DURBAN UNIVERSITY OF TECHNOLOGY

Web Search Engines as Teaching and Research Resources:

A Perceptions Survey of IT and CS staff from Selected Universities of the KwaZulu-Natal and Eastern Cape Provinces of South Africa

By

Paul A Tamba Tamba

20929775

A dissertation submitted in fulfillment of the requirements for the degree of Master in Technology

Department of Information Technology

Faculty of Informatics and Accounting

SUBMISSION APPROVED FOR EXAMINATION

SUPERVISOR	DATE
Prof SD Eyono Obono	

SUPERVISORS PERMISSION TO SUBMIT FOR EXAMINATION

Date: 07 December 2010

Student Name: PA Tamba Tamba

Student no.: 20929775

Dissertation Title: Web Search Engines as Teaching and Research Resources: A Perceptions Survey of IT and CS staff from Selected Universities of the KwaZulu-Natal and Eastern Cape Provinces of South Africa.

As the candidate's supervisor I agree to the submission of this dissertation for examination.

The above student has satisfied the requirements of English Language competency.

Name of Supervisor: Prof SD Eyono Obono

DECLARATION

I Paul A Tamba Tamba, declare that this dissertation represents my work and that it has not been submitted in any form for another degree or diploma at any other institution of higher learning.

Information derived from published or unpublished work of other has been acknowledged in the text and a list of reference is presented.

____

Signature

ACKNOWLEDGEMENTS

A number of individuals had contributed to achieving the completion of this dissertation.

I would like to acknowledge the Almighty God for giving me the strength and inspiration at all times.

I also would like to acknowledge my supervisor, for His support, advices that shaped my career to date.

A special thank also goes to the Eyono-Obono family in Durban for the moral, financial, and caring support throughout these years.

A special thanks is attributed to the work on *Analysis Of The Perceptions Of Expatriate Academics On The Factors Affecting Their Work Performance* By Pauline Ngo Henha from which a considerable contribution was made in guiding this dissertation.

A special thanks to my family for their moral and financial supports.

I thank the department of IT and the Postgraduate office in DUT for the precious resources provided throughout this research

Thanks to all those close and far, who made any sort of support and motivation into this dissertation, I thank you all.

May the good Lord bless you mightily.

ABSTRACT

This study examines the perceived effect of the following factors on web searching ability of academic staff in the computing discipline: demographic attributes such as gender, age group, position held by the academic staff, highest qualification, etc; lecturing experience, research experience, English language proficiency, and web searching experience. The research objectives are achieved using a Likert-scale based questionnaire for 61 academic staff from Information Technology and Computer Science departments from four Universities from the Kwazulu-Natal and Eastern Cape provinces of South Africa. Descriptive and inferential statistics were computed for data analysis from the questionnaire after performing data reliability and validity tests using factor analysis and Cronbach's coefficients methods on the PASW Statistics 18.0 (SPSS).

Descriptive statistics revealed a majority of staff from IT as compared to staff in CS and, a majority of under qualified middle age male staff in junior positions with considerable years of lecturing experience but with little research experience. Inferential statistics show an association between web searching ability and demographic attributes such as academic qualifications, positions, and years of research experience, and also reveal a relationship between web searching ability and lecturing experience, and between web searching ability and English language ability. However, the association between position, English language ability, and searching ability was found to be the strongest of all.

The novelty finding by this study is the effect of lecturing experience on web searching ability which has not been claimed by existing research reviewed. Ideas for future research include mentoring of academic staff by more experienced staff, training of novice web searchers, designing and using semantic search systems both in English and in local languages, publishing more web content in local languages, and triangulating various research strategies for the analysis of the usability of web search engines.

TABLE OF CONTENTS

CHAPTER ONE

Introduction

1.1 Background

The first computers were big machines only accessible to university researchers, government intelligence structures and organisations, and the military. The evolution from these workstations and mainframe computers led to the invention in the 1980's of the Personal Computer (PC). Even more spectacular was the invention of the Internet and the World Wide Web (WWW) in the 1990's which allowed networks of PCs and users to communicate and share knowledge, data, and information from different parts of the world. With the arrival of the Internet, more and more users, including academic staff, connect daily to computer networks for different purposes. Nowadays, the popularity of the internet has migrated to the wireless environment with end users moving around with mobile devices such as cell phones, handsets, PDA, etc: Internet can be accessed anywhere network is available and anytime. Examples of applications and services available on the WWW include emails, web transactions, static pages, social networking and discussion forums, instant messaging, just to name a few.

By 2003, there were over one billion static documents on the WWW accessed by over 200 million users internationally (Fensel et al., 2002). This enormous amount of data makes it increasingly difficult to find, access, present, and maintain the information required by a wide variety of online users.

Internet users can be classified based on the nature of their online activities. There is a minority of Internet users in charge of creating, maintaining, and monitoring Internet applications and services. These users are usually called Advanced Internet Users or Internet (System) Administrators. These administrators are specialised Information

Technology professionals whose job description includes the design of web based information systems. On the other hand, the majority of Internet users are End Users. These are people who use and consume services and applications online. This research focuses on academic staff as internet end users for the online search of teaching and learning resources.

Information Technology and Computer Science academic staff also falls under this category of internet users who utilize online search for their daily work related activities. These staffs are mostly involved in three types of activities: teaching, research and community engagement. Teaching activities include curriculum design and delivery, and assessment. Research activities mostly include the conduct of research projects, publications, and conference attendance. Community engagement involves extracurricular projects and linkages with various organizations such as government, businesses, and civil society. As mentioned earlier, Internet and web search engines in particular were initially invented to facilitate the sharing of information and knowledge between academic staff. This function of the Internet is even more used nowadays with the adoption of various web tools such as digital libraries, online encyclopedia, e-learning platforms, academic and professional association portals, etc.

1.2 Problem Statement

Although the Internet can be accessed easily at any time, a particular web site can only be accessed using a Universal Resource Location (URL). A URL is pointer used for the identification of the location of a web resource. In other words, it represents the exact location of a document on a particular remote server connected to the WWW. A URL must therefore be remembered by end users; this is a very difficult exercise having in mind that there are billions of websites on the WWW. Finding information online therefore becomes a challenging task for users when they do not know the URL of the documents they are looking for. This challenge is in principle alleviated by the use of web search engines in guiding the user towards the right documents. This task becomes even more difficult for novice users because most search engines map users' queries to URL's documents using a syntactic search algorithm whose inputs are keywords representing these queries. Even experienced users do not always succeed in their web search.

Results from existing search engines are usually imprecise as they often yield matches to many thousands of irrelevant web pages, making it difficult for users to find information they are looking for. This leads to a gap between content relevance as expected by users and computed search output relevance as estimated by search engines algorithms (Adda et al. 2009). Moreover, end users face the task of first reading the documents retrieved so that they can extract the desired information. This is because information content is primarily presented in human languages. In addition, it has also been observed that semantic search algorithms in principle, able to output more relevant content are not yet implemented by most search engines (Fensel et al., 2002). Direct consequences of the above highlighted challenges on the user include: waste of time, stress, discouragement, frustration, just to name a few.

1.3 Aim, Objectives and Research Question

The above highlighted problems cannot be generalised to all Internet users otherwise this would mean the total failure of the WWW. In other words, on one hand, certain users are frustrated by their failures on the WWW; but other users enjoy the power of Internet search. This raises the research question as to why certain users are successful online searchers while other users are not? In other words, what are the factors that affect the online searching ability of academic staff? And how can these factors be mitigated?

The aim of this research is to analyse the perceived factors that affect the Internet searching ability of academic staff in computing related disciplines.

In order to achieve the above defined aim, it is first necessary to reach the following research objectives:

a. To model the perceived factors affecting the Internet searching ability of academic staff;

b. To analyse the relationships between the above identified factors;

c. To recommend solutions for the mitigation of Internet searching ability factors for academic staff.

1.4 Rationale

The raison d'être of this research is anchored in the nature of the activities carried out by academic staff that require them to continuously update their knowledge and skills. In particular, academic staffs from the computing field are even a better target for this research because computing is a continuously changing and innovating discipline. It is assumed that academic staff from the computing fields needs the power of Internet search for the upgrading of their skill and knowledge. This research can therefore help to improve their web searching skills towards the effective acquisition of relevant information needed for their teaching, research and community engagement activities. The research is also expected to ease collaboration and information sharing among academic staff allowing them to learn from one another. The research outputs could finally be extended to other disciplines of education for the benefit of a broader audience of academic stakeholders.

1.5 Methodology

This research consists of a questionnaire based survey of users' perceptions on their web searching abilities. The Questionnaire is made up of six sections with each section representing a research variable. The last section of the questionnaire represents the research dependent variable on users' perceived web searching ability. The first section represents users' demographic data. The other four research variables are: perceived academic staff lecturing experience, their perceived research experience, their English language proficiency, and their prior search experience. Five of the six research variables are measured using 5 point Likert scale items, only the first variable on the respondents' demographic data contains nominal and ordinal items.

The research population of this study is made up of academic staff in the computing discipline. This population is represented in the research by a sample of 61 Information Technology and Computer Science staff members spread among four (4) universities (Durban University of Technology-DUT, University of KwaZulu Natal-UKZN, University of Zululand-UZ, and Walter Sisulu University-WSU) in the KwaZulu Natal and Eastern Cape provinces of South Africa. Research data collected from the study is analysed using descriptive and inferential statistics using PASW Statistics 18.0 (SPSS) to respond to the following research questions: What is the English language proficiency of academic staff and how does it affect their perceived web searching abilities? Does teaching and research experience contribute to improved online searching ability? Are academic staffs in the study experienced enough in terms of teaching, research, and web searching? Is there a relationship between web searching experience and web searching ability? How do academic staffs from different demographic background perform with regards to language proficiency, teaching and research experience, and web searching experience and ability?

A detailed description of the research design and methods used in this study is given in chapter four. The next session of this chapter describes the structure of this dissertation followed by assumptions on the limitations of this research just before the conclusion of this chapter.

1.6 Structure of the Dissertation

This dissertation consists of seven chapters. The introductory chapter briefly describes the evolution of the use of computers from the mainframes era to today's information highways. It singles out WWW as a major interface for the use of computers in the Information Age; however, it also highlights challenges faced by users when accessing information from web search engines. The first chapter consequently presents the aim of this study towards the analysis of the factors affecting the perceived searching ability of academic staff; this is done after an explanation on the choice of academic staff as the research population of this study. Chapter one ends with a brief description of the research methodology adopted by the study whereby academic staff from the field of computer science and Information Technology from selected universities are presented as the sample of this research.

The second chapter of this dissertation presents definitions and concepts on the types of web search engines, their history backdated from the Roman and Greek era, the traditional information retrieval models their structure, and their use. The third chapter presents an extensive literature survey of previous work on the factors affecting searching ability of Internet users, the methodologies used by these studies, as well as their recommendations.

The fourth chapter describes the methodology adopted in this research. It presents descriptive and inferential statistical data analysis on research questions such as: Is there a relationship between web searching experience and web searching ability? How do academic staffs from different demographic background perform with regards to language proficiency, teaching and research experience, and web searching experience and ability? The results of these descriptive and inferential statistical tests are presented in chapter five. Chapter six discusses these results compared to existing literature on the factors affecting searching ability of internet users and makes recommendations on implications of the research results on the design and use of web search engines. Finally, chapter seven presents a summary of this study as well as possible avenues for future work emanating from the study.

1.7 Limitations

This research study has some limitations due to its sampling methods, its reliance on users' perceptions, and the nature of the searching activities undertaken by academic staff on Internet. In this study, these activities are restricted to the use of internet for teaching and research, excluding any other personal searching activities such as market information, social networking, general information, etc. On the other hand, the collected research data relies on participants' perceptions, having in mind that it might be necessary to adjust participants' perceptions on the factors affecting the Internet searching abilities compared to the actual values of these factors. Secondly, the choice of specific universities with their own respondents' characteristics may yield results whose outcomes may differ if other universities were chosen. Finally, the majority of the questionnaires items are restricted to syntactic search issues therefore assuming that search engines use syntactic based algorithm.

1.8 Chapter Summary

In this chapter, a presentation on the evolution of use of computers from the era of mainframes after World War II to the advent of Internet and the World Wide Web in the 1990's. The importance of Internet and the role of search engines were thereafter presented before listing its applications such as emails, social networks, static and dynamic websites. This chapter highlighted the growth of the internet sites and its users throughout the years, posing the evident and critical problem of document accessibility associated to the non implementation of semantic search algorithm. It was also mentioned in this chapter the contrast whereby semantic search is more efficient than syntactic search, but most search engines still use syntactic search algorithms. A motivation on the choice of academic staff from computing discipline was also presented as its rationale lies on the fulfilment of their need to permanent update skills and knowledge. The reason for the selection of a survey of 61 academic staff from Information Technology and Computer Science departments from four Universities from the Kwazulu-Natal and Eastern Cape provinces of South Africa was explained in the research methodology. Towards the end of the chapter, the choice of these four universities was identified as a limitation for this research as well as its perceptions analysis nature. The next chapter defines concepts that are used in the research.

CHAPTER TWO

Concepts and Definitions

2.1 Introduction

This chapter presents definitions and concepts on the types of web search engines, their history, their structure, and their use. The chapter starts by defining search engines before presenting a brief overview of their history. The chapter also presents the major types of common search engines, their architecture, as well as their functioning. Because searching activities are different depending on the nature of the search engine at hand, the chapter also highlights major differences between semantic and syntactic search. Furthermore, a description of how most search engines process and retrieve information is given before the presentation of search engines applications.

2.2 Search Engines

Existing literature on web information retrieval defines and classifies search engines based on their functionality and nature of their search results. Tümer et al. (2009) defines a search engine as a program that returns a list of web documents containing users' specified keywords. More often, a search engine is also described as a web site on its own that collects and organizes web content from all over the internet. Search engines continue to attract a large number of web searchers and to consistently be ranked as some of the heavily visited sites worldwide in terms of number of visitors (Netrating, 2002). Existing literature reports that more than 80% of Web visitors use a search engine as a starting point for their internet sessions (Kehoe et al., 1996; Sullivan 2003). Google, Excite, Lycos, AltaVista, Infoseek, Msn, and Yahoo are all examples of popular syntactic search engines. In a study by Tümer et al. (2009), it was found that Google, Yahoo and Msn together make up 93.8 percent of the volume of search ratios in the United States of America (USA) with 71.9%, 17.7% and 4.2% respectively for Google, Yahoo and Msn.

The obvious reason highlighted by users to choose syntactic search engines is that only a keyword is needed to perform a search. But the use of a keyword does not always guarantees a hit because of the complexity and contextual definition of natural languages terms. According to Budzik et al. (2000) and Lawrence (2000), results given by a query are usually identical and independent of the context of the user's request. Glover et al. (2000) re-iterate that this is a huge problem especially because web pages are becoming dynamic, can be automatically generated in real-time (e.g., current stock prices), and may contain multimedia (e.g., sound, video). Tremendous research efforts are being made (Lawrence, 2000; Budzik et al., 2000; Cohen et al., 2000; Fensel et al. 2001; Bergman 2001; Spink, 2003, Hwang et al., 2008) in an attempt to address these contextual and dynamic issues.

2.3 History of Information Retrieval and of Search Engines

The history of information search could be traced as far as before the invention of paper when ancient Romans and Greeks were recording information on papyrus rolls. Some papyrus artifacts from ancient Rome had tags attached to the rolls (Langville et al., 2006). These tags were an ancient form of today's Post-it Note. Langville et al., (2006) also highlight that notable artifacts belonging to information retrieval history are mostly individual library holdings sorted by title, as well as examples of the Dewey decimal system dated as far as 1872, the card catalog in the early 1900s, microfilm in the 1930s, and the MAchine Readable Cataloging (MARC) system of 1960s. In 1989 the storage, access, and searching of document collections was revolutionized by the invention of the World Wide Web by Tim Berners-Lee. Thereafter, the World Wide Web became the ultimate signal of the dominance of the Information Age.

In the earliest days of the Web, there were relatively few documents and web sites. The web was a manageable task to post all documents as static pages. Because all pages were persistent and constantly available, they could be crawled easily by early search engines. In 1994, one of the first web search engines, the World Wide Web Worm (WWWW) (McBryan, 1994) had an index of over hundred thousand of web pages and accessible documents. WWWW was subsequently followed by several other academic search engines, many of which are now public companies. Later in the same year, the *Lycos* search engine went public with a catalogue of 54,000 documents and since then, the compound growth rate in online documents has been on the order of more than 200% annually (Bergman, 2001). Langville et al., (2006) reports that, according to the InformationWeek magazine, apart from checking e-mails, searching information using search engines was the second most popular Internet activity in the early 2000's. In the late 2000's, more than 1,500 different search engines were in existence, but not many are well known or used, except for a few leading top 10 web search engines. As a result, search engines technology had to scale dramatically to keep up with web documents growth.

Due to the rapid growth of the amount of information on the web, as well as the increasing huge number of new end users, enormous challenges arise for the design of effective and efficient search engines. Bergman (2001) highlighted that one of the challenges of the WWW is that most of the Web's information is buried far down in stacks of dynamically generated sites, and most standard search engines almost never find it in the "deep web" because these search engines create their indices by crawling the Web surface, to only discover static pages linked to other pages.

Different search engines have different interfaces, they interpret queries in different ways (Beitzel et al., 2007), and they support different types of advanced search functionalities, and employ different search algorithms (Chu et al., 2005).

2.4. Overview of Traditional Information Retrieval and its basic models

It is important to differentiate between web information retrieval and traditional information retrieval. Web information retrieval is done within the world's largest and linked document collection, whereas traditional information retrieval is done in smaller, more controlled and non-linked collections. Traditional non-linked information retrieval existed even before the birth of the WWW in 1989 and it is still done today mainly for local libraries collections and for intranet documents within an organisation. These documents are not really linked but they are usually organised based on a structure or hierarchy defined by domain experts and they usually stored in books, library documents, and even in web pages. However, the mechanisms for searching for items in these non-linked collections are nowadays almost completely computerized and are referred to as search engines, or virtual machines created by software that enables them to sort through virtual file folders to find relevant documents (Langville et al., 2006). There are three basic computer-aided techniques for searching traditional information retrieval collections: Boolean models, vector space models, and probabilistic models.

2.4.1 Boolean models

The Boolean model of information retrieval, one of the earliest and simplest retrieval methods, uses the notion of exact matching to match documents to a user query. This model of information retrieval operates by considering which keywords are present or absent in a document using Boolean algebra, whereby words are logically combined with AND, OR, and NOT Boolean operators. Thus, a document is classified either as relevant or as irrelevant; there is no concept of a partial match between documents and queries. The drawback of this model is that it cannot return documents whose keywords were not included in the original query but that are semantically related depending for example on the use of synonymy and polysemy. For example, a title search for "Teaching AND experience" on a Boolean engine causes the virtual machine to return all documents that use both words in the title. A relevant document entitled "Lecturing experience" will not be returned because the synonyms teaching and lecturing are not syntactically equivalent.

2.4.2 Vector Space Models

Vector space models transform textual data into numeric vectors and matrices, then employ matrix analysis techniques to discover key features and connections in the document collection. Some advanced vector space models address the common text analysis problems of synonymy and polysemy explained above. The vector space model allows documents to partially match a query by assigning each document a number between 0 and 1, which can be interpreted as the likelihood of relevance to the query. The group of retrieved documents can then be sorted by degree of relevancy. Relevant feedback is therefore an application and advantage of the vector space model. It is also an information retrieval tuning technique as it allows the user to select a subset of the retrieved documents that might be useful in that context. The query is then resubmitted with this additional relevance feedback information, and a revised set of generally more useful documents is retrieved. A drawback of the vector space model is its computational greediness.

2.4.3 Probabilistic Models

This model attempts to estimate the probability that a particular document will be found by a given user. Retrieved documents are ranked by their relevance ratio. The probabilistic model operates recursively and requires the underlying algorithm to perform an initial guess of documents' relevancy based on initial parameters, then to iteratively try to improve this initial guess in order to obtain a final ranking of relevancy probabilities. Unfortunately, probabilistic models can be very hard to build and implement due to the complexity growth and limited scalability of their algorithms. Probabilistic models also require several unrealistic and simplifying assumptions, such as independence between terms as well as documents. For instance, in this dissertation, the most likely word to follow "*searching*" is the word "*ability*", but the independence assumption judges other words such as "*experience", "method"*, or "*books"* to be equally likely to follow the word "*searching*".

These above three searching models grew over decades into new search models resulting in about 3,500 different search engines models as of June 2000 which are all rooted from one, or combinations of these three models (Langville et al., 2006). Modern web information retrieval machine also called web search engines are based on these basic models but due to the manner in which information is retrieved, they are put in another category. Search engines whose structure includes at least one of each of these models are called Meta search engines. This is based on the principle that while one search engine is good, two or more are even better. One search engine may be great at a certain task, while a second search engine is better at another task. Thus, meta-search engines were created to simultaneously exploit the best features of many individual search engines. Meta-search engines send queries to several individual search engines at once and return the results from all of the search engines in one long unified list.

2.5 Type of Search Engines

Search engines are classified in three major categories, spider-based, directory-based, and link-based, based on the way information is retrieved. It is also important to note that other schools of thought group them according to topics or content, business models, and information types.

2.5.1 Spider-based Search Engines

Also called crawler search engines, spider-based search engines are the most commonly used search engines on the Internet. These search engines use a form of software program called spiders or crawlers to find information on the Internet and to store it in giant databases or indexes for users' searching sessions. Some spiders record every single Web site words for their respective indexes, while others only report certain keywords listed in title tags or meta tags. Each website needs to have an identification keyword to uniquely describe its content. These keywords as well as their placement either within actual web sites content or in meta tags, are very important to the effectiveness of the spider. The major drawback of this category of search engine is that website masters need to continuously monitor the results ranking of their own web sites for different search engines. This is due to the fact that spider based search engines often change their criteria for listing (ranking algorithm) different sites, and keywords that cause a site to be listed first in a search on a specific day may not work at all on the next day. As a result, incorrect or misleading information listed in meta tags causes spiders to deliver inaccurate descriptions of web sites to indexes. From the end user point of view, this is what causes irrelevant results when performing web searching. Google for example belongs to this category of search engines. Unlike Google, other search engines do not use spider software.

2.5.2 Directory-based search engines

While some sites use spiders to provide results to searchers, others (for example Yahoo!) use human editors. This means web masters do not rely on technology and keywords to obtain excellent placements, but they must provide content classification (listing) that can be found appealing and valuable by search engines editors. Some of the limitations for this category are that some of the directory-based engines charge a fee for a site to be reviewed for potential listing (Hearst, 2000). For examples, LookSmart, Lycos, AltaVista, MSN, Excite, and AOL search engines rely on providers of directory data to make their search results more meaningful. Other search engines engineers feel that keywords, directories and money cannot be the best way to retrieve information for users.

