

**INTERNET USAGE AMONG GRADUATE STUDENTS OF
INSTITUTIONS OF HIGHER LEARNING: THE ROLE OF
QUESTIONS IN INITIAL WEB SEARCH STRATEGY**

BY

VIJAYA KUMAR JAYARAJ

**Research report submitted in partial fulfillment of the requirements
for the degree of Master of Business Administration.**

MARCH 2004

ACKNOWLEDGEMENTS

MY SINCERE APPRECIATION TO

My fantastic supervisor, Dr Noornina Dahlan and my advisor, Associate Professor Ramayah Thurasamy for their perceptive critique and wise counsel.

My family and friends, for their valuable support and endless encouragements.

My past and present experiences that continue to have an important impact on my thinking.

TABLE OF CONTENTS

Acknowledgement		ii
Table of Content		iii
List of Tables		v
List of Figures		vi
Abstrak		vii
Abstract		viii
Chapter 1	INTRODUCTION	
1.1	Background	1
1.2	Problem Statement	4
1.3	Objective of the Study	5
1.4	Research Questions	6
1.5	Scope of the Study	6
1.6	Significance of the Study	7
1.7	Definition of Key Terms	7
1.8	Organization of the Chapters	10
Chapter 2	LITERATURE REVIEW	
2.1	Introduction	11
2.2	Review of Literature	12
2.2.1	<i>Background - Web IR</i>	12
2.2.2	<i>General Models of IR and Search Strategy</i>	15
2.2.3	<i>Browsing and Navigating on the Web</i>	17
2.2.4	<i>What is a question?</i>	18
2.3	Theoretical Framework and Hypotheses	20
2.3.1	<i>Framework</i>	20
2.3.2	<i>Hypotheses</i>	23
Chapter 3	METHODOLOGY	
3.1	Introduction	25
3.2	Construct Measurement	25
3.3	Questionnaire Design	27
3.4	Research Design	30
3.5	Data Analysis	32

Chapter 4	RESULTS	
4.1	Introduction	33
4.2	Profile of Respondents	34
	4.2.1 <i>General Demographic Profile</i>	34
	4.2.2 <i>Academic Profile</i>	35
4.3	Descriptive Analyses	37
	4.3.1 <i>Experience with Web Search</i>	37
	4.3.2 <i>Frequency of Web Searching</i>	38
	4.3.3 <i>Internet (Web Search) Environmental Dynamism</i>	39
	4.3.4 <i>Web Search Knowledge</i>	40
4.4	Hypothesis Testing	41
	4.4.1 <i>Hypothesis One</i>	41
	4.4.2 <i>Hypothesis Two</i>	42
	4.4.3 <i>Hypothesis Three</i>	44
	4.4.4 <i>Hypothesis Four</i>	45
	4.4.5 <i>Summary of Results</i>	47
Chapter 5	DISCUSSION	
5.1	Recapitulation of the Study	48
5.2	Discussions	49
5.3	Implications	50
5.4	Limitations	51
5.5	Future Research	51
5.6	Conclusion	52
	REFERENCES	53
	APPENDICES	61

LIST OF TABLE

Table 2.1	An Overview of Perspectives on Questions	18
Table 2.2	Characteristics of the Questions Categories	21
Table 2.3	Characteristics of the Search Strategies (Web Tools)	22
Table 3.1	Questions Category Matrix	28
Table 3.2	Sampling Clusters	31
Table 4.1	Demographic Profile	34
Table 4.2	Ethnicity Profile	35
Table 4.3	Academic Profile	36
Table 4.4	Web Experience	37
Table 4.5	Web Frequency	38
Table 4.6	Web Environment	39
Table 4.7	Web Knowledge	40
Table 4.8	Variety of Strategies Used	41
Table 4.9	Chi-Square Assumption	42
Table 4.10	Search Strategy by Type of Question	43
Table 4.11	Chi-Square Tests	43
Table 4.12	Variety of Strategies Used	44
Table 4.13	Reasons Indicated by Students in Malaysia	45
Table 4.14	Reasons Indicated by Students in Finland / American	46
Table 4.15	Summary of the Hypothesis Tests and Results	47

LIST OF FIGURES

Figure 2.1	Background Schema of Web IR	14
Figure 2.2	Buckland (1999) General Model of IR Systems	15
Figure 2.3	Spoerri (1995) Adapted Search Strategy Model	16
Figure 2.4	The Research Model	20
Figure 3.1	Example of an Answer Box in the Questionnaire	28
Figure 3.2	Operational Definition	29

