzhilin, 2005)

Recommendation Approach	Strength	Weakness
Content-based	Recommendation based only on facts that involve the particular user	Recommendation are poor when the system knows little information about the new users
	Derives from the fact that it is based on information for each individual user	Provide false recommendations when the searched item is a gift and not for the user interest
	Capture changes on the user's preferences	Provide to the system false information about the user as it would state other peoples' interests when searching for a gift items
Collaborative filtering	Once the system has found a user's neighbours then it can provide personalized recommendations	The neighbours selection may be controversial and this may result in diverse recommendations additionally may be difficult to find a user's good neighbour. Finally cannot recommend an item or service until several users have rated it

Selecting the appropriate RS filtering technique vary based on the representation of items (Ricci, et al., 2011). Collaborative technique recommends items based on users rating (Xiaoyuan Su & Khoshgoftaar, 2009). Content-based recommendation systems analyze item descriptions to identify items that are of particular interest to the user (Lops, et al., 2011). For the Tourism domain (Kabassi, 2010); Content-based recommendation systems is preferred than Collaborative technique as can be observed in Table 2.2. The nature of tourism items are descriptive (i.e. Most tourism establishments, such as hotels or guesthouses, have websites. Those websites contain necessary information such as contact data, prices, descriptions of offerings and pictures). For the aforementioned reasons our RS selection will consider Content-based recommendation and this thesis goal is not to overcome the disadvantages of recommender systems but merely to integrate the Natural Language Interface and Content-based Recommender System and incorporate Semantic Technology for a Malaysian tourism web service.

2.2.5 Drawbacks of Existing Content-based RS in Tourism Domain

In this subsection we analyze two main drawbacks on the existing Content-based RS in Tourism domain:

- **Semantic Item Description** Items recommended to the user by matching features of the item with the characteristics of the user that are maintained in his user profile. Representing the items by the same set of features and assign the values it may take shape the structured data. For example, in a hotel recommendation application, features selected to describe a hotel are: roomtype, room-price, facilities, location etc. In most content-based RS, features of the item are textual description extracted from Web pages. "*The problem is that traditional keyword-based search features are unable to capture the semantics of user interests because they are primarily driven by a string matching operation*" (Lops, et al., 2011). String matching suffers from problems of :

- POLYSEMY, the presence of multiple meanings for one word, for example the word "book":

  a) Written or printed work consisting of pages

  b) Reserve (accommodation, a place, etc.)

- SYNONYMY, multiple words with the same meaning, for example Health Club and Fitness Centre have the same meaning.

Semantic Technology and its integration in Recommender Systems is one of the pioneering approaches proposed in the coming subsection 2.3 to solve those problems. Semantic Technology provides representing the feature of items in knowledge bases, such as ontologies, in order to obtain a "semantic" interpretation of the items descriptions. The coming subsection 2.3 provides further discussion on the Semantic.

- **User Interface:** Regular RS users query the system by choosing from a fixed set of attributes represented by option sets or dropdown lists. Unfortunately, this diversity of terms results in a dramatically overloaded search interface. The complexity of such overloaded interfaces is an argument in favour of query formulation in the natural language (Berger, et al., 2004). "*Menu-driven navigation and keyword search currently provided by most commercial sites* *have considerable limitations because they tend to overwhelm and frustrate users with lengthy, rigid, and ineffective interactions*" (Chai, et al., 2002). According to Staab, et al (2002) Natural Language Interface is one of the requirements of future systems. According to the interface evaluation conducted in Kaufmann and Bernstein (2007), systems developed to support Natural Language Interfaces (NLI) are perceived as the most acceptable by end-users. Familiarity with the natural language used in these systems is a key to simplify the information retrieval processes. Hence, Natural Language Interface is required. The coming subsection 2.4 provides further discussion on the Semantic.

•

2.3 Semantic Web Technology

The Web is a bunch of massive repositories which contain huge volumes of distributed heterogeneous data. There are plenty of tourism information systems in use which manipulate this data. Systems within the same areas of operations are varying due to the lack of standards in constructing these systems. The challenge is how interoperation between these various systems can take place. Semantic web technologies can help to resolve many challenges in Web system development. Tim Berners-Lee introduced the Semantic Web in 2001(Berners-Lee, et al., 2001), as "extension of the current Web in which information is given the well-defined meaning, better enabling computers and people to work in cooperation". Or, that Semantic Web is a kind of knowledge representation which enables machine to understand terms and relations in a specific domain to support the user in his tasks(Dotsika, 2010; Janev & Vranes, 2009).

The benefit from semantic technology to the field of tourism can be categorized in the following forms:

- **Semantic Interoperability**

Data heterogeneity and interoperability are well known challenges for the electronic tourism market development where companies need to exchange information seamlessly. The meaning and structures of data is an important thing. The challenge is to integrate resources in the tourism domain. Data is represented in different vocabulary and different perspective. For example, when one hotel specifies the presence of an "indoor swimming pool", while another just says "swimming pool" (Fodor & Werthner, 2005). Hence many data providers are unreliable (Jeong, 2011). By semantic interoperability, systems are able to exchange data in such a way that the precise meaning of the data is readily accessible and the data itself can be translated by any system into a form that it understands.

- **Semantic Search**

User in the domain of tourism typically uses the search feature in the web for travel information. Several information retrieval techniques (e.g. for our case Search Engines Google, Yahoo or Bing) are used to retrieve the information relevant to users' requests. Unfortunately, most of these search engines let users search based on keywords and which, numerous queries must be issued in order to achieve the best results. Usually it is very time consuming to exploit and to analyze millions of pages of retrieved information. Users are often puzzled when processing their search results. They need information more suitable for their needs.

The problem is that words can be synonymous (that is, two words have the same meaning) for example: accommodation/housing or polysemous (a single word has multiple meanings) for example: Book, as a reserve accommodation or the written work consisting of pages. Semantic interoperability is a solution for this problem (Cardoso, et al., 2006a). Semantic interoperability is concerned not just with the packaging of data (syntax), but the simultaneous transmission of the meaning with the data (semantics). This is accomplished by adding data about the data (metadata), linking each data element to a controlled, shared vocabulary. The meaning of the data is transmitted with the data itself, in one self-describing "information package" that is independent of any information system. It is this shared vocabulary, and its associated links to an ontology, which provides the foundation and capability of machine interpretation, inference, and logic (Domingue, Fensel, & Hendler, 2011).

Therefore, if the languages used to describe the web pages were semantically interoperable, then the user could specify a query in the terminology that was most convenient, and be assured that the correct results were returned, regardless of how the data was expressed in the sources.

- **Machine readable**

Tourism is a very information intensive industry. Human and tourism agents manipulate and manage this information, and machine support is limited. Enabling machines to exchange and automate processing may became a reality by adapting semantic technology. The idea is to provide the information in a way that computers can understand, so it will allow software agents to analyze the web on our behalf, and produce results more relevant to users' needs. Semantic Technology provide the aforementioned promises with the assist of Metadata, Resource Description Framework (RDF) (RDF, 2002), and the Web Ontology Language (OWL)(Cardoso, et al., 2006a; OWL, 2004).

2.3.1 Metadata

Semantic Web technology is seen as the next generation of web systems, by providing better information retrieval, better services, and enhanced interoperability between different information systems. Metadata is one of the core components for this technology. The idea behind metadata is adding a knowledge level to the data in order to clarify how the data can be used. Describing a resource with metadata allows it to be understood by both humans and machines in ways that promote interoperability. Furthermore, it ensures efficient and reliable sharing and exchange of contents between various data repositories(Allinson, 2008).

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource(National Information Standards Organization, 2004). In simple words, metadata is data about data. It ensures that resources will survive and continue to be accessible into the future. According to the World Wide Web Consortium (W3C), metadata is machine-readable information for the Web. Metadata can describe resources at any level of aggregation, and it enriches resource discovery, interoperability, archiving, and preservation. A simple example is shown in Table 2.4 and Table 2.5. In Table 2.4 we present the traditional sample bookstore data that contains the elements (Title, Language, ISBN, Author), adapted from the Dublin Core Metadata Initiative (DCMI). DCMI is a standard for defining metadata documents (Powell, Nilsson, Naeve, & Johnston, 2005).

Table 2.4 Representing Sample Bookstore Data

Title	Language	ISBN	Author
Alice in Wonderland	English	123456789X	Lewis Carroll

Table 2.5 Metadata represented in XML syntax adapted from Dublin Core Metadata Initiative (DCMI)

```
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
              xmlns:opf="http://www.idpf.org/2007/opf">
          <dc:title>Alice in Wonderland</dc:title>
          <dc:language>en</dc:language>
          <dc:identifier id="BookId" opf:scheme="ISBN">
           123456789X
          </dc:identifier>
          <dc:creator opf:role="aut">Lewis Carroll</dc:creator>
</metadata>
```

In Table 2.5, the representation of the Bookstore Metadata is in XML format. The dc prefix indicates the Dublin Core specification, and the description of each element is shown in the next page:

Metadata Tag		**Description**
dc:title	:	The title of the document, or the name given to the resource.
dc:language	:	Locale, languages used in document.
dc:identifier	:	Text defining a unique identifier of the resource.
dc:creator	:	The authors of the resource.

2.3.2 Resource Description Framework (RDF)

RDF is a standard model for data exchange on the Web; the W3C has developed the RDF (RDF 2002) language to standardize the definition and the use of metadata. RDF has data merging features even for different schemas, and it supports schemas update without changing the data.

The basic structure of the RDF uses the URI (Universal Resource Identifier) to identify the relation between objects as well as the two ends of the link. The relation and the two ends come in the form **<subject, predicate, object>** usually referred to as a RDF "triple" (RDF, 2002).

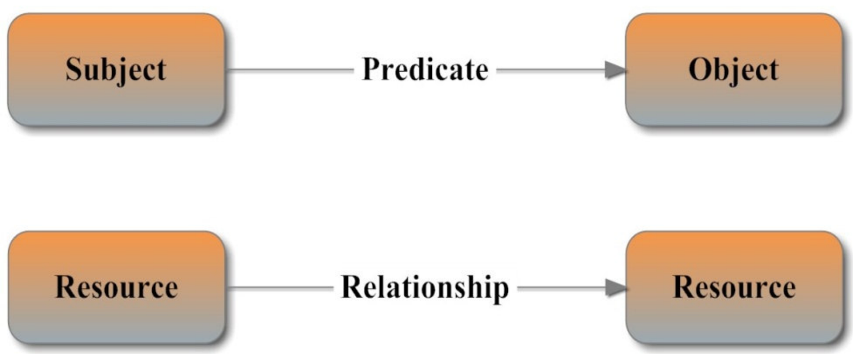


Figure 2.3 RDF Triple

As shown in Figure 2.3, the subject represents the resource; the predicate expresses a relationship between the subject and the object, while the object is the object (another resource or a literal) of this relationship. RDFS is a language for describing the RDF vocabularies in RDF. It has mechanisms to describe RDF classes and properties, such as attributes of resources and relationships between them. RDFS provides a mechanism in which multiple metadata schemas extracted from distributed information can be integrated.

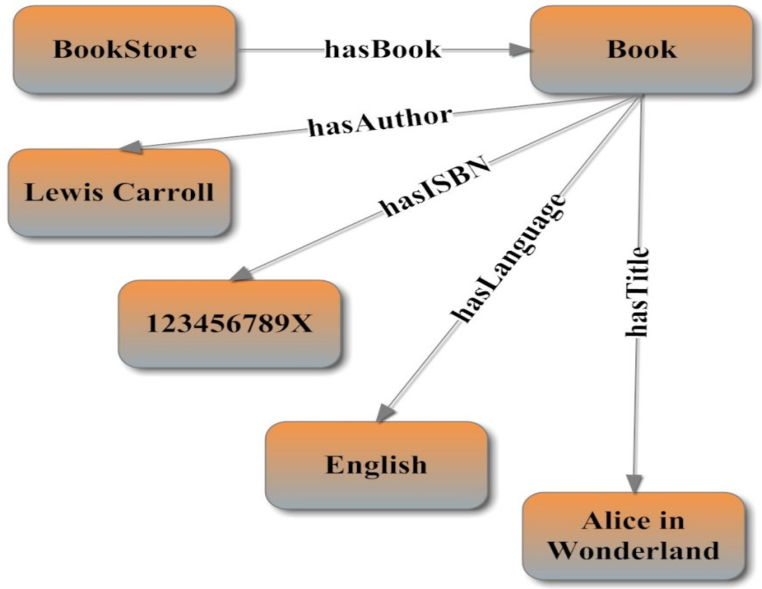


Figure 2.4 RDF representation of DCMI example in Table 2.5

Figure 2.4 illustrates the RDF representation of the data in Table 2.5. RDF is a standard way of interpreting the XML-encoded descriptions of resources. The nature of RDF representation as a logical tree covers all possible representations of the description. In comparison, XML describes elements in order tags each of which would be mapped into different RDF logical trees. However, querying RDF is easier as there is no restriction in the order of the elements as there is in XML. Moreover building a query for RDF representation is independent of the logical tree scheme, where in XML building a query should follow the schema of that XML documents.

RDF is therefore the formal data model for machine-understandable metadata used to provide standard descriptions of Web resources for facilitating data and system integration and interoperability (Cardoso & Sheth, 2006b).

2.3.3 Semantic Modelling and Development

Ontology is a term borrowed from philosophy that refers to the science of describing the kinds of entities in the world and how they are related (W3C OWL Working Group, 2009). In information science, Ontology is a data model that formally represents knowledge of a real-world domain. The task of inferring new knowledge from facts and rules is expressed in an ontology language so it can be used to reason about the most important concepts of that domain, their attributes and relations between concepts.

Some of the reasons to develop an ontology are (Noy & McGuinness, 2001):

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge

An Ontology plays an essential role in realizing the Semantic Web, which is a set of shared, explicit and formal concepts used to organize and classify contents. In addition to the advantages of share-ability and reusability, ontology offers more powerful enhanced facility for representing domain knowledge(C. Choi, et al., 2009)

The most prominent ontology language is OWL (OWL, 2004). It is a semantic markup language for ontology representation. It is resulting from the merging of DAML + OIL languages to extend the RDF syntax. DAML (DARPA Agent Markup Language) was created as part of a research program (www.daml.com) by the United States of America's governmental research organization. OIL (Ontology Inference Layer) was created by the European Union program for information society technologies.

OWL is identical to RDFS with a much greater degree of inference. It provides three increasingly expressive sublanguages: **OWL Full** which was designed for maximal RDF compatibility and uses all the primitives of the language; **OWL DL** (Description Logic) includes all OWL language constructs, but they can be used only under certain logical restrictions and **OWL Lite** a sublanguage of OWL DL that uses a simple restriction. Choosing between these sublanguages depends on the amount of reasoning support, modelling facilities and expressive constructs required (McGuinness & Van Harmelen, 2004).

The semantic web initiative pursues the goal of creating data and metadata in such a way that not only humans but also machines can make use of it. The idea is that the meaning of the data should be expressed in a format which enables it to be processed by computers. Towards this goal, most systems make use of ontologies to describe their data or metadata.

2.3.4 Ontology Query Languages, Reasoning and Tools

- **Query Languages**:

  Various Ontology query languages were developed to query the ontologies in order to retrieve knowledge (ZHANG, 2005). SPARQL is a query language for querying RDF documents (Prud'hommeaux & Seaborne, 2007). It is considered as one of the key technologies of semantic web and has gained standard recommendation by W3C. SPARQL has the capabilities for querying required graph patterns along with their conjunctions and disjunctions. The results of SPARQL queries can be in a result sets or RDF graphs. For instance to query the Author in our Bookstore Data using the SPARQL:

```
PREFIX abc: <http://um.edu.my/exampleOntology#>
SELECT    ? Author
WHERE
     {
              ?Book  abc:title  " Alice in Wonderland "
     }
```

The results of this query will be as follows:-

Author
Lewis
Carroll

SPARQL-DL is a substantial subset of SPARQL with clear semantically based OWL-DL. SPARQL-DL was introduced in (Sirin & Parsia, 2007) as a query language for OWL DL ontologies. SPARQL-DL has the ability to retrieve the two types of queries:

- **TBox**: "A TBox (for terminological knowledge, the basis for T in TBox) is a "terminological component"; that is, a conceptualization associated with a set of facts. TBox statements describe a conceptualization, a set of concepts and properties for these concepts. The TBox is sufficient to describe an ontology" (TechWiki, 2012).

- **ABox**: "An ABox (for assertions, the basis for A in ABox) is an "assertion component"; that is, a fact associated with a terminological vocabulary within a knowledge base. ABox are TBox-compliant statements about instances belonging to the concept of an ontology" (TechWiki, 2012).

SPARQL-DL query language queries the ontology to retrieve triples stored in the OWL-DL format. SPARQL-DL is a significantly expressive language which particularly allows mixing TBox, and ABox queries.

Table 2.6 Examples of TBox (A) and ABox (B) queries in SPARQL-DL

A

TBox
All *Books* Classes *Class(?Books)*
All Data Properties of *Books* *DataProperty(?Books)*
All Object Properties of *Books* *ObjectProperty(?Books)*

B

ABox
All individuals of the class *Books* *Type(?thing, Books)*
All Books written by Lewis Carroll *Type(?y, Books),* *PropertyValue(?y,hasAuthor ,"Lewis Carroll")*
All Properties of class *Books* *PropertyValue(Books, ?p, ?x)*

In Table 2.6, we illustrate the example of TBox and ABox queries in SPARQL-DL as described in the previous paragraph. In column **A**, we show how to retrieve all classes, Data Properties, and Object Properties for the class "*Books"* by using the TBox Query Atoms. In column **B**, we show how to retrieve all individuals under the specific class

"*Books"*, All individuals (*Books*) written by "*Lewis Carroll"*, and All Properties (relations) of the class "*Books"*, column B is the ABox Query Atoms.

- **Reasoning**

  When a knowledge engineer models a domain, the engineer constructs a terminology, by defining new concepts, possibly in terms of others that have been defined before. During this process, it is important to find out whether a newly defined concept makes sense or whether it is contradictory. Checking satisfiability of concepts is a key inference.

  Formal systems provide the ability to deduce new sentences from existing sentences using specific inference rules. This ability, referred to as reasoning, is an essential component of Semantic Web ontology formalism. (Baader & Nutt, 2003; Volz, 2004). For instance, in order to check whether a domain model is correct, or to optimize queries that are formulated as concepts. We can distinguish some of the relationships (assertions) between concepts as follows (Pollock, 2009):

  - **Equivalence**

    Equivalence assertions state that two things are the same. Equivalence can be asserted for classes, properties, and individuals. For instance, in publishing, there is no distinction between author and creator. By asserting the Author concept is equivalent to the Creator concept, anyone who is an Author is also a Creator, and vice versa. This is important because when asking the question "What are all known Authors in this system?" the answer includes both authors and creators in one query. Additionally, the property “hasCreator” has the same meaning as “hasAuthor".

For example, Jane is an author and we want to know what books he has written. If the model relates Jane to one book with hasAuthor, and another book with hasCreator, asking the question “What books hasAuthor Jane?” results in both books being returned. Asserting equivalence should be done very carefully and is usually performed for resolving issues that come about from integrating different modelling contexts. In most practical situations, the description logics–based OWL reasoning system should be relied upon to determine whether concepts are equivalent.

- **Disjointness**

  Disjointness assertions explicitly state that two things are different and not necessary “opposite.” It means only that two things are not the same. Disjointness can be asserted between concepts. So, asserting that two concepts are disjoint states that any member of one concept cannot be a member of the disjoint concept. This means that disjoint concepts can have no common members. In other words, if two concepts are asserted to be disjoint, the OWL reasoner will always conclude that those concepts are not equivalent.

  Consider for the publishing example, the Author concept is asserted to be disjoint with the Book concept. So the “Book-Title” as a member of the Book concept will not be in any case a member for the Author concept and equivalent to the member “Author-Name”. In the previous example, if there is no disjoint assertion on “Book-Title”, an OWL reasoner takes into consideration that the two members could be equivalent, and in fact concludes that they are satisfiably equivalent. As with the equivalent assertion example, this is an important conclusion. It may be a strong indicator that there are inconsistencies in the data that may need resolving.

- **Tools**:

  Building ontologies is considered as a huge and complex task that requires a lot of time and manpower. Consequently, during the last decade communities and research groups build different tools (as shown in Table 2.7) aiming to facilitate the process development and the reuse of ontologies. As a result, a number of tools came to the surface with different purposes and interfaces that help users carry out their development tasks (Gómez-Pérez, Fernández-López, & Corcho, 2004). In an ontology tools survey Perez et al. (2002) had classified tools into development tools, evaluation tools, merge and alignment tools, ontology-based annotation tools, querying tools and inference engines, and learning tools. Moreover, in a comparative study with the help of an evaluation framework, Su & Ilebrekke (2006) had found the most relevant tools to facilitate the development of ontologies. They are listed in Table 2.7 with a summary description, the name of the tool; reference to the developers, and the additional special purposes beside the editing and creating of the ontology.

Table 2.7 List of Ontology Tools

Ontology Tool	Developed by	Special Purposes
Ontolingua	(Farquhar, Fikes, & Rice, 1997)	To ease the development of Ontolingua ontologies in a shared environment between distributed groups
WebOnto	(Domingue, 1998)	To support the collaborative browsing, creation and editing of ontologies
Protégé	(Noy, Fergerson, & Musen, 2000)	To support the graphical software development environment.
OilEd	(Bechhofer, Horrocks, Goble, & Stevens, 2001)	To provide consistency checking functions and automatic concept classifications
OntoEdit	(Sure et al., 2002)	To ease the development in a plug-in architecture
WebODE	(Arpirez, Corcho, Fernandez-Lopez, & Gomez-Perez, 2003)	To support the access services by services and applications plugged in the server

For the purpose of this research selection, we look at Protégé which is an open source standalone application written in Java and provides a plug-and-play environment that specifically supports an OWL editor and reasoner. As shown in Figure 2.5 Protégé OWL plug-in provides a graphic visualization of the classes and properties using different colour codes to help developers distinguish between different types of classes (Breitman, Casanova, & Truszkowski, 2007).

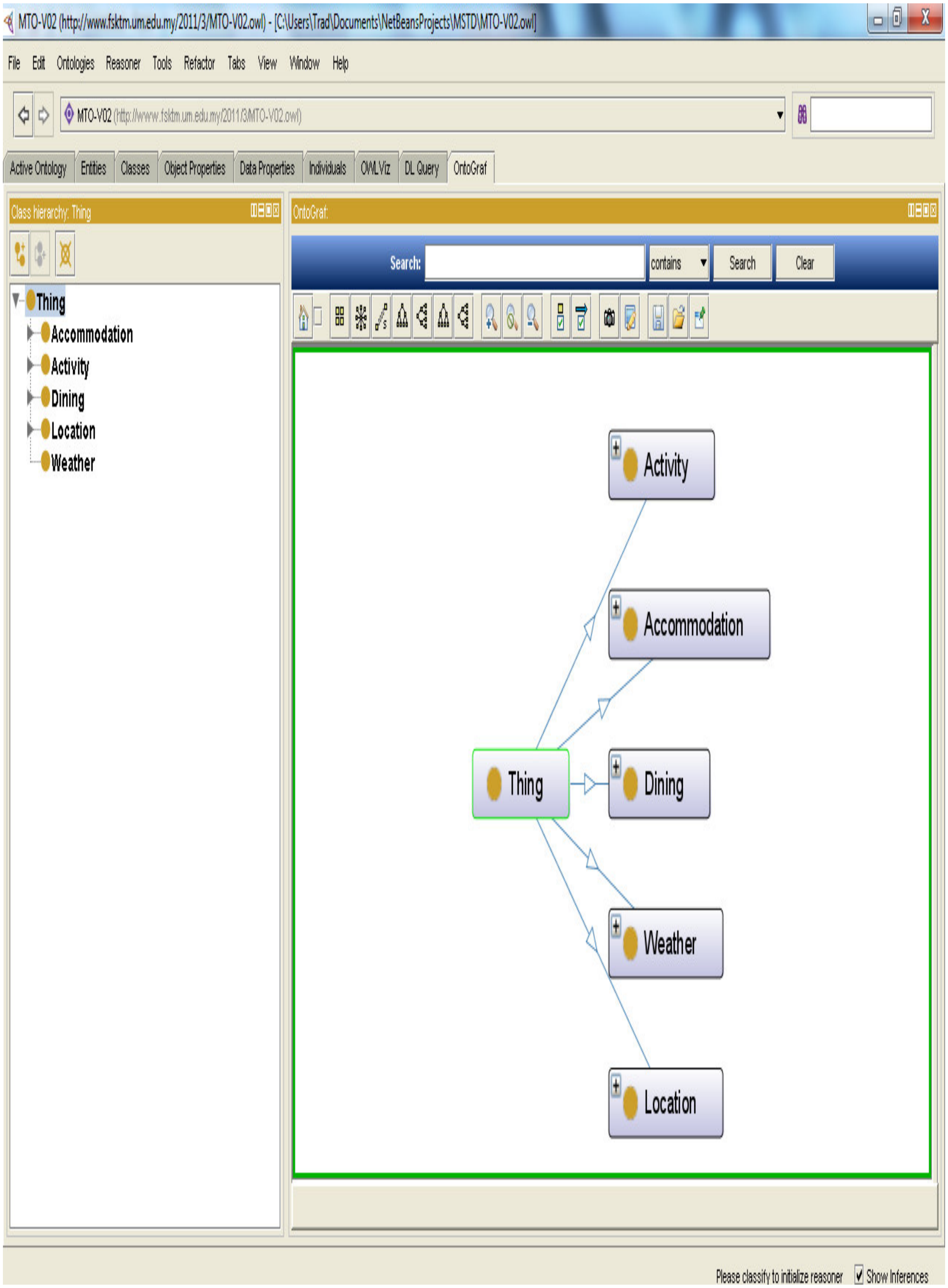


Figure 2.5 Protégé 2000 OWL Graphic Visualization View

2.3.5 Example of Ontologies in the domain of Tourism

- **Harmonise Ontology** was created within the Harmonise Project which is an EU Tourism Harmonisation Network (THN), and then taken over by the DERI E-Tourism Working Group in 2004. It specializes in data exchange (in the tourism domain) to tackle the problem of information interoperability and the aim is to allow Information Systems to cooperate without requiring modifications on their data. The Harmonise Ontology initially covered accommodation, events and activities (Fodor & Werthner, 2005). The project emphasises on the combination of a social consensus process with the application of new technologies. The goal is to allow participating tourism organizations to keep their proprietary data format while cooperating with each other. Specific tourism mediators are dedicated to the "translation" needs between these data sources. Such a mediator looks at information from a higher conceptual semantic level using this level of abstraction for the mapping purpose. Figure 2.6 describe the Harmonise process.

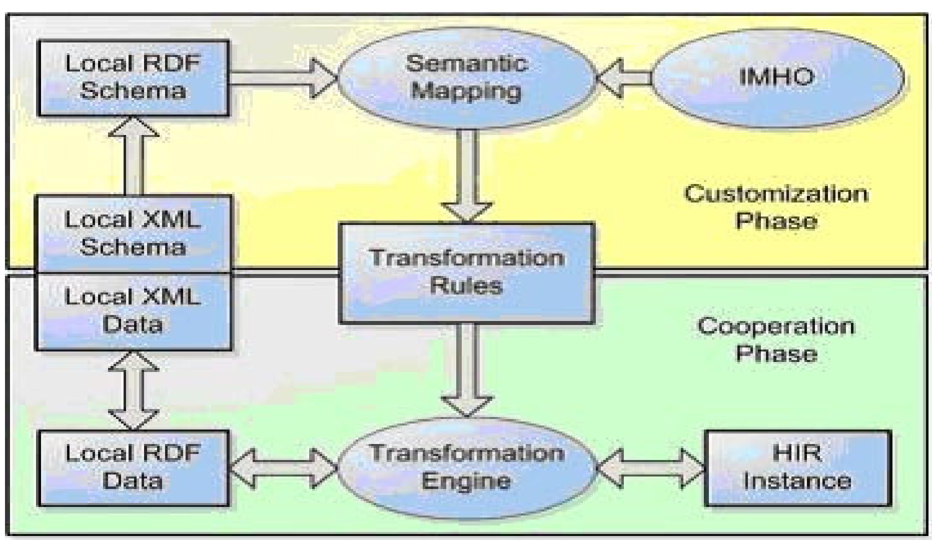


Figure 2.6 The Harmonise process

The data model of a source document, assuming XML, is first lifted to a local conceptual schema (C-Normalization) and then semantically mapped to the terminology specified by the shared ontology, which is built by domain experts. The output of the mapping process is a set of reconciliation rules, which are used in order to transform the local data and to code them according to the ontology content. Harmonise is based on RDF(S) as "language" for representing local conceptual schemata as well as the mediating ontology. However, Harmonise has some limitations: namely, the mapping between different conceptual models of Harmonise does not support ontology reasoning (the importance of reasoning is discussed in 2.3.4), and needed a domain which is based not only on true axioms or facts (Werthner, 2003).

- **The Hi-Touch Ontology** (Mondeca, 2012) was developed during the IST/CRAFT European Program Hi-Touch, which aimed at establishing Semantic Web methodologies and tools for intra-European sustainable tourism. The goal was to formalize knowledge on travelers' expectations and to propose tourism products. The ontology was mainly developed by Mondeca and is encoded in the ontology language OWL. This ontology has the largest number of concepts: it has 1000 concepts including terms for tourism object profiling, tourism and cultural objects, tourism packages and tourism multimedia content. However, the vocabulary of the available tourism ontologies covers a limited set of concepts often describing the domain from different perspectives due to the restricted application (i.e. in French language) scope from which the ontologies have been elicited. Figure 2.7 describe the main concepts in the Hi-Touch Ontology.

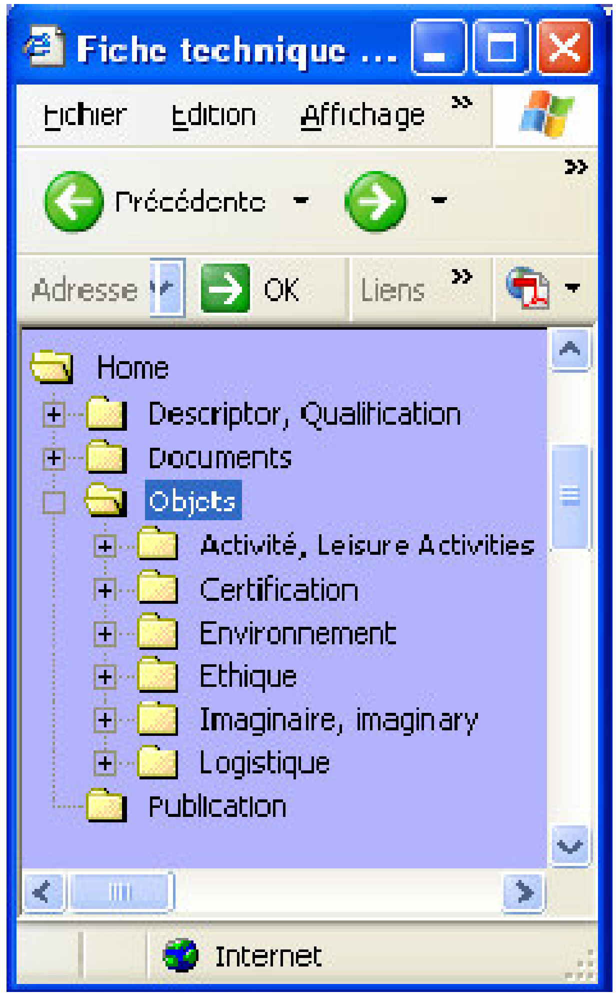


Figure 2.7 Hi-Touch Ontology

- **OnTour Ontology** (Prantner, 2005) is an ontology created especially for the tourism domain as part of the project "OnTour – A Semantic Web Search Assistant". The project consists of different components: There is the OnTour ontology which was designed especially for this project; a knowledge base containing data about tourism providers relying on the structure of the knowledge model, that is the ontology. In order to retrieve data from the knowledge base an inference engine is used; the logic of the system is implemented in a core program in the JAVA language; finally, the user interface is a common HTML website, with an interface to search and to present results. Figure 2.8 visualize the components.