2.5.3 Link-based Search engines

This category of search engines provides results based on hypertext links between sites. Rather than basing results on keywords (like spider-based engines) or on the preferences of human editors (like directory-based engines), sites are ranked based on the quality and quantity of other web sites linked to them. In this case, links serve as referrals between websites, and as a consequence, they increase websites chances of being listed by search engines. The emergence of this kind of search engines called for web masters to develop link-building strategies. These strategies consist of first finding out which sites are listed by search engines for a given product category, in a link-based engine; a company could then contact these sites' owners and ask them for a link. This often involves reciprocal linking, where each company agrees to include links to the other site. The major drawback of this type of search engine is that individual sites require a referral link agreement among each other. For example, companies that are competitors in e-commerce are forced to refer to each other for their website to be retrieved by users.

2.5.4 Other Classifications

In addition to these three major categories (spider-based, directory based, and linked-based), other classifications are based on topic, model, and information type. The next subsections describe these classifications (also see table 2.1).

2.5.4.1 By Topic

These are domain specifics search engines specialized in a particular topic such as education, business, accountancy, enterprise, geographical limited scope, mobile, job, legal, medical, news, people, real property, television, games, etc.

2.5.4.2 By Business Model

These search engines are based on specific business models guiding their crawling policies. Example of search engine business models include: Open source, Semantic browsing, social network, Meta search, and visual search.

2.5.4.3 By Information Type

In this type of engines, search engines are dedicated to a specific kind of information. For example, search engines can be classified as: forums, blogs, multimedia, emails, maps, price, question and answers (Q&A), natural language, etc.

2.5.4.4 Common Search Engines

Search engines that are mostly considered as popular to end users are referred to as common search engines. These search engines are classified on the basis of two ratings used to differentiate various search techniques: precision and recall. Precision is the ratio of the number of relevant documents retrieved to the total number of documents retrieved, and recall measures the ratio of the number of relevant documents retrieved to the total number of relevant documents in the collection (Langville et al., 2006). The higher the precision and recall, the better the search engine. Recall and precision are information retrieval-specific performance measures, but, of course, when evaluating any computer system, time and space are always considered as performance issues in terms of memory-efficiency and processing speed.

Table 2.1 Categories of search engines

Category	Subcategory	Examples
By topic	**General**	Ask.com, Bing, Google, Yahoo! Search
	Business	Business.com, GlobalSpec, Nexis, Thomasnet
	Enterprise	AskMeNow, Concept Searching Limited Dieselpoint dtSearch, dtSearch Web
	Ethnic	RushmoreDrive
	Mobile/ Handhel	Taganode Local Search Engine, Taptu: taptu mobile/social search
	Job	Bixee.com, CareerBuilder.com, Dice.com, Eluta.ca, Hotjobs.com
	Legal	WestLaw, Lexis, Quicklaw, Manupatra
	Medical	Bing Health, Bioinformatic, Harvester, Entrez
	News	Bing News, Google News, Daylife, Topix.net Yahoo! News
	Television	TV Genius
	Video Games	Wazap
	People	PeekYou, Ex.plode.us, InfoSpace, Wink
	Real property	Home.co.uk, HotPads.com, Rightmove, Zillow.com
By Information type	**Forum**	Omgili
	Blog	Amatomu, Bloglines, BlogScope, IceRocket
	Multimedia	Bing Videos, FindSounds, Google Video, Munax's PlayAudioVideo
	Source code	Google Code Search, JExamples, Koders, Krugle
	Maps	Wiki Mapia, Bing Maps, Géoportail, Google Maps
By model	**Open source search engines**	DataparkSearch, Egothor, Grub
	Semantic browsing engines	Evri, Hakia, Yebol
	Metasearch engines	Brainboost, ChunkIt!, Clusty, Excite

Common search engines are Google.com, Yahoo.com, and Bing.com. It is important to note that these search engines have local servers in most regions of the world to localize the processing of queries and also to take into consideration languages and communication challenges. These common engines are mostly commercial and they use syntactic algorithms to crawl the web.

Many web search engines are built on the techniques of traditional search engines although they differ in many important ways. The Web is huge and it is so big that it is hard to get an accurate count of its size. A company called BrightPlanet even sells access to the so-called “Deep Web”, which they estimate to contain over 92,000TB of data spread over 550 billion pages (Bergman, 2001). The Deep Web contains hundreds of thousands of publicly accessible databases that create a collection over 500 times larger than the Surface Web. These Deep Web pages cannot be found by casual, routine surfing. The existence of the Deep Web raises the question to know which type of search processes and algorithms BrightPlanet software uses to access these databases? And why are they not accessible by common search engines?

2.6 Semantic Search versus Syntactic Search

Syntactic search is the process used by search engines to map specific terms (keywords) within web sites content in order to retrieve a particular site as a result index but semantic search seeks to improve search accuracy by understanding searcher intent and the contextual meaning of terms as they appear in the searchable data space, whether on the web or within a closed system, in order to generate more relevant results. Thus a semantic search engine uses semantic search algorithms while a syntactic search engine (also called conventional search engine) simply maps web content to keywords.

The difference between semantic search engines and conventional search engines is that semantic search engines are “meaning-based”. The meaning of a word is usually described in terms of synonyms, antonyms, homonyms, and so one. A semantic search engine could be for example capable to see that words like “Head” or “Chair” are closely related in the context of an academic department but the word “Chair” has an extra meaning in the context of interior architecture. But a conventional keyword search engine will have difficulties seeing how these two words are deemed to be related. Usually, in a semantic search engine, there is a glossary that identifies subjects associated with specific terms, in order to enable indexing content based on the meaning of terms rather than the surface forms in the text as done with syntactic search. Semantic search algorithms are being developed through the formation of the semantic web vision.

2.6.1 Ontology and Semantic Web Vision

Tremendous efforts in WWW research have seen the definition of the Semantic Web as the vision of the next generation of information highways. The backbone of the semantic web is ontology construction. Berners-Lee, the founder of the WWW already had this vision in 1996 that there is a need for data to be available on the web in a machine-readable form with defined semantics (Berners-Lee, 1996). The aim of the Semantic Web is to represent Web content in a form that is more easily machine-processable and to use intelligent search algorithms taking advantage of these representations of the information. Information on the web is well presented, but what lacks is information about content. For machines to be more data-processable, the meaning of data has to be added to the content and formatting of information (Tolksdorf et al., 2005). It is useful to remember the definition of the term ontology as: an explicit specification of a conceptualization (Gruber, 1993). Ontology presents a formal description of concepts and their relationships in some area of interest. Therefore, ontology is a terminology that provides a shared understanding or a finite vocabulary of terms in a specific domain communicated across people and application systems.

2.7 Architecture of Search Engines

Software architecture consists of software components, the interfaces provided by those components, and the relationships between them. The Architecture of a search engine is usually defined by two requirements: effectiveness (quality of results) and efficiency (response time and throughput). The next two sections present the basic components of a search engine and their interface.

2.7.1 Components and Processes

The majority of web search engines are based on a centralized architecture that relies on a set of three key components: the Crawler, the Indexer and the Searcher. Some systems also include a local store copy of the crawler's data. These key components describe the reference architecture that could already be seen in early search engines like the

WWWW and that also exists in most recent ones like Google (See Figure 2.1) (Risvik et al., 2002).

2.7.1.1 The Crawler Module

Web documents collections are stored in a cyber warehouse, a virtual entity that is not limited by geographical constraints and can grow without limit. However, this geographic freedom brings one unfortunate challenge: web search engines must collect information and categorization tasks on their own (Langville et al., 2006). As a result, all web search engines have a crawler module. The crawler is a module aggregating data from the World Wide Web in order to make them searchable. Several heuristics and algorithms exists for crawling, most of them are based upon following web page links (Risvik et al., 2002). This module contains software that collects and categorizes web documents. The crawling software creates virtual robots, called spiders, that constantly scour the Web, gathering new information and web pages, and storing returning documents in a central repository.

2.7.1.2 The Indexer

The indexing module takes each new uncompressed page and extracts only its vital descriptors, creating a compressed description of the page that is stored in various indexes. This module is like a black box that takes the uncompressed page as input and outputs a "Cliffnotes" version of the page. The uncompressed page is then tossed out or, if deemed popular, returned to the page repository. Common practices for the implementation of compressed web pages include inverted files, vector spaces, suffix structures, and sometimes hybrids of these (Risvik et al. 2002).

2.7.1.3 The Searcher

The searcher is working on the output files from the indexer. It accepts user queries expressed in natural languages, converts them into a language that the search system can understand, runs them over the index, and returns computed search results to issuer. A local store copy is a snapshot of the Web at the given crawling time for each document.

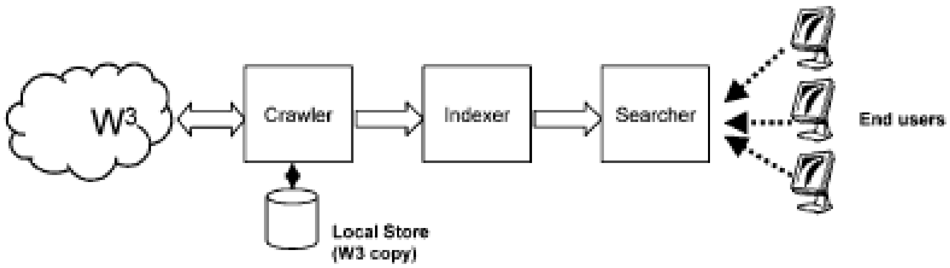


Figure 2.1 Reference search engine model (Risvik et al. 2002).

Most search engines usually run the crawler, indexer, and searcher sequentially in cycles. First the crawler retrieves the content, then the indexer generates the searchable index, and finally, the searcher provides functionality for searching the indexed data. The indexing cycle is run again in order to refresh the search engine. Despite the commonly public use of web search engines, their internal architecture details remain as a black box for most users as their internal design remains hidden in the domain of specialized professionals.

2.7.2 User Interface

Search engines user interface design involves the consideration of many factors including users' age and their experience. This interface is usually based upon simple graphical interface principles due to the consideration of novice end users. Basic widgets and graphics like a text box and button are usually used to start a search session. More complex search engines interfaces may include images and links to other functionalities and features of the engine.

2.8 Use and applications of Search Engines

The web is a dynamic environment with thousands of different search engines. In addition, should the Semantic Web become a reality in spite of its social issues still to be solved; the Web will become XML-based, with appropriate standards for semantic metadata and schemas (Lawrence, 2000). The use of search engines consists of innovative techniques, concepts and applications that enhance web information retrieval. In this regard, the concept of soft computing (SC) was introduced by Baeza-Yates (1990) as a synergy of methodologies which collectively provide a foundation for the conception, design and construction and utilization of information systems and of intelligent systems. Some of the main methodologies of SC are fuzzy logic, genetic algorithms, neural networks, rough sets, Bayesian networks, and other probabilistic techniques (Baeza-Yates 2003).

2.9 Conclusion

The chapter highlighted the definitions and brief history of search engines which reveals that a succession of web search engines deployed between 1994 and today: World Wide Web Worm, *Lycos*, Excite, Google, AltaVista, Infoseek, Msn, Yahoo, etc. It further defined the role and purpose of search engines which is a software helping users to access web documents by formulating a query. The chapter also gave different classifications and examples of common search engines. The chapter also gave different school of thoughts about the taxonomy of web search engines in terms of the types of engines including spider based search engines, directory based search engines, and linked based search engines; and in terms of the classification of these engines by topic, by model, and by information type. Toward the end of the chapter, a more technical report described contrasts between semantic search algorithms and syntactic search algorithms by explaining how semantic takes advantage of the construction of ontology of knowledge that can be understood by intelligent software based on the description of relationships existing between objects meanings. The chapter ended with a brief description of few applications developed to improve the web searching process such as the concepts of concept of soft computing. The following chapter will look at existing literature on factors that influence search ability.

CHAPTER THREE

Literature Review

3.1 Introduction

This chapter reports on the findings of work reviewed by the researcher on the factors affecting the searching ability of internet users. The chapter is subdivided into five (5) sections according to the major themes of existing research on the searching ability of internet users. These research themes include: online searching success, their strategies, searching mechanism, users understanding of the structure of web search engines, temporal searching behavioral trends, and the adoption of domain specific search engines. The chapter also describes the methodologies used by these research studies.

3.2 Search Success

Existing literature considers search success as the mostly used information retrieval (IR) metric. In a study by Nachmias et al. (2001) in which fifty four (54) masters students were asked to accomplish three (relatively simple) search tasks, seven different metrics were used to measure search success: the ability to use Internet tools, knowledge of search techniques, cognitive capacity, the ability to execute search plans, the ability to apply Boolean logic rules, the understanding how information is organized, and knowledge of Internet notations. Results from Nachmias et al. (2001) reveal that search success rates are low, with only 15% of the students succeed in all tasks.

Factors used for the analysis of web search success include: domain familiarity (Bates 2007, Kules et al., 2008, Bilal, 2000, and Hersh et al., 2000), users' knowledge or cognitive abilities (Lee, 1999), users' academic level (Allen 1992), users' web experience (Lazonder, 2000, Stronge et al., 2006, Lee 1999, and Bilal, 2000), users' attitude towards the use of computers (Gugerty et al. 2006 and Jackson, 2001), their gender

(Jackson, 2001) and age (Stronge et al., 2006, Lee, 1999, and Bilal, 2000), their reading (Bilal. 2000) and writing ability (Allen, 1992), their motivation level, their personality (Hersh et al., 2000).

3.2.1 Familiarity of domain

In a study by Bilal (2000), twenty two (22) grade seven learners were asked to search for information using Yahooligans. Yahooligans is a search engine and directory designed for children aged between 7 and 12. The purpose of Bilal (2000) was to examine children's cognitive, physical, and affective behaviour during a fact-based information search task. The research findings from Bilal (2000) reveal that children's domain knowledge does not significantly influence their web search success.

Hersh et al. (2000) also studied users' domain familiarity in the field of medicine. In their study, they used twenty-nine (29) students who completed a questionnaire design based on medical and nurse-practitioner theories. In addition, the participants' computer experiences as well as their cognitive abilities and personality type were measured. Results from Hersh et al. (2000) show that medical students are able to answer more questions correctly than nurse practitioner students before and after searching for information online, meaning that medical students are more knowledgeable in medicine than nurses, irrespective of their use of Internet search.

In a more recent study, Bates (2007) uses twenty (20) university students from various subjects' majors to measure subject and catalogue familiarity when searching for information online. Bates (2007) shows that catalog familiarity has a very significant beneficial effect on search success, and subject familiarity has a slight, but not significant, detrimental effect.

3.2.2 User knowledge

Lee (1999) in his review paper on the usability and accessibility of web search engines identifies the user knowledge as a factor influencing the search success of online users. Results from Lee (1999) indicate that most Internet search sessions are spent on the reading of web content by users and their prior knowledge and web experience are key factors for internet search success. Lee (1999) also associated user's knowledge to other system variables such as internet transmission speed, visual display of characters, and device capabilities as factors affecting web usability. Similarly, Kules et al. (2008) used twenty four (24) sophisticated (experts in their field of work) users to carry out complex tasks to measure the adaptation of search tactics for categorized overviews. Kules et al. (2008) research shows that the more organized the subjects, the deeper it was explored by experts, and that, categorized overviews improves search results. In addition, the study by Kules et al. (2008) found that users experience mild frustration, confusion or doubt when page categories do not match their expectations, and most participants found additional categories of information and the ability to preview and narrow results very beneficial.

3.2.3 User's Cognitive Ability

This section examines the role of cognitive ability in the information search process with regards to users' academic level, their web experience, their reading and writing ability, and their logical, spatial and verbal abilities. According to Kim et al. (2004) and Allen (1992) cognitive abilities defined as the "factors that contribute to intelligence", influence search performance in a variety of information systems.

3.2.3.1 Academic Level

In Allen (1992), an experiment was conducted to examine the relationship between users' cognitive abilities and search tactics. Fifty (50) university students were subjected to eight (8) tests from the Kit of Factor-Referenced Cognitive Tests. Participants were also required to read a print document and to retrieve its online soft copy later. The Factor-Referenced Cognitive Tests mostly measured verbal reasoning, logical reasoning, perceptual speed, spatial scanning, and academic levels. Results from Allen (1992) show that the ability to identify potentially useful citations is associated with the academic levels (senior versus junior students) of the participants. They indicate that the kind of citations chosen by senior students differed from the citations of junior students, and most useful citations are retrieved by senior students, proving that academic level influences search tactics.

3.2.3.2 User's Web Experience

Surprisingly, existing literature does not agree on the definition of the concept of web experience. For Lazonder et al. (1999), users having worked with the Web for less than ten (10) hours are considered as novice users, but that threshold is extended by Palmquist et al. (2000) to two (2) years of online search experience, in the use of online databases in general. Nevertheless, Saito et al. (2001) gives evidence that experience affects web information seeking behaviour as significant differences exist between expert and novice users in relation to solution time, the number of pages searched and the types of pages accessed. This is also supported by Lazonder et al. (2000) whose research indicates that learners with web experience are more proficient in locating web sites than are novice web users. The research participants from Lazonder et al. (2000) were twenty five (25) grade 12 learners (high school final year learners) from two (2) schools. These participants were categorized either as novice users or as expert users based on their web experience. They filled a questionnaire and performed an experimental task on the use of search engines. Observed differences in searching abilities were attributed to the superior skills displayed by expert users in operating web search engines. However, on tasks that required learners to locate information on specific web sites, the performance of experienced and novice users was equivalent.

A study by Bilal (2000) on the use of the Yahooligans by grade seven learners (already described earlier in this dissertation), found that successful children have more experience than unsuccessful ones. Similarly, according to Hersh et al. (2000) (also already presented earlier in the dissertation), the ability to successfully answer medical cognitive questions is associated with having experience in literature searching.

3.1.3.3 Reading Ability, Verbal and Logical Reasoning, Perceptual Speed and Spatial Scanning

Research results by Bilal (2000) indicate that reading ability does not significantly influence search success, but Allen (1992) found that users with lower levels of verbal comprehension seem to have more difficulties in identifying search keywords and they tend to use fewer search expressions. Allen (1992) also found that individuals with high levels of logical reasoning select fewer citations as being potentially useful for internet search. The same research also claims that perceptual speed has an effect on the quality of Internet search as users with low levels of perceptual speed ability find it difficult to cope with many screens of information.

Another study by Teitelbaum-Kronish (1984) emphasizes the importance of logical reasoning in assisting users in the retrieval of better information. This is in line with the results from Allen (1992) showing that logical reasoning influences search tactics. Similarly, Gugerty et al. (2006) supports that verbal and spatial abilities influence users' internet search performance.

3.2.4 Personal Profile

Users' information needs have been recognised as an influential factor in the information seeking process specifically in relation to the effort and time users tend to invest in online searching (Jackson et al., 2001), their attitude towards the use of computers (Gugerty et al. 2006), and their motivation. Existing research also indicates that demographics such as gender and age equally influence the overall success rates in finding information online (Bilal, 2002; Jackson, 2001).

3.2.4.1 User's attitude toward computers

In a study by Gugerty et al. (2006), one hundred and eighty (180) participants were requested to perform a series of internet search tasks. A finding from the analysis of this experience revealed that attitude towards internet has a direct effect on search performance. This finding is supported by Lee (1999) who showed that motivational factors are likely to affect web usability, but it is not supported by Hersh et al. (2000) where it is indicated that there are no personality factors nor attitude associated with improved search success for the use of information retrieval systems by medical students. Moreover, in a study of computer attitude differences between genders by Jackson et al. (2001), it was found that females report more computer anxiety, less computer self-efficacy, and less favorable and less stereotypic computer attitudes than males. These findings were the results of the analysis of data from a questionnaire based survey of six hundred and thirty (630) Anglo American undergraduate students (403 males and 227 females) on their computer anxiety levels.

3.2.4.2 Users' gender

Jackson et al. (2001), while studying users' attitudes, was also interested in motivational factors (e.g. loneliness), cognitive factors (e.g. computer self-efficacy), and gender. The examination of gender in e-mail and web use revealed that females use e-mails more than males, but males use the web more than females. A similar result is presented in a case study by Large et al. (2002) on collaborative web search strategies using sixth grade students organized into same gender groups. These children were requested to search for information on the web to support a classroom assignment and they were videotaped for subsequent analysis where it was found that groups of males are more active on the web than female counterparts: meaning that males' level of interaction with the web is higher than that of their female counterparts.

3.2.4.3 Age and disability

According to Lee (1999), user's characteristics likely to affect web usability in general and searching ability in particular, include among others age and disability limitations in memory and vision. A similar conclusion was reached by Stronge et al. (2006) based on a study of thirty two (32) users divided into two (2) equal size of young and old users. Findings from Stronge et al. (2006) show that older adults are less successful than young adults in finding correct answers to a given search task, and younger adults use more keyword search strategies that older ones. On the contrary, there is no age related difference on advanced search strategies although younger adults use more advanced keyword search strategies than older ones (Stronge et al. 2006).

3.3 Search Strategy

Existing literature reports that web end users plan and execute a type of strategy in their information seeking activity. These strategies are of different types including and are not restricted to the keyword strategy, query reformulation and feedback, and collaboration.

Users' searching abilities also depend on their search strategies with regards to keyword selection (Shneiderman, 1997; Large et al., 2002; Bilal, 2000; Spink ,2003; Fattahi et al., 2008; Wolfram, 2008; Jansen et al., 2001), query reformulation (Anick, 2003; Jansen et al., 2001), and use of help and feedback (Anick, 2003; Shneiderman, 1997).

3.3.1 Keyword Strategy

Keywords are by far the most used strategy by searchers while finding information online. Even children use keywords when finding information on Yahooligan (Bilal, 2000). Keyword search strategies are mostly used on search engines. Existing research on keywords strategies usually examine keywords hit rates, keyword trends, and the number of terms used in search queries.

3.3.1.1 Hit Rate

In a study by Shneiderman (1997), it was found that search engines record zero-hit rates for up to thirty percent (30%) of web searches, clearly highlighting the essential usability problem posed by keywords search interface in the formulation web queries. However, Nachmias et al. (2001) found no significant differences between the strategies in successful tasks and those used in unsuccessful search tasks, except for the final search strategies. In the same study, a relatively larger percentage of failures were also found among students who used search engine as final strategy as opposed to students that used browsing final strategy. In addition, in a study by Stronge et al. (2006), it was found that web experience does not always depend on user's age but it leads to more effective search strategies.

3.3.1.2 Keyword trends

A survey of user's search trends and patterns by Spink (2003) revealed three major findings: (i) a small number of terms are used with high frequency and many terms are used once, (ii) users in general do not really understand the functioning and structure of web search engines and of the Web, (iii) the Web culture is based on a "quick and dirty" approach to searching, rather than on an exploratory and interactive approach.

Fattahi et al. (2008) also did a research on small query expansion in search engines using twenty (20) queries on health and social science related topics. The queries were made up of either Non-Topical Term (NTT) or Topical Term (TT) or Semi-Topical Term (STT). It was found that: (i) Web searching could be greatly enhanced by combining Non-Topical Term (NTT) and Semi- Topical Term (STTs) with TTs in an initial query, (ii) Search results would improve if queries are restricted to the exact title or URL search options, and (iii) the development and implementation of knowledge-based lists of NTTs (and STTs) by both general and specialized search engines to aid query expansion.

