USABILITY OF A KEYPHRASE BROWSING TOOL BASED ON A SEMANTIC CLOUD MODEL

Onaje Omotola Johnston, B.A., M.A.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2006

APPROVED:

Brian O'Connor, Committee Chair
Guillermo Oyarce, Committee Member
Richard Herrington, Committee Member
Herman L. Totten, Dean of the School of Library
and Information Sciences
Sandra L. Terrell, Dean of the Robert B. Toulouse
School of Graduate Studies

Johnston, Onaje Omotola. *Usability of a keyphrase browsing tool based on a semantic cloud model.* Doctor of Philosophy (Information Science), August 2006, 133 pp., 31 tables, 35 figures, references, 54 titles.

The goal of this research was to facilitate the scrutiny and utilization of Web search engine retrieval results. I used a graphical keyphrase browsing interface to visualize the conceptual information space of the results, presenting document characteristics that make document relevance determinations easier.

ACKNOWLEDGEMENTS

I would like to thank my committee: Brian O'Connor, Guillermo Oyarce and Richard Herrington for all the help and guidance I received. I also extend thanks for their support to my family: Enid St. Kitts, Beryl, Anthony, Omotayo and Ihuoma Johnston and to my friends: Amy Wolf, Fatih Oguz, Jenine Lurie, Winsome Demetrius and Phyllis Mire Romano.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This study investigates if a keyphrase browsing application using information visualization, via context views, can be a useful adjunct to standard information retrieval techniques; specifically, how to improve user access to information and relevance determination, by developing a keyphrase browsing interface, for visualizing information. It enables a user to explore, manipulate, and connect concepts in an information space to their information need, by examining documents in context, in their retrieved results from a search engine. Further, it develops a conceptual visual model to illustrate semantic relationships in the information space; uses a semantic analysis technique to associate related documents that do not contain words in common but do share a semantic relationship.

Background

Most research and commercial efforts have focused on producing effective systems for retrieving and ranking relevant documents in response to a query, little attention has been paid to ease the process of result inspection and query reformulation, especially in the Web domain (Berenci et al., 1998). Search engines provide a degree of access to information published on the Internet. Searchers need tools that will enable them to decide if the documents matching their search query fulfill their information need. Web users are precision oriented - they prefer a small set of documents containing a good proportion of useful documents to a large set of documents that contains a lot of useful information, but a fair amount of irrelevant information as well.

The objective of this research is to build a keyphrase browsing application that enables a user, via context 'windows' or 'views', to examine documents in the retrieved set via single and multi-word keyphrases (throughout this thesis keywords and keyphrases are used interchangeably and mean the same thing for the purposes of my discussion). Users would be able to quickly determine which documents in the retrieved set are most relevant. Since naive keyword searches fail to discover the semantic relationships between documents that do not share words in common but may actually be related, contextual networks graphs will be used to locate similar documents in conjunction with keyword search techniques.

"Language and meaning are important to the study of information systems." "Requests for information can be clear or ambiguous, precise or imprecise, just as any statement in natural language can" (Blair, in process). According to Blair, "Our current most widespread model of information systems" implemented is the logical computer data model of information. The logical model works "well for the precise, highly determinate content of our data bases - things like names, addresses, phone numbers, account balances, etc. … less determinate information - things like written text, images, audio recordings, etc. … are not as well managed by the logical data model of information, as one can easily see when trying to find some specific subject matter using an Internet search engine on the World Wide Web" (Blair, in process).

There is a large amount of data on the Web. Lyman and Varian (2003) estimated that "The World Wide Web contains about 170 terabytes of information on its surface; in volume this is seventeen times the size of the Library of Congress print collections." Only a portion is indexed by search engines, meaning that at the onset of a search the searcher does not have optimal conditions. There are potentially many 'relevant' documents that will not be found because they are excluded from the search engine collection. Within the collection, 'relevant' documents that do not contain the query words are missed.

Information retrieval (IR) systems have problems of representation. Choices have to be made about the makeup of a collection. A collection may not include all useful documents that a searcher would need. "A key issue for information retrieval (IR) and all other content-based text management applications is document indexing, i.e. the task of automatically constructing an internal representation of a text." (Caropreso, Matwin and Sebastiani, 2000) The features selected for the representation may not be the features used by a searcher.

Blair believes, "ordinary language is the best medium for us to express our information needs, and any subset of ordinary language that may be used as an access language to an information system will be correspondingly less effective than ordinary language for searching." This suggests "that the best information systems are those which can understand the subtle meanings and nuances of information requests stated in ordinary language." To this end many search engines employ keywords to search their collection of documents, where they attempt to find documents containing the words in a user query. Researchers are now working on the 'semantic Web.' It is a conceptual information space with the practical objective to be able to express meaning in a way that machines will understand so that they can then find documents more efficiently.

JC Hertz "maintains that any scheme that requires humans to input metadata [data about data] with their data will fail." (Hardin, 2005a) She points to Weblogs that use simple technology (folksonomy) to classify their content. Users typically prefer to complete tasks in the simplest manner. Most people will not take the time to learn a metadata coding scheme for their content if it's too hard to learn, and learn quickly. Downes (2004) suggests that the Semantic Web has not taken the shape or has not been implemented the way its designers envisioned because the technologies it uses must be "(a) capable of being understood by millions of people, and (b) easily implemented by these same people." When asked "How much metadata will people need to input on the Semantic Web? [Tim] Berners-Lee responded that most of this data is in databases; in proprietary file formats. Connections haven't been well made" (Hardin, 2005b).

At this time there are separate and sometimes competing efforts at linking data. There is the original hypertext model, which is the most prevalent. The semantic model is supposed to provide a standard framework but has not matured to that point. Instead there appears to be more fragmentation in implementing the semantic Web than cohesion. The problem of finding the data that you need in Web pages still remains.

Definitions[1]

Cronbach's alpha - reliability is measured in terms of the ratio of true score variance to observed score variance, the relationship between true score and observed score should be strong.

Collection* - a group of items or documents usually selected based on a collection management plan.

Corpus - In principle, any collection of more than one text can be called a corpus, (corpus being Latin for 'body,' hence a corpus is any body of text). See collection.

Dichotomous question - generally a "yes/no" question can be used to "screen out" those who do not match a criterion.

Document - a unit of retrieval; it can be a paragraph, a section, a chapter, a Web page, an article, or a whole book. Image or music files can also be considered 'documents.'

Document surrogate* - a representation of a document such as the title, the author, the date of publication and a short summary or image; surrogates are often used to display the answers to a user query.

Feature* - information extracted from an object or document and used during query processing.

Folksonomy** - A term coined by Thomas Vander Wal to refer to a collaboratively generated, open-ended labeling system that enables Internet users to categorize content such as pages, online photographs, and Web links. The freely chosen labels – called tags – help to

[1] These are the working definitions used for this research. Those marked with * are adapted from the glossary of Modern Information Retrieval and those with ** from Wikipedia, The Free Encyclopedia.

improve search engine's effectiveness because content is categorized using a familiar, accessible, and shared vocabulary. The labeling process is called tagging.

HTML - Hypertext markup language used to create Web pages.

Human-Computer Interface* (HCI) – ‘the study of interfaces which assist the user with information seeking related tasks such as: query formulation, selection of information sources, understanding of search results, and tracking of the retrieval task.’

Hypertext model* - ‘a model of information retrieval based on representing document relationships as edges of a generic graph in which the documents are the nodes.’

Index - a data structure, usually a list of terms extracted from or associated with a document, that serves as a guide, pointer, indicator or otherwise facilitates reference when searching.

Index term* (or keyword) – ‘a pre-selected term which can be used to refer to the content of a document. Usually, index terms are nouns or noun groups.’ The index can also be a list of all the words in a document.

Logical view of documents* - ‘the representation of documents and Web pages adopted by the system. The most common form is to represent the text of the document by a set of indexing terms or keywords.’

Metadata* - data about data; attributes of data or a document, usually broken into categories or facets.

Noun - Grammatical construct common to most romance and Germanic languages that refers to places, or things. Nouns can be objects or subjects in a sentence, can be described using adjectives, can be replaced by pronouns (such as "it"), and are generally preceded by a direct or indirect article (such as "the" or "a" in English).

Private Web - Web pages that are not indexed by search engines. They are blocked by the robot exclusion standard (a robots.txt file); typically a dynamic page (e.g. CGI, PHP ASP etc.), hidden or password protected, or behind a firewall (i.e. company intranet).

Public Web - Web pages that are open to the public, meaning they can be indexed by search engines.

Query* - ‘the expression of the user information need in the input language provided by the information system. The most common type of input language allows simply the specification of keywords and of a few Boolean connectives.’

Query expansion* - ‘a process of adding new terms to a given user query in an attempt to provide better contextualization (and hopefully retrieve documents which are more useful to the user)’.

Semantic indexing - used in this research to refer to techniques that examine document content instead of metadata i.e. patterns of co-occurrence relationships derived from counting words, and their distribution across a document collection.

Set - a well-defined collection of objects or elements.

Stemming* - 'a technique for reducing words to their grammatical roots.'

Stopwords* - 'words which occur frequently in the text of a document. Examples of stopwords are articles, prepositions, and conjunctions.' They are contained in a stoplist.

Tag cloud** (known as a weighted list in the field of visual design) – 'a visual depiction of content tags used on a Website. Often, more frequently used tags are depicted in a larger font or otherwise emphasized, while the displayed order is generally alphabetical. Thus both finding a tag by alphabet and by popularity is possible. Selecting a single tag within a tag cloud will generally lead to a collection of items that are associated with that tag.' Also referred to as a topic cloud.

Topic - the main organizing principle of a discussion, usually verbal or written (could also be visual). It is the focus or theme which governs the subject matter under consideration

Tf.Idf - the term weight: the term frequency (Tf) is the frequency of the occurrence of a term or phrase in a single document. The Index document frequency (Idf) is the frequency of a term or phrase in all documents in a collection is often calculated as Idf = log (N/n). *N* being the total number of documents in a collection and *n* the number of documents containing the term. Weight for a term in a document calculated as *w* = Tf x Idf

User relevance feedback* - 'an interactive process of obtaining information from the user about the relevance and the non-relevance of retrieved documents.'

Summary and Structure of the Study

This study analyzed the user interaction satisfaction for a keyphrase browsing application called KeyView. The dissertation is composed of six chapters. This chapter has provided an introduction to the research problem. Chapter 2 reviews related literature discussing the information retrieval process and retrieval models: specifically the three classic models and their derivatives such as the spreading activation model and drawbacks of the model. It also discusses the selection of features for indexing and statistical vs. syntactic keyphrase extraction. Interface design Web information retrieval evaluation, and keyphrase browsing interfaces are also covered. Chapter 3 proposes a conceptual framework that guided the development of the keyphrase browsing application. Chapter 4 describes the research design methodology, including sources of data and data analysis procedures. It also includes pilot study results. Chapter 5 presents the result of data analysis within the conceptual framework of this study. Chapter 6 discusses the key findings of this study and offers suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

Information Retrieval Process

Information retrieval (IR) is about the collection, representation, storage, organization, accessing, manipulation and display, of the information items necessary to satisfying a searcher's information needs. The majority of Web search engines use a simple information retrieval interaction model (see Appendix A). This model does not accommodate the users dislike for many retrieval results that "do not directly address their information needs. It also contains an underlying assumption that the user's information need is static and the information seeking process is one of successively refining a query until it retrieves all and only those documents relevant to the original information need" (Hearst, 1999) .

Keyphrases provide a simple mechanism for supporting collection evaluation, topic or information space exploration, and query exploration. Associative information retrieval finds relevant information by retrieving information that is 'associated' with already retrieved information that is know it to be relevant. The system determines associations between information items through interaction with the user, for example by retrieving documents that are similar to documents the user points out to be relevant.

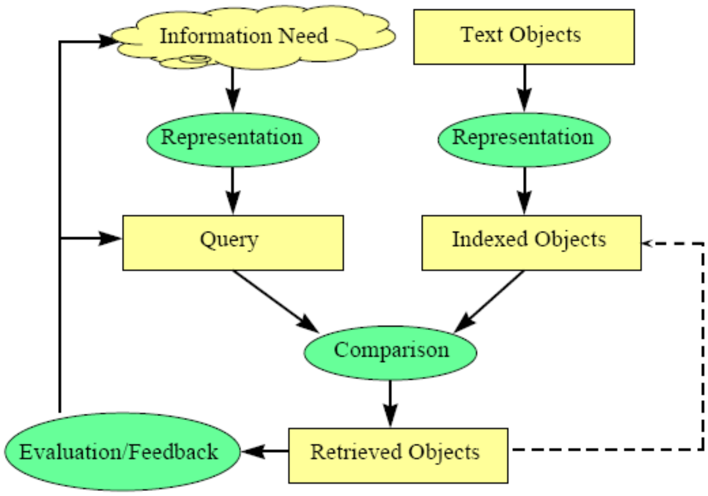


Figure 1. Representation of an informal retrieval system (IRS) (Hearst, 1999).

What is a Retrieval Model?

A model is an embodiment of the theory in which we define a set of objects about which assertions can be made and restrict the ways in which classes of objects can interact. A retrieval model specifies the representations used for documents and information needs, and how they are compared (Turtle & Croft, 1992). An IR model specifies the representations used for documents and information needs, and how they are compared. An information retrieval model has a representation (index) for documents and a method of representing queries. It has a modeling framework for the relationships among documents and queries. It can also have a ranking or similarity function to order documents with respect to a query match.

Exact-Match System

Document features (such as document terms) are binary variables (0, 1). A document is found or not found (retrieved or not retrieved) based on the success of a Boolean query expression. The query specifies precise retrieval criteria. The result is a set of unordered matching documents.

Best-Match System

The query expression describes good or “best” matching document. Every document matches query to some degree. The result is a ranked list of documents.

Mixing the Two Systems

Most search engines often provide a mixture of both systems: ranking mechanism for the result set (best of both worlds) and best-match query language that incorporates exact-match operators.

Three ‘Classic’ IR Models and Derivatives

The three ‘classic’ models and their derivatives referenced in modern information retrieval are the Boolean model, vector space model and probabilistic model. They are described briefly below with greater attention given to neural network based models. This research considers the Boolean model an "exact match" model while the others are "best match" models.

Boolean Model

The Boolean model is based on set theory and Boolean algebra. Documents are expressed in sets of terms (indexing terms) and queries are Boolean expressions on terms. Boolean model only retrieves an exact match. It can retrieve too little or too much. There is no ranking of documents to indicate closest match.

Statistical Model

The vector space and probabilistic models are the two major examples of the statistical retrieval approach. Both models use statistical information in the form of term frequencies to determine the relevance of documents with respect to a query. Although they differ in the way they use the term frequencies, both produce as their output a list of documents ranked by their estimated relevance.

Vector Space Model

The vector space model (VSM) represents the documents and queries as vectors in a multidimensional information space, of terms based features. Features (usually term co-occurrence) are selected from document collection or corpus and make up a term-document matrix (TDM). The distance between feature vectors determine similarity or relevance.

Term-Document Matrix

In the TDM rows represent documents while columns represent document terms. For instance if a document corpus had only the following two (short) documents:

Document 1: “I like butterflies”

Document 2: “I hate butterflies,” then the document-term matrix would be as shown in Table 1.

Table 1

Term Document Matrix

	I	like	hate	butterflies
D1	1	1	0	1
D2	1	0	1	1

Showing which documents contain which terms and how many times they appear. A more sophisticated weight scheme could be used beyond simple binary weights; such as the tf-idf (term frequency - inverse document frequency).

Cluster Model

Clustering allows users to quickly get the general idea of a collection by grouping like items together, ideally in ways that will make sense to human beings. Clustering consists of partitioning a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait - often similarity or proximity for some defined distance measure.

Latent Semantic Indexing

Latent semantic indexing (LSI) model uses an index of document features similar to the term-document matrix of the vector model but uses neural or graph theory techniques to find similarities. The assumption is that there is ‘latent’ structure in the associative term-document relationship. Queries can retrieve documents even if they have no words in common.

Probabilistic Model

The probabilistic retrieval model is based on the probability ranking principle. Documents are ranked in decreasing order of probability of relevance to the information need. The principle acknowledges that there is uncertainty in the representation of the information need therefore documents may not match exactly.

Linguistic or Natural Language Processing Model

Natural language processing (NLP) is a subfield of artificial intelligence and linguistics. It studies the problems inherent in the processing and manipulation of natural language, and, natural language understanding devoted to making computers "understand" statements written in human languages.

Neural Network Model

The neural network model, frequently used in artificial intelligence, is based on the neurological functions of biological organisms. This model is very good at pattern recognition. Doyle (1961) proposed information retrieval by means of "semantic road maps" to "increase the mental contact between the reader and the information store, so that the reader can proceed unerringly and swiftly to identify and receive the [information] he is looking for."

Network Models of Semantic Memory

Basic Assumptions

1. Concepts are represented in a network of interconnecting nodes.
2. Distance between the nodes represents similarity between the items.
3. Definition of a concept is in terms of the connections with other concepts.

Principle of Semantic Relatedness

- The distance between two nodes in a network is determined by semantic relatedness.
- Concepts close in meaning/highly related (e.g., doctor, nurse) are stored close together in memory.
- Unrelated concepts (doctor, truck) are stored far away.

- Time to verify relationships depends on "distance" and the likelihood that activation has spread through pathways connecting the concepts

Semantic Memory Network

It is widely believed that people store, access and retrieve information through semantic networks. Many theories of semantic memory contend that memory traces for semantically associated information can be characterized in terms of organized networks. In the late 1960's, Ross Quillian introduced semantic networks as a method of modeling the structure and storage of human knowledge in the shape of a graph. Quillian wanted his system to explore the meaning of English words by the relationships between them.

Semantic Network Cognition Model

In this model of cognition, all known concepts are connected in a network, called a semantic network. A concept is directly connected by links to its most-closely-related concepts. When a concept receives an activation signal from one of its neighbors, it in turn passes along the activation signal to each of its other neighbors until the activation energy is spent. This is referred to as spreading activation. "Spreading activation helps explain how thinking of one topic can bring to mind related topics" (Withrow, 2003).

Quillian's Architecture

- A node (or point or vertex) represents a concept or word that the system "knows" about.
- An arc (or edge) between nodes represents a relation between two concepts.
- A plane is a definition composed of a concept (node) and its relations.

Associative Retrieval or Connectionist Search

Vannevar Bush (1945) describes “associative indexing, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another.” In associative retrieval associations among information items are often represented as a network, where information items are represented by nodes and associations by links connecting nodes; often implemented by means of a technique called spreading activation.

Crestani (1997b) dates studies on associative retrieval date as early as the 1960s, originating in “statistical studies of associations among terms and/or documents in a collection.” The ‘associative linear retrieval model’ is an early model built on the concept of associative retrieval. This model expands an “original query using statistically determined term–term, term–document, and document–document associations. This technique is based on the assumption that there are statistically determinable relations among terms, among documents, and among documents and terms” (Crestani, 1997b).

The Spreading Activation Model

The spreading activation model (see Figure 2; designed by Collins and Loftus in 1975) measures *semantic distance* in the mind between words. Schvaneveldt, Durso & Mukherji (cited in Brooks, 1998) consider semantic distance a psychological construct that locates “concepts along various dimensions of meaning.” Semantic distance determines the degree of semantic similarity, or relatedness, between two or more concepts. The spreading activation model assumes that human memory is an associative network in which the most similar memory items have the strongest connections. Memory is retrieved by spreading activation.

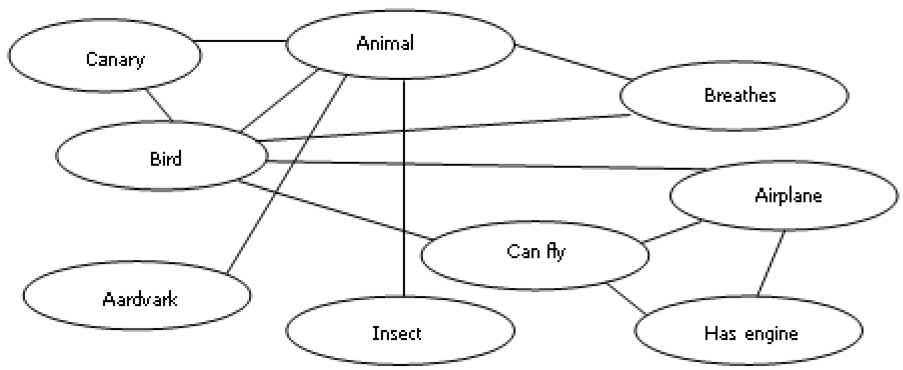


Figure 2. Spreading activation model example.

Crestani (1997b) proposed a conceptual model whereby the application of the classification mechanism to the IR data works at three different levels of abstraction: documents, index terms, and concepts (see Figure 3). Links connect nodes to express a semantic relation between them. The semantics of a link can be made explicit by placing a label on the link and/or can be measured for importance via weights. There are links connecting objects of the same type (on the same level) and links connecting objects of different type. For example a link connecting two index terms indicates that the two terms are similar or occur quite often together. A link connecting an index term with a document indicates that the document has been indexed using that term or that that term occurs in the document. A link between an index term and a concept indicates that the concept can be expressed using that index term. Links on the document level represent bibliographical citations or similarity between documents.

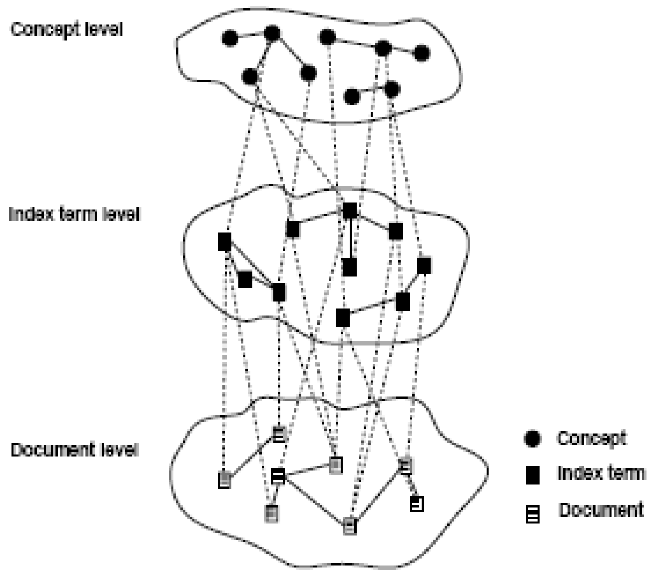


Figure 3. A three-level hypertext (Crestani, 1997).

The Pure Spreading Activation Model

The SA model in its *pure* form (see Figure 4) is made up of a conceptually simple processing technique on a network data structure. The processing technique is defined by a sequence of iterations that can be halted by the user or by the triggering of some termination condition. An iteration or pulse consists of:

1. Preadjustment
2. Spreading
3. Postadjustment
4. Termination check

The pre-adjustment and post-adjustment phases are optional; activation decay can be applied to the active nodes. These phases are used to avoid retention of activation from previous pulses, enabling control of activation on single nodes and overall network activation (see

Appendix D for an alternate illustration of the SA model process). They serve as a form of ‘loss of interest’ in nodes that are not continually activated (Crestani, 1997a, 1997b).

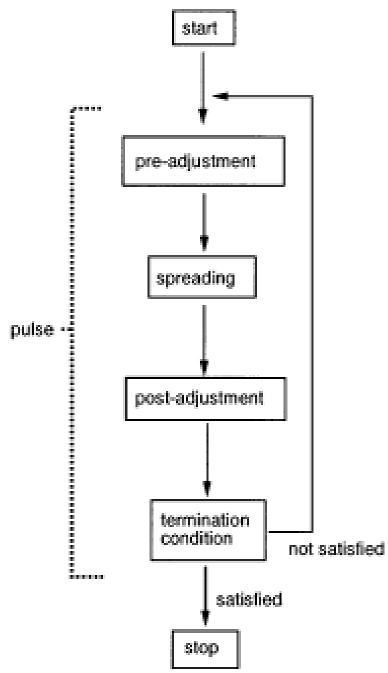


Figure 4. The pure SA model (Crestani, 1997).

