USING SOCIAL ANNOTATIONS TO IMPROVE WEB SEARCH

by

Worasit Choochaiwattana

B.B.A. (Marketing), Chulalongkorn University, Thailand, 1996

M.Sc. (Tech of Info Sys Mgt), Mahidol University, Thailand, 2000

M.Sc. (Computer Science), Chulalongkorn University, Thailand, 2003

Submitted to the Graduate Faculty of

School of Information Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

SCHOOL OF INFORMATION SCIENCES

This dissertation was presented

by

Worasit Choochaiwattana

It was defended on

April 17, 2008

and approved by

Peter Brusilovsky, PhD, Associate Professor, School of Information Sciences

Daqing He, PhD, Assistance Professor, School of Information Sciences

Stephen Hirtle, PhD, Professor, School of Information Sciences

Brian Butler, PhD, Associate Professor, Joseph M. Katz Graduate School of Business and College of Business Administration

Dissertation Advisor: Michael Spring, PhD, Associate Professor, School of Information Sciences

USING SOCIAL ANNOTATIONS TO IMPROVE WEB SEARCH

Worasit Choochaiwattana, PhD

University of Pittsburgh, 2008

Web-based tagging systems, which include social bookmarking systems such as Delicious, have become increasingly popular. These systems allow participants to annotate or tag web resources. This research examined the use of social annotations to improve the quality of web searches. The research involved three components. First, social annotations were used to index resources. Two annotation-based indexing methods were proposed: annotation based indexing and full text with annotation indexing. Second, social annotations were used to improve search result ranking. Six annotation based ranking methods were proposed: Popularity Count, Propagate Popularity Count, Query Weighted Popularity Count, Query Weighted Propagate Popularity Count, Match Tag Count and Normalized Match Tag Count. Third, social annotations were used to both index and rank resources. The result from the first experiment suggested that both static feature and similarity feature should be considered when using social annotations to re-rank search result. The result of the second experiment showed that using only annotation as an index of resources may not be a good idea. Since social Annotations could be viewed as a high level concept of the content, combining them to the content of resource could add some more important concepts to the resources. Last but not least, the result from the third experiment confirmed that the combination of using social annotations to rank the search result and using social annotations as resource index augmentation provided a promising rank of search results. It showed that social annotations could benefit web search.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation could not have completed without the generous support of my advisor, Prof. Michael Spring. He was always there when I needed him. He has patiently given me guidance over the past five years. I would not have made it without him. I greatly appreciate all the support and suggestions from my dissertation committee members, Prof. Peter Brusilovsky, Prof. Daqing He, Prof. Stephen Hirtle, and Prof. Brian Butler.

I am thankful for my friends, Tanapat Anusas-amornkul, Pattarana Sae-Chew and Nutchanon Satanyasuwan. They were always there for me as I prepared my presentations for the comprehensive examination, proposal defense and dissertation defense. They also helped me with everything while I was in Pittsburgh. I would like to thank Chirayu Wongchokprasitti for helping me set up Indri and showing me how to use Indri to create indexes of web resources; Jon Walker for proofreading my writing in this dissertation; Jonathan Grady, Sue Yeon Syn, Nathan Sulinski, and Jumpol Polvichai for their opinions and suggestions.

I would really like to thank my parents, Aphichat Chuchaiwattana and Jiralad Chuchaiwattana, my sister, Piyamon Chuchaiwattana, my brother and my sister-in-law, Thossaporn Chuchaiwattana and Tassawan Chuchaiwattana, and my 11-month-old niece, Aksarapak Chuchaiwattana for their never-ending support and encouragement. They flew all the way from Bangkok, Thailand to Pittsburgh, USA to attend my dissertation defense. I always have a good time when I am with them. I would like to thank my best friend, Kitti Pooltavee, for

his never-ending support and encouragement. He was always with me when there was something wrong.

Finally, I would like to thank Dr. Nucharee Premchaiswadi and Dhurakij Pundit University for providing financial support for my doctoral studies at the University of Pittsburgh.

1.0 INTRODUCTION

Web-based tagging systems, which include social bookmarking systems such as Delicious, allow participants to annotate or tag a particular resource. Historically, annotations have been used in several ways. Students annotate their books to emphasize interesting sections, to summarize ideas and to comment on what they have read (Marshall, 1998, Wolfe, 2000, O'Hara and Sellen, 1997). Davis and Huttenlocher (1995) suggested that shared annotations in the educational context can serve as a communication tool among students and between students and instructors. There can be threaded discussions around class materials that directly and precisely link to the class material (Brush *et al.*, 2002, Kurhila *et al.*, 2003). Farzan and Brusilovsky (2005, 2006) made use of annotation as an indicator of the page relevance for a group of learners in an online learning system.

Web systems that allow for *Social Annotation* can provide useful information for various purposes. Dmitriev *et al.* (2006) explored the use social annotation to improve the quality of enterprise search. Freyne *et al*. (2007) made use of social annotation to re-rank research paper search results. Hotho *et al.* (2006) proposed a formal model and a new search algorithm for folksonomies. Bao *et al.* (2007) explored the use of social annotation to improve web search.

Social annotations have the potential to improve searching for resources (Marlow *et al.*, 2006). However, published research on using social annotations to improve web search is sparse. Bao *et al.* (2007) explore the use of social annotation to improve web searches. In the following

section, a brief description of Delicious is presented and then the motivation for integrating social annotations into web searches is discussed.

1.1 DELICIOUS

Delicious (http://del.icio.us/), is a web based bookmark sharing system developed by Joshua Schachter in late 2003 and is now part of Yahoo. It can be viewed as a non-hierarchical keyword categorization system. The main objective of the system is to allow participants to store, share, and discover web bookmarks – which point at resources. They can create tags for the resources with any number of freely chosen keywords.

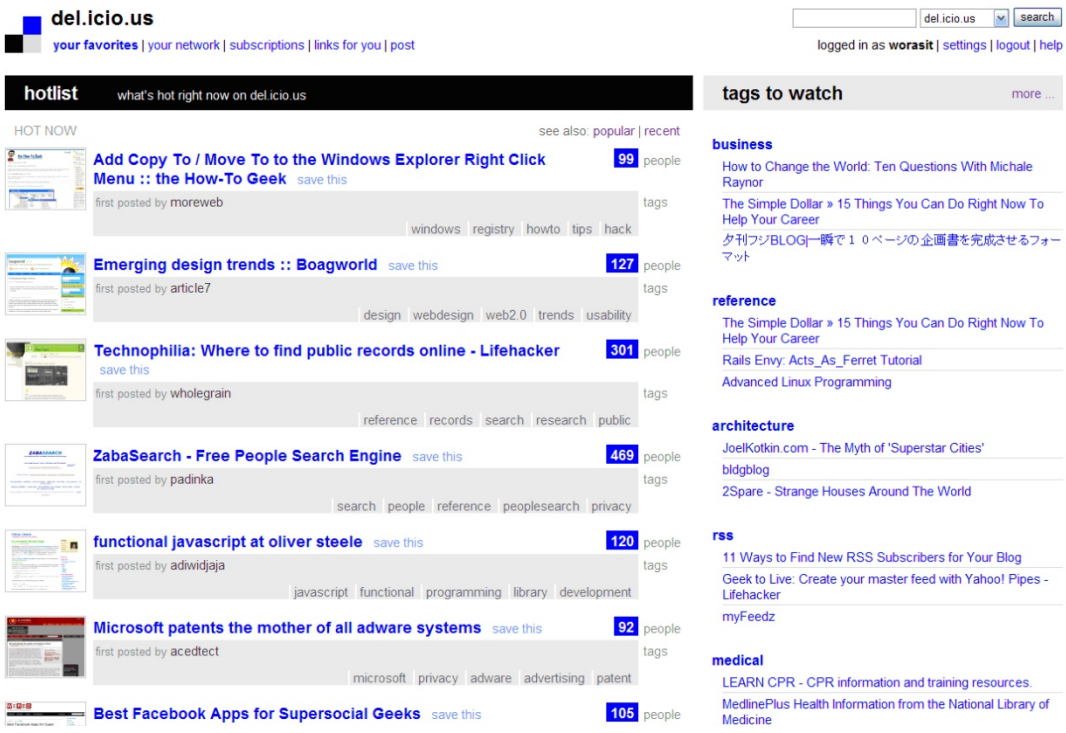


Figure 1. The main screen of Delicious

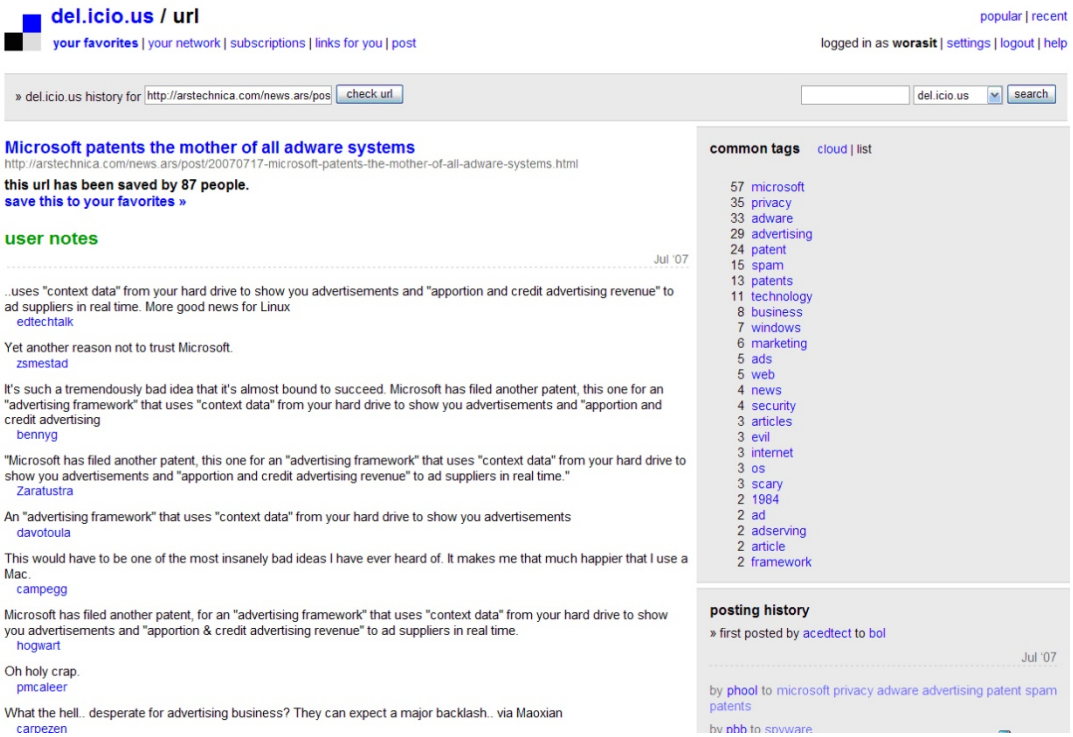


Figure 2. A screen showing the details of a web resource bookmarking on Delicious

The Delicious website has more than one million registered participants[1]. The tags or annotations created by participants provide information that can be used in various ways. With a combination of everyone's bookmarks, interesting resources can be identified. Figure 1 and figure 2 show a main screen of Delicious and a screen showing the details of a web resource bookmarking on Delicious respectively.