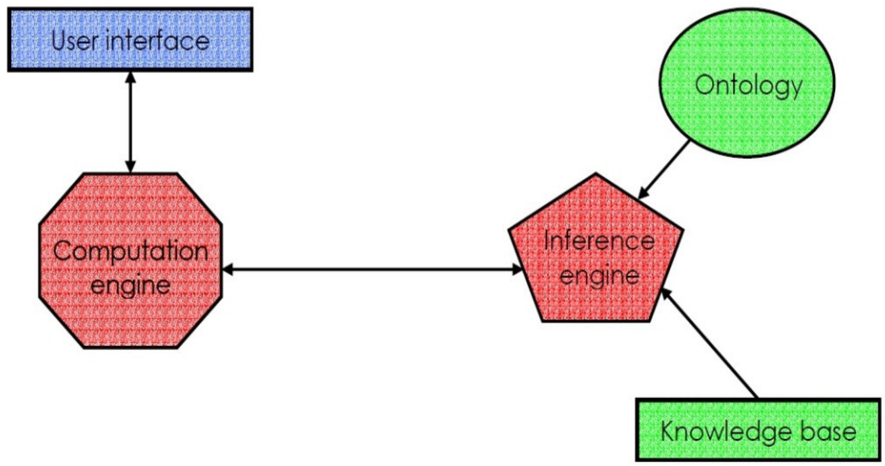


Figure 2.8 OnTour System Design Overview

The main focus is the description of accommodations, infrastructure, activity and concepts that describe leisure activities and geographical data. The ontology language used is the OWL-DL. The drawback of the OnTour ontology is in defining many properties as Boolean values. For example the property *hasPhone* is defined as Boolean, defining properties as Boolean is not compatible to NLI. In NLI the query would be in the form of "what is the phone number of Time square?" and a Boolean value (True or False) would not provide the answer for this question.

2.4 Natural Language Interfaces

Ontology and semantic web are becoming the essential method to represent domain-specific conceptual knowledge; the need to make these ontologies accessible to end-users become vital as the amount of information stored in the ontology-based knowledge bases steadily increases. Hence, specific research in developing interfaces for accessing structured knowledge has been advanced in the past couple of years. For instance, faceted search, menu-guided, form-based and graphical query language interface (Damljanović, 2011).

According to the interface evaluation conducted in Kaufmann and Bernstein (2007), systems developed to support Natural Language Interfaces (NLI) are perceived as the most acceptable by end-users. Familiarity with the natural language used in these systems is a key to simplify the information retrieval processes. NLI often referred as closed-domain Question Answering (QA) systems (Damljanovic, Agatonovic, & Cunningham, 2010), have a very important role as they are intuitive for the end users and are preferred to keyword-based, menu-based or graphical interfaces. Nevertheless, NLI provides a familiar and convenient means of query access to Semantic Web data for casual end-users. Several studies have shown that NLI can achieve high retrieval performance as well as domain independence (Kaufmann & Bernstein, 2007).

In the domain of tourism a Natural Language Query Interface for Tourism Information work is proposed (Dittenbach, Merkl, & Berger, 2003). In this work Dittenbach and his group have described a multilingual natural language database interface. The interface allows queries to be formulated in two languages: German and English. The query is detected using a text based classification approach with a spell checker to correct the typing errors. After analysis the concepts show that they are transformed into SQL statement. This type of systems is categorised under NLI to Database.

2.4.1 NLIDB: Natural Language Interfaces to Databases

NLI have been in existence since the late 60s, with early NLI systems developed to allow users to ask interesting questions about well-structured data sets. BASEBALL answered questions about baseball games played in the American league over one season (Green, Wolf, Chomsky, & Laughery, 1961), and LUNAR answered questions about the analysis of rock samples from the Apollo moon missions (Woods, 1997). Both LUNAR and BASEBALL are examples of what have been described as natural language interfaces to databases (NLIDB), that is, their source of information was a database that contained the relevant information about the topic. The user's question was converted into a database query, and the database output was given as the answer.

Two well known NLI are ELIZA (Weizenbaum, 1966) and SHRDLU (Winograd, 1972), ELIZA one of the first programs to respond to natural language in a seemingly intelligent interaction, simulated a conversation as if the user is talking with a psychiatrist. ELIZA was able to converse on any topic by resorting to very simple rules that identified important keywords in the conversation, the program rank these keyword and search through the rules to find the appropriate response to the user. SHRDLU maintain a simulation of a robot in blocks world domain, and it offered the possibility to ask the robot to move things around in the world in response to user commands expressed in the natural language. The specific domain and a very simple world with rules of physics made SHRDLU a landmark program. The difference between these systems and systems such as LUNAR and BASEBALL are their ability to carry out a dialogue.

Since the 1960s, the progress and success of NLI has been mediocre. During the 70s and 80s there was great effort made in the development of theoretical bases for computational linguistics. The use of NLIDB, however, did not gain the expected rapid and commercial acceptance mainly due to the emergence of friendly graphical and form-based interfaces (Androutsopoulos, Ritchie, & Thanisch, 1995). Several NLI prototype systems appeared at that time such as (e.g., Ladder, Chat-80, Janus), but none were extremely impressive or radically different from one another. As a result, the interest in NLIs died down in the 1990s and the research was no longer as impressive as in previous decades. Notable work in Tourism domain includes:

2.4.2 Semantic ontology-based Natural Language Interfaces

The need for successful NLI became more acute in the early 2000s as the amount of information stored tremendously increased, and the need to search and query these information become popular amongst nontechnical users that want to access a wide range of repositories through web browsers, PDAs, cell phones, etc. (Kaufmann & Bernstein, 2007; Popescu, Etzioni, & Kautz, 2003). The rising popularity of semantic web and the emergence of large-scale semantics added to the reignition in the interest of NLI. The necessity of NLI to ontology-based repositories has become more acute, as casual users are typically overwhelmed by the formal logic of the semantic web and prefer to use a NL interface to query the ontology(Kaufmann & Bernstein, 2007). The main role of using NLI to ontology is to transfer the human natural language such as English into a structural computer language such as OWL. Moreover, NLI and ontology approaches rely both on string similarity. In this research string similarity approach is based on Levenshtein distance (Levenshtein, 1966).

The Levenshtein distance between two strings is the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution of a single character. Scores vary in the range of 0 to 1. It determines the relatedness of two strings in terms of the number of insert, remove, and replacement operations to transform one string str_1 into another string str_2. This edit distance is defined as $xform(str_1; str_2)$. As a normalization factor, the worst case transformation cost $xform_{wc}(str_1; str_2)$ is calculated replacing all parts of str_1 with parts of str_2, then deleting the remaining parts of str_1, and inserting additional parts of str_2. The final similarity between str1 and str_2 is calculated by:

$$sim_{lev}(str_1; str_2) = 1 - \frac{xform(str_1; str_2)}{xform_{wc}(str_1; str_2)}$$

turning the normalized edit distance into a similarity score.

For instance, if in a query 'list cities located in Europe', identified key concepts would be cities and Europe, the first referring to the class City, and the latter referring to an instance of the class Continent, the text given between these concepts (located in) will be compared with names of all defined properties between identified concepts. If the property with name locatedIn is present in the ontology, according to Levenshtein distance of the two strings 'locatedIn' and 'located In' will be similar.

There are a number of studies that provide natural language interfaces to ontologies. Notable works include:

ORAKEL by Cimiano and colleagues (2007) is a portable NLI to knowledge bases that shared a general lexicon, where for example words such as “what” or “which” are stored. A part of the domain specific lexicon is created automatically from the domain ontology and is called ontological lexicon. Another part of the domain specific lexicon is created manually and contains mappings of subcategorisation frames to relations, as specified in the domain ontology. Subcategorisation frames are essentially linguistic argument structures, e.g. verbs with their arguments, nouns with their arguments, etc. For example, a verb “write” requires a subject and an object, as it is a transitive verb. This ’triple’ of subject-verb-object in this case could be considered a subcategorisation frame, and could be mapped to an ontology relation “writes”. Subcategorisation frames are created by the person in charge of customizing the system. He does not have to be familiar with computational linguistics, although he is expected to have basic knowledge of subcategorisation frames. The adaptation of the NLI is performed in several iterative cycles in the user interaction sessions, based on the questions which the system fails to answer. In this way, the coverage of the lexicon is being increased each time. ORAKEL supports “WH” type questions such as *Who*, *What*, *Where*, etc. Factual here means that answers are ground facts as found in the knowledge base, and not complex answers to *Why* or *How* questions that require explanation. Users in this system will be engaged in a dialog, until the system learns enough to be able to automatically suggest the correct answer. Figure 2.9 describes the design of ORAKEL system.

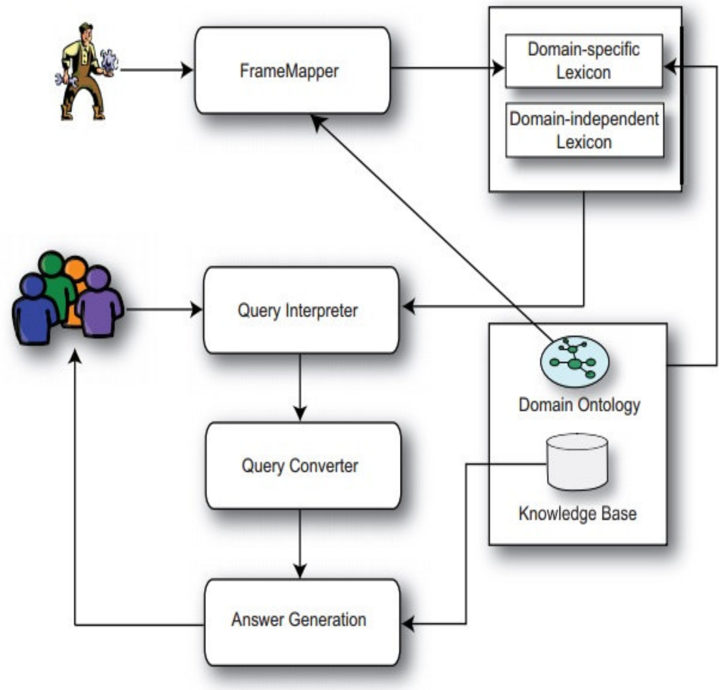


Figure 2.9 Overview of the ORAKEL system

1. **AquaLog** (Lopez, Uren, Motta, & Pasin, 2007) is a portable question-answering system which allows the user to ask Natural Language queries with respect to the universe of discourse covered by the ontology. With a controlled language, such as that used by AquaLog for querying ontologies, users can create "WH" type questions such as *What*, *Which*, *Who* and the like. The system works by converting the natural language query into a set of ontology-compatible triples that are then used to extract information from a knowledge store. It utilizes shallow parsing on the user natural language question by using a Java Annotation Patterns Engine (JAPE) grammars (Cunningham, Maynard, & Tablan, 2000) and other tools such as GATE. The General Architecture for Text Engineering (GATE) was developed at the University of Sheffield in 1995 and was used in several text mining applications (Maynard et al., 2000). GATE is completely written in Java, and it has a better

semantic understanding of text compared to other text mining tools like RapidMiner (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006) and Pimiento (Adeva & Calvo, 2006). GATE includes an information extraction system called ANNIE (A Nearly-New Information Extraction System). ANNIE produces syntactic annotations related to the user query. These annotations passed to JAPE grammars to identify terms, relations, and question indicators (who, what, etc.). Evaluation conducted by Damljanović and Bontcheva (2009) indicated that AquaLog suffer from low performance as affected by the level of complexity of the ontology structure.

2. **PANTO** (Wang, Xiong, Zhou, & Yu, 2007) is a portable NLI to Ontologies that translate natural language queries into SPARQL and executes a corresponding SPARQL query on a given ontology model.

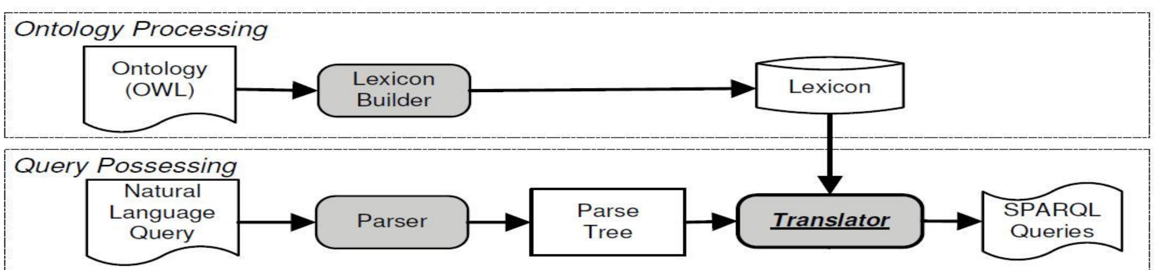


Figure 2.10 PANTO Architecture

Figure 2.10, describes the design of PANTO. First the system utilizes WordNet (Fellbaum, 2010) to build the lexicon from the ontology. Upon receiving user inputs in the form of natural language query, the StanfordParser (Klein & Manning, 2003) is used to produce the parse tree, the translator transforms the parse tree into a SPARQL query with the help of the Lexicon. According to (Wang, et al., 2007), the type of questions supported are not clear, although Wang et al (2007) expect that PANTO supports questions similar to that supported by AquaLog, as PANTO parsed and analyzed the 170 sample queries presented on the AquaLog web site (http://technologies.kmi.open.ac.uk/aqualog/).

3. **Querix** (Kaufmann, Bernstein, & Zumstein, 2006) is a domain-independent NLI that translates generic natural language queries into SPARQL. The idea behind Querix is using clarification dialogs to query ontologies without using any complex semantics based technologies. However, natural language ambiguities are solved by asking the user for clarification in a dialog window if an ambiguity occurs in the input query. The user interface allows the user to choose an ontology and enter full NL queries. The ontology in Querix is enhanced by obtaining synonyms from WordNet. The StanfordParser is used to provide a parse tree for the NL query. Querix require complete "WH" type questions such as "Which," "What," "How many," "How much," "Give me" or "Does" and end with a question mark or full stop.

The proposed prototype in this research will be evaluated based on the following:

2.4.3 Natural Language Interfaces Evaluation

NLI provide a familiar and easy query access to the Semantic Web data (Kaufmann & Bernstein, 2010). High-quality retrieval performance is an important aim of NLI systems as mentioned in this section. For decades, the two most frequent and basic measures for information retrieval effectiveness are precision and recall as shown in Figure 2.11 (Mandl, 2008; Manning, Raghavan, & Schtze, 2008; Zuva & Zuva, 2012).

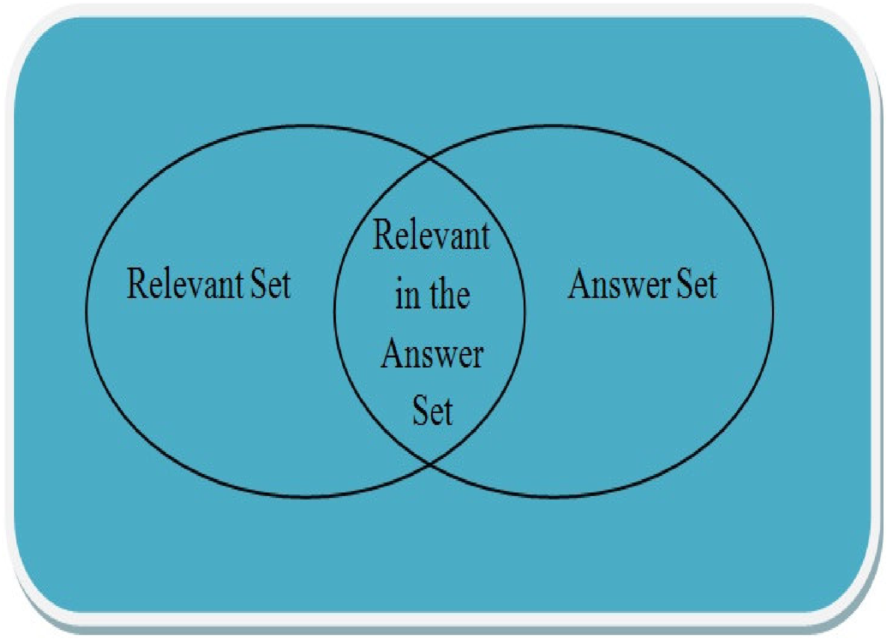


Figure 2.11 Diagram showing Precision and recall for a given query

- Recall- refers to the number of relevant answers returned out of the total number of relevant answers available in the ontology being searched. High recall means that the system returned most of the relevant answers. Within the scope of this research relevant answers are correct answers.

$$\text{Recall} = \frac{|\{Relevant\ Answers\} \cap \{Retrieved\ Answers\}|}{|\{Relevant\ Answers\}|}$$

- Precision- refers to the number of relevant answers returned to the total number of answers returned. High precision means that the system returned more relevant results.

$$\text{Precision} = \frac{|\{Relevant\ Answers\} \cap \{Retrieved\ Answers\}|}{|\{Retrieved\ Answers\}|}$$

To the user the scalar value of recall indicates the ability of the system to find relevant items as per query from the collection of different items and precision ability to output high matched relevant items as per query. In general the user is interested in the relevant retrieved items thus the measures of precision and recall concentrate the evaluation on the relevant output of the system. The lower the values indicates bad performance of the system and the higher the values the more the user is encouraged to use the system due to the anticipation of getting more of the relevant search items. These evaluation measures are inter-dependent measures in that as the number of retrieved items increases the precision usually decreases while recall increases (Mandl, 2008; Manning, et al., 2008; Zuva & Zuva, 2012).

In conclusion, the power of knowledge representation in ontologies enhances the performance of Ontology-based NLI as it analyzes users' queries more perfectly. It is observed, that Ontology-based NLI systems implement a wide range of components (e.g.; WordNet: lexical database; Sesame: architecture for querying RDF; Stanford: unlexicalized natural language parser) that support the development of the NLI. Comparing the NLI to database systems with Ontology-based NLI systems, there is no reliable comparison that can be made due to lack of standards that can govern the evaluation as different datasets (ontologies) and different evaluation measures are implemented in order to evaluate these systems separately (Damljanović, 2011).

Damljanović & Bontcheva (2009) reviewed several NLIs to KBs and reported on their performance and customisation issues. In order to give as objective a comparison as possible, we show on which dataset the system was evaluated, how the process of customisation was performed, and the recall and precision values. This section only covers a sub-set of NLIs to KBs, i.e., those that were reported in the evaluation results. A brief overall summary is shown in Table 2.8, subdivided by dataset, as no reliable comparison of precision and recall can be made across different datasets. The main conclusion to be drawn from this table is that although systems with zero customisation tend to have reasonable performance, it varies significantly across systems in general. The more complex the supported queries are, the lower the performance is.

Table 2.8 Natural Language Interfaces to Knowledge Bases, adapted from Damljanović & Bontcheva (2009)

Dataset	System	Precision	Recall	Portability
Geographical facts about Germany	ORAKEL	80.60-84.23%	45.15%-53.7%	customised
Software engineering ontology	AquaLog	86.36%	59.37%	0 customisation
Mooney: restaurants	PANTO	90.87%	96.64%	0 customisation
Mooney: geography	Querix	86.08%	87.11%	0 customisation

3.0 Solution Design

3.1 Introduction

The extensive range of tourism information available on the Internet is a challenge for the contemporary tourism industry across the globe. Information is one of the biggest needs for tourists. The search for information cost time, effort and resources, even when tourists need a small piece of information, a large quantity must be examined as aforementioned in the previous chapter discussion.

A Recommender system may prove to be efficient in solving the overload problem (Ricci, et al., 2011). Nevertheless, Semantic Technology is another solution to answer tourists' query expressed in the Natural Language without having to learn any formal query languages. Semantic representation plays the role of modelling the probability between Natural language and formal query language (Wang, et al., 2007).

The proposed approach is developed by employing a combination of different technologies for the benefit of the tourism industry. We selected Open Source software (protégé and J2EE) that has a freely-readable source code as this source code can be modified, improved, and tested. Hence, our prototype will not only answer tourists' question semantically, but it also enriches the semantic answers with a recommendation based on the tourists' preferences using interfaces that are easy to use (Kaufmann & Bernstein, 2007). Our definition of "easy to use" is the ability to use the system with little or no training.

This chapter describes in detail the architecture and design of the proposed approach and elaborate the implementation steps by using an example in the domain of Malaysian tourism. The chapter presents an overview of SMTRS. It gives insight into its strategy and underlying assumptions, its process architecture, and its main sub-processes. It consists of three components (Natural Language Interface, MTO and Recommender system based on Content-based filtering). The reasons behind selecting these components are discussed in the following:

3.2 SMTRS Architecture

SMTRS is a system aimed at understanding users' query and provides semantic answers from the Malaysian tourism domain based on the tourists' preferences. SMTRS is intended to be a service for foreigners to become familiar with the Tourism Cities in Malaysia and to help plan leisure activities. The SMTRS architecture is shown in Figure 3.1. SMTRS Architecture is built to solve the aforementioned challenges of ICT in Tourism domain (Information Overload, Poor Knowledge Representation and the Absence of Personalization and User Profile Utilization) as discussed in 2.1.4.

Information is considered as one of the biggest needs for tourists. If they have the appropriate information, it will help them in making their choices about (what to do, where to stay, and how to get there) the trip (Siricharoen, 2008). However, data providers represent the information in their own vocabulary. Pan and Fesenmaier (2006) found that the "language of tourism" (Dann, 1996) is extremely rich; further, their study indicated that the vocabularies used on destination marketing organization websites differed substantially from those of potential users. As such, they concluded that the richness in language and the differences in perspectives make it very difficult for Internet users to have a satisfying online search experience. The challenge in the tourism domain is that tourists fall victims to poor communication and lack of a common understanding (Cardoso, 2005). SMTRS Architecture uses a locally stored Knowledge Base (Malaysia Tourism Ontology). MTO was built in this research to represent the Malaysian Tourism information. Ontologies provide a good basis for reasoning and classifying the various information in the tourism domain. Such a representation, provide uniform definitions and therefore increase knowledge sharing, remove semantic ambiguity, and are fundamental to automated knowledge extraction on the Web (Gruber, 1995). Nevertheless, the goal of MTO is to serve as knowledge base and contain hundreds of tourism items (such as Hotels, Service Apartments, Activities, etc.) The next step is concerned with how to extract knowledge from the MTO without overloading the tourist with too many choices of tourism items.

SMTRS extract knowledge from MTO from three components as shown in Figure 3.1. **First**, the Natural Language Processing component analyze tourists' query to recognize the semantic relations in the query, and identify a query triple (triple e.g. hotels, has-location and location-BukitBintang). **Second**, the Ontological Entity Mapping component map the identified triple with the MTO contents by using string similarity (refer to 2.4.2) to provide semantic answers to the tourist's query. **Finally**, the Content-based Recommendation component refines the semantic answers (such as several Hotels in BukitBintang) to find which one of these semantic answers matches the tourists' preferences and recommend the best match to the tourist. Each step of the recommendation is performed by exploiting knowledge about the tourist (from the user profile), and knowledge about the contents (from MTO).

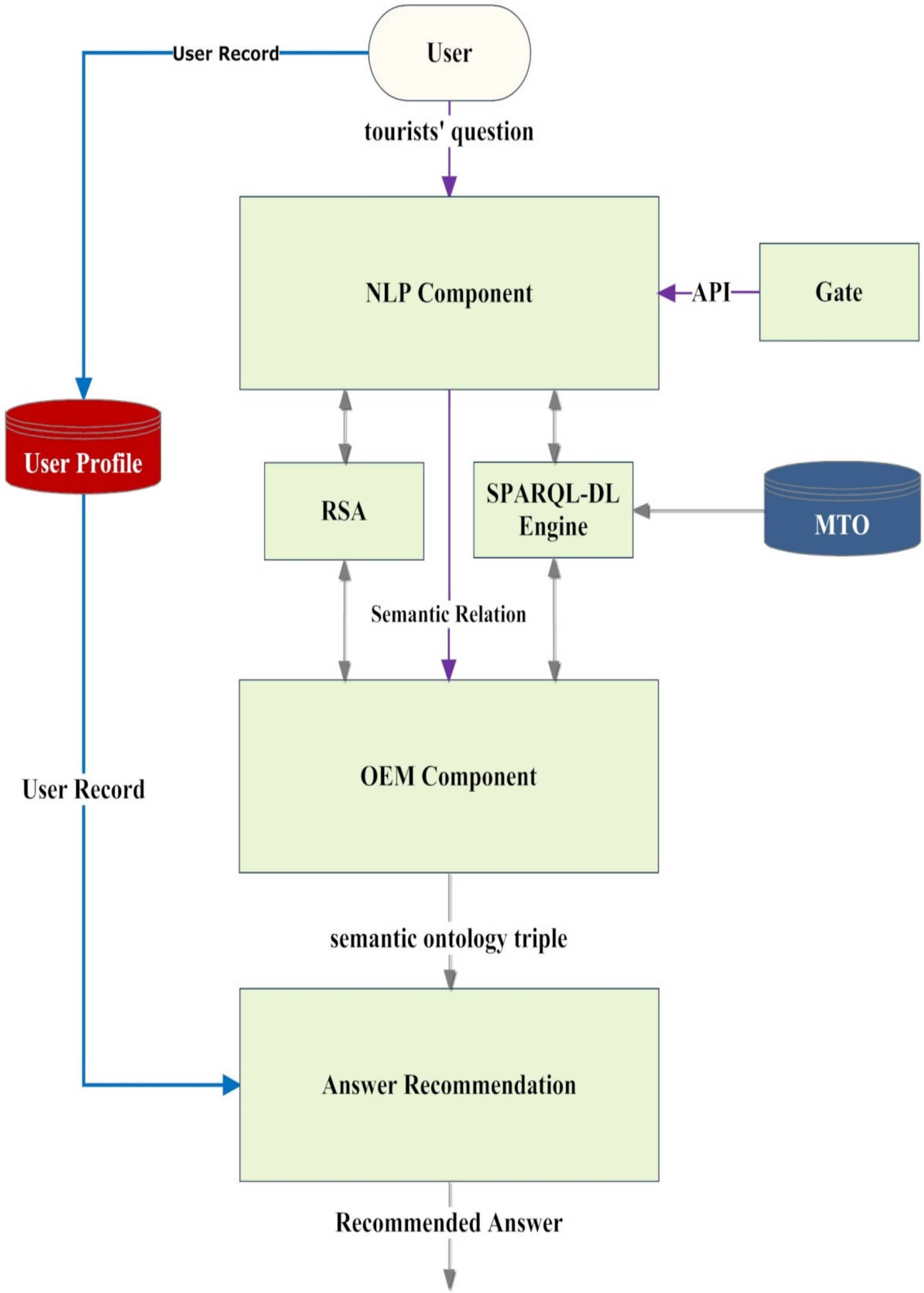


Figure 3.1 SMTRS Architecture

Terms	:	Definition
Relation Search Algorithm (RSA)	:	set of rules to tokenize user question and match it with MTO
Potential Relation (PR)	:	set of candidate terms as relations extracted from user question
Semantic Relation (SR)	:	set of Potential Relations matches relations in MTO
Ontological Entity Mapping (OEM)	:	set of Semantic Relations with related Triples
Semantic Ontology Triple (SOT)	:	set of OEM with terms similar to user question terms
Recommended Answer (RA)	:	set of SOT that matches the user preferences

3.3 System Input

3.3.1 User Profile

The tourism User Profile is a structured representation of the tourists' preferences. Implementing explicit User Profile has the advantage of letting users specify directly their interests (Middleton, et al., 2004). We created a user form for the users to closely match their needs in order for the SMTRS to build a recommendation. The users are asked to fill an online registration form when visiting the web site for the first time.

The literature reveals several tourism activities that may be classified into four main categories of goods and services namely (refer to 2.1.3), **Accommodation, Transportation, Activity** and **Food** (Schmidt-Belz & Poslad, 2003). In the SMTRS architecture our selection considers three of these categories, namely, **Accommodation, Activity** and **Food**. Murphy, Pritchard and Brock (2000) mentioned that visitors consume the products of a destination; therefore, the products must be something the visitor wants and needs (such as daily budget for determining the hotel). Also literature indicates that the food products of a country can be among its most important cultural expressions (Handszuh, 2003; Rand & Heath, 2006; Rand, Heath, & Alberts, 2003). Thus, the information in the user profile is used to generate those three types of recommendations. As shown in Figure 3.2, in order to recommend food for tourists coming to Malaysia two types of information that the tourist should fill in the user profile are type of food and preferred cuisine.

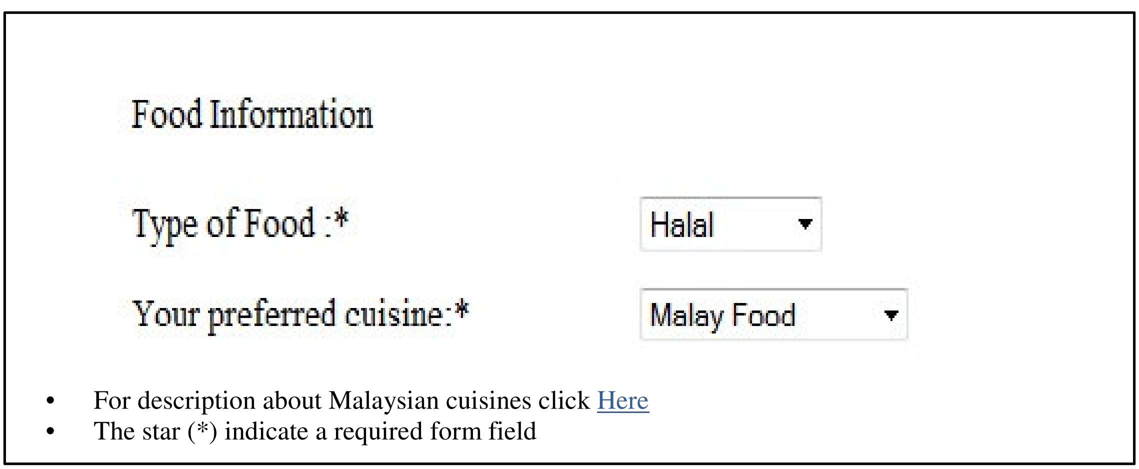
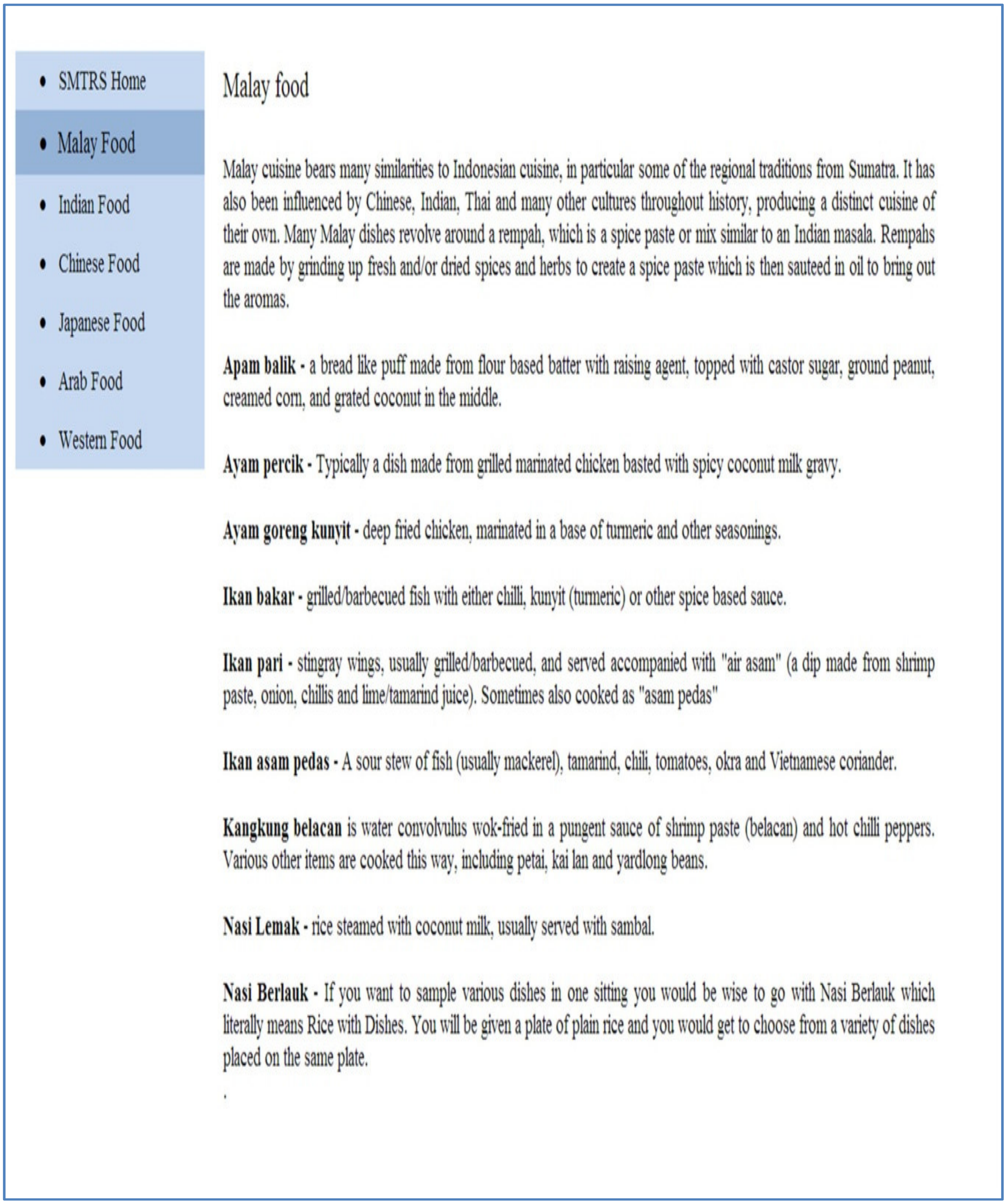


Figure 3.2 SMTRS User Online Registration Form (the Part of Food Information)



- SMTRS Home
- Malay Food
- Indian Food
- Chinese Food
- Japanese Food
- Arab Food
- Western Food

Malay food

Malay cuisine bears many similarities to Indonesian cuisine, in particular some of the regional traditions from Sumatra. It has also been influenced by Chinese, Indian, Thai and many other cultures throughout history, producing a distinct cuisine of their own. Many Malay dishes revolve around a rempah, which is a spice paste or mix similar to an Indian masala. Rempahs are made by grinding up fresh and/or dried spices and herbs to create a spice paste which is then sauteed in oil to bring out the aromas.