Constrained Spreading Activation

Crestani et. al. (1997a, 1997b) used ‘constrained’ spreading activation because the pure SA model has several serious drawbacks:

- If the pre-adjustment and the post-adjustment phases are missing or not well implemented the activation ends up spreading all over the network;
- There is not a complete use of the information provided by labels or weights associated to links;
- It is difficult to implement forms of inference based on the semantics or weights of associations.

The solution is to consider the relationships among nodes in the SA process can have diverse significance i.e. using labels or weights on links and by processing links in special ways according to them. Complex forms of heuristics can be employed, and activation spread on the network according to complex inference rules. These rules are *constraints* on the spreading. Some constraints used in SA models are identified by Crestani (1997a, 1997b):

Distance constraints: the spreading of activation should cease when it reaches nodes that are too far away in terms of links covered to reach them from the initially activated ones. The strength of the relation between two nodes decreases with their semantic distance.

Fan-out constraints: the spreading of activation should stop at nodes with very high connectivity (fan-out), i.e. at nodes connected to a very large number of other nodes.

Path constraints: activation should spread using preferential paths, using application dependent inference rules. This can be depend on link weights or using link labels, to divert the activation flow to a particular semantic path while stopping it from following other less meaningful paths.

Activation constraints: control the spreading of activation on the network by using the threshold activation function at a single node level. The threshold value could automatically change in relation to the total level of activation over the entire network during a single pulse. In addition, each node or set of nodes can have different threshold levels assigned "in relation to their meaning in the context of the application" (Crestani, 1997b). In the pure SA model, these constraints act during the pre-adjustment phase (this is the case for distance, fan-out, and path constraints) or during the post-adjustment phase (activation constraints).

A node can represent anything: terms, documents, authors etc. Links indicate the association of a node with another node, as, for instance, an author with a document or a document with a document with a cited document. An example of a fragment of a document collection representation is shown in Figure 5. "Specific link types include term occurrence, document publication, term assignment by indexing, document authorship, document assignment to classification, document citation, and so forth. The set of node types and link types is determined by the data available and by the purpose of the application" (Crestani, 1997a).

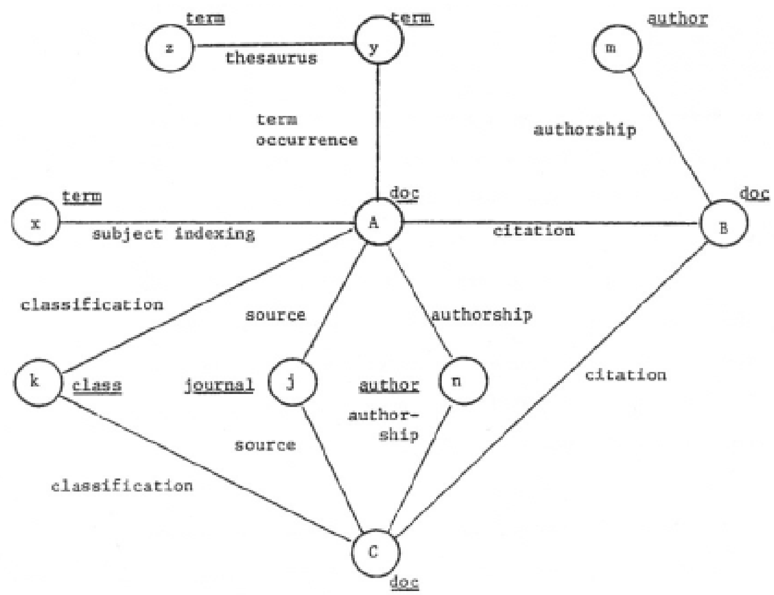


Figure 5. The structure of a bibliographic collection network representation (Preece, 1981). [2]

IR systems that use the SA techniques (see Figure 6) usually do not implement the pure SA model. Crestani (1997a) noted that Preece's 1981 research showed improved retrieval using constrained activation. Some differences identified in the constrained SA model:

- The activation level of a node reached by the spreading of activation is determined by the starting activation level and the type of nodes and links traversed before reaching it.
- Distance constraints are usually imposed by stopping or degrading the activity at some specified distance from the original node.
- Nodes with a large branching ratio (fan-out), that is, nodes connected to many other nodes, may receive special treatment in the activation process in order to avoid or boost a large spreading of activation on the network.
- The activation process follows specified rules that try to mimic some sort of inference in the process of associative retrieval. (Crestani, 1997a)

[2] An adapted version of this is found in "Application of Spreading Activation Techniques in Information Retrieval," Crestani, F. (1997), Artificial Intelligence Review, Volume 11(Issue 6).

Preece (1981) drew a parallel between spreading activation, the antecedent of CNG, and the vector model. He stated it was "a combination of a network view of the data and a processing paradigm based on spreading through that network. Since the network can be realized as a connectivity matrix, we could specify the processing as a matrix equation, with the status of the network indicated at any given time as a vector of activations. If we project that vector onto the subset of the node set corresponding to the particular node types, we obtain weight vectors." He also showed it "is possible to implement the Boolean model, the vector space model and use various forms of weighting for associative retrieval. Moreover, he showed how, using relevance feedback, SA can be used for automatic classification and indexing, and for concept building" (Crestani, 1997a).

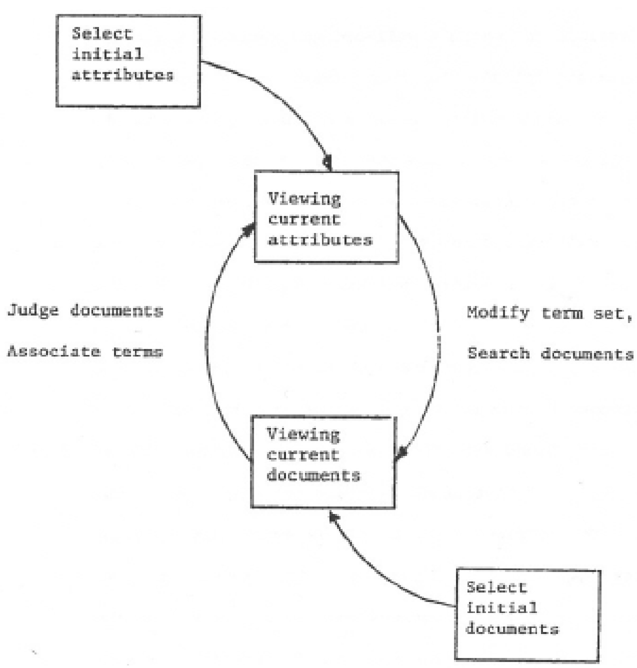


Figure 6. Basic search process using the spread activation model (SAM). The entry point for the search can be either query terms or sample documents (Preece, 1981).

Drawbacks

The common drawback found by many researchers was in the manual construction of semantic network representations. Automatic means are more efficient because they save time. Crestani (1997a) noted at the time his review was conducted that there was a decline in interest for using spreading activation on large networks (e.g. semantic networks) but predicted that interest in the SA technique itself would resurge. The primary stumbling block was the construction of network associations among information items (documents, citations, images, Web pages etc.). The process is time consuming when the size of the document collection is very large, requiring lots of computing power. Early research in associative retrieval used small document collections, where semantic associations were often created manually or semi-automatically. Large document collections were not conducive to using those two techniques. The advances in computing power, and the lower cost, make it possible to construct associative networks from large document collections automatically. Currently there is an intense focus on the semantic Web, spurring research in semantic search techniques; spreading activation being one.

Contextual Network Graphs

Contextual network search (CNS) or contextual network graphs (CNG) is the National Institute for Technology and Liberal Education (NITLE) term for an alternative to latent semantic indexing. It was developed in 2003 as a method of similarity searching. It is based on the work of Scott Preece, who suggested it in his 1981 dissertation. As previously mentioned, Preece's research has its foundation in 'spreading activation search,' which has its origins in semantic network theory.

CNG uses “a term-document matrix (TDM), which is essentially a weighted lookup table of term frequency data for the entire document collection” (Ceglowski et al., 2003). It figures out which words might be shared between documents that contain the keyword in a query and documents that do not, and their relative importance. The TDM is represented as

> a bipartite graph of term and document nodes where each non-zero value in the TDM corresponds to an edge connecting a term node to a document node.… every term is connected to all of the documents in which the term appears, and every document has a link to each term contained in that document. The weighted frequency values in the TDM correspond to weights placed on the edges of the graph. We call this construct a contextual network graph. (Ceglowski et al., 2003) [as illustrated in Figure 7]

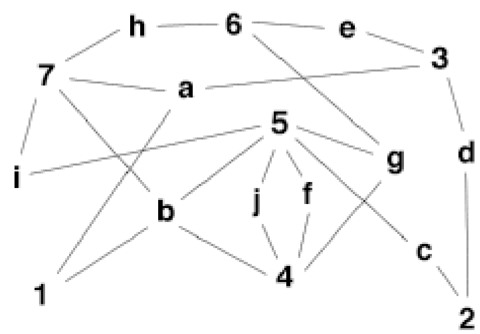


Figure 7. Sample contextual network graph indicating connections between documents and content terms (Ceglowski et al., 2003).

Each edge in a contextual network graph has a strength assigned. The strength of the edge depends on the choice of local and global term weighting scheme used in generating the TDM. All edge weights must fall in the interval (0, 1). The collection, represented as a graph is searched by energizing a query node. The energy spreads to other nodes along the edges of the graph based on a set of rules.

> The total energy deposited at any given node in the graph will depend both on the number of paths between it and the query node, as well as the relative strength of the connections along those paths. This corresponds to the intuition that documents that share many rare terms are likely to be semantically related. (Ceglowski et al., 2003)

And “a query on a particular keyword may still reach a document that does not contain the word itself, but is closely linked to other documents that do” (Ceglowski et al., 2003).

Table 2 typifies the major retrieval methods in terms of lexical, morphological, syntactic and semantic concerns.

Table 2

Characteristics of the Major Information Retrieval Methods in Terms of Lexical, Morphological, Syntactic and Semantic Issues

Linguistic Level	Boolean Retrieval Models	Statistical Retrieval Models	Linguistic Retrieval Models
Lexical	Stop word list	Stop word list	Lexicon
Morphological	Truncation symbol	Stemming	Morphological analysis
Syntactic	Proximity operators	Statistical phrases	Grammatical phrases
Semantic	Thesaurus	Clusters of co-occurring words	Network of words/phrases in semantic relationships

Selecting Features (Indexing Model)

The main concern, when creating an index, is feature selection. Which terms should be used to index (describe) a document? How will you represent documents in your information retrieval system? Index terms should be rich enough to cover your entire document corpus but not so broad that non-context bearing words (common words that typically appear in a language, often placed in a ‘stop list’ of words to excluded from an index) are included.

Words vs. Terms vs. "Concepts"

Simple indexing is based on words or word stems. More complex indexing could include phrases or thesaurus classes. Concept-based retrieval is often used to imply something beyond word indexing. In virtually all systems, a concept is a name given to a set of recognition criteria or rules. Words, phrases, synonyms, linguistic relations can all be evidence used to infer presence of the concept (see Figure 8).

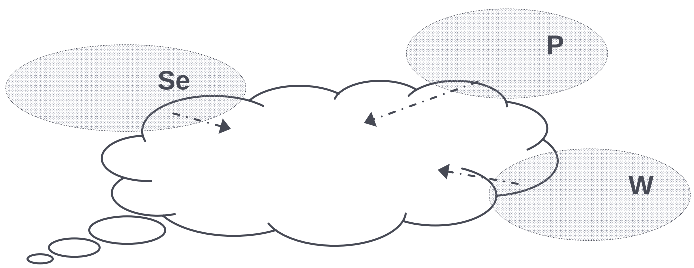


Figure 8. Possible concept formation.

Finding Context

Context is that which surrounds, and gives meaning to, something else. The root meaning comes from the Latin word *contextus*: to join together, to weave.

1. Linguistic context -- The part of a text or statement that surrounds a particular word or passage and determines its meaning.
2. The circumstances in which an event occurs; a setting.[3]

There are two general schools of thought in relation to finding context in documents: statistical and semantic. The statistical approach determines context by looking at statistical patterns within documents and across documents in a collection. A variety of techniques examine word frequencies, word co-occurrence etc. The other is the semantic approach. Context is

[3] Dictionary.com

determined by a complex network of relationships, that relates words and therefore documents together.

Keyphrase Indexing

Keyphrases can have multiple applications: (1) for categorization and summarization of a journal article. The reader can swiftly determine whether the given article is pertinent. (2) Indexing - they enable the reader to speedily find a relevant article when the reader has a particular need. (3) Precision - labeled keywords enable the reader to make a search more precise. Documents matching a keyword produce a smaller, higher quality set of documents than a full text search of a collection.

N-gram Language Model

An n-gram is a string of n, usually adjacent, characters extracted from a section of continuous text. The n-gram language model: Given a corpus of text, the *n*-grams are the sequences of *n* consecutive words that are in the corpus. N-grams are defined as a continuous or non-continuous sequence of character or words, occurring in some proximity to each other in a corpus. Continuous n-grams occur directly next to each other in a corpus while non-continuous n-grams or positional n-grams, are n-grams located within a specified proximity of each other rather than directly next to each other.

Phrases in information retrieval have been researched as a means of improving retrieval effectiveness. Multi-word terms usually contain more meaningful information than a single word. Search engine indexes often contain single words, but not multi-word phrases. “The use of phrases as part of a text representation or indexing language has been investigated since the early days of information retrieval research.... Certainly, there has always been the feeling that phrases, if used correctly, should improve the specificity of the indexing language and, consequently, the quality of the text representation” (Croft et al., 1991). Salton suggests "term phrases instead of single terms. A phrase is likely to be more specific than a single term, and more specific content descriptions may lead to better discrimination between useful and extraneous items, hence increasing the search precision" (Salton, 1998). “These results have been very mixed, ranging from small improvements in some collections to decreases in effectiveness in others” (Croft et al., 1991).

> A phrase can be defined as a concatenation of two or more words which must occur in text separated only by white space and do not range across paragraph or sentence bounds. As a unit of information, a phrase is more content bearing than the sum of its constituent words and it is precisely because it is a richer representation of content than single words, phrase-based representations of document and query content can lead to improvements in retrieval effectiveness (Smeaton & Kelledy, 1998).

A phrase is a textual unit usually larger than a word but smaller than a full sentence. A statistical phrase is defined by constraints on the number of occurrences and co-occurrences of its component words and/or the proximity between occurrences of components in a document (Croft et al., 1991). Using phrases in an indexing scheme suggest several advantages according to Caropreso et al. (2000):

- Phrases come closer than individual words to expressing structured concepts;
- Phrases have a smaller degree of ambiguity than their constituent words;
- Using phrases as index terms, a document that contains a phrase that occurs in the request would be ranked higher than a document that just contains its constituent words in unrelated contexts; phrase extraction is easier with current natural language processing technology

Statistical vs. Syntactic

Mitra, Buckley, Singhal, and Cardie (1997) noted that phrases were found to be useful indexing units in research using Text Retrieval Conferences data for performance evaluations of IR systems. In their review, they considered statistical and syntactic noun phrases.

1. Statistical phrases defined as any pair (or triple, quadruple, etc.) of non-function words that occur contiguously often enough in a corpus constitute a phrase.
2. Syntactic phrases as any set of words that satisfy certain syntactic relations or constitute specified syntactic structures make up a phrase. Syntactic phrases capture actual linguistic relations between words rather than the simple juxtaposition of words, and are expected to be semantically more meaningful (Mitra et al., 1997).

They found at the time, TREC retrieval effectiveness using single terms increased over a five year period, but any added effectiveness due to statistical phrases was down from 7% to only around 1%.

Smeaton & Kelledy (1998) reviewed the series of experiments comparing the effectiveness of document retrieval based on using linguistic phrases, done by “Fagan and later repeated by Mitra et al. using a larger data set and improved term weighting,” where Mitra concluded that adding statistical phrases to retrieval only improves performance by a small amount which was “contrary to Fagan’s findings, when phrases only (i.e. no single word terms) are used then syntactic phrases perform better than their statistical equivalent” Smeaton & Kelledy (1998).

Document-Keyphrase Association

There are two dominant approaches to associating keyphrases with documents: keyphrase assignment and keyphrase extraction. Keyphrase assignment (also known as text categorization) uses a controlled vocabulary in the keyphrase selection process ensuring that “similar documents are classified consistently, and documents can be associated with concepts that are not explicitly mentioned in their text. However, there are also disadvantages: potentially useful keyphrases are ignored if they are not in the vocabulary; and controlled vocabularies require expertise and time to build and maintain, so are not always available" (Jones, 2001).

Keyphrase extraction, the text of a document is analyzed and the most appropriate words and phrases that it contains are identified and associated with the document. Every phrase that occurs in the document is a potential keyphrase of the document. This approach does not require a predefined vocabulary, and is not restricted to the concepts in such a vocabulary. However, the keyphrases assigned to each document are less consistent, and it is not easy to identify the 'most appropriate' words and phrases.

Keyphrase Extraction

Techniques for keyphrase identification and extraction include syntactic, statistical and manual. Syntactic techniques select words by their parts-of-speech and usually extract sequences of nouns. Noun phrases (NPs) are believed to contain rich semantic information. "Statistical methods rely on frequency counts and measures of word relatedness such as co-occurrence" (Pickens & Croft, 2000). Statistical analysis ignores parts of speech. A stop-word list is often used to avoid finding phrases made up of function words (e.g. of, the, and, then). Keyphrase determination process generally means ranking the candidates based on co-occurrence criteria and then selecting the top-ranked phrases. Two common criteria for ranking candidates are simple frequency (normalized by the length of the document), and TF·IDF measures, which compare the frequency of a phrase in the document with its rarity in general use. Smeaton & Kelledy (1998) stated the "majority of the work reported in the IR literature on exploiting phrases for document retrieval, determining the phrases used to represent documents and queries is done as a static, automated process, with no user involvement."

Phrase Weighting

Caropreso et al. (2000) noted that Furnkranz, Mladenic and Grobelnik extracted n-grams of length up to 5 in their research on statistical phrases. Mladenic and Grobelnik found that n-grams of length up to 4 are beneficial while 5-grams are not. Furnkranz saw performance improvements with n-grams of length up to two and three and also confirmed that larger n-grams were not very useful. Phrases and words share the same correlation to documents. The more one is seen in the collection, the more it is found in relevant documents.

> There are differences, however. The first difference is the distribution of the data points. In the word plot, most words are found in the mid-to-high range frequencies, whereas with phrases, most phrases are found in the mid-to-low range. The second difference is that when the phrase values do flatten out at very low frequencies, they become somewhat erratic, and more so than words. Even with these differences, it is remarkable that phrases and words have similar relevance values at similar IDF values. (Pickens & Croft, 2000)

Interface Design and Relevance Feedback

The design principles of Ben Shneiderman (cited in Hearst, 1999) are guidelines for creating a human interface. A user interface should offer informative feedback to users about the query-document and document-documents relations. It should reduce working memory load. Information access is a dynamic and iterative process, the goals of which shift and change as information is encountered.

Information access interfaces provide the user with mechanisms to remember selections made during the search process. "This includes suggestions of related terms or metadata, and search starting points including lists of sources and topic lists" (Hearst, 1999). Relevance feedback offers users a method of communicating, to search engines, which documents are relevant or useful, and which are not. The results presented to the user via this process should reflect the feedback provided, for example 'more like this' and 'iterative searches'.

Relevance Feedback

An important part of the information access process is query reformulation, and a proven effective technique for query reformulation is relevance feedback. In its original form, relevance feedback refers to an interaction cycle in which the user selects a small set of documents that appear to be relevant to the query, and the system then uses features derived from these selected relevant documents to revise the original query. This revised query is then executed and a new set of documents is returned. Documents from the original set can appear in the new results list, although they are likely to appear in a different rank order.

Visualization

Information visualization attempts to provide a visual interpretation of an information spaces. "Visualization of inherently abstract information is more difficult, and visualization of textually represented information is especially challenging. Language is our main means of communicating abstract ideas for which there is no obvious physical manifestation" (Hearst, 1999).

According to Murtaugh (2003) visualization is a natural evolution of the user interface. Visualizing stored information and data whether in databases, unstructured or semi-structured collections, is significant for the following reasons:

1. It allows the user to have some idea before submitting a query as to what type of outcome is possible. Hence visualization is used to summarize the contents of the database or data collection (i.e. information space).

2. The user's information requirements are often fuzzily defined at the outset of the information search. Visualization helps the user in their information navigation, by pointing out related items, by displaying relative density of information, and by inducing a (possibly fuzzy) categorization on the information space.

3. Visualization can help the user before an interaction with the information space, and during this interaction.

Context

Hearst (1999) discusses techniques for placing a document in context:

1. Document Surrogates

A ranked document list is the typical output for a query, in order of their computed relevance to the query. The document list could contain "the document's title and a subset of important metadata, such as date, source, and length of the article. In systems with statistical ranking, a numerical score or percentage is also often shown alongside the title, where the score indicates a computed degree of match or probability of relevance. This kind of information is sometimes referred to as a document surrogate" (Hearst, 1999).

The document surrogate could be shown in short format, as several lines of text or detailed (summary or abstract) format. "In most interfaces, clicking on the document's title or an iconic representation of the document shown beside the title will bring up a view of the document itself, either in another window on the screen, or replacing the listing of search results" (Hearst, 1999).

2. Query Term Hits within Document Content

Highlighting terms matching the user query as part of the display mechanism, draws "the user's attention to the parts of the document most likely to be relevant to the query. Highlighting of query terms has been found … a useful feature for information access interfaces [landauer93], [march] (p.31)" (Hearst, 1999). Terms are seen in the 'context' of adjacent words.

KWIC, KWOC and KWAC

A concordance is usually an alphabetical list of all important or meaningful words in a book. Keyword-in-context (KWIC), also referred to as highlighting, is an indexing technique developed in 1958 by Hans Peter Luhn. Concordances generated via computer can be KWIC (see Table 3), keyword-out-of-context (KWOC) (see Table 4) and KWAC (Keyword-alongside-context) (see Table 5). The later two are variations of KWIC.

"A KWIC extract shows sentences that summarize the ways the query terms are used within the document. This display can show not only which subsets of query terms occur in the retrieved documents, but also the context they appear in with respect to one another" (Hearst, 1999). KWIC applications typically extract full sentences, fragments or groups of sentences from a document that contains a matching query term. "The KWIC facility is usually not shown in Web search result display, most likely because the system must have a copy of the original document available from which to extract the sentences containing the search terms. Web search engines typically only retain the index without term position information. Systems that index individual Websites can show KWIC information in the document list display" (Hearst, 1999).

Table 3

Example KWIC Display

KWIC
Running can be a very tiring sport. Adequate hydration…
a very tiring sport. Adequate hydration is important for a successful marathon…
Adequate hydration is important for a successful marathon…

Keyword-out-of-context (KWOC) application will not display a keyword in the context of the words preceding and following it, instead the keyword is displayed followed by the context for the keyword. Each keyword in a document will occupy the first (left) position.

Table 4

Example KWOC Display

KWOC
Runing Running can be a very tiring sport. Adequate hydration…
sport Running can be a very tiring sport. Adequate hydration…
hydration. Running can be a very tiring sport. Adequate hydration…

Keyword-alongside-context (KWAC) application also displays a keyword first but it does not show the context of the words preceding and following it, only the words following.

Table 5

Example KWAC Display

KWAC
Runing can be a very tiring sport. Adequate hydration…
Adequate hydration is important for a successful…
hydration is important for a successful marathon…

Web Information Retrieval Evaluation

Evaluate the results of interactive searches such as the number of documents followed to estimate the amount of retrieved information by user, time spent on traversing the result set as a measure of search efficiency, and the click distance, which is physical distance between finding relevant documents (Zamir and Etzioni, 1999). The trouble with the number of documents followed is that it does not consider the degrees of relevance in followed documents. Users follow links based on documents in the retrieved set or find the relevant documents by reading the document summary without following a link to examine a full document. Time spent is another problematic measure. Hard to differentiate between time spent on network delays, document evaluation and traversing links. Click distance for ranked lists and cluster displays have different interpretations and implications so it is difficult to accurately compare the cost of finding documents between systems.