This research explores whether tags created by participants can be used to improve the quality of web search.

[1] http://blog.del.icio.us/ (Mar 29, 2007)

1.2 MOTIVATION

To understand challenges of using social annotation to improve web search, this section reviews motivation of this research.

1.2.1 Using Social Annotations as Resource Indexes

Full text indexing is of the basis of information retrieval. Numerous variations on indexing have been tried over the years. Modern search engines use several methods to find additional metadata information to improve a resource indexing for enhancing the performance of the similarity ranking. As examples, document title (Hu *et al.*, 2005), anchor text (Brin and Page, 1998, Caswel *et al*., 2001, Westveld *et al*., 2002, Eiron and McCurley, 2003) and user query log (Xue *et al*., 2005) have all been used.

Craswell *et al.*, (2001) and Westerveld *et al.*, (2002) explored the use of links and anchors for web resource retrieval. They pointed out that anchor text helps improve the quality of search results significantly. The anchor text can be viewed as web page creator annotation. This suggests that annotation can be used to support document indexing.

Social tagging systems, e.g. Delicious, allow participants to add keywords that are tags to a web resource. These tags can be viewed as user annotations of a web resource. Dmitriev *et al.* (2006) explored the use of user annotation as intranet document indexes. Yanbe *et al*. (2007) converted a tag and its frequency to be a vector that represents a page's content.

Taken together, this suggests that there is potential value to investigate how well social annotations, which are viewed as user annotations, contribute to the search results when they are used as resource indexes.

Figure 3. Social annotations of two web resources on Delicious

1.2.2 Search Result Ranking with Social Annotation

Because of the large number of documents returned using full text indexing, ranking of search results becomes critical in improving the quality of the search experience for the users. Similarity ranking, or query-dependent ranking, measures the match between a query and the content of the

web resource. Various approaches to ranking have been used. Static ranking, also known as query-independent ranking, measures the quality of the web documents e.g. PageRank (Page *et al.*, 1998) and fRank (Richardson *et al.*, 2006).

Resource searching in social tagging systems such as Delicious is limited. In Delicious, there is no content driven ranking of the results returned from searches. The results are ordered in reverse chronological order (Hotho *et al.*, 2006). Hotho *et al.* proposed *Adapted PageRank*, which is based on the idea that a URL which is tagged with important tags by important users becomes important itself. In addition, Bao *et al.* (2007) proposed a novel algorithm, called *SocialPageRank*. The idea behind the algorithm is that high quality web pages are usually popularly annotated by many up-to-date users with popular annotation.

Both *Adapted PageRank* and *SocialPageRank* use the count of tags as a main element to create the static ranking. They do not consider other factors such as annotation spamming. Using only the count of tags can be really harmful when annotation spamming has occurred as was the case for Google. Because of the complexity of the algorithms, the speed of the algorithm may slow down if the scale of social annotations keeps growing exponentially. The challenge is to propose a new practical ranking algorithm, which exploits the use of social annotations.

1.3 RESEARCH OBJECTIVES

This research proposes a new approach to the use of social annotations to improve the quality of web search. *The main objective of the research is to design and implement a framework for effective use of social annotations to improve user experience of web search focused on ranking search results.* The framework exploits the use of social annotations for resource index

augmentation and as a search result ranking method. To accomplish this objective, a method for using social annotations as web resource indexes is developed and a social annotation based ranking is proposed.

1.4 ORGANIZATION OF THE PAPER

This dissertation is organized as follows. The second chapter reviews and describes related work on annotation and using annotation to improve search results. The third chapter addresses research question, describes the research design, and shows the results from a preliminary study. The fourth chapter describes and discusses the results from the experiments. The final chapter provides a conclusion and describes a future work.

1.5 DEFINTIONS OF TERMS

1.5.1 Annotation

Annotation, as defined by Webster's Universal Encyclopedic Dictionary, "*is a note added for commenting or explanation*". Nagao (2003) defined annotation as "*a sort of commentary or explanation, or the act of producing such a commentary.*" He concluded that annotation could be viewed as mark up for extra information in a context. Euzenat (2002) defined annotation as "*a set of formal representations that attach to the content of the document*". In W3C's Amaya project, "*annotations are comments, notes, explanations, or other types of external remarks that*

can be attached to a Web document or a selected part of the document" (Vatton *et al*., 2004). Petkovic *et al.* (2004) wrote "*annotation can represent comments or remarks that users create for themselves or for others, referring to a specific piece of content (word, paragraph, image region etc.)*". This definition highlights the locality of a specific piece of content in the document, as well as the fact that an annotation can be used by those who created it or by others.

In summary, an annotation is a piece of information created by an individual that is associated with the whole or part of an information artifact[2] and may be directed at the author, another individual, or a group to the end of communication, clarification, summarization or some similar function. This will be the definition of annotation that will be used for the rest of this paper.

1.5.2 Anchor and Anchor text

In context of annotation systems, anchor can be referred as position, area or time range to which an annotation on an artifact is directed. Its form depends on the form of artifact. In a document, anchor text is the series of characters associated with the anchor.

1.5.3 Link

In hypertext system, a link is a reference from one document to another document. A link in an annotation system is a reference between an annotation and an anchor.

---

[2] According to Webster's Universal Encyclopedic Dictionary, an artifact means *something created by humans usually for a practical purpose*. An information artifact, in this paper, means something created by humans which is intended to store or communicate information. Common forms include documents, pictures, drawings, image, voice, animation, video, etc. Some of them, e.g. document, picture, and image may exist in physical formats. However, most of them are currently available in digital format.

1.5.4 Metadata

The most common definition of Metadata is data about data. In WWW, metadata is structured data used to describe Web resources. In this particular paper, an annotation, a description attached to an artifact, can be viewed as a form/type of metadata.

1.5.5 Collaboration

Collaboration is a process in which two or more people work together to accomplish a common task.

1.5.6 Folksonomy

A folksonomy is a user-generated taxonomy used to categorize and retrieve resources on the WWW using open-ended labels called tags. Flickr and Delicious are examples of folksonomy systems.

1.5.7 Social Tagging/Social Annotation

Social tagging/Social Annotation is collaborative metadata generated by a group of users, also known as collaborative tagging

1.5.8 Synonymy and Polysemy

Synonymy occurs when there are different words with similar or identical meanings (e.g. car and automobile). Polysemy occurs when there is a word or phrase with multiple, related meanings.

1.5.9 Noise

In social annotation, noise refers as a tag or an annotation that irrelevant to the content of the document, or misspelled.

1.5.10 Similarity Ranking and Static Ranking

In web searching, similarity ranking, also known as query dependent ranking, is a similarity measurement between query and resources. A static ranking, also known as query independent ranking, is a measurement of quality of resources separated from the specifics of a particular query.

1.5.11 Resource and URL

In this dissertation, resource is a web document and URL (Unique Resource Locator) is an address of a resource.

2.0 RELATED WORK

This chapter reviews research on annotation generally and on using annotation to improve web searching.

2.1 BACKGROUND

As written language evolved, so did the ability to comment on that written communication. It is likely that the history of annotation parallels the history of written communication. Yet research on annotation is primarily a modern phenomenon.

The acquisition of Galileo's work on sunspots by the British Library in 1998 alludes to the development of research on annotation (Jackson, 2001). The annotations on Galileo's manuscript are commentary in Italian dating from the early seventeenth century. They are in the margins throughout the book. These annotations are associated with Galileo's observations and findings. However, no one knows whose annotations they are.

Jackson (2001) noted that the tradition of annotation began before the seventeen century. Philosophers in the past wrote manuscripts to express their knowledge. These manuscripts left room between lines and in the margins for editor's or reader's notes. Rosenthal (1997) gathered 242 early annotated books. Some of them were published during the sixteenth century. One of the sixteenth-century books contains annotations of several kinds including emendations,

explanations of unfamiliar terms, paraphrases of the author's meaning, citation of literary sources, historical notes, and references to other commentaries.