ABSTRAK

Kajian ini berusaha untuk menggunakan metodologi kualitatif dan kuantitatif untuk menganalisis perbezaan di antara pelajar lepasan ijazah yang sedang melanjutkan pengajian sarjana mereka di empat buah universiti utama dengan pilihan strategi (alamat utama, direktori subjek, dan enjin pencarian) pencarian laman pertama mereka dan rasional pilihan tersebut. Tinjauan ini menggunakan soal selidik untuk mengkaji pengaruh empat jenis soalan, iaitu soalan tertutup/terbuka dan soalan boleh dijangka/ tidak dijangka untuk mencari sumber jawapan muktamad. Walaupun keputusan kajian tidak menunjukkan perbezaan di antara strategi pencarian laman pertama dengan jenis soalan. Kajian ini telah menunjukkan bahawa terdapat perbezaan yang ketara antara kelakuan pencarian laman Barat and Timur, ini menimbulkan soalan tentang kajian masa depan untuk memahami dengan lebih teliti pelakuan pencarian laman tempatan. Hasil kajian ini mempunyai implikasi serantau dalam rekaan mekanisma akses laman pencarian.

ABSTRACT

This research employs both qualitative and quantitative methodology to analyze differences between graduate students studying in four leading public universities and their choice of initial Web search strategies (direct address, subject directory, and search engines) and their reasoning underlying these choices. The research investigates, via an adapted questionnaire, the differences for four types of questions with two variables: closed/open and predictable/unpredictable source of answer. Although the results found no significant differences, between initial search strategies and types of questions, the considerable difference in Web search styles between Western and Eastern cultures may lead to further research to better understand local Web behavior. This study raises issues concerning regional assessment in designing web search access mechanisms and interoperability.

Chapter 1

INTRODUCTION

1.1 Background

Enzer (2004) defined the Internet as the vast collection of inter-connected networks that are connected using the TCP/IP protocols and that evolved from the ARPANET of the late 60's and early 70's. The Internet connects tens of thousands of independent networks into a vast global internet and is probably the largest Wide Area Network in the world.

Given its exponential growth (Aikat, 1998; Fife & Pereira, 2002) and the consequential information overflow in cyberspace (Lenssen, 2003); we are now confronted with excessive retrievability of information that can often paralyze even the simplest search activity.

Eco (1990) philosophically illustrates the Internet's information overflow, "Wanting connections, we found connections - always, everywhere, and between everything. The world exploded in a whirling network of kinships, where everything pointed to everything else, everything explained everything else..."

As a result, the importance of information retrieval has become a critical area of research. Web search is arguably one of the most challenging problems with using the Internet, particularly as an information gateway tool.

The Internet is also an important gateway that bridges the digital divide between developed countries and developing nations (Chen & Wellman, 2003). In a study of Internet use in 26 countries, Creed (2001) reported that Malaysia was still lagging behind other technologically advanced nations and attributed this to the low Internet penetration rate. This is despite the fact that Malaysia's Internet penetration rate had risen steadily with a compound annual growth rate of 13 percent (IDC Market Research, 2003) and more than doubled from 10.5 percent in 2001 to 25 percent in 2003 (Ciolek, 2003).

The Malaysian government is playing a significant role in promoting the ICT industry (Paynter & Lim, 2001), given that the introduction of the Internet in Malaysia was relatively late and Internet usage is very much limited to urban areas (Sharifah, 2003) with 65 percent of the users located in three of the thirteen states; Selangor, Wilayah Persekutuan and Pulau Pinang (Yap, 2000). The latest data shows that Internet users in Malaysia had leaped from 83,718 users in 1987 to 5,700,000 users in 2002.

Information tools, such as the personal computer and the Internet, are increasingly becoming critical to economic success and personal advancement (National Telecommunications and Information Administration, 1999). Ciolek (2003) asserted that the energetic uptake of Internet technologies across Eastern Asia has placed fresh social, cultural and political demands on the region. Regional responses to these new pressures vary in both direction and strength. A new hierarchy has emerged but Malaysia is still some distance behind the leading countries like Japan, Korea, and Taiwan.

Nevertheless, it is important to look at not only who uses the Internet, but to also distinguish varying levels of online skills among individuals. Skill, in this context, is defined as the ability to efficiently and effectively find information on the Web. Hargittai (2002) described this phenomenon as a second-level digital divide, which is relative to Web users' specific abilities to successfully use the Internet.

Although the promotion of higher access rates and further investment in Internet infrastructure would improve Malaysia's position. Some observers argue that improving the human factor, particularly with eradicating cyber-ignorance, is essential for Malaysia to achieve its vision to be an informed and computer-literate society. Khoo (2000) commented:

I reckon the I-Factor, the i for ignorance, will still continue to plague an infant industry taking its first steps. Take it from us. We are quite ignorant when it comes to the Internet's potentials. We are still waiting for someone to lead the way allowing us, in our humble Malaysian demeanor, to follow. Ignorance is not an excuse nor is being led by the nose something to be proud of...