Web Search Behavior Patterns

Jansen, Sink, Bateman & Saracevic (cited in Choo, Detlor and Turnbull, 1998) noted in their research that searchers tend to use very short queries; they exert minimum effort in evaluating or refining searches. Most searchers browsed only the first page of search results (the Google search engine typically displays only ten results per page) and do not engage in search refinement or query reformulation (Silvertein et al., 1998).

Different techniques are implemented to search for and present documents, in current search engine but most return a ranked list of documents. Chen et al. (cited in Wu et al., 2002) found users rarely examine more than the top 40 retrieved documents. There is a ‘sequentiality’ effect whereby “users are satisfied if the relevant results are shown first, and users are dissatisfied when they see irrelevant ones among top retrieved documents” (Wu et al., 2002).

Marchionini suggested that interfaces for end-user information seeking should reflect the assumption that “End users want to achieve their goals with a minimum of cognitive load and a maximum of enjoyment.... humans seek the path of least cognitive resistance and prefer recognition tasks to recall tasks; most people will trade time to minimize complexity” (Anick, 1997).

According to Nielsen and Loranger (2006), "the search engine result page is usually referred to as a SERP ... Most users don't see more than one SERP per query" In their study 93 percent of searches only visited the first SERP, "which usually held ten search results plus a number of ads. In only seven percent of cases did users [visit] a second SERP, and the number who visited three SERPs for a single query were too small to provide a firm estimate, but it was likely less than one percent." They found that users usually "didn't even bother reviewing the entire page. Only 47 percent of users scrolled the first SERP, which means that 53 percent saw only those search URLs that were 'above the fold.' (Originally a newspaper term, 'above the fold' refers to what part of a Web page is visible on a screen without scrolling.) On the most widely used search engine, Google, users can only see four or five results above the fold, if they view the page in the most common screen resolution of 1024-by-768 pixels, like those in our study [Nielsen and Loranger]" (p. 39).

Keyphrase Browsing Interfaces

Browsing applications support interactive navigation of index terms and provide direct access to the original documents via the index terms. Terms are presented to users in ways that allow them to either ‘drill down’ from a shorter, more general term to longer, more specific ones or to navigate from one term to other related ones via graphical interfaces.

A major advantage of these systems is that they provide users with index terms instead of requiring users to devise the terms themselves, an especially difficult task for an information seeker looking for information about an unfamiliar domain. Because the terms are extracted from documents, phrase-browsing systems can be easily integrated with full-text search and are complementary to standard information retrieval systems. Browsing systems are distinct from organizational systems based on ontologies that do not correspond directly to collection content.

IBM’s Textract mining system, identifies proper names and technical terms. Textract also automatically identifies term relations; including named relations such as similar-to and unnamed relations based on mutual information. The browsing interface is a graphical display of phrases obtained from the document, connected by named and unnamed relationships.

DR-LINK system (Liddy and Myaeng 1993) is a phrase browsing application that used linguistically motivated units such as noun phrases and proper names to help users select documents that meet a specified information need.

Intell-Index, a phrase browsing system designed for testing the effectiveness of natural language processing techniques for automatic identification of index terms.

Kea algorithm automatically extracts keyphrases documents using simple lexical processing. It uses index document frequency (TF.IDF) and where in the document the phrase first occurs. Applications using Kea include:

Kniles is a Web-based system for inserting topic-based hypertext links into existing, large-scale digital library collections. The links allow the browsing of document collections that do not already have embedded links. The hypertext links inserted by Kniles are based on author keywords, and on keyphrases automatically extracted by Kea.

Phind is a phrase browsing interface that has been integrated into the Greenstone digital library software. Searchers browse through the phrases that occur in a document collections in a hierarchy.

Phrasier is designed to assist users in making sense of documents read online. Phrasier provides four views: 1) the topic overview presents a list of terms that characterize overall document content; 2) the topic location view shows users where in a document a chosen phrase frequently occurs; 3) the document content view allows the user to control the degree of differentiation between the keyphrases and document content; and 4) the summary view displays sentences that have been identified as significant in the document.

Search Tactics

Foraging

> Foraging for information on the Web and foraging for food share common features: both resources tend to be unevenly distributed in the environment, uncertainty and risk, characterize resource procurement, and all foragers are limited by time and opportunity costs as they choose to exploit one resource over another (Sandstrom 1994). Successful foragers are those who adopt strategies that maximize their harvest rates and their chances of survival. … Information foraging refers to activities associated with assessing, seeking, and handling information sources. (Choo et al., 1998)

Kearns (2003) discusses foraging tactics, which move a search forward, and foraging strategies, which are the search methods used to locate data. Foraging strategies are observed “in action in everyday situations” (p.126). Several strategies are defined by Kerns. This research describes only two as likely avenues of relevance determination: browsing and grazing. “Any of the strategies can be used to forage for whatever relevance is being sought. The strategy employed depends on one's depth within the search process and the specificity of information that is required" (p.126).

Browsing

Gutwin, C., Paynter, Gordon., Witten, Ian., Nevill-Manning, Craig., & Frank, Eibe (1999) said users cannot easily determine what is in a collection, how well a particular topic is covered, or what kinds of queries will provide useful results. In browsing, the user interacts with a collection and carries out searches, without having in mind a specific document or document set. Browsing is characterized by a succession of queries (an interactive activity) rather than a single search - that is, people determine their next query partly based on the results of the previous one (relevance feedback).

"Browsing takes place in many situations in which related information has been grouped together according to subject affinity" Chang and Rice (cited in Choo et al., 1998) define browsing as 'the process of exposing oneself to a resource space by scanning its content (objects or representations) and/or structure, possibly resulting in awareness of unexpected or new content or paths in that resource space."

Choo et al. (2000) stated that Marchiononi observed three general types of browsing (see Table 6):

Table 6

Three Types of Browing and Related Task Questions (Gutwin et al., 1999)

Task Type	Questions
Collection evaluation	What's in this collection? What topics does this collection cover?
Subject exploration	How well does this collection cover area X? What topics are available in area X?
Query exploration	What kind of queries will succeed in area X? How can I specialize or generalize my query?

Directed browsing occurs when browsing is systematic, focused, and directed by a specific object or target. Examples include scanning a list for a known item, and verifying information such as dates or other attributes. Semi-directed browsing occurs when browsing is predictive or generally purposeful: the target is less definite and browsing is less systematic. An example is entering a single, general term into a database and casually examining the retrieved records. Finally, undirected browsing occurs when there is no real goal and very little focus. Examples include flipping through a magazine and 'channel-surfing' (Choo et al., 2000)

O'Connor (2003) describes browsing as searching for "an unknown among uncertainties" (p.128). Users employ various tactics in an information search. The discovered information that solves a problem or answers the question is 'relevant.' Although commonly viewed as a 'browsing' application, I hesitate to unequivocally characterize the proposed search engine application as one designed solely to facilitate 'browsing,' due to the uncertainty factor contained in many definitions of browsing. The grazing search tactic could be pertinent.

Grazing

O'Connor (2003) defines the foraging strategy of grazing as searching for relevance "among knowns or certainties" (p.129). Searchers employ various tactics in an information search. The discovered information that solves our problem or answers the question is 'relevant.' "Similar to browsing, grazing is foraging in a space where evaluation and supply are not issues. … One grazes when one forages in an area where a subject is determined, as in the library stacks" Kearns (2003). She also distinguished a grazing tactic as one that "implies that some other force is responsible for ensuring supply" (p.129). The other force supplying data for this application is the search engine that provides the results of the initial user query.

Browsing seems a suitable foraging strategy to employ when searching on a topic that is unfamiliar while grazing seems suited to searching on a familiar topic.

Conclusion

Associative search implemented via spreading activation has been used in conjunction with the Boolean, vector and probabilistic models of IR. The CNG technique straddles the statistical and semantic schools (see Table 3). Keyphrase applications can be good tools for information access, showing context and allowing concept mapping. Their utility should be improved with semantic analysis techniques.

CHAPTER 3

CONCEPTUAL FRAMEWORK

This chapter describes the ideas underlying the construction of a visual model for a semantic cloud and its relationship to searching the domain of Web pages and retrieved subsets of those pages. I developed a conceptual visual model served as a framework to consider the user's information need, their query formulation, search strategy and the semantic relationships in the information space retrieved for their search. The differences of keyword search versus a semantic search, the study of semantics, and concepts and categories as they relate to semantic mapping is briefly mentioned.

Keyword Search Engine

A keyword search engine when given a user-input 'query' (a list of words) attempts to find documents matching the query. Keyword search works best when a query is as specific as possible. The user should have a very good idea about what s/he is searching for, ideally knowing the keywords used in the document collection. The major disadvantage of keyword search: it finds only documents containing the words the user queries for unless it incorporates other techniques.

Semantic Approach

A Semantic Web Search

Including metadata in documents would allow us to search the World Wide Web semantically. A semantic search engine should discover if two documents are similar even if they do not have any specific words in common and reject documents that do not have any commonality – find documents with similar concepts instead of similar words. Semantic mapping is an important concept in the evolution of the Web.

Semantics – The Meaning of Meaning

Semantics is a philosophical discipline, which deals with the nature and the organization of reality. In general, semantics (from the Greek semantikos, or "significant meaning," derived from sema, sign) is the study of meaning, in some sense of that term. In linguistics, semantics examines the ways in which words, phrases, and sentences can have meaning (see figure 8). Semantics is a subfield of linguistics that is traditionally defined as the study of meaning of (parts of) words, phrases, sentences, and texts.

Humans require words (or at least symbols) to communicate efficiently. The mapping of words to things is only indirectly possible. We do it by creating concepts that refer to things. The relation between symbols and things has been described in the form of the meaning triangle (see figure 9):

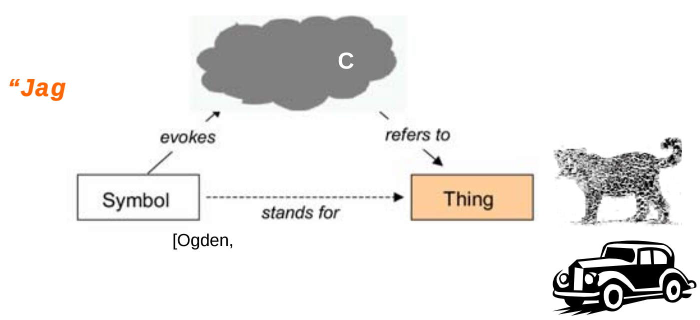


Figure 9. Meaning triangle (Klimesch, 1994).

The semantic triangle refers to the three part association among a referent, a reference, and a symbol. The referent is the object. The reference is the thought of the object. The symbol is the word for the object. The semantic triangle permits uncertainty; one person's understanding is not the same as another.

What are Concepts?

"Human concepts are probably…like hooks or nodes in a network from which many different properties hang. The properties hanging from a node are not likely to be all equally accessible; some properties are more important than others, and so may be reached more easily or quickly… Thus a concept would be a set of interrelationships among other concepts…everything is defined in terms of everything else…like a dictionary" (Collins & Quillian, 1972).

> Concepts are not learned or represented in isolation. Instead, they are born within a Web of concepts, and the relationships between concepts can influence the representation of individual concepts, as well as the classification of individual exemplars as being members of one concept or another. (Rosch and Lloyd, 1978)

Features of a Concept: Defining vs Characteristic

Defining features are features absolutely essential to the concept. Defining features appear at the top of each feature list. Characteristic features are features that are common but not essential to the meaning of a concept. Characteristic features appear at the bottom of each feature list. "We discover that once a good basic ranking scheme is being used, the use of phrases does not have a major effect on precision at high ranks. Phrases are more useful at lower ranks where the connection between documents and relevance is more tenuous" (Mitra, 1997).

Mental Categories

From the user's perspective, a mental category is a grouping mechanism, a way to bring together items or concepts through some unifying characteristic(s) or attribute(s) (Withrow, 2003). According to Rosch (1978) natural categories occur in the real world of our experience and have a complex internal structure. Natural categories have fuzzy boundaries; category membership is a matter of degree. The central or core instance of a category is called the prototype.

Concept Mapping

A particularly good way to organize information about a problem or subject is to construct a 'concept map' shown in Figure 10. Construction of concept maps helps us pull together information we already know about a subject and understand new information as we learn. Concept maps consist of nodes and lines. The node represents a concept and is drawn with a circle around the term or concept. Lines between nodes show inter-concept relationships. The label on the line tells how or in what way the concepts are related.

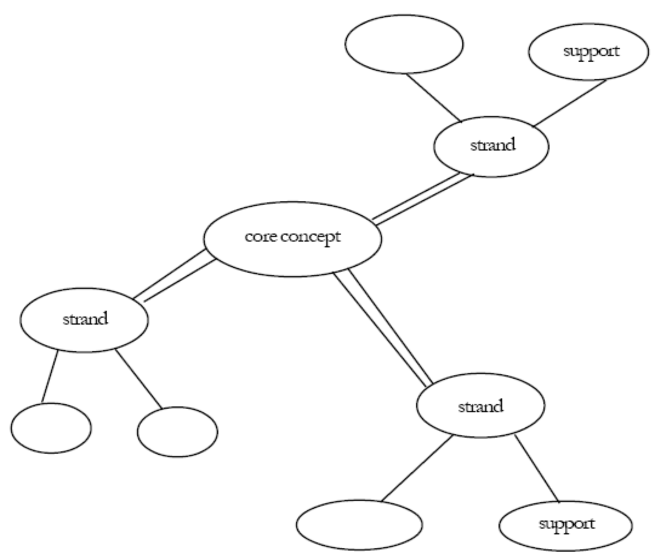


Figure 10. Concept map components.

Semantic Mapping

Semantic mapping is a strategy for graphically representing concepts; the construction of a two dimensional representation of concept relationships. Semantic maps (aka mind maps, concept maps, graphic organizers etc. shown in Figure 11) portray the schematic relations that compose a concept. General or broad concepts are typically found at the top of a concept map while more specialized concepts are at the bottom. The map assumes that there are multiple relations between a concept and the knowledge that is associated with the concept.

There are three components to a semantic map:

1. Core question or concept (prototype): a keyword or phrase that is the main focus of the map.
2. Strands: subordinate ideas that help explain or clarify the main concept.
3. Supports: details, inferences and generalization that are related to each strand. Supports clarify the strands and distinguish one strand from another.

The sequence of semantic mapping is as follows:

1. Select a word central to the topic.
2. Display the target word.
3. Generate as many words as possible that relate to the target word.
4. Write the generated words in categories.
5. Label categories.
6. From this list, construct a map.

Ideas and concepts for fish

- Live in water
- Have gills
- Have fins
- Make good pets
- Eat plants
- Used for food
- Are alive

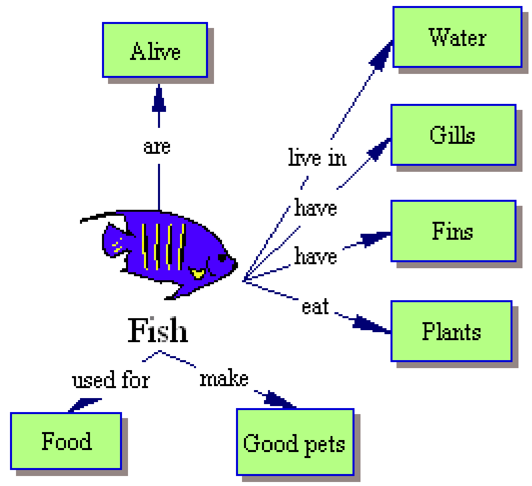


Figure 11. Semantic map for fish.

Semantic and Network Cloud

In the literature on the semantic web, the term ‘semantic cloud’ is often seen but I could not locate a definition. The origin of the term could lie in telecommunications’ ‘network cloud.’ A *network cloud* or *cloud* is the unpredictable part of any network through which data passes between two end points. Clouds exist because between any two points in packet-switched networks, the physical path on which the packet travels can vary from one packet to the next (see Figure 12). In circuit-switched networks, the specific circuit that is set up can vary from one connection to the next. The variability and amorphous form of the Internet can be represented via the image of a cloud. Hyperlinks between Websites can therefore have semantic implications.

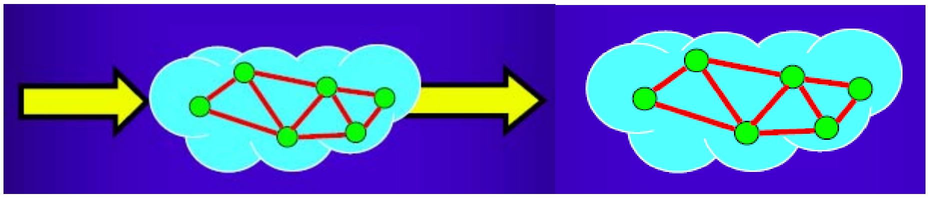


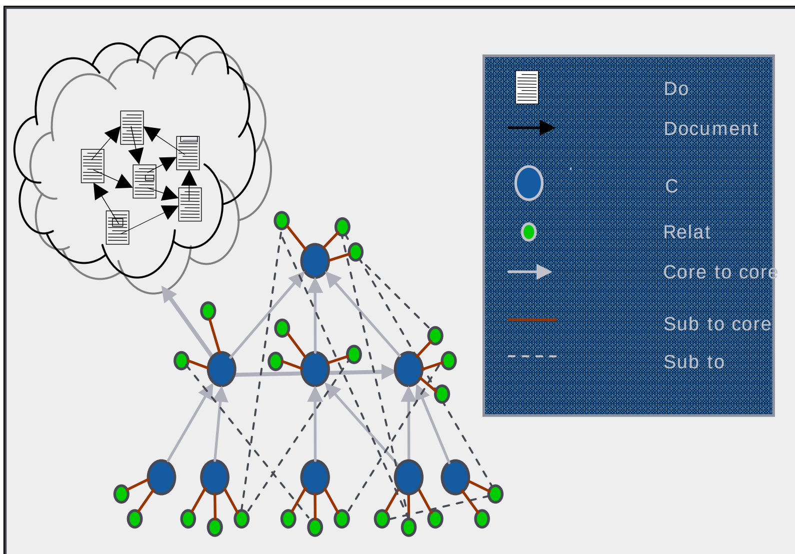
Figure 12. The unpredictable part of the network where data passes through two ports.

Atomic Analogy

The de-compositional perspective of meaning theorizes that the meaning of words can be analyzed by defining meaning atoms or primitives, which establish a language of thought. Guillermo Oyarce's conceptualization of the 'semantic cloud' (G. Oyarce, personal communication, 2004) reflects a relational nature of concepts. In his view, problems of information representation and controlled vocabularies using keyword descriptors correspond with an atomic analogy in the concept space. A concept is atomic and not isolated but together with other concepts form part of the conceptual space thus each concept is understood as supporting or being part of other concepts. The concept space is built from simple to complex, and that is how learning takes place. Specific concepts form a cloud-like structure in the concept space where some atomic supporting concepts are related to each other at different degrees, thus forming subspaces of more or less density. Concepts are not discrete elements in the sense that they have no clear boundaries where one ends and another starts. Concepts seem to be of a fuzzy nature.

Visual Model

A visual model was developed (see Figure 13), by the research, as a metaphor that could apply to the domain of Web pages and a retrieved subset of those pages as an information space. Words in the space will be concepts that will have core and sub-core (meaning that they have a partial semantic association) relationships. Core concepts for a document or set of documents will have relations to other core concepts and sub-core concepts will be related to other sub-core or core concepts in the information space. The information space created by the results of a query will arise from the document to document links or relationships present in the retrieved result set.



The “semantic cloud” as a network of concepts: the path or activation connecting concepts varies for each user.

Figure 13. Visual illustration of atomic conceptual model of semantic cloud (Johnston, 2006).

CHAPTER 4

PILOT TEST

This chapter describes the prototype engineering of a keyphrase browsing application based on a contextual phrase analyzer engine and the results of a heuristic evaluation of the prototype.

Prototype Development Background

Oyarce (2005a) in "A Contextual Phrase Analyzer" describes "a contextual phrase analyzer engine builds a contextual tree at different levels of specificity from existing data, e.g. data extracted from one or more documents, thus synthesizing an information universe [or information space] and reducing the cognitive volume [or cognitive load of the user of an information retrieval (IR) system] to process."

One implementation of a contextual phrase analyzer is the kontextual interface search system, KISS, a contextual interactive support system described in Oyarce (2005a) that "provides techniques for analyzing and evaluating document sets … provide a context for words, combinations of words, and/or concepts present in the document sets."

The original impetus for the development of the search tool was an effort to create a dynamic Web version of the Kontextual Interface Search System (KISS) (Oyarce, 2004). KISS as implemented in the prototype is a straightforward Boolean matching system; it only finds documents that contain a selected keyphrase.

Functional Minimum Prototype

The functional minimum for this search tool was developed via keywords in context prototype without employing any semantic techniques. A user enters a query into a search form; the search is submitted to the AltaVista search engine. Out of the results retrieved by AltaVista, a subset is created (the subset becomes the data corpus): the first 50 URLs (uniform resource locators) are extracted from the AltaVista results and crawled (each page visited and a text version of the page without any HTML code created). Searchers select keywords in an iterative interactive process to locate documents that are most relevant to their area of interest (see Appendix C for a process illustration).

Keyphrase Index Creation

Indexes of variable length keyphrases are constructed from the subset. These indexes are used to create 'windows' or 'views' of the most frequently occurring words in the subset. Criteria for creating variable length keyphrase indexes: unigrams or uniterms (1-grams) single words in the corpus. A bigram (2-gram) is a sequence of two words. A trigram (3-gram) is a sequence of three words and a tetragram (4-gram) a sequence of four words. The tf x id weighting scheme was deemed appropriate for index creation.

The semantic information space is navigated by queries instead of following explicit document links. The system does not allow the user to enter new search phrases while exploring the retrieved set. Query formulation or reformulation is done only with system determined search terms. After query selection, the program will give the user a list of documents matching their selection with the chosen keywords in 'context,' context being a range of words before and after the keyword or phrase.

Prototype Analysis

Human computer interface (HCI) evaluation can be done by experts or users. User-based evaluation basically is evaluation through user participation i.e. evaluation that involves the people for whom the system is intended. These methods can be conducted in the laboratory and/or in the field. Expert-based evaluation uses HCI experts. Each expert separately (or in groups) reviews an interface and categorizes and justifies problems based on a short set of heuristics (rules of thumb). The heuristics come from known/standard cognitive principles or empirical results.

Pros:

- Uses experts; gives reviewers common language; reasonably fast
- Expert analysis methods can be used at any stage in the life cycle
- Expert analysis methods are relatively cheap

Cons

- Heuristics used in expert analysis methods do not assess the actual use of the system by a user.

Heuristic Evaluation

Heuristic evaluation (HE) as defined by Jakob Nielsen (1994) "is a usability engineering method for finding the usability problems in a user interface design so that they can be attended to as part of an iterative design process." Experts scrutinize the interface and its elements against established usability principles. The experts should have some background knowledge or experience in HCI design and usability evaluation. Three to five experts are considered to be sufficient to detect most usability problems.

Experts are provided with the proper roles (and scenarios to use) to support them when interacting with the system/prototype under evaluation. They then evaluate the system/prototype individually. This is to ensure an independent and unbiased evaluation by each expert. They assess the user interface as a whole and also the individual user interface elements. When all the experts are through with the assessment, they come together and compare and appropriately aggregate their findings.

Usability Heuristics and Expert Review

There are many "usability heuristics" that can be applied in the design of user interfaces. However they are not necessarily proven and can conflict with one another. For these reasons and in the interest of convenience, expert review (ER) was the method used to evaluate the prototype. Expert review uses people having sufficient experience or an advanced degree in a related discipline (HCI experts) critique a product separately or in groups to determine areas in need of improvement. In ER the experts/reviewers use no guidelines, scenarios or tasks. Therefore, it may be difficult to standardize or categorize rationale for design changes. The iterative design process for the prototype used ER. I tested four different versions (the original design plus three redesigns) before halting expert testing. All versions were tested separately with three users with varying degrees of exposure to usability heuristics. Although HE was not used to evaluate the prototype, I looked at the prototype design and problems identified via ER to determine how well they correspond to the ten usability heuristics developed by Nielsen (1994) that could be used in HE (see Figure 14):

- *Visibility of system status:*

The system should always keep users informed about what is going on, via appropriate feedback within reasonable time.