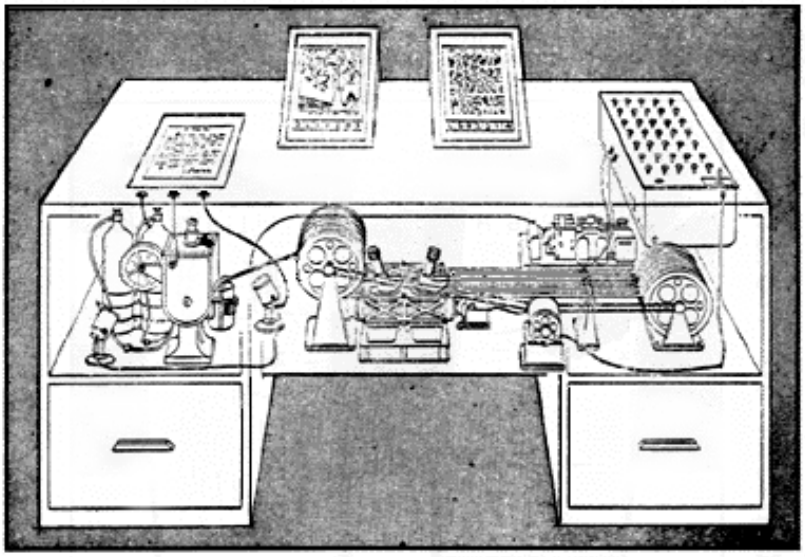
Pryde (1882) proposed a systematic marking of text in his book, "*Highways of Literature; or What to Read and How to Read*". He advocates the use of annotation as a means of forming judgment. Adler (1940) wrote a well-known article, "How to Mark a Book" to suggest why it is so important and how to annotate a book. He addresses annotations as a conversation between reader and the author. Thus, annotating a book is literally an expression of reader's differences, or agreements of opinion, with the author. He suggests several forms of annotation and their functions as follows;

> *"There are all kinds of devices for marking a book intelligently and fruitfully. Here's the way I do it:*
> *1. Underlining: of major points, of important or forceful statements.*
> *2. Vertical lines at the margin: to emphasize a statement already underlined.*
> *3. Star, asterisk, or other doo-dad at the margin: to be used sparingly, to emphasize the ten or twenty most important statements in the book. (You may want to fold the bottom corner of each page on which you use such marks. It won't hurt the sturdy paper on which most modern books are printed, and you will be able to take the book off the shelf at any time and, by opening it at the folded-corner page, refresh your recollection of the book.)*
> *4. Numbers in the margin: to indicate the sequence of points the author makes in developing a single argument.*
> *5. Numbers of other pages in the margin: to indicate where else in the book the author made points relevant to the point marked; to tie up the ideas in a book, which, though they may be separated by many pages, belong together.*
> *6. Circling of key words or phrases.*
> *7. Writing in the margin, or at the top or bottom of the page, for the sake of: recording questions (and perhaps answers) which a passage raised in your mind; reducing a complicated discussion to a simple statement; recording the sequence of major points right through the books. I use the end-papers at the back of the book to make a personal index of the author's points in the order of their appearance."*

In *As We May Think*, Bush (1945) proposed a machine, called "*Memex*", that would store textual and graphical information and which would be able to link any piece of information to any other piece. Using the Memex, users could add marginal notes and comments on all contents stored in the system. The users also could create a trail of links among pieces of information.

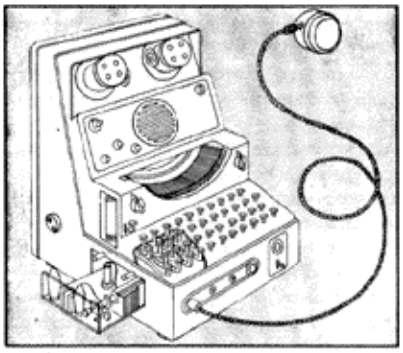
This trail can be kept and used for future reference. The following is an excerpt from the article outlining Bush's vision on *Memex*.

> "*Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and to coin one at random, ``memex'' will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.*
>
> *It consists of a desk, and while it can presumably be operated from a distance, it is primarily the piece of furniture at which he works. On the top are slanting translucent screens, on which material can be projected for convenient reading. There is a keyboard, and sets of buttons and levers. Otherwise it looks like an ordinary desk.*
>
> *In one end is the stored material. The matter of bulk is well taken care of by improved microfilm. Only a small part of the interior of the memex is devoted to storage, the rest to mechanism. Yet if the user inserted 5000 pages of material a day it would take him hundreds of years to fill the repository, so he can be profligate and enter material freely.*
>
> *Most of the memex contents are purchased on microfilm ready for insertion. Books of all sorts, pictures, current periodicals, newspapers, are thus obtained and dropped into place. Business correspondence takes the same path. And there is provision for direct entry. On the top of the memex is a transparent platen. On this are placed longhand notes, photographs, memoranda, all sort of things. When one is in place, the depression of a lever causes it to be photographed onto the next blank space in a section of the memex film, dry photography being employed.*
>
> *There is, of course, provision for consultation of the record by the usual scheme of indexing. If the user wishes to consult a certain book, he taps its code on the keyboard, and the title page of the book promptly appears before him, projected onto one of his viewing positions. Frequently-used codes are mnemonic, so that he seldom consults his code book; but when he does, a single tap of a key projects it for his use. Moreover, he has supplemental levers. On deflecting one of these levers to the right he runs through the book before him, each page in turn being projected at a speed which just allows a recognizing glance at each. If he deflects it further to the right, he steps through the book 10 pages at a time; still further at 100 pages at a time. Deflection to the left gives him the same control backwards.*
>
> *A special button transfers him immediately to the first page of the index. Any given book of his library can thus be called up and consulted with far greater facility than if it were taken from a shelf. As he has several projection positions, he can leave one item in position while he calls up another. He can add marginal notes and comments, taking advantage of one possible type of dry photography, and it could even be arranged so that he can do this by a stylus scheme, such as is now employed in the teleautograph seen in railroad waiting rooms, just as though he had the physical page before him*."



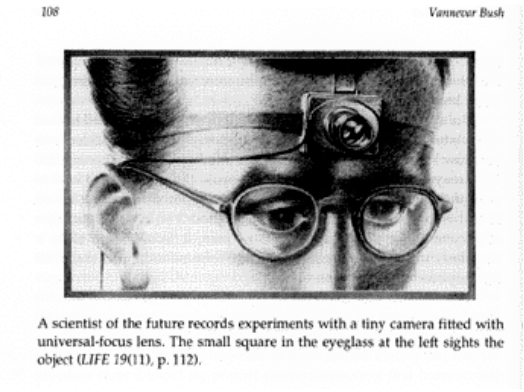
Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE* 19(11), p. 123).

(a) Vannevar Bush's MEMEX



Supersecretary of the coming age, the machine contemplated here would take dictation, type it automatically and even talk back if the author wanted to review what he had just said. It is somewhat similar to the Voder seen at the New York World's Fair. Like all machines suggested by the diagrams in this article, it is not yet in existence (*LIFE* 19(11), p. 114).

(b) Vannevar Bush's MEMEX voice input output device



108 *Vannevar Bush*

A scientist of the future records experiments with a tiny camera fitted with universal-focus lens. The small square in the eyeglass at the left sights the object (*LIFE* 19(11), p. 112).

(c) Vannevar Bush's MEMEX head camera

Figure 4. Drawing of Memex [3]

Bush's vision inspired a generation of researchers to conduct both hypertext and annotation research.

[3] From http://www.acmi.net.au/AIC/BUSH_BERRNIER.html

2.1.1 Annotation Forms and Functions

Students frequently create annotations while they are reading a textbook. Paper-based annotation includes underlined text, highlighted text, note adding and ink marks, which are added to a document or a book. Marshall (1997, 1998) conducted some of the first modern research studies on annotation, finding that students normally annotate a book by underlining or highlighting text to emphasize the important concepts and by writing down short notes to summarize the important content or to give an opinion about, or comment on, the content they have read.

O'Hara and Sellen (1997) found that annotating while reading helps the readers to understand deeply. The annotations also allow them to create a summary more easily. Wolfe concluded that the annotations improve recall of emphasized items and decrease chances of unnecessary summarizing (Wolfe, 2000).

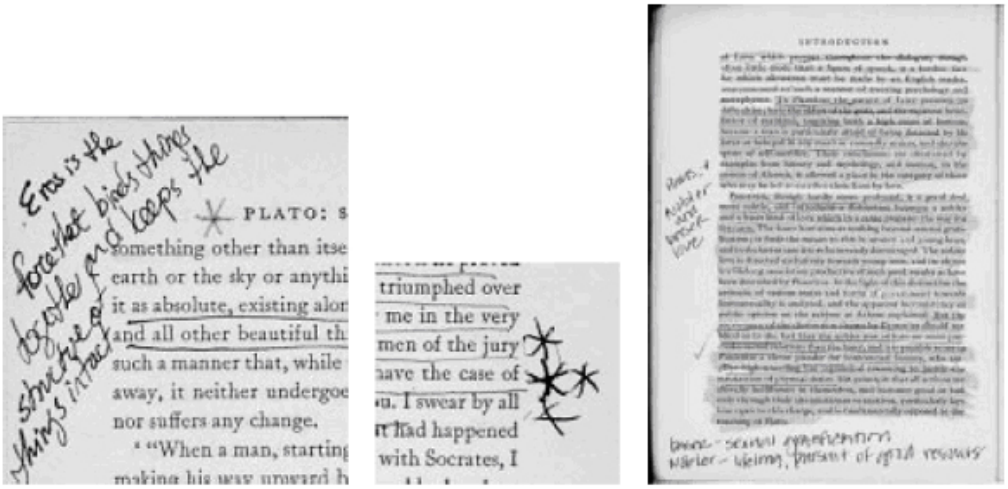


Figure 5. Examples of paper-based annotations [4]

[4] From (Marshall, 1998)

Marshall (1997) examined fifteen different sets of used books, more than 150 books. She found out that the annotations can be created within the text and on marginal or blank space. They can be "telegraphic marginal symbols" or textual. She also noted that the annotations can be created in different forms, such as underlining or highlighting, Short notes, and drawing. Table 1 and Table 2 show the characteristics and forms of annotation and its function.

Table 1. Characteristics of annotations written on the books[5]

Form	Within-text	Marginal or blank space
Telegraphic	Underlining; Highlighting; Circles and boxes around words and phrases	Brackets, Angle Brackets, and Braces; Asterisks, and Stars; Circles and boxes around whole pages; Arrows and other deictic devices to connect within-text markings to other marginal markings
Explicit	Brief notes written between lines, especially translation of words in foreign language texts	Short phrases in margin; Extended notes in margin; Extended notes on blank pages in the front of the book; Problems worked in margins

Table 2. Mapping annotation form into function[5]

Form	Function
Underlining or highlighting higher level structure (like section headings); telegraphic marginal markings like asterisks.	Procedural signaling for future attention.
Short highlighting; circled words or phrases; other within-text markings; marginal markings like asterisks.	Placemarking and aiding memory.
Appropriate notation in margins or near figures or equations.	Problem-working.
Short notes in the margins; longer notes in other textual interstices; words or phrases between lines of text.	Interpretation.
Extended highlighting or underlining.	Tracing progress through difficult narrative.
Notes, doodlings, drawings, and other such markings unrelated to the materials themselves.	Incidental reflection of the material circumstances of reading.