To conjure up a remedy for cyber-ignorance, it is crucial to examine Web users' behavior. Recognizing the dilemma, Prime Minister Datuk Seri Abdullah Ahmad Badawi recently initiated a call for a mindset change among Malaysian, moving away from the physical landmarks "hardware" of development to the "software" part of development with emphasis on human capital (Wong & Wong, 2004), particularly with education and learning (Asia-Pacific Development Information Programme, 2003).

Three out of four, or about 73 percent, of US college students use the Internet for their research more than a conventional library (Pew Internet & American Life Project, 2002). Inevitably, the Internet will be the main source of information and overshadow other conventional channels (Asirvatham, 2003) for information retrieval in academia.

The use of Internet in institutions of higher learning cannot be doubted, even in a developing country like Malaysia (Ramayah, Jantan, & Aafaqi, 2003). Students today have great potential to drive and transform the future of the digital economy (Jones, 2002), and to spearhead an innovative and knowledge-based society (TeAM, 2003).

1.2 Problem Statement

Several research efforts had attempted to incorporate the nature of questions in their design (Clarke, Cormack & Lynam, 2002; Dumais, Banko, Brill, Lin, & Ng, 2002; Jansen, 2000; Kwok, Etzioni, & Weld, 2001) nevertheless; they were not systematically focused on the effects of specific question-related variables on search performance. Chen and Wellman (2003) asserted that the increase in Internet usage has sparked more Web research but emphasized that little research is focused on how Internet access and usage fits into everyday life in developing countries.

Further research to explore the evident research gap, particularly on a local context, is advantageous and meaningful; especially with establishing regional interoperability of Web search applications. Therefore this study investigates the influence of question-related variables on local Web users' initial search strategy.

1.3 Objective of the Study

Early research aimed at determining behavioral characteristics and demographics of Web users via self-selection (Pitkow & Recker 1994) were highly informative, and provided estimations of who is using the Web, but fail to provide detailed information on exactly how the Web is being used.

With the prolific growth of the World Wide Web, there has been an increased demand for a wider understanding of the Web audience (Catledge & Pitkow, 1995). Other areas of audience analysis, such as users interaction with Web resources (Wang, Hawk, & Tenopir, 2000) were being researched.

However, specific search strategy behavior still remains unstudied. Actual user behavior as determined by an analysis of the influence of question-related variables (White & Iivonen, 2001) can supplement the understanding of Web users with more concrete data.

This research also corresponds with Iivonen and White's (2001) study that compared cross-national differences between American and Finnish graduate students' Web search behavior. They strongly suggested further examination across other regional populations because Web users behavioral studies (Kling, 2000) are becoming increasingly important, especially in understanding cross-national diversity (Scull, Milewski, & Millen, 1999) within the borderless cyberspace.

1.4 Research Questions

This study attempts to answer specifically the following questions:

1. What initial search strategy do Web users prefer?
2. Do Web users differentiate their initial choice of search strategies based on the type of questions?
3. What reasons enter into the Web users' initial choice of search strategies?

White and Iivonen (2001) defined initial Web search strategy as a choice among using the direct address, utilizing a subject directory, or searching via a search engine.

1.5 Scope of the Study

The scope of the study is limited to a specific decision point in the Web search strategy behavior, namely the choice of the initial search strategy, of graduate students studying at institutions of higher learning in Malaysia. This is only the first step in a Web search strategy process and it has implications for the continuation of the search (White & Iivonen, 2001).

Students studying at institutions of higher learning are a unique population and have been at the forefront of social change since the end of World War II (Jones, 2002). The selection of the universities as the focal of this study were drawn from their business schools MBA ranking (Asiaweek, 2000) and their respective campuses that are geographically located at high Internet usage sites (Yap, 2000).

1.6 Significance of the Study

This study extends Iivonen and White's (2001) suggestion for further research on question-related variables and search performance, in a different regional and cultural setting.

Research in this area would shed light on local Web users' search behavior and provide vital information for both policymakers and technopreneurs to construct and improve on the regional interoperability of current Web search applications.

1.7 Definition of Key Terms (Retrieved from www.google.com)

ARPANET (Advanced Research Projects Agency Network)

The precursor to the Internet. Landmark packet-switching network established in 1969 by the US Department of Defense as an experiment in wide-area-networking that would survive a nuclear war.