- *Match between system and the real world:*

The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms [use simple and natural language]. Follow real-world conventions, making information appear in a natural and logical order.

- *User control and freedom:*

Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

- *Consistency and standards:*

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

- *Error prevention:*

Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

- *Recognition rather than recall:*

Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate e.g., visible menus for systems functionalities/commands.

- *Flexibility and efficiency of use:*

Accelerators - unseen by the novice user - may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions e.g., shortcuts through function keys and key combinations.

- *Aesthetic and minimalist design:*

Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

- *Help users recognize, diagnose, and recover from errors:*

Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

- *Help and documentation:*

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Developer Analysis and ER Results

Using the keyword indexes in the search interface via a browser was a slow process. This could have been due to several factors, the programming language used for the interface (PHP) or the limitations of the computer the prototype was installed on (Pentium III 800 megahertz with 128 megabytes RAM). The amount of memory allocated to the PHP script had to be increased from the default eight megabytes to 20 megabytes. When tested on a newer computer (Pentium 4 1.3 gigahertz with 512 megabytes RAM) with more memory and processing power, the prototype performance increased, but the decision was made to limit significant keywords in the initial context window to the first hundred for the purpose of testing the prototype. This dramatically sped up the prototype performance. A review of the literature indicates that in the interest of efficiency, the version readied for user testing could further limit the number of keywords seen by the user at one time. An option to see the next set of relevant keywords (in ranked order of most to least significant) could be included. The evaluators were given a brief description of the purpose of the software application, and were then asked to interact with the application, noting any interface features that could be problematic for a general user.

Evaluators identified the following issues in the interface design:

1. Topic exploration was originally an uninterrupted process: single keyword -> two word keyphrase -> three word keyphrase -> four word keyphrase -> retrieves documents. Evaluators wanted to be able to stop and retrieve documents at any stage in the process.

   Result: options to 'get documents' or 'refine search' further.

2. Searchers had to rely on memory to know the keyphrase(s) chosen at each stage. Evaluators wanted to know their selection(s) from the previous stages.

   Result: system tells the user the search terms used in the previous stage.

3. Evaluators wanted the option to initiate a new search or go back to a prior stage after selecting terms.

   Result: only the option to return to Stage 1 or 'start a new search' added.

4. The stages had domain specific labels i.e. choose bigrams, trigrams etc. in stage descriptions. Evaluators felt this would be confusing to users not already familiar with those terms.

   Result: the messages were removed with the intention to add more general messages in later design iteration.

Visibility of system status (system feedback)	**Yes** - users know here they are in the search process
Match between system and the real world (use simple and natural language)	**Yes** - uses terms on buttons and links that should be familiar to users of search systems. **No** - language describing stage of process, uses terms not familiar to a general user, for example.
User control and freedom (undo, redo, exit options)	**Yes** - users can always exit the current search and start over or begin a completely new search. To exit the process completely, they can navigate to another website or close their browser window.
Consistency and standards (words/actions mean the same thing; maintain platform conventions)	**No** - users cannot resubmit at each stage of the search process, they only have the option to refine their search.
Error prevention	**No** - there are no confirmation dialogs for the user chosen query terms; it is assumed the user meant to make the selections.
Recognition rather than recall (user should not have to remember information)	**Yes** - all options for each stage are visible; user is reminded of chosen terms for each stage; instructions are present for each stage.
Flexibility and efficiency of use (options for expert vs. novice; system shortcuts)	**No** - there is no advanced option for an expert versus novice user; it is assumed the interface as designed, will work for all users. Nor is there an option for users to tailor frequent actions
Aesthetic and minimalist design	**Yes** - the design is very minimalist with short messages, no graphics and only a back and white colour scheme.
Help users recognize, diagnose, and recover from errors	**No** - if a user does not select at least one query term, they receive no error message; redirected to stage one.
Help and documentation	**No** - there is no separate help or documentation aside from messages to the user contained within the system; it is assumed system can be used without documentation

Figure 14. Correspondence to Nielson's Ten Usability Heuristics (Johnston, 2006).

CHAPTER 5

RESEARCH DESIGN METHODOLOGY

In chapter 4, a conceptual framework was proposed as a guide for development of the keyphrase browsing application. This chapter covers the following topics: a) areas of research in keyphrase browsing, b) the research question, c) material selection, d) pilot test, e) instrumentation, f) data collection, g) treatment of data, and h) anticipated results.

Areas of Research

Background Issues in Keyphrase Browsing

Several researchers (cited in Wacholder et al., 2001) concluded keyphrase browsing systems had been "adequately evaluated for usability" pointing to the studies of Gutwin et al., Jones and Paynter, Jones, Peñas et al. (cited in Wacholder et al., 2001). The workshop participants also agreed effective strategies were developed for resolving "the tension between efficiency and quality of phrases" (Wacholder et al., 2001). They saw the need for further investigation of information access since it is undetermined "whether providing information access via phrases actually speeds up or otherwise improve the results of the search process" (Wacholder et al., 2001).

The most significant problem in phrase browsing system identified by the researchers was efficiency versus quality:

1. "Development of efficient techniques and methods for navigating and browsing phrases that provide access points to full text documents" and

2. "Development and analysis of techniques" for effective extraction of index terms, "determining which phrases are most useful, and for organizing and classifying terms in ways that that take into account human searching behavior and knowledge of language" (Wacholder et al., 2001).

Other issues for further research were also identified among them

- How to present complex information seekers in ways that is helpful but not distracting;
- What criteria are most suited for evaluating phrase browsing technology?
- How can the list of terms presented to the user be flexibly adapted to user's level of domain expertise and to corpus characteristics? (Wacholder et al., 2001)

Research Question

The general research question in this work is will a searcher find a keyphrase 'browsing' interface a valuable tool in evaluating search engine results?

Expected Outcomes

- Searchers unfamiliar with the domain of a search topic will find the application useful for gaining an overview of the information space.
- Searchers familiar with the domain of a search topic will find the application useful for quickly isolating their area of interest.
- Searchers will not find longer keyphrases as valuable as shorter phrases.

Material Selection

Requirement to support main hypothesis: a document corpus, each link to a document in the set is followed and its text saved; document keyphrase indexes, method to create semantic relations; a method of visualizing the retrieved result set; a group of users; and an instrument for measuring perceived user satisfaction.

A popular measure of data now available electronically is the size of Google's Web index, which as of June 2005 stood at 8.05 billion Web pages, 1.3 billion images, and over one billion Usenet messages - in total, approximately 10.4 billion items[4]. However, this number is only a fraction of the total, when you consider journal articles, news stories, electronic texts, and scientific data sets, most of which are not freely available on the Web. Note that the Web consists of the public or surface Web (fixed Web pages) and the private or the deep Web (usually database driven Websites that create Web pages on demand).

The data pulled in from the search engines are somewhat ‘unstructured.’ This refers to any data, for example Web pages, that does not have categories, keywords, or other descriptors attached to it. Documents can have a built-in structure, for example HTML documents, but it is not always available in a defined format. Metadata attempts to overcome this by providing a structure for unstructured data. This goes by many names - categories, taxonomies, controlled vocabulary, or ontology.

KeyView Phrase Browsing Application

Petroski (cited in O'Connor, Copeland and Kearns, 2003, p.105) said engineering is a human endeavor and, thus, is subject to error. Failure considerations and proactive failure analysis are essential for achieving success.

The phrase browsing application used in the full study, sought to address key issues identified in the pilot test: The ability to stop and retrieve documents at any stage in the process is accomplished via ‘get documents’ function. The user is also able, via navigation links, to immediately start their keyphrase search at a level (unigram, bigram, trigram and tetragram) they specify. Searchers no longer rely on memory to know the keyphrase chosen. The application reports which keyphrases the user selected from the previous search stage. There is an option to

[4] Source: Wikipedia, the free encyclopedia. Retrieved 12/2/2005 from http://en.wikipedia.org/wiki/Google_(search_engine)

initiate a new search or return to a prior search stage after selecting keyphrases. This is accomplished via navigation links and also use of the browser's back button to return to a previous search stage.

I added a ‘related documents’ feature utilizing the contextual network graph technique (which is based on spreading activation), in order to exploit semantic relationships beyond just the local context of matching words. This will locate documents that are similar to the document matching the selected term. Specifically the Search::ContextGraph module, a Perl implementation of CNG search will be used for semantic document association.

As described in four, user enters a query into a search form; the search is submitted to a search engine (in this case the Google Search Engine). Out of the total number of results retrieved by Google, a subset is created (the subset becomes the data corpus): the first one hundred URLs (uniform resource locators) are extracted and crawled. Searchers are then directed to search these results using select keywords in an iterative interactive process to locate documents that are most relevant to their area of interest.

KeyView Search Process

The user first submits a query or search terms to a search engine then waits for the application to extract the first one hundred Web pages in the result set supplied by the search engine (see figures 15-17).

Figure 15. Submit a search to the search engine.

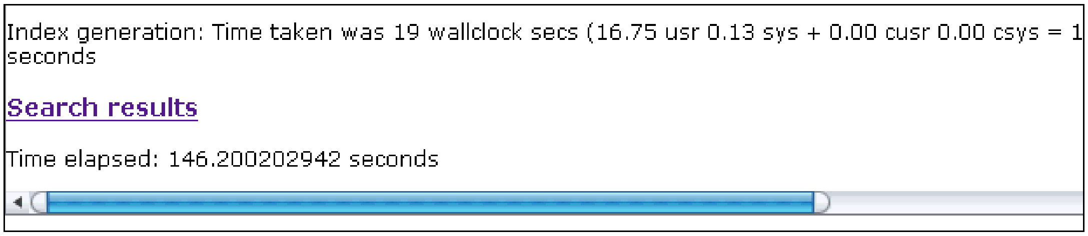


Figure 16. Click on "Search results" to examine the keyphrase(s).

Results matching the searcher’s query are examined via variable length keyphrases. The application begins at the single ‘unigram’ keyword level by default.

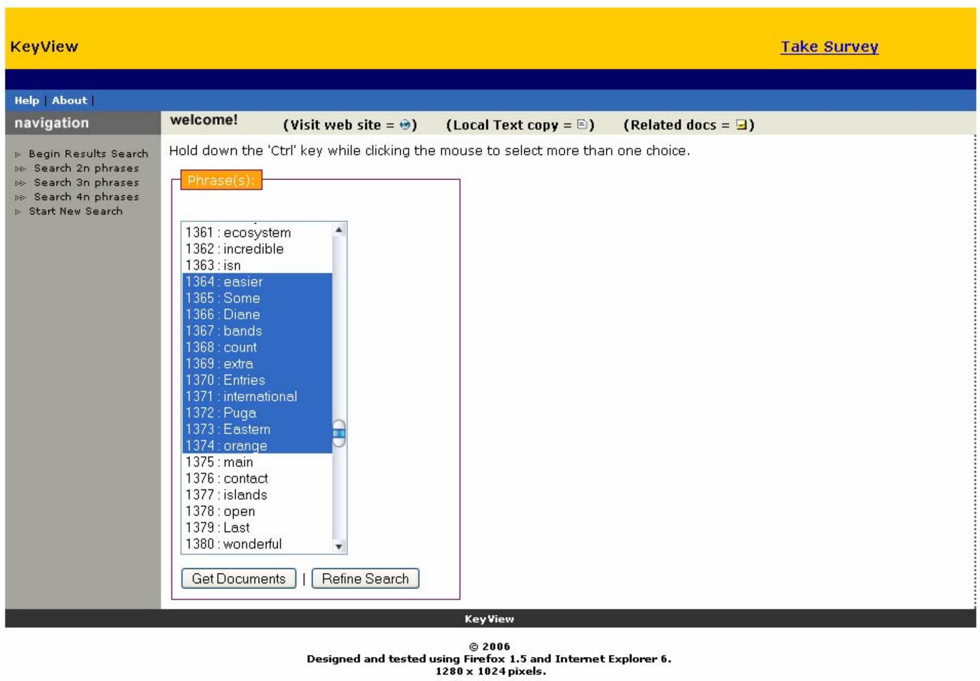


Figure 17. Selecting keyphrases.

You can go directly to multiple word terms by using the left navigation (see figure 18):

- 'Search 2n phrases' means a keyphrase of two words
- 'Search 3n phrases' means a keyphrase of three words
- 'Search 4n phrases' means a keyphrase of four words
- 'Start New Search' begins a new search session.

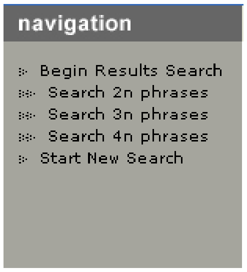


Figure 18. KeyView navigation.

The user can select only a single term or a multiple single from the list. Terms are arranged in order of decreasing frequency.

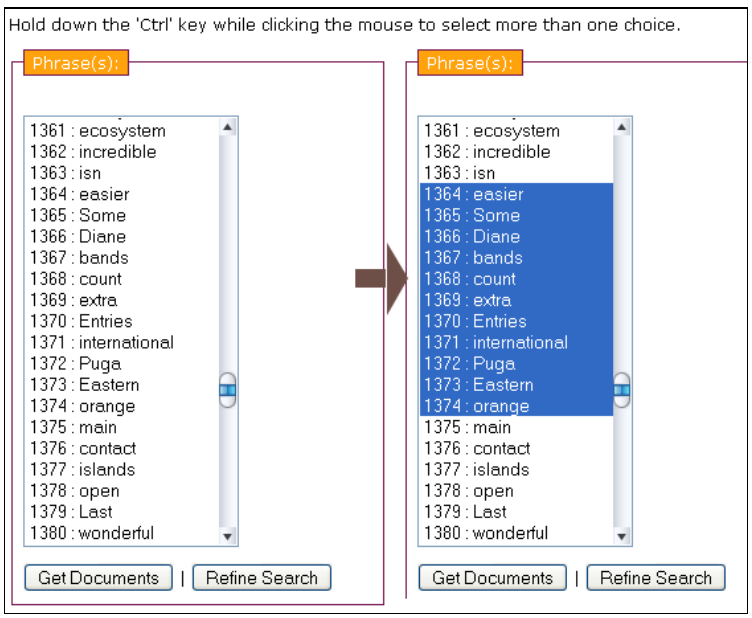


Figure 19. Keyphrase or term selection.

If the user chose to 'get documents' matching their selection, they can click on 'expand all' to see all documents matching selection. Or click on individual keyphrase(s) or terms to see only the documents that contain those keyphrase(s) or term.

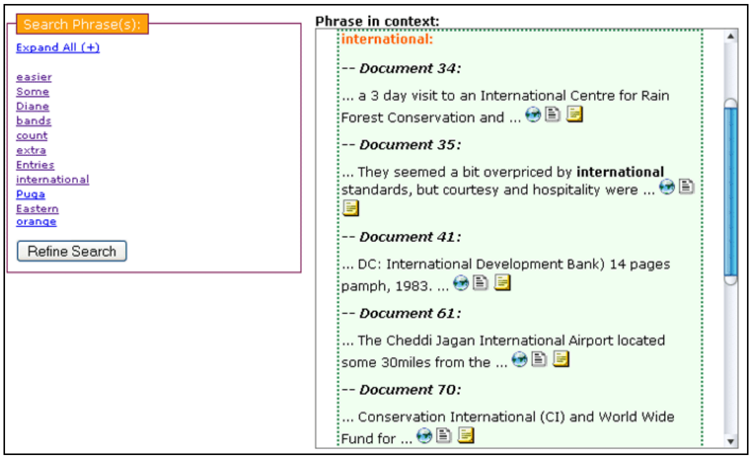


Figure 20. Documents containing keyphrase(s) and the phrase(s) in context.

You can visit the Website of the document matching your result, examine a local text copy or discover related documents.

Figure 21. Document options.

If the application could not find the term(s) selected, it will display 'could not grab document snippet' in the phrases in content window. Click on the 'related docs' icon to discover if the application found similar documents.

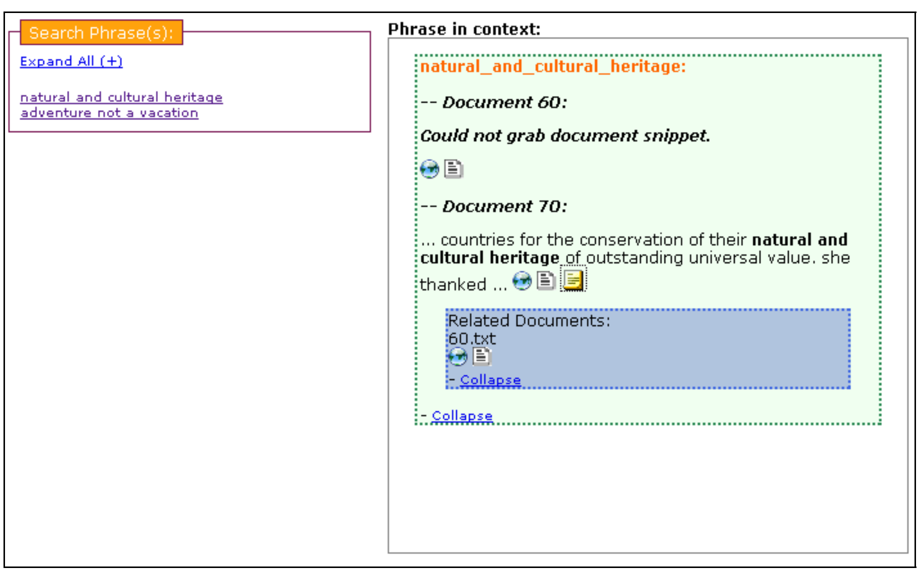


Figure 22. Discovering similar documents.

'Refine search' display a list of the terms previously selected and a new list of terms containing the prior terms. The searcher can refine further or explore the matching document.

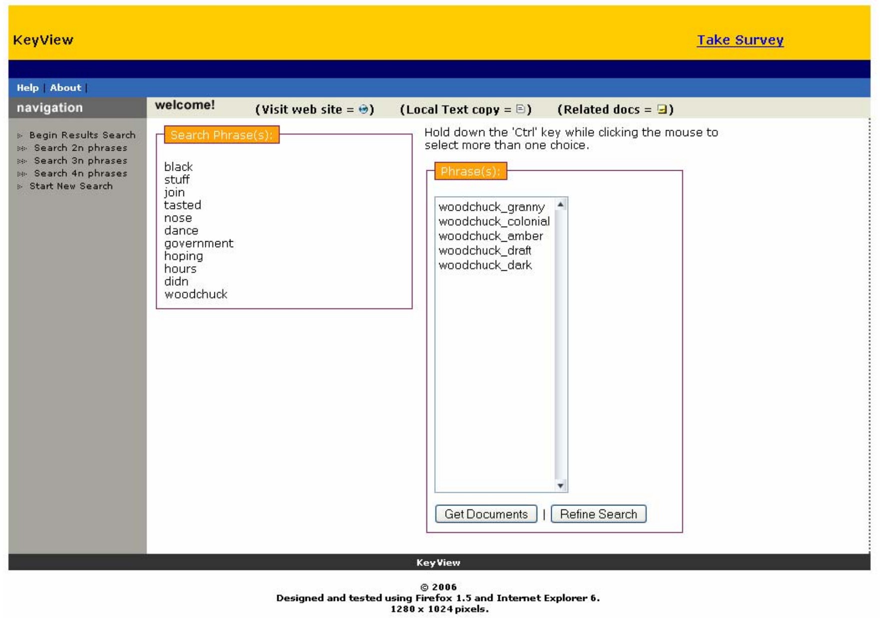


Figure 23. Refining a keyphrase search.

Methodology of Full Study

A descriptive research methodology will be used for this study. The term 'survey' is commonly applied to a research methodology designed to collect data from a specific population, or a sample from that population, and typically utilizes a questionnaire or an interview as the survey instrument. Surveys are widely accepted and used to obtain data from individuals about themselves, their households, or about larger social institutions (school boards). Sample surveys are an important tool for collecting and analyzing information from selected individuals.

Instrumentation

Questionnaire for User Interface Satisfaction (QUIS)

An adapted version (short form) of the Questionnaire for User Interaction Satisfaction (QUIS) version 7.0 was administered. The short form of QUIS is usually administered because the long form is only used for very diagnostic examinations. Sections or items of QUIS can be dropped and question added as needed. QUIS is designed to assess users' subjective satisfaction with specific aspects of a human-computer interface. The purpose of the questionnaire is to guide in the design or redesign of systems (in this study, an application), provide a tool for assessing potential areas of application improvement, provide researchers with a validated instrument for conducting comparative evaluations, and serve as a test instrument in usability labs (HCIL University of Maryland). It was developed using psychological test construction methods to ensure proper construct and empirical validity and reliability of the items (Chin, Diehl, & Norman, 1988). Because of the previous studies done by Chin, Diehl & Norman (1988), and Harper, B., Slaughter, L., & Norman, K. (1997) a pilot study was not carried out by this researcher to test the survey.

QUIS asks general questions about usability and aspects of the interface. For this reason, it usually appropriate for most applications (HCIL University of Maryland). The best time to administer QUIS is immediately after a user has interacted with the system being tested. The short form of the QUIS is divided into 5 sections. The sections are designed to assess 1) overall user reactions, 2) screen design and layout, 3) terminology and system messages, 4) learning, and 5) system capabilities. Users rate each question on a scale from 0 (the lowest rating) to 9 (the highest) rating. Assessed by a semantic differential scale (the Semantic Differential (SD)) measures people's reactions to stimulus words and concepts in terms of ratings on bipolar scales defined with contrasting adjectives at each end. The semantic differential scale does not have a neutral or middle selection. A person must choose, to a certain extent, one or the other adjective) of 0 - 9 anchored with bipolar adjectives: positive adjectives anchor the right side and negative adjectives anchor the left.

The study population will be a convenience sample of University of North Texas (UNT) School of Library and Information Science (SLIS) graduate students and random sample Internet users. Sampling group will consist of adults over the age of 18. Projected sample size is 50 students and Internet users. The respondents were solicited to take the adapted short form of the survey from UNT SLIS graduate students enrolled and attending classes during the summer semester of 2006 and from the Internet. Access to students was obtained via the SLIS mailing list and SLIS Village in WebCT and emails to professors teaching summer courses. Access to general Internet users was obtained via Web forums, mailing lists, newsgroups and email. The selected 56 questions took the respondents approximately 10 to 20 minutes to complete.

The survey instrument consisted of a recruitment cover letter and the instrument survey. The specific demographic information obtained from the instrument included age, education, gender, and familiarity with using an information retrieval system (IRS)).

User attitudes toward KeyView were assessed using the QUIS. Included in the QUIS are five different domains designed to examine several relevant areas of interest. The information obtained from these instruments included:

Table 7

QUIS Instrument Scales

Scale Dimension	Number of items
Overall User reactions	13
Screen design & layout	9
Terms & system information	13
Learning	10
System capabilities	6

Each of the instruments from which the items were taken has demonstrated high reliabilities and adequate levels of validity, as demonstrated in studies done by Chin, Diehl and Norman (1988). After data collection, the scale reliabilities were calculated for the modified instrument used in this study. Reliabilities that were considered to be acceptable for consideration in this study did not fall below 0.9.

Survey Data

Respondents used the KeyView application online, and then took the survey after searching on a familiar and unfamiliar topic. Raw survey data was collected and analyzed using SPSS (originally, Statistical Package for the Social Sciences). The methodology for this study was analytical in nature. Reliability coefficients were calculated for each of the five scales used in the instrument. Non parametric tests were used to compare survey respondents based on groups because group populations were not equal or roughly equal as required for parametric test. Information derived from the demographics was used to analyze any possible difference due to amount of education and gender. Other demographic information derived from the questionnaire may be used for exploratory purposes to guide later studies.

Summary

Chapter 4 describes the specific processes that were used to conduct the research in this study. The study used as a convenience sample of University of North Texas School of Library and Information Science graduate students and a random sample Internet users to determine users' subjective satisfaction with the KeyView keyphrase browsing application. Five subscales from an existing survey were used to assess user attitudes. The results are discussed in chapter 6.