Apam balik - a bread like puff made from flour based batter with raising agent, topped with castor sugar, ground peanut, creamed corn, and grated coconut in the middle.

Ayam percik - Typically a dish made from grilled marinated chicken basted with spicy coconut milk gravy.

Ayam goreng kunyit - deep fried chicken, marinated in a base of turmeric and other seasonings.

Ikan bakar - grilled/barbecued fish with either chilli, kunyit (turmeric) or other spice based sauce.

Ikan pari - stingray wings, usually grilled/barbecued, and served accompanied with "air asam" (a dip made from shrimp paste, onion, chillis and lime/tamarind juice). Sometimes also cooked as "asam pedas"

Ikan asam pedas - A sour stew of fish (usually mackerel), tamarind, chili, tomatoes, okra and Vietnamese coriander.

Kangkung belacan is water convolvulus wok-fried in a pungent sauce of shrimp paste (belacan) and hot chilli peppers. Various other items are cooked this way, including petai, kai lan and yardlong beans.

Nasi Lemak - rice steamed with coconut milk, usually served with sambal.

Nasi Berlauk - If you want to sample various dishes in one sitting you would be wise to go with Nasi Berlauk which literally means Rice with Dishes. You will be given a plate of plain rice and you would get to choose from a variety of dishes placed on the same plate.

Figure 3.3 Description about cuisine available in Malaysia

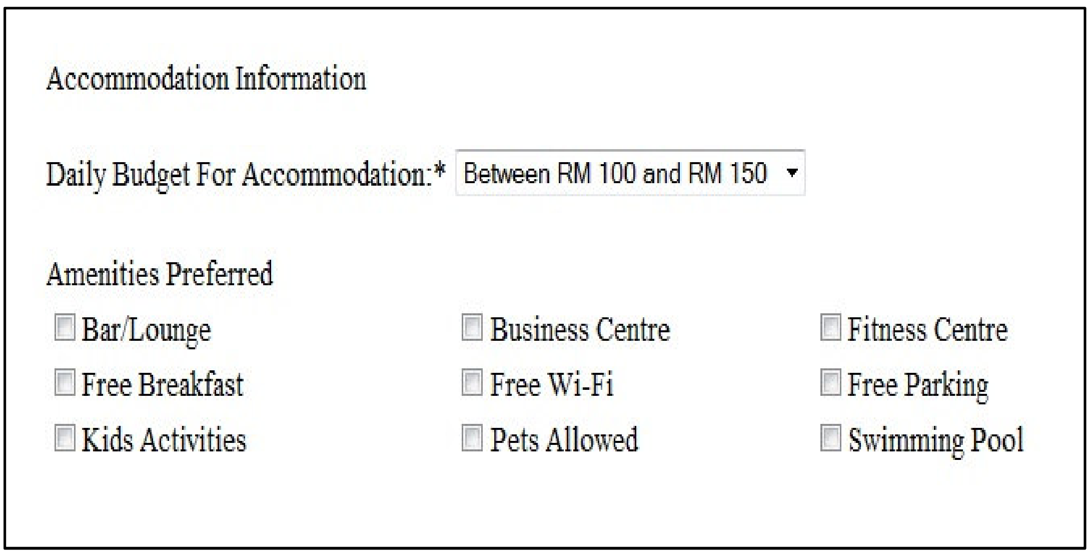
Figure 3.3 shows the Malaysian cuisine description provided for users who may not be familiar with the type of food in Malaysia.

- *Type of food:* the majority of tourists in Malaysia come from the Middle East, West Asia and China. In fact, most of them are Muslim (Samori & Rahman, 2013). Malaysia is trying to attract Muslim tourists from all over the world offering facilities in accordance with the religious beliefs of Muslim tourists. According to AFP (AFP, 2013) Malaysia has been rated the world's top Muslim-friendly holiday destination in a survey carried out by a Singapore-based Muslim travel consultancy Crescentrating; who ranked countries on how well they cater to the growing number of Muslim holidaymakers seeking Halal or Islam-compliant food and services. Thus the Malaysian type of food can be categorised as, Halal and Nonhalal; Halal foods are foods that are allowed under Islamic dietary guidelines, while Nonhalal foods are not allowed by Islamic dietary guidelines.
- *Preferred cuisine*: Malaysia has different restaurants which offer different cuisines, Many food websites in Malaysia (see Table 3.1) define their structure on the basis of various ethnic groups (Malay, Chinese, Indian, etc). The tourist should state the type of cuisine preferred in order for the system to recommend specialized restaurant.

Table 3.1 Food websites in Malaysia

	Website
1	http://food.malaysiamostwanted.com
2	http://www.travellersworldwide.com
3	http://en.wikipedia.org/wiki/Malaysian_cuisine#Food_types
4	http://www.malaysianfood.net/
5	www.foodmsia.com
6	http://malaysia.travel-culture.com/cuisine.shtml
7	http://www.bestmalaysianfood.com/

For the Hotel recommendation, different people have different preferences. To recommend accommodation that reflects the user needs of personalized service, SMTRS use two types of information concerned with the Accommodation characteristics (Ghose, Ipeirotis, & Li, 2012; Xiong & Geng, 2010) that should be included in the user profile. Figure 3.4 shows the Accommodation information needed in the User profile


Accommodation Information

Daily Budget For Accommodation:* Between RM 100 and RM 150

Amenities Preferred

Bar/Lounge	Business Centre	Fitness Centre
Free Breakfast	Free Wi-Fi	Free Parking
Kids Activities	Pets Allowed	Swimming Pool

Figure 3.4 SMTRS User Online Registration Form (the Part of Accommodation Information)

- *Daily budget for accommodation:* the tourist should select the budget for renting accommodation per night; recommendation will be based on rooms with the range of price of the selection.
- *Accommodation amenities preferred:* Accommodation may provide many types of amenities that a tourist may prefer. For instances, kids activities, Wi-Fi, free car parking, swimming pool, etc.

Finally, in order to recommend Activities for tourists coming to Malaysia the information that should be filled in the user profile is shown in Figure 3.5. One or more activities can be selected for the user profile.

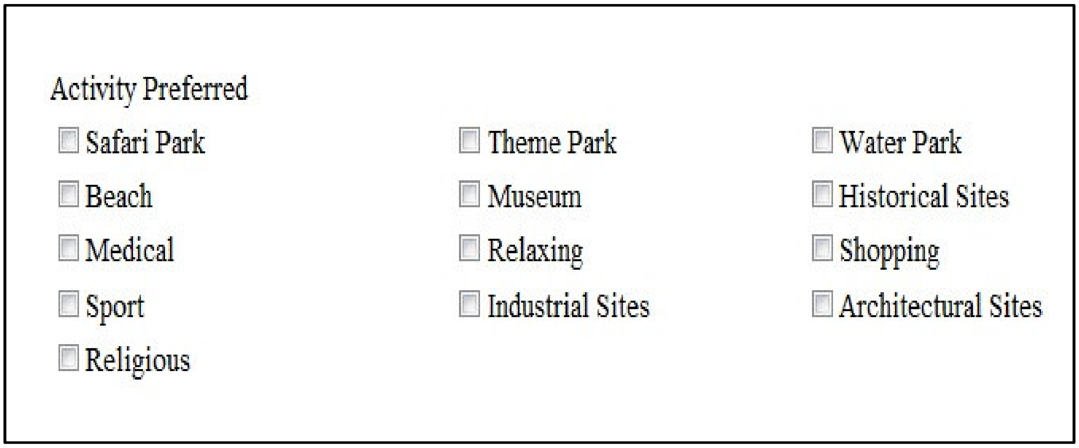


Figure 3.5 SMTRS User Online Registration Form (the Part of Activity Information)

- *Activity preferred:* we categorize different types of activities in Malaysia such as historical sites. The tourist can then select one or more activities in order to get recommendation that is close to their preferences.

3.3.2 Malaysia Tourism Ontology (MTO)

Representing the Malaysian tourism information is a core component of our prototype. The two prominent languages used in building the semantic ontology are RDF and OWL. OWL supports high expressivity in modelling and reasoning compared to the RDF (refer to 2.3.2 and 2.3.3). For our prototype, the ontology is written in OWL-DL to gain the required reasoning support. OWL-DL is based on formalizing the semantics in Description Logic.

OWL-DL is recommended by the W3C for the Semantic Web and it allows automatic reasoning and support Open World Assumption. Open World Assumption means that, what is not stated is regarded as unknown. For example when there is no information about the availability of free parking in the *x hotel* does not mean that there is no free parking, but that there is no information about this. Furthermore, the following steps describe the method for developing the MTO. We construct the MTO using Ontology Development 101(Noy & McGuinness, 2001).

According to Noy & McGuinness (2001), there is no absolutely one correct way or methodology for developing ontologies. However there are some fundamental rules in ontology design that can help to make wise design decisions. These are given as follows:

- There is no one correct way to model a domain- there are always viable alternatives. The best solution almost always depends on the application that one has in mind and the extensions that are anticipated.

- Ontology development is necessarily an iterative process.

- Concepts in the ontology should be close to objects (physical or logical) and relationships in the domain of interest. These are most likely to be nouns (objects) or verbs (relationships) in sentences that describe the domain.

Step 1: Determine the Domain and Scope of the Ontology

As described in architecture 3.1, our prototype interacts with users by using the natural language which is very complex, since the natural language query is converted into semantic triple (subject, predicate and object) and matched with the return triples of the ontology within SMTRS in order to satisfy the query.

However, the need for clear representation of the knowledge (i.e. concept, relations, individual, etc.) in the tourism domain will support the process of extracting and retrieving information. The development of the MTO considers the usage of this representation in order to answer the tourist's query as discussed in 2.1.3, such as (*what to do in Kuala Lumpur)* or (*where to stay in Bukit Bintang)*.

The domain of MTO is inclusive to the tourism in Malaysia, for instance: accommodation (including hotels and apartments), food (halal and nonhalal), food cuisine types, and activities (including adventure, beach, and historical events).

To define the domain and scope of MTO, we had to answer several selected questions, which had been suggested by the Ontology Development 101 (Noy & McGuinness, 2001) to help determine the following goals: purpose, usage, type of information, and who will need the MTO. Table 3.2 illustrates the questions & answers used:-

Table 3.2 Questions & Answers determine MTO's domain & scope

Question	Answer
What is the domain the Ontology will cover?	The purpose of building this ontology is to cover the Malaysian Tourism as a Domain & we call it **MTO.**
What is MTO going to be used for?	MTO is built to be used as an infrastructure for Semantic Technology regardless of which application uses it with the intention of Semantic matching between tourism information and user profile via Natural Language Interface. With respect to the tourism domain, a core domain ontology needs to describe Tourism related information, which is typically provided by tourist services in order to satisfy the information needs of tourists visiting Malaysia.
What type of answers should MTO provide?	**MTO** needs to provide an understandable, conceptualized and linked vocabulary required by the Malaysian Tourism Domain. The main concepts to be defined are tourism related objects (i.e., Accommodation, food, Activity, Location, and Weather).
Who will use MTO?	The **MTO** is used here as a knowledge base for the SMTRS users (i.e. foreigner tourists to Malaysia). Also the use of MTO intended to engender knowledge reuse and semantic interoperability within specific tourism domain.

Step 2: Consider Reusing Existing Ontologies

Many ontologies (as shown in 2.3.5) that have been built in the field of tourism, aim to enhance the information extraction and retrieval, not only for human access but also for machine access. In the early stages of building the SMTRS prototype, we used the OnTour ontology (refer to 2.3.5). Figure 3.6 visualize the OnTour ontology.

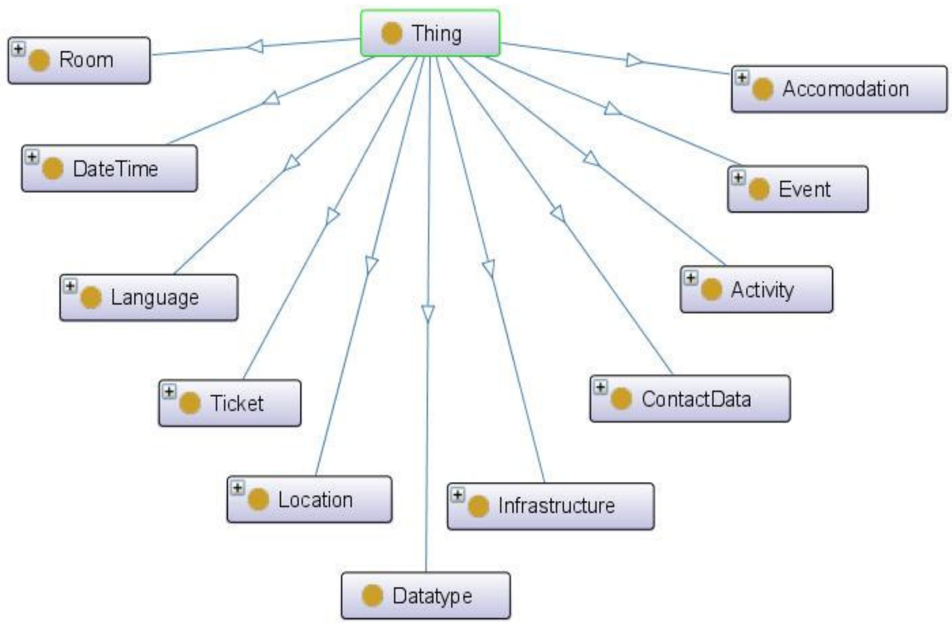


Figure 3.6 OnTour Ontology

The main focus is the description of accommodations, infrastructure, activity and concepts that describe leisure activities and geographical data. We choose OnTour ontology as it is built using the OWL-DL language. In order to use OnTour in SMTRS, we customized the OnTour ontology to fit the Malaysian tourism nature, such as (Hotels names, City names, activities available in Malaysia, food served in Malaysian restaurants, etc), the customization covers the instances without changing any of the concepts or the structure of OnTour.

In the development of SMTRS, we face many drawbacks in the OnTour ontology which needed changing the structure of the OnTour, such as defining many properties as Boolean values. For example the property hasPhone is defined as Boolean, defining properties as Boolean is not compatible in NLI. The development of the MTO is mostly concerned with answering tourists' query and semantic matching between tourism items and tourist preferences.

From our review (2.3.5) existing tourism ontologies neglect building the tourism ontology to match users' preferences such as categorizations for tourism activity, locations and distance between tourism items, which are fundamental to provide personalized information about tourism items. Another Tourism ontology worth mentioning here is Harmonise Ontology. Harmonise is formed by a limited number of the most representative concepts of the tourism industry, so allowing for information exchange between tourism actors. Moreover, Harmonise has a limitation with ontology reasoning.

To date, there is no published works on a Malaysian tourism ontology. There is a Malaysian tourism ontology which was developed as part of a final year project (Safiin, 2006) and placed on the web (URL). Also, there is no access for the public to the ontology and no published article describing the ontology and despite contacting the researcher, there has been no forthcoming response. Although the public is informed of the existing of this project through the web, no details have been provided.

In 2009 another Malaysia Tourism ontology (Zakaria, Hall, & Lewis, 2009) was published in the web with the goal of modelling images for Semantic description. It consists of two main roots which are Attraction and Event. Figure 3.7 visualise the ontology, and shows how Concepts were identified by the approach for *the St. Paul's Hill.* The images are linked to the identified information in the knowledge bases. All the description in this ontology are pertaining to specific images only. The number of images and the scope of the ontology was not provided.

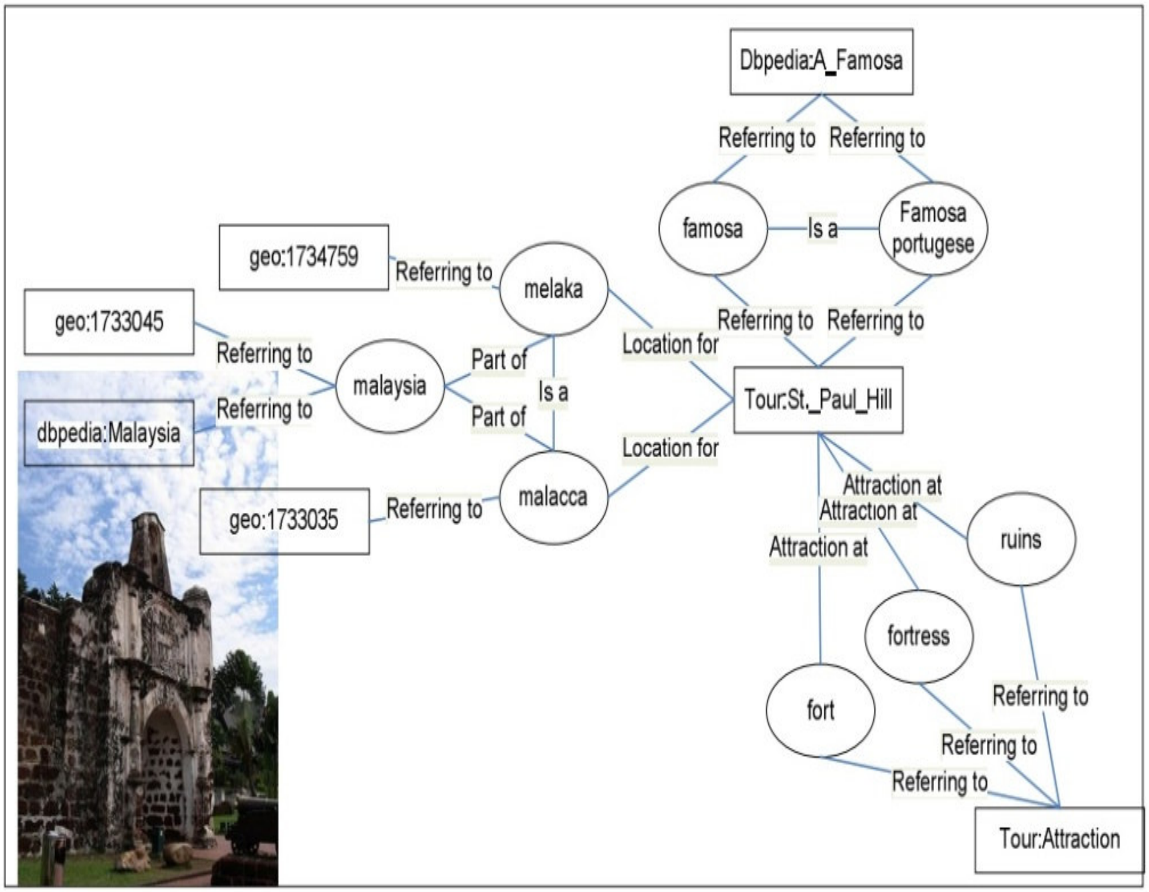


Figure 3.7 Malaysia Tourism ontology (Zakaria, Hall, & Lewis, 2009)

As a conclusion, there are two Tourism ontologies (Harmonise discussed in section 2.3.5 and OnTour discussed in this section) that can be considered for reuse. However, both ontologies can not fulfil our aim for the following:

1. OnTour: defining many properties as Boolean values.
2. Harmonise: the limitation of concepts and ontology reasoning.

Step 3: Enumerate Important Terms in the Ontology

The concepts of MTO was built based upon the Thesaurus on Tourism and Leisure Activities, published by the World Tourism Organization (WTO, 2001) as a specialized agency of the United Nations and an international standard for the sector of e-tourism. WTO serves as a global forum for tourism policy issues and practical source of tourism know-how. The Thesaurus is a standard terminology built to ensure the consistency of the tourism resources (Domingue, et al., 2011). Figure 3.8 shows a sample from Thesaurus on Tourism and Leisure Activities defines two concepts used in Tourism domain (Theme Events and Theme Park)

THEME EVENTS
MANIFESTATION THEMATIQUE / ACONTECIMIENTO TEMATICO
10.05
SN: Coordinated activities which together follow a story line or a popular, historic, cultural or fantasy experience for the enjoyment of the participants
BT: TOURISM EVENTS

THEME PARKS
PARC A THEME / PARQUE TEMATICO
05.05.06
SN: A park where the entertainment offered is focussed on one or more themes, eg: Disneyland
BT: LEISURE PARKS

Figure 3.8 Sample from Thesaurus on Tourism and Leisure Activities

According to the literature review (refer to 2.1.3), tourists search for information in four main categories (Accommodation, Transportation, Activity and Food). The work done in this thesis focused on three of these main categories (Accommodation, Activity and Food). The Transportation category was not considered as it may be misleading because of the broad range of terms that sometimes led to too detailed concepts, which had to be taken out in the later stage of the development, (i.e. many concepts involved in the transportation domain and relations between these concepts, e.g., all kinds of transporting modes in water, land and air, correlated administrative area, landmarks, meteorological information knowledge and so on). The inclusion of this will expand the ontology to an unmanageable size.

However, some of the concepts such as the top level concepts (*Accommodation*, *Activity* and *Location*) are universal concepts of tourism and can be borrowed from other ontologies such as Harmonise and OnTour as shown in Figure 3.9. The fields that the ontology cover identified by these concepts, but it is not necessary to borrow the sub classes of these concepts.

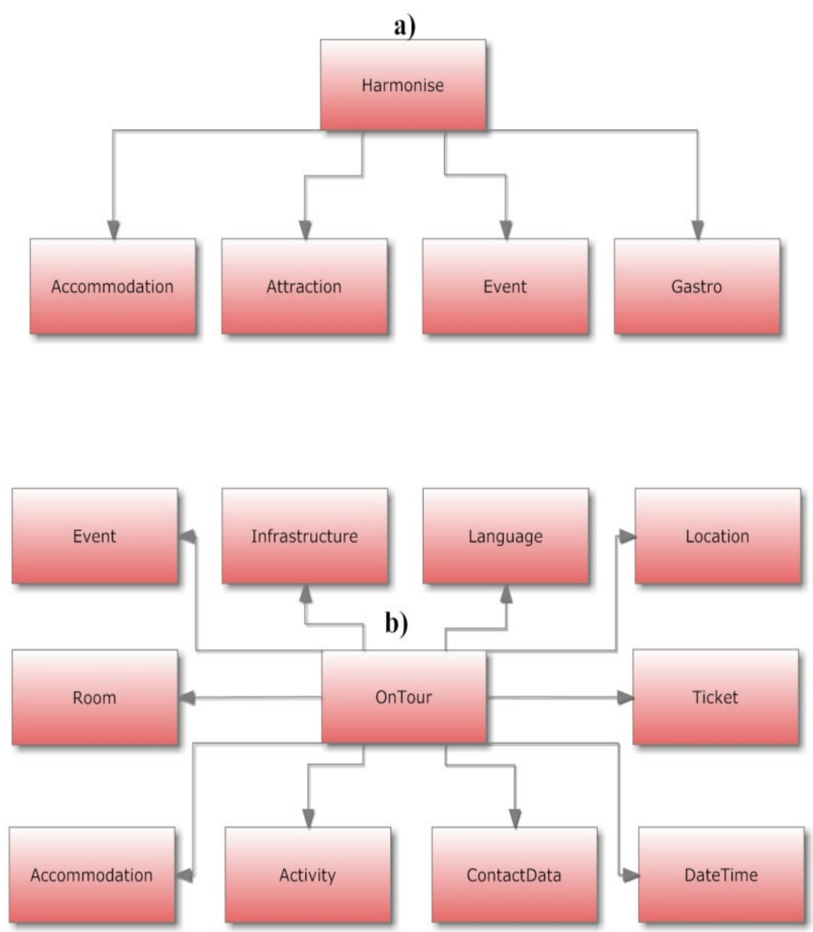


Figure 3.9 top level concepts for Harmonise and OnTour ontologies

The structure of the ontology defines the sub classes according to the appropriateness of the systems' functionality, for example:

- The concept Accommodation can be divided into two sub classes (such as Hotels and Serviced Apartments) as shown in Figure 3.10. Each instance in Malaysia (such as JW-Marriott and Times-Square) and are classified under a certain sub class based on its type.

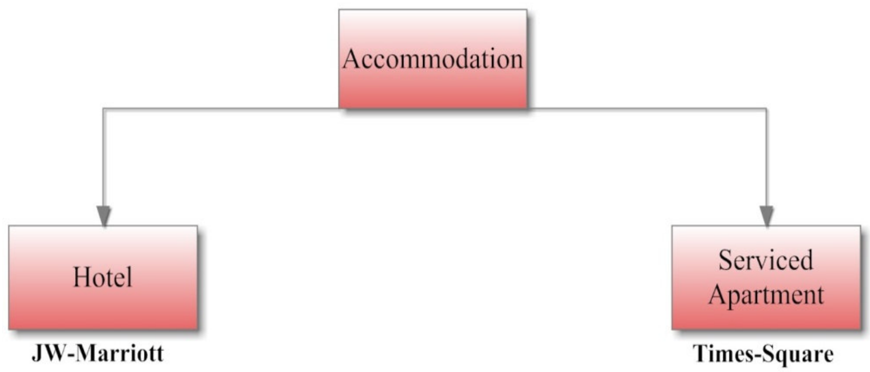


Figure 3.10 Dividing the concept Accommodation into Two sub classes

- The concept Accommodation conclude all the instances without dividing the Accommodation into sub classes as shown in Figure 3.11, and link each instance by a relation (is-a) to other instances (such as Hotel or Services Apartment)

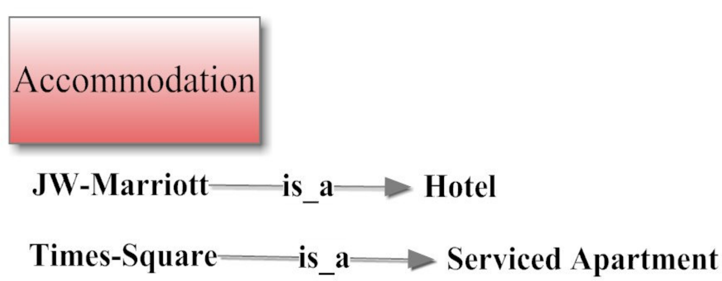


Figure 3.11 The concept Accommodation conclude instances with different types

We added the *Dining* category to the top level of the MTO in order to include the restaurants that tourists might query about. Also in order to support climatic information (i.e. expected conditions), we added the *Weather* category as some activities for example cannot be undertaken if it is raining, and the tourist may want to know this in advance. As a conclusion, MTO consists of five main categories as shown in Figure 3.12.

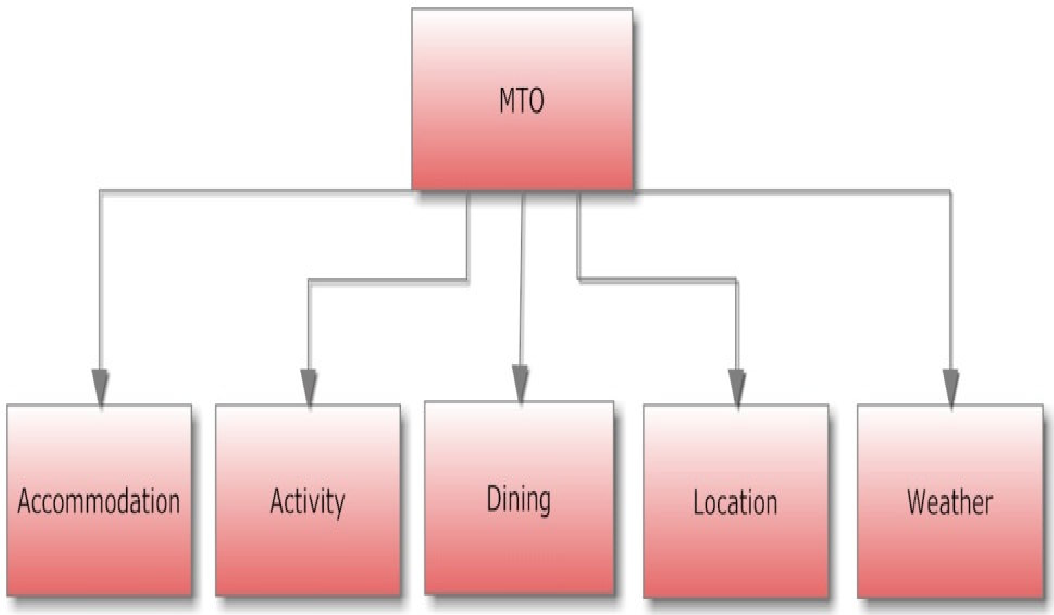


Figure 3.12 The five main categories of MTO

Step 4: Define the Classes and the Class Hierarchy

There are several possible approaches in developing a class hierarchy (Uschold and Gruninger 1996): We selected the top-down development process, starting with the definition of the most general concepts in the domain and subsequent specialization of the concepts. In order to generate the skeleton of the MTO, we selected the top-down approach so the hierarchy starts with the definition of the most general concepts in the domain followed by the subsequent specialization of the concepts as described briefly below:-

1) Categorisation of the main concepts according to the general classification as shown in Table 3.3:-

Table 3.3 Definition and general classification of MTO

General Classification	Definition
Accommodation	Terms include all facilities like hotels, guest houses and apartments
Activity	Terms include events and attractions that tourist can do e.g. adventure, sightseeing and shopping
Dining	Terms include food at several kinds of restaurants (such as local cuisines and familiar chain restaurants) and type of food available (such as Halal and Non-Halal)
Location	Terms include different ways for describing the geographical location of an entity
Weather	Terms include describing the weather at a particular place (such as humidity, temperature and rain)

2) Identify the sub and sub-sub concepts of the high level concepts (the general classifications) as shown in Table 3.4:-

Table 3.4 Identifying the sub and sub-sub concepts of MTO terms

Classes	Sub-Classes	Sub-sub class
Activity	Adventure	SafariPark
		ThemePark
		WaterPark
	Beach	
	Cultural	AsianArt Galary
		IslamicArt Galary
		NationalArt Galary
	Medical	
	Shopping	Indoor
		Outdoor
Accommodation	Apartments	
	Hotel	
Location	Kuala Lumpur	
	Malacca	
	Penang	
Dining	Halal	ArabRestaurant
		ChineseRestaurant
		IndianRestaurant
		JapaneseRestaurant
		MalayRestaurant
		WesternRestaurant
		VegetarianRestaurant
	Nonhalal	JapaneseRestaurant
		ChineseRestaurant
		IndianRestaurant
Weather		

Moreover, we have used Protégé (Noy, et al., 2000) which is one of the most widely used ontology development editor that defines ontology concepts (classes), properties, taxonomies, various restrictions and class instances. It also supports several ontology representation languages, including OWL-DL. A view of Protégé illustrating the top level concepts in MTO is shown in Figure 3.13.