3.3.1.3 Number of queries keywords terms

A research by Wolfram (2008) to measure search characteristics with regards to term usage, query formulation, and session duration recorded transaction logs from four (4) different Web-based information retrieval environments (a bibliographic databank, OPAC, a search engine, and a specialized search system). It was found that: (i) the high usage of articles, prepositions, and conjunctions (e.g., "AND" as a Boolean operator or otherwise) is common across all datasets, (ii) in a bibliographic databank, queries are short and are of academic interest. Sessions are short and brief in duration, indicating a quick assessment of query results, (iii) in OPAC, queries are longer and reflect more academic interests. Sessions are longer with longer inter-query times, indicating more review and evaluation of retrieved content, (iv) in search engine: queries are of mid-length and of general or popular interest. Sessions are short but reflect the longest time period of review between queries, (v) in specialized search service, queries are short and focused. Sessions contain a larger number of queries but reflect very short durations, indicating quick assessment of the returned results for each query.

Similarly, in a study of transaction logs from 51,473 Web user queries from 18,113 users on *Excite*, Jansen et al. (2000) found that: (i) most users do not perform many queries per search session (mean number of queries per user is 2.8), (ii) web queries are short averaging approximately 2 terms per query, and less than 4% of the queries are comprised of more than 6 terms, (iii) most users search one query only and they do not follow it with other successive queries, (iv) relevant feedback is rarely used, and (v) the

`+' and `-' modifiers that specify the mandatory presence or absence of a term are used more than Boolean operators, and Boolean operators are seldom used.

A literature review by Jansen et al. (2001) on Web search session durations and query terms concluded that (i) search analysis depends on the type of queries or sessions chosen by researchers, (ii) query level analysis focuses on query length, query complexity, and failure rate, and (iii) query terms depend on specific information retrieval systems that also define the use of possible query delimiters.

Since the first query terms chosen by users do not always lead to a success hit, a query reformulation is usually necessary. The next paragraph reviews research papers describing changes to initial queries.

3.3.2 Query Reformulation, Feedback and Use of Help

Query reformulation refers to changes made by users from their initial queries. Reformulation might include adding or subtracting keyword or Boolean terms, or combining more than one of these possibilities. Query reformulation may be influenced by feedback received by web users. Feedback usually comes in the form of a message written by the search engine software to the user and it may include suggested terms for a particular query.

In a study by Anick (2003) on the behavior of web searchers with regards to query reformulation, data from log sessions for two groups of users interacting with variants of the AltaVista search engine revealed that on average, about 56% of search sessions involve some degree of refinement. Within this subset of sessions, it was found that uptake of feedback increases to 25%. The same research showed that users achieve the same level of search success regardless of whether feedback is offered or not. However, in Anick (2003), the feedback group was slightly more likely to do at least one refinement within a session and less likely to click on a result immediately after the initial query. The research also observed that 28% of feedback refinements occur directly after the initial query, 21% occur after a result click, and users prefer to choose feedback phrases containing query terms. Jansen et al. (2001) also found that for the

Excite search engine, most users usually search for one query only and do not usually follow it up with any other queries reformulation, and relevant feedback given to the user is rarely re-used in subsequent queries.

3.4 Web Search Engines Output formats

Existing research on web search engines formats indicate that the presentation format of results from a search engine heavily influences users' selection of search results. More recently Höchstötter et al. (2009) examined search engine results pages by analyzing five hundred (500) queries for four major search engines (*Google.com, Yahoo.com, Live.com, and Ask.com*). It was found that search engines use different approaches to results pages composition, therefore, users get to see quite different results sets depending on the search engine and search query used. Even more surprising is that some search engines show Wikipedia results quite often, except from Google.com and Yahoo.com that prefer results from their own web pages.

Cumbersome search results lists generated by traditional search engines are a well recognized problem in information retrieval and providing the user with a means of viewing groups of similar search results potentially enhances Web search effectiveness. However, there has been little research into Web searchers' interactions with clustered search engine results except from a study by Koshma et al. (2006) where users' web sessions were analyzed with regards to their interactions with clustered web pages and their *Vivisimo* web search engine. Transaction logs were analyzed over a period of two (2) weeks and from the analysis of these transaction logs it was found that almost half of user interactions with clusters consisted of simply displaying a cluster's result set, and only a small percentage of interactions showed cluster tree expansion.

3.5 Multitasking, Multitopics, and Searching Approaches

As already mentioned earlier in this chapter, according to Spink (2003), current approaches to web searching are far from being neither exploratory nor interactive; instead, they are quick and fix exercises whose mechanism is not well defined. Hersh et

al. (2003) also attempted to identify search mechanics factors associated with successful use of information retrieval system. Unfortunately, their data did not uncover factors that could be used to guide improvements in search mechanisms or would lead to further benefits for users except from improving their knowledge of their domain.
Existing literature on searching approaches and mechanisms also indicate that multitasking and multitopics are search strategies that are used by advanced web user to enhance and speed up their search (Spink et al., 2006; Koshma et al., 2006).

Koshma et al. (2006) while analyzing the application of clustering to the *Vivisimo* web search engine transaction logs realized that about eleven percent (11%) of search sessions was multitasking searches, and there are a broad variety of search topics in multitasking search sessions. Similarly, Spink et al. (2006) did a qualitative analysis of two types of users' queries on *AltaVista* transaction logs from 2002 -2004 to measure the degree of multitasking search and information topic switching. Findings from Spink et al. (2006) reveal that (i) about eighty one percent (81%) of two-query sessions included multiple topics, (ii) about ninety one percent (91%) of three or more query sessions included multiple topics, (iii) there are a broad variety of topics in multitasking search sessions, (iv) three or more query sessions sometimes contained frequent topic changes. These findings show that multitasking and multitopics are found to be a growing approach in web searching.

3.6 Changes in Behavioural trends

Over the years, the web has grown into a vital channel of communication and it has become an important vehicle for information dissemination and retrieval. This therefore is "exerting power over the evolution and development of information-seeking behavior" (Nahl et al. 1998). Behavioural changes in information search processes can be analyzed in terms of session duration (Nahl et al., 1998, Jansen et al., 2005), query and terms used (Wang et al., 2003, Jansen et al., 2006), and the number of result pages viewed (Spink et al., 2001).

3.6.1 Session Duration and Frequency

Major web search engines such as *Google.com, Yahoo.com*, and *AltaVista* are essential tools in the quest to locate information online (Jansen et al., 2005). Session duration is defined as the average time spent by a user when searching for information online. Jansen et al. (2005) described observed patterns of user's session duration on AltaVista between 1998 and 2002 using the engine transaction logs. Their findings reveal that there was an increase in the percentage of users viewing more than the first results page, which when combined with other increased interactions, may indicate greater user persistence in locating relevant results. Recently, a study by Wolfram (2008) of search characteristics for different types of web-based information retrieval environments made also showed that sessions are short and brief in duration for both search engines and specialised search service libraries. Similarly, Koshma et al. (2006) while studying the application of clustering to web search engines, also found that a high percentage of search sessions contains one (1) query and lasts for less than one (1) minute in duration.

Term frequency is defined as the average term mostly used in a session. Spink et al. (2003) in their analysis of log of transaction records for over one million web queries by 200000 anonymous users of the *Excite* search engine found that close to half of the users look at two or less result pages only, and only a small number of search terms are used with high frequency, and a great many query terms are unique; concluding that the language of web queries is distinctive.

3.6.2 Query and Terms Lexicon

A lexicon usually refers to a dictionary or a list of words in a particular language. Jansen et al. (2005) found that the percentage of longer sessions increased over the years (1998-2002) with 32% of users submitting three or more queries per session. A study of web user queries trends by Wang et al. (2003) consisting of the analysis of 541,920 user queries submitted to and executed in an academic website during a four-year period (May 1997 to May 2001) using a relational database showed that: (i) most queries are short with an average of two words or 13 character positions, (ii) single word queries are likely to produce higher numbers of hits with lower precision, (iii) the size of the vocabulary (lexicon) grows as the number of queries increases, (iv) the increase in the number of words is much slower than the increase in the number of queries, (v) the percentage of misspelled words is as high as 26.0% of the total term used in the queries. Similarly Spink et al. (2003) research findings also show that: (i) most people use few search terms, few modified queries, and rarely use advanced search features, and (ii) some 48.4% of users submit a single query, 20.8% of them submit two queries and about 31% of users enter three or more unique queries. Jansen et al. (2006) also found that searchers on United State (US)-based web search engines use more query operators than searchers on European-based search engines.

3.6.3 Number of Result Pages Viewed

Spink et al. (2003) reveal that most people use few modified queries and view few web pages, and about one in every three modified queries has the same number of terms as the preceding query. Jansen et al. (2006) also studied the influence of the number of pages viewed on search success. Their findings reveal that users view fewer result pages and there are statistically significant differences between use of Boolean operators and result pages viewed. On the other hand, Spink et al. (2001), reports that less than 5% of all queries use Boolean operators (AND, OR, NOT) but there is a relationship between the use of Boolean operators and the number and accuracy of result pages viewed. The observations above clearly reveal that users do not view enough pages which is assumed to be an additional task in their quest to find the relevant information.

3.7 Adoption of Domain Specific Search Engines

This section reviews existing literature on users' preferences between domain specific digital libraries (e.g. *Google Scholar, ACM and IEEE* for academics, etc) and general search engines (e.g. *Google.com, Yahoo.com, Bing,* etc) especially for academics and for health practitioners.

3.7.1 Academics

In study by Griffiths (2005), university students were surveyed to measure their preference between general search engines or academic resources. The research found that: (i) students prefer to locate information or resources via a search engine and their use of academic resources is low; (ii) *Google.com* is the search engine of choice of most students. These findings clearly show that university students prefer search engines which tend to find general URL's documents unlike a dedicated academic digital library which is more likely to be specific to their studies and discipline.

3.7.2 Health Practitioners

Ash (1997) studied sixty seven (67) academic health sciences centres for a total of 1335 staff members to measure the diffusion and infusion of end user online literature searching using different factors such as communication, participative decision making, existence of champions or reward systems, planning, and top management support. It was found that: (i) communication and existence of champions influence both electronic mail (email) interaction with computer-based patient records and the infusion of online end user searching, (ii) top management support and existence of reward systems only affect the infusion of online end user searching, and (iii) participative decision making only influences electronic mail (email) interaction with computer-based patient records,

3.8 Conclusion

The chapter has reviewed previous studies on factors that influence users' searching ability. These factors were divided into topics search success, search strategy, the search mechanism, the users' basic understanding of web engines, the temporal trend in searching, and domain specific search engine. This chapter did not only report on findings on these factors but an equal emphasis was made on the description of the methodologies used in order to support findings. Each of the section's topics reported different methodologies included analysis of web servers' transaction logs, questionnaire based surveys (from various demographic user groups), experimental searching tasks, literature surveys, video recordings, just to name a few.

Major findings from existing literature indicated that searching hit rates are generally low, the number of terms used in a query average 2 to three, searching ability depends on a number factors that relate to the user's biographical attributes such as academic levels, gender, age, and disability, and also the users' levels of experience and knowledge of the web and domain. Literature also reported that searching ability is affected by cognitive factors such familiarity with the search domain, users' logical reasoning, and language related factors such as reading, communication abilities.

This chapter also reported that existing literature has no evident framework dictating novice web users' ways to search rather the web searching trends remain the traditional use of keyword as query terms, query refinement, although search sessions are quick and brief in duration and multitasking and multi-topics are practised by advanced users to optimise their search. Other challenges recorded in literature include users' neglect of web engines feedback and help for queries refinement, their choice of general search engine instead of the domain and discipline specific engines. The following chapter presents the detailed research methodology.

CHAPTER FOUR

Research Design

4.1. Introduction

This chapter provides a description of the research design used by this study. It describes methods and procedures used for the construction of the research sample, the collection of data, the verification of data reliability and validity, the analysis of data, and the overall research strategy used by this study. The purpose of the chapter is to fully describe how the research was conducted without showing the research results that are instead presented in the next chapter. Readers might feel overwhelmed by the level of details in the description of the data analysis steps by PASW Statistics 18.0 (SPSS) (for example see section 4.6.2.3 of this chapter), but this was done to allow other researchers to be able to repeat these steps with the hope of finding results similar to the ones of this current research. Most of the tables, diagrams and figures resulting from the conduct of the research are not presented in this chapter but in the appendices located at the end of the dissertation. This chapter also contains some elements of literature on research methodology in order to justify some of choices made by the researcher during the conduct of the research.

4.2. Research Strategy

Research design is usually defined as a general plan of how one intends to answer the research questions and it is informed by the research objectives. The objectives of this research project are to identify, analyse, and mitigate factors influencing the ability of

Computer Science and Information Technology academics to effectively use of online search engines as teaching and research resources.

The researcher chose to achieve this research objective by undertaking a survey of a sample of academic staff from different institutions of higher learning. A questionnaire was distributed to these research participants in order to capture their perceptions on possible factors that may influence their searching abilities for their teaching and research activities.

The choice of a survey as a research strategy for this research is mainly justified by the fact that surveys are usually considered as a simple and cost effective research strategy whereby a large amount of data can be collected within a reasonable time frame (Van Schaik et al., 2007), and secondly because surveys are suitable for perceptions analysis studies as it is the case for this research. As it is the case for all surveys, this research therefore does not intend to establish causal relationships between its research variables, instead it identifies and analyses relationships between these research variables either individually or as combinations.

The conduct of this research follows the standard protocol adopted by questionnaire based surveys: selection of the research population and sampling, questionnaire design, data collection, data analysis, and results reporting (Kelley at al., 2003) as described in the following sections.

4.3. Research Population and Sampling

The target population of this research consists of academic staff in the field of Information Technology (IT) and Computer Science (CS), two of the five main classifications of computing disciplines (based on ACM). For convenience, this academic staffs were selected from four universities around the area where the research took place. The research was conducted in the department of Information Technology at the Durban University of Technology located in the KwaZulu-Natal province of South Africa. However, because the researcher was an academic staff of the IT department of the Walter Sisulu University located in the Eastern Cape province of South Africa, it was convenient to select respondents from IT and CS departments from universities from both provinces. The KZN province has four universities and three of them were selected in this research. The fourth university was not included in the research because their IT department does not offer postgraduate programmes. The Eastern Cape Province also has four universities and only one of them was selected in this research due to geographical distance constraints between these universities. The total number of staff from the above departments could be estimated to be 90. A total of 61 lecturers or academic staff at the four different universities' departments participated in the research.

In this study, research participants were to be selected from the chosen institutions using a non-probabilistic sampling method. It was not possible to choose a probabilistic sampling method because the researcher did not have beforehand a complete list of potential research respondents from which it would have been possible to operate a random selection. Thus the sampling method of this research was a mixture of convenience sampling and self-selection sampling: In some cases, the research participants were approached at the end of a departmental meeting, and in some other instances, the researcher met individual respondents in their offices without any prior arrangement. Not all the academic staff accepted to participate in the research. Those who accepted to participate in the research were requested to fill in a questionnaire. The questionnaire distribution was done at different times over a period of three weeks. Appendix A shows the detailed questionnaire items.

4.4. Data Collection

The data of this study were collected in the form of numerical and standardised data. The questionnaire consisted of rating questions except for the first section of the questionnaire on the respondents' biographical data. The questionnaire consisted of the following five sections: background biographical information, lecturing experience, language ability, prior research experience, and searching ability. The first four sections of the questionnaire represented factors that were assumed to have a possible impact of the searching abilities of an academic staff. Therefore these four sections represent the independent variables while the last one represents the dependent variable. A summarised description of the questionnaire sections is given below.

Section A: Background Information (Biographical Data)

Section A of the questionnaire consisted of questions requesting research participants to provide data on their following biographical profile attributes: the name of their academic department, their gender, age group, their position, the name of their institution, their research and teaching experience, their highest qualification, the language in which they received their primary and secondary education, and the language in which they received their tertiary education. Respondents were not required to write their names on the questionnaire; but to simply indicate their biographical details by marking a cross in an appropriate block option.

Section B: Lecturing Experience

Section B of the questionnaire consisted of 10 items. Each item was a statement intended to measure how an academic staff assesses his or her level of experience in a specific aspect of the teaching profession such as the design of course materials, the setting up of question papers, and the handling of heavy lecturing loads. For each statement, each participant was requested to auto-assess his or her experience using a 5-point Likert rating scale ranging from '*Strongly Agree*' (5) to '*Strongly Disagree*' (1), by simply putting a cross in an appropriate block option.

Section C: Research Experience

Section C of the questionnaire also consisted of 10 items. Each item was a statement intended to measure how an academic staff assesses his or her level of experience in a specific aspect of a research project such as the writing of papers, making oral presentations at conferences, publishing articles for journals, publishing book chapters or books, etc. For each statement, each participant was requested to auto-assess his or her experience using a 5-point Likert rating scale ranging from '*Strongly Agree*' (5) to '*Strongly Disagree*' (1), The questionnaire was designed in a way that the respondents could just put a cross in an appropriate block to indicate their choice of an answer question.

Section D: English Language Ability

Section D of the questionnaire consisted of 10 items. Each item was a statement intended to measure how an academic staff assesses his or her level of aptitude to use the English language with regards to their writing, speaking and reading abilities, how easily the participant spot mistakes while reading, the extent to which the participant uses the language out of his or her work environment, etc. For each statement, each participant was requested to auto-assess his or her language proficiency using a 5-point Likert rating scale ranging from '*Strongly Agree*' (5) to 'S*trongly Disagree*' (1). Like all Likert scale items in this research, here again, respondents were required to simply put a cross in an appropriate block to indicate their answer question.

Section E: Prior Search Experience

Section E of the questionnaire also consisted of 10 items. Each item was a statement intended to measure how an academic staff assesses his or her level of searching experience in the use of search engines in terms of the number of years the participant had been using search engines, their knowledge of specialised computing digital libraries, their technical knowledge of search engines, etc. For each statement, each participant was requested to auto-assess his or her prior searching experience using a 5-point Likert rating scale ranging from '*Strongly Agree*' (5) to '*Strongly Disagree*' (1). Here again, respondents were required to simply put a cross in appropriate block options to indicate their answers.

Section F: Searching Ability

Section F of the questionnaire also consisted of 10 items. Each item was a statement intended to measure how an academic staff assesses his or her searching ability to find information online including the ability to find relevant information with ease, and to choose suitable keywords to use for a specific search, etc. For each statement, each participant was requested to auto-assess his or her searching ability through a 5-point

Likert rating scale ranging from '*Strongly Agree*' (5) to '*Strongly Disagree*' (1). Here again, respondents were required to simply put a cross in appropriate block options to indicate their answers.

The questionnaire thus includes a total of 60 items all written in English in which only sticks or crosses were required from the respondents.

4.5. Data Analysis

In order to ensure that all the questions were clearly understood by the research participants, a pilot test was carried out on 5 academic staff before the final questionnaire was distributed. The purpose of the pilot test was to minimize the likelihood of respondents having problems in answering the questions. During the pilot pretesting of the questionnaire, the following few changes were made to its biographical section:

- An additional age group was added to the age group item;
- The Associate Director option was added to the position's item;
- The isiXhosa/isiZulu language option was added as additional language for primary and secondary education, and also for tertiary education. IsiXhosa and isiZulu are the two main languages spoken by the natives of the KwaZulu Natal and Eastern Cape provinces of South Africa, the two provinces where this current research took place.

The final version of the questionnaire was distributed to the research participants and the collection of these questionnaire yielded research data that were later on analysed by the researcher. The first phase of the data analysis consisted of the testing of the reliability and validity of the collected data. Then a number of statistical tests were executed on the data with the help of the PASW STATISTICS 18.0 (SPSS) Statistics software package.

4.5.1. Data Reliability and Validity for each research Variable

A questionnaire is only useful when its data is reliable and valid hence the need to always test reliability and validity of research data. Data reliability for each section of the questionnaire was established in this study using the Cronbach's coefficient alpha (α) method. Section A (biographical data) of the questionnaire, not being based on Likert scale items, was not considered for reliability and validity tests. For all the other sections of the questionnaire, reliability and validity tests were performed. According to Field A. (2005:254), the Cronbach's coefficient alpha (α) stipulates that reliability coefficients less than 0.700 are considered to be poor; while α between 0.700 and 0.799 are acceptable and α above 0.800 are regarded as good. For variables with a poor α, there is a need to perform a further factor analysis in order improve the reliability of the variable by removing items spreading on more than two components.

According to Field A. (2005:258), for a variable to be considered as valid, its factor analysis requires to have a Determinant D greater than 0.00001 and a KMO and Bartlett's coefficient test of sphericity greater than 0.500. Factor analysis also allows the testing of construct validity by validating constructs reduced into a single component. Therefore, in the case of variables with poor Cronbach's coefficient alpha, factor analysis ultimately leads to a situation whereby all non-suitable items of the questionnaire are removed to the point where the variable in question became reliable and valid.

In the case of variables with good Cronbach's coefficient alpha, factor analysis was also performed to lead to the removal of non-suitable items in order to make the variable both reliable and valid. The following paragraphs give the details of the steps followed in SPSS to test the construct validity of the first variable through factor analysis. The details of the factor analysis for the other variables are not described in this chapter because it follows the same procedure.

Step 1: Calculation of α Cronbach's coefficient alpha

The calculation of Cronbach's coefficients alpha (α) is done in PASW STATISTICS 18.0 (SPSS) using the menu *Analyze* > *Scale*> *Reliability Analysis* and by selecting only the *Alpha model* option in the *Reliability Analysis* window.

Step 2: Setting up the factor analysis model

Factor analysis is done in PASW STATISTICS 18.0 (SPSS) by selecting the menu Analysis > Dimension Reduction > Factor. This menu then opens the *Factor Analysis* window where the user is requested to specify a certain number of options. This window has the following five buttons: Descriptives, Extraction, Rotation, Scores, and Options. In the case of this study, the following three radio buttons in the *Descriptives* window were selected: the *Initial solution*, *Determinant,* and *KMO and Bartlett's test of sphericity*. The following four items in the *Extraction* window were also selected: *Principal Components*, *Correlation matrix*, *Unrotated factor solution*, and *Eingenvalue* greater than 1. The following two items in the *Rotation* frame were selected: *Varimax, and Rotated solution*. No change was made in the *Scores* window. The following two items in the *Options* window were selected: *Exclude cases pairwise*, and *Absolute value* below 0.450.