CHAPTER 6

DATA ANALYSIS

Introduction

> Statistical significance is the least interesting thing about the results. You should describe the results in terms of measures of magnitude - not just, does a treatment affect people, but how much does it affect them.
>
> Gene V. Glass

The purpose of the purpose of this survey was to investigate user satisfaction as it relates to using a keyphrase search application (KeyView). This study is a qualitative exploratory study that primarily uses descriptive statistics and effect size to describe observed results. Though observed *p*-values are reported for results, it is not recommended that these *p*-values be taken as definitive statements about the replicability of the present findings if this study were to be repeated. The sample size is small because of logistical and pragmatic reasons beyond my control.

The findings in this chapter are organized into three sections: the first section depicts the characteristics of the survey population, the second discusses the scale responses and the third discusses the qualitative finding for comments submitted by survey respondents.

Survey Findings

A total of 35 usable survey results were collected. Demographic and Level I analysis were conducted with descriptive statistics (i.e., percentages, frequency distribution) for survey items. Level II analysis includes Cronbach's alpha, Spearman's rho and Mann-Whitney tests. Responses to the open-ended item is included in the section entitled Qualitative Analysis.

Qualitative data (optional written comments) were collected and analyzed in an effort to enhance and further validate the survey findings. Content analysis was conducted to determine the opinions of respondents (see section entitled Qualitative Analysis).

Level I: Frequency Analysis of Item Responses

The total sample included 36 respondents. A breakdown of sample demographics is presented below. Percentages were rounded to the nearest integer.

Characteristics of the Population

The survey queried participants on 4 demographic questions and 52 user application satisfaction questions (see Appendix B). Demographic questions consisted of gender, age, education, and previous information retrieval experience. The demographics portion of the survey had two dichotomous questions that asked respondents their gender and if they used an information retrieval system (IRS).The majority of respondents were female (63%). Overall, most respondents were between the ages of 30-39 (43%) and with only two respondents (6%) 20 years old or younger. Twenty-nine percent of the respondents had a master's degree whereas only 11% had a doctorate or professional degree, 31% had a baccalaureate and 17% had only some college or less. Ninety-four percent had prior experience using an IRS. See Tables 9-14 and Figures 24-28 for a complete listing of demographic information.

Table 8

Four Demographic Questions

		Age	Sex	Education	Information Retrieval System Experience
N	Valid	36	36	36	35
	Missing	0	0	0	1
Mean		4.50	1.39	6.53	1.06
Std. Error of Mean		.351	.082	.302	.040
Median		4.00	1.00	6.00	1.00
Mode		4	1	6	1
Std. Deviation		2.104	.494	1.812	.236
Skewness		.438	.476	-.825	3.989
Std. Error of Skewness		.393	.393	.393	.398
Kurtosis		-.274	-1.881	.939	14.752
Std. Error of Kurtosis		.768	.768	.768	.778
Minimum		1	1	1	1
Maximum		9	2	9	2

Table 9

Gender Demographics

	Frequency	Percent	Valid Percent	Cumulative Percent
Female	22	61.1	61.1	61.1
Male	14	38.9	38.9	100.0
Total	36	100.0	100.0	

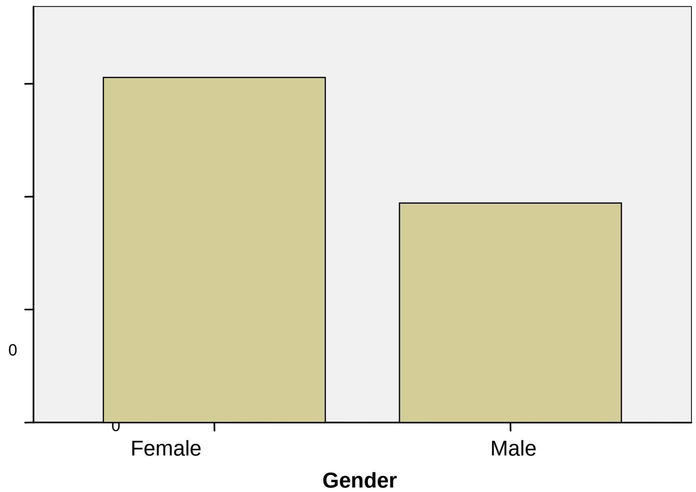


Figure 24. Gender demographics.

Table 10

Age Demographics

	Frequency	Percent	Valid Percent	Cumulative Percent
18 to 20	2	5.6	5.6	5.6
21 to 24	5	13.9	13.9	19.4
25 to 29	4	11.1	11.1	30.6
30 to 34	9	25.0	25.0	55.6
35 to 39	6	16.7	16.7	72.2
40 to 44	4	11.1	11.1	83.3
45 to 49	2	5.6	5.6	88.9
50 to 54	2	5.6	5.6	94.4
55 to 59	2	5.6	5.6	100.0
Total	36	100.0	100.0	

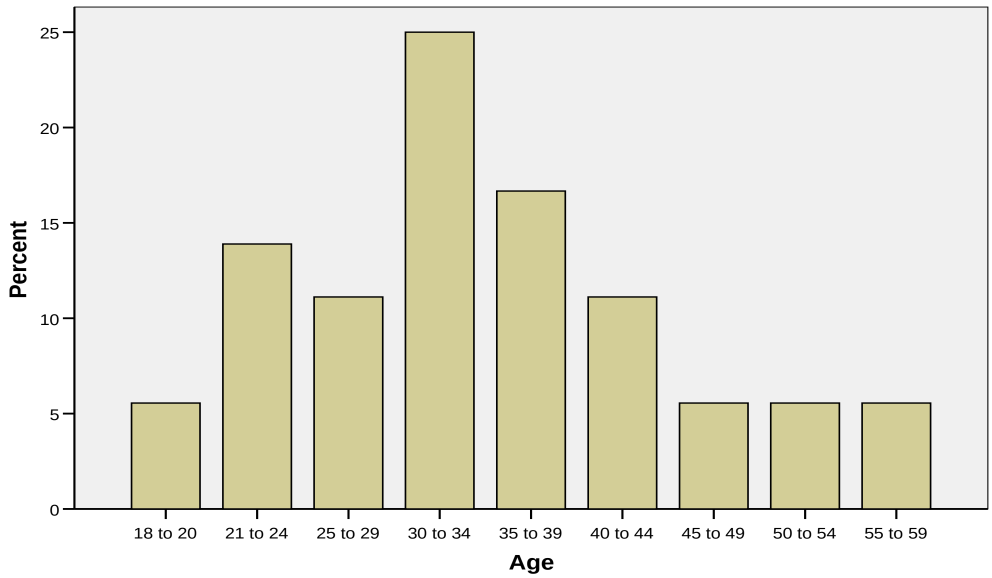


Figure 25. Age demographics.

Table 11

Education Demographics

	Frequency	Percent	Valid Percent	Cumulative Percent
High school or less	1	2.8	2.8	2.8
Some college	5	13.9	13.9	16.7
Vocational school graduate	1	2.8	2.8	19.4
College graduate	12	33.3	33.3	52.8
Some post-graduate	3	8.3	8.3	61.1
Postgraduate degree - Masters or equivalent	10	27.8	27.8	88.9
Postgraduate degree - Ph.D. or equivalent	4	11.1	11.1	100.0
Total	36	100.0	100.0	

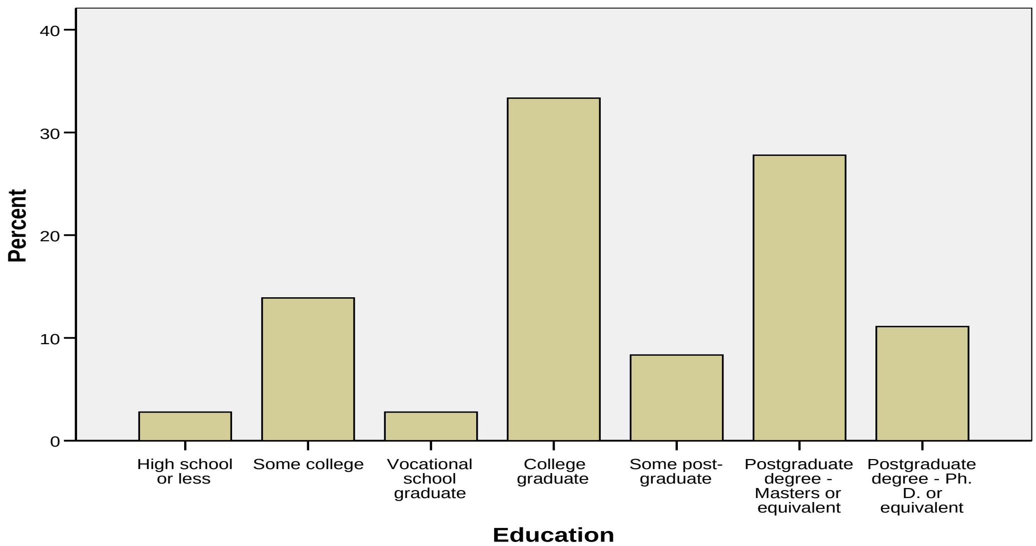


Figure 26. Education demographics.

The respondents' demographic education data was split into two groups: those that had Ph.D. to college graduate level education and those with vocational or some college to high school level education.

Table 12

Education Group Demographics

	Frequency	Percent	Valid Percent	Cumulative Percent
Vocational school graduate to High school or less	7	20.0	20.0	20.0
Ph.D. or equivalent to College graduate	28	80.0	80.0	100.0
Total	35	100.0	100.0	

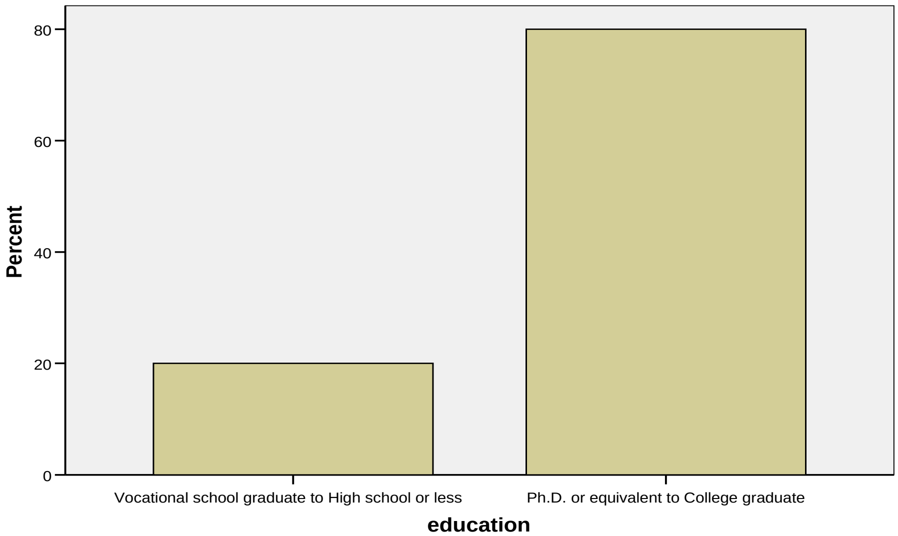


Figure 27. Education group demographics.

Table 13

Information Retrieval System Experience Demographics

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	33	91.7	94.3	94.3
	No	2	5.6	5.7	100.0
	Total	35	97.2	100.0	
Missing	System	1	2.8		
Total		36	100.0		

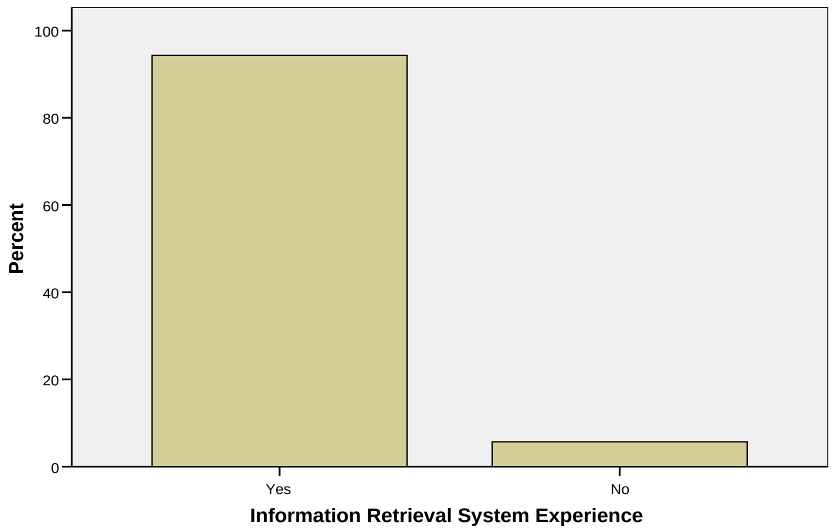


Figure 28. Information retrieval system experience demographics.

Level II: Scale Analysis

Reliability Analyses

There were five scales or domains of interest in this particular investigation. User satisfaction was assessed by examining the scale dimensions of

1. Overall user reactions
2. Screen design and layout
3. Terminology and system messages
4. Learning
5. System capabilities

Cronbach's alpha (α) was computed for each scale and can be seen in Tables 15-19. The scales demonstrate a high level of reliability with alphas exceeding 0.90.

Conceptual Summary of Survey Scales: Definition and Content

Overall user reactions: This scale considered users' opinions and attitudes toward

KeyView. The scale included items relevant to software perception, usefulness of search results, and determination of document relevance and length of keyphrases.

Table 14

Overall User Reactions Reliability Scale

Case Processing Summary

	N	%
Valid	28	77.8
Excluded[a]	8	22.2
Total	36	100.0

[a] Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.965	.965	13

Item Statistics

	Mean	Std. Dev.	*N*
Overall Reaction (terrible/wonderful)	5.82	1.786	28
Overall Reaction (frustrating/satisfying)	6.14	2.189	28
Overall Reaction (dull/stimulating)	5.61	2.266	28
Overall Reaction (difficult/easy)	5.96	2.317	28
Overall Reaction (inadequate power/adequate power)	6.46	2.117	28
Overall Reaction (rigid/flexible)	5.50	2.134	28
Search results (unhelpful/helpful)	6.43	2.284	28
Determine document relevance to search topic (difficult/easy)	6.75	2.171	28
Keyphrase search on an unfamiliar topic (unhelpful/helpful)	6.14	2.353	28
Keyphrase search on a familiar topic	6.75	2.303	28
Determine document relevance to search topic (too slow/fast enough)	6.04	2.442	28
Shorter keyphrases (unhelpful/helpful)	5.96	2.252	28
Longer keyphrases (unhelpful/helpful)	6.21	2.425	28

The scale included 13 items and produced a Cronbach's alpha coefficient (α) of .965.

Screen design and layout: This scale examined users' perceptions of screen layout, including use of highlighting, the amount of information displayed on screen, the arrangement of information on screen and screen sequence.

Table 15

Screen Design and Layout Reliability Scale

Case Processing Summary

	N	%
Valid	30	83.3
Excluded[a]	6	16.7
Total	36	100.0

[a] Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	*N*
.937	.939	9

Item Statistics

	Mean	Std. Dev.	*N*
Characters on the computer screen (hard to read/easy to read)	6.67	2.090	30
Highlighting on the screen (unhelpful/helpful)	7.33	1.668	30
Use of bolding (unhelpful/helpful)	7.20	1.627	30
Screen layouts were helpful (never/always)	6.40	2.175	30
Amount of information that can be displayed on screen (inadequate/adequate)	6.23	2.112	30
Arrangement of information on screen (illogical/logical)	6.50	1.815	30
Sequence of screens (confusing/very clear)	6.57	1.942	30
Next screen in a sequence (unpredictable/predictable)	6.80	1.883	30
Going back to the previous screen (impossible/easy)	7.00	1.554	30

The scale included 9 items and produced a α of .937.

Terminology and application/system messages: This scale was developed to examine users' reactions of the use of terminology throughout application, including messages on screen, instructions for commands, application predictability, and error messages.

Table 16

Terminology and Application/System Messages Reliability Scale

Case Processing Summary

	N	%
Valid	22	61.1
Excluded[a]	14	38.9
Total	36	100.0

[a] Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	*N*
.981	.982	13

Item Statistics

	Mean	Std. Dev.	*N*
Use of terminology throughout application (inconsistent/consistent)	6.55	2.241	22
Task related terminology (inconsistent/consistent)	6.82	2.015	22
Messages which appear on screen (inconsistent/consistent)	6.86	2.210	22
Position of instructions on the screen (inconsistent/consistent)	7.00	2.093	22
Messages which appear on screen (confusing/clear)	6.50	2.502	22
Instructions for commands or functions (confusing/clear)	5.91	2.580	22
Instructions for correcting errors (confusing/clear)	5.82	2.260	22
Application keeps you informed about what it is doing (never/always)	6.77	2.617	22
Performing an operation leads to a predictable result (never/always)	6.86	2.336	22
Length of delay between operations unacceptable/acceptable)	6.23	2.487	22
Error messages (unhelpful/helpful)	6.73	2.164	22
Error messages clarify the problem (never/always)	6.55	2.558	22
Phrasing of error messages (unpleasant/pleasant)	6.68	2.418	22

The scale included 13 items and produced a α of .981.

Learning: This scale is designed to assess users' perceptions relative to learning to operate the KeyView; exploration of features by trial and error, steps to complete a task follows a logical sequence, and receiving feedback on the completion of the steps.

Table 17

Learning Reliability Scale

Case Processing Summary

	N	%
Valid	28	77.8
Excluded[a]	8	22.2
Total	36	100.0

[a] Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	*N*
.973	.973	10

Item Statistics

	Mean	Std. Dev.	*N*
Learning to operate the system (difficult/easy)	6.46	2.380	28
Getting started (difficult/easy)	5.68	2.229	28
Learning advanced features (difficult/easy)	6.36	2.248	28
Time to learn to use the application (slow/fast)	6.18	2.450	28
Exploration of features by trial and error (discouraging/encouraging)	5.93	2.227	28
Remembering names and use of commands (difficult/easy)	6.61	2.424	28
Tasks can be performed in a straight-forward manner (never/always)	6.36	2.527	28
Number of steps per task (too many/just right)	6.21	2.183	28
Steps to complete a task follow a logical sequence (never/always)	6.79	2.200	28
Feedback on the completion of the steps (unclear/clear)	6.93	2.292	28

The scale included 10 items and produced an α of .973.

Application/system capabilities: This scale assessed a variety of elements of application capabilities. Specifically, the scale referred to application speed, response time for most operations and level of expertise affecting ease of operation, and other items.

Table 18

Application/System Capabilities Reliability Scale

Case Processing Summary

	N	%
Valid	29	80.6
Excluded[a]	7	19.4
Total	36	100.0

[a] Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	*N*
.918	.919	6

Item Statistics

	Mean	Std. Dev.	*N*
Application speed (too slow/fast enough)	6.14	2.601	29
Response time for most operations (too slow/fast enough)	6.55	2.558	29
Rate information is displayed	6.86	2.356	29
The application is reliable (never/always)	6.83	2.300	29
Correcting your mistakes (difficult/easy)	6.34	2.126	29
Ease of operation depends on your level of experience (never/always)	6.45	2.369	29

The scale included 6 items and produced an α of .918.

Correlations were conducted on the five scale dimensions. As Table 20 indicates, of the scales demonstrate a high level of reliability with all alphas exceeding .90 and are associated with each other.

Table 19

Cronbach's Alphas for Five Scale Dimensions

	Number of items	Alpha
Overall User reactions	13	.965
Screen design & layout	9	.937
Terms & system information	13	.981
Learning	10	.973
System capabilities	6	.918

Survey respondents mean scores on the five scales tended to be range from 6.1 to 6.7, the more positive end of the semantic distance scale of 0-9 (see Table 20).

Table 20

Mean User Response for Each QUIS Scale Section

	N	Min	Max	Mean	Mean Std. Error	Std. Dev.
Overall user reactions	28	.00	8.77	6.1374	.35559	1.88159
Screen design and layout	30	1.78	8.67	6.7444	.28105	1.53937
Terminology and application/system messages	22	.00	8.54	6.5594	.45320	2.12569
Learning	28	.00	8.60	6.3500	.39294	2.07926
Application/system capabilities	29	.00	9.00	6.5287	.37388	2.01338
Valid N (listwise)	13					

Min = minimum; Max = maximum

Scale correlations of overall software satisfaction (as measured by TOverall) and the other scales (as measured by TScreen, TApp, TLearn, TCap) was investigated using Spearman's rho (see Table 22) to gage the strength and direction of relationships among the scales. Correlation examines relationships between variables. When two or more variables vary together (a change in one is accompanied by a change in the other), they are correlated. If a relationship has positive correlation, then an increase in one variable will also indicate an increase in another. A negative correlation means that as one variable increase, the other decreases. The correlation coefficient expresses quantitatively the extent to which the variables are related. The correlation coefficient (r) is calculated by squaring r (r^2)

Table 21

Interpreting Correlation Coefficient Direction and Strength

r =	□	(0 □ 1)
	Sign	Magnitude
	Gives direction	Gives strength

Cohen (1988) has suggested the following interpretations for correlations:

Table 22

Assessing Small, Medium and Large Correlations

Correlation	Negative	Positive
Small	−0.29 to −0.10	0.10 to 0.29
Medium	−0.49 to −0.30	0.30 to 0.49
Large	−0.50 to −1.00	0.50 to 1.00

There was a strong positive correlation between the scales, with high levels on one scale being associated with high levels on another, to different degrees. Overall satisfaction (TOverall) was most strongly correlated with Terms & System Information (TApp), helping to explain 83% of the variance in respondents' scores on the Terms & System Information scale. The second strongest correlated scale was Learning (TLearn). Scores on Overall satisfaction scale helped to explain 67% of the score variance on the Learning scale. Screen Design & Layout (TScreen) was most strongly correlated with overall satisfaction (49% of score variance) and System Capabilities (TCap) (42% of score variance). Terms & System Information was most strongly correlated with Learning (89% of score variance) and Overall satisfaction. Learning was most strongly correlated with Terms & System Information and Overall satisfaction. System Capabilities was strongly correlated to Terms & System Information (64% of score variance) and Learning (50% of score variance).

Table 23

Correlation of Five Scales

		TOverall	TScreen	TApp	TLearn	TCap
TOverall	Corr. Coef.	1.000	.702(**)	.908(**)	.821(**)	.574(**)
	Sig. (2-tailed)	.	.000	.000	.000	.003
	N	28	25	18	23	24
TScreen	Corr. Coef.		1.000	.611(**)	.506(*)	.651(**)
	Sig. (2-tailed)		.	.005	.012	.000
	N		30	19	24	27
TApp	Corr. Coef.			1.000	.944(**)	.797(**)
	Sig. (2-tailed)			.	.000	.000
	N			22	19	20
TLearn	Corr. Coef.				1.000	.708(**)
	Sig. (2-tailed)				.	.000
	N				28	24
TCap	Corr. Coef.					1.000
	Sig. (2-tailed)					.
	N					29

Corr coef = correlation coefficient. **Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Seven questions were added (the "overall user reaction" scale has six original rating questions) to the QUIS measuring users' overall satisfaction with specific aspects of KeyView. One would expect this question to be most highly correlated with the other questions in the overall user reaction category (72% of score variance). These seven questions were considered a sub-scale (called KeyView Questions, see Table 25) of overall satisfaction for correlation purposes, in order to answer the research questions. The KeyView sub-scale questions as a whole were most strongly correlated to the Terms & System Information (TApp) scale (53% of score variance), followed by Screen Design & Layout (TScreen) scale (38% of score variance) with only a medium correlation to the Learning (TLearn) scale (24% of score variance) and System Capabilities (TCap) scale (22% of score variance).

Table 24

KeyView Subscale in Comparison to Four Other Scales

		KeyView Questions
TOverall	Correlation Coefficient	.851(**)
	Sig. (2-tailed)	.000
	N	28
TScreen	Correlation Coefficient	.606(**)
	Sig. (2-tailed)	.001
	N	26
TApp	Correlation Coefficient	.728(**)
	Sig. (2-tailed)	.000
	N	19
TLearn	Correlation Coefficient	.490(*)
	Sig. (2-tailed)	.015
	N	24
TCap	Correlation Coefficient	.467(*)
	Sig. (2-tailed)	.019
	N	25

**Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

KeyView sub-scale individual questions in comparison to each other and the four other scales (see tables 30-31 in Appendix E) also examined for correlation. The search results item was most strongly correlated with the ease of determining document relevance, accounting for 70% of score variance. Second strongest correlation was to searching for a familiar topic (62% of score variance) then followed by speed of determining relevance (60% of score variance). There was a strong relationship to the Learning and System Capabilities scales.