Blogs (Web Logs)

A journal that is available on the Web and are typically updated daily using software that allows people with little or no technical background to maintain the blog.

Cyber-Ignorance

The lack of knowledge or education relating to the Internet, computers or other high-tech technologies.

Digital Divide

The gap in opportunities experienced by those with limited accessibility to technology, especially the Internet. This includes accessibility limitations in social, cultural, disability, economic, and learning issues.

Digital Economy

An economy based on the digitization of information and the respective information and communications infrastructure.

ICT (Information and Communications Technology)

The amalgam of computing, telecommunications technologies, and the Internet; within which information and digital media are created, distributed and accessed.

Internet Penetration Rates

Internet penetration rates provide indicators concerning the proportion of people in a specific area are accessible to the Internet.

Interoperability

The ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner. There are three aspects of interoperability: semantic, structural and syntactical.

Information Retrieval (IR)

Information retrieval or document retrieval is the systematic manipulation of textual information so that it can be easily found again (retrieved). On the WWW, the most important method of IR is the indexing of free-form text. IR exhibits similarities to (but is not the same as) other areas of information processing, such as expert systems and data base management systems.

Knowledge-based Society

The concept means that knowledge and information are viewed as global public goods, and they are tools to enrich the learning environment, support everyday experience, and augment instructional resources.

Second-level Digital Divide

A sub-set of the digital divide, distinguishing varying levels of online skills and the ability to efficiently and effectively find information on the Web.

Technopreneurs

A person who uses the application of digital technology to industrial or commercial objectives or one who organizes, operates, and assumes the risk for a business venture in the Digital Economy

World Wide Web (WWW)

A part of the Internet designed to allow easier navigation of the network through the use of graphical user interfaces and hypertext links between different Uniform Resource Locators (URLs) - frequently called "Web".

1.8 Organization of the Chapters

This paper is organized into five chapters. The first chapter gives a brief background to the subject matter, the research problem, the objectives, the significance, and definition of key terms of the study. Chapter two discusses the literature review on related works done by previous researchers followed by the theoretical framework and hypotheses. Chapter three describes the research methodology. In chapter four, data analysis approaches used to validate the hypotheses is presented, followed by the findings of the study. Finally, in chapter five, this research paper concludes with a discussion of the findings, limitation, suggestion for future research directions, and conclusion of the study.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

Surprisingly, a search for literature revealed no other comparable studies looking at the effect of specific question-related variables on search performance with the exception of White and Iivonen's research papers (1999, 2001, 2002). The gap and prospect to explore this area, particularly on a local context, is evident.

However, the numerous studies on broader Information Retrieval (IR) topics are nevertheless valuable in providing relevant comparisons and to sketch an overview of current research efforts.

The scope of IR literature is broad and diverse, encompassing technical research papers (Bartolini, Ciaccia, & Waas, 2001; Chang & Li, 2002; Fuhr, 1994; Herlocker, Konstan, Borchers, & Riedl, 1999) and extends across the spectrum to communicative articles (Sherman, 2000), library guides (Bishop, 2001), and even Web logs or blogs (KM Blog, 2003) which are typically unsanctioned.

This chapter presents a brief background of Web IR and related discussion on works done by previous research, followed by the theoretical framework and the hypotheses generation.

2.2 Review of Literature

2.2.1 *Background - Web IR*

Vannevar Bush, the "Godfather" of the Internet, inspired many early researchers (Zachary, 1997), by his 1945 groundbreaking article "As We May Think" that conceptualizes the Memex (Keep, McLaughlin, & Parmar, 2000), describing a machine that is remarkably similar to the World Wide Web. Cerami (2003) associated Vannevar Bush's descriptions of "trails between documents" with hyperlinks, "indexes of trails" to search engines, and "publication of trails with commentary" to Blogs.

Suel (2003) described IR as a subfield of computer science, but with roots in library science, information science, and linguistics. The term "Information Retrieval" was actually coined by Calvin N. Mooers in 1948 (Austin, 2001; Gay, 2000). However, only with the advent of modern computers and computer networks did traditional information retrieval truly embark on a new era.

During the 1960s and 1970s, IR systems began to mature into real systems (Lesk, 1995) with the development of computer typesetting and word processing, which meant that lots of text was now available in machine-readable format.

Strangely enough, Lesk (1995) commented that the rise of computer science had negative effects on IR studies because research interest and funding was instead channeled into computer science studies.

Even though information retrieval has existed long before the Internet (Larsen, 1998), the high rate of expansion of the Internet between the 1980s and 1990s (Ratnatunga, 2003) together with the steady decrease in the price of disk space (Grochowski & Halem, 2003) meant that more and more information was available via the Internet and it coincidentally rekindle renewed interest in IR studies.