[5] From (Marshall, 1997)

Marshall (1998) conducted a study to examine characteristics of annotations from the community of annotators. She concluded that there are several dimensions that allow us to describe forms of annotation. She also defined the functions of annotation from reader's points of view, and the roles of annotations that are used to communicate with others. The table 3 shows her "dimensions of annotation."

Table 3. List of some dimensions of annotation and its objective[6]

Dimension of annotation	Its objective
Formal vs. Informal annotation	*Formal annotation*: for specifying metadata to ensure interoperability. *Informal annotation*: for the annotator only as a journal article.
Explicit vs. Tactic Annotation	*Explicit annotation*: understandable for everyone. *Tactic annotation*: understandable only for the annotator.
Annotation as writing vs. Annotation as reading	*Annotation as writing*: commentary, explanation of what they think *Annotation as reading*: organization, interpretation , conclusion of what they read
Permanent vs. Transient Annotation	*Permanent annotation*: bring value to future reader including the annotator. *Transient annotation*: reflect the reader's engagement for a current period of time.
Published vs. Private Annotation	*Published annotation*: publicly available for everyone. *Private annotation*: only available for the annotation or the group of authorized reader.

2.1.2 Purposes of Annotation

2.1.2.1 Annotation for Memory

When people annotate a document, their primary purpose is to communicate with themselves or other people. Marshall found that college students check out the dirtiest copy of the book rather than the cleanest one (Marshall, 1998). The dirtiest copy of a book contains underlined or

[6] From (Marshall, 1998)

highlighted content, notes summarizing the main idea and marks emphasizing important paragraphs. Wolfe (2000) found that annotations help students recall emphasized items and decreased unnecessary summarizing. O'Hara and Sellen (1997) suggested that annotating while reading allowed readers to understand the text deeply and extract structure of the document creating a plan for writing. This kind of annotation helps locate interesting sections of documents and improves recall of the important concepts in a document.

2.1.2.2 Annotation for Communication

While many people annotate to remember, others use annotations to communicate. Davis and Huttenlocher (1995) observed that shared annotations, especially in the educational context, provide benefits to both students and instructors. The shared annotations provide opportunities for communication outside of the classroom or computer laboratory. Both students and instructors can communicate via the shared annotations. From the instructors' point of view, because the annotations are available to everyone in the class, instructors can correct misunderstandings about the content of the class or the assignment and publicly announce them.

Davis and Huttenlocher (1995), Brush *et al.* (2002) and Kurhila *et al*. (2003) illustrated how shared annotation provides benefits in an educational environment. Annotations can be kept personal, or can be shared and can serve as anchors for threaded discussions. Discussions around class materials outside the class room are recorded for others to see and they are directly and precisely linked to the class material.

Frazan and Brusilovsky (2005) make use of annotation to develop personalized navigation support techniques based on user feedback in an online learning system. They call these techniques "*annotation-based social navigation support (Annotation-based SNS)*". The annotations are implicitly used as indicators of page relevance for the current group of learners.

Their techniques are implemented in the Knowledge Sea II (KSII). They found that using the annotation approach provides favorable results for navigation support in the online learning system.

2.1.2.3 Annotation for Collaboration

Annotations play an important role in collaboration. When people work together, communications between them are very important. Neuwitrth *et al.* (1997) studied annotations as a factor in collaboratively authoring. In collaborative writing, writer, consultant, editor, and reviewer use the annotation as a tool to provide feedback (Baecker *et al.*,1993). Willms *et al*. (1997) and Spring *et al*. (1999) confirmed that annotations are helpful for collaborative authoring and document editing.

Other than collaborative writing, the annotation can be used for discussion purposes. Conklin and Begeman (1988) concluded that annotations are helpful for collaborative discussion among users. Mashayekhi and his colleagues (1994) identified several advantages of annotation for collaboration. They found that annotations are useful for software engineers when they do collaborative inspection of software (Mashayekhi *et al.*, 1994). Sapsomboon (2000) studied the effects of annotations for shared defect detection in asynchronous software inspection. He concluded that annotations help the software inspector in both the defect detection task and the defect review task.

Perry (2002) described scenarios where annotation helps in scientific collaboration in real time environment. Scientists in different locations can collaboratively work and discuss the working document in the real time fashion. One scientist annotates the document while others see the annotation immediately.

Chiueh and Katz (1991) looked at annotations in design tasks. Designers annotate their design objects to provide more information or give feedback to other people in their team.

2.1.2.4 Annotation for Description

Annotations can also be used for describing an object. Many researchers use annotations for classification. Shneiderman and Kang (2000) used annotations to label human photos for easier retrieval. This kind of annotation adds more information describing an object, an image in this case. In addition, Wilhelm and his colleagues (2004) developed a framework, called Mobile Media Metadata, which enables image annotation at the point of capture using Nokia 3650 camera phone.

Handschuh and Staab (2002) proposed a framework, called CREAM (CREAting Metadata). They used the annotation as marked content to describe objects on a web page (Handschuh and Staab, 2002). Hua *et al.* (2005) introduced a system that can automatically extract semantic knowledge for the annotation of an image.

Others have made use of descriptive annotations to improve retrieval results and document clustering. Ginsburgh (1998) views annotations as extra descriptions of the documents that can be used as clues to guide the user during a search session. Denoue and Vignollet (2000) found that annotations can improve automatic clustering of web pages and, at the same time, improve information access and retrieval.

Linguistic researchers use annotations to dissect parts of a sentence. Bird and Liberman (1999) suggest that linguistic annotation covers any descriptive or analytic notation applied to raw language data. The annotation may include transcriptions of phonetics, phonology, morphology, syntax, semantics, pragmatics, and discourse structure.

2.2 DIGITAL ANNOTATION SYSTEMS

Annotation of digital documents is inspired by paper-based annotation. Digital annotation is similar to paper-based annotation but it provides additional benefits that paper-based annotation cannot provide.

Digital annotation is not physically limited to available space, can be added concurrently by multiple individuals, and is separable from the target (Rodriguez, 2001, Sannomiya *et al*., 2001, Ramachandran and Kashi, 2003). Paper-based annotation is limited by the space available on the page in a document and is permanently attached with a document, while digital annotation can be treated as a separate document that is linked to the annotated artifact and can be kept separately from the original document. Multiple individuals can annotate the same piece of content in the document without interfering with other users' annotations.

Paper-based annotation is difficult to remove from the document, while digital annotation can be easily removed. Digital annotation is combinable and it can be recursive. Users can be allowed to annotate other users' annotations (Marshall and Brush, 2002). Finally, digital annotation can easily be mined (Wang *et al*., 2006).

As mentioned in the previous section, Bush's vision inspired a generation of researchers to develop both hypertext and annotation systems. There have been a number of academic prototypes and a number of commercial products that provide annotation functionality on artifacts ranging from text-only documents to three dimension vector data.

2.2.1 Systems that Used Annotations for Collaborative Purposes

2.2.1.1 NoteCard

NoteCards is an early hypertext system developed by Randall Trigg, Frank Halasz and Thomas Moran at Xerox PARC in 1984. It is a tool to organize, manage, and display a collection of textual and graphical information. NoteCards provide users with electronic note cards interconnected by typed links in a "semantic network". Each note card containe text, graphics, images or some other editable substance. NodeCards include a filing mechanism for building hierarchical structures using the note cards and the links.

NoteCards is designed for collaboration. However, it does not allow multiple users to simultaneously access a notefile, a project file in NoteCards. Users share the notefile by taking turns.

The comments on each other's work seem to be the most common activity in notecards (Trigg *et al.*, 1986). The comments are kept in *Annotation cards* and linked to the note cards being annotated with the link icon. The annotation cards have a title describing their substance. Link types show the relationship between the annotation card and the card annotated. Annotative link types include Comment, Response and Argument.

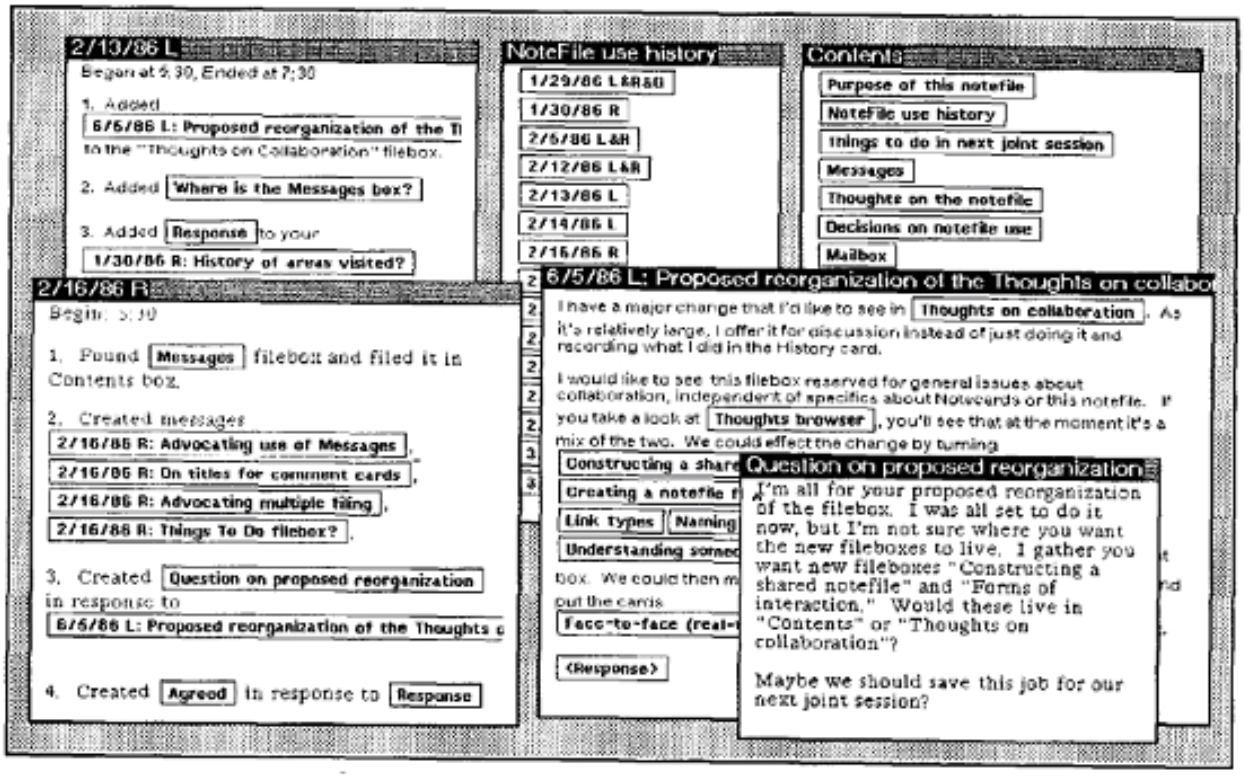


Figure 6. Annotation card "Question on proposed reorganization" in NoteCards [7]