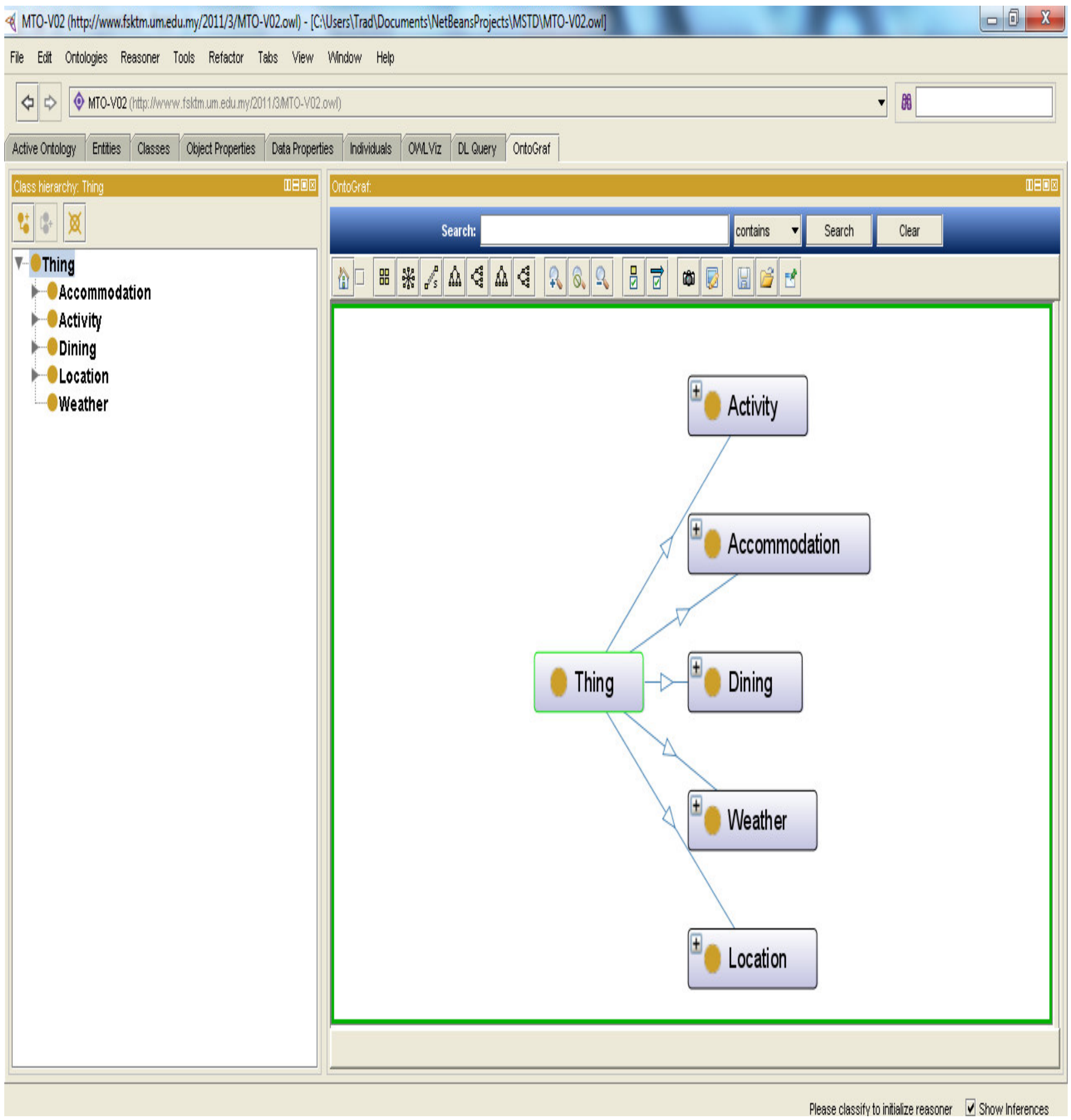


Figure 3.13 The Five main concept in MTO

Step 5: Define the Object Properties of Classes

While classes are universal across ontologies, sub classes and sub sub classes are particular to Malaysia tourism as provided by MTO. In this step, after identifying the classes we have to define the Object properties as classes alone will not provide enough information without a proper description of the concepts internal structure. Object Properties link individuals from the domain to individuals from the range. For example "*hasHotel*" is an object property that links the domain "*Location*" with the range "*Accommodation*" as shown in Figure 3.14. Therefore, the "*hasHotel*" object property links the "*BukitBintang*" an instance of the domain class "Location" with the "*JWMarriott*" an instance of the range class "*Accommodation*".

Figure 3.14 Object Property for hasHotel

We illustrate the MTO Object Properties with their domains and ranges in Table 3.5.

Table 3.5 MTO Object Properties with their domain and range

Object Property	Domain	Range
hasHotel	Location	Accommodation
hasLocation	Accommodation Activity Dining	Location
hasDining	Location	Dining
hasWeather	Location	Weather
hasActivity	Location	Activity

Step 6: Define the Data Properties of Classes

Data type properties represent relationships between an individual to an XML Schema Data type value or an RDF literal. In MTO, we identified several data properties to answer the tourists' questions in details for example (names, street address, date, time, and phone number).

Figure 3.15 shows that the individual "*JWMarriott*" has phone number "0123456789".

Figure 3.15 Data Property for hasPhone

In Table 3.6 we show some of the data type properties. MTO covers the general data properties of "*Accommodation, Activity, and Dining*" classes. The domain field in the second column shows the respective class of the data property, while the range in the third column shows the types of this data.

Table 3.6 MTO Data Properties with their domain and range

Data Property	Domain	Range
hasAddress	Accommodation	string
hasCloseTime	Activity Dining	dateTime
hasEmail	Accommodation Activity	string
hasOpenTime	Activity	dateTime

hasPhone	Dining	integer
hasPlace	Activity Dining	string
hasRate	Accommodation	integer
hasSeason	Activity	string
hasWebsite	Accommodation Activity	string
isCloseTo	Accommodation Activity Dining	string
hasFitnessCenter	Accommodation	boolean
hasKidsActivities	Accommodation	boolean
hasFreeWiFi	Accommodation	boolean

Step 7: Create Instances

The last step of developing the MTO is creating individual instances of the classes. For the purpose of testing our prototype, we randomly selected instances from the Malaysian tourism website (http://www.tourism.gov.my). These instances such as accommodations, accommodation facilities, Activities and Restaurants are inserted into the MTO. Additionally, Objects and Data properties for each instance are inserted into the MTO. Examples of these properties are Phone Number, Address, hotel names and any other information required. The instances with their classes are listed in appendix **B**.

3.4 System Process

3.4.1 NLP Component

Based on our literature review and to the best of our knowledge at the time of writing this lines, SMTRS is the first approach that integrates the Natural Language Interface with Content-based filtering in a Semantic environment. Before presenting SMTRS method to answer queries, we clarify existing NLI methods of answering queries. The query terms are tagged with class information, i.e. the relevant concepts of the domain (e.g. "*address*" as a type of hotel-address and "*JW-Marriot*" as hotel-name).

Original Query: What is the address of JW-Marriot?

The system algorithm analysis the original query and form the following tagged query.

Tagged Query: What is the address (hotel-address) of JW-Marriot (hotel-name)?

Depending on the Tagged terms found in the query, the system generate the SQL statement and the parameters are substituted with the appropriate values. A query about the address of JW-Marriot hotel produces the following statement.

SELECT hotels."hotel-address" FROM hotels

WHERE hotels.name = hotels."@hotel-name".

In SMTRS, the task of the NLP component is to recognize the semantic relations in the Natural Language query, and identify a query triple, considered as the simplest form of sentence structure (Subject, Predicate, and Object). Then, SMTRS measures the string similarity (refer to 2.4.2) between the query triple and the ontology triples. The result from this process is the best match for the answers. The following example describes a resource with statements.

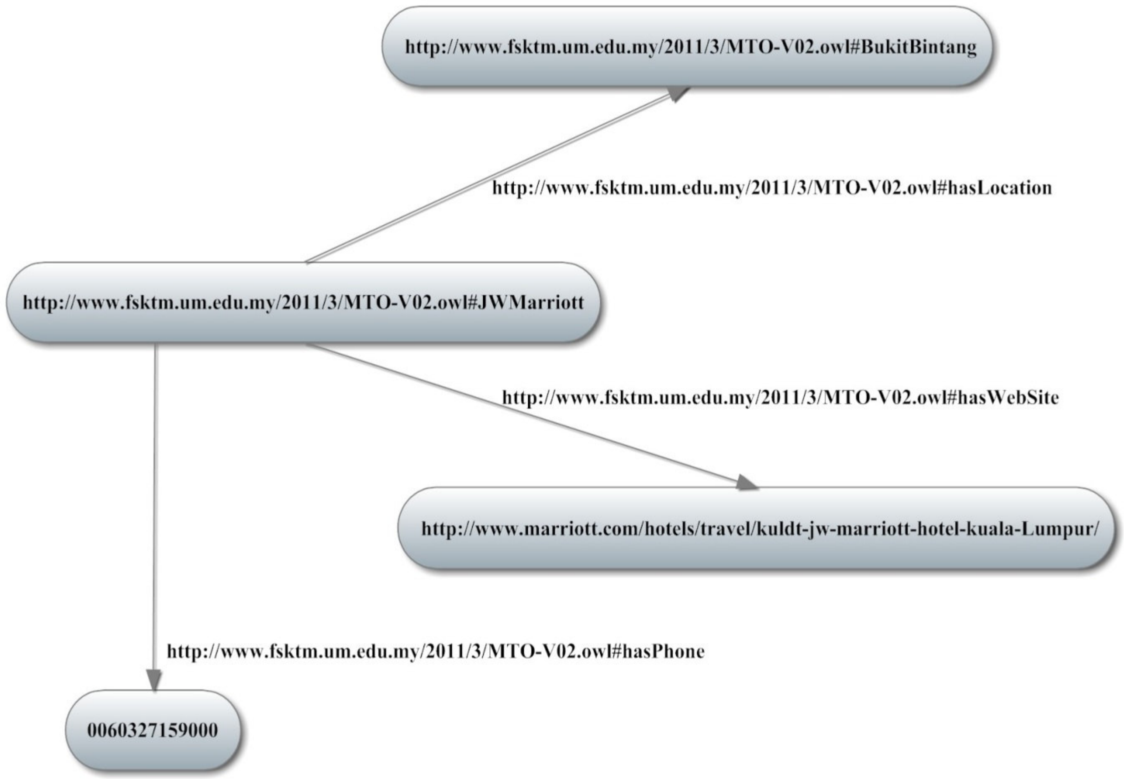


Figure 3.16 Example describe a resource in triples

Figure 3.16 visualized the description of the item “JWMarriott” hotel in triples; "there is a Hotel identified by (http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#JWMarriott) whose website address (http://www.marriott.com/hotels/travel/kuldt-jw-marriott-hotel-kuala-Lumpur/), whose location (http://www.marriott.com/hotels/travel/kuldt-jw-marriott-hotel-kuala-Lumpur/), and whose phone number is “0060327159000".

Furthermore, we generate the triples by using the General Architecture for Text Engineering (GATE) as discussed in (Cunningham, Maynard, Bontcheva, & Tablan, 2002). And to overcome the limitation of the GATE, for instance, if GATE is not able to capture the potential relation or terms to form the triple we built additional algorithm and we call it Relation Search Algorithm (RSA). Our solution utilizes the idea behind the structure of the Semantic Technology, since the natural language query needs to be converted into semantic triple (subject, predicate, object).

As shown in Figure 3.17. Semantic Technology is built based on the relations between the terms, where each term represents a concept and semantic relations between the terms capture their meaning. So there is a way to start processing the tourist question by finding those semantic relations.

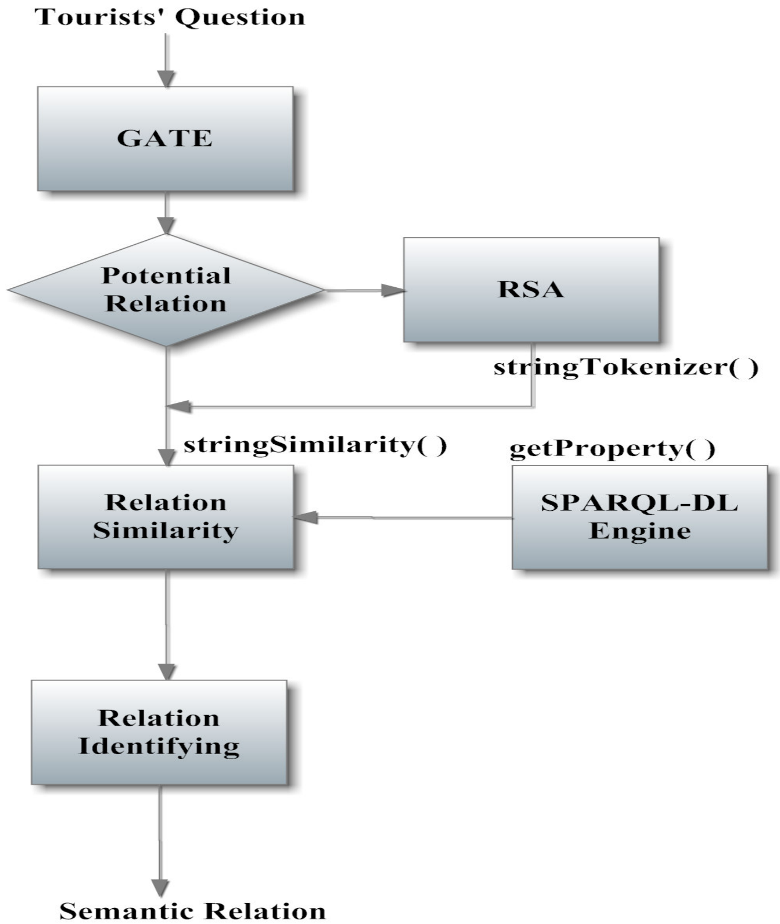


Figure 3.17 NLP Activity Diagram

For elaboration, we used GATE (refer to 2.4.2) as it was implemented by PowerAqua (López, Motta, & Uren, 2006). So far, GATE is able to handle many categories of queries based on pure linguistic criteria. Terms and relations extracted by GATE is completely domain independent i.e. the NL processing has nothing to do with the MTO. The process is entirely based on analyzing tourists' questions and representing these in triple forms.

In the following steps we provide an example to describe how a tourist's question is analyzed in SMTRS:

1. The tourist's question is "*What are the hotels in Bukit Bintang Area?*" (in case of compound words such as *Bukit Bintang* even if the user used space between the words, SMTRS will resolve it by deleting the space before proceeding)
2. GATE includes an information extraction system called ANNIE (A Nearly-New Information Extraction System). ANNIE produces syntactic annotations related to the user query. These annotations passed to JAPE grammars to identify terms, relations, and question indicators (who, what, etc.). GATE captures the potential relation or terms and give the result in this form:

   The first term is *What*

   The second term is *BukitBintang Area*

   The relation is *Hotels*

   From the previous example we can see that, GATE successfully identify the triple. Nevertheless, GATE is a third party component; hence we do not control the output. In certain situations GATE fails to retrieve the required relation, as some of the annotations created by GATE for nouns or verbs are not correct, usually if the relations in the question are in noun form, GATE will fail in giving the required relation. For example, if the tourist's question is "*Where to stay in Bukit Bintang?*"

   The first term is *Hotel*

   The second term is *Bukit Bintang*

   The relation is ***Not Found***

Obviously from the answers provided GATE fails to retrieve the relation between the first term and the second term for this tourist's question. Gate did not find a relation, where it should be (hasHotel). Although, GATE performs well in general questions but this is not always the case. For some questions analyzed by GATE we notice Linguistic failures. These Linguistic failures also accounted by Lopez, et al.(2007). There is no facility to trace and identify the failure through the GATE API. Therefore, in order to overcome this failure and taking advantage of GATE, we create the RSA which is discussed as follows. SMTRS uses RSA only if GATE could not successfully capture the potential relation or terms.

3. Relation Search Algorithm (RSA) was created as a domain dependent algorithm. It uses the MTO to identify the potential relation. The MTO helps us to reformulate and understand the query in terms of concepts, instances, values and relations between them. The process tokenizes the tourist' question and search each word in the MTO relations (either object or data property). The algorithm is as follows:

```
Algorithm RelationSearchAlgorithm()

    Input: String userQuestion
    Output: String potentialRelation

uQ← userQuestion

MTO_terms = MTO.getProperty()// Return all properties from
MTO

while uQ hasTokens do

    Token ← uQ.nextToken()

    while MTO_terms hasNext do
```

```
   Term ← MTO_terms.next ()

   if isSimilar(Term, Token) then // Compare term with Token

   potentialRelation ← Token
Return potentialRelation
else Return not found
```

An example of the RSA algorithm is described in the following steps:

a. Split the tourists' question "*Where to stay in Bukit Bintang?*" into tokens.

   Where / to/ stay / in/ BukitBintang

b. Identify each token with serial number for future identification as shown in below:-

Where	:	*01*
to	:	*02*
stay	:	*03*
In	:	*04*
BukitBintang	:	*05*

4. The next step, all relations are retrieved from the MTO ontology by using the SPARQL-DL Query to return all the properties in the MTO in an ontological property list. Then the set of tokens is compared separately to the ontological property list. SMTRS calculate the string similarity between each token and ontological property list based on Levenshtein distance.

   The Levenshtein distance between two strings is the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution of a single character. Scores vary in range from 0 to 1 were 0 is distantly similar and 1 is closely similar.

In order to correct spelling mistakes committed by users the string similarity is tuned to 90% match between the token and the ontological property. This process is described in the following steps:

a. Retrieve all relations from the MTO by executing the following SPARQL-DL query

   SPARQL-DL query.

```
public QueryResult getProperty(){

      return result = processQuery("SELECT * WHERE
    (Property(?x")));
       }
```

Where the variable (*x*) is used to retrieve all relations

b. Calculate the string similarity between each token and ontology relations based on Levenshtein distance (refer to 2.4.2).

c. Identify the **Semantic Relation (SR)** from the token list if the string similarity result is higher than 90%.

In the assumed tourists' question the extracted semantic relation was "**hasHotel**" as shown in the Figure 3.18.

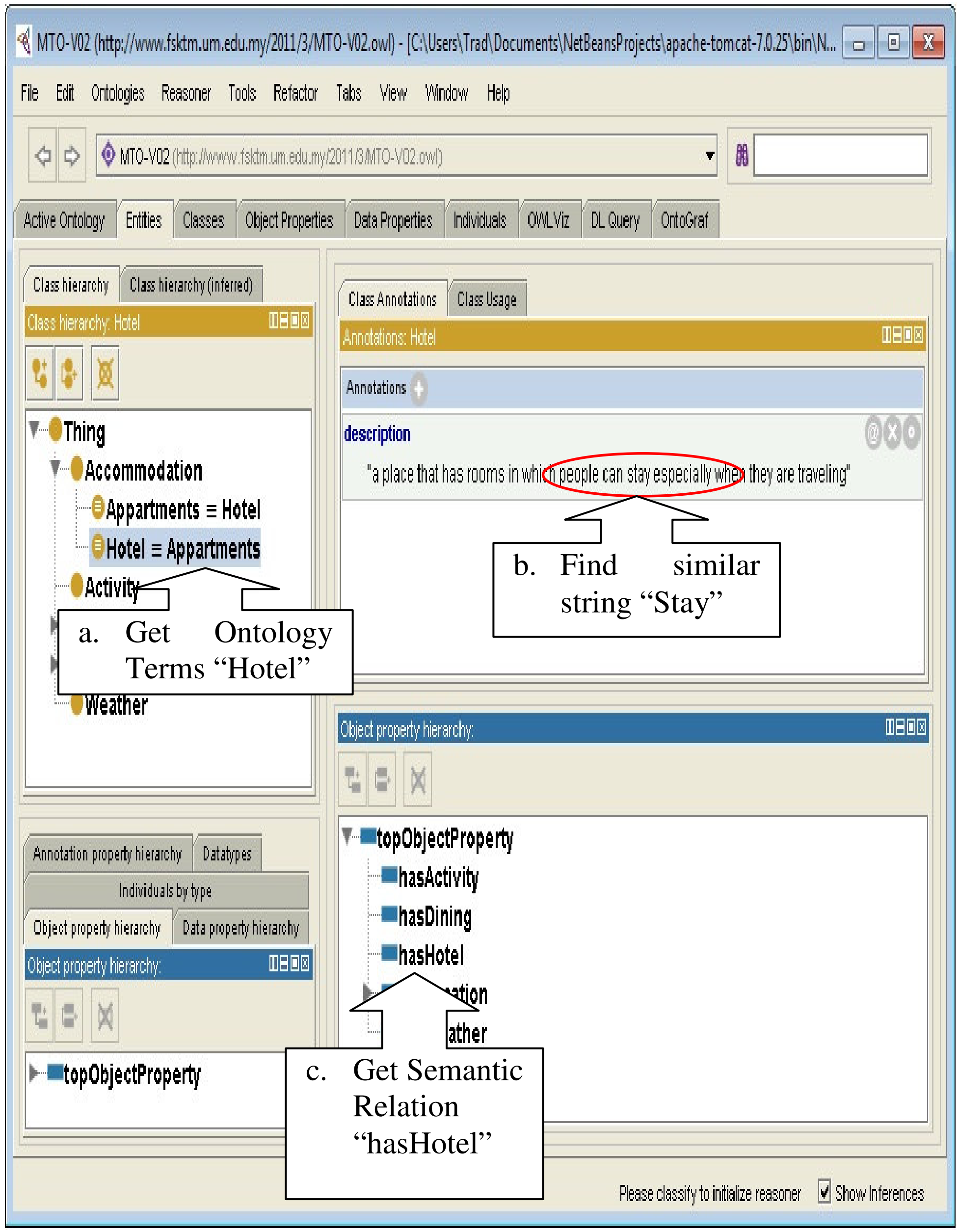


Figure 3.18 Extracted Semantic Relation from MTO

This SR "hasHotel" will be used in the following steps to show how we retrieve the semantic answer for the tourists' question.

3.4.2 Ontological Entity Mapping (OEM)

As described in the previous section, SMTRS use two methods in order to extract the relation from the tourist' question: First, by using Gate which extract the relation based on linguistic criteria. Second, by using RSA which extract the relation based on the MTO. The task of this component is to identify the entities (terms) in the ontology to the possible semantic relation found in the tourists' question and extracted by the NLP component. Figure 3.19 describe the steps.

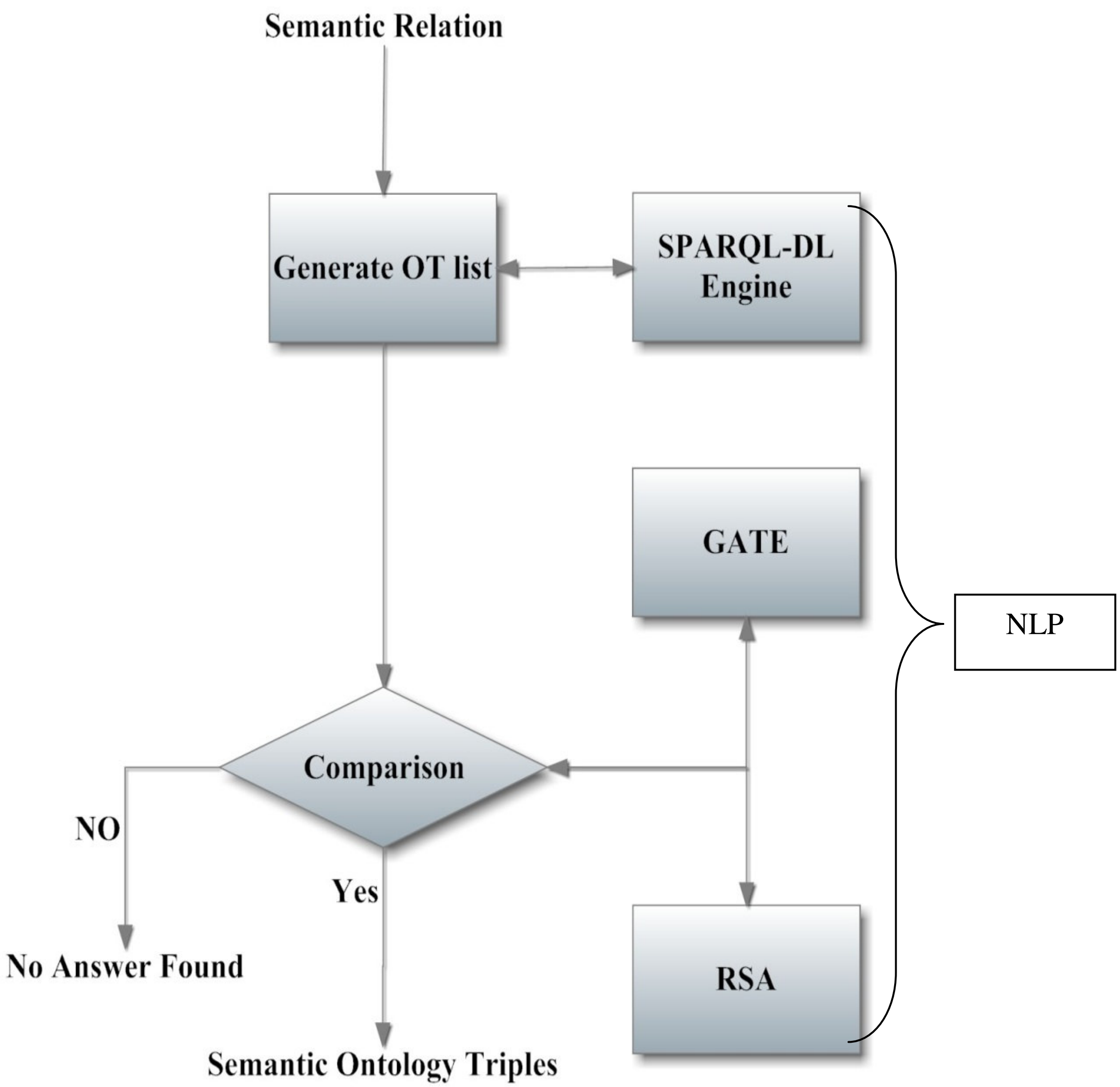


Figure 3.19 OEM Activity Diagram

As shown in figure 3.19, the Semantic Relation will be the main input for the OEM to start the process. There are two main procedures in the OEM as described in the next page:

I. **Generate Ontology Triple List:**

In order to retrieve the ontology triple list for the Semantic Relation which is "*hasHotel*" in our previous example, SPARQL-DL engine was used. SPARQL-DL aims to produce a query language that have the powerful and clear semantic to deal with OWL-DL (refer to 2.3.4). In this step our semantic concern by using the SPARQL-DL is to retrieve the ontology triples from MTO according to the given SR. For instance, we show a complete SPARQL-DL query, in order to get all the ontology triples:

```
PREFIX MTO: <http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#>
        SELECT ?subject  ?object
          WHERE
          {
          PropertyValue(?subject  , MTO: SemanticRelation, ?object)
          }
```

In Table 3.7 we illustrate the list of sample of the related classes and instances for the Semantic Relation (hotel) that had been retrieved from the SPARQL-DL Engine. In the 1st column of the mentioned table a sample of the classes which is semantically related to "hasHotel" has been retrieved. In the 2nd column we show the Semantic Relation that was used to retrieve the data. Finally in the 3rd column some sample of the instances, which are semantically related to the SR and connected to their respective classes, was retrieved.

Table 3.7 List of sample of the classes and instances for the SR

Classes	Semantic Relation	Instances
BukitBintang	hasHotel	TimeSquare
BukitBintang	hasHotel	JWMarriott
KLCC	hasHotel	Prince
SriHartamas	hasHotel	GrandMillennium
KLCC	hasHotel	Shangri-la
KLCC	hasHotel	Traders

Each row in the table represents an ontology triple in the MTO. The remaining step is to filter these triples to match the tourists' question terms. Hence, the next sub-heading shows the comparison process.

II. Comparison:

We illustrate in Table 3.8 the comparison between the terms found in the NLP component 3.3.1 and the Ontology Triples list in the previous subheading. As shown in the 1st and 2nd columns we used the extracted terms from the two methods GATE and RSA to compare it with the Classes and Instances of the ontology triples retrieved from the Semantic Relation "hasHotel". In the comparison process we used the Levenshtein distance between terms and the Ontology Triples to find the minimum number of operations needed to transform one into the other. The range of score is from 0 to 1 were 0 is distantly similar and 1 is closely similar.

Table 3.8 Comparison between terms from tourists' question and OT

Tourists' Questions Terms Extracted Method	Terms	Classes/ Individuals	Classes/ Individuals	Range of Similarity
GATE	*What*	Null	Null	0
	BukitBintang Area	BukitBintang	TimeSquare	1.0
		BukitBintang	JWMarriott	1.0
RSA	*What*	Null	Null	
	Are	Null	Null	
	The	Null	Null	
	Hotels	Null	Null	
	In	Null	Null	
	BukitBintang	BukitBintang	TimeSquare	1.0
		BukitBintang	JWMarriott	1.0
	Area			

III. Semantic Ontology Triples:

The last step in this process we generate the semantic ontology triple (SOT) list as shown in Table 3.9.

Table 3.9 Shows Semantic Ontology Tripels

Classes	Semantic Relation	Instances
BukitBintang	hasHotel	TimeSquare
BukitBintang	hasHotel	JWMarriott

3.4.3 Content-Based Filtering

Content-based filtering is a development of information filtering based on the classification of the user profile. It involves recommending unseen items that are similar to the preferences of the user. The similarity of items is calculated based on the features associated with the compared items (Ricci, et al., 2011). SMTRS is not only answering the tourists' question semantically but it enriches the semantic answers with recommended items that matched the tourists' preferences. For instance, if a user has preferred accommodations with business facilities while creating his/her profile, then the system will give a higher rate priority to the retrieved items in the same category and recommend these items. For example, the tourist's question is "*What are the hotels in Bukit Bintang Area?*" SMTRS provide the tourist with the hotels in Table 3.10 and tagged the JW-Marriot hotel as recommended based on the tourist preferences.