Step 3: PASW STATISTICS 18.0 (SPSS) viewer graphs and values

After setting up the factor analysis window, the ten (10) questionnaire items (11 to 20) of the first Likert scale based research variable were moved into the *variables* box in PASW STATISTICS 18.0 (SPSS) . The results of the factor analysis for these 10 items are described by appendix B. Even though the correlation matrix is not shown as per request of the researcher, the value of the determinant is still meaningful. For the KMO and Barthlett's test table, only the Kaiser-Meyer-Olkin Measure of Sampling Adequacy is of interest for data validity tests. The other values on the KMO and Barthlett's test table (Chi-Square, df, and Sig. values) are not important at this point. Appendix B also shows a table explaining the total variance of components whereby it is obvious that there are only three components with *Eingenvalues* greater than one. All these three components contribute for more than 60% of the total component spreading.

Appendix B ends with three tables: a component matrix, a rotated component matrix, and finally a component transformation matrix. Only the rotated component matrix is useful for the purpose of data validity testing. The rotated matrix shows some blank space due to the prior exclusion of spreading coefficients less than 0.450. These blanks therefore mean that the spreading of the item in question is negligible for the specified component.

Any item spreading on more than one component is considered as ambiguous because it measures more than one construct. That is for example the case of item 18 in the rotated component matrix in appendix B. Item 18 weights 0.583 on component 1 and 0.650 on component 2 therefore it is split between two constructs and it makes sense to discard it in order to improve the construct validity of the variable under study. Once ambiguous items are discarded from the rotated component matrix, there are still three components to be considered. At this point there are two possibilities for construct validity: to break the variables into three or to only select one component (Principal component). The latter option is preferred in order to avoid having too many research variables. Thus items 12, 13, 14 and 15 are the only ones remaining for the first Likert based scale research variable of this study.

Step 4: Repeating validity testing for the extracted items

In this step, once the researcher had assessed D and KMO values and had observed the spreading of research variable items into only one component by the factor analysis, it became clear that the these items were valid for the variable under study; but the Cronbach's reliability coefficient for these four items still need to be calculated. The resulted Cronbach's reliability coefficient was weak, therefore triggering the need for the removal of more items from the variable. After the removal of the item with the lowest coefficient (0.544) on the component matrix, the Cronbach's reliability coefficient becomes satisfactory.

Step 5: Test reliability and validity for the last selected items (12, 14, 15)

Item 13 was therefore removed from the item list and the researcher ran validity and reliability tests for the remaining three items (12, 14, 15) and acceptable validity and reliability values were finally obtained D=0.449, KMO = 0.654 and Cronbach's coefficient alpha (α) of 0.757.

In summary, the first variable, lecturing experience consisting of items 12, 14, 15 is valid and reliable. Reliability and validity tests on other Likert scale based research variables also indicated that: items 23, 24, 25, 26, 27, 28 are valid and reliable for the research experience variable, items 31, 32, 33, 34, 36 are valid and reliable for the English language ability, items 42, 43, 44, 45 are valid and reliable for the searching experience variable, and items 51, 53, 54, 57, 58, 59 are valid and reliable for the searching ability variable.

4.5.2. Overall Data Validity

The researcher needed to establish the overall data validity of the questionnaire Likert scale based independent variables after its data cleaning as described by the above. This overall validity was done using the factor analysis for all valid and reliable Likert scale items for the independent research variables namely 12, 14, 15, 23, 24, 25, 26, 27, 28, 31, 32, 33, 34, 36, 42, 43, 44, and 45 (see details in appendix B).

4.6. Statistical Tests

After completion of validity and reliability tests, the research data was analysed using descriptive and inferential statistical methods. Descriptive statistical methods using frequencies, means, summation, standard deviations and variances were done to count the number of participants in the sample for the different biographical attributes groups, and to measure meaningful values for the Likert scale based research variables. Inferential statistics were also calculated using correlations, Analysis Of Variance (ANOVA) and linear regression analysis in order to interpret the data by showing associations among research variables; these were associations between users' searching ability and their lecturing experience, research experience, language ability, and prior searching experience.

4.6.1. Variables Measurement Classification

The first step before the running of reliability, validity, descriptive and inferential statistical tests was the coding of the data especially for section A of the questionnaire. It was also necessary to categorise variables according to the "NOIR" classification or measures. According to Field A. (2005:254), there are four kinds of levels of measurement for variables namely Nominal (N), Ordinal (O), Interval (I) or Ratio (R). N and O are Non-parametric data while I and R are parametric data. This is important because the measure of the dependent variable determines the nature of statistical tests.

Table 4.1 gives the details of the measures and codes for all the research variables of this study.

For variables items representing academic staff age groups, positions, teaching experience, research experience, highest qualification, the Ordinal (O) measure was chosen because these variables present some kind of order. For example, the staff positions vary from junior lecturers to professors. One can therefore put an order from junior staff (lowest position) to professor (highest position).

The next set of variables (staff academic departments, genders, university names, language mostly used in primary, secondary, and tertiary education) do not present any sense of order; they are simply nominal (N) as they represent different categories of items. For example there is no particular order between different genders: a staff can either be a male or a female.

The last sets of variables are Ratio measures because they are Likert scale or computed values. A Likert scale item is designed to help researchers understand how respondents perceive a particular view or product or service by asking them to rank their perceptions on a gradual scale with specific points or ratios.

4.6.2. Descriptive Statistics

The following types of descriptive statistics were used in this research during data analysis in order to summarise the data: frequencies, means, standard deviations and variances. A detailed description of these statistics is presented in appendix C. However, the following paragraphs present the three major steps used in this study during descriptive statistical analysis in PASW STATISTICS 18.0 (SPSS).

Step 1: Summation of variables items

Except for the variables representing the respondents' biographical data (section A), all reliable and valid items were summed according to their variables as indicated by table 4.1. This summation is possible since all the answers are Likert scale values ranging from 1 to 5 and it allows the reduction of variables into the following:

Table 4.1 Variables, their coded values and their "NOIR" measure types

Variable Name	Measure	Questionnaire Section	Coded value
Question 1	Nominal	A	1 = "IT" 2 = "IS" 3 = "CS" 4 = "Other"
Question 2	Nominal	A	1 = "Male" 2 = "Female"
Question 3	Ordinal	A	1 = "25 or less" 2 = "26 to 35" 3 = "36 to 45" 4 = "46 to 55" 5 = "56 and more"
Question 4	Ordinal	A	1 = "Other" 2 = "Junior Lecturer" 3 = "Lecturer" 4 = "Senior Lecturer" 5 = "Associate Director" 6 = "Associate Professor" 7 = "Professor"
Question 5	Nominal	A	1 = "DUT" 2 = "UKZN" 3 = "WSU" 4 = "Zululand"
Question 6	Ordinal	A	1 = "2 or Less" 2 = "3 to 6" 3 = "7 to 10" 4 = "11 to 14" 5 = "15 and more"
Question 7	Ordinal	A	1 = "2 or Less" 2 = "3 to 6" 3 = "7 to 10" 4 = "11 to 14" 5 = "15 and more"
Question 8	Ordinal	A	1 = "ND" 2 = "BTech" 3 = "BSc" 4 = "BSc(Hons)" 5 = "Masters" 6 = "Doctorate"
Question 9	Nominal	A	1 = "Zulu/Xhosa" 2 = "African" 3 = "Afrikaans" 4 = "English" 5 = "French" 6 = "Other"
Question 10	Nominal	A	1 = "Zulu/Xhosa" 2 = "African" 3 = "Afrikaans" 4 = "English" 5 = "French" 6 = "Other"
Lecturing Experience	Ratio	B	
Research Experience	Ratio	C	
Language Ability	Ratio	D	
Prior Search Experience	Ratio	E	
Searching Ability (Dependent Variable)	Ratio	F	

Lecturing_Exp_Total_Score,

Research_Exp_Total_Score,

English_Ability_Total_Score,

Prior_Search_Abil_Total_Score, and

Search_Abil_Total_Score_DV.

Computing a new variable in PASW STATISTICS 18.0 (SPSS) is done by selecting the menu *Transform* > *Compute Variable*. The *Compute Variable* window opens and it comprises a panel of buttons and boxes namely *Target variable, Numeric Expression, Function group*, and *Functions and Special Variables*. The name of the new variable is written in the *Target variable* box and the computation is defined by typing an expression in the *numeric expression* box. In the case of this research, the researcher simply added the items of each section for which the variable was valid and reliable as described in section 4.5.1. The summation details of each research variable are shown in table 4.2.

Table 4.2 Summation for each variable

Target Variable	**Numeric Expression**
Lecturing_Exp_Total_Score	*I12+I14+I15.*
Research_Exp_Total_Score	I23+I24+I25+I26+I27+I28
English_Ability_Total_Score	I31+I32+I33+I34+I36
Prior_Search_Abil_Total_Score	I42+I43+I44+I45
Search_Abil_Total_Score_DV	I51+I53+I54+I57+I58+I59

Step 2: Frequency analysis

Frequencies are calculated in PASW STATISTICS 18.0 (SPSS) by selecting the menu *Analyze* > *Descriptives Statistics* > *Frequencies* where three options are presented: *Statistics, Charts,* and *Format*. In the case of this study, the following items were selected from the *Statistics* window: *Mean, Std. deviation*, and *Variance*. On the C*harts* window, different *charts type* buttons were selected depending on the type of frequency analysis needed; therefore one of either the *Bar charts*, or the *Pie charts*, or even *the Histograms* was selected. The following items were also selected from the *Format* window: *Ascending values*, and *Compare variables*.

Step 3: Descriptive type analysis

Descriptive statistics (mean, variance, standard deviation) are analysed in PASW STATISTICS 18.0 (SPSS) by selecting the menu *Analyze* > *Descriptive Statistics* > *Descriptives* ... For this research, the following *Options* were selected*: Mean, Minimum, Maximum, Variance, Range, Variable list.*

The above presented descriptive statistics only describe the range and summary of the research data. A further analysis needs to be made on the variables' associations using inferential statistics.

4.6.3. Inferential Statistics

Inferential statistics compute relationships between two or more variables, differences between different subgroups, and the account of independent variables towards the variance of the dependent variable (Kelley at al., 2003). For the purpose of this research and based on the nature of the dependent variable (Searching Ability) at hand, the following inferential statistics were used: Pearson's Correlation Test, Multiple Regressions Analysis, ANalysis Of VAriance (ANOVA), and ANalysis Of COVAriance (ANCOVA).

4.6.2.1 Pearson's Correlation Test

The Pearson's correlation test is the measure of how two variables are related. Pearson's correlation coefficients (ρ) values range from -1 to 1. The sign of this coefficient indicates the direction of the relationship (positive or negative) between the two variables. Absolute values of the correlation coefficient indicate the strength of the relationship, with larger absolute values indicating stronger relationships. For the purpose of this study, the researcher performed Pearson's correlation tests to identify the variables that have a linear correlation with the staff searching abilities.

Pearson's correlation tests are performed in PASW STATISTICS 18.0 (SPSS) by selecting the menu *Analyze > Correlate > Bivariates*. The *Bivariate* correlations window shows the following settings: *correlation coefficients, Variables, Test of Significant,* and *Options*. For this research, the following options were selected: *Pearson, Two-tailed,* and *Flag significant correlations*. On the *Options* window the following were also selected: *Means and standard deviations, and Exclude case pairwise*.

The following variables were analysed for correlation test: academic staff lecturing experience, their research experience, their English language ability, and their prior search experience against the searching ability. These test resulting tables and graphs are presented in appendix D.

4.6.2.2 Multiple Regression Analysis

Multiple regression analysis is usually performed in the case whereby more than one independent variable is used to predict the behaviour of the dependent variable. The following paragraph presents the major steps followed in this research for the purpose of performing multiple regression analysis using PASW STATISTICS 18.0 (SPSS).

Multiple linear regression in PASW STATISTICS 18.0 (SPSS) is performed by selecting the menu *Analyze > Regression > Linear*. This opens the *linear Regression* window. This window shows different settings to be made by the user: *Dependent variable, Independent(s) variables, Selection Variable, Case Labels, WLS Weight, Statistics, Plots,* and *Options*. For the purpose of the current study, the following settings were made:

- Declaring Searching ability as the *Dependent variable* as well as the five independent variables: lecturing experience, research experience, English language ability, and prior search experience.
- Setting parameters for the *Regression Coefficient* in the *Statistics* window: *Estimates, Confidence interval level (%)*-set to 95, and *Model fit*.

- Setting parameters for the *Standardized Residual Plots* in the *Plots* window by selecting the *Normal probability plot* option.

- Setting parameters for the *Predicted Values* as *Unstandardized* in the *Save* window.

- Instructing PASW STATISTICS 18.0 (SPSS) to use 0*.05 and 0.10 respectively as the Entry* and *Removal* value for the calculation of the *Stepping method criteria* as requested *by the* Options window. In the same window, the following two options were also selected: *Include constant in equation, and Exclude cases pairwise*.

4.6.2.3 The One-way ANOVA

The ANOVA analysis is performed in order to show and predict the behaviour of the dependent variable against each of the demographic factors. There are different types of ANOVA analysis depending on the number of factors predicting the dependent variable. In this research, a one-way ANOVA was performed for each of the ten (10) items of the biographical background section of the questionnaire.

ANOVA was performed in PASW STATISTICS 18.0 (SPSS) by selecting the menu *Analyze > General Linear Model > Univariate whereby* the following items were set up for this study: *Model, Contrasts, Plots, Post Hoc, Save, and Options*. The following settings were made while using the PASW STATISTICS 18.0 (SPSS) menus:

- Dependent variable: *Search_Abil_Total_Score_DV*
- Covariate: *left empty*
- Fixed Factors: any *Item among the 10 of biographical information*
- *FullFactorial Model:*
- *Plots: One of the fixed Factor to* the horizontal axis
- *Post Hoc: Tukey* procedure, and *include the fixed factor* as the Post Hoc Tests variables.

- *Options: Homogeneity tests*, and 0.05 for the significance level (95% CI corresponds to a 5% (0.05).

ANOVA outputs are presented in the form of tables or graphs in the PASW STATISTICS 18.0 (SPSS) Statistics viewer as showed on appendix D.

4.6.2.4 The five-way ANCOVA

From the ANOVA analysis above described, the researcher found that five (5) biological factors were predictors of staff searching ability. Furthermore the Pearson's Correlation analysis (section 4.6.2.1) showed a correlation between the dependent variable (searching ability) and two (2) other independent variables. A five way ANCOVA was therefore performed in this study to predict the behaviour of searching abilities when these significant biographical factors and co-variables were combined.

There are different types of ANOVA analysis depending on the number of factors predicting the dependent variable. A one way ANOVA has only one predictor while a two way ANOVA has two predictors. In the case where at least one predictor is as covariant, the name of the ANOVA analysis changes to ANCOVA (ANalysis of COVAriance). Consequently it is appropriate to indicate that the researcher performed a five way ANCOVA in this study as described below. The following paragraph presents the major steps of this analysis as measured using PASW STATISTICS 18.0 (SPSS) .

ANCOVA was performed in PASW STATISTICS 18.0 (SPSS) by selecting the menu *Analyze > General Linear Model > Univariate* whereby the following settings were made for this study: *Model, Contrasts, Plots, Post Hoc, Save, and Options. The following changes were made*

- Dependent variable: *Search_Abil_Total_Score_DV*
- Covariate: *English_Ability_Total_Score, Lecturing_Exp_Total_Score*
- Fixed Factors: 5 *Items of biographical information*
- *Fullfactorial Model:*
- *Plots:* 5 *Items of biographical information* for the horizontal axis
- *Post Hoc: Tukey* procedure, and 5 *Items of biographical information* as the Post Hoc Tests variables.

- *Options: Homogeneity tests*, and 0.05 for the significance level (95% CI corresponds to a 5% (0.05).

ANCOVA outputs are presented in PASW STATISTICS 18.0 (SPSS) in the form of tables and graphs as shown in appendix D.

4.7. Conclusion

This chapter presented the methodology used by the study, as guided the research objective. A questionnaire based survey served as a research instrument which was successively tested for reliability and validity. The chapter described the convenience sampling that was used for the construction of a sample of 61 Information Technology and Computer Science lecturers from four different Universities in both the KwaZulu-Natal and Easter Cape provinces. Out of 61 questionnaires distributed, all 61 questionnaires were returned making it a 100% response rate. These participants were requested to simply indicate their choice from a 5 point Likert-scale questions. The questionnaire's first section contained demographic items while the other sections constituted the research variables namely lecturing experience, research experience, English language proficiency, and web searching experience with the last section, searching ability capturing the dependent variable. Different data analysis techniques are presented in this chapter: descriptive statistics (frequency distributions, mean and Standard deviation); and inferential statistics (Pearson's correlation analysis, multiple regression, ANOVA, and ANCOVA). The next chapter presents the results obtained from these analyses.

CHAPTER FIVE

Presentation of Results

5.1. Introduction

This chapter presents the results of the data analysis tests described in the previous chapter. The chapter starts by presenting reliability and validity tests results for the different Likert scale based sections of the questionnaire. Secondly, frequency distributions tables and Charts are presented in order to describe the biographical profile of the research respondents. The chapter thereafter presents results on the relationships between the different research variables using the inferential statistics methods described in chapter 4 (Pearson's correlation analysis, multiple regression, one way ANOVA, and ANCOVA). The structure of this chapter is similar to the structure of chapter four because the purpose of this current chapter is simply to present the results obtained from the conduct of the research as per the research design described in chapter four.

5.2. Reliability and validity results.

As explained in section 4.5.1 of chapter four, reliability testing of questionnaire data was done by calculating the Cronbach's coefficient alpha (α) for each Likert scale based section of the questionnaire. Data from a questionnaire section is deemed to be reliable only when its Cronbach's coefficient alpha is greater than 0.700. On the other hand, the validity of data from a section of the questionnaire is performed by doing factor analysis as described in the previous chapter. Factor analysis consists of a KMO and Bartlett's test as well as the computing of the component matrix Determinant (D). For the purpose of this study, data from a questionnaire section is deemed valid when the following three conditions are met:

- D is greater than 0.00001
- Kaiser Meyer Olkin (KMO) measure of sampling adequacy is greater than 0.500
- Only one component is extracted by the section's items after factor analysis

It is important to recall that each section of the questionnaire represents a research variable: section A represents respondents' biographical profiles, section B measures staff lecturing experience, section C represents staff research experience, section D measures staff English language ability, section E represents staff searching experience, and section F measures staff searching ability. This chapter presents in details the results of reliability and validity tests for the second research variable (lecturing experience). The results of the reliability and validity tests for the other research variables are then simply presented in the form of a table (see table 5.1 below).

The second variable, lecturing experience, initially consisted of ten questionnaire items numbered from item 11 to item 20. These groups of ten items gave an initial Cronbach's coefficient alpha (α) of 0.719 but three components were extracted from the initial factor analysis validity test as described in section 4.5.1 of chapter four. From these ten questionnaire items, only the following three items were able to reduce the variable into a single component through factor analysis: items 12, 14, and 15. For these three items, a factor analysis was performed with the following results: the Cronbach's coefficient (α=0.757) was greater than 0.700, the KMO value (0.651) was greater than 0.500 and the determinant value (D=0.449) was also greater than 0.00001. Therefore, data related to the second research variable of the questionnaire was deemed to valid when only items 12, 14, and 15 were considered. Table 5.2 also shows that the questionnaire data is reliable and valid for all the other research variables of this study. More details are available on factor analysis tables (Correlation Matrix, KMO and Bartlett's Test, communalities, Total variance explained, and Component matrix) and on reliability statistics tables in appendix B.

TABLE 5.1 Reliability and validity tests results for the questionnaire data

Variables (numbers of valid and reliable items)	Selected questionnaire Items	Cronbach Alpha coefficient α	KMO	Determinant (D)
Lecturing Experience (3)	12, 14, 15	0.757	0.651	0.449
Research Experience (6)	23, 24, 25, 26, 27, 28	0.925	0.833	0.006
Language Ability (5)	31, 32, 33, 34, 36	0.783	0.805	0.085
Prior search Experience (4)	42, 43, 44, 45	0.806	0.767	0.27
Searching Ability (6)	51, 53, 54, 57, 58, 59	0.862	0.822	0.069

The researcher also performed the testing of the overall validity of the questionnaire data by factor analysing all the reliable and valid items from the Likert scale based independent variables and the results were as follows:

- Number Items = 18
- Number of Components extracted is 4 (representing our four independent variables)
- KMO = 0.697

Figure 5.1 shows the rotated matrix of the reliable and valid items from the Likert scale based independent variables (see a detailed description in appendix B). The extraction of four components from this factor analysis confirms that the Likert scale based independent variables of the questionnaire are indeed valid to represent four independent variables.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
11				.812
12				.699
15				.847
23	.875			
24	.869			
25	.835			
26	.876			
27	.784			
28	.855			
31		.725		
32		.819		
33		.870		
34		.775		
36		.648		
42			.807	
43			.843	
44			.811	
45			.583	

Figure 5.1 Rotated Matrix for the four Likert scale based independent variables

5.3. Descriptive Statistics

Biographical data collected from the questionnaire is used in this section to produce a set of descriptive statistics that can be analysed to highlight interesting characteristics of the sample of academic staff that participated in this research. Descriptive statistics are also performed on all Likert-scale based research variables in order to have a general idea on these values for the population under study. There are many forms of descriptive statistics namely mean and mode analysis, variance and standard deviation analysis, and frequency distribution analysis; but because the data in section A of the questionnaire is mostly nominal and ordinal, only the mode and frequency distribution analysis are relevant in this section.

5.3.1. Respondents Biographical Profiles

This section looks at the biographical profile of the research respondents for items described on the first section of the questionnaire: the name of their academic department, their gender and age group, their position, the name of their institution, their research and teaching experience, their highest qualification, the language in which they received their primary and secondary education, and the language in which they received their tertiary education.

Before the presentation of the descriptive statistics it is important to note that the response rate for this research was 100% since all respondents who agreed to fill up a questionnaire returned it completed fully. It is also important to recall the estimated number of staff members for each of the Information Technology (IT) and Computer Science (CS) university departments under study as showed on table 5.2.

According to table 5.2, the estimated total number of staff in the departments under study is 90. The number of respondent's in this research is 61 corresponding to sixty eight percent (68%) of the total number of staff members in the departments under study.

Table 5.2 Staff complement for the computing departments of the universities under study

University	Department	Number of Staff
Durban University of Technology (DUT)	IT	40
University of KwaZulu Natal (UKZN)	CS	10
Walter Sisulu University (WSU)	IT, CS	35
Zululand (UNIZUL)	CS	5
Total		90

5.3.1.1. Academic Department

Table 5.3 and figure 5.2 represent the academic departments the respondent belongs to. They show that respondent's academic departments are categorized into two disciplines: Computer Science (CS) and Information Technology (IT), with 23% of the staff coming from the CS discipline while 77% are from the IT discipline.