The exponential growth of the Internet (Fife & Pereira, 2002) had also inevitably established a new branch of information retrieval study. Jansen, Spink, Bateman, and Saracevic (1999) and MacCall (1999) both suggested that Internet information retrieval (Web IR) is very different from traditional IR.

Generally, there are two central focal on Web search studies. One looks at characterizing the Web by the use of content and structural analysis of raw data obtained by Web crawlers (Bray, 1996; Kehoe, Pitkow, Sutton, Aggarwal, & Rogers, 1999; Lawrence & Giles, 1998).

The other looks at characterizing the Web searcher behavior through Web users' surveys (Wang, Hawk, & Tenopir, 2000; White & Iivonen, 2001) or analysis of search engine logs (Jansen, 2000; Spink, Wolfram, Jansen, & Saracevic, 2001).

For instance Jansen, Spink, Bateman, and Saracevic (1998) reported that Web searchers tend to use short queries, exert minimum effort in refining their searches, and most browsed the first page of search results only; which is consistent with Marchionini (1992) findings that Web searchers generally do not want to adopt high involvement in the information retrieval process.

1940s – 1950s

- Vannevar Bush “As We May Think”
- Calvin Mooers coined the term Information Retrieval (IR)

1960s – 1970s

- Mature into real systems - Computer typesetting & word-processing
- Computer Science overshadowing IR studies

1980s – early 1990s

- Expansion of the Internet + Decreasing cost of disk space
- Rekindle research interest in IR studies (Web IR)

late 1990s – 2004

- Two focus (Web IR)

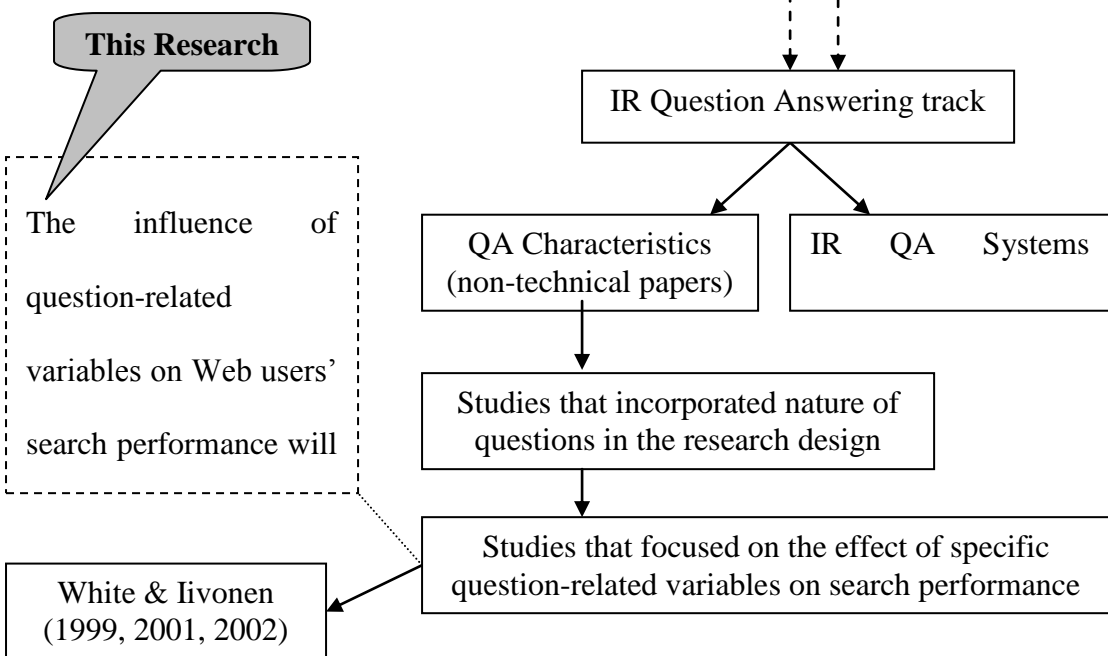


Figure 2.1. Background schema of Web IR.

2.2.2 General Models of IR and Search Strategy

Soboroff (2002) categorized the classical IR models as the Boolean model, the Vector Space model, and the Probabilistic model. However Buckland (1999) suggested a generalized model for information retrieval (see figure 2.2) by excluding all the additional descriptive detail that IR researchers typically insert to test their theories.