Table 3.10 SMTRS answer for the Hotels located in Bukit Bintang with the recommended one

Classes	Semantic Relation	Instances	Tagged
BukitBintang	hasHotel	JWMarriott	**Recommended for you**
BukitBintang	hasHotel	TimeSquare	

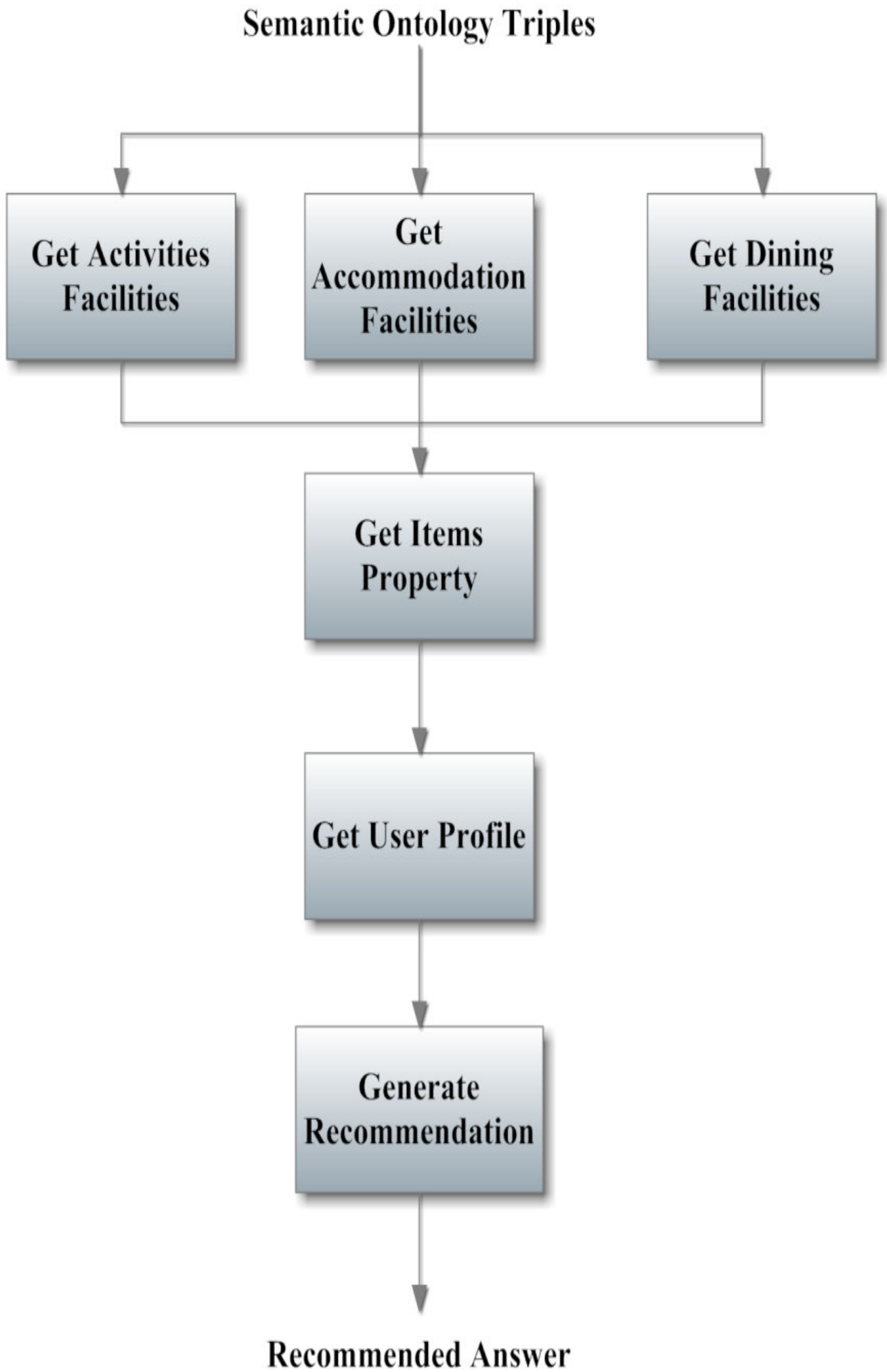


Figure 3.20 Recommendation Activity Diagram

Figure 3.20 is a visualization of the recommendation activity diagram, the SMTRS recommendation process basically consists of the following steps:

I. Filter the Semantic Ontology Triples

We laminate three types of recommendation: accommodations, activities and dining. By using the previous question "*What are the hotels in BukitBintang Area?*" the SOT list shown in Table 3.9 represents the general answer for that question. Hence, SMTRS identifies the items and their amount. If the number of items is more than one the recommendation algorithm is invoked to find which of the items in the list matches the tourists profile to be recommended.

To generate a recommended list for the specific tourist, SMTRS will identify the type of the recommendation. As the items are of the type "*hotels*" then the recommendation will be for the "*Accommodation*" type. As shown in the SOT list there are more than one item, therefore SMTRS will return all the properties of these items as discussed in the following step.

II. Get properties for each instance (item) found in the SOT list.

Each item in the MTO has several properties, these properties range between the Object and Data property as discussed in 3.3.2 steps 5 and 6, and provide information about the item as shown in Table 3.11 for the specific item "*JWMarriott*". This information is required to be used in the matching process with the user preferences. Therefore in the next step we retrieve the user profile information.

Table 3.11 Item property details

Item	Relation	Details
JWMarriott	hasLocation	BukitBintang
JWMarriott	hasWebSite	http://www.marriott.com/hotels/travel/kuldt-jw-marriott-hotel-kuala-Lumpur
JWMarriott	hasMinimumRoomCharge	230
JWMarriott	hasPhone	0327159000
JWMarriott	isCloseTo	Pavilion-Mall
JWMarriott	isCloseTo	Midvally-Mega-Mall
JWMarriott	hasMaximumRoomCharge	900
JWMarriott	hasAddress	183 Jalan Bukit Bintang Kuala Lumpur, 55100 Malaysia
JWMarriott	hasFitnessCenter	true
JWMarriott	hasKidsActivities	true
TimeSquare	hasLocation	BukitBintang
TimeSquare	hasWebSite	http://www.timesuites.com/
TimeSquare	hasMinimumRoomCharge	250
TimeSquare	hasPhone	012-928 2880
TimeSquare	isCloseTo	Pavilion-Mall
TimeSquare	isCloseTo	Low-Yat-Mall
TimeSquare	hasMaximumRoomCharge	
TimeSquare	hasAddress	A22-16, Berjaya Times Square, No.1, Jalan Imbi, 55100 Kuala Lumpur
TimeSquare	hasFitnessCenter	true
TimeSquare	hasBusinessCentre	true
TimeSquare	hasFreeWiFi	true
TimeSquare	hasFreeParking	true

III. Get User profile preferences

In this step the recommendation component retrieve the user profile information. The information will return the preferences of the user based on the type of recommendation identified in the previous step (i.e. Accommodation). For illustration purposes, the preferences represent the users' wishes, interests and facilities of the Accommodation type to be. In Table 3.12 we show an example of the preferences that users can select in the user profile for the accommodation section.

Table 3.12 Accommodation User Preferences list

Type	Facilities	Items	User
Accommodation	Amenities	Business Centre	
		Bar And Lounge	
		Fitness Centre	✓
		Free Breakfast	
		Free WiFi	
		Free Parking	
		Kids Activities	✓
		Pets Allowed	
		Swimming Pool	
	Daily budget	Below RM 100	
		Between RM 100 and RM 149	
		Between RM 150 and RM 199	
		Between RM 200 and RM 299	✓
		Over RM 300	

IV. Match up the attributes of the user profile with the SOT items

In order to find matches between user and items, we compare the preferences attributes of the user profile with the properties related to the item in the SOT list. The matching process include direct string matching such as *"Fitness Centre"* from the user profile with the property *"hasFitnessCentre"*, and indirect matching such as linking the user preference budget *"Between RM 200 and RM 299"* with the property *"hasMinimumRoomCharge =230"*. Table 3.13 show examples of the matches found between a random user and the hotel property found in the SOT.

Table 3.13 Matching list of User preferences with SOT items

User preference	Matched Property	Items for the matched property
Fitness Centre	hasFitnessCentre	JWMarriott
	hasFitnessCentre	TimeSquare
Kids Activities	hasKidsActivities	JWMarriott
Between RM 200 and RM 299	hasMinimumRoomCharge =230	JWMarriott
	hasMinimumRoomCharge =250	TimeSquare

V. Generate Recommended Items

The result is a relevance judgment that represents the user's level of interest in that object. In this step we calculate the total matches. The highest total represents the item that matches the user's preferences as shown in Table 3.14.

Table 3.14 Total Matches of each Item

Item name	Property	Total Matches
JWMarriott	1. hasFitnessCentre 2. kidsActivities 3. hasMinimumRoomCharge =230	3
TimeSquare	1. hasFitnessCentre 2. hasMinimumRoomCharge =250	2

As shown in Table 3.15 the highest matched item will be tagged as "*Recommended for you*" and the final SOT list will be displayed to the user as the Recommended Answer.

Table 3.15 Recommended Answer

Classes	**Semantic Relation**	**Instances**	**Tagged**
BukitBintang	hasHotel	JWMarriott	*Recommended for you*
BukitBintang	hasHotel	TimeSquare	

VI. Generate Recommended Items of the same question for different user with different profile.

Using the previous question "*What are the hotels in BukitBintang Area?*". In this step the recommendation component retrieve the second user profile information. The information will return the preferences of the user based on the type of recommendation identified in the previous step (i.e. Accommodation). For illustration purposes, the preferences represent the users' wishes, interests and facilities of the Accommodation type to be. In Table 3.16 we show an example of the preferences that the second user selected in the user profile in the accommodation section.

Table 3.16 Amenities chosen by Tourist 2

Type	Facilities	Items	User 2
Accommodation	Amenities	Business Centre	✓
		Bar And Lounge	
		Fitness Centre	
		Free Breakfast	
		Free WiFi	✓
		Free Parking	✓
		Kids Activities	
		Pets Allowed	
		Swimming Pool	
	Daily budget	Below RM 100	
		Between RM 100 and RM 149	
		Between RM 150 and RM 199	
		Between RM 200 and RM 299	✓
		Over RM 300	

In order to find matches between the second user and items, we compare the preferences attributes of the user profile with the properties related to the item in the SOT list. Table 3.17 show the matches found between the second user and the hotel properties found in the SOT. Obviously, the matches list in Table 3.17 differs from matches list in Table 3.13.

Table 3.17 Matching list of User 2 preferences with SOT items

User preference	Matched Property	Items for the matched property
Business Centre	hasBusinessCentre	TimeSquare
Free WiFi	hasFreeWiFi	TimeSquare
Free Parking	hasFreeParking	TimeSquare
Between RM 200 and RM 299	hasMinimumRoomCharge =250	TimeSquare
	hasMinimumRoomCharge =230	JWMarriott

The highest total represents the item that matches the second user's preferences is shown in Table 3.18.

Table 3.18 Total Matches of each Item with User 2 preferances

Item name	Property	Total Matches
JWMarriott	1. hasMinimumRoomCharge =230	1
TimeSquare	1. hasBusinessCentre	4
	2. hasFreeWiFi	
	3. hasFreeParking	
	4. hasMinimumRoomCharge =250	

As shown in Table 3.19 the highest matched item will be tagged as "*Recommended for you*" and the final SOT list will be displayed to the user as the Recommended Answer.

Table 3.19 Recommended Answer for User 2

Classes	Semantic Relation	Instances	Tagged
BukitBintang	hasHotel	JWMarriott	
BukitBintang	hasHotel	TimeSquare	*Recommended for you*

As a conclusion, the same query "*What are the hotels in BukitBintang Area?*" asked by two users with different preferences will generate different recommendation by SMTRS, namely the first user was recommended JWMarriott while the second user two was recommended Time Square.

3.5 SMTRS Sequences of Operations

The typical sequences of operations within the SMTRS components are as follows:

Firstly, the interface takes the tourists' question in the form of a query (English language). The Natural Language Processing Component (NLPC) analyzes the query by using GATE or Relation Search Algorithm (RSA) and the output is a **Potential Relation (PR)**.

Secondly, we map the PR to the ontological properties retrieved from the Malaysian Tourism Ontology (MTO). In the process we use a SPARQL-DL query to retrieve the ontological property and form an ontological property list. The list is compared with the PR to find the similarity and if there is a 90% similarity, the PR will be considered semantically equivalent and is called as the **Semantic Relation (SR)**.

Thirdly after the SR is identified the **Ontological Entity Mapping (OEM)** process takes place. In this process the Query Generator in SPARQL-DL Engine component builds a SPARQL-DL query to return all Ontology Triples available based on the SR. From the found triples we form the **Ontology Triple (OT)**.

Fourthly, the Answer Retrieval in the SPARQL-DL Engine component compares the Ontology Triples with the tourists' question by calculating the string similarity based on Levenshtein distance. This results in the **Semantic Ontology Triple (SOT)**.

Finally, if there is more than one item in the SOT the Answer Recommendation component will use the user profile information to generate a recommended list. The list will contain the nearest matched items to the user's preference based on content base filtering and subsequently form the Recommended Answer.

4.0 Evaluation

This work accomplished answering a tourist question (using Natural Language Interface) by giving a semantic answer (i.e. different ways of wording similar questions result in the same answer as show in Tables 4.2 and 4.3 questions 7 and 9). The answer contains a content-based recommendation based on tourists' preferences (as an example if two users with different profiles ask the same question the system give different answers as described in 3.4.3,VI). For instance, if a user has preferred accommodations with business facilities while creating his/her profile, then the SMTRS will give a higher rate priority to the retrieved items in the same category and recommend these items (example provided in 3.4.3-V).

This chapter describes the design of the evaluations strategies of SMTRS and their results. The evaluation aims to establish the validity of the ontology, the retrieval performance and the usability.

4.1 Test Plan

In order to investigate the research goals, the SMTRS was evaluated in a series of equipment.

- **System-based retrieval performance evaluation:** Precision and recall is used to measure the information retrieval effectiveness.

- **Domain specific knowledge (Evaluation of MTO):** Pellet and Fact++ reasoners are used to evaluate the MTO inference, they enhance reasoning by extending the meaning behind relationships among concepts. Both reasoners specialize in the DL technology (refer to 2.3.3 and 2.3.4).
- **User-based usability evaluation:** The questionnaire used for evaluating SMTRS was driven from the System Usability Scale (SUS) (Brooke, 1996), with the addition of specific questions (Q11 to Q14 in 4.4) to measure the users' satisfactions of the recommendation feature.

4.2 System-based Retrieval Performance Evaluation

In general, the aim for information retrieval systems is to optimise both precision and recall (refer to 2.4.3), this optimisation ensure the quality of the retrieval performance. Furthermore our concern about precision and recall is to determine the relevance of the output compared to the MTO concepts and relations. We have evaluated our retrieval performance approach using sample questions obtained from "Answers Yahoo" (Answer Yahoo, 2012), by searching the following keywords. The first keywords (1 to 4) represent some of the main topics of the MTO classes (refer to 3.3.2 step 4.) The remaining keywords (5 to 6) represent the domain location.

1. "hotel"
2. "restaurant"
3. "activity"
4. "recommended"
5. "Malaysia"
6. "Kuala Lumpur"

From the sample questions obtained from Yahoo we reformulate some of these questions in order to give it a Malaysian perspective without changing the structure of the original questions. When choosing these questions, the main classes of the MTO (refer to 3.3.2) are covered. Table 4.1 shows the original questions extracted from (Answer Yahoo, 2012) and the reformulated questions using Malaysian cities and places.

Table 4.1 Original Q Vs Reformulated Q

	Original Question	Reformulated Question
1	What to do in Kuala Lumpur?	Using original question
2	What is the number of the Swiss Garden hotel in Kuala Lumpur?	What is the phone number of Swiss Garden hotel?
3	What to do and where to stay in London England?	Where to stay in Bukit Bintang?
4	What are the hotels near Sepang Circuit?	Using original question
5	What are some tourist attractions in Kuala Lampur?	Using original question
6	What is the location of Hotel Eldritch Shareon in Singapore?	What is the location of Times Square?
7	What is there for tourists to visit in London?	What is there for tourists to visit in Kuala Lumpur?
8	What are the activities in the country Swiss?	What is the activity in Penang?
9	Where can I get the best Western restaurant in Kuala Lumpur?	Using original question
10	What is the best restaurant (food) in Maryland?	What is the best restaurant (food) in Kuala Lumpur?
11	Where is a nice place to eat in Kuala Lumpur, Malaysia?	Using original question
12	What is the address of SEMBOL HOTEL in Istanbul?	What is the address of the Capitol hotel?
13	What are the names of all the hotels located on Carlton Street in Melbourne, Australia?	What are the hotels located in Bukit Bintang?
14	Where are the interesting places in Gold Coast Australia?	Where is the interesting place in Kuala Lumpur?
15	Which restaurant is the best in Epcot?	Which restaurant is the best in Kuala Lumpur?
16	What are some five star Hotels in Montego Bay Jamaica?	What are the five star hotels in Bukit Bintang?

From Table 4.1, obviously "WH" type questions are selected as these are the desired questions as revealed by the literature (refer to 2.1.3), but not complex answers to, why or how questions, that require explanation. Moreover, for the questions that have comparative keywords (nice, best, interesting, etc.) SMTRS eliminate them and give the recommendation based on the tourist preferences.

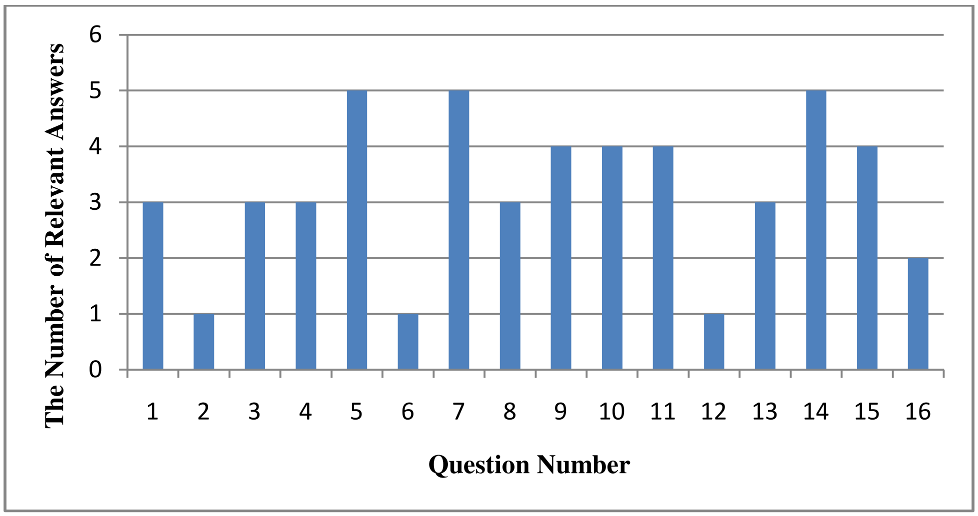


Figure 4.1 Histogram of number of relevant answers for each question in the sample set

In order to assess the relevant answers (i.e. correct rate), we compare the SMTRS retrieved answers with the manually generated SPARQL queries' retrieved answers. The metrics we used are recall and precision. For each domain, recall means the percentage of relevant answers that SMTRS produced an output; precision refers to the percentage of relevant answers that SMTRS produced an output in the total retrieved answers. SMTRS was able to find the relevant answers in relations to the main topics of the MTO classes as shown in Figure 4.1 after processing the sample questions. Obviously, the number of relevant answers varies among the questions since the number of items (i.e. Hotel, Restaurant and Activity) varies in our MTO.

Table 4.2 Precision and Recall for the sample Questions

	Question	Retrieved	Relevant	Relevant & Retrieved	Recall	Precision
1	What to do in Penang?	3	3	3	1	1
2	What is the phone number of JWMarriott hotel?	1	1	1	1	1
3	Where to stay in Bukit Bintang?	3	3	3	1	1
4	What are the hotels near KLCC?	3	3	3	1	1
5	What are tourist attractions in Kuala Lumpur?	5	5	5	1	1
6	What is the location of Times Square?	1	1	1	1	1
7	What is there for tourists to visit in Kuala Lumpur?	5	5	5	1	1
8	What is the activity in Penang?	3	3	3	1	1
9	Where can i get restaurant in Kuala Lumpur	4	4	4	1	1
10	What is the best restaurant (food) in Kuala Lumpur?	4	4	4	1	1
11	Where to eat in Kuala Lumpur?	4	4	4	1	1
12	What is the address of the capitol hotel?	1	1	1	1	1
13	What are the hotels located in Bukit Bintang?	3	3	3	1	1
14	Where is the interesting place in Kuala Lumpur?	5	5	5	1	1
15	Which restaurant is the best in Kuala Lumpur?	4	4	4	1	1
16	What are the five star hotels in Bukit Bintang?	3	2	2	1	0.6667
	Average				**1**	**0.9791**

The sample set of questions is shown in Table 4.2. From the table the average precision is 97.91% and the average recall is 100%. Obviously, questions 9 and 11 have two different structures with the same meaning and for the same set of answers as shown in Table 4.3. The precision and recall results show that the answers retrieved by SMTRS to those questions are similar, which show the semantic retrieving process of SMTRS.

Table 4.3 The answer set for questions 9 and 11

Concept	Relation	Concept
Songket	hasLocation	KualaLumpur
Nasi Lemak Tanglin	hasLocation	KualaLumpur
Makan Kitchen	hasLocation	KualaLumpur
Restoran Queen's	hasLocation	KualaLumpur

Moreover, as can be seen in question 16, SMTRS retrieved irrelevant answer as well as relevant answers. To illustrate, the question was "*What are the five star hotels in Bukit Bintang?*". Bear in mind that this question contains two facts, hotels located in Bukit Bintang; and hotels with five star rates. SMTRS returned "*Capitol*" as one of the retrieved answers. While, "*Capitol*" is located in Bukit Bintang but it is not a five-star hotel and we considered it as a wrong or irrelevant answers. SMTRS supports factual questions and factual here means that answer are ground facts as found in the MTO.

As a conclusion, the precision and recall results show that SMTRS is capable of retrieving recommended relevant answers in relation to the concepts available in MTO and user profile. Hence, this proves that SMTRS is competent to provide tourists, visiting Malaysia, with the relevant information to help them in planning their vacations efficiently. We come to the conclusion that SMTRS overcomes the information overload problem aforementioned in chapter **1.2.** comparing to other available systems as mention in Table 2.8.

4.3 Evaluation of MTO

MTO is an OWL-DL ontology built to represent the Malaysian tourism information. Pellet and Fact++ are the state of the art reasoners in ontology inference engines specializing in DL technology. Both reasoners were plugged-in with Protégé as shown in Figure 4.2. Both support the inference services during the ontology development stages. Some of the services these reasoners support are identifying inconsistency and classifying taxonomy in the ontologies.

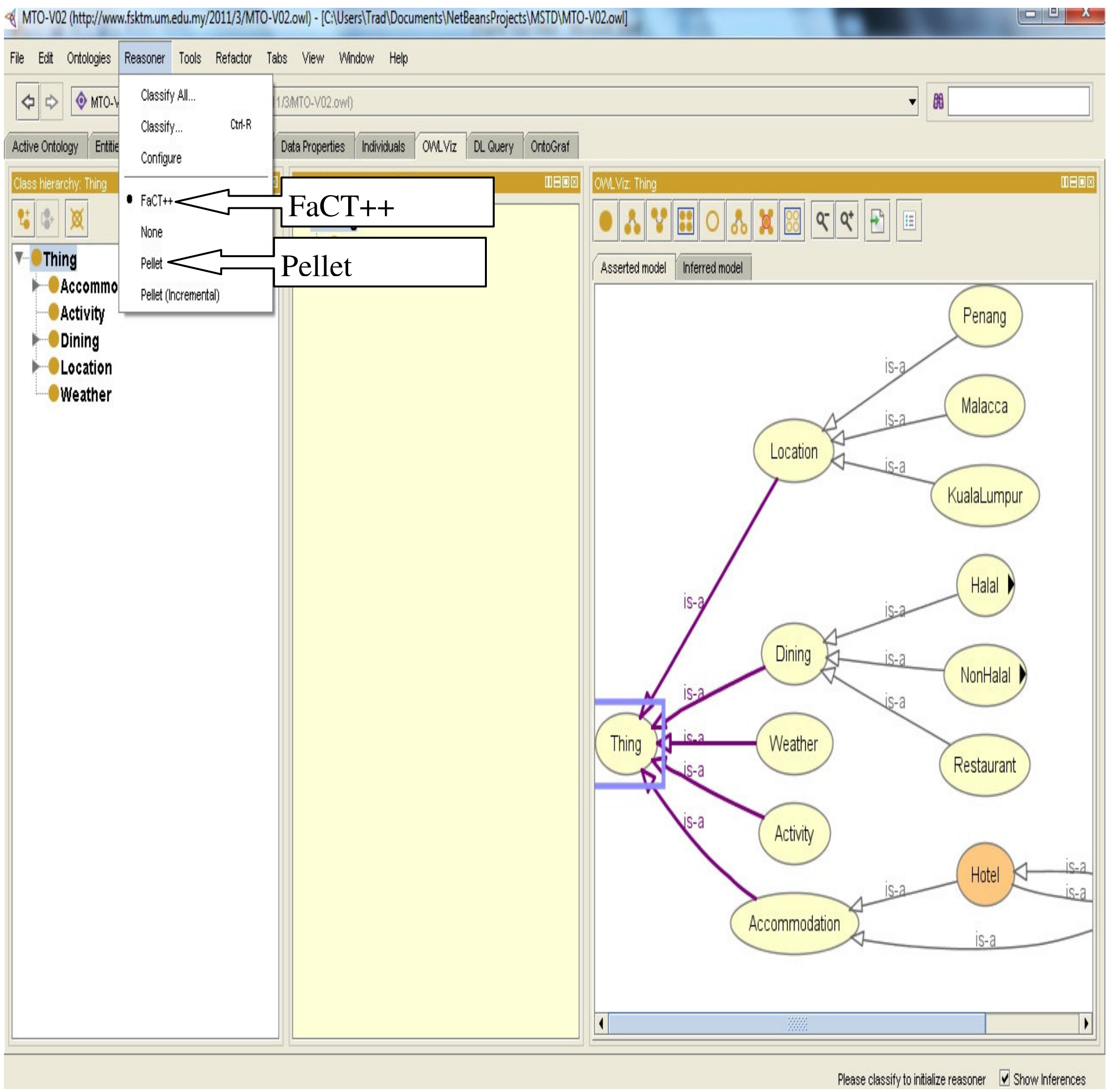


Figure 4.2 Reasoners Used to evaluate the MTO

The Reasoners performs the inferred class hierarchy automatically. This process is also known as classifying the ontology. To test for an unsatisfiable status, the member "Kuala Lumpur" of the concept "Location" was inserted as an equivalent concept to "Hotel". The concept "Hotel" is inserted as disjoint concept to "Location" and all members of both concepts cannot be inserted as equivalent concepts (as reasoning discussed in 2.3.4). Therefore, unsatisfiably classified classes generate an error message box which means error in the taxonomy and inconsistent reasoning as shown in Figure 4.3. This was because of the erroneous equivalence that was made between "Kuala Lumpur" and "Hotel", where it should be an equivalence of "Location" instead. Furthermore, "Hotel" is disjoint to "Location", although this is correct.

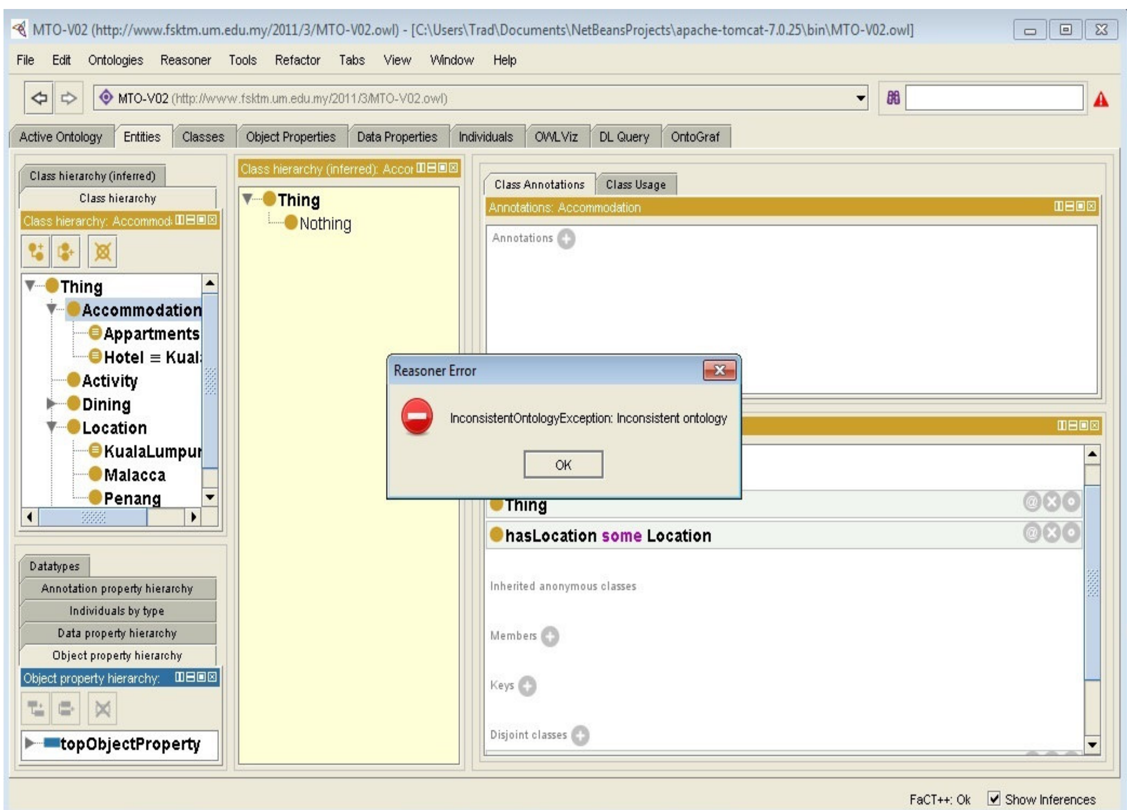


Figure 4.3 Inconsistent texonomy error message

The next subsection describes how each reasoner works:-

Pellet (Parsia & Sirin, 2005): We choose the pellet reasoner from the reasoner menu, which will automatically check the ontology classification. Protégé presents the inferred class hierarchy in a separate tab after the reasoner finishes classifying. Figure 4.4 presents the inferred hierarchy graph showing the consistent classes.

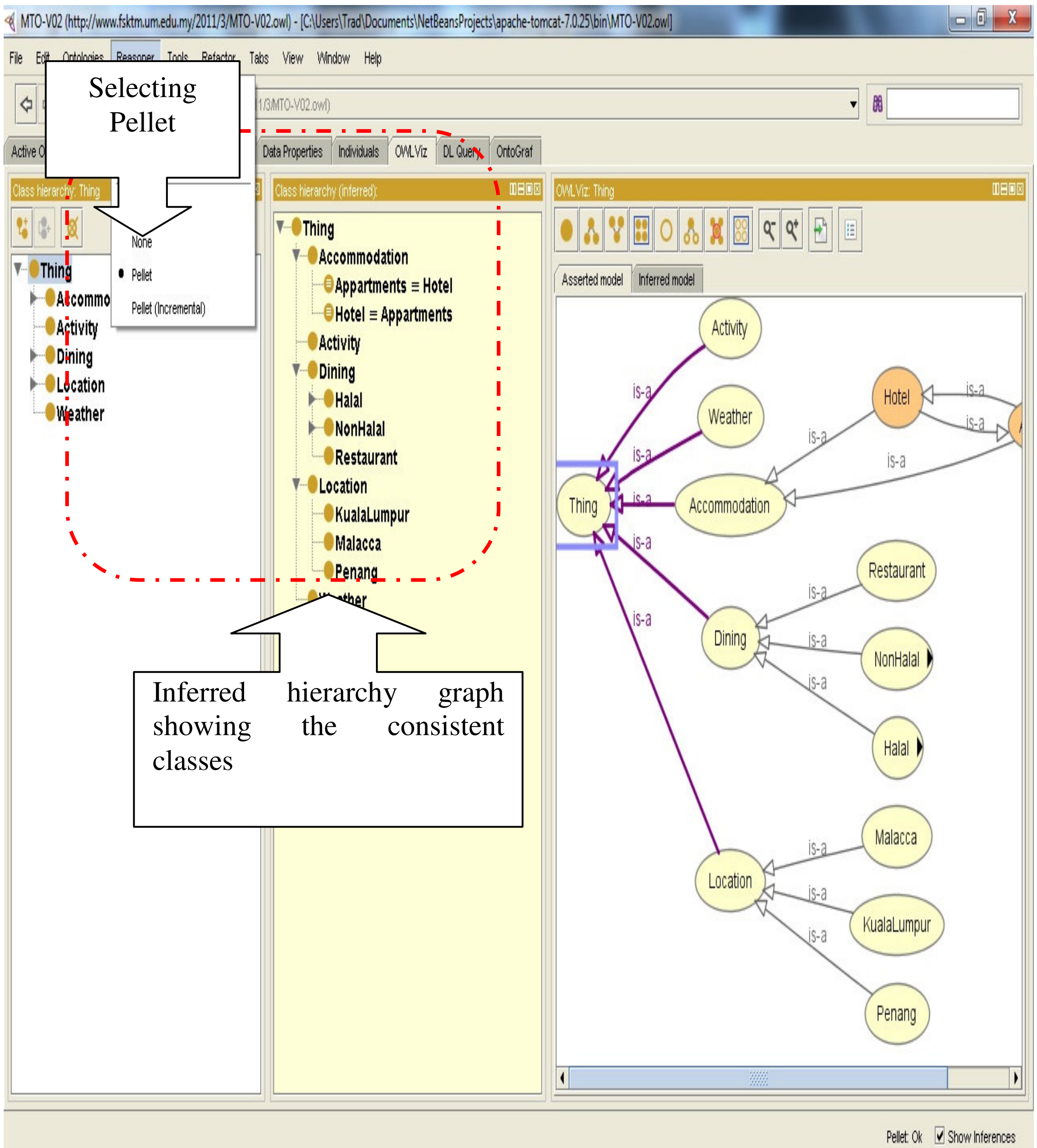


Figure 4.4 Pellet Reasoner

FaCT++(Tsarkov & Horrocks, 2006): We used FaCT++ to evaluate the MTO as it employs a wide range of performance enhancing optimization techniques. These are namely, the reasoner performs classification, i.e., computes and caches the subsumption partial ordering (taxonomy) of named concepts in the MTO. Several optimisations are applied here, mainly involving choosing the order in which concepts are processed so as to reduce the number of subsumption tests performed. The classifier uses a knowledge base satisfiability checker in order to decide subsumption problems for given pairs of concepts. This is the core component of the FaCT++ reasoner, and the most highly optimised one. Figure 4.5 shows the inferred hierarchy graph showing the consistent classes with the FaCT++ reasoner in protégé proving the validity of MTO.