Table 5.3. Academic Department

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Information Technology	47	77.0	77.0	77.0
	Computer Science	14	23.0	23.0	100.0
	Total	61	100.0	100.0	

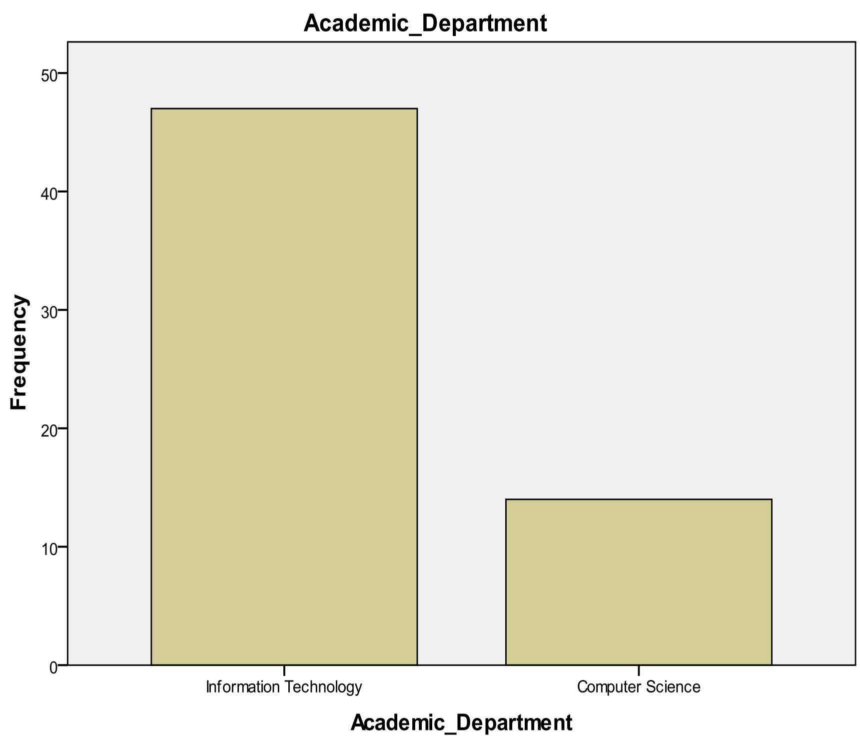


Figure 5.2 Frequency distribution of the staffs' academic departments

5.3.1.2. Gender

Table 5.4 represents the gender distribution of the respondents. The table shows that 41% of the respondents are females and a 59% of them are males. A graphical representation of this distribution is depicted by Figure 5.3.

Table 5.4 Gender frequency distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	36	59.0	59.0	59.0
	Female	25	41.0	41.0	100.0
	Total	61	100.0	100.0	

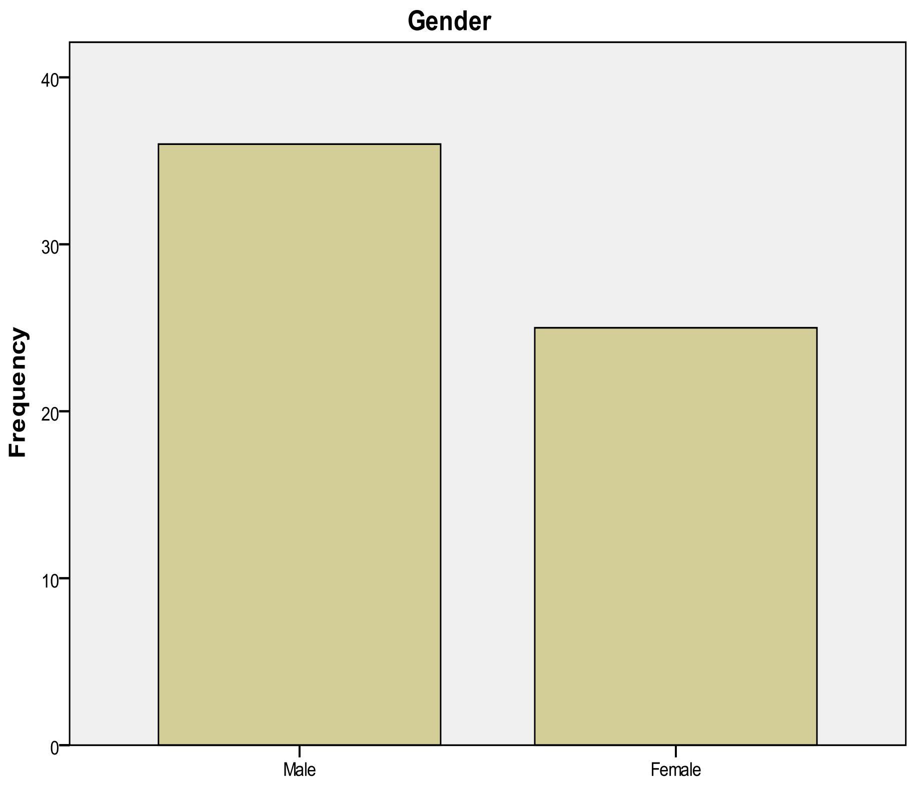


Figure 5.3: Gender distribution

5.3.1.3. Age groups

Table 5.5 and Figure 5.4 represent the different age groups of the respondent. They show the distribution of the respondent's ages across all the age groups specified in the questionnaire. The histogram in figure 5.4 shows that the lowest frequencies are for the less than twenty five year old age groups (25 or less) and for the more than fifty six years old age group (56 and more). The figure also reveals that three medium age groups present much higher frequencies. Thus the age groups distribution of the respondents sample tends to match the shape of a normal distribution curve as shown on the histogram.

Table 5.5 Age group frequency distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	25 or less	5	8.2	8.2	8.2
	26 to 35 yrs	22	36.1	36.1	44.3
	36 to 45 yrs	17	27.9	27.9	72.1
	46 to 55 yrs	12	19.7	19.7	91.8
	56 and more	5	8.2	8.2	100.0
	Total	61	100.0	100.0	

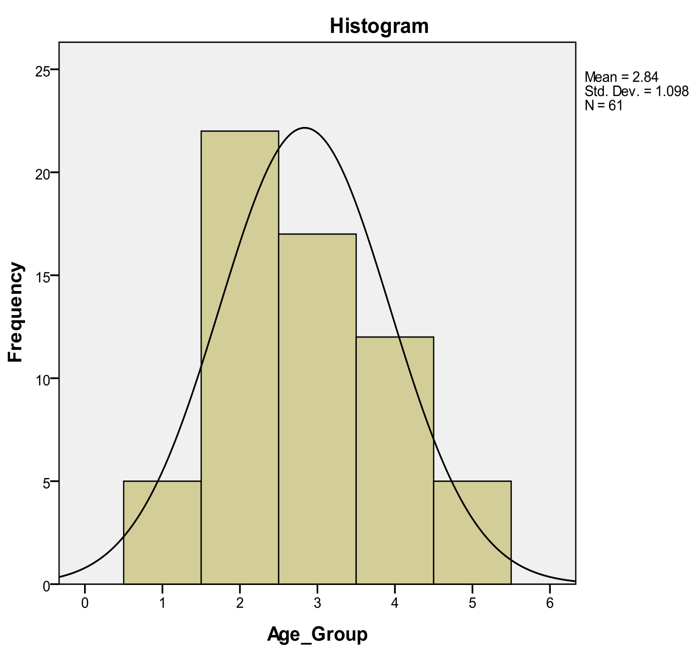


Figure 5.4 Age distribution

5.3.1.4. Position held

Table 5.6 and Figure 5.5 represent the different staff positions in their respective universities. They show the distribution of these positions in the range of positions specified by the questionnaire. The histogram in figure 5.5 shows a gradual increase in the staff numbers until the position of lecturer (52 %). Thereafter, figure 5.5 shows a decrease in frequencies from senior lecturer down to the position of professor. Thus this figure reveals that the position distribution of the sample staff tends to have the shape of a normal distribution as expected from the population.

Table 5.6 Positions frequency distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Others	3	4.9	4.9	4.9
	Jn. lecturer	13	21.3	21.3	26.2
	Lecturer	32	52.5	52.5	78.7
	Sr. Lecturer	5	8.2	8.2	86.9
	Associate Dir	4	6.6	6.6	93.4
	Associate Prof	3	4.9	4.9	98.4
	Professor	1	1.6	1.6	100.0
	Total	61	100.0	100.0	

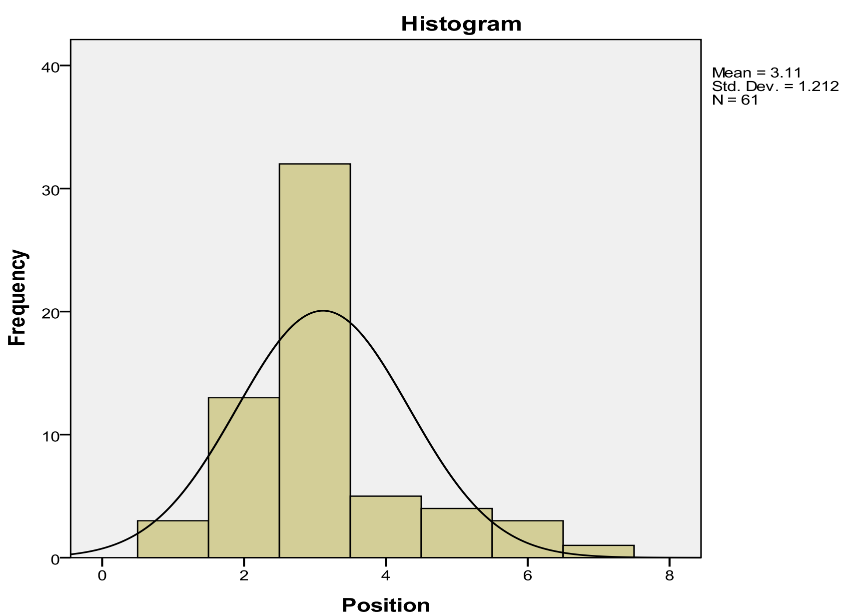


Figure 5.5 Respondents positions

5.3.1.5. University of Affiliation

Table 5.7 and Figure 5.6 represent the number of respondents per university. The chart reveals a majority (45.9%) of the respondents from WSU in the EC province where the researcher resides, followed by respondents from DUT (37.7) from the KZN province.

Table 5.7 University affiliation frequency distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Durban University of Technology(DUT)	23	37.7	37.7	37.7
	University of KwaZulu Natal (UKZN)	4	6.6	6.6	44.3
	Walter Sisulu University (WSU)	28	45.9	45.9	90.2
	University of Zululand	6	9.8	9.8	100.0
	Total	61	100.0	100.0	

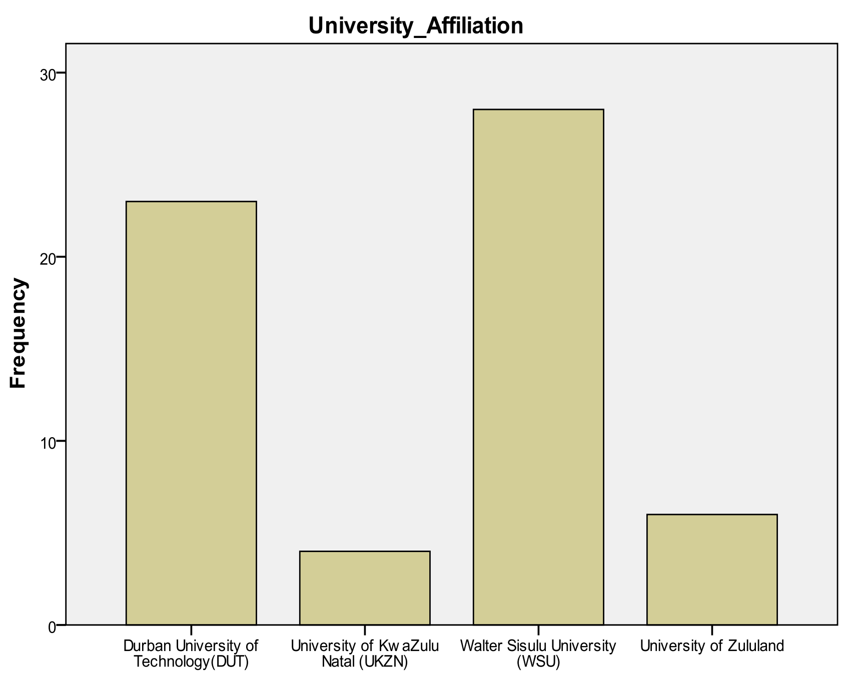


Figure 5.6 Frequencies of staff members per institution

5.3.1.6 Years of Lecturing Experience

Table 5.8 and figure 5.7 represent the distribution of the respondent years of lecturing experience. They show a decrease in frequencies as the years of lecturing experience increases. They reveal that:

- 14.8% are novice in lecturing (with less than 3 years in lecturing experience),
- 47.7% have a considerable number of year of lecturing experience, and
- 37.7 % have more than ten (10) years of lecturing experience.

These numbers show that the sample used resembles the population of aging universities lecturers.

Table 5.8 Staff years of lecturing experience

Years		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 or Less	9	14.8	14.8	14.8
	3 to 6	16	26.2	26.2	41.0
	7 to 10	13	21.3	21.3	62.3
	11 to 14	12	19.7	19.7	82.0
	15 and more	11	18.0	18.0	100.0
	Total	61	100.0	100.0	

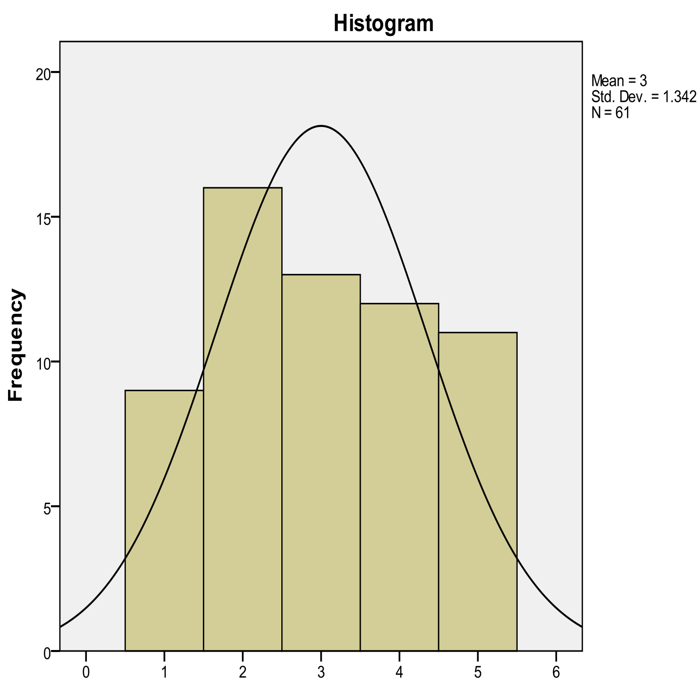


Figure 5.7 Staff's number of year experience in lecturing

5.3.1.7 Years of research experience

Table 5.9 and figure 5.8 represent the distribution of the respondents years of research experience. The table reveals that more that eighty two (82.2) percents of the participants are young (less than 7 years) researchers while only about eighteen (actually 17.8) percent of them are senior (7 years or more) researchers.

Table 5.9 Years of Research Experience

Years		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2 or Less	33	54.1	54.1	54.1
	3 to 6	17	27.9	27.9	82.0
	7 to 10	4	6.6	6.6	88.5
	11 to 14	4	6.6	6.6	95.1
	15 and more	3	4.9	4.9	100.0
	Total	61	100.0	100.0	

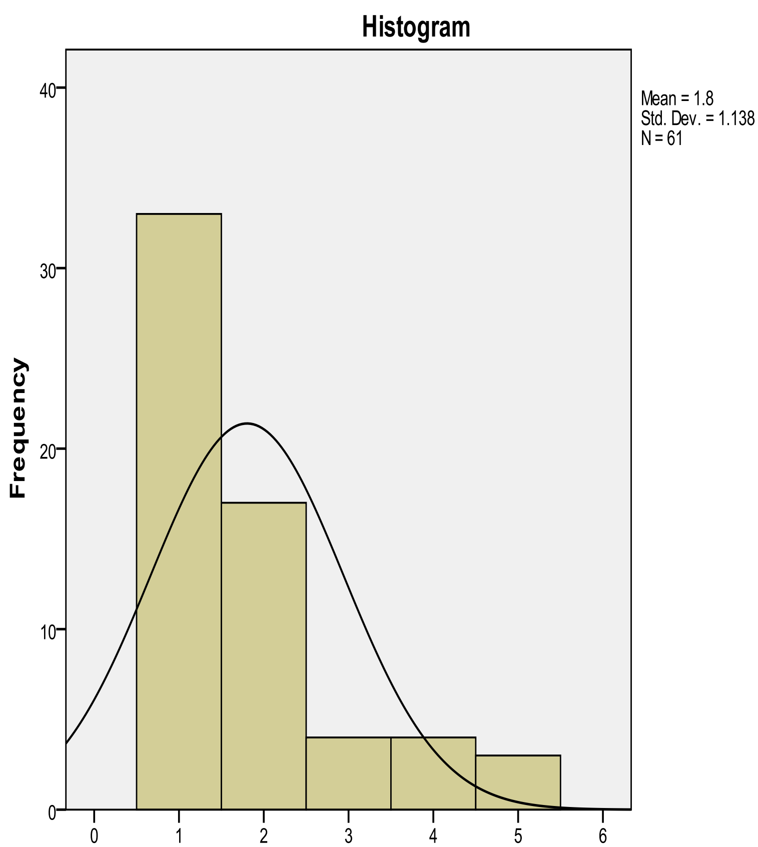


Figure 5.8 Staff's years of research experience

5.3.1.8 Highest qualification

Table 5.10 and figure 5.9 represent the distribution of the respondents' highest qualification. The sample's qualifications range from National Diploma (ND) to the Doctorate. The table shows a 32.8 % of masters' holders followed by a 24.6 % of BTech holders as the highest frequencies. This sample is composed of less than 10% (9.8 %) of Doctorate holders. It is also important to note that although BTech and BSc (Hons) qualifications both take 4 years to be completed from the end of high school, these qualifications are strictly speaking not equivalent in the South African educational system. A choice was therefore made in this research to keep the BTech and BSc (Hons) as two different types of qualifications. However, their combination into a single qualification could have yielded a different frequency distribution of staff qualifications.

Table 5.10 Staff Highest Qualification

Qualification		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ND	3	4.9	4.9	4.9
	BTech	15	24.6	24.6	29.5
	BSc	7	11.5	11.5	41.0
	BSc(Hons)	10	16.4	16.4	57.4
	Masters	20	32.8	32.8	90.2
	Doctorate	6	9.8	9.8	100.0
	Total	61	100.0	100.0	

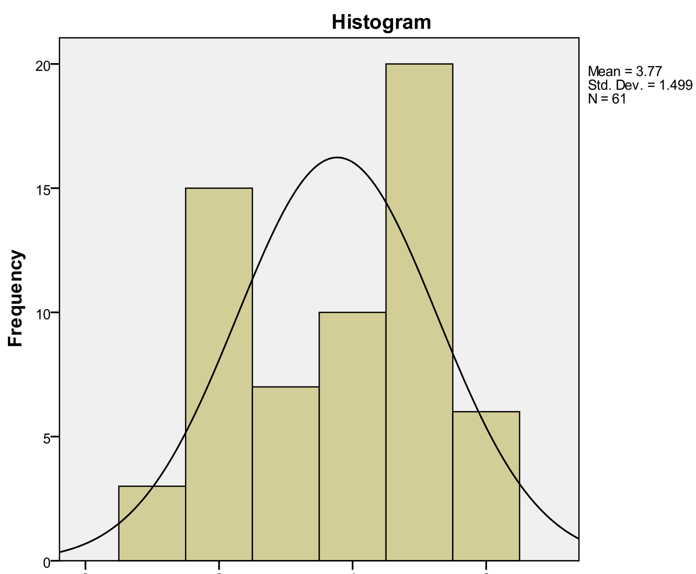


Figure 5.9 Staff highest qualification

5.3.1.9 Primary and Secondary education Language of staff

Table 5.11 and figure 5.10 represent the participants' most spoken language in their primary and secondary education. Of all seven (7) options for languages options available in the questionnaire, only four (4) were chosen by respondents namely English, French, isiZulu/isiXhosa, and Others. The histogram in figure 5.10 reveals that most of the respondents (82 %) received their primary and secondary education in English. Less than 10 % (8.2 %) received their primary and secondary education in the two most popular local languages (isiXhosa and isiZulu) of the provinces where the study took place. Other staff members (about 10 %) received their education in French and Asian languages.

Table 5.11 Primary and second education language of staff

Language		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Other	3	4.9	4.9	4.9
	French	3	4.9	4.9	9.8
	English	50	82.0	82.0	91.8
	isiZulu/isiXhosa	5	8.2	8.2	100.0
	Total	61	100.0	100.0	

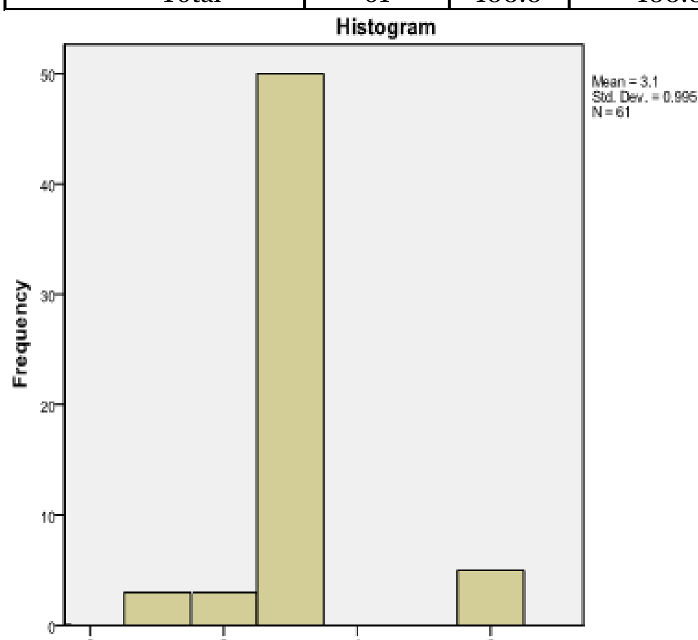


Figure 5.10 Primary and secondary education language of staff

5.3.1.10 Tertiary Education Language of staff

Table 5.12 and figure 5.11 represent the participants' tertiary education language. Of all seven (7) languages options available in the questionnaire, only three (3) were chosen by respondents namely English, French, and isiZulu/isiXhosa. 95.1 % of the respondents studied in English at the university level, and the rest studied either in French or in isiXhosa.