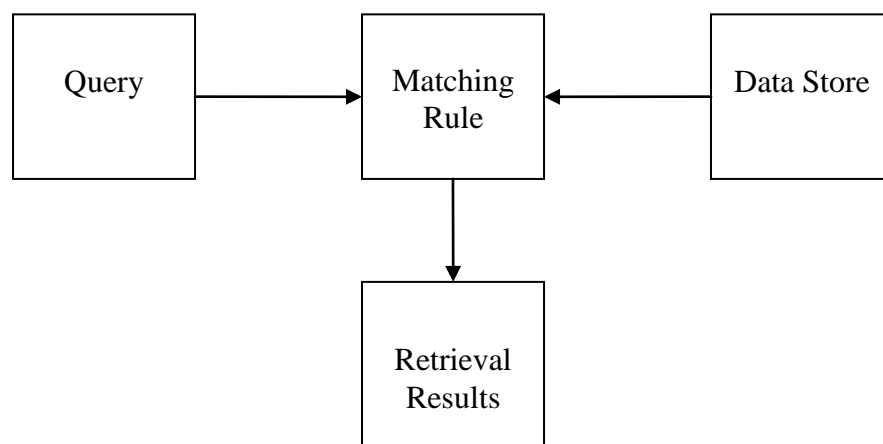


Figure 2.2. Buckland (1999) general model of IR systems.

The general IR model is applicable for both traditional and Web IR systems, However, Web IR acquire data via crawling the Web. Additionally, Grossi (2003) explains that Web IR data store collections are much larger; approximately 3 billion pages or 50 Terabytes (Suel, 2003) and the documents are of very mixed quality and types.

Since Web queries are very short (Jansen et al. 1988), Web IR researchers suggest that Web user must employ search engine manipulation (Suel, 2003) or in other words, a Web “search strategy” for optimal retrieval results.

Spoerri (1995) proposed a general search strategy model by adapting Lancaster and Warner (1993) information retrieval process model. The adapted model is represented by a pyramid diagram (see figure 2.3), where the peak is made visible by users in the form of a conceptual query, that is the result of a conceptual analysis, that operates on the information need, which may be both well or vaguely defined in the user's mind.

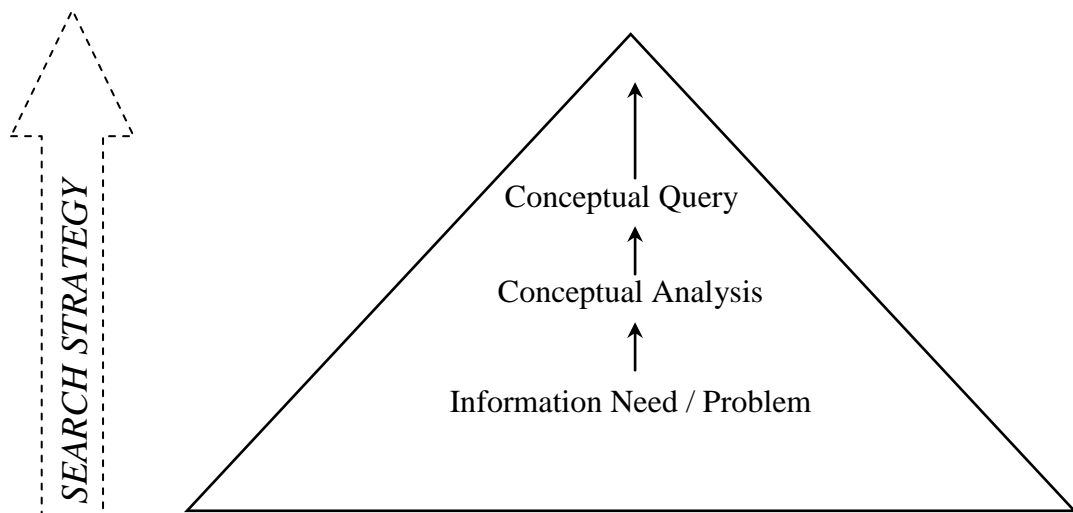


Figure 2.3. Spoerri (1995) adapted search strategy model.

The main aim of an information retrieval system is to effectively provide documents, textual and non-textual, to the information searchers. Nonetheless, human searchers are still required to formulate their search strategy (Wilms, 1988) in a format that can be understood by the particular retrieval mechanism.

Information seeking is a form of problem solving (Marchionini, 1992) where users have to develop knowledge about various sources of information, skills in defining search problems and applying search strategies, and competence in using electronic search tools.

2.2.3. *Browsing and Navigating on the Web.*

Browsing or navigating is a characteristic search strategy on the Web and Web users mainly navigate with frequent use of backtracking (Catledge & Pitkow, 1995) and making quick decisions about which link to click next. Many preliminary studies have addressed user strategies and usability of hypermedia systems, databases and library information systems (Caramel, Crawford, & Chen, 1992) and distinguish between browsing and searching.