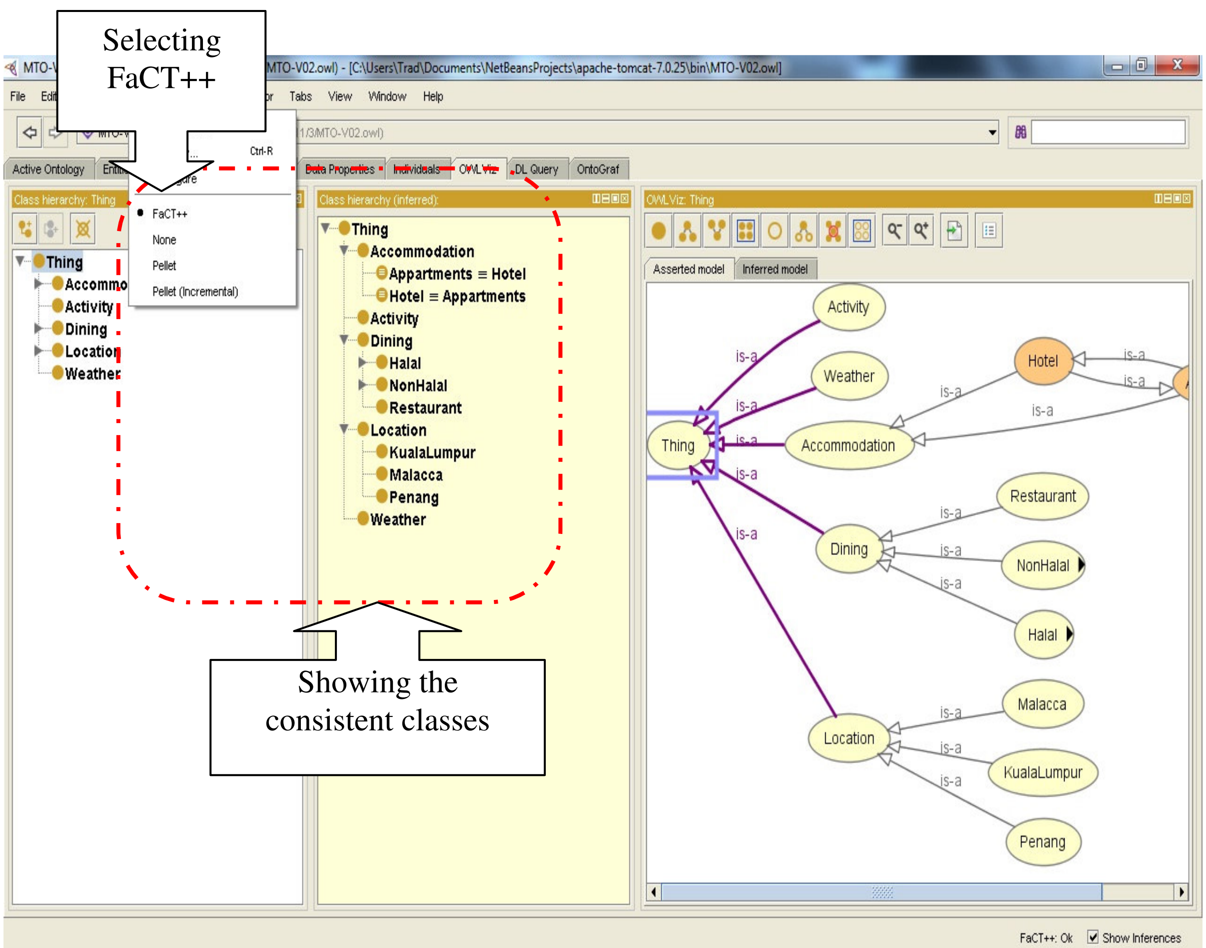


Figure 4.5 FaCT++ Reasoner

MTO is a fundamental component in SMTRS, which provides explicit tourism information in Malaysia domain. From the results shown in Figures 4.4 and 4.5 the concepts are consistent and the taxonomies are classified in MTO.

As a conclusion, the proposed tourism information representation relies on MTO. First, to solve the aforementioned problems of the tourism information representation, disseminated tourism information shall be standardized and integrated in a centralized ontology. Second, to understand this information, the metadata is efficiently used to explain tourism concepts such as Accommodations, Activities, Dining, Location or Weather. In addition to this explanation, an interrelation among each tourism instance would be described by MTO. With our proposed ontology, SMTRS generates a representation for the tourism knowledge as described in chapter 1.2.

4.4 User-based Usability Evaluation

SMTRS was built in order to: search information about tourism in Malaysia, by using tourists' query expressed in Natural Language and extracts this information with a recommended item using Content-based filtering. In order to know if the system is usable, we evaluated the system's usability by adapting the System Usability Scale (SUS) as discussed in (Brooke, 1996). SUS is one of the most popular questionnaires containing a standardized collection of questions. The Measurements of usability have several different aspects:

- Effectiveness (can users successfully achieve their objectives)

- Efficiency (how much effort and resource is expended in achieving those objectives)
- Satisfaction (was the experience satisfactory)

In general the aim of measuring the usability of SMTRS is to evaluate the core features (i.e. registration, login, retrieving answers, etc.). Moreover specific questions (Q11 to Q14) were added to measure the users' satisfactions of the recommendation feature. The recommended items are measured from the user's point of view. Hence, participants were requested to respond to the questionnaire to express their views on SMTRS usability.

- **Sample Selection:** A total of 30 participants performed the usability test. All of them were international visitors from various countries who were familiar with tourism sites in Malaysia. Therefore, they were able to give a good feedback regarding the system usability and the recommended items as they have been in Malaysia for quite some time and knew the places SMTRS recommended.

- **Experiment**: The methodology we used to perform the experiment with the participants and the prototype is as follows:

    - Users were taken through a 15 minutes tutorial session at the commencement of the experiment. Then each participant was asked to fill the registration page first, to create a personal profile (refer to Appendix E and Appendix F) and to log in via the log-in page as shown in Figure 4.6.

Figure 4.6 SMTRS Log-in Page

- Participants were given a private session with the system to navigate through the system features. We focused the evaluation of the ontology on only tourist sites in Kuala Lumpur since it is the capital of Malaysia and is the major visiting site. In order to evaluate sites within the whole of Malaysia would take a long time and may not be possible within the scope of this research. Subsequently, participants were asked in the tutorial session to formulate questions in English language to find Hotel, Restaurant and Activity (as they are the main categories of tourism activities refer to 2.1.3) within Kuala Lumpur area and review the retrieved answers with the recommended items. A sample question with the retrieved answers done by the researcher to show the prototype interface is shown in Figure 4.7.

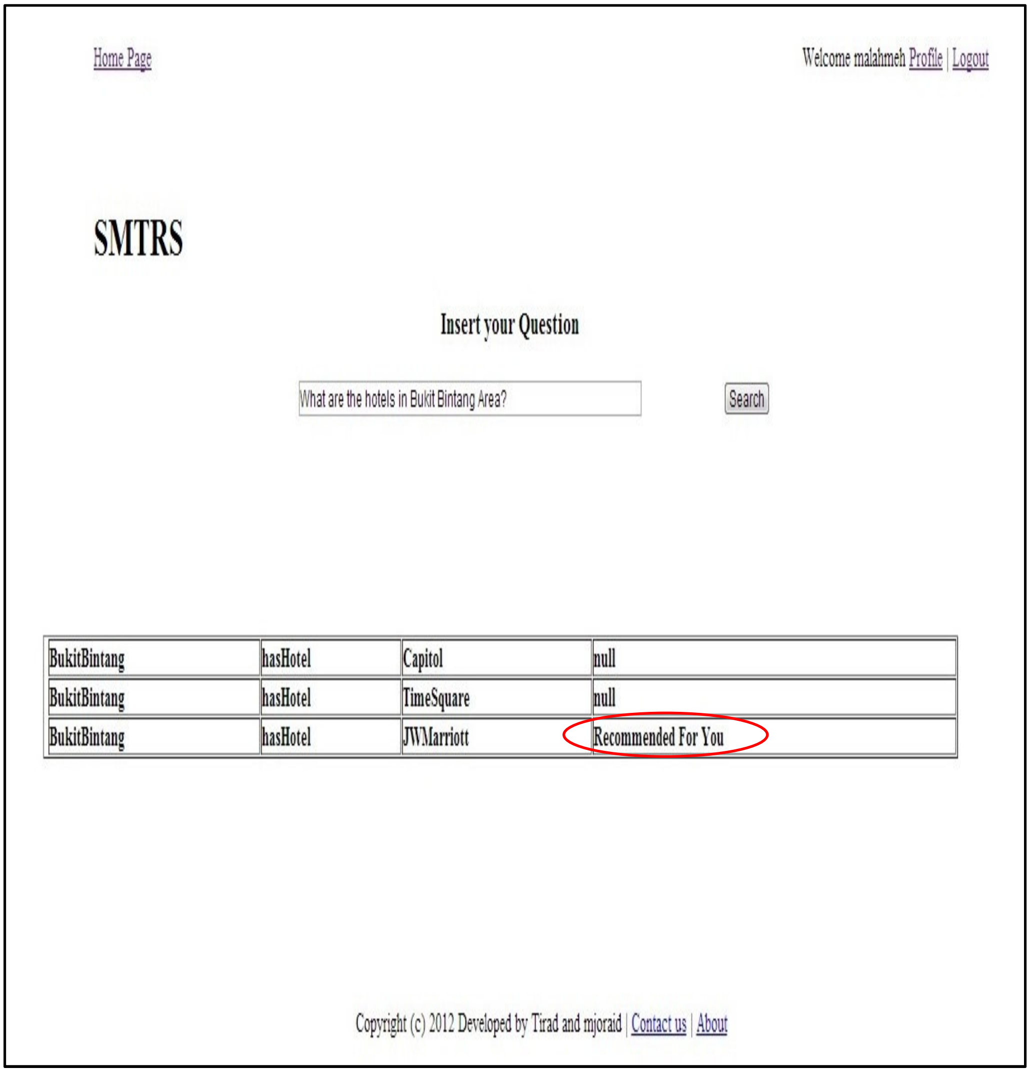


Figure 4.7 Retrieved answers with the recommended item

- **Results Evaluation:** The participants were given the questionnaire (refer to Appendix D) after the private session. The result of the questionnaire is a value between 1 and 100, where 1 signifies that a user found the system absolutely useless and 100 that a user found a system optimally useful. The participants were required to give each answer from a 5-point Likert scale, i.e. rate the system with a scale of 1 as *strongly disagree* to 5 as *strongly agree* based on the following questions:

Q1. I think that I would like to use this system frequently. Q2. I found the system unnecessarily complex. Q3. I thought the system was easy to use. Q4. I think that I would need the support of a technical person to be able to use this system. Q5. I found the various functions in this system were well integrated. Q6. I thought there was too much inconsistency in this system. Q7. I would imagine that most people would learn to use this system very quickly. Q8. I found the system very cumbersome to use. Q9. I felt very confident using the system. Q10. I needed to learn a lot of things before I could get going with this system. Q11. I found the recommendation provided by the system met my interests. Q12. I am not interested in the recommended items. Q13. Information provided in the SMTRS is personalized. Q14. I could not find new items through the recommended items.

- **Results Calculation:**
  - **The Likert scale** which is in the form of 1-5 needs to be converted to 0-4 of the SUS score. Therefore, results are calculated as (score minus one) so the 1 (Strongly Disagree) scored 0 points, 2 (Disagree) scored 1, and so on, where 5 (Strongly Agree) scored 4. To calculate the SUS score, first we sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For the positive questions

1,3,5,7,9,11 and 13 the score contributions are the scale position (participant response) minus one. For the negative questions 2, 4, 6, 8, 10, 12 and 14 the contributions are five minus the scale position.

- **Normalising** the score on the scale from 0 to 100. The original SUS has 10 questions and to obtain the overall value we need to multiply the sum of the scores by (2.5).

  Number of questions (10)* Maximum score per question (4) = 40

  100 / 40 = 2.5

  In our case, we added 4 questions therefore to obtain the overall value we multiplied the sum of the scores by (1.78571).

  Number of questions (14)* Maximum score per question (4) = 56

  100 / 56 = 1.78571

- Sample questioner result calculation is illustrated in Table 4.4.

Table 4.4 Sample questioner result calculation

Question	scale position (A)	Calculation	Score
Q1. I think that I would like to use this system frequently.	4	A-1=	3
Q2. I found the system unnecessarily complex.	1	5-A=	4
Q3. I thought the system was easy to use.	5	A-1=	4
Q4. I think that I would need the support of a technical person to be able to use this system.	1	5-A=	4
Q5. I found the various functions in this system were well integrated.	4	A-1=	3
Q6. I thought there was too much inconsistency in this system.	1	5-A=	4
Q7. I would imagine that most people would learn to use this system very quickly.	5	A-1=	4
Q8. I found the system very cumbersome to use.	2	5-A=	3
Q9. I felt very confident using the system.	4	A-1=	3
Q10. I needed to learn a lot of things before I could get going with this system.	2	5-A=	3
Q11. I found the recommendation provided by the system meeting my interests.	4	A-1=	3
Q12. I am not interested in the recommended items.	2	5-A=	3
Q13. Information provided in the SMTRS is personalized.	5	A-1=	4
Q14. I could not find new items through the recommended items.	3	5-A=	2
Total Score =			**47**
SUS Score =	**47 * 1.78571 =**		**83.92837**

The usability results from the questionnaire of the SMTRS are shown in Figure 4.8.

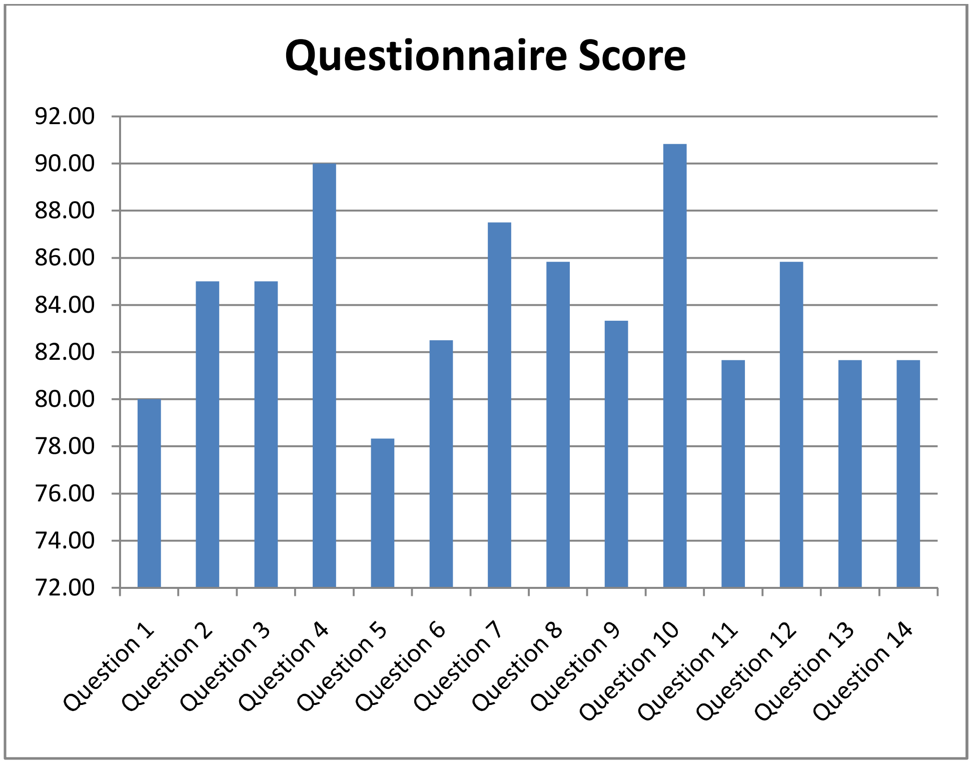


Figure 4.8 Questionnaire Results

From a range of 0 to 100, the participants gave SMTRS an average score of 84.23 (refer to Appendix G). The interpretation of the scores describing the acceptability of SMTRS is EXCELLENT as shown in Figure 4.9.

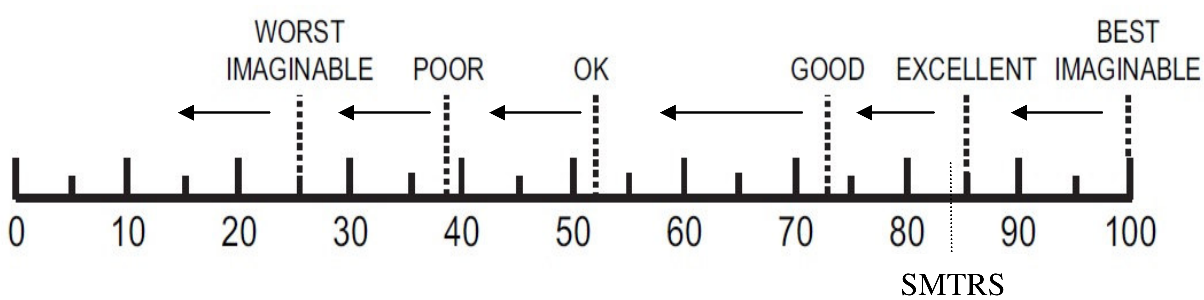


Figure 4.9 The Acceptability of SUS Score Adapted from (Bangor, Kortum, & Miller, 2008)

- **Result Analysis:** The result indicate that the participants found SMTRS usable and provide the participants with information specially recommended for them based on the score. From these, SMTRS has overcome the problem of irrelevant information retrieval in the tourism systems (by giving satisfying answers) aforementioned in chapter 1.2. The SMTRS information recommendation consists of content-based filtering recommendation based upon a description of the items and a profile of the user's interests. Our results show that these compositions are satisfactory to users.

5.0 Conclusion

This research has discussed that the fast development of ICT and the increase of Internet users have reshaped the tourism Industry around the world. ICT has been the backbone of many innovative processes. As a result, incorporating ICT in the tourism business enrich the field with greater productivity, decreased costs, increased revenues and improved customer service. However, information is generated and published by multiple official or unofficial tourism sources. Searching for information for desired spot of vacation is usually difficult and very time-consuming.

For example, tourists usually have problems when trying to find the available choices of accommodation and what is the most suitable to their needs. Moreover, the open and distributed nature of the Web makes it more difficult for Web search engines to find information related to user information. Users feel overwhelmed before finding the intended information, as they cannot process all the provided information. Tourists are often overloaded by relevant and irrelevant information and resources. Providing information with particular interest to the user is still a challenging task for the tourism information systems.

In setting out how we achieved the stated objectives of this thesis, this research involved three major parts:-

I. The first part is concerned with the development of an ontological model for representing the tourism resources.

II. The second part is concerned with the development of a specific architecture for answering tourists' queries based on semantic natural language interface to implement the SMTRS.

III. The third part is concerned with enriching answers to tourist queries (from the second part) by incorporating Content-based recommendation based on the classification of the user preferences.

5.1 Discussion

Literature reveals that the full-sentence query option was significantly preferred to any other method (as has been discussed in 2.2.5). The main purpose of this research was to unravel the tourism information finding process, to assist tourist with relevant information. To achieve this goal, the Semantic Malaysia Tourism Recommender System (SMTRS) was designed and developed. SMTRS adapts earlier techniques (i.e. GATE) and include new techniques (i.e. RSA as described in 3.4.1). It analyzes the tourist full-sentence query by capturing the semantics between the terms in the query and converting it to semantic triples. Then it matches these triples with the MTO in order to retrieve answers. These answers are analyzed and prepared, so they could be matched with the user profile. Finally, the system creates customized recommendation specific to the tourist's profile. This involves recommending items that are similar to the preferences of the user. The similarity of items is calculated based on the features associated with the compared items (as discussed in 3.4.3-V). With this, we give the tourist the relevant information and the necessary recommendation to gain their satisfaction. Consequently, tourist satisfaction will lead to more tourism development.

To achieve our goal, we came out with questions that have been answered in the previous chapters. In Table 5.1, we illustrate where in this thesis these questions have been clarified and answered.

Table 5.1 Sections map showing where in thesis research questions answered

Questions	Chapters	Sections
Q1. What do we understand about the challenges of the current information systems in tourism domain? And what kind of technologies is nominated to overcome those challenges	**2**	**2.1-2.4**
Q2. What is the appropriate recommender system for tourism domain? Is there any need for improvement? And what type of improvement need to be considered?	**2/3**	**2.2, 2.4, 3.4**
Q3. How can we use the semantic technology to allow users the freedom to build natural language questions for Tourism information enquiry?	**3**	**3.3, 3.5**
Q4. How to evaluate the SMTRS efficiency and the MTO?	**4**	**4.2-4.4**

In order to compare SMTRS with other applications in Tourism domain, Table 5.2 illustrate the main features provided by SMTRS (NLI, RS, Semantic Technology and ontology) and the features provided by other systems:-

Table 5.2 Features provided by SMTRS compared to other applications in Tourism domain

Application Name	NLI	RS Approach	Semantic Technology	Ontology	Location	Reference
Harmonise	x	x	x	√	Europe	(Fodor & Werthner, 2005)
OnTour	x	x	√	√	Europe	(Prantner, 2005)
Trip@dvice	x	CBR	x	x	Europe	(Venturini & Ricci, 2006)
Triplehop's TripMatcher	x	CBF	x	x	Europe	(Ricci, et al., 2003)
VacationCoach	x	CBF	x	x	Europe	(Staab, et al., 2002)
mITR	x	CF & CBF	x	x	Italy	(Nguyen, 2004)
Yahoo	√	x	x	x	Worldwide	(Yahoo, 2014)
AquaLog	√	x	√	√	UK	(Lopez, et al., 2007)
SMTRS	√	√	√	√	**Malaysia**	

This thesis presented the evaluation of the SMTRS model as follows:-

- ✓ Results were reported in terms of recall and precision that the SMTRS is capable of retrieving relevant answers with high performance. (Ref to 4.2)

- ✓ Results were reported in terms of building and using MTO as a fundamental component in SMTRS. MTO succeeded in representing the Malaysian tourism knowledge domain and retrieving the knowledge. (Ref to 4.3)

- ✓ Results were reported in terms of system usability, SMTRS achieved users satisfactions as it utilize the user profile in the information retrieval, and SMTRS provide personalized information with particular interest to the user. (Ref to 4.4)

These results demonstrate that SMTRS is applicable in the real-life simulation.

5.2 Contribution

The contributions of this thesis apply to the tourism domain; it demonstrates the benefits of combining Content-base filtering with Natural Language Interface to query the tourism ontology. The contributions of this dissertation are divided into three parts as follows:

- ✓ Supporting the Malaysian tourism industry with MTO, a domain specific knowledge base that provides a classification of the main types of tourism related terms. MTO can be reused and shared for information retrieval systems that involve semantic inference capabilities. Nevertheless, MTO enhance the process of generating tourists' recommendation based on a facilitated knowledge structure that allows matching user preferences with tourism related terms.

- ✓ Natural Language Process Component that converts questions asked in the natural language to query triples. In order to ensure the quality of this conversion semantic and linguistic (RSA & GATE) techniques were employed. GATE is a third party component, which been used to extract information. RSA is a domain dependent algorithm contributed to this research to overcome the GATE limitation. Also, a Semantic approach was used for mapping User words into MTO terms, employing the SPARQL-DL engine in SMTRS model which allows mixing TBox and ABox queries particularly. All these lead to a high retrieval performance refer to 2.3.4.

- ✓ Integrating the Content-based filtering recommender and the user profile to enrich the semantic answers based on NLI tourist query in the SMTRS architecture. The model correlate parameters in the user profile with the same list of parameters of the semantic items content (as discussed in 3.4.3-V).

The SMTRS approach merges the semantic NLPC with Content-based filtering; resulting in recommending machine readable and relevant information. This machine readable and relevance is achieved having answering the tourists' question semantically, and the enrichment of the semantic answers with recommendations based on tourists' preferences. The promising results show that users are satisfied with the system which may ultimately add to the benefits of the tourism industry of Malaysia.

5.3 Future Work

Suggestions for future research lies in four directions:

- **The Ontology**: Enrich the ontology with additional Malaysian tourism related concepts. It is also possible to create an ontology management tool which can be used by information providers.

- **Dynamic User Profile:** While our user profile is static and entered by the user, dynamic user profile accurately captures the changes of user's interest. It will ensure the accuracy of obtaining users interest and capture the dynamic change of user's interest.

- **System Performance**: The experiment conducted in this research consists of a sample of the wide range of tourists sites in Malaysia as collecting all Tourism sites is beyond the time scope limit of this research, in the near future we need to experiment the system by running over a large scale ontology, which will show the implementation time and retrieval performance changes.

- **Query for Services**: The idea for these services is to generate a scheme (query pattern) for the input and output information. The scheme can be discovered after running the system for some time. This scheme can be a major component in aiding answering complex tourism queries.

References

Abowd, G., Atkeson, C., Hong, J., Long, S., Kooper, R., & Pinkerton, M. (1997). Cyberguide: A mobile context-aware tour guide. *Wireless Networks, 3*(5), 421-433.

Adeva, J. J. G., & Calvo, R. (2006). Mining Text with Pimiento. *Internet Computing, IEEE, 10*(4), 27-35.

Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach. *ACM Transactions on Information Systems, 23*(1), 103-145.

Adomavicius, G., & Tuzhilin, A. (2005). 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.

AFP. (2013, Jan). Malaysia rated most Muslim-friendly destination, *Agence France-Presse (AFP)*. Retrieved from http://www.google.com/hostednews/afp/article/ALeqM5ieqSBlkjpMRfgcVLgrMWKcB4e_Yg

Aldebert, B., Dang, R. J., & Longhi, C. (2011). Innovation in the tourism industry: The case of Tourism. *Tourism Management, 32*(5), 1204-1213.

Allinson, J. (2008). Describing Scholarly Works with Dublin Core: A Functional Approach. *Library Trends, 57*(2), 221--243.

Androutsopoulos, I., Ritchie, G. D., & Thanisch, P. (1995). Natural language interfaces to databases – an introduction. *Natural Language Engineering, 1*(01), 29-81.

Answer Yahoo. (2012). Answer Yahoo Retrieved 20/September/2012, from http://answers.yahoo.com/

Ardissono, L., Goy, A., Petrone, G., Segnan, M., & Torasso, P. (2003). INTRIGUE: Personalized recommendation of tourist attractions for desktop and handset devices. *Applied Artificial Intelligence, 17*(8), 687-714.

Arpirez, J. C., Corcho, O., Fernandez-Lopez, M., & Gomez-Perez, A. (2003). WEBODE in a Nutshell. *AI MAGAZINE, 24*.

Asean Affairs. (2010). Malaysian tourism reaps returns. *Asean Affairs, Time international management enterprisesCO., LTD, 19 October 2010*.

Baader, F., & Nutt, W. (2003). *Basic description logics.* Paper presented at the Description logic handbook.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction, 24*(6), 574-594.

Barta, R., Feilmayr, C., Pröll, B., Grün, C., & Werthner, H. (2009). *Covering the semantic space of tourism: an approach based on modularized ontologies*. Paper presented at the Proceedings of the 1st Workshop on Context, Information and Ontologies, Heraklion, Greece.

Bechhofer, S., Horrocks, I., Goble, C., & Stevens, R. (2001). OilEd: A Reason-able Ontology Editor for the Semantic Web. In F. Baader, G. Brewka & T. Eiter (Eds.), *KI 2001: Advances in Artificial Intelligence* (Vol. 2174, pp. 396-408): Springer Berlin Heidelberg.

Belkin, N. J., & Croft, W. B. (1992). Information filtering and information retrieval: two sides of the same coin? *Commun. ACM, 35*(12), 29-38.

Berger, H., Dittenbach, M., & Merkl, D. (2004). An Accommodation Recommender System Based on Associative Networks *Information & Communication Technologies in Tourism ENTER 11* (pp. 216-227): Springer.

Berka, T., & Plößnig, M. (2004). *Designing recommender systems for tourism.* Paper presented at the The 11th International Conference on Information Technology in Travel and Tourism, ENTER, Cairo, Egypt.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*. Retrieved from http://www.sciam.com/article.cfm?id=the-semantic-web.

Bilbao, S., Lejarazu, A., & Herrero, J. (2010). Dynamic Packaging Semantic Platform for Tourism Intermediaries. In U. Gretzel, R. Law & M. Fuchs (Eds.), *Information and Communication Technologies in Tourism 2010* (pp. 617-628): Springer Vienna.

Bio, L. F. (2013). Semantic Web vs. Semantic Technologies Retrieved 11/10/2013, 2013, from http://www.cambridgesemantics.com/semantic-university/semantic-web-vs-semantic-technologies

Breitman, K., Casanova, M. A., & Truszkowski, W. (2007). *Semantic web: concepts, technologies and applications*. Verlag: Springer.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Weerdmeester, A. Thomas & I. L. McLelland (Eds.), *Usability evaluation in industry*: Taylor and Francis.

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet--The state of eTourism research. *Tourism Management, 29*(4), 609-623.

Buhalis, D., & Licata, M. C. (2002). The future eTourism intermediaries. *Tourism Management, 23*(3), 207-220.

Buono, P., Costabile, M., Guida, S., & Piccinno, A. (2002). Integrating User Data and Collaborative Filtering in a Web Recommendation System. In S. Reich, M. Tzagarakis & P. E. Bra (Eds.), *Hypermedia: Openness, Structural Awareness, and Adaptivity* (Vol. 2266, pp. 315-321): Springer Berlin Heidelberg.

Burke, R. (1999). Knowledge-based Recommender Systems. In M. Dekker (Ed.), *Encyclopedia of Library and Information Systems* (Vol. 69). New York.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction, 12*(4), 331-370.

Burke, R. (2005). Hybrid Systems for Personalized Recommendations. In B. Mobasher, S. Anand & R. Burke (Eds.), *Intelligent Techniques for Web Personalization* (Vol. 3169, pp. 133-152): Springer Berlin / Heidelberg.

Cardoso, J. (2005). E-Tourism: Creating Dynamic Packages using Semantic Web Processes, from http://www.w3.org/2005/04/FSWS/Submissions/16/paper.html

Cardoso, J., & Sheth, A. P. (2006b). Developing An Owl Ontology For e-Tourism. In R. Jain & A. Sheth (Eds.), *Semantic Web Services, Processes and Applications* (Vol. 3, pp. 247-282-282): Springer US.

Cardoso, J., Sheth, A. P., & Sheth, A. (2006a). The Semantic Web and Its Applications. In R. Jain & A. Sheth (Eds.), *Semantic Web Services, Processes and Applications* (Vol. 3, pp. 3-33): Springer US.

Cernat, L., & Gourdon, J. (2012). Paths to success: Benchmarking cross-country sustainable tourism. *Tourism Management, 33*(5), 1044-1056.

Chai, J., Horvath, V., Nicolov, N., Stys, M., Kambhatla, N., Zadrozny, W., & Melville, P. (2002). Natural language assistant – A dialog system for online product recommendation *AI Magazine, 23*, 63--75.

Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? *Intelligent Systems and their Applications, IEEE, 14*(1), 20-26.

Chen, J. S., & Gursoy, D. (2000). Cross-cultural comparison of the information sources used by first-time and repeat travelers and its marketing implications. *International Journal of Hospitality Management, 19*(2), 191-203.

Cheverst, K., Davies, N., Mitchell, K., Friday, A., & Efstratiou, C. (2000). *Developing a context-aware electronic tourist guide: some issues and experiences*. Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems, The Hague, The Netherlands.

Chi-Ok, O. (2005). The contribution of tourism development to economic growth in the Korean economy. *Tourism Management, 26*(1), 39-44.