Table 5.12 Tertiary education language of staff

Language		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	French	2	3.3	3.3	3.3
	English	58	95.1	95.1	98.4
	isiZulu/isiXhosa	1	1.6	1.6	100.0
	Total	61	100.0	100.0	

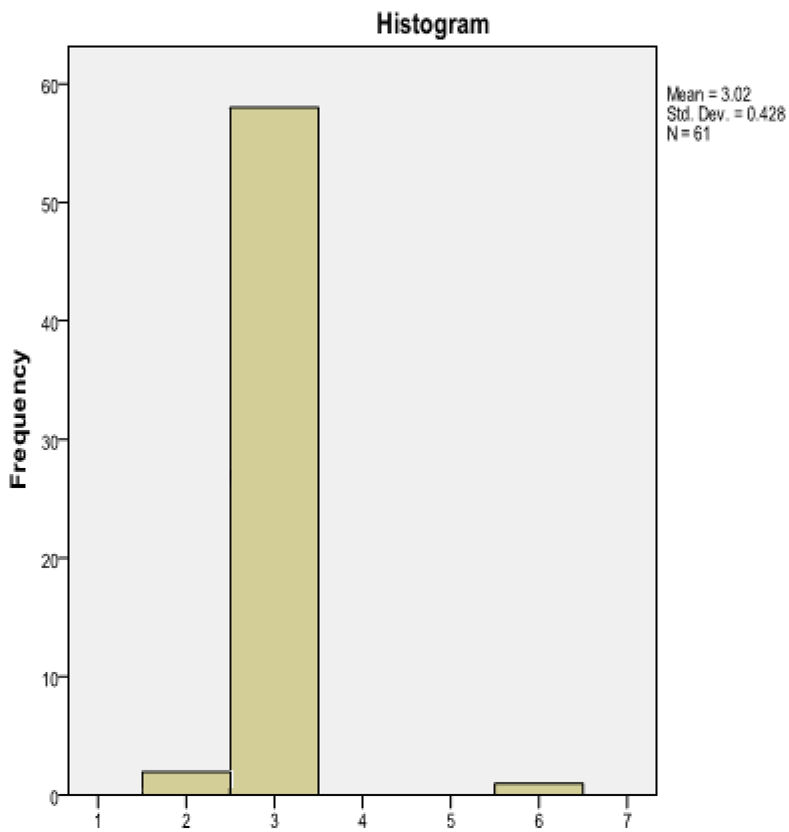


Figure 5.11 Tertiary education language of staff

5.3.2 Mean Analysis of the Likert scale based Research Variables

Table 5.13 represents calculated average for the Likert scale based research variables. It shows the mean, standard deviation, variance, minimum and maximum values for the research variables Lecturing experience, Research Experience, English Language Ability, Prior Search ability and Searching Ability.

Table 5.13 Likert scale based research variables descriptives statistic

		Lecturing experience	Research Experience	English Language Ability	Prior Search Experience	Searching Ability
N	Valid	61	61	61	61	61
	Missing	0	0	0	0	0
Mean		**11.69**	**12.23**	**21.57**	**11.00**	**23.30**
Std. Error of Mean		.351	.827	.370	.477	.537
Std. Deviation		2.742	6.456	2.889	3.724	4.197
Variance		7.518	41.680	8.349	13.867	17.611
Minimum		3	6	12	4	14
Maximum		**15**	**30**	**25**	**20**	**30**

Table 5.13 shows that the average perceived lecturing experience, English language ability and searching ability of the respondents are high (respectively 11.69 out of 15.00, 21.57 out of 25.00, and 23.30out of 30.00). On the other hand, the average perceived research experience and prior search experience are moderate (respectively 12.23 out of 30.00, and 11.00 out of 20.00). Likewise variance is low for perceived lecturing experience (7.518) and English language ability (8.349), but high for the variance searching ability (17.611), research experience (41.68), and prior searching experience (13.867). A further discussion on these statistics will be given later in the study.

As a summary, the frequency distribution gave an indication of the group size and a summary of its general distribution. A more analysis of these results is thus needed to predict the interdependencies between the variables through inferential statistics.

5.4. Inferential Statistics

This section presents the results of the inferential statistical tests described in section 4.6.3 of chapter four. It gives the results of the Pearson's correlation test, multiple regression analysis tests, 1 way ANOVA tests, and of the 5-way ANCOVA test.

5.4.1 Pearson's Correlations

For the purpose of this research, a two-tailed Pearson's correlation test was performed as described in section 4.6.2.1 of chapter four. There were many tables resulting from this test. All these tables are presented in appendix D except for the correlation table which is presented in this section (table 5.12). This correlation table displays the Pearson's correlation coefficients (ρ), significance values (p), and the number of cases with non-missing values (N) for each variable. The correlation coefficients on the main diagonal are always equal to 1, because each variable has a perfect positive linear relationship with itself. The significance value (p) of each correlation coefficient is also displayed in the correlation table. If the significance value (p) is very small (less than or equal to 0.05), then the correlation is significant, and the two variables in question are deemed to be linearly related. If p is relatively large (greater than 0.05) then the correlation in question is not significant, and the two variables are not linearly related.

Table 5.12 uses shaded cells to shows variables with a significant correlation value. A shaded cell is the result of existence of a correlation between the variable represented by a particular row and the variable represented by a particular column. Even though there are five (5) two-tailed correlations on table 5.12 (see underlined values), for the purpose of the study, the researcher is more interested in correlations between the dependent variable and four other research variables.

Table 5.12 reveals that staff's lecturing experience and their English language ability present a significant p-value when compared to searching ability. Thus each of these variables has a linear correlation against searching ability variable. The researcher also observed from table 5.12 that the correlation between English language ability and searching ability is significant at 99% (p=0.000) because its Pearson's coefficient has a double star superscript (**) while other correlations in the shaded cells are only marked with one star (*). The one star (*) marked correlations carry a level of confidence 95% while the two stars (**) marked correlations carry a level of confidence 99%.

On the other hand, the following variables do not show a linear correlation with the searching ability variable because their p-value is relatively large: Research experience, and Prior searching experience.

Table 5.12 Correlations between variables

		lecturing experience	Research experience	English language ability	Prior searching experience	searching ability
lecturing experience	Pearson Correlation	1				
	Sig. (2-tailed)					
	N	61				
Research experience	Pearson Correlation	.240	1			
	Sig. (2-tailed)	.062				
	N	61	61			
English language ability	Pearson Correlation	**.288***	.071	1		
	Sig. (2-tailed)	**.024**	.584			
	N	**61**	61	61		
Prior searching experience	Pearson Correlation	-.106	**-.277***	**-.301***	1	
	Sig. (2-tailed)	.416	**.031**	**.019**		
	N	61	**61**	**61**	61	
searching ability	Pearson Correlation	**.299***	.090	**.376****	-.135	1
	Sig. (2-tailed)	**.019**	.489	**.003**	.298	
	N	**61**	61	**61**	61	61

*. Correlation is significant at the 0.05 level (2-tailed)

**. Correlation is significant at the 0.01 level (2-tailed)

It is also worth noting from table 5.12 that there is a correlation between research experience and prior searching experience, between English language ability and prior searching experience, and between Lecturing experience and English language ability.

5.4.2 Regression Analysis

There are many tables resulting from the regression analysis tests performed in this study (see appendix D). However only the Model Summary table (Table 5.14) and the coefficient table (Table 5.15) are presented in this study because the analysis of the Model Summary table allows one to measure significance of the regression analysis test and the coefficient table identifies research variables linked to the dependent variable through a regression correlation equation of the type $Y = B\,X + Constant$.

Table 5.14 Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.568[a]	.343	.123	3.929

a. Predictors: (Constant), Prior_Searching_Exp_Total_Scores, Lecturing_Exp_Total_Scores, Research_Exp_Total_Scores, English_Language_Abil_Total_Scores

b. Dependent Variable: Searching_Ability_Total_Scores_DV

The Model summary in Table 5.14 shows that the regression model is significant because its R value is greater than 0.500 (or $R^2>0.250$). In the coefficient table shown by table 5.15, the model reveals that only the English language ability (*English_Ability_Total_Score variable*) is a predictor directly influencing the searching ability of staff members as its p-value is 0.022 ($p<0.050$). Its row is shaded to show its significance. This table also reveals that the linear regression equation is positive because the slope (B coefficient) is +0.452. Therefore the regression equation between these two variables could be written as:

Searching_Ability_Total_Scores_DV = *0.452 * English_Ability_Total_Score + 9.967*   *equation 5.1*

Table 5.15 Regression's Coefficient table

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig. (p)
		B	Std. Error	Beta		
1	(Constant)	9.967	5.019		1.986	.052
	Lecturing_Exp_Total_Scores	.313	.199	.204	1.574	.121
	Research_Exp_Total_Scores	.009	.084	.014	.112	.911
	English_Language_Abil_Total_Scores	.452	.192	.311	2.359	.022
	Prior_Searching_Exp_Total_Scores	-.018	.148	-.016	-.123	.903

a. Dependent Variable: Searching_Ability_Total_Scores_DV

5.4.3 1-way ANOVA

Each of the demographic background data items in section A of the questionnaire could form a predicting factor of searching ability. This is checked by a one way ANOVA analysis whose results are hereby presented based on the methodology described by section 4.6.2.3 of chapter four. The ANOVA analysis results to a number of tables and graphs presented in appendix D but this chapter only presents the results of the Levene's tests for equality of variance and of the tests between subjects' effect.

Table 5.16 represents the Levene's Test of Equality of Error Variances, and it reveals that the test is significant because its p value is 0.015 (Sig. = 0.015<0.05). This shows that the model used is deemed valid for ANOVA analysis. Table 5.17 also shows that out of the ten (10) items, only five (5) namely position (0.003), years of research experience (0.005), highest qualification (0.022), primary and second language (0.001), and tertiary language (0.013) have a p-value less than 0.05.

Table 5.16 Levene's Test of Equality of Error Variances

F	df1	df2	Sig.
3.406	4	56	.015

Table 5.17 Tests of Between-Subjects Effects

Dependent Variable:Searching_Ability_Total_Scores_DV

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	789.862[a]	33	23.935	2.422	.010
Intercept	727.640	1	727.640	73.629	.000
Academic_Department	13.174	1	13.174	1.333	.258
Gender	5.680	1	5.680	.575	.455
Age_Group	26.399	4	6.600	.668	.620
Position	259.544	6	43.257	4.377	.003
University_Affiliation	71.366	3	23.789	2.407	.089
Yrs_Of_Experience_Lecturing	43.557	4	10.889	1.102	.376
Yrs_Of_Experience_Research	189.545	4	47.386	4.795	.005
Highest_Qualification	156.286	5	31.257	3.163	.022
Primary_Second_Language	201.588	3	67.196	6.800	.001
Tertiary_Language	101.296	2	50.648	5.125	.013
Error	266.827	27	9.882		
Total	34159.000	61			
Corrected Total	1056.689	60			

a. R Squared = .747 (Adjusted R Squared = .439)

5.4.4 5-way ANCOVA

The Pearson's correlation analysis of the searching ability of respondents revealed a correlation between lecturing experience (*Lecturing_Exp_Total_Score variable)* and searching ability and between English language ability (*English_Ability_Total_Score variable* at 99% Confidence Interval) and searching ability. The regression tests also confirmed that English language ability variable is a predictor of staff searching ability; therefore there is a need to further analyze differences in searching abilities, lecturing experience, and English language ability within and between the different groups for each of the observed five (5) fixed factors.

Table 5.18 5 way ANCOVA Test Between Subjects Effect

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	811.618[a]	36	22.545	2.208	.022
Intercept	66.072	1	66.072	6.471	.018
Lecturing_Exp_Total_Scores	1.375	1	1.375	.135	.717
English_Language_Abil_Total_Scores	45.480	1	45.480	4.454	.045
Position	188.744	4	47.186	4.621	.007
Yrs_Of_Experience_Research	32.879	3	10.960	1.073	.379
Highest_Qualification	81.489	5	16.298	1.596	.199
Primary_Second_Language	61.347	2	30.674	3.004	.069
Tertiary_Language	73.163	1	73.163	7.165	.013
Position * Yrs_Of_Experience_Research	.382	1	.382	.037	.848
Position * Highest_Qualification	20.914	2	10.457	1.024	.374
Position * Primary_Second_Language	.206	1	.206	.020	.888
Position * Tertiary_Language	.000	0	.	.	.
Yrs_Of_Experience_Research * Highest_Qualification	1.327	1	1.327	.130	.722
Yrs_Of_Experience_Research * Primary_Second_Language	.000	0	.	.	.
Yrs_Of_Experience_Research * Tertiary_Language	.000	0	.	.	.
Highest_Qualification * Primary_Second_Language	.000	0	.	.	.
Highest_Qualification * Tertiary_Language	.000	0	.	.	.
Primary_Second_Language * Tertiary_Language	.000	0	.	.	.
Position * Yrs_Of_Experience_Research * Highest_Qualification	.000	0	.	.	.

Position * Yrs_Of_Experience_Research * Primary_Second_Language	.000	0	.	.	.	
Position * Yrs_Of_Experience_Research * Tertiary_Language	.000	0	.	.	.	
Position * Highest_Qualification * Primary_Second_Language	.000	0	.	.	.	
Position * Highest_Qualification * Tertiary_Language	.000	0	.	.	.	
Position * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.	
Yrs_Of_Experience_Research * Highest_Qualification * Primary_Second_Language	.000	0	.	.	.	
Yrs_Of_Experience_Research * Highest_Qualification * Tertiary_Language	.000	0	.	.	.	
Yrs_Of_Experience_Research * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.	
Highest_Qualification * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.	
Position * Yrs_Of_Experience_Research * Highest_Qualification * Primary_Second_Language	.000	0	.	.	.	
Position * Yrs_Of_Experience_Research * Highest_Qualification * Tertiary_Language	.000	0	.	.	.	
Position * Yrs_Of_Experience_Research * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.	
Position * Highest_Qualification * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.	

Yrs_Of_Experience_Research * Highest_Qualification * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.
Position * Yrs_Of_Experience_Research * Highest_Qualification * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.
Error	245.071	24	10.211		
Total	34159.000	61			
Corrected Total	1056.689	60			

a. R Squared = .768 (Adjusted R Squared = .420)

Table 5.18 presents the PASW STATISTICS 18.0 (SPSS) result of the analysis between subject and their combined effect to predict the change on the dependent variable. This table shows at the bottom that R Square value is 0.768 which is greater than 0.500 therefore the model used is deemed valid for ANOVA analysis. Table 5.17 also shows that out of the combination effect of the model, only the position (0.007), tertiary language (0.013) and the English language ability variable (0.045), are significant. Thus the table 5.18 reveals that after the ANCOVA test, taking into account the combined effect of demographic data and of Likert-scale based co-variables, searching ability of academic staff can be predicted by their position, the language of their tertiary education and by their English language ability.

5.4.4 Conclusion

This chapter presented the research results obtained from the different tests performed in chapter four, starting from the testing of the reliability and validity of the questionnaire. The chapter thereafter presents the results of the descriptive statistics frequencies for the biographical data of the questionnaire an indication of the characteristics of the research population such as gender equity, average qualification of staff from Information Technology as compared to those in Computer Science, a majority of under qualified middle age male staff in junior positions with considerable years of lecturing experience but with little research experience. In addition, the descriptive statistics analysis revealed that the participants' perceived lecturing experience, English language proficiency and searching abilities are as high, but their prior searching experience is perceived as slow.

The chapter also presented results of the following inferential statistics tests: Pearson's correlation analysis, multiple regressions, one way ANOVA, and five way ANCOVA. These tests results reveal an association between academic staff searching ability and their biographical attributes such as highest academic qualifications, positions, and years of research experience. The findings also show a direct relationship between searching ability and lecturing experience, and between searching ability and English language proficiency. Ultimately, the ANCOVA test proves that position, English language skills are predictors of searching ability. The next chapter will discuss and give recommendations from these research findings.

CHAPTER SIX

Discussion, Recommendations, and Future Research

6.1 Introduction

This chapter discusses the research results presented in the previous chapter. It aims at presenting the major findings of this study in comparison with results from other previous similar studies from literature. The chapter also presents some recommendations and future research ideas emanating from the results of discussion. These results are categorised into descriptive results and inferential results. Five recommendations are outlined together with the possible emanating future works.

6.2 Descriptive results

6.2.1 Search Experience

The analysis of this research survey's results shows that the prior searching experience of academic staff is average (see table 5.13). Existing literature also reports that users' search experience is usually associated to their familiarity with a domain (White et al., 2005) and to the number of terms used in query formulation (Jansen et al., 2001). Other previous studies (Fenichel, 1981; Hsieh-yee, 1993; Bates et al., 1993; Spink et al.; 1997) reveal that the mean number of search terms in query formulation as being higher for experienced searchers and relatively low for novice users. In line with these findings, studies by Klein (1998) and Shim et al. (2001), and by Kim et al. (2004) also found that previous search experience with internet shopping is a good predictor of online search behaviour.

<u>Recommendation One and Ideas for future research</u>: The above discussion suggests the need to encourage research among academic staff so as to improve their searching experience based on the assumption that searching is a crucial aspect of research. This can be achieved by encouraging the mentoring of inexperienced academic staff by experienced researchers as a way to share online searching skills among academic staff. However, it will be interesting to conduct future research on the assessment of this form of mentoring on the searching abilities of academic staff.

6.2.2 Search Ability

The analysis of this research survey results shows that the perceived searching ability of academic staff is high (see table 5.13). Contrary to these findings, existing literature reveals that the overall success in the online searching of information was low (Nachmias et al., 2001). Nachmias et al. (2001) used fifty four (54) masters' students who were asked to accomplish three (relatively simple) search tasks experiments. The outcomes of this experiment revealed that the overall success in searching information was low, and only about 15% of the students succeed in all three tasks. The findings of the current research and Nachmias et al. (2001)'s results are different due to many factors. Firstly the type of research participants (academic lecturers) used in this current research and those (students) used by Nachmias et al. (2001) are different; secondly, academic staff in this current research had to simply complete a survey questionnaire while students from Nachmias et al.(2001) executed three experimental tasks which might had contributed in their low overall success.

<u>Recommendation two and Ideas for future research</u>: The above contradiction between the findings of Nachmias et al. (2001) and the result from this current research calls for the conduct of a new research project able to measure the searching ability of academic staff not through a questionnaire but using experimental online searching tasks.

6.3 Inferential results

6.3.1 Biographical Profiles and Searching Ability

According to the findings of the current research, the perceived searching ability of academic staff is affected by each of the following biographical factors: staff position, their years of research experience, their highest qualification, primary and secondary education language, and tertiary education language. This result is however weakened by the choice of its research sample and its generalisation cannot be inferred. On the other hand, according to literature, biographical profile attributes mostly examined while measuring users searching abilities include users academic levels (Allen, 1992), gender (Jackson, 2001; Large et al., 2002), and age (Lee, 1999; Stronge et al., 2006).

In Allen (1992), participants were made up of fifty (50) university students whose reading abilities were measured using eight different tests and who were requested to retrieve from the web a document that they just read. These students were categorized into senior and junior students depending on their academic levels of study. Allen (1992) found that the perceptions of the kind of citations the senior students would want for term papers differed from the perceptions of junior students, and most useful citations were retrieved by senior students. Thus, similar to Allen (1992)'s research, this current research shows that highest qualifications influence searching ability. It is also important to note that these results are similar although their methodologies are different.

According to Jackson (2001) and Large et al., (2002) the gender of the user has an influence on their searching ability. Jackson (2001) used 630 American undergraduate (403 males and 227 females) students to fill in a questionnaire related to the use of different internet services including email and use of the web. The outcome showed that females use e-mails more than males and males use the web for other search and retrieval activities more than females. Results from this current research do not show gender as a significant predictor of searching ability although it uses a research methodology similar to Jackson (2001) (questionnaire). These differences in findings seem to exist despite the type of participants (academic staff versus undergraduate students), and the similar male/female ratios in the two studies (around 60% of males in both studies).

Large et al. (2002) also studied gender differences in collaborative web search strategies using sixth grade students organized into same gender groups. In their research, children were requested to search for information on the web to support a classroom assignment. Their research data was collected using a case study approach and children search activities were captured on videotape for subsequent analysis. It was found that groups of male are more active on the web than females: meaning that male's level of interaction is higher than their female counterparts. The outcome from Large et al. (2002) research differs from that of this current research probably because completely different methodologies were used.

Lee (1999) and Stronge et al., (2006) show that age influences search ability but not for advanced search. The latter research, based of 16 young and 16 old users revealed that older adults were less successful than young adults in finding correct answers to a given search task. On the contrary, there is no age related difference on advanced search strategies although younger adults use more advanced keyword search strategies than older ones (Stronge et al., 2006). The difference observed between Stronge et al. (2006) and the current research is on the number of participants' age groups (two age groups versus five age groups).

In addition to the biographical profile attributes above highlighted by literature, this current research has found that years of research experience, staff position, primary and secondary education language, and tertiary education language influence searching ability. However, position and years of research experience seem related to the academic level examined by Allen (1992). It is important to note that in this current research, the position factor was the only factor besides the English language ability factor that was found to impact on searching abilities even when considering the effect of co-variants. The impact of language on searching ability is not discussed in this section as it will be discussed towards the end of this chapter while the research experience factor was already discussed above.

<u>Recommendation Three and Ideas for future research</u>: The above discussion on the impact of academic qualifications, levels, and positions on searching ability suggests that academic training involves practices favourable to the enhancement of web searching skills. It is therefore recommended to enrol novice online searchers into web searching training programs. The designs of such training programs as well as the evaluation of their efficiency are possible ideas for future research.

6.3.2 Lecturing Experience and Searching Ability

According to the findings of this current research, lecturing experience affects the perceived searching ability of academic staff. This is a new finding compared to existing literature as a review of existing literature does not present enough findings on the direct correlation between academic staff lecturing experience and their searching ability. However, it is important to note that research experience, highest qualification and language were captured as biographical profile data in this current research while some other researcher captures them as independent variables that can correlate with the dependent variables.

<u>Recommendation Four and Ideas for future research</u>: The above discussion suggests the need to encourage the mentoring of inexperienced lecturers by more experienced ones. It will be interesting for future research to look at the format in which this type of mentoring can place so as to cover all the important lecturing duties that can affect web searching abilities.

6.3.3 English Language Ability and Searching Ability

The current research findings suggest that English language ability has an effect on the perceived searching ability of academic staff. The content of a particular web page is usually written in a particular language. According to Bilal (2000), reading and writing ability of a user influence his or her search success but reading ability alone does not significantly influence searching ability. Furthermore, research by Allen (1992) shows that when combined with logical reasoning, reading and writing in a particular language usually predicts the search success behaviour. Another verbal ability research done by Gugerty et al. (2006) on the analysis of internet search accuracy and speed, found that verbal abilities or ability to speak a language influences internet search performance. These findings by Gugerty et al. (2006) were also obtained by analyzing questionnaire items on English language ability.

The current research therefore confirms results from Allen (1992), Bilal (2000), and Gugerty et al. (2006) by supporting that searching ability is influenced by their English language ability. It is important to recall that in this current study English language ability was found to affect searching ability first as a biographical factor, then as a Likert scale based research variable with a 99% level of confident. It is also important to note that in this current research, the English language ability factor was the only factor besides the position factor that was found to impact on searching abilities even when considering the effect of co-variants.

Recommendation Five and Ideas for future research: The above discussion calls for the need for the improvement of the design and use of semantic search engines able to understand users' queries even when these users do not master the command of human languages such as English and other languages used on the Internet. It is also recommended to publish more web content in local languages and to design web engines interfaces in these languages; not only in the English language at it seems to be the case now. The study of language usability remains a wide avenue for future research not only for web search engines but also for other information systems considering the diversity of natural languages used worldwide.