Cove and Walsh (1988) describe browsing strategy as:

1. Search browsing or directed search - where the goal is known
2. General purpose browsing - consulting sources that have a high likelihood of items of interest
3. Serendipitous browsing - purely random

Catledge and Pitkow (1995) rationalized that this continuum provides a middle ground to distinguish between browsing and would appear that browsing and searching are not mutually exclusive activities.

Bates (1989) found that Web user's search strategy is constantly evolving through browsing but Palmquist and Kim (2000) argued that Web searchers cognitive styles influence their browsing styles. All the same, White and Iivonen (1999, 2001) had limited their study to a specific decision point in the Web search strategy behavior, namely the first step in a Web search strategy process. This has implications for the continuation of the search, regardless of the browsing strategies employed.

2.2.4 What is a question?

Loos, Anderson, Day, Jordan, and Wingate (1999) defined a “question” as:

- i. A question is an illocutionary act that has a directive illocutionary point of attempting to get the addressee to supply information.
- ii. A question is a sentence type that has a form (labeled interrogative) typically used to express an illocutionary act with the directive illocutionary point mentioned above. It may be actually so used (as a direct illocution), or used rhetorically

Alternatively, IR researcher Pomerantz (2003) defined a “question” for his doctoral study by constructing an overview (see table 2.1) of the IR perspectives on questions.

Table 2.1
An Overview of Perspectives on Questions

Theory	Overview of the Theory	Perspective on Questions
Erotetic logic	Formal logic of questions	A question may be decomposed into a subject and a request.
Speech Act Theory	Decomposition of any utterance into locutionary, illocutionary, & perlocutionary acts.	A question is one step in a larger conversation.
Discourse Analysis	Communication as a process of semantic collaboration.	A question is part of a larger conversation.
Information Needs and Uses	Individuals need information to bridge gaps in their understanding of their situation.	Questions are “the observable behavioral indicators of information needs.”

Given that this research intends to study Web search behavior, “questions” from the perspective of information needs and uses is the most applicable and relevant.

The importance of developing IR question answering (QA) systems got a boost when the Text REtrieval Conference (TREC) added a QA track in 1999 and has since took an entirely different approach to responding to questions as compared to traditional IR QA systems (MacCall, 1999; Pomerantz, 2003).

IR studies are constantly suggesting the importance of considering question characteristics as a factor in retrieval performance. Although several Web search studies had integrated the nature of the questions in their research design (Clarke, Cormack & Lynam, 2002; Dumais, Banko, Brill, Lin, & Ng, 2002; Kwok, Etzioni, & Weld, 2001; Saracevic, Kantor, Chamis, & Trivison, 1988; Wang, Hawk, & Tenopir, 2000), they had neglected to systematically focus on the effects of specific question-related variables on search performance.

Up till now, only White and Iivonen (1999, 2001, 2002) had investigated the influence of question-related variables on Web users' choices of search strategy by analyzing both Finnish and American students' Web searchers behavior.

White and Iivonen (2001) indicated that Web searchers displayed a relatively high degree of familiarity with the initial search options, used different search strategies, and were influenced in their choice of an initial search strategy by question-related characteristics especially with predictable/unpredictable source of the answer. Iivonen and White (2001) also indicated that reasons collected varied across countries raises the importance of considering cultural differences in designing web search access mechanisms.

2.3 Theoretical Framework and hypotheses

2.3.1 Framework

The main objective of this research is to study the influence of question-related variables on Web users' choices of the initial search strategy. Accepting its general validity, a research model was adopted and experimentally tested in the Malaysian context, specifically in institutions of higher learning.