Chin, D., & Porage, A. (2001). Acquiring User Preferences for Product Customization. In M. Bauer, P. Gmytrasiewicz & J. Vassileva (Eds.), *User Modeling 2001* (Vol. 2109, pp. 95-104): Springer Berlin Heidelberg.

Choi, C., Miyoung Cho, Junho Choi, Myunggwon Hwang, Jongan Park, & Pankoo Kim. (2009). *Travel Ontology for Intelligent Recommendation System*. Paper presented at the Proceedings of the 2009 Third Asia International Conference on Modelling and Simulation.

Choi, S., Lehto, X. Y., & Oleary, J. T. (2007). What does the consumer want from a DMO website? A study of US and Canadian tourists' perspectives. *International Journal of Tourism Research, 9*(2), 59-72.

Cimiano, P., Haase, P., & Heizmann, J. (2007). *Porting natural language interfaces between domains: an experimental user study with the ORAKEL system*. Paper presented at the Proceedings of the 12th international conference on Intelligent user interfaces, Honolulu, Hawaii, USA.

Console, L., Torre, I., Lombardi, I., Gioria, S., & Surano, V. (2003). Personalized and Adaptive Services on Board a Car: An Application for Tourist Information. *Journal of Intelligent Information Systems, 21*(3), 249-284.

Corallo, A., Lorenzo, G., & Solazzo, G. (2006). *A semantic recommender engine enabling an etourism scenario*. Paper presented at the Proceedings of the 10th international conference on Knowledge-Based Intelligent Information and Engineering Systems - Volume Part III, Bournemouth, UK.

Costa, H., & Macedo, L. (2013). Emotion-Based Recommender System for Overcoming the Problem of Information Overload. In J. Corchado, J. Bajo, J. Kozlak, P. Pawlewski, J. Molina, V. Julian, R. Silveira, R. Unland & S. Giroux (Eds.), *Highlights on Practical Applications of Agents and Multi-Agent Systems* (Vol. 365, pp. 178-189): Springer Berlin Heidelberg.

Coyle, L., & Cunningham, P. (2003). *Exploiting Re-ranking Information in a Case-Based Personal Travel Assistant*. Paper presented at the The ICCBR'03 Workshop on MixedInitiative Case-Based Reasoning, Trondheim, Norway.

Crotts, J. C. (2000). Consumer decision making and prepurchase information search. In A. Pizam & Y. Mansfeld (Eds.), *Consumer behavior in travel and tourism* (pp. 149-168). New York: Haworth Hospitality Press.

Cunningham, H., Maynard, D., Bontcheva, K., & Tablan, V. (2002). *GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications.* Paper presented at the Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02).

Cunningham, H., Maynard, D., & Tablan, V. (2000). JAPE: a Java Annotation Patterns Engine. *Department of Computer Science, University of Sheffield*.

Damljanovic, D. (2009). Semantic Web and E-Tourism In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology, Second Edition* (pp. 3426-3432).

Damljanović, D. (2011). *Natural Language Interfaces to Conceptual Models.* Doctoral dissertation, The University of Sheffield.

Damljanovic, D., Agatonovic, M., & Cunningham, H. (2010). *Natural Language Interfaces to Ontologies: Combining Syntactic Analysis and Ontology-based Lookup through the User Interaction.* Paper presented at the In Proceedings of the 7th Extended Semantic Web Conference (ESWC 2010), Heraklion, Greece.

Damljanović, D., & Bontcheva, K. (2009). Towards Enhanced Usability of Natural Language Interfaces to Knowledge Bases. In V. Devedžić & D. Gašević (Eds.), *Web 2.0 & Semantic Web* (Vol. 6, pp. 105-133): Springer US.

Damljanović, D., & Devedžic, V. (2008). Applying semantic web to e-tourism. In Z. Ma & H. Wang (Eds.), *The Semantic Web for Knowledge and Data Management: Technologies and Practices* (pp. 243-265). IGI Global, New York.

Dann, G. M. S. (1996). *The language of tourism: a sociolinguistic perspective*. Wallingford, UK: CAB International, First edition (January 4, 1996).

Dittenbach, M., Merkl, D., & Berger, H. (2003). A Natural Language Query Interface for Tourism Information. In A. Frew, M. Hitz & P. O'Connor (Eds.), *Information and Communication Technologies in Tourism 2003* (pp. 152-162): Springer Vienna.

Domingue, J. (1998). *Tadzebao and WebOnto: discussing, browsing, and editing ontologies on the Web.* Paper presented at the Eleventh Workshop on Knowledge Acquisition,Modeling and Management, Banff, Alberta, Canada.

Domingue, J., Fensel, D., & Hendler, J. A. (2011). *Handbook of semantic web technologies Vol. 1, Foundations and technologies*: Springer.

Dotsika, F. (2010). Semantic APIs: Scaling up towards the Semantic Web. *International Journal of Information Management, 30*(4), 335-342.

Duhan, N., & Sharma, A. K. (2011). QUESEM: Towards building a Meta Search Service utilizing Query Semantics. *IJCSI International Journal of Computer Science Issues, 8*(1).

Eirinaki, M., & Vazirgiannis, M. (2003). Web mining for web personalization. *ACM Trans. Internet Technol., 3*(1), 1-27.

Engel, J. F., Blackwell, R. D., & Miniard, P. W. (Eds.). (1995). *Consumer Behavior* (Vol. 8). Orlando, FL: The Dryden Press.

Farquhar, A., Fikes, R., & Rice, J. (1997). The Ontolingua Server: a tool for collaborative ontology construction. *International Journal of Human-Computer Studies, 46*(6), 707-727.

Fellbaum, C. (2010). WordNet. In R. Poli, M. Healy & A. Kameas (Eds.), *Theory and Applications of Ontology: Computer Applications* (pp. 231-243): Springer Netherlands.

Fink, J., & Kobsa, A. (2002). User Modeling for Personalized City Tours. *Artif. Intell. Rev., 18*(1), 33-74.

Fodness, D., & Murray, B. (1997). Tourist information search. *Annals of Tourism Research, 24*(3), 503-523.

Fodor, O., & Werthner, H. (2005). Harmonise: a step toward an interoperable e-tourism marketplace. *International Journal of Electronic Commerce, 9*(2), 11-39.

García-Crespo, A., Chamizo, J., Rivera, I., Mencke, M., Colomo-Palacios, R., & Gómez-Berbís, J. M. (2009). SPETA: Social pervasive e-Tourism advisor. *Telematics and Informatics, 26*(3), 306-315.

Garzotto, F., Paolini, P., Speroni, M., Proll, B., Retschitzegger, W., & Schwinger, W. (2004). *Ubiquitous Access to Cultural Tourism Portals.* Paper presented at the DEXA '04 Proceedings of the Database and Expert Systems Applications, 15th International Workshop.

Gavalas, D., & Kenteris, M. (2012). Evaluation of a web recommender system in electronic and mobile tourism. *International Journal of Web Engineering and Technology, 7*(1), 4-21.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science, 31*(3), 493-520.

Goeldner, C. R., & Ritchie, J. R. B. (2003). *Tourism: Principles, Practices, Philosophies* (9 ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*: Springer Verlag.

Gonzalez, R. (2013). NLP and the Semantic Web Retrieved 11/10/2013, 2013, from http://www.cambridgesemantics.com/semantic-university/nlp-and-the-semantic-web#nlp-definition

Green, B. F., Wolf, A. K., Chomsky, C., & Laughery, K. (1961). *Baseball: an automatic question-answerer*. Paper presented at the Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference, Los Angeles, California.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies, 43*(5–6), 907-928.

Gursoy, D., & Terry Umbreit, W. (2004). Tourist information search behavior: cross-cultural comparison of European union member states. *International Journal of Hospitality Management, 23*(1), 55-70.

Haller, M., Pröll, B., Retschitzegger, W., Tjoa, A. M., & Wagner, R. R. (2000). Integrating Heterogeneous Tourism Information in TIScover — The MIRO-Web Approach. In D. Fesenmaier, S. Klein & D. Buhalis (Eds.), *Information and Communication Technologies in Tourism 2000* (pp. 71-80): Springer Vienna.

Handszuh, H. (2003). *Local Food in Tourism Policies.* Paper presented at the WTO-CTO local food & tourism international conference, Larnaka, Cyprus.

Hinze, A., & Buchanan, G. (2005). Context-awareness in Mobile Tourist Information Systems: Challenges for User Interaction. *in 'Proc. Workshop on Context in Mobile HCI, in conjunction with Mobile HCI', Salzburg, Austria.*

Hinze, A., & Voisard, A. (2003). Location- and Time-Based Information Delivery in Tourism. In T. Hadzilacos, Y. Manolopoulos, J. Roddick & Y. Theodoridis (Eds.), *Advances in Spatial and Temporal Databases* (Vol. 2750, pp. 489-507): Springer Berlin Heidelberg.

Hojeghan, S. B., & Esfangareh, A. N. (2011). Digital economy and tourism impacts, influences and challenges. *Procedia - Social and Behavioral Sciences, 19*(0), 308-316.

Hong, D., Park, Y.-K., Lee, J., Shin, V., & Woo, W. (2005). *Personalized Information Retrieval Framework.* Paper presented at the ubiComp workshop (ubiPCMM).

Horner, S., & Swarbrooke, J. (1996). *Marketing Tourism, Hospitality and Leisure in Europe* London: International Thomson Business Press.

Horrocks, I. (2007). *Semantic web: the story so far*. Paper presented at the Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), Banff, Canada.

Huang, Y., & Bian, L. (2009). A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet. *Expert Systems with Applications, 36*(1), 933-943.

Husain, M., McGlothlin, J., Masud, M. M., Khan, L., & Thuraisingham, B. M. (2011). Heuristics-Based Query Processing for Large RDF Graphs Using Cloud Computing. *Knowledge and Data Engineering, IEEE Transactions on, 23*(9), 1312-1327.

Hyde, K. F. (2008). Information processing and touring planning theory. *Annals of Tourism Research, 35*(3), 712-731.

Inversini, A., & Buhalis, D. (2009). Information Convergence in the Long Tail: The Case of Tourism Destination Information. In W. Höpken, U. Gretzel & R. Law (Eds.), *Information and Communication Technologies in Tourism 2009* (pp. 381-392): Springer Vienna.

Iwasaki, H., Mizuno, N., Hara, K., & Motomura, Y. (2007). *User-Adapted Car Navigation System using a Bayesian Network -Personalized Recommendation of Content.* Paper presented at the Telecommunications, 2007. ITST '07. 7th International Conference on ITS.

Janev, V., & Vranes, S. (2009). Semantic Web Technologies: Ready for Adoption? *IT Professional, 11*(5), 8-16.

Jeong, H. (2011). *Hybrid filtering in Semantic Query Processing.* Ph.D. Dissertation, George Mason University, Virginia.

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration. *Information Systems Research, 15*(2), 194–210.

Kabassi, K. (2010). Personalizing recommendations for tourists. *Telematics and Informatics, 27*(1), 51-66.

Kamar, A. (2003). Mobile tourist guide (m-ToGuide). *Deliverable 1.4, Project final report IST-2001-36004, Mobile Tourist Guide Consortium*.

Kaufmann, E., & Bernstein, A. (2007). *How useful are natural language interfaces to the semantic web for casual end-users?* Paper presented at the Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference, Busan, Korea.

Kaufmann, E., & Bernstein, A. (2010). Evaluating the usability of natural language query languages and interfaces to Semantic Web knowledge bases. *Web Semantics: Science, Services and Agents on the World Wide Web, 8*(4), 377-393.

Kaufmann, E., Bernstein, A., & Zumstein, R. (2006). *Querix: A Natural Language Interface to Query Ontologies Based on Clarification Dialogs*. Paper presented at the 5th International Semantic Web Conference (ISWC 2006).

Kim, H. J., Chen, M.-H., & Jang, S. S. (2006). Tourism expansion and economic development: The case of Taiwan. *Tourism Management, 27*(5), 925-933.

Klein, D., & Manning, C. D. (2003). *Accurate unlexicalized parsing*. Paper presented at the Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, Sapporo, Japan.

Krösche, J., Baldzer, J., & Boll, S. (2004). MobiDENK-Mobile Multimedia in Monument Conservation. *IEEE MultiMedia, 11*(2), 72-77.

Lam, C., & McKercher, B. (2013). The tourism data gap: The utility of official tourism information for the hospitality and tourism industry. *Tourism Management Perspectives, 6*(0), 82-94.

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *SOVIET PHYSICS-DOKLADY, 10*(8).

Linden, G., Hanks, S., & Lesh, N. (1997). *Interactive assessment of user preference models: the automated travel assistant.* Paper presented at the User Modeling: Proceedings of the Sixth International Conference, UM97, Vienna, New York.

López, V., Motta, E., & Uren, V. (2006). *PowerAqua: Fishing the Semantic Web.* Paper presented at the 3rd European Sematic Web Conference, Budva, Montenegro.

Lopez, V., Uren, V., Motta, E., & Pasin, M. (2007). AquaLog: An ontology-driven question answering system for organizational semantic intranets. *Web Semantics: Science, Services and Agents on the World Wide Web, 5*(2), 72-105.

Lops, P., Gemmis, M., & Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. In F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 73-105-105): Springer US.

Mandl, T. (2008). Recent Developments in the Evaluation of Information Retrieval Systems: Moving Towards Diversity and Practical Relevance *Informatica, 32*, 27–38.

Manning, C. D., Raghavan, P., & Schtze, H. (2008). *Introduction to Information Retrieval*. California: Cambridge University Press.

Marchiori, E., & Cantoni, L. (2011). The Online Reputation Construct: Does it Matter for the Tourism Domain? A Literature Review on Destinations' Online Reputation. *Information Technology & Tourism, 13*(3), 139-159.

Marcus, A., & Chen, E. (2002). Designing the PDA of the future. *interactions, 9*(1), 34-44.

Maw, S., & Naing, M. (2006). Personalization agent for multi-agent tourism system. *Proceedings of the 4th international conference on computer application*, 117–124.

Maynard, D., Cunningham, H., Bontcheva, K., Catizone, R., Demetriou, G., Gaizauskas, R., . . . Wilks, Y. (2000). A Survey of Uses of GATE: University of Sheffield.

McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. *W3C recommendation, 10*(2004-03), 10.

Middleton, S. E., Shadbolt, N. R., & Roure, D. C. D. (2004). Ontological User Profiling in Recommender Systems. *ACM Transactions on Information Systems, 22*(1), 54-88.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). *YALE: rapid prototyping for complex data mining tasks*. Paper presented at the Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, PA, USA.

Mistilis, N., & Buhalis, D. (2012). Challenges and potential of the Semantic Web for tourism. *e-Review of Tourism Research (eRTR), 10*(2).

Mohammad Abooyee Ardakan, K. M. (2009). Applying Design Research Method to IT Performance Management: Forming a New Solution. *Journal of Applied Sciences, 7*, 11. doi: 10.3923/jas.2009.1227.1237

Mondeca. (2012). Mondeca, Retrieved 17/12/2012, 2012, from www.mondeca.com

Montaner, M., López, B., & de la Rosa, J. (2003). A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review, 19*(4), 285-330.

Murphy, P., Pritchard, M. P., & Smith, B. (2000). The destination product and its impact on traveller perceptions. *Tourism Management, 21*(1), 43-52.

National Information Standards Organization. (2004). Understanding Metadata: NISO Press.

Nguyen, Q. N. (2004). *Interactivity and Decision Making for On-Tour Travel Recommender Systems.* Paper presented at the In Proceedings of the Ph.D. Workshop at the ENTER 2004 Conference, January 25, 2004, Cairo, Egypt.

Niaraki, A. S., & Kim, K. (2009). Ontology based personalized route planning system using a multi-criteria decision making approach. *Expert Syst. Appl., 36*(2), 2250-2259.

Nicholas, D., Huntington, P., Jamali, H. R., & Dobrowolski, T. (2007). Characterising and evaluating information seeking behaviour in a digital environment: Spotlight on the 'bouncer'. *Information Processing & Management, 43*(4), 1085-1102.

Noy, N. F., Fergerson, R. W., & Musen, M. A. (2000). *The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility*. Paper presented at the Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management.

Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology *Stanford Knowledge Systems Laboratory Technical Report KSL-OI-05*. Stanford University, USA.

O'Grady, M. J., & O'Hare, G. M. P. (2004). Gulliver's Genie: agency, mobility, adaptivity. *Computers & Graphics, 28*(5), 677-689.

Orgun, M., Thornton, J., Bohnert, F., & Zukerman, I. (2007). Using Viewing Time for Theme Prediction in Cultural Heritage Spaces *AI 2007: Advances in Artificial Intelligence* (Vol. 4830, pp. 367-376): Springer Berlin / Heidelberg.

OWL. (2004). OWL Web Ontology Language Reference. *W3C Recommendation, World Wide Web Consortium, http://www.w3.org/TR/owl-ref/.*

Pan, B., & Fesenmaier, D. R. (2006). Online Information Search: Vacation Planning Process. *Annals of Tourism Research, 33*(3), 809-832.

Park, J.-Y., & Jang, S. (2013). Confused by too many choices? Choice overload in tourism. *Tourism Management, 35*(0), 1-12.

Parsia, B., & Sirin, E. (2005). *Pellet: An owl dl reasoner.* Paper presented at the Proceedings of the International Workshop on Description Logics(DL2004), Whistler, BC, Canada.

Pashtan, A., Blattler, R., Heusser, A., & Scheuermann, P. (2003). *CATIS: A Context-Aware Tourist Information System.* Paper presented at the Proceedings of IMC 2003, 4th International Workshop of Mobile Computing, Rostock, Germany.

Pazzani, M., & Billsus, D. (2007). Content-Based Recommendation Systems. In P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), *The Adaptive Web* (Vol. 4321, pp. 325-341): Springer Berlin Heidelberg.

Perez, A. G., Angele, J., Lopez, M. F., Christophides, V., Stutt, A., & Sure, Y. (2002). A survey on ontology tools. *OntoWeb deliverable 1.3, Universidad Politecnia de Madrid*.

Pollock, J. T. (2009). *Semantic Web for dummies*: John Wiley & Sons.

Poon, A. (1993). *Tourism, Technology and Competitive Strategies*. Wallingford: CAB International.

Popescu, A.-M., Etzioni, O., & Kautz, H. (2003). *Towards a theory of natural language interfaces to databases*. Paper presented at the Proceedings of the 8th international conference on Intelligent user interfaces, Miami, Florida, USA.

Poslad, S., Laamanen, H., Malaka, R., Nick, A., Buckle, P., & Zipl, A. (2001). CRUMPET: creation of user-friendly mobile services personalised for tourism. *IEE Conference Publications, 2001*(CP477), 28-32.

Powell, A., Nilsson, M., Naeve, A., & Johnston, P. (2005). DCMI Abstract Model.

Prantner, K. (2005). Ontour. The ontology, DERI Innsbruck from http://e-tourism.deri.at/ont/docu2004/OnTour%20-%20The%20Ontology.pdf

Prud'hommeaux, E., & Seaborne, A. (2007). SPARQL Query Language for RDF (Working Draft).

Rand, G. E. D., & Heath, E. (2006). Towards a Framework for Food Tourism as an Element of Destination Marketing. *Current Issues in Tourism, 9*(3), 206-234.

Rand, G. E. D., Heath, E., & Alberts, N. (2003). The role of local and regional food in destination marketing: A South African situation analysis. *Journal of Travel & Tourism Marketing, 14*(3-4), 97-112.

Rathod, M. D., Prajapati, M. R., & Singh, M. A. (2012). Applying Semantic Web Mining Technologies In Personalized E-Learning. *International Journal of Engineering, 1*(3).

RDF. (2002). Resource Description Framework (RDF). *http://www.w3.org/RDF/*. Retrieved 2011/3/31, 2011

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM, 40*(3), 56-58.

Rhodes, B. J., & Maes, P. (2000). Just-in-time information retrieval agents. *IBM Systems Journal, 39*(3.4), 685-704.

Ricci, F., Fesenmaier, D. R., Mirzadeh, N., Rumetshofer, H., Schaumlechner, E., Venturini, A., . . . Zins, A. H. (2006). DieToRecs: A Case-based Travel Advisory System. In D. R. Fesenmaier, K. W. Wöber & H. Werthner (Eds.), *Destination recommendation systems: behavioral foundations and applications* (pp. 227-239). Wallingford: CABI Publisher International.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to Recommender Systems Handbook. In F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 1-35-35): Springer US.

Ricci, F., Venturini, A., Cavada, D., Mirzadeh, N., Blaas, D., & Nones, M. (2003). Product Recommendation with Interactive Query Management and Twofold Similarity. In K. Ashley & D. Bridge (Eds.), *Case-Based Reasoning Research and Development* (Vol. 2689, pp. 479-493): Springer Berlin Heidelberg.

Roth, J. (2002). *Context-aware web applications using the pinpoint infrastructure.* Paper presented at the Proceedings of the IADIS International Conference on WWW/Internet; Nov2002, p3.

Safiin, N. L. B. (2006). *Developing ontology for Malaysian tourism*. Universiti Teknologi Petronas. Seri Iskandar ,Tronoh Perak, Malaysia. Retrieved from http://utpedia.utp.edu.my/4619/

Samori, Z., & Rahman, F. A. (2013). *Towards the formation of shariah compliant hotel in malaysia: an exploratory study on its opportunities and challenges.* Paper presented at the The 2013 WEI International Academic Conference Proceedings, Istanbul, Turkey.

Scherp, A., & Boll, S. (2004). *Generic support for personalized mobile multimedia tourist applications*. Paper presented at the Proceedings of the 12th annual ACM international conference on Multimedia, New York, NY, USA.

Schiaffino, S., & Amandi, A. (2009). Building an expert travel agent as a software agent. *Expert Systems with Applications, 36*(2, Part 1), 1291-1299.

Schmidt-Belz, B., & Poslad, S. (2003). User Validation of a mobile Tourism Service. *Workshop "HCI in mobile Guides", Udine (Italy).*

Schul, P., & Crompton, J. L. (1983). Search Behavior of International Vacationers: Travel-Specific Lifiestyle and Sociodemographic Variables. *Journal of Travel Research, 22*(2), 25-30.

Siricharoen, W. V. (2008). *Learning semantic web from e-tourism*. Paper presented at the Proceedings of the 2nd KES International conference on Agent and multi-agent systems: technologies and applications, Incheon, Korea.

Sirin, E., & Parsia, B. (2007). *SPARQL-DL: SPARQL Query for OWL-DL*. Paper presented at the OWLED, Innsbruck, Austria,.

Siva, M., & Poobalan, A. (2012). Semantic Web Standard in Cloud Computing.

Smyth, B., Coyle, M., & Briggs, P. (2011). Communities, Collaboration, and Recommender Systems in Personalized Web Search

In F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 579-614): Springer US.

Snepenger, D., Meged, K., Snelling, M., & Worrall, K. (1990). Information Search Strategies B y Destination-Naive Tourists. *Journal of Travel Research, 29*(1), 13-16.

Soe, Y. M., Naing, M. M., & Ni Lar, T. (2006, 5-8 Nov. 2006). *RPCF Algorithm for Multi-Agent Tourism System.* Paper presented at the IEEE International Symposium on Micro-NanoMechatronics and Human Science, 2006, Nagoya, Japan.

Srisuwan, P., & Srivihok, A. (2008). Personalized Trip Information for E-Tourism Recommendation System Based on Bayes Theorem. In L. Xu, A. M. Tjoa & S. Chaudhry (Eds.), *Research and Practical Issues of Enterprise Information Systems II* (Vol. 255, pp. 1271-1275): Springer US.

Srivihok, A., & Sukonmanee, P. (2005). *Intelligent agent for e-tourism: Personalization travel support agent using reinforcement learning*. Paper presented at the Proceedings of the WWW 2005.

Staab, S., Werther, H., Ricci, F., Zipf, A., Gretzel, U., Fesenmaier, D. R., . . . Knoblock, C. (2002). Intelligent systems for tourism. *Intelligent Systems, IEEE, 17*(6), 53-66.

Su, X., & Ilebrekke, L. (2006). A Comparative Study of Ontology Languages and Tools. In A. Pidduck, M. Ozsu, J. Mylopoulos & C. Woo (Eds.), *Advanced Information Systems Engineering* (Vol. 2348, pp. 761-765): Springer Berlin / Heidelberg.

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Adv. in Artif. Intell., 2009*, 2-2.

Sugiyama, K., Hatano, K., & Yoshikawa, M. (2004). *Adaptive web search based on user profile constructed without any effort from users*. Paper presented at the Proceedings of the 13th international conference on World Wide Web, New York, NY, USA.

Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R., & Wenke, D. (2002). OntoEdit: Collaborative Ontology Development for the Semantic Web. In I. Horrocks & J. Hendler (Eds.), *The Semantic Web — ISWC 2002* (Vol. 2342, pp. 221-235): Springer Berlin Heidelberg.

Tan, E. M.-Y., Goh, D. H.-L., Theng, Y.-L., & Foo, S. (2007). *An analysis of services for the mobile tourist*. Paper presented at the Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology, Singapore.

TechWiki. (2012). Semantic Technologies Glossary Retrieved 16/12/2012, from http://techwiki.openstructs.org/index.php/Semantic_Technologies_Glossary

Tourism Malaysia. (2010). *Tourist arrivals & receipts to Malaysia, Avalible at http://corporate.tourism.gov.my/corporate/research.asp*.

Tsarkov, D., & Horrocks, I. (2006). FaCT++ Description Logic Reasoner: System Description. In U. Furbach & N. Shankar (Eds.), (Vol. 4130, pp. 292-297): Springer Berlin / Heidelberg.

Tung, H.-W., & Soo, V.-W. (2004, 28-31 March 2004). *A personalized restaurant recommender agent for mobile e-service.* Paper presented at the e-Technology, e-Commerce and e-Service, 2004. IEEE.

UNWTO. (2010). UNWTO tourism highlights, Avalible at: http://www.unwto.org/facts/eng/pdf/highlights/UNWTO_Highlights10_en_HR.pdf. (21-Feb-2011).

UNWTO. (2011). *UNWTO Tourism Highlights* *http://mkt.unwto.org/sites/all/files/docpdf/unwtohighlights11enlr_1.pdf*.

Uschold, M. (2010). The Semantic Web: Down to Business. Washington, DC: A session at Taxonomy Bootcamp.

Vaishnavi, V., & Kuechler, W. (2007). Design research in information systems. *wwwisworldorg, 22*(2), 1-16.

Vaishnavi, V. a. K., W. (2004, August 16, 2009). Design Research in Information Systems Retrieved 2-feb-2010, 2010, from http://desrist.org/design-research-in-information-systems/

Venturini, A., & Ricci, F. (2006). *Applying Trip@dvice Recommendation Technology to www.visiteurope.com*. Paper presented at the Proceedings of the 2006 conference on ECAI 2006: 17th European Conference on Artificial Intelligence August 29 -- September 1, 2006, Riva del Garda, Italy.

Volz, R. (2004). *Web ontology reasoning with logic databases.* PhD thesis, AIFB, Karlsruhe, 2004.

W3C OWL Working Group. (2009). OWL 2 Web Ontology Language Document Overview, W3C Recommendation 27 October 2009. *W3C*.

Wang, C., Xiong, M., Zhou, Q., & Yu, Y. (2007). PANTO: A Portable Natural Language Interface to Ontologies *The Semantic Web: Research and Applications, 4th*

European Semantic Web Conference, ESWC 2007, Innsbruck, Austria (pp. 473-487).

Weizenbaum, J. (1966). ELIZA-a computer program for the study of natural language communication between man and machine. *Commun. ACM, 9*(1), 36-45.

Werthner, H. (2003). *Intelligent systems in travel and tourism.* Paper presented at the Proceedings of the 18th international joint conference on Artificial intelligence, Acapulco, Mexico.

Werthner, H., & Klein, S. (1999). *Information technology and tourism: a challenging relationship*. New York: Springer Verlag Wien.

Werthner, H., & Ricci, F. (2004). E-commerce and tourism. *Commun. ACM, 47*(12), 101-105.

Winograd, T. (1972). Understanding natural language. *Cognitive Psychology, 3*(1), 1-191.

Woods, W. A. (1997). Conceptual Indexing: A Better Way to Organize Knowledge: Sun Microsystems, Inc.

WTO. (2001). *Thesaurus on Tourism & Leisure Activities*: World Trade Organization

WTOBC. (2001). E-Business for tourism: Practical guidelines for tourism destinations and businesses. *Madrid:World TourismOrganization Business Council.*

Xiang, Z., Gretzel, U., & Fesenmaier, D. R. (2009). Semantic Representation of Tourism on the Internet. *Journal of Travel Research, 47*(4), 440-453.

Xiong, Y.-n., & Geng, L.-x. (2010, Aug.). *Personalized Intelligent Hotel Recommendation System for Online Reservation--A Perspective of Product and User Characteristics.* Paper presented at the 2010 International Conference on Management and Service Science (MASS).

Yahoo. (2014) Retrieved 22 / February / 2014, from www.yahoo.com

Yang, Y., & Marques, N. (2005). User Group Profile Modeling Based on User Transactional Data for Personalized Systems. In C. Bento, A. Cardoso & G. Dias (Eds.), *Progress in Artificial Intelligence* (Vol. 3808, pp. 337-347): Springer Berlin Heidelberg.

Yıldırım, H., Chaoji, V., & Zaki, M. J. (2012). GRAIL: a scalable index for reachability queries in very large graphs. *The VLDB Journal, 21*(4), 509-534.

Zakaria, L. Q., Hall, W., & Lewis, P. (2009). *Modelling image semantic descriptions from web 2.0 documents using a hybrid approach*. Paper presented at the Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services, Kuala Lumpur, Malaysia.

ZHANG, Z. (2005). *Ontology Query Languages for the Semantic Web, A Performance Evaluation.* the Graduate Faculty of The University of Georgia, ATHENS, GEORGIA.

Zuva, K., & Zuva, T. (2012). Evaluation of Information Retrieval Systems. *International Journal of Computer Science & Information Technology (IJCSIT), 4*(3).