6.5 Conclusion

This chapter presented a discussion of the findings of this current research and comparing them with the previously related research study. It also suggested appropriate recommendations and future research ideas from these discussions. The chapter highlighted mostly similarities, a novelty, and a contradiction in findings of the current research and existing literature. As far as the literature surveyed during the conduct of this research, a novel finding is the effect of lecturing experience on searching ability and the contradiction was on the searching success rates which is low in most research and seem to be high in the current study. Ideas for future research suggested by this chapter include mentoring of academic staff by more experienced staff in order to transfer searching skills, training of novice web searchers, the design and use semantic search systems both in English and in local languages, publishing more web content in local languages.

CHAPTER 7

Summary

7.1 Introduction

This is a chapter by chapter summary of the most important aspects of this research dissertation. It summarises content from the existing body of knowledge on the use and design of web search engines within the scope delimited by the objectives of this study, while taking into account new contributions made by this research. The focus of this chapter is on literature reviews, research methods, and on research findings and recommendations.

7.2 Summary of Chapter One

The Introductory chapter of this dissertation starts with the presentation of some background on the evolution of use of computers from the era of mainframes after World War II to the advent of Internet and the World Wide Web in the 1990's after the introduction of personal computers (PC's) in the 1980's. The growth of Internet is then described in terms of its plethora of applications such as emails, social networks, static and dynamic websites and in terms of its total number of users of more than 200 millions. This chapter then turns its attention to the document access problems posed by this web traffic where users usually feed keywords to syntactic web search engines for the retrieval of documents needed by them for a variety of activities such as learning, teaching, research, business, entertainment, etc. A description by the introductory chapter of the objectives and research questions of this study reveals that this research aims at the analysis of the factors affecting the online searching of academic staff from the computing field, with the rationale that the Web is a good tool for this category of professionals to fulfil its need for permanent update of skills and knowledge. Chapter one also briefly describes how the research objectives will be achieved using a survey of 61 academic staff from Information Technology and Computer Science departments from four Universities from the Kwazulu-Natal and Eastern Cape provinces of South Africa. The choice of these four universities is also identified by the first chapter as a limitation for this research as well as its perceptions analysis nature.

7.3 Summary of Chapter two

Chapter two starts with the definition of search engines borrowed from existing literature as software serving web documents containing keywords input by users. A brief presentation of the history of the Web reveals a succession of web search engines deployed between 1994 and today: World Wide Web Worm, *Lycos*, Excite, Google, AltaVista, Infoseek, Msn, Yahoo, etc. The second chapter also gives taxonomy of web search engines in terms of the types of engines including spider based search engines, directory based search engines, and linked based search engines; and in terms of the classification of these engines by topic, by model, and by information type. Chapter two also contrasts semantic search as opposed to syntactic search by explaining how semantic search takes advantage of the construction of ontology of knowledge that can be understood by intelligent software based on the description of relationships existing between objects meanings. Chapter two ends with a description of the architecture of typical web search engines components such as user interfaces, searchers, indexers, crawlers, local stores, etc.

7.4 Summary of Chapter Three

The third chapter is a review of existing literature on factors affecting online search ability. This chapter does not only report on findings on these factors but it also describes methodologies used by researchers to support their findings. These methodologies include participants' surveys for various demographic groups (e.g. children, students, young people, old people, academics, health practitioners, etc), literature surveys, experimental searching tasks, and analysis of web transaction logs. Findings from chapter three indicate searching hit rates are generally low and that searching ability depends on demographic attributes such as academic levels, gender, age, and disability. It is also reported that searching ability is affected by cognitive factors such familiarity with the search domain, user knowledge and web experience, and language related factors such as reading ability and verbal reasoning. Web trends are also presented in chapter three where it is reported that query terms are usually short, that there is a small number of keywords with a high frequency in query terms, that searching sessions are quick and brief in duration, and that multitasking and multi-topics searching is only practised by advanced users, therefore highlighting the current "quick and dirty" approach of the Web instead of it being interactive and exploratory. Users neglect of web engines feedback and help is also reported by the third chapter but it stresses that the use of feedback leads more queries refinement.

7.5 Summary of Chapter Four

Chapter four describes the questionnaire based survey used by this study to reach its research objectives. It is a survey of 61 academic staff selected from Information Technology and Computer Science departments from four universities from the Kwazulu-Natal and Eastern Cape provinces of South Africa using a systematic sampling method. These participants were requested to fill a questionnaire whose first section contained nominal and ordinal demographic items such as age, gender, academic qualifications, education language, years of lecturing experience, etc. The other sections of the questionnaire were all Likert-scale based with only one section on searching ability, the dependent variable, and three sections on the following four independent variables: research experience, lecturing experience, English language proficiency, and web searching experience. Descriptive and inferential statistics were computed for data analysis after performing data reliability and validity tests using factor analysis and Cronbach's coefficients methods. These descriptive statistics included means and proportions analysis while inferential statistics included Pearson's correlations analysis, regression analysis, ANOVA, and ANCOVA.

7.6 Summary of Chapter Five

Chapter five is dedicated to the presentation of the research results obtained by this research. It confirms the reliability and validity of the research data as checked by the reliability and validity tests described in the chapter four. Descriptive statistics gathered during data analysis revealed a majority of staff from Information Technology as compared to Computer Science, a majority of under qualified middle age male staff in junior positions with considerable years of lecturing experience but with little research experience. These participants perceive their lecturing experience and their English language and searching abilities as high, but their prior searching experience is perceived as slow. Inferential statistics gathered by this study show an association between searching ability and demographic attributes such as academic qualifications, positions, and years of research experience. They also show a relationship between searching ability and lecturing experience, and between searching ability and English language ability. However, the association between position, English language ability, and searching ability was found to be the strongest of all.

7.7 Summary of Chapter six

Chapter six discusses the findings of this research compared to existing literature and suggests appropriate recommendations from these discussions as well as ideas for future research. It is reported that most of the findings of this research are supported by existing literature except for the one on searching success rates and the one on the effect of gender and age on searching ability. Chapter six also mentions the novelty of a finding by this study on the effect of lecturing experience on searching ability. Ideas for future research and recommendations suggested by the sixth chapter include mentoring of academic staff by more experienced staff, training of novice web searchers, designing and using semantic search systems both in English and in local languages, publishing more web content in local languages, and triangulating various research strategies for the analysis of the usability of web search engines.

7.8 Conclusion

This study adds evidence to the existing body of knowledge in support of the claim that web searching ability is affected by biographical factors, language factors, and by cognitive factors related to the expertise and experience in the domain in which searching is taking place. The limitations of this study are mainly related to its perceptions analysis nature. However, most of its research findings are supported by existing literature, and plausible explanations can be made when they are not. The novelty of this study resides in the choice of its research sample and on the identification of new factors and recommendations on web searching ability.

References

Adda M., Missaoui R., Valtchev P. (2009). *Toward Feedback-Based Web Search Engine*. International Conference on Advanced Information Networking and Applications Workshops. IEEE computer Society. USA. (2009), pp. 756-761.

Allen B. (1992). *Cognitive differences in End User Searching of a CD-ROM Index*. 15th Ann. Int'1 Special Interest Group in Information Retrieval (SIGIR), 92/Denmark-6/92.

Anick P. (2003). *Using Terminological Feedback for Web Search Refinement - A Log-based Study*. ACM Special Interest Group in Information Retrieval (SIGIR) 2003, July 28–August 1, Toronto, Canada.

Ash J. (1997). *Organizational Factors that Influence Information Technology Diffusion in Academic Health Sciences Centers*. Journal of the American Medical Informatics Association Volume 4 Number 2 Mar /Apr 1997.

Baeza-Yates, R. (2003). *Information retrieval in the Web: beyond current search engines*. International Journal of Approximate Reasoning 34 (2003) 97–104.

Baeza-Yates, R., Ribeiro-Neto B. (1999). *Modern Information Retrieval*. Addison-Wesley, England, 1999.

Bates M. J. (2007). *Factors affecting subject catalog search success*. Journal of the American Society for Information Science and Technology Volume 28 Issue 3, Pages 161 – 169 Published Online: 22 Mar 2010

Bergman M.K. (2001). *The Deep Web: Surfacing Hidden Value*. White paper available online at http://beta.brightplanet.com/deepcontent/tutorials/DeepWeb/index.asp [last accessed 1 December 2010]

Berners-Lee T. (1996). *The World Wide Web: Past, Present and Future*. IEEE Computer special issue of October 1996. Available on line at http://www.w3.org/People/Berners-Lee/1996/ppf.html. Last accessed July 9, 2009

Berners-Lee T. (1999). *Waving the Web*. Orion Business Books, London, 1999.

Bilal D. (2000). *Children's Use of the Yahooligans! Web Search Engine: Cognitive, Physical, and Affective Behaviors on Fact-Based Search Tasks*. Journal Of The American Society For Information Science. 51(7):646–665, 2000.

Birchal, J. (2005). *Sampling and Samples*. Retrieved 06 09, 2009, from Market Research World: available on http://www.marketresearchworld.net/index.php?option=com_content&task=view&id=23&Itemid=41&limit=1&limitstart=2

Budzik J., Hammond K.J. (2000). *User interactions with everyday applications as context for just-in-time information access*. In Proceedings of the 2000 International Conference on Intelligent User Interfaces, New Orleans, Louisiana, 2000. ACM Press.

Callan J. (2000). *Searching for Needles in a World of Haystacks*. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, September 2000. Vol. 23 No. 3, pp 33-37

Chakrabarti S., Van den Berg M., Dom B. (1999). *Focused crawling: A new approach to topic-specific web resource discovery*. In Proceedings of The Eighth International World Wide Web Conference (WWW-99), Toronto, 1999.

Cho J., Garcia-Molina H., Page L. (1998). *Efficient crawling through URL ordering*. In Proceedings of the 7th World Wide Web Conference (WWW7), Brisbane, Australia, April 1998.

Cohen W., McCallumx A., Quass D. (2000). *Learning to Understand the Web*. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, September 2000 Vol. 23 No. 3. Pp. 17-24

Dembo M. H., Gibson S. (1985). *Teachers' Sense of Efficacy: An Important Factor in School Improvement*. The Elementary School Journal, Vol. 86, No. 2 (Nov. 1985), pp. 173-184.

Fattahi R., Wilson C. S., Cole F. (2007). *An alternative approach to natural language query expansion in search engines: Text analysis of non-topical terms in Web documents*. Information Processing and Management Vol. 44 (2008), pp. 1503–1516.

Fenichel, Hansen C. (1981). *Online Searching: Measures that discriminate among users with different types of experience*. Journal of the American Society of Information Science. Vol. 32, Issue 1 (1981), pp. 23-32.

Fensel D., Hendler J., Lieberman H., Wahlster W. (2002). *Creating of Semantic Web*. Available on line http://www.cs.vu.nl/~dieter/ftp/paper/mit.introduction.pdf, Last accessed July 9, 2009

Field A. (2005). *Discovering Statistics Using SPSS (Introducing Statistical Methods)*. 2nd Edition, Sage Publications Inc. CA. ISBN 978-1-84787-906-6

Giunchiglia F., Kharkevich U., Zaihrayeu I. (2008). *Concept Search: Semantics Enabled Syntactic Search*. SemSearch 2008, CEUR Workshop Proceedings, ISSN 1613-0073, online at CEUR-WS.org/Vol-334.

Glover E.J., Lawrence S., Gordon M. D., Birmingham W. P., Lee Giles C. (2000). *Web Search – Your Way*. Communications of the ACM on Web Information and Data Management 2000. Vol. 59, No. 2, (2000), pp 34-39.

Griffiths J. R., Brophy P. (2005). *Student Searching Behavior and the Web: Use of Academic Resources and Google*. Library trends, Vol. 53, No. 4, (*"The Commercialized Web: Challenges for Libraries and Democracy," edited by Bettina Fabos, pp. 539–554),* University of Illinois. USA.

Gruber T.R. (1993). *A translation approach to portable ontology specifications*. Knowledge Acquisition, (pp 199-220). Standford University, California. Available online at http://tomgruber.org/writing/ontolingua-kaj-1993.pdf, Last accessed July 9, 2009

Gugerty L., Treadaway C., Rubinstein J. S. (2006). *Individual Differences in Internet Search Outcomes and Processes*. *CHI 2006,* April 22–27, 2006, Montréal, Québec, Canada. ACM 1-59593-298-4/06/0004.

Hearst N. A. (2001). *Next Generation Web Search: Setting Our Sites*. In proceedings of the IEEE 17th International Conference on Data Engineering Heidelberg, Germany. Vol. 23 No. 3, (September 2001), pp. 38-42.

Hersh W. R., Crabtree M. K., Hickam D. H., Sacherek L., Rose L., Friedman C.P. (2000). *Factors associated with successful answering of clinical questions using an information retrieval system*. Bulletin Medical Library Association Vol. 88 No. 4 (2000).

Höchstötter N., Lewandowski D. (2009). *What users see – Structures in search engine results pages*. Journal of Information Sciences Vol. 179 (2009) pp. 1796–1812.

Hwang G. J., Tsai P. S., Tsai C.C., Tseng J.C.R. (2008). *A novel approach for assisting teachers in analyzing student web-searching behaviors*. Computers & Education Vol. 51, (2008), pp. 926–938.

Jackson L.A., Ervin K. S., Gardner P. D., Schmitt N. (2001). *Gender and the Internet: Women communicating and men searching*. Sex Roles, Vol. 44, No 5/6, 2001. Academic Research Library. Michigan University, USA.

Jansen B. J., Pooch U. (2001). *A review of Web searching studies and a framework for future research*. Journal of the American Society for Information Science and Technology, Vol. 52, No. 3, (2001), pp 49-54

Jansen B. J., Spink A., Saracevic T. (2000). *Real life, real users, and real needs: a study and analysis of user queries on the web*. Information Processing and Management Vol. 36 (2000), pp. 207-227.

Jansen B.J., Spink A. (2006). *How are we searching the World Wide Web? A comparison of nine search engine transaction logs*. Information Processing and Management. Vol. 42, (2006), pp. 248–263.

Jansen B.J., Spink A., Pedersen J. (2005). *A Temporal Comparison of AltaVista Web Searching*. Journal Of The American Society For Information Science And Technology. Vol. 56, No. 6, (2005) pp. 559–570.

Kelley K., Clark B., Brown V., Sitzia J. (2003). *Good practice in the conduct and reporting of survey research*. International Journal for Quality in Health Care. Vol. 15, No. 3, (2003), pp. 261-266.

Kim J., Lee H.C., Kim H.J. (2004). *Factors Affecting Online Search Intention and Online Purchase Intention*. Seoul Journal of Business. Vol. 10, No. 2, (2004), pp. 127-139.

Kontopoulos E., Vrakas D., Kokkoras F., Bassiliades N., Vlahavas I. (2008). *An ontology-based planning system for e-course generation*. Expert Systems with Applications. Vol. 35, (2008), pp. 398–406.

Koshma S., Spink A., Jansen B. J. (2006). *Web Searching on the Vivisimo Search Engine*. Journal Of The American Society For Information Science And Technology. Vol. 57, No. 14, (2006), pp. 1875–1887.

Kules B., Shneiderman B. (2008). *Users can change their web search tactics: Design guidelines for categorized overviews*. Information Processing and Management. Vol. 44, (2008), pp. 463–484.

Langville N.A., Meyer, C. D. (2006). *Google's PageRank and Beyond*. Princeton University Press. US.

Large A., Beheshti J., Rahman T. (2002). *Gender differences in collaboration Web searching behavior: An elementary school study*. Information Processing and Management. Vol. 38, (2002), pp. 427-443.

Lawrence S. (2000). *Context in Web Search*. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering. Vol. 23, No. 3, (2000), pp. 25-32.

Lazonder W., Biemans Harm J.A., Wopereis I.G.J.H. (2000). *Differences between Novice and Experienced Users in Searching Information on the World Wide Web*. Journal of the American society for Information Science. Vol. 51, No. 6, (2000), pp. 576–581.

Lee A. T. (1999). *Web Usability A Review of the Research*. SIGCHI Bulletin. Vol. 31, No. 1, (1999), pp. 38-40.

Liaw S.S., Huang H. M. (2003). *An investigation of user attitudes toward search engines as an information retrieval tool*. Computers in Human Behavior. Vol. 19, (2003), pp. 751–765.

Martzoukou K. (2004). *A review of Web information seeking research: considerations of method and foci of interest*. *Information Research*. Vol. 10, No. 2, (2004), paper 215. [Available at http://InformationR.net/ir/10-2/paper215.html]

Nachmias R., Gilad A. (2001). "*Needle in a Hyperstack: Searching Information on the World Wide Web*". Journal of Research on Computing in Education 2001.

Nahl D. (1998). *Learning the Internet and the Structure of Information Behavior*. Journal of the American Society of Information Science. Vol. 49, Issue 11 (1998), pp. 1017-1023.

Palmquist, R.A., Kyung-Sun K. (2000). *Cognitive Style and On-line Database search experience as Predictors of Web Search Performance*. Journal of the American Society of Information Science. Vol. 51, Issue 6 (2000), pp. 558-566.

PASW Statistics 18, *Release Version 18.0.0* (SPSS, Inc., 2009, Chicago,IL, www.spss.com)

Risvik K. M., Michelsen R. (2002)."Search engines and Web dynamics". Computer Networks Vol. 39, (2002), pp. 289–302.

Sevindik T., Kay K., Ünlükahraman O. (2010). *Semantic web and its reflections in distance education*. Procedia Social and Behavioral Sciences Vol. 2, (2010), pp. 5109–5113.

Shneiderman, B., Byrd D., Croft, B.W. (1997). *Clarifying search: A user Interface framework for Text searches*. D-lab magazine online www.dlib.org.

Sillivan D. (2003).*How to search engine rank pages*. ACM Special Interest Group in Information Retrieval (SIGIR) 2003, Toronto, Canada.

Spink A. (2003). *Web Search: Emerging Patterns*. IEEE, Library Trends, Vol.52, No. 2, (2003), pp. 299-306.

Spink A., Park M., Jansen B.J., Pedersen J. (2006). *Multitasking during Web search sessions*. Information Processing and Management. Vol. 42, (2006), pp. 264–275.

Spink A., Wolfram D., Jansen B. J., Saracevic T. (2001). *Searching the Web: The Public and Their Queries*. Journal Of The American Society For Information Science And Technology. Vol. 52, No. 3, (2001), pp. 226-234.

Stronge A.J., Rogers W. A., Fisk A. D. (2006).*Web-Based Information Search and Retrieval: Effects of Strategy Use and Age on search success*. Human Factors, ProQuest Medical Library Vol. 48, No. 3 (2006), pp. 434-446.

Tolksdorf R., Nixon L., Paslaru Bontas E., Nguyen D. M., Liebsch F.(2005). *Enabling real world Semantic Web applications through a coordination middleware*. In Proceeding of the 2nd European Semantic Web Conf. Hawaii, ESWC'05, 2005. IEEE computer Society.

Tümer D., Shah M. A., Bitirim Y. (2009). *An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia*. Fourth International Conference on Internet Monitoring and Protection (pp. 51-55). IEEE computer Society.

Van Schaik P., Ling J. (2007).*Design Parameters of Rating Scales for Web Sites*. ACMTransactions on Computer-Human Interaction, Vol. 14, No. 1, (2007), Article 4.
Volery T., Lord D. (2000). *Critical success factors in online education*. The International Journal of Educational Management Vol. 14, No. 5, (2000), pp. 216-223.

Vrasidas C., McIsaac M. S. (1999). *Factors Influencing Interaction in an Online Course*. The American Journal Of Distance Education Vol. 13 No. 3 (1999).

Wang P., Berry M. W., Yang Y. (2003). *Mining Longitudinal Web Queries: Trends and Patterns*. Journal Of The American Society For Information Science And Technology. Vol. 54, No. 8, (2003), pp. 743-758.

White R.W., Ruthven I., Jose M. J., Van Rijsbergen C. J. (2005). *Evaluating Implicit Feedback Models Using Searcher Simulations*. ACM Transactions on Information Systems, Vol. 23, No. 3, (2005), pp. 325–361.

Wolfram D. (2008).*Search characteristics in different types of Web-based IR environments: Are they the same?* Information Processing and Management Vol. 44, (2008), pp. 1279–1292.

Appendix A: Questionnaire

DURBAN UNIVERSITY of TECHNOLOGY

Department of Information Technology

Research Questionnaire

The information in this questionnaire is required for research purpose and your identity will not be known or used therefore please answer as fully and honestly as possible

This questionnaire is designed to capture data that will be used to analyse the factors influencing the ability of computer Science and Information Technology academics to effectively use online search engines as a teaching, learning, and research resource.