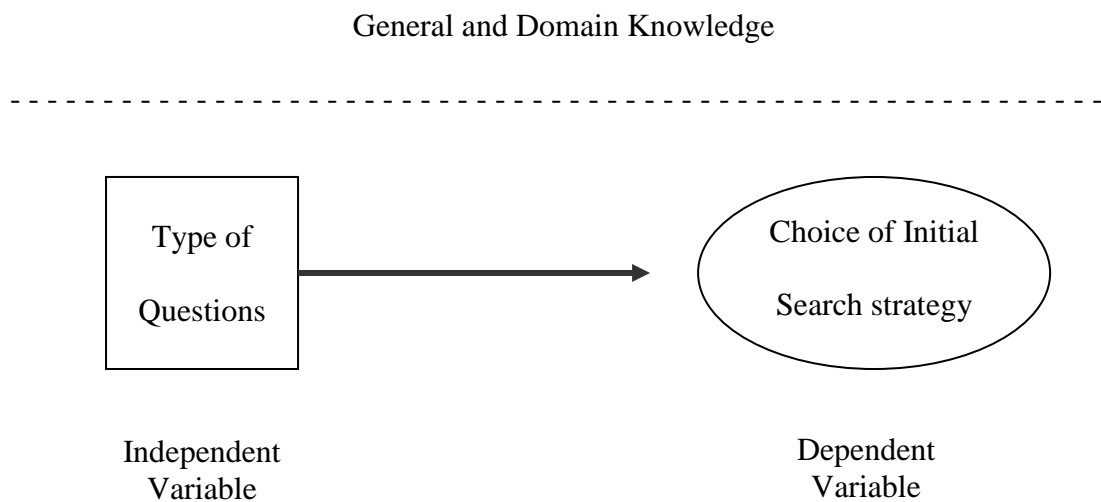


Figure 2.4. The research model.

This research model was adopted from the Choice Behavior Model (White & Iivonen, 2001) that investigated question-related variables on Web users' behavior. This is consistent with the researchers (Iivonen & White, 2001) suggestion for future research in assessing the nature of differences among Web users in countries and cultures that are more dissimilar, particularly Eastern cultures.

For the purpose of this study, the variables listed in the model are defined as follows.

Independent Variable

Types of questions, as articulated by White and Iivonen (2001), are expressions of the information needs raised by information problems, and the way in which they are formulated and phrased has implications for information retrieval and provides insights into an individual's understanding of the problem and of information necessary to address it (see figure 3.2 for the operational definition).

Two questions characteristics are used in this study are:

- i. The open/closed nature of the questions
- ii. The predictable/unpredictable source of the answer

Table 2.2
Characteristics of the Questions Categories

Question Category	Characteristics
Closed	Requiring exact answers and searchers have little discretion in judging correctness or in determining alternatives.
Open	Having no one exact answers and searchers have to apply judgment in developing acceptable responses, sometimes incorporating information from multiple sources.
Predictable Source	Searchers know or can presume with a high probability of success where relevant information can be found.
Unpredictable Source	Searchers may require some searching to locate the appropriate source, if they exist.

Dependent Variable

White and Iivonen (2001) defined initial search strategy to be a choice between searching with a search engine, the direct address, or a subject directory to reach the potentially relevant Web site.

Table 2.3
Characteristics of the Search Strategies (Web Tools)

Web Tools	Characteristics
Search Engines	<p>Multi-step process to reach intended site.</p> <p>A search engine indexes are based on programs that spider the Web continuously locating information without human intervention or evaluation.</p> <p>Searchers use keywords to find Web sites that may contain the information sought. For example, Google.com</p>
Direct Address	<p>Typically takes only one step to reach the intended site.</p> <p>The location, or URL, of a website, file, or resource on the Internet. For example, http://www.usm.my/</p>
Subject Directories	<p>Multi-step process to reach intended site.</p> <p>Similar to conventional library catalogues, classifying Web documents typically organized according to relevant topics and subtopics by human selection and intercession.</p> <p>More selective than search engines and involves recognition rather than recall. For example, Yahoo!</p>

The selection of these search strategies often translates into a decision influenced by considerations about the amount of information, degree of pre-screening, probability of relevance, and amount of effort involved in creating a viable search statement.

2.3.2 Hypotheses

Based on the research model and all the preceding discussion, appropriate hypotheses will be developed to answer the research questions.

Given that the Web contains many potential sources (UIC, 2001), it is important to first examine whether Web users do vary their strategy when searching the Web (About.com, 2004; UNE, 2003), thus leads to the generation of the first hypothesis:

Hypothesis 1: *Web users vary their initial search strategy in searching the Web.*

White and Iivonen (2001) inferred that the distinction between questions, on the basis of their open/closed nature and predictability of the sources of the answers, has significant impacts on Web search strategy choices. As a result, it is proposed that:

Hypothesis 2: *Web users differentiate their initial choice of search strategies based on the type of questions.*

It would be interesting to observe whether one strategy is favored more than the others. Since White and Iivonen (2001) conducted their study in 1998, there has been a huge upsurge in Web search engine's popularity (Online Publishing News, 2002; Sullivan, 2003). With this in mind, the following hypothesis is proposed:

Hypothesis 3: *Web users prefer using search engines for their initial search strategy.*

Consistent with the digital divide theory (Ciolek, 2003; Hargittai, 2002), this study further examines the regional differences of Web users as suggested by Iivonen and White (2001) and postulates the following hypothesis:

Hypothesis 4: There is a difference between the variety and number of reasons indicated by students studying at institutions of higher learning in Malaysia compared to their US and Finnish counterparts.