Recording of each entry into the notefile and the changes made during each session were kept in *History Cards*. This provides an historical record of activity in the notefile, allowing users to see what others had done during the session. Direct questions or suggestions can be sent to other users via *Message cards*. Unlike annotation cards, message cards can be read only by the recipient of that card. NoteCards was used extensively at Xerox PARC in idea processing tasks ranging from writing research papers to designing parts of hardware (Halsz *et al.*, 1987).

[7] From (Trigg *et al.*, 1986).

2.2.1.2 InterNote

InterNote is another early hypertext system with annotation facilities. It is developed by Timothy Catlin, Paulette Bush and Nicole Yankelovich from Institute for Research in Information and Scholarship, Brown University. The purpose of this system is to extend their hypermedia framework to support annotative collaboration. It allows users to create annotations as commentary or suggest specific editing changes to the working document in collaborative authoring tasks (Catlin *et al.*, 1989).

It also provides "Incorporate Annotation" to allow users to revise the document with the suggested changes from the annotation. This allows users to keep track of which annotations had been incorporated. It helps alleviate conflicts if there are many users working on the same document and suggesting conflicting revision.

One thing that makes InterNote different from other hypermedia annotation systems is "*Warm Linking*". It allows a user to transfer data across the links between two documents using either "Push" or "Pull" command. The Push command copies the content of the link anchor associated with the selected link marker and pastes it at the other end of the link. The Pull command copies the contents of the remote link anchor and replaces the contents of the link anchor associated with the selected marker. With this capability, one coauthor can create an annotation and copy some portion of the document and rewrite them, while another can easily import it back to the original document.

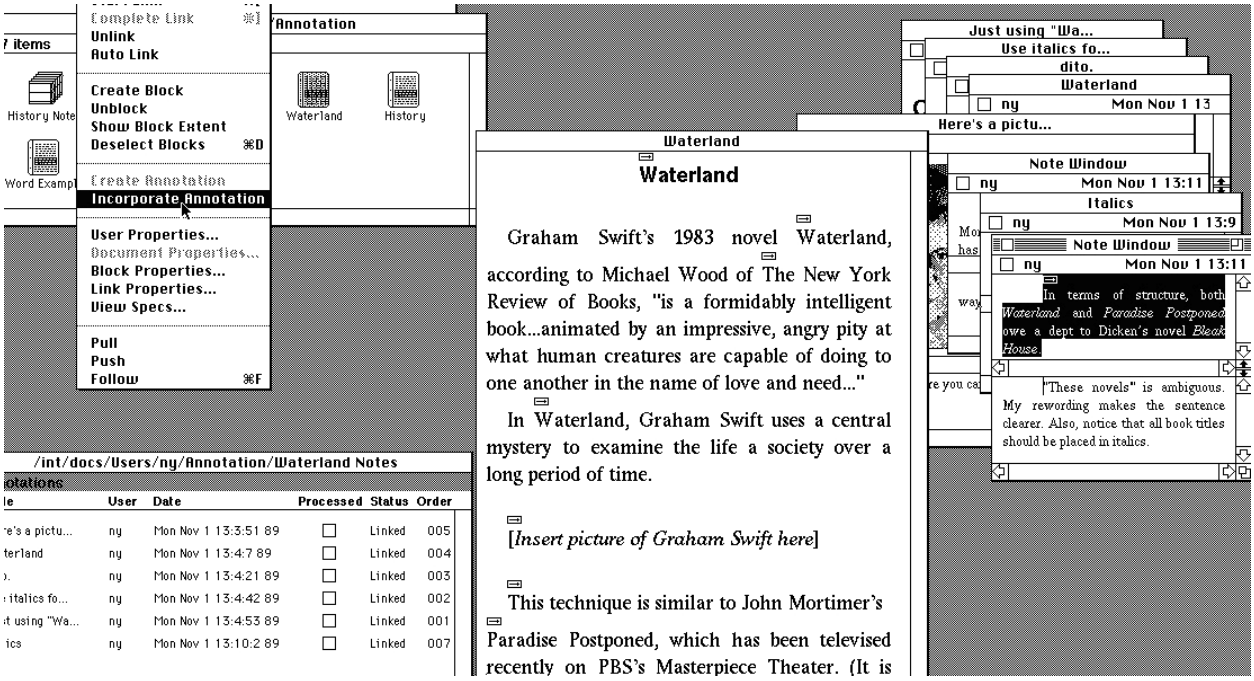


Figure 7. InterNote screen shot [8]

To make a comment or a suggestion on the working draft, a coauthor selects one or more text, graphics or timeline objects and issues a "Create Annotation" command. A link is established between the coauthor's selection in the draft and an insertion point in the Note's Frame as illustrated in figure 8. The marker will be placed to indicate the link between the draft and the note. Then, the coauthor's selection in the draft is pulled across the link into the link anchor in the Incorporation Frame as shown in figure 8. Coauthor can create textual commentary in the Commentary Frame in the lower portion of the Note window.

[8] From http://www.scholars.nus.edu.sg/cpace/ht/HTatBrown/InterNote_378.html

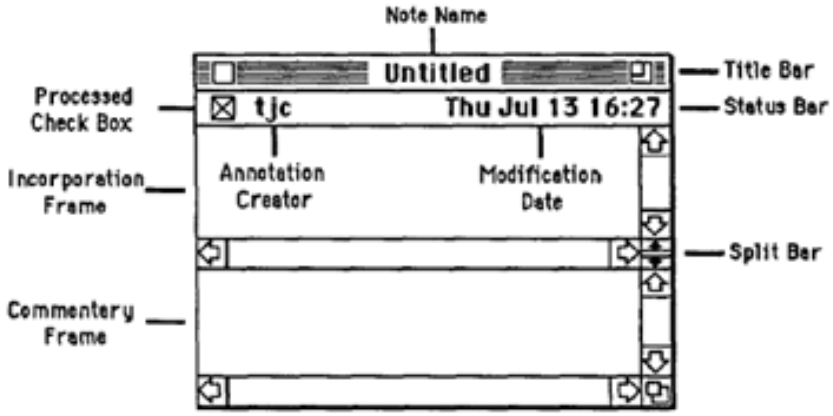


Figure 8. A note window [9]

InterNote was evaluated informally by a group of users. Catlin *et al.* (1989) concluded that creating annotations in InterNote was simple and effective. Users can easily make suggestions for editorial changes in the document. However, it is difficult to suggest structural changes. They recognize that a different style of annotation interface might be more appropriate. Drawing lines, arrows and circles directly onto the document was suggested by users. Also, facilities for synchronously communicating with other users currently working on the document was recommended.

2.2.1.3 CASCADE

CASCADE (Computer Augmented Support for Collaborative Authoring and Document Editing) is a collaborative document production system developed at the University of Pittsburgh (Spring et al., 1997). The system focuses on providing a working environment among co-authors to collaboratively produce a document. Five main functions of CASCASE are document editing, document reviewing, annotation and commenting, document balloting/approving, and

[9] From (Catlin *et al.*, 1989)

user/authority management. CASCADE allows user to specify personal preferences for everything from colors and fonts to editors and display software

For each collaborative document production task, project administrator can assign authority to individual user or a group of users with an access level and can specify the audit level and revision control level. CASCADE provides a document version control to keep track of the changes made by users. CASCADE also allows project administrator to define annotation classes and types and make use of color to differentiate them. In one project, the maximum number of different classes is four and there can be ten types in each class. All activities performing by CASCADE user are logged. The report on document activity, on user activity and on comments is created automatically and is dynamically linked to the appropriate artifacts. In addition, with query builder provided by CASCADE, project activities can be search by action, by user, by group of user and by document. Thus all user actions can be tracked.

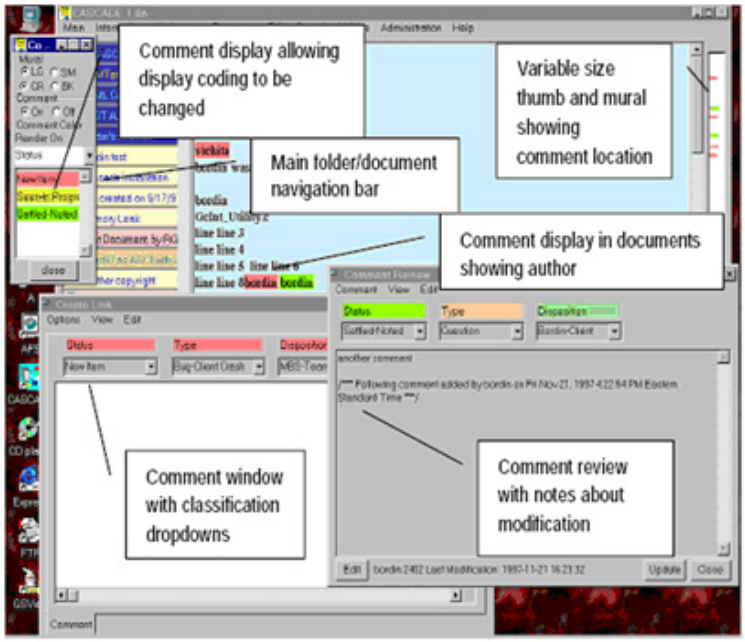


Figure 9. CASCADE main screen with comment dialogs [10]

[10] From http://www.sis.pitt.edu/~cascade/papers/design.pdf

Another interesting feature of CASCADE is that it provides a visualization of the document and annotation location on the right of the main screen. This helps user easier locate where annotations are on a document.