Ease of determining document relevance most strongly correlated to helpfulness of search results, followed by the utility of longer keyphrases (60% of score variance). It was most strongly correlated with the System Capabilities scale.

Keyphrase search on an unfamiliar topic was most strongly correlated to searching on a familiar topic (63% of score variance). This suggests that the perceived helpfulness of the keyphrases is not affected very much whether searching on familiar or unfamiliar topics. There was a strong correlation to the utility of longer keyphrases (56% of score variance). There was a strong correlation to all the scales, the strongest being Terms & System Information (TApp) scale (67% of score variance).

Keyphrase search on a familiar topic was most strongly correlated to searching on an unfamiliar topic followed by helpfulness of search results. There was a strong correlation to the Terms & System Information (TApp) scale (60% of score variance) and only a medium correlation to the System Capabilities (TCap) scale.

Speed of document relevance determination was strongly correlated to helpfulness of longer keyphrases (63% of score variance) followed by helpfulness of search results (60% of score variance). Strongest correlations were to Terms & System Information (TApp) scale (45% of score variance) and Learning (TLearn) scale(37% of score variance).

Shorter keyphrases was most strongly correlated to helpfulness of search results (48% of score variance) followed by searching on a familiar (46% of score variance) and searching on an unfamiliar (42% of score variance). Most strongly correlated to Terms & System Information (TApp) scale searching on a familiar (75% of score variance) and had only a small correlation to System Capabilities (TCap) of 4% of score variance.

Longer keyphrases was most strongly correlated to speed of determining relevance (63% of score variance) followed by ease of document relevance determination (60% of score variance). This item had a large correlation to the Terms & System Information (TApp) scale (45% of score variance), followed by System Capabilities (TCap) (25% of score variance).

Examining Group Difference

Exploring the relationship between the scales for two groups of differing education levels: comparing those with a college degree and higher vs. those that do not.

Table 25

Comparing Education Group by Scales

Education Groups			TOverall	TScreen	TApp	TLearn	TCap
Vocational school graduate to High school or less	TOverall	Corr. coef.	1.000	.200	.975(**)	.986(**)	.600
		Sig. (2-tailed)	.	.747	.005	.000	.285
		N	6	5	5	6	5
	TScreen	Corr. coef.		1.000	.632	.154	.600
		Sig. (2-tailed)		.	.368	.805	.285
		N		5	4	5	5
	TApp	Corr. coef.			1.000	.975(**)	.949
		Sig. (2-tailed)			.	.005	.051
		N			5	5	4
	TLearn	Corr. coef.				1.000	.667
		Sig. (2-tailed)				.	.219
		N				6	5
	TCap	Corr. coef.					1.000
		Sig. (2-tailed)					.
		N					5

(table continues)

Table 25 *(continued).*

Education Groups			TOverall	TScreen	TApp	TLearn	TCap
Ph.D. or equivalent to College graduate	TOverall	Corr. coef.	1.000	.760(**)	.928(**)	.796(**)	.613(**)
		Sig. (2-tailed)	.	.000	.000	.000	.005
		N	22	20	13	17	19
	TScreen	Corr. coef.		1.000	.676(**)	.595(**)	.607(**)
		Sig. (2-tailed)		.	.006	.007	.003
		N		25	15	19	22
	TApp	Corr. coef.			1.000	.925(**)	.830(**)
		Sig. (2-tailed)			.	.000	.000
		N			17	14	16
	TLearn	Corr. coef.				1.000	.710(**)
		Sig. (2-tailed)				.	.001
		N				22	19
	TCap	Corr. coef.					1.000
		Sig. (2-tailed)					.
		N					24

Corr. coef. = correlation coefficient; **Correlation is significant at the 0.01 level (2-tailed).

The overall satisfaction of respondents with an education level of less than a college graduate was most strongly correlated with learning the application (97% of score variance) and had strong correlation with the other scales except screen design & layout which only accounted for 4% of score variance. Respondents with more than a college education had overall satisfaction scores strongly correlated to all other scales, with terms & system information scale the strongest.

Screen Design & Layout for less educated group had only small correlations to overall satisfaction (4% of score variance) and learning (2% of score variance). For the more educated group, screen design & layout had the largest correlation to overall satisfaction (57% of score variance).

Terms & System Information for the less educated group had two equally strong correlations to overall satisfaction and learning accounting for 95% of the variance of scores on those two scales. For the more educated group, terms & system information was also most strongly correlated with overall satisfaction and learning accounting for 86% of the variance of scores on both scales.

Learning for the less educated group had the strongest correlation to overall satisfaction, followed by terms & system information (95% of score variance) and only a small correlation to screen design & layout. Learning for the more educated group was also strongly correlated to terms & system information, followed by overall satisfaction (63% of score variance).

System Capabilities for the less educated group for the less educated group had strong correlations to all scales. The strongest was terms & system information (90% of score variance), followed by learning; overall satisfaction and screen design & layout had equally strong correlations (36% of score variance). System Capabilities for the more educated group also had the strongest correlation to terms & system information (69% of score variance), followed by learning, overall satisfaction and screen design & layout.

Table 26

Comparing Education Group by KeyView Subscale by Scales

Education Groups			TOverall	TScreen	TApp	TLearn	TCap
Vocational school graduate to High school or less	Search results (unhelpful/helpful)	Corr. coef.	.353	.791	.763	.403	.474
		Sig. (2-tailed)	.492	.111	.133	.428	.420
		N	6	5	5	6	5
	Determine document relevance to search topic (difficult/easy)	Corr. coef.	.353	.580	.684	.418	.264
		Sig. (2-tailed)	.492	.306	.203	.410	.668
		N	6	5	5	6	5
	Keyphrase search on an unfamiliar topic (unhelpful/helpful)	Correlation Coefficient Corr. coef.	.383	.211	.973(**)	.493	.632
		Sig. (2-tailed)	.454	.734	.005	.321	.252
		N	6	5	5	6	5

(table continues)

Table 26 *(continued).*

Education Groups			TOverall	TScreen	TApp	TLearn	TCap
Vocational school graduate to High school or less *(cont.)*	Keyphrase search on a familiar topic (unhelpful/helpful)	Corr. coef.	.577	.447	.947(*)	.647	.894(*)
		Sig. (2-tailed)	.231	.450	.014	.165	.041
		N	6	5	5	6	5
	Determine document relevance to search topic (too slow/fast enough)	Corr. coef.	.617	.289	.459	.548	-.289
		Sig. (2-tailed)	.192	.638	.437	.260	.638
		N	6	5	5	6	5
	Shorter keyphrases (unhelpful/helpful)	Corr. coef.	.177	-.791	.342	.269	-.474
		Sig. (2-tailed)	.738	.111	.573	.607	.420
		N	6	5	5	6	5
	Longer keyphrases (unhelpful/helpful)	Corr. coef.	.706	.158	.553	.672	-.158
		Sig. (2-tailed)	.117	.800	.334	.144	.800
		N	6	5	5	6	5
Ph.D. or equivalent to College graduate	Search results (unhelpful/helpful)	Corr. coef.	.812(**)	.576(**)	.717(**)	.637(**)	.638(**)
		Sig. (2-tailed)	.000	.003	.002	.002	.001
		N	22	24	16	21	23
	Determine document relevance to search topic (difficult/easy)	Corr. coef.	.609(**)	.602(**)	.648(**)	.609(**)	.586(**)
		Sig. (2-tailed)	.003	.002	.009	.003	.004
		N	22	23	15	21	22
	Keyphrase search on an unfamiliar topic (unhelpful/helpful)	Corr. coef.	.823(**)	.751(**)	.808(**)	.606(**)	.680(**)
		Sig. (2-tailed)	.000	.000	.000	.005	.000
		N	22	24	17	20	23
	Keyphrase search on a familiar topic (unhelpful/helpful)	Corr. coef.	.807(**)	.721(**)	.757(**)	.593(**)	.493(*)
		Sig. (2-tailed)	.000	.000	.000	.004	.014
		N	22	25	17	22	24
	Determine document relevance to search topic (too slow/fast enough)	Corr. coef.	.812(**)	.429(*)	.734(**)	.655(**)	.504(*)
		Sig. (2-tailed)	.000	.032	.001	.001	.012
		N	22	25	17	22	24
	Shorter keyphrases (unhelpful/helpful)	Corr. coef.	.862(**)	.372	.710(**)	.646(**)	.290
		Sig. (2-tailed)	.000	.067	.001	.001	.169
		N	22	25	17	22	24
	Longer keyphrases (unhelpful/helpful)	Corr. coef.	.778(**)	.565(**)	.770(**)	.488(*)	.652(**)
		Sig. (2-tailed)	.000	.004	.000	.029	.001
		N	22	24	17	20	23

Corr. coef. = correlation coefficient; **Correlation is significant at the 0.01 level (2-tailed); *Correlation is significant at the 0.05 level (2-tailed).

Table 27

Comparing KeyView Subscale Response by Gender

Gender			Search results (unhelpful/helpful)	Determine document relevance to search topic (difficult/easy)	Keyphrase search on an unfamiliar topic (unhelpful/helpful)	Keyphrase search on a familiar topic	Determine document relevance to search topic (too slow/fast enough)	Shorter keyphrases (unhelpful/helpful)	Longer keyphrases (unhelpful/helpful)
Female	Search results (unhelpful/helpful)	Pearson Corr.	1	.761(**)	.749(**)	.899(**)	.798(**)	.846(**)	.822(**)
		Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
		N	21	19	19	21	20	20	18
	Determine document relevance to search topic (difficult/easy)	Pearson Corr.	.761(**)	1	.583(*)	.769(**)	.843(**)	.748(**)	.798(**)
		Sig. (2-tailed)	.000		.011	.000	.000	.000	.000
		N	19	20	18	20	19	19	17
	Keyphrase search on an unfamiliar topic (unhelpful/helpful)	Pearson Corr.	.749(**)	.583(*)	1	.777(**)	.874(**)	.801(**)	.893(**)
		Sig. (2-tailed)	.000	.011		.000	.000	.000	.000
		N	19	18	20	20	19	19	19
	Keyphrase search on a familiar topic (unhelpful/helpful)	Pearson Corr.	.899(**)	.769(**)	.777(**)	1	.852(**)	.816(**)	.734(**)
		Sig. (2-tailed)	.000	.000	.000		.000	.000	.000
		N	21	20	20	22	21	21	19
	Determine document relevance to search topic (too slow/fast enough)	Pearson Corr.	.798(**)	.843(**)	.874(**)	.852(**)	1	.897(**)	.883(**)
		Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
		N	20	19	19	21	21	21	19
	Shorter keyphrases (unhelpful/helpful)	Pearson Corr.	.846(**)	.748(**)	.801(**)	.816(**)	.897(**)	1	.779(**)
		Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
		N	20	19	19	21	21	21	19
	Longer keyphrases (unhelpful/helpful)	Pearson Corr.	.822(**)	.798(**)	.893(**)	.734(**)	.883(**)	.779(**)	1
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
		N	18	17	19	19	19	19	19
Male	Search results (unhelpful/helpful)	Pearson Corr.	1	.901(**)	.890(**)	.861(**)	.801(**)	.705(**)	.873(**)
		Sig. (2-tailed)		.000	.000	.000	.001	.005	.000
		N	14	14	14	14	14	14	14
	Determine document relevance to search topic (difficult/easy)	Pearson Corr.	.901(**)	1	.882(**)	.891(**)	.732(**)	.668(**)	.848(**)
		Sig. (2-tailed)	.000		.000	.000	.003	.009	.000
		N	14	14	14	14	14	14	14

(table continues)

Table 27 *(continued).*

Gender			Search results (unhelpful/helpful)	Determine document relevance to search topic (difficult/easy)	Keyphrase search on an unfamiliar topic (unhelpful/helpful)	Keyphrase search on a familiar topic	Determine document relevance to search topic (too slow/fast enough)	Shorter keyphrases (unhelpful/helpful)	Longer keyphrases (unhelpful/helpful)
Male *(cont.)*	Keyphrase search on an unfamiliar topic (unhelpful/helpful)	Pearson Corr.	.890(**)	.882(**)	1	.886(**)	.729(**)	.749(**)	.882(**)
		Sig. (2-tailed)	.000	.000		.000	.003	.002	.000
		N	14	14	14	14	14	14	14
	Keyphrase search on a familiar topic (unhelpful/helpful)	Pearson Corr.	.861(**)	.891(**)	.886(**)	1	.796(**)	.834(**)	.873(**)
		Sig. (2-tailed)	.000	.000	.000		.001	.000	.000
		N	14	14	14	14	14	14	14
	Determine document relevance to search topic (too slow/fast enough)	Pearson Corr.	.801(**)	.732(**)	.729(**)	.796(**)	1	.694(**)	.847(**)
		Sig. (2-tailed)	.001	.003	.003	.001		.006	.000
		N	14	14	14	14	14	14	14
	Shorter keyphrases (unhelpful/helpful)	Pearson Corr.	.705(**)	.668(**)	.749(**)	.834(**)	.694(**)	1	.675(**)
		Sig. (2-tailed)	.005	.009	.002	.000	.006		.008
		N	14	14	14	14	14	14	14
	Longer keyphrases (unhelpful/helpful)	Pearson Corr.	.873(**)	.848(**)	.882(**)	.873(**)	.847(**)	.675(**)	1
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.008	
		N	14	14	14	14	14	14	14

Pearson Corr. = Pearson's correlation. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

No significant difference was found between the groups since z value is lower than 1.96 and the p value of significance is not less than or equal to 0.05. The results were also not significant when compared by gender alone (see Table 27).

For the less educated group there is significance difference in overall satisfaction and learning scores between men and women. Finding significant difference for men vs. women might not be meaningful due to small case population. The total cases or sample size analyzed is six for both scales. This means that one out of the total seven cases for the less educated group was not used because it had missing values. The probability of an outcome depends upon the sample size. A large effect can be non-significant if the sample size is small. Even a very small effect can produce a very small probability if the sample size is large. There is no significant difference between men and women for the more educated group.

Table 28

Comparing All Scales' Responses by Gender

Gender			TOverall	TScreen	TApp	TLearn	TCap	Keyview questions
Female	TOverall	Corr. coef.	1.000	.739(**)	.872(**)	.810(**)	.573(*)	.944(**)
		Sig. (2-tailed)	.	.001	.000	.001	.032	.000
		N	16	16	11	13	14	16
	TScreen	Corr. coef.		1.000	.731(**)	.751(**)	.849(**)	.630(**)
		Sig. (2-tailed)		.	.003	.001	.000	.009
		N		20	14	16	18	16
	TApp	Corr. coef.			1.000	.963(**)	.797(**)	.832(**)
		Sig. (2-tailed)			.	.000	.001	.001
		N			14	11	13	11
	TLearn	Corr. coef.				1.000	.544(*)	.678(*)
		Sig. (2-tailed)				.	.036	.011
		N				17	15	13
	TCap	Corr. coef.					1.000	.376
		Sig. (2-tailed)					.	.185
		N					18	14
	Keyview questions	Corr. coef.						1.000
		Sig. (2-tailed)						.
		N						16

(table continues)

Table 28 *(continued).*

Gender			TOverall	TScreen	TApp	TLearn	TCap	Keyview questions
Male	TOverall	Corr. coef.	1.000	.550	.929(**)	.924(**)	.524	.737(**)
		Sig. (2-tailed)	.	.125	.003	.000	.120	.006
		N	12	9	7	10	10	12
	TScreen	Corr. coef.		1.000	.205	-.120	.325	.558
		Sig. (2-tailed)		.	.741	.776	.394	.094
		N		10	5	8	9	10
	TApp	Corr. coef.			1.000	.874(**)	.505	.578
		Sig. (2-tailed)			.	.005	.248	.133
		N			8	8	7	8
	TLearn	Corr. coef.				1.000	.696(*)	.325
		Sig. (2-tailed)				.	.037	.330
		N				11	9	11
	TCap	Corr. coef.					1.000	.492
		Sig. (2-tailed)					.	.124
		N					11	11
	Keyview questions	Corr. coef.						1.000
		Sig. (2-tailed)						.
		N						14

Corr. coef. = Correlation coefficient. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

The only significant difference detected, based on gender, for the five scales was for Terms & System Information. Females had higher scores than males.

Discovering if a significant difference exists between females vs. males with less than a college level education and females vs. males with a college degree and greater.

Table 29

Comparing Scale Response by Education by Gender

Gender	Education Groups			TOverall	TScreen	TApp	TLearn	TCap	KeyView questions
Female	Vocational school graduate to High school or less	TOverall	Corr. coef.	1.000	1.000 (**)	1.000 (**)	1.000 (**)	1.000 (**)	.500
			Sig. (2-tailed)						.667
			N	3	3	3	3	3	3
		TScreen	Corr. coef.	1.000 (**)	1.000	1.000 (**)	1.000 (**)	1.000 (**)	.500
			Sig. (2-tailed)						.667
			N	3	3	3	3	3	3
		TApp	Corr. coef.	1.000 (**)	1.000 (**)	1.000	1.000 (**)	1.000 (**)	.500
			Sig. (2-tailed)						.667
			N	3	3	3	3	3	3
		TLearn	Corr. coef.	1.000 (**)	1.000 (**)	1.000 (**)	1.000	1.000 (**)	.500
			Sig. (2-tailed)						.667
			N	3	3	3	3	3	3
		TCap	Corr. coef.	1.000 (**)	1.000 (**)	1.000 (**)	1.000 (**)	1.000	.500
			Sig. (2-tailed)						.667
			N	3	3	3	3	3	3
		Keyview questions	Corr. coef.	.500	.500	.500	.500	.500	1.000
			Sig. (2-tailed)	.667	.667	.667	.667	.667	
			N	3	3	3	3	3	3
	Ph.D. or equivalent to College graduate	TOverall	Corr. coef.	1.000	.749 (**)	.850 (**)	.833 (**)	.597	.938 (**)
			Sig. (2-tailed)	.	.003	.007	.003	.053	.000
			N	13	13	8	10	11	13
		TScreen	Corr. coef.	.749 (**)	1.000	.762 (**)	.899 (**)	.829 (**)	.658 (*)
			Sig. (2-tailed)	.003		.006	.000	.000	.015
			N	13	17	11	13	15	13
		TApp	Corr. coef.	.850 (**)	.762 (**)	1.000	.928 (**)	.820 (**)	.946 (**)
			Sig. (2-tailed)	.007	.006		.001	.004	.000
			N	8	11	11	8	10	8
		TLearn	Corr. coef.	.833 (**)	.899 (**)	.928 (**)	1.000	.596 (*)	.723 (*)
			Sig. (2-tailed)	.003	.000	.001		.041	.018
			N	10	13	8	14	12	10

(table continues)

Table 29 *(continued).*

Gender	Education Groups			TOverall	TScreen	TApp	TLearn	TCap	KeyView questions
Female *(cont.)*	Ph.D. or equivalent to College graduate *(cont.)*	TCap	Corr. coef.	.597	.829 (**)	.820 (**)	.596 (*)	1.000	.427
			Sig. (2-tailed)	.053	.000	.004	.041		.190
			N	11	15	10	12	15	11
		KeyView questions	Corr. coef.	.938 (**)	.658 (*)	.946 (**)	.723 (*)	.427	1.000
			Sig. (2-tailed)	.000	.015	.000	.018	.190	
			N	13	13	8	10	11	13
Male	Vocational school graduate to High school or less	TOverall	Corr. coef.	1.000	1.000	1.000	.866	-1.000	.500
			Sig. (2-tailed)				.333	1.000	.667
			N	3	2	2	3	2	3
		TScreen	Corr. coef.	1.000 (**)	1.000			-1.000	-1.000
			Sig. (2-tailed)					1.000	1.000
			N	2	2	1	2	2	2
		TApp	Corr. coef.	1.000 (**)		1.000	1.000		1.000
			Sig. (2-tailed)						.
			N	2	1	2	2	1	2
		TLearn	Corr. coef.	.866		1.000 (**)	1.000		.866
			Sig. (2-tailed)	.333					.333
			N	3	2	2	3	2	3
		TCap	Corr. coef.	-1.000	-1.000			1.000	1.000
			Sig. (2-tailed)	1.000	1.000				
			N						
		Keyview questions	Corr. coef.	.500	-1.000	1.000 (**)	.866	1.000 (**)	1.000
			Sig. (2-tailed)	.667	1.000		.333		
			N	3	2	2	3	2	4
	Ph.D. or equivalent to College graduate	TOverall	Corr. coef.	1.000	.714	1.000 (**)	.857 (*)	.563	.790 (*)
			Sig. (2-tailed)		.071		.014	.146	.011
			N	9	7	5	7	8	9
		TScreen	Corr. coef.	.714	1.000	.316	-.116	.414	.801 (*)
			Sig. (2-tailed)	.071		.684	.827	.355	.017
			N	7	8	4	6	7	8

(table continues)

Table 29 *(continued).*

Gender	Education Groups			TOverall	TScreen	TApp	TLearn	TCap	KeyView questions
Male *(cont.)*	Ph.D. or equivalent to College graduate *(cont.)*	TApp	Corr. coef.	1.000 (**)	.316	1.000	.943 (**)	.638	.667
			Sig. (2-tailed)	.	.684	.	.005	.173	.148
			N	5	4	6	6	6	6
		TLearn	Corr. coef.	.857 (*)	-.116	.943 (**)	1.000	.721	.301
			Sig. (2-tailed)	.014	.827	.005	.	.068	.468
			N	7	6	6	8	7	8
		TCap	Corr. coef.	.563	.414	.638	.721	1.000	.726 (*)
			Sig. (2-tailed)	.146	.355	.173	.068	.	.027
			N	8	7	6	7	9	9
		KeyView questions	Corr. coef.	.790 (*)	.801 (*)	.667	.301	.726 (*)	1.000
			Sig. (2-tailed)	.011	.017	.148	.468	.027	.
			N	9	8	6	8	9	10

Corr. coef. = Correlation coefficient. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

Research Questions Outcomes

Expected outcomes to research questions were tested by examining the correlation of KeyView subscale questions (see Table 30).

> Expected Outcome 1 - *Searchers unfamiliar with the domain of a search topic will find the application useful for gaining an overview of the information space*.

The items examined were the helpfulness of search results, ease of determining document relevance, speed of relevance determination and keyphrase search on an unfamiliar topic. There is strong correlation between searching on an unfamiliar topic and the helpfulness of search results (49% score variance). Searching on an unfamiliar topic correlated with ease of determining document relevance - 49% score variance. Searching on an unfamiliar topic correlated with speed of relevance determination - 45% score variance. The correlations are all taken as an indicator that 'searchers unfamiliar with the domain of a search topic will find the application useful for evaluating the retrieved result set.'

> Expected Outcome 2 - *Searchers familiar with the domain of a search topic will find the application useful for quickly isolating their area of interest*.

The items examined were the helpfulness of search results, ease of determining document relevance, speed of relevance determination and keyphrase search on a familiar topic. Searching on a familiar topic had the strongest correlation to the helpfulness of search results (65% score variance). The speed of relevance determination (40% score variance) was lower than the correlation for ease of determining document relevance (46% score variance). This is still taken as an indicator that 'searchers familiar with the domain of a search topic will find the application useful for quickly isolating their area of interest.'

Expected Outcome 3 - *Searchers will not find four word keyphrases as valuable as shorter phrases*.