Appendixes

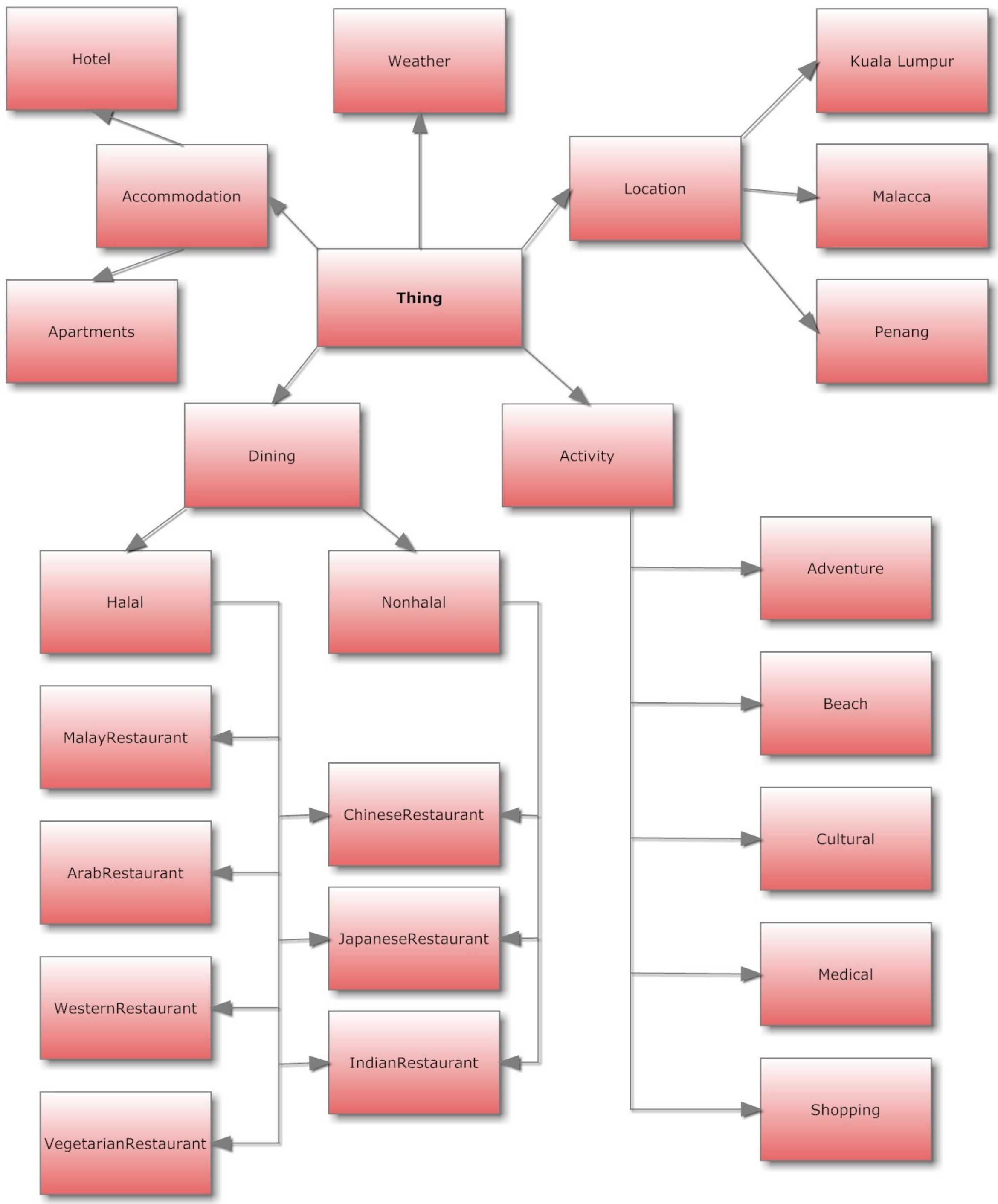
APPENDIX A: User Profile Selection

Accommodation	Amenities	businessCenter
		barAndLounge
		FitnessCenter
		freeBreakfast
		freeWiFi
		freeParking
		kidsActivities
		petsAllowed
		swimmingPool
	Daily budget	Below RM 100
		Between RM 100 and RM 149
		Between RM 150 and RM 199
		Between RM 200 and RM 299
		Over RM 300
Activity	Like to do in trip	SafariPark
		ThemePark
		WaterPark

		Beach
		Museums
		HistoricalSites
		Medical/Relaxing
		Shopping
		Sport
		IndustrialSites
		ArchitectureSites
Dining	Cuisine type	Malay
		Chinese
		Indian
		Western
		Japanese
		Vegetarian
		Arab

APPENDIX B: MTO Entities

MTO ontology



Object Properties

Object Properties	• range • domain • Restriction	hasActivity hasDining hasHotel hasLocation hasWeather isCityin

hasActivity	<owl:ObjectProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasActivity">
hasDining	<owl:ObjectProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasDining">
hasHotel	<owl:ObjectProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasHotel">
hasLocation	<owl:ObjectProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasLocation">
hasWeather	<owl:ObjectProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasWeather"/>
isCityin	<owl:ObjectProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#isCityin">

Datatype Properties

Datatype-Properties	• range • domain • Restriction	hasAddress hasArchitectural hasBeach hasCloseTime hasCuisine hasEmail hasFitnessCenter hasHalal hasHistorical hasIndustrial hasMaximumRoomCharge hasMedical hasMinimumRoomCharge hasMuseum hasOpenTime hasPhone hasPlace hasRate hasRecipes hasRelaxing hasReligious hasReservation hasSafari hasSeason hasShopping hasSmoking hasSport

		hasTheme
		hasView
		hasWater
		hasWaterPark
		hasWebSite
		isCloseTo

hasAddress	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasAddress"/>
hasArchitectural	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasArchitectural">
hasBeach	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasBeach">
hasCloseTime	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasCloseTime">
hasCuisine	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasCuisine">
hasEmail	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasEmail"/>
hasFitnessCenter	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasFitnessCenter"/>
hasHalal	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasHalal"/>
hasHistorical	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasHistorical"/>
hasIndustrial	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasIndustrial">
hasMaximumRoomC	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasMaximumRoomCharge"/>

harge	
hasMedical	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasMedical">
hasMinimumRoomCharge	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasMinimumRoomCharge"/>
hasMuseum	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasMuseum"/>
hasOpenTime	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasOpenTime"/>
hasPhone	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasPhone"/>
hasPlace	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasPlace">
hasRate	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasRate">
hasRecipes	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasRecipes">
hasRelaxing	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasRelaxing">
hasReligious	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasReligious"/>
hasReservation	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasReservation">
hasSafari	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasSafari">
hasSeason	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasSeason"/>

hasShopping	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasShopping">
hasSmoking	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasSmoking">
hasSport	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasSport">
hasTheme	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasTheme">
hasView	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasView">
hasWater	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasWater">
hasWaterPark	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasWaterPark"/>
hasWebSite	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#hasWebSite"/>
isCloseTo	<owl:DatatypeProperty rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#isCloseTo"/>

Example of classes

Description	Class Name	OWL/ RDFS
Name of Class	**Accommodation**	<owl:Class rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Accommodation">
subClassOf	Thing	<rdfs:subClassOf rdf:resource="&owl;Thing"/>
someValuesFrom	Location	<owl:someValuesFrom rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Location"/>
disjointWith	Activity	<owl:disjointWith rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Activity"/>
disjointWith	Dining	<owl:disjointWith rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Dining"/>

Description	Class Name	OWL/ RDFS
Name of Class	**Hotel**	<owl:Class rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Hotel">
subClassOf	Accommodation	<rdfs:subClassOf rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Accommodation"/>
equivalentClass	Appartments	<owl:equivalentClass rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl# Appartments "/>

Description	Class Name	OWL/ RDFS
Name of Class	**Malacca**	<owl:Class rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Malacca">
subClassOf	Location	<rdfs:subClassOf rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Location"/>

Appendix B:

Example of Individuals

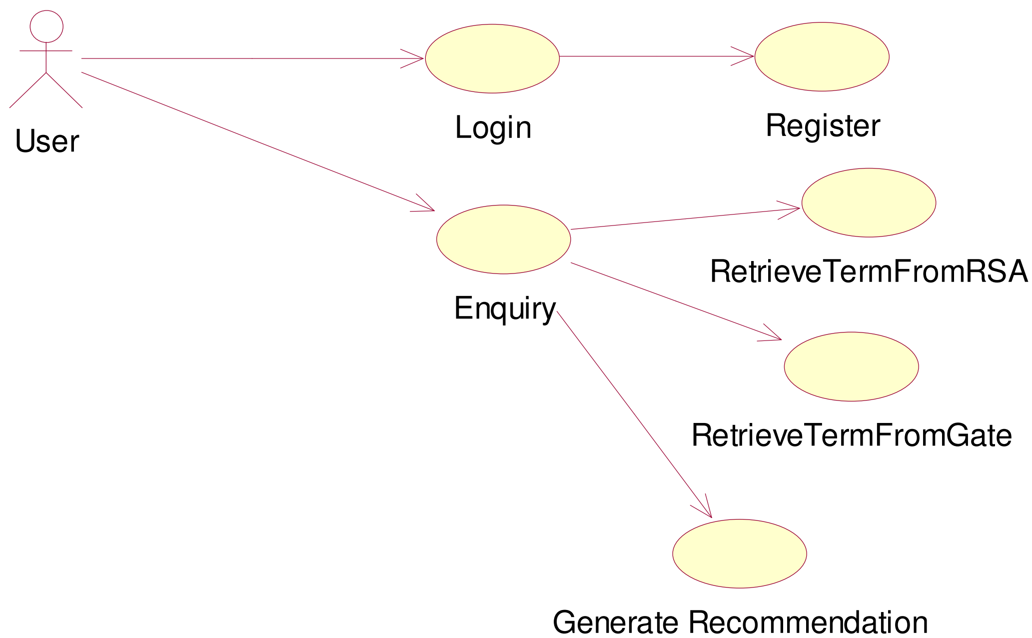
Description	Class Name	OWL/ RDFS
Individual Name	**AFamosa**	<owl:NamedIndividual rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#AFamosa">
Individual type	Activity	<rdf:type rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Activity"/>
hasSafari	SafariPark	<hasSafari rdf:datatype="&xsd;string">SafariPark</hasSafari>
hasWater	WaterPark	<hasWater rdf:datatype="&xsd;string">WaterPark</hasWater>
hasLocation	Malaca	<hasLocation rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Malaca"/>
hasActivity	WaterPark	<hasActivity rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#WaterPark"/>

Description	Class Name	OWL/ RDFS
Individual Name	**BukitBintang**	<owl:NamedIndividual rdf:about="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#BukitBintang">
Individual type	Location	<rdf:type rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl# Location "/>
hasHotel	Capitol	<hasHotel rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Capitol"/>

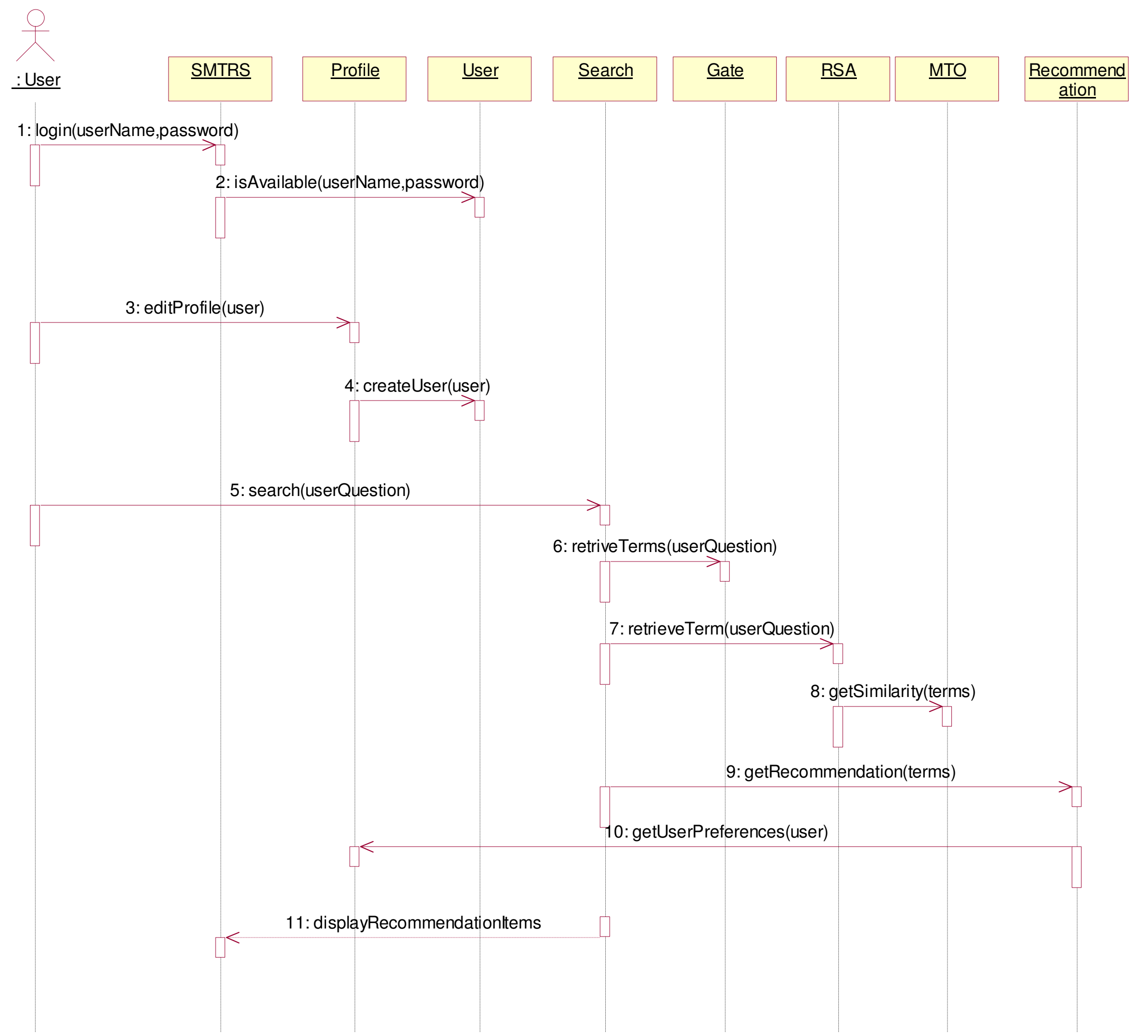
hasHotel	JWMarriott	<hasHotel rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#JWMarriott"/>
hasHotel	TimeSquare	<hasHotel rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#TimeSquare"/>
isCityin	Kuala_Lumpur	<isCityin rdf:resource="http://www.fsktm.um.edu.my/2011/3/MTO-V02.owl#Kuala_Lumpur"/>

APPENDIX C: SMTRS Documentation

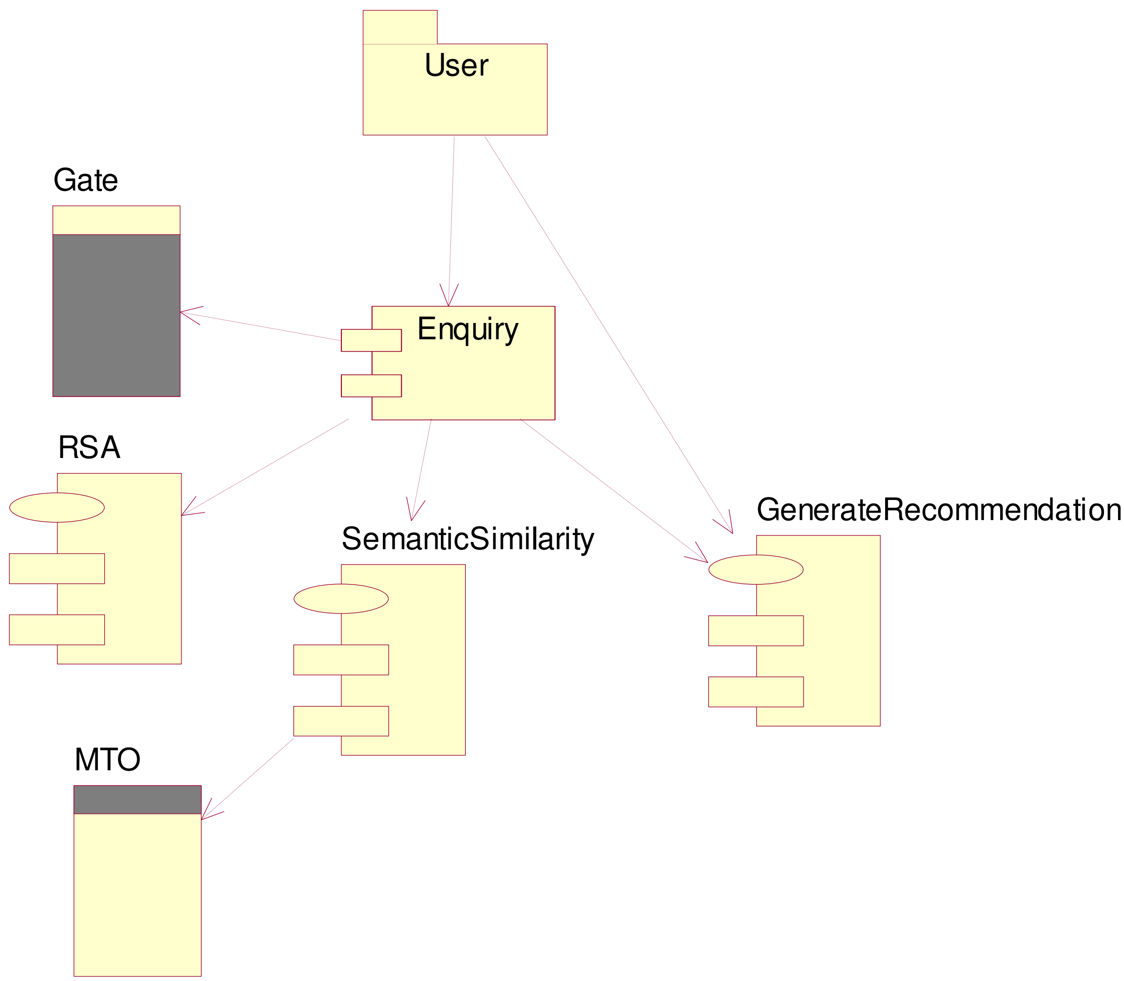
1.0 Use Case



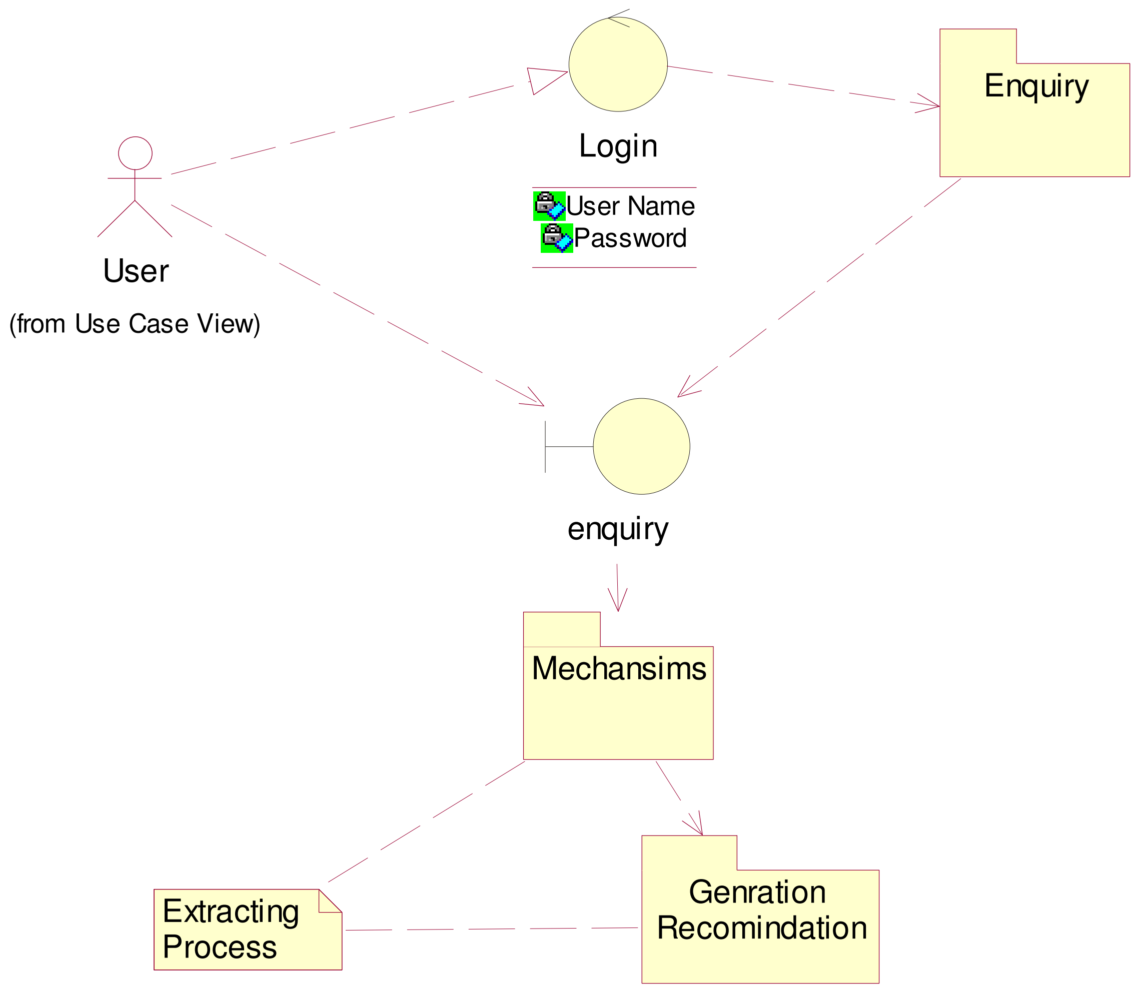
2.0 Sequence Diagram



3.0 Active Component



4.0 Logical View



APPENDIX D: Survey Form

Dear Sir / Madam,

We are conducting this survey to find out your perception of SMTRS usability, so we can better the Malaysia tourism information retrieval facilities, and to understand your needs and offer better service.

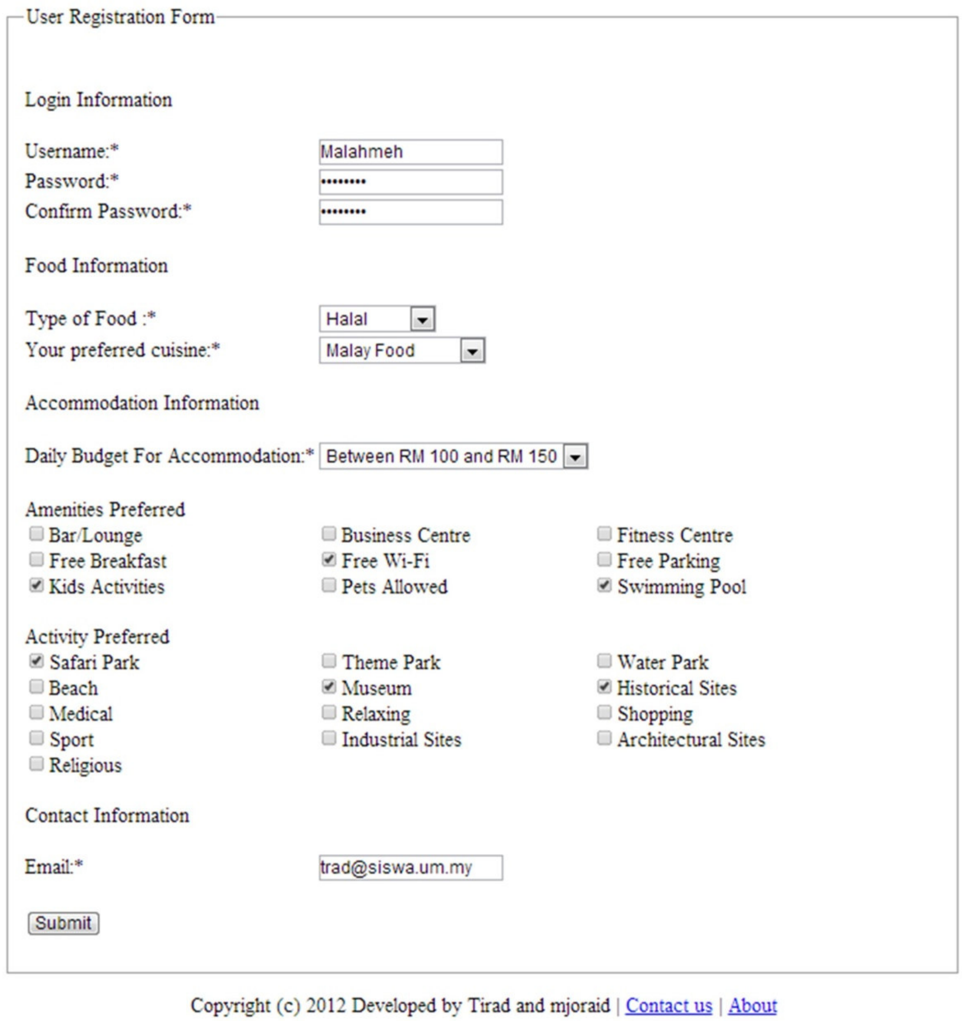
Thank you for taking the time to participate in this activity.

	Questions	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
Q1	I think that I would like to use this system frequently.	1	2	3	4	5
Q2	I found the system unnecessarily complex.	1	2	3	4	5
Q3	I thought the system was easy to use.	1	2	3	4	5
Q4	I think that I would need the support of a technical person to be able to use this system.	1	2	3	4	5
Q5	I found the various functions in this system were well integrated.	1	2	3	4	5
Q6	I thought there was too much inconsistency in this system.	1	2	3	4	5

Q7	I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
Q8	I found the system very cumbersome to use.	1	2	3	4	5
Q9	I felt very confident using the system.	1	2	3	4	5
Q10	I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5
Q11	I found the recommendation provided by the system meeting my interests.	1	2	3	4	5
Q12	I am not interested in the recommended items.	1	2	3	4	5
Q13	Information provided in the SMTRS is personalized.	1	2	3	4	5
Q14	I could not find new items through the recommended items.	1	2	3	4	5

Thank you for your participation

APPENDIX E: User Profile Form



User Registration Form

Login Information

Username:* Malahmeh
Password:* ••••••••
Confirm Password:* ••••••••

Food Information

Type of Food :* Halal
Your preferred cuisine:* Malay Food

Accommodation Information

Daily Budget For Accommodation:* Between RM 100 and RM 150

Amenities Preferred

- [ ] Bar/Lounge
- [ ] Business Centre
- [ ] Fitness Centre
- [ ] Free Breakfast
- [x] Free Wi-Fi
- [ ] Free Parking
- [x] Kids Activities
- [ ] Pets Allowed
- [x] Swimming Pool

Activity Preferred

- [x] Safari Park
- [ ] Theme Park
- [ ] Water Park
- [ ] Beach
- [x] Museum
- [x] Historical Sites
- [ ] Medical
- [ ] Relaxing
- [ ] Shopping
- [ ] Sport
- [ ] Industrial Sites
- [ ] Architectural Sites
- [ ] Religious

Contact Information

Email:* trad@siswa.um.my

Submit

Copyright (c) 2012 Developed by Tirad and mjoraid | Contact us | About

APPENDIX F: User Profile Data

User Number	Type of Food	Preferred Cuisine	Daily Budget for Accommodation	Amenities Preferred									Activity Preferred												
				Bar And Lounge	Business Center	Fitness Center	Free Breakfast	Free WiFi	Free Parking	Kids Activities	Pets Allowed	Swimming Pool	Safari Park	Theme Park	Water park	beach	museum	historical	Medical	Relaxing	shopping	Sport	industrial	Architectural	Religious
1	1	4	4		●		●	●				●			●					●					●
2	1	1	2				●	●				●	●					●			●				
3	1	1	1			●	●	●									●				●				
4	1	3	2					●	●			●		●	●							●			
5	2	5	3	●	●			●				●									●		●		
6	2	5	2			●															●				●
7	1	4	1				●						●							●		●	●		
8	1	4	3					●		●			●		●						●		●		●
9	2	2	4			●		●	●				●	●	●										
10	1	7	2			●		●				●					●						●		●
11	1	1	2			●	●	●										●			●	●			
12	1	2	2						●	●				●							●				●
13	1	3	1				●	●				●			●		●	●		●	●				
14	1	1	3			●		●				●	●								●	●			
15	2	2	3		●							●			●					●	●		●	●	
16	1	1	1				●														●				●
17	1	7	4		●	●			●					●	●						●				●
18	1	4	2				●	●	●			●					●					●			
19	2	2	3				●	●		●					●			●			●	●			●
20	2	6	2			●	●	●					●					●			●				
21	1	3	3				●		●			●					●					●	●		
22	1	3	3					●	●					●							●	●			●
23	1	1	1				●		●			●		●			●				●	●	●		
24	1	4	4				●	●				●		●			●	●			●	●			●
25	1	4	4			●																			
26	2	2	2			●	●	●				●	●		●							●			●
27	1	3	2				●		●	●			●	●	●							●			●
28	2	2	3			●				●					●						●	●			
29	1	2	1				●	●				●	●		●						●	●			
30	1	1	2					●	●			●		●	●						●	●			

Code Description in Next Table

Code Description for the Codes used in User Profile Data Table

Information Type	Code	Descriptions
Type of Food	1	Halal
	2	Nonhalal
Preferred Cuisine	1	Malay Restaurants
	2	Chinese Restaurants
	3	Indian Restaurants
	4	Arab Restaurants
	5	Japanese Restaurants
	6	Western Restaurants
	7	Vegetarian Restaurants
Daily Budget for Accommodation	1	Below RM 100
	2	Between RM 100 and RM 149
	3	Between RM 150 and RM 199
	4	Between RM 200 and RM 299
	5	Over RM 300

APPENDIX G: Questionnaire Score

User Number	Question	scale position	Score
User 1	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	3	2
	Total Score =		**47**
	SUS Score =		**83.92837**
	Question	**scale position**	**Score**
User 2	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	1	4

	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	3	2
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 3	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	3	2

	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	3	2
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	5	4
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 4	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3

	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**47**
	SUS Score =		**83.92837**
	Question	**scale position**	**Score**
User 5	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	3	2
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	5	4
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**49**
	SUS Score =		**87.49979**

	Question	scale position	Score
User 6	Q1. I think that I would like to use this system frequently.	3	2
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**45**
	SUS Score =		**80.35695**
	Question	**scale position**	**Score**
User 7	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3

	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	3	2
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 8	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	5	4

	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	5	4
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**52**
	SUS Score =		**92.85692**
	Question	**scale position**	**Score**
User 9	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	3	2
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	3	2
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	3	2
	Q14. I could not find new items through the recommended items.	3	2

	Total Score =		**39**
	SUS Score =		**69.64269**
	Question	**scale position**	**Score**
User 10	Q1. I think that I would like to use this system frequently.	3	2
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	5	4
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 11	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	2	3

	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	5	4
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**49**
	SUS Score =		**87.49979**
	Question	**scale position**	**Score**
User 12	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3

	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 13	Q1. I think that I would like to use this system frequently.	3	2
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3

	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**47**
	SUS Score =		**83.92837**
	Question	**scale position**	**Score**
User 14	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	5	4
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**51**
	SUS Score =		**91.07121**

	Question	**scale position**	**Score**
User 15	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	3	2
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**47**
	SUS Score =		**83.92837**
	Question	**scale position**	**Score**
User 16	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	3	2
	Q3. I thought the system was easy to use.	4	3

	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	5	4
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 17	Q1. I think that I would like to use this system frequently.	3	2
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4

	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 18	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	5	4
	Q12. I am not interested in the recommended items.	2	3

	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**49**
	SUS Score =		**87.49979**
	Question	**scale position**	**Score**
User 19	Q1. I think that I would like to use this system frequently.	3	2
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**

User 20	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	5	4
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 21	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3

	Q5. I found the various functions in this system were well integrated.	5	4
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**49**
	SUS Score =		**87.49979**
	Question	**scale position**	**Score**
User 22	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	5	4

	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	3	2
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**46**
	SUS Score =		**82.14266**
	Question	**scale position**	**Score**
User 23	Q1. I think that I would like to use this system frequently.	3	2
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	3	2
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3

	Total Score =		**44**
	SUS Score =		**78.57124**
	Question	**scale position**	**Score**
User 24	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**51**
	SUS Score =		**91.07121**
	Question	**scale position**	**Score**
User 25	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	1	4

	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	4	3
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	3	2
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**47**
	SUS Score =		**83.92837**
	Question	**scale position**	**Score**
User 26	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	2	3

	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**48**
	SUS Score =		**85.71408**
	Question	**scale position**	**Score**
User 27	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	4	3
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	5	4
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3

	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**51**
	SUS Score =		**91.07121**
	Question	**scale position**	**Score**
User 28	Q1. I think that I would like to use this system frequently.	3	2
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	2	1
	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	5	4
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	5	4
	Q14. I could not find new items through the recommended items.	1	4
	Total Score =		**48**
	SUS Score =		**85.71408**

	Question	scale position	Score
User 29	Q1. I think that I would like to use this system frequently.	5	4
	Q2. I found the system unnecessarily complex.	1	4
	Q3. I thought the system was easy to use.	5	4
	Q4. I think that I would need the support of a technical person to be able to use this system.	1	4
	Q5. I found the various functions in this system were well integrated.	4	3
	Q6. I thought there was too much inconsistency in this system.	1	4
	Q7. I would imagine that most people would learn to use this system very quickly.	4	3
	Q8. I found the system very cumbersome to use.	2	3
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	2	3
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	1	4
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**49**
	SUS Score =		**87.49979**
	Question	**scale position**	**Score**
User 30	Q1. I think that I would like to use this system frequently.	4	3
	Q2. I found the system unnecessarily complex.	2	3
	Q3. I thought the system was easy to use.	4	3

	Q4. I think that I would need the support of a technical person to be able to use this system.	2	3
	Q5. I found the various functions in this system were well integrated.	5	4
	Q6. I thought there was too much inconsistency in this system.	3	2
	Q7. I would imagine that most people would learn to use this system very quickly.	5	4
	Q8. I found the system very cumbersome to use.	1	4
	Q9. I felt very confident using the system.	5	4
	Q10. I needed to learn a lot of things before I could get going with this system.	1	4
	Q11. I found the recommendation provided by the system meeting my interests.	4	3
	Q12. I am not interested in the recommended items.	2	3
	Q13. Information provided in the SMTRS is personalized.	4	3
	Q14. I could not find new items through the recommended items.	2	3
	Total Score =		**46**
	SUS Score =		**82.14266**
	Total SUS Score =	**84.23**	