2.2.1.4 Collaborative Clinical Trial Protocol Writing System

A Collaborative Clinical Trial Protocol Writing System is a collaborative document production system developed by Weng and colleagues at the University of Washington. They define four user roles: 1) *Author*, who is authorized to edit a document, 2) *Reviewer*, who adds comments, 3) *Editor*, who creates and owns a document, 4) *Manager*, who monitors the schedule and progress of the document development (Weng *et al.*, 2004).

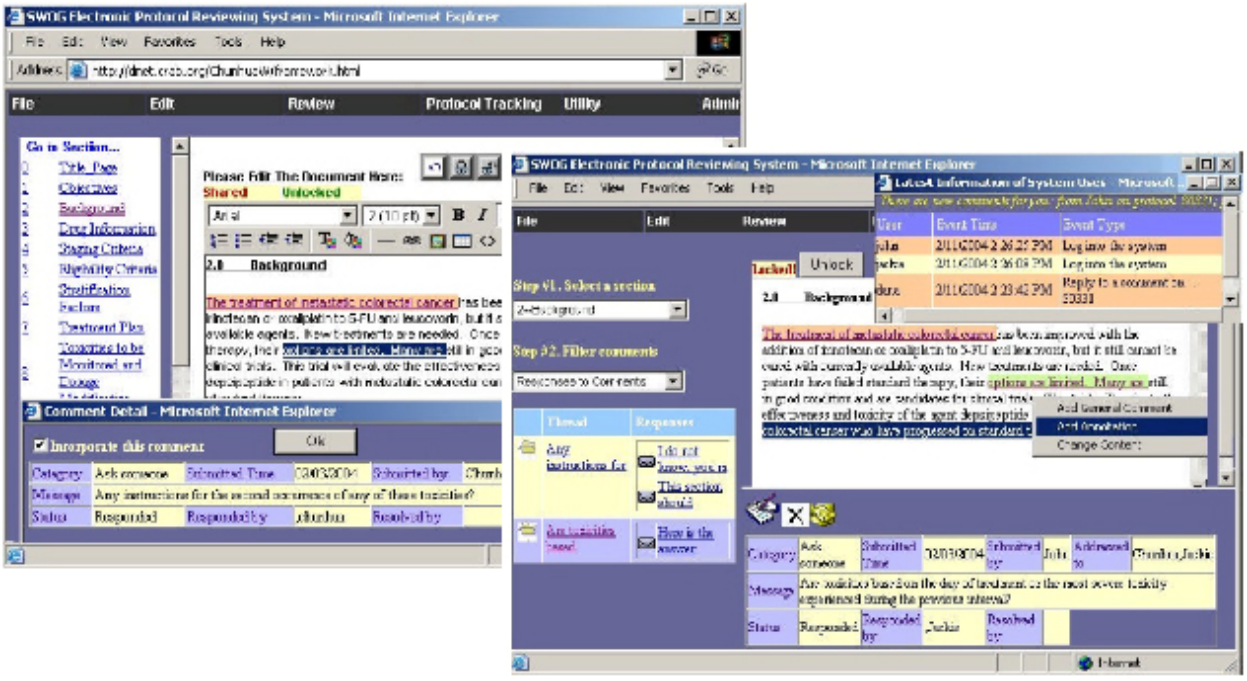


Figure 10. Collaborative clinical trial protocol writing system screen shot [11]

[11] From (Weng *et al.*, 2004)

The system includes a comment module to facilitate the communication among users. With this module, reviewers can specify a group of document authors/editors who they want to read or response to their comments. The authors/editors will be notified by email or message box when they log into the system that a new comment on their working document is posted. Editors can change status of comments when they revise the document. Therefore, reviewers will know whether or not the comments have been incorporated into new versions of the document. Weng *et al*. (2004) conduct an informal user study of the system. Users give positive feedback and show enthusiastic acceptance of the system.

2.2.1.5 Col•laboració

Col•laboració is a web-based collaborative writing tool developed by Rodríguez from the Royal Institute of Technology of Sweden. The main objective of the system is to provide a shared space for co-authors working on a document (Rodríguez, 2001). The system is composed of two main modules; *Document Development Module* and *Author Management Module*. The document development module includes adding a section to a document, editing the document, and adding comment to a section or a document. The author management module includes creating a new document, joining as co-author to a document, and setting preferences for notification. The system has been used to support collaborative writing in eight different tasks including workshop poster creation, short paper development and masters' thesis proposal. Most of the comments that users made were about how to improve the system interface. Users reported the orphaning of annotations as a significant problem within the system.

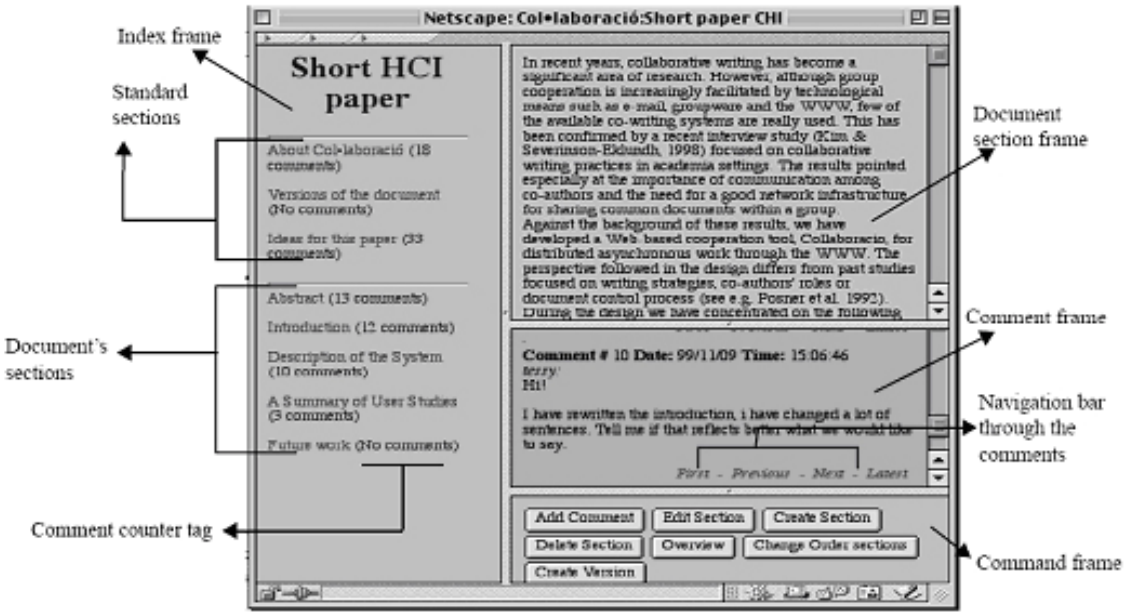


Figure 11. Document development module screen shot [12]

Figure 12. Co-author's management module screen shot[12]

[12] From (Rodríguez , 2001)

2.2.2 Systems that Used Annotations for Memory and Communication Purposes

2.2.2.1 WebAnn

WebAnn is a web annotation tool to support discussion in an educational environment. It is developed by Brush from the Department of Computer Science and Engineering, University of Washington. It is embedded in Microsoft Internet Explorer. To create an annotation, users can select the text to be annotated from a web page and choose to create an annotation from a popup menu. Users have a option to make their annotation private or public. All annotations are listed as threaded discussions on the left side in the browser as shown in the figure 13.

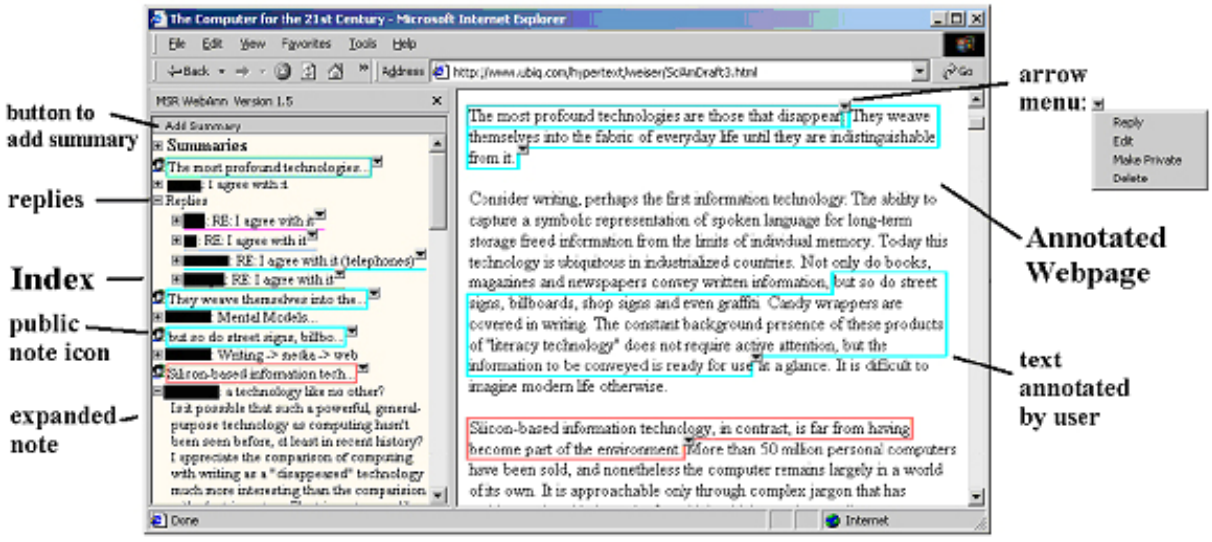


Figure 13. Screen shot of WebAnn [13]

Annotations are stored on an annotation server. The server is part of the common annotation framework (CAF) implemented by Davide Bargeron (Bargeron *et al.*, 2001). WebAnn uses URLs to retrieve all attached annotation from the server. It then highlights the

[13] From (Brush, 2002)

annotated text on that web page according to anchor information that is stored on the server when the user creates the annotation.