The items examined were the helpfulness of search results, ease of determining document relevance, speed of relevance determination, longer keyphrase and shorter keyphrases. Shorter and longer keyphrases both had strong correlations to helpfulness of search results, with shorter accounting for 48% of score variance, while longer keyphrases for 54% of score variance. Shorter keyphrases had 33% score variance vs. Longer keyphrases at 60% score variance, with ease of determining document relevance. Shorter keyphrases had 45% score variance vs. Longer keyphrases at 63% score variance, with ease of speed of relevance determination. Shorter keyphrases with Longer keyphrases - 29% score variance. Since longer keyphrases account for more score variance, this is taken as an indicator that 'Searchers will find longer keyphrases as valuable as shorter phrases or more valuable.' The assumption for the expected outcome, stemmed from searchers' tendency to use very short queries and their desire to exert minimum effort. It was thought that longer keyphrases might require more mental effort.

Qualitative Analysis of Survey Comments

There was one open ended question asking for comments about KeyView. The comments were analyzed by identifying positive, negative and mixed themes. Data was analyzed for 35 respondents, and 11 submitted written comments on the survey. The demographic breakdown is: five males and six females who wrote comments. Education level for comments: two respondents had some college (18%). Three were college graduates (27%). One had some post-graduate education (9%). Three had a postgraduate degree - masters or equivalent (27%). And two had a postgraduate degree - Ph.D. or equivalent (18%).

Findings

KeyView Criticisms

Five comments were critical of KeyView. Some respondents thought the application needed more work, and should provide more instruction or did not find it useful. Sample comments are:

- It needs work! I am not too sure what you are trying to accomplish with this...
- The lists of words have no known sequence that I can decipher and are way too long...
- Dull.

KeyView Positives

The three comments listed reflect the opinions of those respondents who felt positive about using the KeyView application. The comments are:

- It was very informative.
- I liked using the longer phrases, I found pages faster.
- My first keyphrase was very long - guyanese restaurants in georgia -- I did not get results for this. I shortened it to: guyanese restaurants. I found a restaurant in plano, texas!!

KeyView Mixed Reactions

The three comments listed reflect the opinions of those respondents that had both positive and negative reactions to the application. The comments are:

- I think it can be useful if you are looking for as much information as possible on a topic. Search time is a bit long though.
- Instruction was not very clear in a scene of explaining where application performs a search (all Web, some test collection, domain?). Term "results" was used referring to the phrases. User would think results are documents. Very useful application.

The following mixed comment was the longest comment submitted by a respondent:

1. The interface itself is very nice, nice lay-out and colors.
2. In the “Phrases” window one has to scroll down, concentrate and search through the list of matching multiple words. If one does not know exactly what term he/she is looking for, it’s hard to search for it. Besides, they are not in alphabetic order.
3. Saved Search Results would be more convenient to use if they were in alphabetic order.
4. The tool bar over the screen with icons Visit Website, Local text content, Related doc is very helpful
5. The results seem to be listed not in an order of reference ((there are many examples of this). The results are just not about the search terms.
6. Refine search brought even further from the target of search.

Data Analysis Summary

The data analysis presented in chapter 6 provided a basis for developing the conclusions and recommendations that follow in chapter 7. All usable surveys were analyzed and the respective information was discussed in narrative form and presented in tables where appropriate. The five sections of this chapter provided an in-depth look at survey demographic findings, the characteristics of the population, reliability analyses, correlations and Mann-Whitney tests. The chapter provided specific findings for the research questions, which will be summarized and discussed in chapter 7.

CHAPTER 7

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction

The purpose of this chapter is to summarize the significance, literature, methodology, procedures, and results of this study. On the basis of the results, conclusions and recommendations are provided in support of the following research questions:

1. Will a searcher find a keyphrase 'browsing' interface a valuable tool in evaluating search engine results? Analysis of the data found the answer to this question to be true overall.

   Sub-questions:

   a) Will searchers unfamiliar with the domain of a search topic find the application useful for gaining an overview of the information space?

The expected outcome was found to be true: 'Searchers unfamiliar with the domain of a search topic will find the application useful for evaluating the retrieved result set.'

   b) Will searchers familiar with the domain of a search topic find the application useful for quickly isolating their area of interest?

The expected outcome was found to be true: 'Searchers familiar with the domain of a search topic will find the application useful for quickly isolating their area of interest.'

2. Will searchers find longer keyphrases as valuable as shorter phrases?

The expected outcome, 'Searchers will not find longer keyphrases as valuable as shorter phrases,' was not found to be true. Data analysis indicates that searchers found longer keyphrases as valuable as shorter phrases; and perhaps more valuable.

Summary

The purpose of this study was to determine whether or not searchers would find a keyphrase browsing application a useful tool to evaluate their search engine results. The study was performed to examine user satisfaction of the KeyView phrase browsing application. The review of related literature showed that previous studies primarily used linguistic or syntactic techniques instead of statistical techniques to extract keyphrases. Prior studies had found keyphrase browsing applications useful. The results of this research also indicate keyphrase browsing applications are useful; whether searching on a familiar or unfamiliar topic.

Significance of the Study

The amount of information published on the Internet continues to increase exponentially each year. Search engines provide a degree of access to this information. Searchers need tools that will enable them to decide if the document matching their search query fulfills their information need. This technique can be applied to 'structured data' collections in proprietary formats and to 'unstructured data' on the Internet.

Limitations and Key Assumptions

The strength of this search tool is dependent on the data it collects. If the external search engine (AltaVista, Google, Yahoo etc.) is unavailable, then the tool cannot extract a list of URLs. If there are many missing pages or URLs whose content the data collector cannot 'grab,' the indexes may not be good representations of the retrieved set information space. This also applies to Web pages that use images, Flash, Shockwave etc. in place of simple text. The application assumes the presence of words.

The following portions of this chapter include the conclusions and subsequent recommendations drawn from the findings of these statistical analyses.

Conclusions

The conclusions of this study were based on the survey data collected and are assumed to be characteristic of the students in SLIS and general Internet users. The anticipated result of this study was that keyphrases would be found useful by searchers. This assumption was found to be true. It is hoped that the results of this study well help to inform the development of other keyphrase browsing applications.

Future Work and Recommendations

A prevalent complaint of the users of the KeyView application was the length of time taken to gather the search results. The application is not a fully fledged search engine, nor is it meant to be one. Ideally, the functionality of a keyphrase browser would be a built in feature of an information retrieval system, such as a search engine. The length of time taken to gather the results was dependent on network traffic and server loads. The busier the network or the server CPU, the longer the application takes to gather the contents. This was the bottleneck affecting the speed of the application. Index generation after all the pages were gathered was a short process. Once indexes were generated the results could be examined relatively fast. One way of addressing speed complaints in future development iterations; would be to return to the 50 URL limit used in the prototype or some other lower number. Another option would be to let the user decide; one user indicated that the application is useful for discovering ‘as much information as possible on a topic.’ Setting the limit on URLs too low would decrease the usability of the application for this type of searcher.

An unexamined question in this study was the prevalence of nouns in statistically generated phrases. Most examples of phrase browsing applications found during the literature review used a syntactic technique for phrase extraction, which indicated nouns as the dominant part of speech. User search queries were saved as part of their interaction with the KeyView application. The Wordnet::QueryData and Lingua::EN::Tagger Perl modules could be used to examine the parts of speech found in the keyphrase indexes.

Applications

Real World Applications of a Semantic Search Engine

Adding semantic information to a search engine improves basic searching. Information filtering, such as such as filtering spam, could also be affected. Improved information gathering software agents that scan the Web and locates pages of interest i.e. similar to RSS feeds that continually update when Websites or topics in the feed produce new content. Interactive information extraction or query expansion (the search tool developed in this research could be characterized as such): an initial user query is used to instantiate a list of generic extraction phrase queries automatically.

APPENDIX A

INFORMATION ACCESS PROCESSES

Standard Interaction Process:

1. Start with an information need.

2. Select a system and collections to search on.

3. Formulate a query.

4. Send the query to the system.

5. Receive the results in the form of information items.

6. Scan, evaluate, and interpret the results.

7. Either stop, or,

8. Reformulate the query and go to step 4.

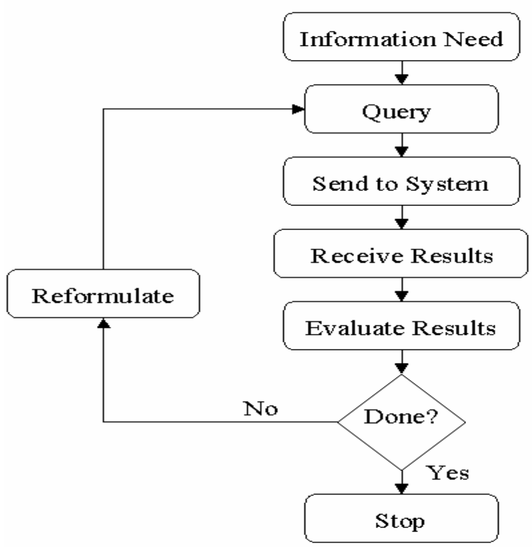


Figure 29. A simplified diagram of the standard model of information access processes (Hearst, 1999).

APPENDIX B

KEYVIEW QUESTIONNAIRE FOR USER INTERACTION SATISFACTION

KeyView Questionnaire for User Interaction Satisfaction

Welcome! This questionnaire assumes that you have used the KeyView Applications to obtain a collection of documents which were examined via keyphrases to determine document relevance to a search topic.

The intent of this survey is to get feedback on current and future development directions. Your input will help provide a better information retrieval tool.

First, a few background questions before starting the questionnaire.

Please answer the following questions about yourself

Into which of the following age groups do you fall?

○ 18 to 20
○ 21 to 24
○ 25 to 29
○ 30 to 34
○ 35 to 39
○ 40 to 44
○ 45 to 49
○ 50 to 54
○ 55 to 59
○ 60 to 64
○ 65 and over

Gender: ○ Female ○ Male

What is the highest level of education you completed?

○ High school or less
○ High school graduate
○ Some vocational school
○ Some college
○ Vocational school graduate
○ College graduate
○ Some post-graduate
○ Postgraduate degree - Masters or equivalent
○ Postgraduate degree - Ph.D. or equivalent

Have you used an information retrieval system before?
Examples include Google, Yahoo, MSN Search, Ask Jeeves, Lexis Nexsis, PubMed, Open Directory, Northern Light, Complete Planet, Library Catalog or other system.

○ Yes
○ No

OVERALL REACTION TO THE SOFTWARE		0	1	2	3	4	5	6	7	8	9		NA
1.	terrible	○	○	○	○	○	○	○	○	○	○	wonderful	○
2.	frustrating	○	○	○	○	○	○	○	○	○	○	satisfying	○
3.	dull	○	○	○	○	○	○	○	○	○	○	stimulating	○
4.	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
5.	inadequate power	○	○	○	○	○	○	○	○	○	○	adequate power	○
6.	rigid	○	○	○	○	○	○	○	○	○	○	flexible	○
7. Search results	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
8. Determine document relevance to search topic	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
9. Keyphrase search on an unfamiliar topic	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
10. Keyphrase search on a familiar topic	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
11. Determine document relevance to search topic	too slow	○	○	○	○	○	○	○	○	○	○	fast enough	○
12. Shorter keyphrases	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
13. Longer keyphrases	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○

Item		0	1	2	3	4	5	6	7	8	9		NA
SCREEN		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
14. Characters on the computer screen	hard to read	○	○	○	○	○	○	○	○	○	○	easy to read	○
15. Highlighting on the screen	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
15.1. Use of bolding	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
16. Screen layouts were helpful	never	○	○	○	○	○	○	○	○	○	○	always	○
16.1. Amount of information that can be displayed on screen	inadequate	○	○	○	○	○	○	○	○	○	○	adequate	○
16.2. Arrangement of information on screen	illogical	○	○	○	○	○	○	○	○	○	○	logical	○
17. Sequence of screens	confusing	○	○	○	○	○	○	○	○	○	○	very clear	○
17.1. Next screen in a sequence	unpredictable	○	○	○	○	○	○	○	○	○	○	predictable	○
17.2. Going back to the previous screen	impossible	○	○	○	○	○	○	○	○	○	○	easy	○
TERMINOLOGY AND APPLICATION INFORMATION		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
18. Use of terminology throughout application	inconsistent	○	○	○	○	○	○	○	○	○	○	consistent	○
18.1. Task related terminology	inconsistent	○	○	○	○	○	○	○	○	○	○	consistent	○
19. Messages which appear on screen	inconsistent	○	○	○	○	○	○	○	○	○	○	consistent	○
19.1. Position of instructions on the screen	inconsistent	○	○	○	○	○	○	○	○	○	○	consistent	○
20. Messages which appear on screen	confusing	○	○	○	○	○	○	○	○	○	○	clear	○
20.1. Instructions for commands or functions	confusing	○	○	○	○	○	○	○	○	○	○	clear	○
20.2. Instructions for correcting errors	confusing	○	○	○	○	○	○	○	○	○	○	clear	○
21. Application keeps you informed about what it is doing	never	○	○	○	○	○	○	○	○	○	○	always	○
21.1. Performing an operation leads to a predictable result	never	○	○	○	○	○	○	○	○	○	○	always	○
21.2. Length of delay between operations	unacceptable	○	○	○	○	○	○	○	○	○	○	acceptable	○
22. Error messages	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
22.1. Error messages clarify the problem	never	○	○	○	○	○	○	○	○	○	○	always	○
23 Phrasing of error messages	unpleasant	○	○	○	○	○	○	○	○	○	○	pleasant	○
LEARNING		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
24. Learning to operate the application	difficult	○	○	○	○	○	○	○	○	○	easy		
24.1. Getting started	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
24.2. Learning advanced features	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
24.3. Time to learn to use the application	slow	○	○	○	○	○	○	○	○	○	○	fast	○
25. Exploration of features by trial and error	discouraging	○	○	○	○	○	○	○	○	○	○	encouraging	○
26. Remembering names and use of commands	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
27. Tasks can be performed in a straight-forward manner	never	○	○	○	○	○	○	○	○	○	○	always	○
27.1. Number of steps per task	too many	○	○	○	○	○	○	○	○	○	○	just right	○
27.2. Steps to complete a task follow a logical sequence	never	○	○	○	○	○	○	○	○	○	○	always	○
27.3. Feedback on the completion of the steps	unclear	○	○	○	○	○	○	○	○	○	○	clear	○
APPLICATION CAPABILITIES		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
28. Application speed	too slow	○	○	○	○	○	○	○	○	○	○	fast enough	○
28.1. Response time for most operations	too slow	○	○	○	○	○	○	○	○	○	○	fast enough	○
28.2. Rate information is displayed	too slow	○	○	○	○	○	○	○	○	○	○	fast enough	○
29. The application is reliable	never	○	○	○	○	○	○	○	○	○	○	always	○
30. Correcting your mistakes	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
31. Ease of operation depends on your level of experience	never	○	○	○	○	○	○	○	○	○	○	always	○
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**

32.Please write your comments about KeyView here: (Optional)

Thanks for participating.

APPENDIX C

SEARCH PROCESS

Information Need

Submit a query to a search engine:

- Save *n* from the retrieved set, i.e. first 30 results, as a subset
- Visit each page and save a copy stripped of HTML
- Create variable length key phrase indexes

Stage 1

Use a term weighting formula to find frequently used terms in the subset:

- Rank terms, display a list of one word terms
- Get documents that contain term(s) or refine search

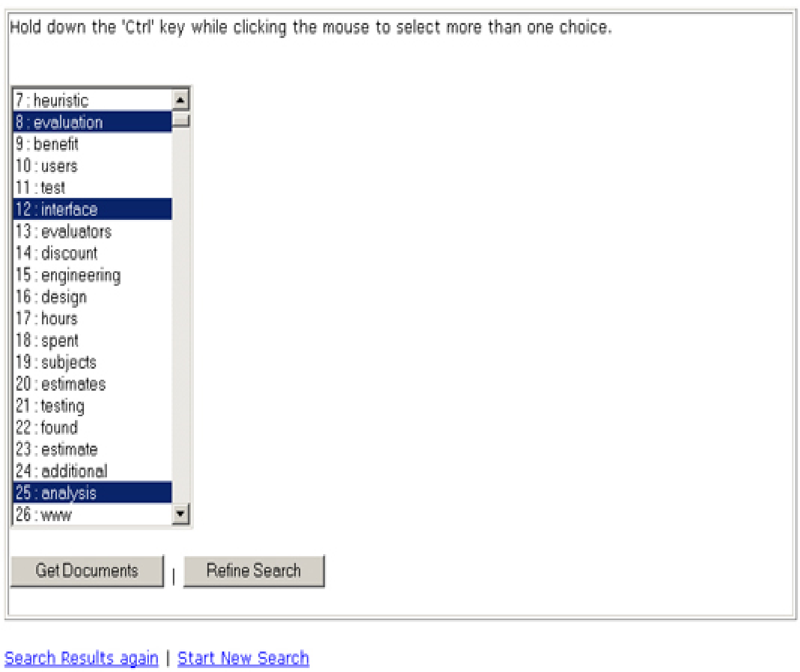


Figure 30. Stage 1 - selecting keyphrases.

Stage 1 Choice: Retrieving Documents

Get Documents:

- Show the search term in context
- Examine document via link to Web page or look at local copy stripped of HTML

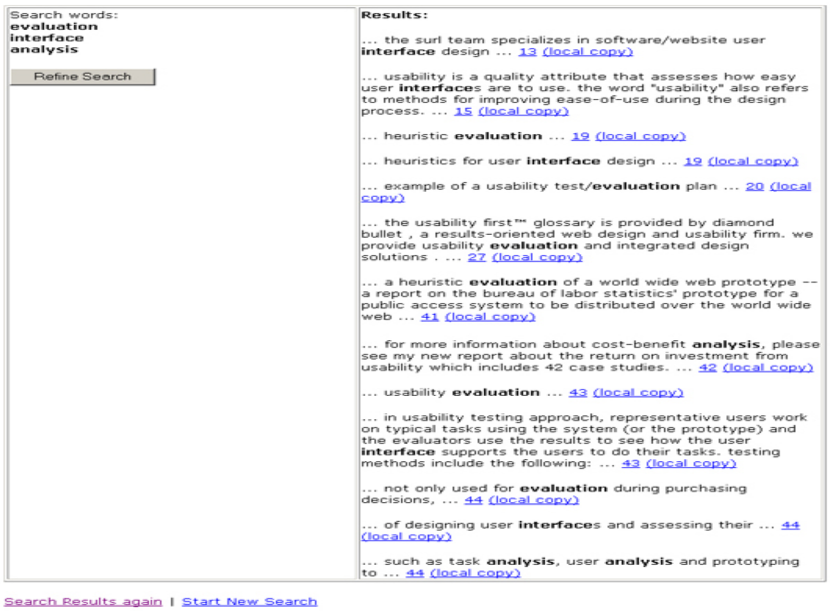


Figure 31. Stage 1 - retrieving keyphrases in contxt from documents.

Stage 1 Choice: 'Refine Search' leads to Stage 2

Stage 2 choices: Get Documents or Refine Search further

- Display words that contain the term(s) (single or multiple words)
- Get documents or refine search further

Note: Options to re-examine the subset by choosing from the initial list or begin a brand new search are always present.

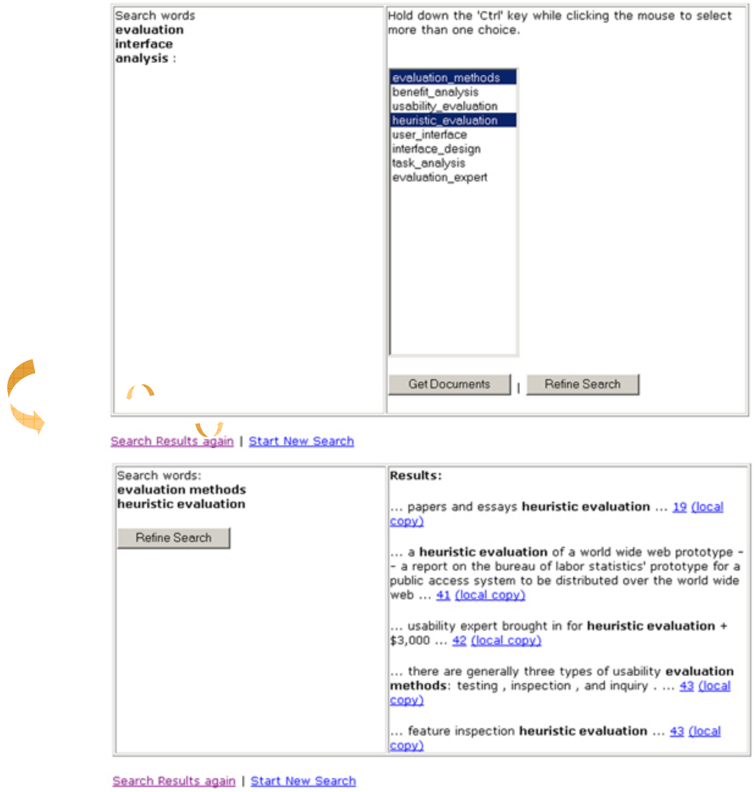


Figure 32. Stage 2 - choices lead to query feedback or document surrogate display.

Stage 3

Search further:

- Continue refining search by refining search term or displaying relevant documents
- Display ‘no match’ if terms cannot be refined further

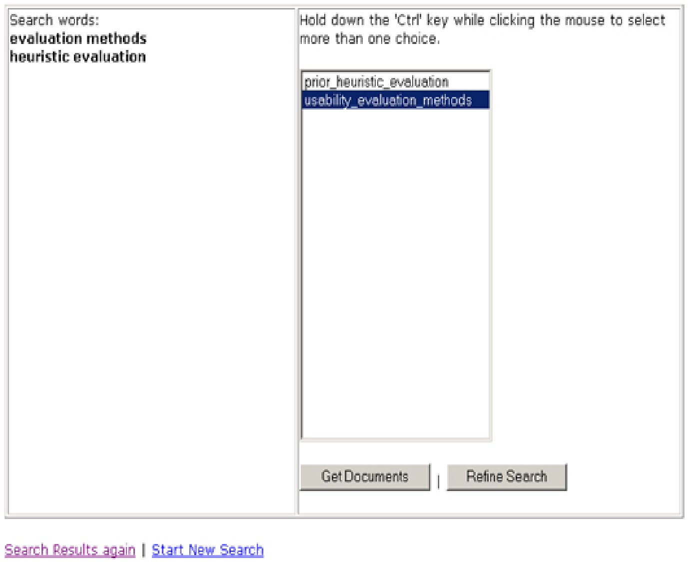


Figure 33. Stage 3.

Stage 4

Search further:

Once limit on the length of terms is reached, automatically display relevant document

- Examine document via link to Web page or look at local copy stripped of HTML
- Re-examine the subset by choosing from the initial list or begin a brand new search are always present

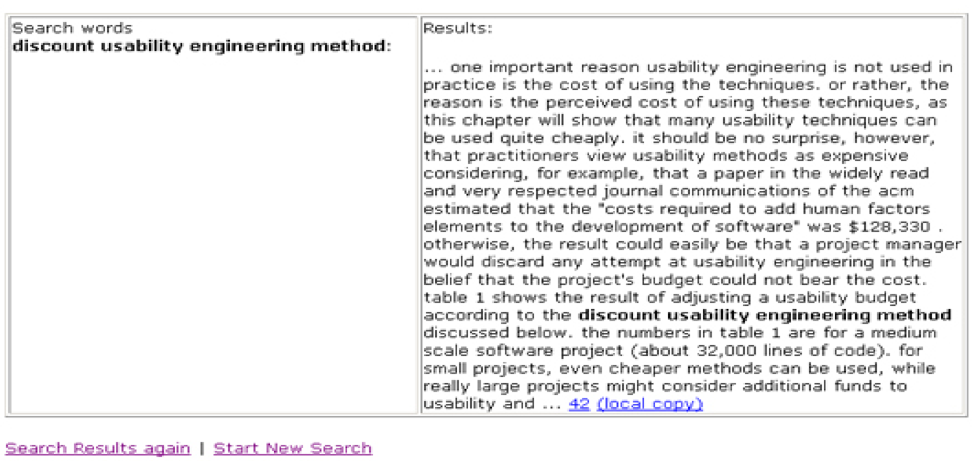


Figure 34. Stage 4.