Section A: Background Information

1. Indicate your Academic Department	☐ IT ☐ IS ☐ CS ☐ Other ____________
2. Indicate your gender	☐ Male ☐ Female
3. Indicate your Age group	☐ 25 or less ☐ 26 to 35 ☐ 36 to 45 ☐ 6 to 55 ☐ 56 and more
4. Indicate your designation	☐ Prof ☐ Ass.Prof ☐ Associate Director ☐ Senior Lecturer ☐ Lecturer ☐ Jn Lecturer ☐ Others
5. Indicate your University	☐ DUT ☐ UKZN ☐ WSU ☐ Zululand
6. Indicate the number of years experience in lecturing	☐ 2 or less ☐ 3-6 ☐ 7-10 ☐ 11-14 ☐ 15 and more

7. Indicate the number of years experience in research ☐ 2 or less ☐ 3-6 ☐ 7-10 ☐ 11-14 ☐ 15 and more
8. Indicate your highest qualification ☐ Doctorate ☐ Masters ☐ BSc(Honours) ☐ BSc ☐ BTech ☐ ND
9. Indicate the language in which you recieved most of your primary and secondary education isiZulu/isiXhosa ☐ African ☐ Afrikaans ☐ English ☐ French ☐ Others______________
10. Indicate the language in which you recieved most of your tertiary education isiZulu/isiXhosa ☐ African ☐ Afrikaans ☐ English ☐ French ☐ Others______________

Section B: Lecturing experience

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
11. I am currently involved in lecturing					
12. I have lectured different courses so far in my career.					
13. I have lectured at different Higher Education institutions					
14. I can lecture any new topic in computing on a short notice					
15. I have lectured many advanced topics in computing					
16. I usually give advice to other colleagues on their lecturing					

17. I usually lecture without referring to my lecture notes					
18. I easily set up my exams and test question papers					
19. I have always enjoyed my lecturing job					
20. I have always managed heavy lecturing workloads					

Section C: Research experience

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
21. I am currently involved in some research projects					
22. I have written many research books					
23. I have written several research book chapters					
24. I have published many papers as conference proceedings					
25. I have made many oral presentations at conferences					
26. I have published many papers in journals					
27. I am currently supervising postgraduate research					
28. I usually serve as a reviewer for research publications					
29. I use specialized computing digital libraries such as ACM, IEEE, Springer for research					
30. I am currently registered for a postgraduate qualification					

Section D: English Language Ability	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
31. I am good at writing English					
32. I am good at speaking English					
33. I am good at reading English					
34. I read other materials apart from computing related ones					
35. I write other periodicals apart from computing related ones					
36. I usually participate in discussions involving topics outside of the computing field					
37. I always use English while on duty					
38. I easily spot English mistakes in conversations or while reading					
39. I will be comfortable Lecturing a basic English course					
40. I usually speak English out of my work environment					

Section E: Prior searching experience

	Strongly Agree	Agree	undecided	Disagree	Strongly Disagree
41. I have been using search engines for many years					
42. I cannot design a syllabus of a course without using online search engines					
43. I cannot design an exam paper without using online search engines					
44. I cannot compile lecture notes for my courses without using search engines					
45. I prefer online teaching materials compared to books					
46. I directly go to other universities websites to use their educational materials					
47. I use specialized computing digital libraries such as ACM for teaching purposes					
48. I regularly use specialized educational search engines such as Google scholar					
49. I take advantage of advanced search features provided by search engines					
50. I have a technical knowledge of the design and implementation of search engines					

Section F: Searching ability

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
51. I am able to find relevant information on a topic online with ease					
52. I now know most of the search engine to refer to when searching					
53. When searching for information, I always know the right keyword to use					
54. My search results are always successful once I am satisfied with my keyword					
55. I usually use many natural languages to perform a search					
56. I usually only make use of a few number of keywords to perform a search					
57. I usually get relevant search results in a short time					
58. I find search sites user friendly					
59. At the end I always get what I am looking for					
60. I have improved my searching skill over the years					

Thank you

Appendix B Reliability and Validity

B1. Lecturing Experience variable tests: Items 12, 14, 15

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.757	.757	3

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.651
Bartlett's Test of Sphericity	Approx. Chi-Square	46.520
	df	3
	Sig.	.000

Correlation Matrix[a]

		12	14	15
Correlation	12	1.000	.393	.532
	14	.393	1.000	.605
	15	.532	.605	1.000

a. Determinant = .449

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.025	67.498	67.498	2.025	67.498	67.498
2	.613	20.433	87.931			
3	.362	12.069	100.000			

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

a. Only one component was extracted. The solution cannot be rotated.

Component Matrix[a]

	Component
	1
12	.768
14	.813
15	.879

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

B2. Research Experience Variable Tests Items 22, 24, 25, 26, 27, 28

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.925	.930	6

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.833
Bartlett's Test of Sphericity	Approx. Chi-Square	292.026
	df	15
	Sig.	.000

Correlation Matrix[a]

		23	24	25	26	27	28
Correlation	23	1.000	.654	.605	.803	.620	.746
	24	.654	1.000	.819	.723	.691	.680
	25	.605	.819	1.000	.770	.555	.631
	26	.803	.723	.770	1.000	.606	.681
	27	.620	.691	.555	.606	1.000	.727
	28	.746	.680	.631	.681	.727	1.000

a. Determinant = .006

Component Matrix[a]

	Component
	1
23	.859
24	.886
25	.850
26	.891
27	.809
28	.865

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.442	74.027	74.027	4.442	74.027	74.027
2	.549	9.147	83.174			
3	.457	7.610	90.784			
4	.250	4.171	94.954			
5	.186	3.102	98.056			
6	.117	1.944	100.000			

Extraction Method: Principal Component Analysis.

B3 English Language Ability Variable Tests: Items 31, 32, 33, 34, 36

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.783	.849	5

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.805
Bartlett's Test of Sphericity	Approx. Chi-Square	141.559
	df	10
	Sig.	.000

Correlation Matrix[a]

		31	32	33	34	36
Correlation	31	1.000	.701	.693	.529	.244
	32	.701	1.000	.733	.524	.323
	33	.693	.733	1.000	.587	.437
	34	.529	.524	.587	1.000	.524
	36	.244	.323	.437	.524	1.000

a. Determinant = .085

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.159	63.178	63.178	3.159	63.178	63.178
2	.879	17.588	80.767			
3	.420	8.393	89.160			
4	.288	5.767	94.926			
5	.254	5.074	100.000			

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component
	1
31	.822
32	.849
33	.887
34	.792
36	.592

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

B4 Prior searching experience variable Tests: Items 42, 43, 44, 45

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.806	.799	4

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.767
Bartlett's Test of Sphericity	Approx. Chi-Square	88.300
	Df	6
	Sig.	.000

Component Matrix[a]

	Component
	1
42	.861
43	.867
44	.864
45	.553

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Correlation Matrix[a]

		42	43	44	45
Correlation	42	1.000	.695	.659	.290
	43	.695	1.000	.650	.328
	44	.659	.650	1.000	.372
	45	.290	.328	.372	1.000

a. Determinant = .217

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.547	63.670	63.670	2.547	63.670	63.670
2	.797	19.923	83.593			
3	.354	8.838	92.431			
4	.303	7.569	100.000			

Extraction Method: Principal Component Analysis.

B5. Searching Ability Variable Tests: Items 51, 53, 54, 57, 58, 59

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.862	.862	6

Correlation Matrix[a]

		51	53	54	57	58	59
Correlation	51	1.000	.551	.583	.548	.458	.337
	53	.551	1.000	.657	.615	.398	.472
	54	.583	.657	1.000	.511	.514	.452
	57	.548	.615	.511	1.000	.607	.557
	58	.458	.398	.514	.607	1.000	.401
	59	.337	.472	.452	.557	.401	1.000

a. Determinant = .069

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.822
Bartlett's Test of Sphericity	Approx. Chi-Square	152.648
	df	15
	Sig.	.000

Component Matrix[a]

	Component
	1
51	.754
53	.807
54	.809
57	.835
58	.726
59	.684

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.568	59.462	59.462	3.568	59.462	59.462
2	.704	11.733	71.195			
3	.635	10.586	81.780			
4	.449	7.482	89.262			
5	.406	6.772	96.034			
6	.238	3.966	100.000			

Extraction Method: Principal Component Analysis.

B6.All independent Variable validity tests

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.568	59.462	59.462	3.568	59.462	59.462
2	.704	11.733	71.195			
3	.635	10.586	81.780			
4	.449	7.482	89.262			
5	.406	6.772	96.034			
6	.238	3.966	100.000			

Extraction Method: Principal Component Analysis.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.754
Bartlett's Test of Sphericity	Approx. Chi-Square	681.221
	df	153
	Sig.	.000

Correlation Matrix[a]

	12	14	15	23	24	25	26	27	28	31	32	33	34	36	42	43	44	45
Correlation	1.000	.393	.532	-.083	.148	.190	.105	.113	.052	-.045	.025	.103	.033	.108	.028	-.093	-.015	.007
	.393	1.000	.605	.177	.190	.120	.194	.193	.217	.209	.411	.403	.199	.389	-.044	-.146	-.050	.106
	.532	.605	1.000	.144	.293	.208	.236	.278	.188	.161	.210	.123	.100	.143	-.174	-.318	-.129	.040
	-.083	.177	.144	1.000	.654	.605	.803	.620	.746	.137	.234	.184	-.096	.018	-.314	-.133	-.092	-.096
	.148	.190	.293	.654	1.000	.819	.723	.691	.680	.136	.210	.101	-.138	-.078	-.306	-.167	-.139	-.243
	.190	.120	.208	.605	.819	1.000	.770	.555	.631	.019	.067	.099	-.028	.051	-.266	-.116	-.100	-.101
	.105	.194	.236	.803	.723	.770	1.000	.606	.681	.112	.189	.179	.054	.069	-.338	-.186	-.130	-.147
	.113	.193	.278	.620	.691	.555	.606	1.000	.727	.184	.097	.158	-.157	-.015	-.206	-.245	-.174	-.234
	.052	.217	.188	.746	.680	.631	.681	.727	1.000	.088	.062	.053	-.168	.105	-.238	-.220	-.155	-.165
	-.045	.209	.161	.137	.136	.019	.112	.184	.088	1.000	.701	.693	.529	.244	-.405	-.391	-.400	-.305
	.025	.411	.210	.234	.210	.067	.189	.097	.062	.701	1.000	.733	.524	.323	-.383	-.236	-.312	-.043
	.103	.403	.123	.184	.101	.099	.179	.158	.053	.693	.733	1.000	.587	.437	-.228	-.196	-.274	-.109
	.033	.199	.100	-.096	-.138	-.028	.054	-.157	-.168	.529	.524	.587	1.000	.524	-.193	-.261	-.220	-.126
	.108	.389	.143	.018	-.078	.051	.069	-.015	.105	.244	.323	.437	.524	1.000	.032	-.043	-.203	.065

	.028	-.044	-.174	-.314	-.306	-.266	-.338	-.206	-.238	-.405	-.383	-.228	-.193	.032	1.000	.695	.659	.290
	-.093	-.146	-.318	-.133	-.167	-.116	-.186	-.245	-.220	-.391	-.236	-.196	-.261	-.043	.695	1.000	.650	.328
	-.015	-.050	-.129	-.092	-.139	-.100	-.130	-.174	-.155	-.400	-.312	-.274	-.220	-.203	.659	.650	1.000	.372
	.007	.106	.040	-.096	-.243	-.101	-.147	-.234	-.165	-.305	-.043	-.109	-.126	.065	.290	.328	.372	1.000

a. Determinant = 2.73E-006

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.465	30.359	30.359	5.465	30.359	30.359	4.536	25.202	25.202
2	3.435	19.084	49.444	3.435	19.084	49.444	3.279	18.216	43.418
3	2.125	11.808	61.252	2.125	11.808	61.252	2.752	15.289	58.707
4	1.601	8.893	70.145	1.601	8.893	70.145	2.059	11.437	70.145
5	.894	4.966	75.111						
6	.882	4.903	80.014						
7	.741	4.116	84.130						
8	.512	2.842	86.971						
9	.449	2.493	89.464						
10	.395	2.195	91.659						
11	.303	1.685	93.343						
12	.272	1.511	94.854						
13	.219	1.219	96.073						
14	.197	1.096	97.170						
15	.171	.952	98.121						
16	.148	.823	98.944						
17	.107	.594	99.538						
18	.083	.462	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
12				.812
14				.699
15				.847
23	.875			
24	.869			
25	.835			
26	.876			
27	.784			
28	.855			
31		.725		
32		.819		
33		.870		
34		.775		
36		.648		
42			.807	
43			.842	
44			.811	
45			.583	

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Appendix C Descriptive statistics

C1.Independent Variable statistics

Statistics

		Lecturing_Exp_Total_Scores	Research_Exp_Total_Scores	English_Language_Abil_Total_Scores	Prior_Searching_Exp_Total_Scores	Searching_Ability_Total_Scores_DV
N	Valid	61	61	61	61	61
	Missing	0	0	0	0	0
Mean		11.69	12.23	21.57	11.00	23.30
Std. Error of Mean		.351	.827	.370	.477	.537
Std. Deviation		2.742	6.456	2.889	3.724	4.197
Variance		7.518	41.680	8.349	13.867	17.611
Minimum		3	6	12	4	14
Maximum		15	30	25	20	30

Appendix D Inferential Statistics

D1. Correlations

Correlations

		Lecturing_Exp_Total_Scores	Research_Exp_Total_Scores	English_Language_Abil_Total_Scores	Prior_Searching_Exp_Total_Scores	Searching_Ability_Total_Scores_DV
Lecturing_Exp_Total_Scores	Pearson Correlation	1	.240	.288*	-.106	.299*
	Sig. (2-tailed)		.062	.024	.416	.019
	N	61	61	61	61	61
Research_Exp_Total_Scores	Pearson Correlation	.240	1	.071	-.277*	.090
	Sig. (2-tailed)	.062		.584	.031	.489
	N	61	61	61	61	61
English_Language_Abil_Total_Scores	Pearson Correlation	.288*	.071	1	-.301*	.376**
	Sig. (2-tailed)	.024	.584		.019	.003
	N	61	61	61	61	61
Prior_Searching_Exp_Total_Scores	Pearson Correlation	-.106	-.277*	-.301*	1	-.135
	Sig. (2-tailed)	.416	.031	.019		.298
	N	61	61	61	61	61
Searching_Ability_Total_Scores_DV	Pearson Correlation	.299*	.090	.376**	-.135	1
	Sig. (2-tailed)	.019	.489	.003	.298	
	N	61	61	61	61	61

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

D2. Regression Tests

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Prior_Searching_Exp_Total_Scores, Lecturing_Exp_Total_Scores, Research_Exp_Total_Scores, English_Language_Abil_Total_Scores[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Searching_Ability_Total_Scores_DV

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	192.103	4	48.026	3.111	.022[a]
	Residual	864.585	56	15.439		
	Total	1056.689	60			

a. Predictors: (Constant), Prior_Searching_Exp_Total_Scores, Lecturing_Exp_Total_Scores, Research_Exp_Total_Scores, English_Language_Abil_Total_Scores

b. Dependent Variable: Searching_Ability_Total_Scores_DV

Coefficients[a]

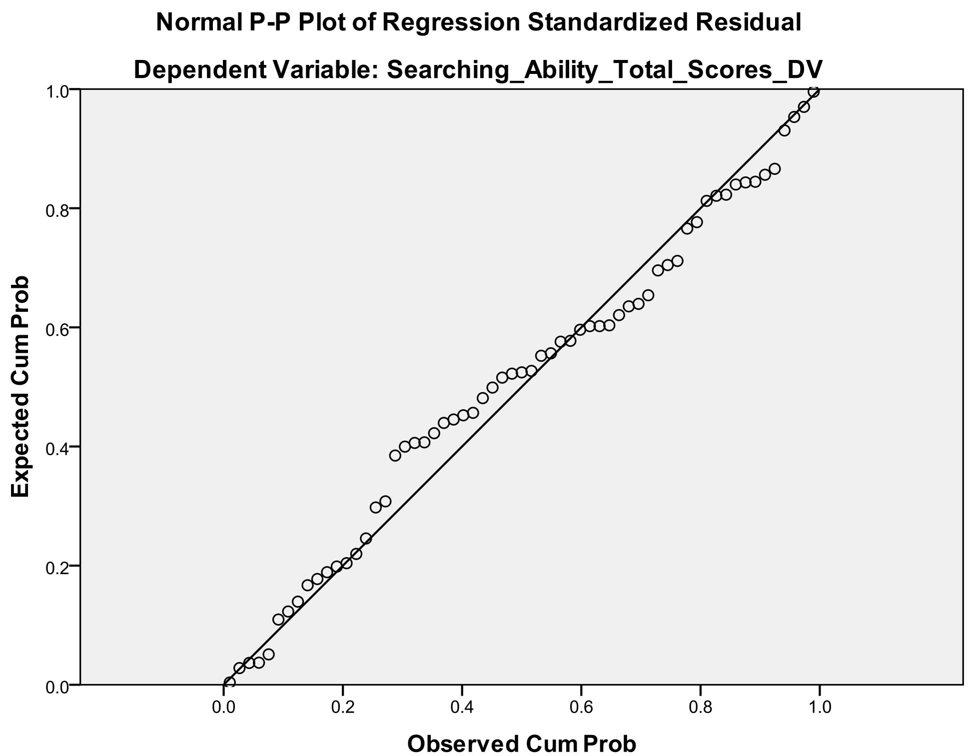
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.967	5.019		1.986	.052
	Lecturing_Exp_Total_Scores	.313	.199	.204	1.574	.121
	English_Language_Abil_Total_Scores	.452	.192	.311	2.359	.022
	Research_Exp_Total_Scores	.009	.084	.014	.112	.911
	Prior_Searching_Exp_Total_Scores	-.018	.148	-.016	-.123	.903

a. Dependent Variable: Searching_Ability_Total_Scores_DV

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	18.01	26.00	23.30	1.789	61
Residual	-10.489	10.321	.000	3.796	61
Std. Predicted Value	-2.953	1.510	.000	1.000	61
Std. Residual	-2.669	2.627	.000	.966	61

a. Dependent Variable: Searching_Ability_Total_Scores_DV


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Searching_Ability_Total_Scores_DV
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

D3. One Way ANOVA tests

Between-Subjects Factors

		Value Label	N
Academic_Department	1	Information Technology	47
	2	Computer Science	14
Gender	1	Male	36
	2	Female	25
Age_Group	1	25 or less	5
	2	26 to 35 yrs	22
	3	36 to 45 yrs	17
	4	46 to 55 yrs	12
	5	56 and more	5
Position	1	Others	3
	2	Jn lecturer	13
	3	Lecturer	32
	4	Sr Lecturer	5
	5	Associate Dir	4
	6	Associate Prof	3
	7	Professor	1
University_Affiliation	1	Durban University of Technology(DUT)	23
	3	University of KwaZulu Natal (UKZN)	4

	4	Walter Sisulu University (WSU)	28
	5	University of Zululand	6
Yrs_Of_Experience_Lecturing	1	2 or Less	9
	2	3 to 6	16
	3	7 to 10	12
	4	11 to 14	12
	5	15 and more	12
Yrs_Of_Experience_Research	1	2 or Less	33
	2	3 to 6	17
	3	7 to 10	4
	4	11 to 14	4
	5	15 and more	3
Highest_Qualification	1	ND	3
	2	BTech	15
	3	BSc	7
	4	BSc(Hons)	10
	5	Masters	20
	6	Doctorate	6
Primary_Second_Language	1	Other	3
	2	French	3
	3	English	50
	6	Zulu/Xhosa	5
Tertiary_Language	2	French	2
	3	English	58
	6	Zulu/Xhosa	1

Tests of Between-Subjects Effects

Dependent Variable:Searching_Ability_Total_Scores_DV

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	789.862[a]	33	23.935	2.422	.010
Intercept	727.640	1	727.640	73.629	.000
Academic_Department	13.174	1	13.174	1.333	.258
Gender	5.680	1	5.680	.575	.455
Age_Group	26.399	4	6.600	.668	.620
Position	259.544	6	43.257	4.377	.003
University_Affiliation	71.366	3	23.789	2.407	.089
Yrs_Of_Experience_Lecturing	43.557	4	10.889	1.102	.376
Yrs_Of_Experience_Research	189.545	4	47.386	4.795	.005
Highest_Qualification	156.286	5	31.257	3.163	.022
Primary_Second_Language	201.588	3	67.196	6.800	.001
Tertiary_Language	101.296	2	50.648	5.125	.013
Error	266.827	27	9.882		
Total	34159.000	61			
Corrected Total	1056.689	60			

a. R Squared = .747 (Adjusted R Squared = .439)

D4. Five Way ANOVA Tests

Between-Subjects Factors

		Value Label	N
Position	1	Others	3
	2	Jn lecturer	13
	3	Lecturer	32
	4	Sr Lecturer	5
	5	Associate Dir	4
	6	Associate Prof	3
	7	Professor	1
Yrs_Of_Experience_Research	1	2 or Less	33
	2	3 to 6	17
	3	7 to 10	4
	4	11 to 14	4
	5	15 and more	3
Highest_Qualification	1	ND	3
	2	BTech	15
	3	BSc	7
	4	BSc(Hons)	10
	5	Masters	20
	6	Doctorate	6
Primary_Second_Language	1	Other	3
	2	French	3
	3	English	50
	6	Zulu/Xhosa	5
Tertiary_Language	2	French	2
	3	English	58

Between-Subjects Factors

		Value Label	N
Position	1	Others	3
	2	Jn lecturer	13
	3	Lecturer	32
	4	Sr Lecturer	5
	5	Associate Dir	4
	6	Associate Prof	3
	7	Professor	1
Yrs_Of_Experience_Research	1	2 or Less	33
	2	3 to 6	17
	3	7 to 10	4
	4	11 to 14	4
	5	15 and more	3
Highest_Qualification	1	ND	3
	2	BTech	15
	3	BSc	7
	4	BSc(Hons)	10
	5	Masters	20
	6	Doctorate	6
Primary_Second_Language	1	Other	3
	2	French	3
	3	English	50
	6	Zulu/Xhosa	5
Tertiary_Language	2	French	2
	3	English	58
	6	Zulu/Xhosa	1

Tests of Between-Subjects Effects

Dependent Variable:Searching_Ability_Total_Scores_DV

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	811.618[a]	36	22.545	2.208	.022
Intercept	66.072	1	66.072	6.471	.018
Lecturing_Exp_Total_Scores	1.375	1	1.375	.135	.717
English_Language_Abil_Total_Scores	45.480	1	45.480	4.454	.045
Position	188.744	4	47.186	4.621	.007
Yrs_Of_Experience_Research	32.879	3	10.960	1.073	.379
Highest_Qualification	81.489	5	16.298	1.596	.199
Primary_Second_Language	61.347	2	30.674	3.004	.069
Tertiary_Language	73.163	1	73.163	7.165	.013
Position * Yrs_Of_Experience_Research	.382	1	.382	.037	.848
Position * Highest_Qualification	20.914	2	10.457	1.024	.374
Position * Primary_Second_Language	.206	1	.206	.020	.888
Position * Tertiary_Language	.000	0	.	.	.
Yrs_Of_Experience_Research * Highest_Qualification	1.327	1	1.327	.130	.722
Yrs_Of_Experience_Research * Primary_Second_Language	.000	0	.	.	.
Yrs_Of_Experience_Research * Tertiary_Language	.000	0	.	.	.
Highest_Qualification * Primary_Second_Language	.000	0	.	.	.
Highest_Qualification * Tertiary_Language	.000	0	.	.	.

Primary_Second_Language * Tertiary_Language	.000	0	.	.	.	
Position * Yrs_Of_Experience_Research * Highest_Qualification	.000	0	.	.	.	
Position * Yrs_Of_Experience_Research * Primary_Second_Language	.000	0	.	.	.	
Position * Yrs_Of_Experience_Research * Tertiary_Language	.000	0	.	.	.	
Position * Highest_Qualification * Primary_Second_Language	.000	0	.	.	.	
Position * Highest_Qualification * Tertiary_Language	.000	0	.	.	.	
Position * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.	
Yrs_Of_Experience_Research * Highest_Qualification * Primary_Second_Language	.000	0	.	.	.	
Yrs_Of_Experience_Research * Highest_Qualification * Tertiary_Language	.000	0	.	.	.	
Yrs_Of_Experience_Research * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.	

Highest_Qualification * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.
Position * Yrs_Of_Experience_Research * Highest_Qualification * Primary_Second_Language	.000	0	.	.	.
Position * Yrs_Of_Experience_Research * Highest_Qualification * Tertiary_Language	.000	0	.	.	.
Position * Yrs_Of_Experience_Research * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.
Position * Highest_Qualification * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.
Yrs_Of_Experience_Research * Highest_Qualification * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.
Position * Yrs_Of_Experience_Research * Highest_Qualification * Primary_Second_Language * Tertiary_Language	.000	0	.	.	.
Error	245.071	24	10.211		

Total	34159.000	61				
Corrected Total	1056.689	60				

a. R Squared = .768 (Adjusted R Squared = .420)