The interesting feature of WebAnn is "Keyword Anchoring". This algorithm associates an annotation with key words in the text and makes the anchor text somewhat resistant to minor edits (Brush *et al.*, 2001, Brush, 2002). It helps improve anchoring and locating of annotation. When there is significant editing of the anchor, WebAnn indicates that the annotation is "orphaned" and presents a best guess candidate anchor.

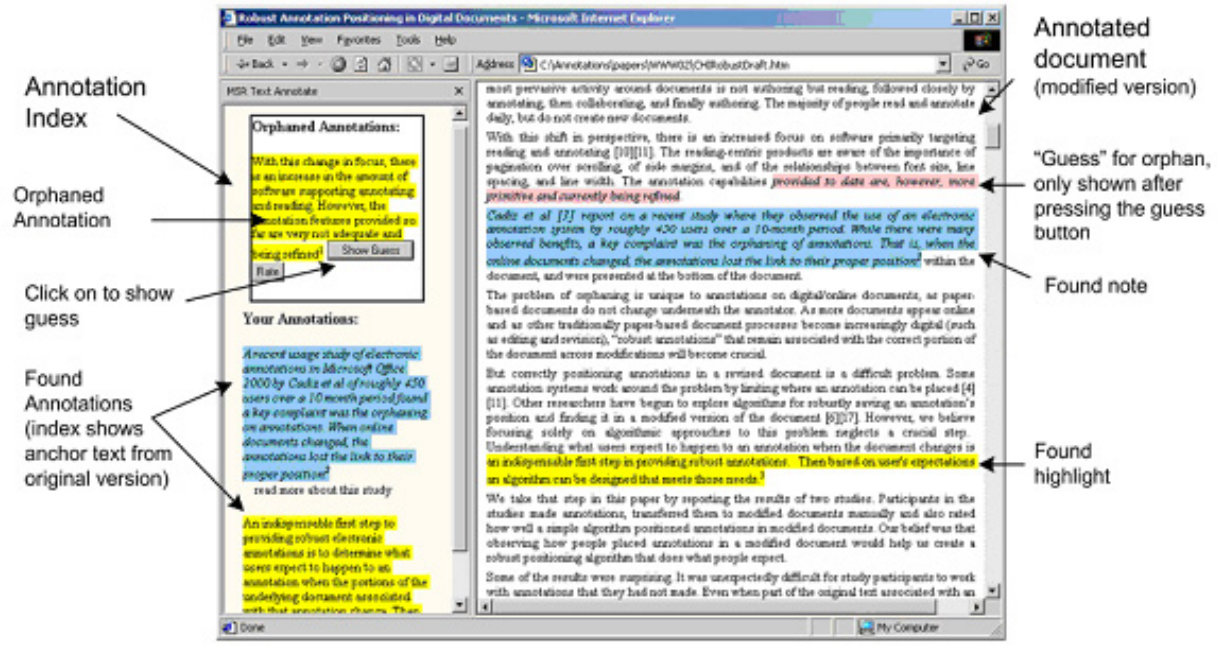


Figure 14. Screen shot of WebAnn showing orphaned annotation[14]

Brush conducted a study to assess the algorithm. She concluded that the Keyword Anchoring worked well in several situations. When the annotation's anchor text does not change and when it moves, the algorithm can perfectly locate it in the modified document assuming that

[14] From (Brush, 2002)

it is long enough to be unique. When the annotation's anchor text is modified e.g. words are deleted or the text is reworded, the algorithm can still locate it in the modified document. However, the algorithm fails when the annotation's anchor text is changed dramatically.

2.2.2.2 EDUCOSM

EDUCOSM is a collaborative learning environment developed by Miettinen and his colleagues from Helsinki Institute for Information Technology, Finland. The primary components of EDUCOSM are annotations, hierarchical newsgroup discussions and publication of students' writings. It provides the community with a shared view of the Web. When users request a web page, the EDUCOSM server reads it from the original location and inserts annotations, menus and other application specific data before sending it to the client (Miettinen *et al.*, 2003).

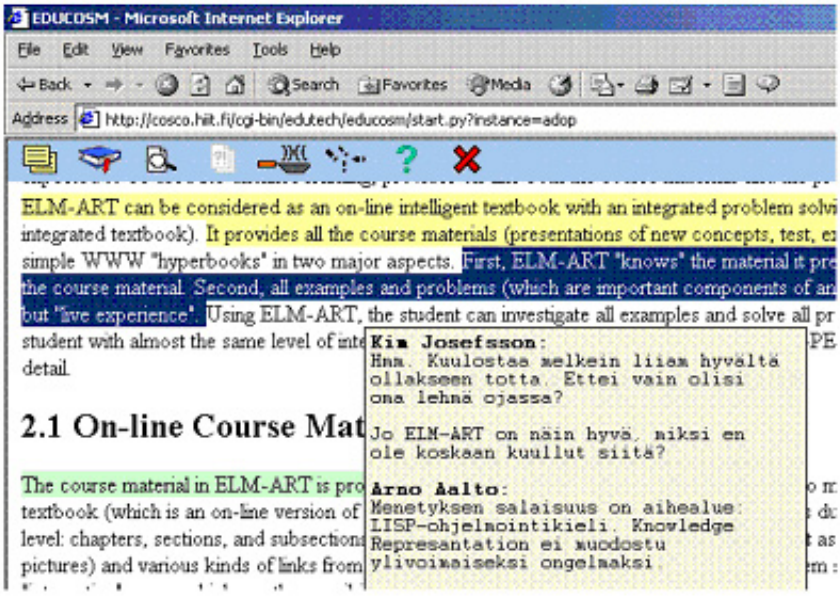


Figure 15. Example of annotations in EDUCOSM interface [15]

[15] From (Kurhila *et al.*, 2003)

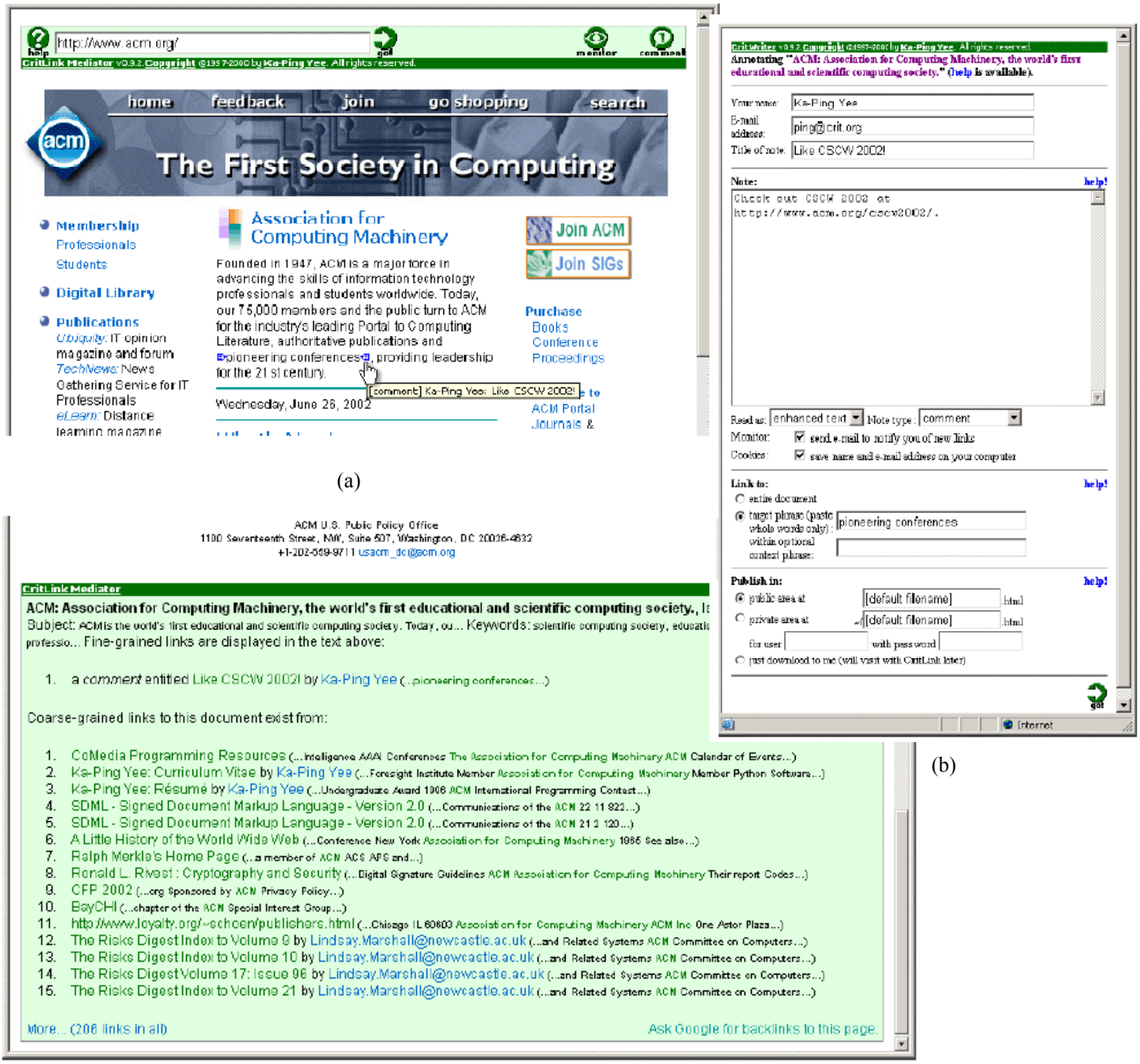
In EDUCOSM, there are two types of annotation; Highlight and Comment as shown in the above figure. Users can select parts of texts and apply highlighting to them or add a comment to them. These collaborative annotations form a hierarchical newsgroup discussion. In addition, users can publish their document for discussion and feedback. EDUCOSM has an annotation filter that allows users to see only annotations made by a given group or during a certain period. Splitting and merging of overlapping annotation are offered. When the document has been changed, the surrounding text is used to find the right position of annotations. An approximate matching algorithm based on comparing sets of n-grams is responsible for this task.