APPENDIX D

SPREADING ACTIVATION MODEL

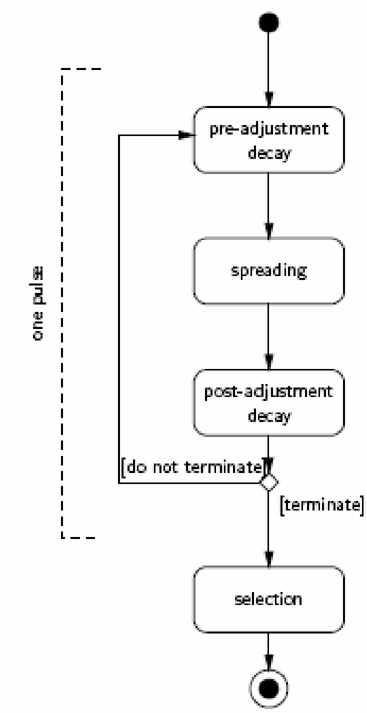


Figure 35. Spreading activation model.

- Iterative process
- Each iteration has

    1. One or more pulses: each pulse has three phases
    2. Pre-adjustment, decay: Output energy of a node is computed from activation level in the previous iteration
    3. Spreading: Input energy is accumulated for each node on the network
    4. Post-adjustment, decay: New activation level is computed from input energy and activation level in previous iteration
    5. Termination check: After a fixed number of iterations, or when other conditions are met, iteration stops.

APPENDIX E

KEYVIEW SUBSCALE QUESTIONS IN COMPARISON TO EACH OTHER

Table 30

KeyView Subscale Questions in Comparison to Each Other

		Search results (unhelpful/helpful)	Determine document relevance to search topic (difficult/easy)	Keyphrase search on an unfamiliar topic (unhelpful/helpful)	Keyphrase search on a familiar topic	Determine document relevance to search topic (too slow/fast enough)	Shorter keyphrases (unhelpful/helpful)	Longer keyphrases (unhelpful/helpful)
Search results (unhelpful/ helpful)	Corr. coef.	1.000	.836(**)	.718(**)	.787(**)	.768(**)	.691(**)	.743(**)
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000
	N	35	33	33	35	34	34	32
Determine document relevance to search topic (difficult/easy)	Corr. coef.	.836(**)	1.000	.686(**)	.653(**)	.679(**)	.549(**)	.772(**)
	Sig. (2-tailed)	.000	.	.000	.000	.000	.001	.000
	N	33	34	32	34	33	33	31
Keyphrase search on an unfamiliar topic (unhelpful/ helpful)	Corr. coef.	.718(**)	.686(**)	1.000	.792(**)	.659(**)	.648(**)	.745(**)
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000
	N	33	32	34	34	33	33	33
Keyphrase search on a familiar topic (unhelpful/ helpful)	Corr. coef.	.787(**)	.653(**)	.792(**)	1.000	.628(**)	.681(**)	.587(**)
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000
	N	35	34	34	36	35	35	33
Determine document relevance to search topic (too slow/fast enough)	Corr. coef.	.768(**)	.679(**)	.659(**)	.628(**)	1.000	.669(**)	.791(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000
	N	34	33	33	35	35	35	33
Shorter keyphrases (unhelpful/ helpful)	Corr. coef.	.691(**)	.549(**)	.648(**)	.681(**)	.669(**)	1.000	.536(**)
	Sig. (2-tailed)	.000	.001	.000	.000	.000	.	.001
	N	34	33	33	35	35	35	33
Longer keyphrases (unhelpful/ helpful)	Corr. coef.	.743(**)	.772(**)	.745(**)	.587(**)	.791(**)	.536(**)	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.001	.
	N	32	31	33	33	33	33	33

Corr. coef. = Correlation coefficient. **Correlation is significant at the 0.01 level (2-tailed).

Table 31

KeyView Subscale in Comparison to Five Scales

		TOverall	TScreen	TApp	TLearn	TCap
Search results (unhelpful/helpful)	Correlation Coefficient	.757(**)	.522(**)	.695(**)	.561(**)	.551(**)
	Sig. (2-tailed)	.000	.004	.000	.002	.002
	N	28	29	21	27	28
Determine document relevance to search topic (difficult/easy)	Correlation Coefficient	.592(**)	.519(**)	.616(**)	.534(**)	.478(*)
	Sig. (2-tailed)	.001	.005	.004	.004	.012
	N	28	28	20	27	27
Keyphrase search on an unfamiliar topic (unhelpful/helpful)	Correlation Coefficient	.776(**)	.669(**)	.821(**)	.555(**)	.583(**)
	Sig. (2-tailed)	.000	.000	.000	.003	.001
	N	28	29	22	26	28
Keyphrase search on a familiar topic (unhelpful/helpful)	Correlation Coefficient	.763(**)	.647(**)	.773(**)	.599(**)	.470(*)
	Sig. (2-tailed)	.000	.000	.000	.001	.010
	N	28	30	22	28	29
Determine document relevance to search topic (too slow/fast enough)	Correlation Coefficient	.807(**)	.355	.673(**)	.608(**)	.369(*)
	Sig. (2-tailed)	.000	.054	.001	.001	.049
	N	28	30	22	28	29
Shorter keyphrases (unhelpful/helpful)	Correlation Coefficient	.774(**)	.287	.607(**)	.544(**)	.193
	Sig. (2-tailed)	.000	.124	.003	.003	.316
	N	28	30	22	28	29
Longer keyphrases (unhelpful/helpful)	Correlation Coefficient	.749(**)	.483(**)	.671(**)	.463(*)	.501(**)
	Sig. (2-tailed)	.000	.008	.001	.017	.007
	N	28	29	22	26	28

**Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

APPENDIX F

IRB APPROVAL LETTER

UNT
UNIVERSITY OF
NORTH·TEXAS

DISCOVER THE POWER OF IDEAS

Office of Research Services

March 20, 2006

Onaje Johnston
School of Library and Information Sciences
University of North Texas

RE: Human Subjects Application No. 06-059

Dear Ms. Johnston:

Your proposal titled "KeyView Questionnaire for User Interaction Satisfaction" has been approved by the Institutional Review Board as permitted under federal law and regulations governing the use of human subjects in research projects 45 CFR 46.101. **Federal policy 45 CFR 46.109(e) stipulates that IRB approval is for one year only.**

Enclosed is the consent document with stamped IRB approval. Please copy and **use this form only** for your study subjects.

It is your responsibility according to U.S. Department of Health and Human Services regulations to submit annual and terminal progress reports to the IRB for this project. Please mark your calendar accordingly. The IRB must also review this project prior to any modifications.

Please contact Shelia Bourns, Research Compliance Administrator, ext. 3940 or Boyd Herndon, Director of Research Compliance, ext. 3941, if you wish to make such changes or need additional information.

Sincerely,

Scott Simpkins, Ph.D.
Chair
Institutional Review Board

SS:sb

P.O. Box 305250
Denton, Texas 76203-5250
940.565.3940 TEL
940.565.4277 FAX
940.369.8652 TTY
www.unt.edu

Informed Consent Notice for Internet Users

My name is Onaje Johnston and I am a graduate student in the School of Library and Information Sciences Department at the University of North Texas. I am conducting an online study that involves a web based search application that uses document keyphrases instead of keywords, to determine the relevance of a document to a search topic. The purpose of this study is to investigate user satisfaction as it relates to using a keyphrase search application (KeyView).

If you agree to take part in this study, you will be asked to complete a questionnaire about using KeyView. It will take approximately 30 minutes to complete. Your responses may help to provide a greater understanding of the factors contributing to the design of an information retrieval application to allow faster relevance judgments of documents found by the application.

Participation in this study is completely voluntary. You have the right to skip any question you choose not to answer. There are no foreseeable risks involved in this study; however, if you decide to withdraw your participation you may do so at any time by simply leaving the web site.

Your name will not be requested in this study so your responses will be anonymous. All research records will be kept confidential by the Principal Investigator. No individual responses will be disclosed to anyone because all data will be reported on a group basis. If you have any questions about the study, please contact *Onaje Johnston* at telephone number 940-565-3561 or *Dr. Brian O'Connor*, UNT Department of SLIS, at telephone number 940-565-2347.

This research project has been reviewed and approved by the UNT Institutional Review Board (IRB). Please contact the UNT IRB at 940-565-3940 with any questions regarding your rights as a research subject.

If you agree to participate, you may print this document for your records.

By clicking below, you are confirming that you are at least 18 years old and you are giving your informed consent to participate in this study.

Click Here To Enter Study

APPROVED BY THE UNT IRB
FROM 3/20/06 TO 3/19/07

APPENDIX G

QUIS LICENSE AGREEMENT

1. Definitions: "QUIS^{tm}" means the "Questionnaire for User Interaction Satisfaction" (Copyright © 1984, 1993, 1998. University of Maryland. All rights reserved.). "Licensed Materials" means the electronic and paper versions of QUIS^{tm} and all documentation included in this package and any modifications or updates of said materials delivered to Licensee. "Licensor" means the University of Maryland. "Licensee" means the individual or organization licensing and opening this package.

2. Grant of License: In consideration of the payment of the fees and charges paid and the obligations undertaken by Licensee, Licensor grants to Licensee a nonexclusive, nontransferable license to the Licensed Materials.

3. Scope and Limitations of Rights:

QUIS^{tm} SITE LICENSE: Licensor grants to Licensee the right to copy, modify, distribute and use the Licensed Materials throughout a single enterprise located within one building, or buildings, all of which are addressable by only a single postal address.

QUIS^{tm} INTERNET LICENSE: In addition to the rights granted under the Site License, the Internet License grants to Licensee, IF AND ONLY IF LICENSEE HAS FULLY PAID FOR A QUIS^{tm} INTERNET LICENSE, the right to display QUIS^{tm} on its Website but only for the purpose of gathering information on Licensee's publicly distributed software products, and explicitly not for downloading. IF LICENSEE DISPLAYS QUIS^{tm} ON THE INTERNET, LICENSEE AGREES TO DISPLAY IN A CONSPICUOUS AND LEGIBLE FASHION THE COPYRIGHT NOTICE AND A WARNING THAT QUIS^{tm} IS PROTECTED BY COPYRIGHT AND THAT TO DOWNLOAD QUIS^{tm} WOULD BE UNLAWFUL.

4. Support and Operation:

4.1 Licensor's obligation for support and maintenance of the Licensed Materials shall be limited to that specified herein. Licensor will support only the most current version of the Licensed Materials. Licensee is responsible for the application and implementation of new releases, computer program code corrections, and updates to the documentation issued to Licensee by Licensor. Licensor is not responsible for use of superseded, outdated or uncorrected versions of the Licensed Materials nor for obsolescence of the Licensed Materials that may result from changes in Licensee's requirements or software or equipment not supplied by Licensor.

4.2 Licensor's obligation for support and maintenance shall only apply to correction of verifiable and reproducible errors and shall be limited to providing commercially reasonable efforts to correct such errors. Licensor shall not be responsible for the correction of any error attributable to Licensee's misuse or improper use of Licensed Materials, nor shall Licensor be responsible for maintaining computer program code which has been modified after delivery by Licensor.

5. Proprietary Protection:

5.1 Licensee acknowledges that the Licensed Materials are a commercially valuable proprietary product of Licensor and that Licensor is the owner of all right, title and interest in and to the Licensed Materials and all modifications and enhancements thereto (including ownership of all trade secrets, copyrights and patents pertaining thereto and subsisting therein) subject only to the rights and privileges expressly granted by Licensor. This agreement grants

Licensee only a right of limited use, revocable in accordance with the terms hereof. Licensee shall keep the Licensed Materials free and clear of all claims, liens, and encumbrances.

5.2 Licensee shall not use, distribute, allow access to, copy, or modify the Licensed Materials, or any copy, adaptation, transcription or merged portion thereof, except as expressly authorized by this Agreement.

5.3 Licensee's obligations hereunder shall survive termination of this Agreement and shall remain in effect for as long as Licensee continues to use, possess or have access to the Licensed Materials.

6. Disclaimer of Warranty and Limitation of Liability:

6.1 THE LICENSED MATERIALS ARE MADE AVAILABLE ON AN "AS IS" BASIS. EXCEPT AS EXPRESSLY SET FORTH IN THIS AGREEMENT, LICENSOR DISCLAIMS ANY AND ALL PROMISES, REPRESENTATIONS AND WARRANTIES BOTH EXPRESS AND IMPLIED WITH RESPECT TO THE LICENSED MATERIALS AND ANY SUPPORT SERVICES PROVIDED HEREUNDER, INCLUDING THEIR CONDITION, CONFORMITY TO ANY REPRESENTATION OR DESCRIPTION, THE EXISTENCE OF ANY PATENT DEFECTS THEREIN, AND THEIR MERCHANTABILITY OR FITNESS FOR A PARTICULAR USE OR PURPOSE.

6.2 In no event shall Licensor be liable for any business expense, machine down time, loss of profits, any incidental, special, exemplary or consequential damages, or any claims or demands brought against Licensee or Licensee's customers, even if Licensor has been advised of the possibility of such claims or demands. This limitation upon damages and claims is intended to apply without regard to whether other provisions of this Agreement have been breached or have proven ineffective.

7. Term and Termination:

7.1 This agreement shall automatically terminate if Licensee uses, distributes, allows access to, copies, or modifies the Licensed Materials, or if Licensee transfers possession of any copy, adaptation, transcription or merged portion of the Licensed Materials to any other party in any way not expressly authorized by this Agreement.

7.2 Termination of this agreement shall not relieve either party of the obligations arising hereunder before termination. Upon termination, Licensee must return, or certify the destruction of (and erasure from any storage device), all copies of the Licensed Materials.

8. General:

8.1 It is agreed that the laws of the State of Maryland will govern without reference to conflict of law principles. Any and all legal actions must be brought in the courts in the State of Maryland or in the U.S. District Court for the District of Maryland. Licensee consents to the jurisdiction of said courts.

8.2 No modification of this Agreement shall be binding unless it is written and signed by an authorized representative of the party against whom enforcement of the modification is sought.

8.3 Neither this Agreement, the License granted hereby, nor any part thereof may be assigned or otherwise transferred by Licensee.

8.4 In the event that any of the terms of this Agreement is, or becomes, or is declared to be invalid or void by any court or tribunal of competent jurisdiction, such term or terms shall be null and void and shall be deemed severed from this Agreement and all the remaining terms of this Agreement shall remain in full force and effect.

8.5 No action, regardless of form, arising out of this Agreement, except an action by the University of Maryland for enforcement of its intellectual property rights or damages resulting from Licensee's breach thereof, may be brought by either party more than one (1) year after the cause of action has arisen.

Should you have any questions concerning this Agreement, or if you wish to contact the University of Maryland for any reason, please write:

Office of Technology Commercialization
University of Maryland
6200 Baltimore Avenue, Suite 300
College Park, Maryland 20742-9520

REFERENCES

Anick, P. G. (1997). Exploiting clustering and phrases for context-based information retrieval. Proceedings of the 20th annual international ACM SIGIR conference on research and development in information retrieval (pp. 314 - 323). Philadelphia, Pennsylvania, United States.

Baeza-Yates, R., Ribeiro-Neto, Berthier. (1999). *Modern information retrieval* (1st ed.): Addison Wesley Longman Publishing Co. Inc.

Berenci, E., Carpineto, C., and Giannini, V. (1998). Improving the effectiveness of Web search engines using selectable views of retrieval results. *Journal of Universal Computer Science, 4*(9), 737-747.

Blair, D. C. (in progress). Wittgenstein, language and information-"Back to the rough ground!"

Brooks, T. A. (1998). The semantic distance model of relevance assessment. Paper presented at the 61st Annual Meeting of the American Society for Information Science and Technology (ASIS 1998): Information access in the global information economy, Pittsburgh, PA.

Bush, V. (1945). As We May Think. *The Atlantic Monthly, 176*(1), 101-108.

Caropreso, M. F., Matwin., Stan., and Sebastiani, Fabrizio. (2000). Statistical phrases in automated text categorization, Technical Report IEI-B4-07-2000: Pisa, IT.

Ceglowski, M., Coburn, Aaron. & Cuadrado, John. (2003). Semantic search of unstructured data using contextual network graphs: National Institute for Technology and Liberal Education, Middlebury College.

Choo, C. W., Detlor, Brian and Turnbull, Don. (1998). A behavioral model of information seeking on the Web - Preliminary results of a study of how managers and IT specialists use the Web. Paper presented at the 61st Annual Meeting of the American Society for Information Science and Technology (ASIS 1998): Information access in the global information economy, Pittsburgh, PA.

Choo, C. W., Detlor, Brian and Turnbull, Don. (2000). Information seeking on the Web: An integrated model of browsing and searching, *First Monday, 5*.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.): Lawrence Erlbaum Associates.

Crestani, F. (1997a). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review, 11*(6), 453 - 482.

Crestani, F. (1997b). Retrieving documents by constrained spreading activation on automatically constructed hypertexts. Paper presented at the EUFIT ‘97 - 5th European Congress on Intelligent Techniques and Soft Computing, Aachen, Germany.

Croft, B., Turtle, H. and Lewis, D. (1991). The use of phrases and structured queries in information retrieval. Proceedings of the 14th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval (pp. 32-45). Chicago, IL.: ACM Press.

Downes, S. (2004, 6/17/2004). Whither the semantic Web. Retrieved 3/18/2005, from http://www.downes.ca/cgi-bin/website/view.cgi?dbs=Article&key=1087499522

Doyle, L. B. (1961). Semantic road maps for literature searchers. *Journal of the ACM, 8*(4), 553-578.

Folksonomy. (2006, August 2). In *Wikipedia, The Free Encyclopedia*. Retrieved 23:02, August 4, 2006, from http://en.wikipedia.org/w/index.php?title=Folksonomy&oldid=67200104.

Gutwin, C., Paynter, Gordon., Witten, Ian., Nevill-Manning, Craig., & Frank, Eibe. (1999). Improving browsing in digital libraries with keyphrase indexes. *Journal of Decision Support Systems, 27*(1-2), 81-104.

Hardin, S. (2005a). JC Herz addresses first 2004 Anuual Meeting Plenary Session. *Bulletin of the American Society for Information Science and Technology, 31*(3), 10-11.

Hardin, S. (2005b). Tim Berners-Lee: The semantic Web - Web of machine-processable data. *Bulletin of the American Society for Information Science and Technology, 31*(3), 12-13.

Harper, B., Slaughter, L., & Norman, K. (1997). Questionnaire administration via the WWW: A validation and reliability study for a user satisfaction questionnaire. Paper presented at the WebNet 97, Association for the Advancement of Computing in Education, Toronto, Canada.

Hearst, M. (1999). User interfaces and visualization. In R. a. R.-N. Baeza-Yates, Berthier. (Ed.), *Modern information retrieval.* Addison-Wesley-Longman Publishing co.

Human-computer Interaction Lab (HCIL). Quis information. Retrieved 12/08/2005 from University of Maryland, Human-computer Interaction Lab: www.cs.umd.edu/hcil/quis/

Johnston, O. (2006, January). Engineering a semantic search tool. Works-in-Progress Poster Session presented at 2006 ALISE Annual Conference.

Jones, S., Paynter, Gordon W. (2001). Human evaluation of Kea, an automatic keyphrasing system. Paper presented at the Proceedings of the 1st ACM/IEEE-CS joint conference on Digital Libraries, Roanoke, Virginia, United States.

Kerns, J. (2003). Foraging for Relevance. In *Hunting and gathering on the information savanna: conversations on modeling human search abilities*. Lanham, Md: Scarecrow Press.

Klimesch, W. (1994). *The Structure of long term memory: A connectivity model of semantic processing*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

Kline, Rex B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Lyman, P., and Varian, Hal R. (2003, 10/27/2003). How much information? 2003. Retrieved 8/45/2005, 2005, from http://www.sims.berkeley.edu/how-much-info

Mitra, M., Buckley, Christopher., Singhal, Amit., and Cardie, Claire. (1997). An analysis of statistical and syntactic phrases. Paper presented at the Proceedings of RIAO-97, 5th International Conference, Montreal, CA.

Murtagh, F., Taskaya, T., Contreras, P., Mothe, J. and Englmeier, K. (2003). Interactive visual user interfaces: A Survey. *Artificial Intelligence Review, 19*(4), 263 - 283.

Nielsen, J., Loranger, Hoa. (2006). *Prioritizing Web usability* (1 ed.): New Riders Press.

Nielsen, J. (1994a). How to conduct a evaluation. Retrieved 11/28/2005, from http://www.useit.com/papers/heuristic/heuristic_evaluation.html

Nielsen, J. (1994b). Ten usability heuristics. Retrieved 11/28/2005, from http://www.useit.com/papers/heuristic/heuristic_list.html

O'Connor; Brian Clark., Copeland, Jud H., and Kearns, Jodi L. (2003). *Hunting and gathering on the information savanna: Conversations on modeling human search abilities*. Lanham, Md: Scarecrow Press.

Oyarce, G. A. (2004, November 13 - 18). Kontextual interactive support system. Poster Session at the Proceeding of the 67th Annual Meeting of the American Society for Information Science and Technology (ASIST 2004), Providence, Rhode Island.

Oyarce, G. A. (2005a). "A contextual phrase analyzer" Pending U.S. Patent Application No. 11/374,452. Washington, DC: U.S. Patent and Trademark Office.

Oyarce, G. A. (2005b). "A contextual interactive support system" Pending U.S. Patent Application No. 11/373,886. Washington, DC: U.S. Patent and Trademark Office.

Paynter, G. W., Witten, Ian H., Cunningham, Sally Jo. and Buchanan, George. (2000). Scalable browsing for large collections: a case study. Paper presented at the Proceedings of the Fifth ACM International Conference on Digital Libraries.

Pickens, J., and Croft, B. (2000). An exploratory analysis of phrases in text retrieval. Proceedings of RIAO 2000 Conference (pp. 1179-1195). Paris, France.

Preece, S. E. (1981). A spreading activation model for information retrieval. PhD Dissertation. University of Illinois, Urbana-Champaign.

Rosch, E., & Lloyd, B.B. (1978). Principles of categorization. In *Cognition and Categorization*. Hillsdale, NJ.: Lawrence Erlbaum.

Salton, G. (2000). Automatic text indexing using complex identifiers, Proceedings of the ACM conference on Document processing systems (pp. 135 - 144). Santa Fe, New Mexico, United States.

Salton, G., Buckley, Chris. (1988). On the use of spreading activation methods in automatic information retrieval. Paper presented at the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Grenoble, France.

Sloman, S. A., Lagnado, D.A. (2005). The problem of induction. In K. H. R. Morrison (Ed.), *The cambridge handbook of thinking and reasoning*: Cambridge University Press.

Smeaton, A. F., Kelledy, F. (1998). User-chosen phrases in interactive query formulation for iInformation retrieval. Paper presented at the 20th Annual BCS-IRSG Colloquium on IR, Autrans, France.

Spoerri, A. (1995). InfoCrystal: A visual tool for information retrieval. Massachusetts Institute of Technology.

Tag cloud. (2006, August 4). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:59, August 4, 2006, from http://en.wikipedia.org/w/index.php?title=Tag_cloud&oldid=67724577.

Turney, P. (2003). Coherent keyphrase extraction via Web mining. Paper presented at the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03), Acapulco, Mexico.

Turtle, H., and Croft, W. Bruce. (1992). A comparison of text retrieval models. *The Computer Journal, 35*(3), 279-290.

Wacholder, N., Nevill-Manning, Craig G. (2001). The technology of phrase browsing applications: Workshop held in conjunction with the first ACM-IEEE joint conference on digital libraries. *SIGIR Forum, 35*(1), 16-20.

Withrow, J. (2003, 8/11/2003). Cognitive psychology & IA: From theory to practice. Retrieved 3/29/2005, from http://www.boxesandarrows.com//archives/cognitive_psychology_ia_from_theory_to_practice.php

Wu, Y.-F. B., Rakthin, Chatchai. and Li, Changzhi. (2002, 7/18/2002). Improve searching in large document collections using automatic table-of-contents interface. Paper presented at the Proceedings of the Workshop on Document Search Interface Design and Intelligent Access in Large-Scale Collections, JCDL'02, Portland, OR.

Zamir, O., Etzioni, Oren. (1999). Grouper: A dynamic clustering interface to Web search results. Paper presented at the Eighth International World Wide Web Conference, Toronto, Canada.