CHAPTER THREE

METHODOLOGY

Research Method

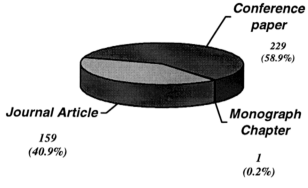
The present study adopts a descriptive research approach, primarily because it is an exploratory study and no local studies could be located on the subject, to describe the quantity, characteristics and productivity of Malaysian publications in the fields of computer science and information technology. The data collected for this study was obtained from three international CD-ROM databases in the area of computer science and information technology with authors affiliated to institutions with Malaysian work address. Bibliometric analysis and regression analysis are also employed to show and determine the spread trend and character of the literatures collected.

Data Collection

The data for this study which was taken from 3 international CD-ROM databases, i.e., *COMPENDEX* (1987-1999), *IEL (IEE/IEEE Electronic Library)* (1988-1999), and *INSPEC* (1990-1998), by using combinations of keywords such as Malaysia, computer, information technology, information systems, and their synonyms, consists of all published works listed in the above databases between the years 1990 and 1998. The process of data collection involved identifying all bibliographic records of Malaysian authors, according to the fields of Affiliation in *INSPEC* and Corporate Source in *COMPENDEX* which reflected the affiliation status of only the first author. Duplicate records were identified and eliminated. The search contributed a total of 389 bibliographic records which consisted of 159 (40.9%) journal articles, 229 (58.9%)

conference papers and 1 (0.2%) monograph chapter. Figure 3.1 shows the distribution of the bibliographic records retrieved.

Figure 3.1: Distribution of Bibliographic Records Retrieved



In view of a variety of differences in spelling of authors' names, for example, confusion of surname with given name, invert order, initial signature, etc., the same single author was recognized, identified and unified manually. Non-Malaysian co-authors were ignored for this current study. The bibliographic detail of information about each journal, such as the location of publisher, was downloaded from the *Ulrich's International Periodicals Directory* (CD-ROM) and the Internet. *Malaysian Research and Development Classification System* (1998) produced by the Malaysian Science and Technology Information Centre (MASTIC) was applied when searching under research subjects.

Malaysian Research and Development Classification System

The Malaysian Research and Development Classification System (MRDCS) (1998) is a document produced by the Malaysian Science and Technology Information Centre (MASTIC) for the purpose of categorising research and development (R&D) activities

carried out in Malaysia. This classification is used by MASTIC in its national R&D survey to categorise related R&D activities undertaken by government research institutes, institutes of higher learning, private sector and non profit organisation in Malaysia.

The present study used the classification system produced by MRDCS. This will not only facilitate the systematic categorisation of R&D information but will also indicate the progress of various R&D activities in the Malaysia. In addition, it allows a more meaningful international benchmarking, as the classification was adapted from the Organisation for Economic Co-operation and Development (OECD) Proposed Standard Practice for Survey of Research and Experimental Development, 'Frascati Manual', 1993.

The classification allows R&D activities in Malaysia to be evaluated by the Field of Research (FOR). The FOR classification is arranged in a hierarchical order. The fields include recognized academic disciplines and related major sub-fields taught at universities or tertiary institutions, major fields or research investigated by national research institutions and organisations, and emerging areas of study. It reflects the expansion of research activity particularly in information, computer and communication technologies, the applied science and technologies, the biological and other life science, mathematics, the social science and the humanities. There are 2 divisions, 15 categories and 126 groups (Malaysian Science and Technology Information Centre, 1998). The research subjects analyzed in the current study refer to:

Division 1: Natural Sciences, Technologies and Engineering relevant to computer science and information technology

Category F10500 Information, Computer & Communication Technologies

Group F10501 Information systems

Group F10502 Hardware

Group F10503 Software

Group F10504 Current information technology

Group F10505 Communication

Group F10506 Security system

Group F10599 Other information, computer & communication technologies not elsewhere classified

Category F10600 Applied Sciences and Technologies

Group F10602 Manufacturing and process technologies and engineering

Group F10604 Educational technology

Database Design

The database created in this study contains collected bibliographic data in order to manage the data more effectively and efficiently. Table 3.1 shows the structure of the database created by using Microsoft Access 97. The Table Record accommodates the 389 bibliographic records of data collected from CD-ROM databases. The database includes 12 fields, i.e.,

1. IDNo – the AutoNumber primary key of the table. A unique sequential (incremented by 1) number assigned by Microsoft Access whenever a new record is added to a table. AutoNumber fields cannot be updated;
2. Title of publication – the English translation, if the original title in other language;
3. Author – all contributors of a specific publication;
4. Number of Author – the sum account of contributors for an individual publication;
5. Author Affiliation – the corporation of the first author;
6. Source of publication – the titles of journals, proceedings or monograph;
7. Issue Number – the identification of publication issuance;
8. Publication Year;
9. Country of Publication;
10. Document Type – the option of journal article, conference paper or monograph chapter;
11. Number of Reference; and
12. Subject of research.

The database provides the following information for analysis:

1. The total number of contributions in computer science and information technology authored by Malaysian researchers submitted in the span of 1990-1998;
2. The subject content of the publications in the discipline of computer science and information technology researched by Malaysian authors;
3. The type of publications authored (journal article or conference paper);

4. The total number of publications in accordance to the geographical distribution of the source;
5. The number of references cited by each publication;
6. The Malaysian authors' names which authored the publications;
7. The co-authorship patterns; and
8. The all authors' affiliations.

Table 3.1: Structure of Database

Database Object	Name	Description
Tables	Record	Store 389 records with 12 fields
Reports	Record	Print Table Record
	Papers by Subject	Print records by Subject
	Papers by Year	Print records by Year
	Papers by Corporate	Print records by Corporation
	Papers by Country of Publication	Print records by Country of Publication
	Papers by Document Type	Print all records by Document Type
	Author	Print all authors with affiliation
	Subject	Print all subjects
	Journal	Print all journal titles
	Conference	Print all conference titles
	Publication	Print all publication titles
	Country of Publication	Print all countries of publisher

Bibliometrics and Statistics Analysis

The Bibliometrics Toolbox (Version 1.4), a small application software developed by T.A. Brooks in 1987, is used to analyze the data retrieved from the Access database. Data saved in a text file was read into this program which automatically performs bibliometric analysis and, then, writes the results to a previously declared file. The outputs can include a brief summary and a bibliograph, a complete listing by groups, a minimum Bradford zonal analysis.

Regression analysis, a form of statistical analysis used for forecasting, estimates the relationship between variables, so that a given variable can be predicted from one or more other variables. By using regression analysis, a trendline is extended in a chart forward or backward beyond the actual data to show a trend and a moving average is created in order to smooth out fluctuations in data and show the pattern or trend more clearly. This method can graphically represent the trends in data series. The linear equation $y = mx + b$, where m is the slope and b is the intercept, calculates the least squares fit for a trendline. Moving average is a sequence of averages computed from parts of a data series. The number of points in a moving average trendline equals the total number of points in the series minus the number specified for the period.

Summary

This Chapter elaborates the research method used to collect and analyze data. In the following chapter, Malaysian computer science and information technology research publication from 1990-1998 will be analyzed using the data collected from the three above-mentioned international databases.