EDUCOSM (Nokelainen *et al*., 2004) was used in an advance computer science class at the Helsinki Institute for Information Technology. Nokelainen conducted an evaluation study on how EDUCOSM affects learner-centered collaborative learning and concluded that a shared document-based annotation tool provided several benefits. It helps users elaborate on what they are doing and produces higher quality learning outcomes.

2.2.2.3 CritLink

CritLink is a web document annotation system developed by Ka-Ping Yee from University of California at Berkeley. It allows users to create annotations on any public web pages or even on annotations. The annotations can be public or private. Whenever users request any web pages via CritLink, it reads the web pages from that server, adds annotation links to the page and sends it to the web browser. Users can register for notification when the web page is annotated.

CritLink allows extrinsic linking (Yee, 2002). This type of link supports better collaboration since it allows other parties to concurrently contribute links relating a document to other documents.



(a)

(b)

(c)

Figure 16. CritLink screen shot[16]

(a) A home page with annotation; (b) Annotation entry form;

(c) Extrinsic link appended to the home page

[16] From (Yee, 2002)

2.2.2.4 Annotea

Annotea is a Web-based annotation system developed by Kahand and Koivunen from W3C. The system is based on an RDF infrastructure. Annotea allows users to annotate a web document. Users have three choices for creating an annotation; annotating a whole document, annotating at any position in a document, and annotating selected text in a document. They can choose to keep their annotations private or to publish them. All annotations are stored on the Annotea Server.

In Annotea, annotation consists of the body of the annotation, which contains the textual or graphical content of the annotation, the link to the annotated document with a location within the document, an annotation author, and additional property information about annotation, for instance annotation type, and annotated date/time (Kahan and Koivunen, 2001). Annotations are typed with the predefined types being *Advice*, *Change*, *Comment*, *Example*, *Explanation*, *Question*, and *SeeAlso*. The explanation of what each type is for is illustrated in the table 4. Users also are allowed to define their own type.

Table 4. Annotation types in Annotea

Annotation type	Explanation
Advice	Advice to the reader
Change	Document or proposed a change to the source document
Comment	Describe annotations that are comments
Example	Represent example
Explanation	Explanation about the content
Question	Question about the content
SeeAlso	Reference to another resource

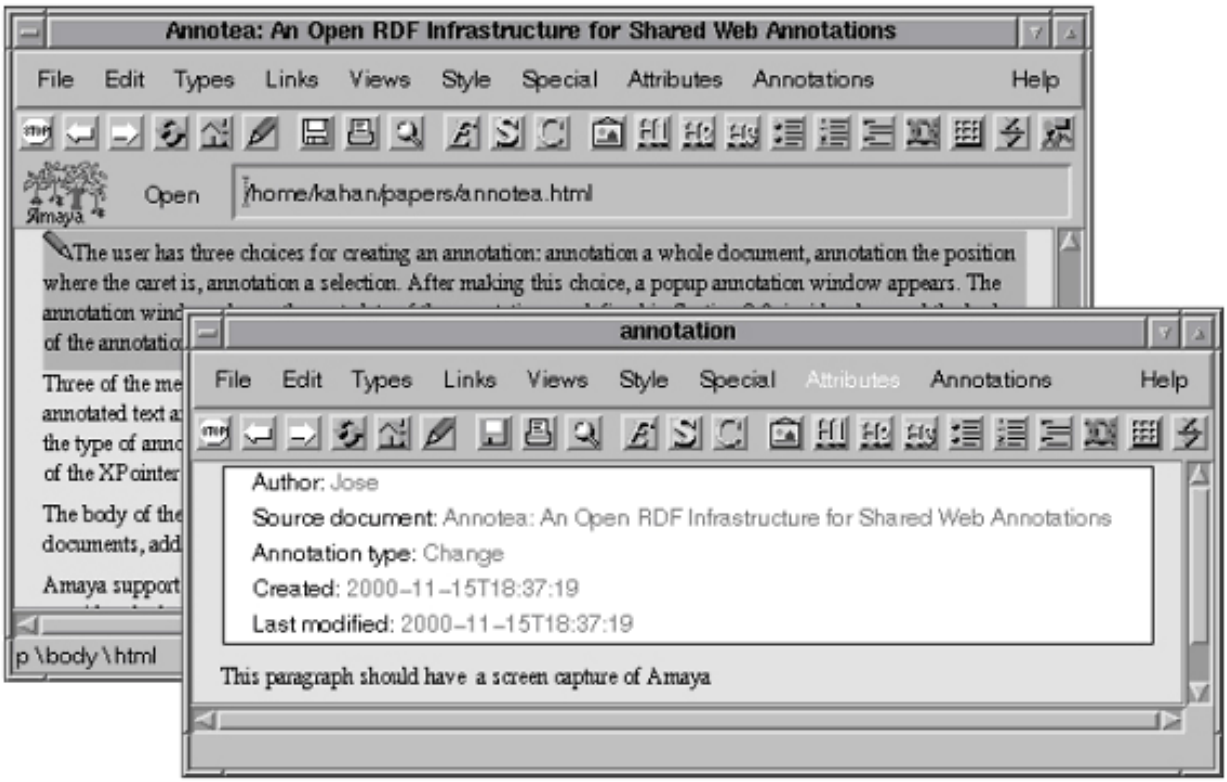


Figure 17. Annotea's paragraph annotation with Amaya Browser [17]

Users can also filter annotations by author name, by annotation type, and by annotation server. With this feature, the user can focus on the annotations of interest. Furthermore, the users can temporarily disable the annotation server when they would not like to read any annotations on that web document.

The major problem with the annotation of web documents is that some annotations may be orphaned when a web document is modified. Annotea doesn't address this problem. It shows the orphan annotations in a separate window to notify users.

[17] From (Kahan and Koivunen., 2001)

2.2.2.5 Microsoft Office

Microsoft Office 2000 is one of the first commercial products to support annotation for workgroups (Cadiz *et al.*, 2000). Microsoft Office provides an annotation feature called "Web Discussions". This feature allows team members to create annotations on any web page. Annotations will be kept on an annotation server which resides on a company's intranet. The server communicates with web browsers via WebDAV (Web Document and Versioning Protocol). After a web page is downloaded, it is checked by the annotation server for annotations. Annotations then are inserted at the appropriate places on that web page. However, when the server fails to match annotation with location, the "orphaned" annotations are displayed at the bottom of the browser.

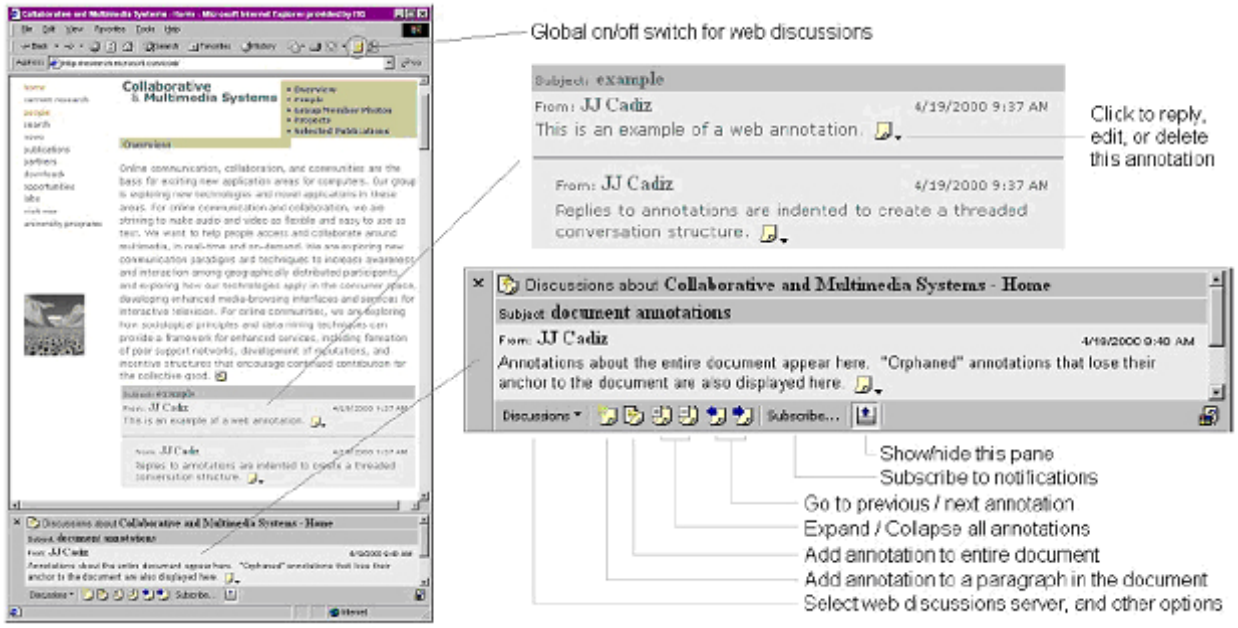


Figure 18. An annotated web page by Microsoft Office [18]

[18] From (Cadiz *et al.*, 2000)

Annotations consist of annotation author, annotation content, data and time, annotation topic and annotation anchor. When users would like to create a new annotation, the system displays all the possible places where an annotation can be made. Users, then, can select and create the annotation. Annotation authors can edit or delete the annotation. Users are allowed to expand. Collapse, or filter the set of annotations by person or time period. A notification function will send an email to a user when annotations on that webpage are created or modified. Users have an option whether to have the notification sent for every change, or to have a daily or weekly summary of changes.

2.2.2.6 Microsoft Research Annotation System

Microsoft Research Annotation System (MRAS) is a prototype of multimedia annotation web-based system for asynchronous educational environments developed by Bargerin and his colleagues from Microsoft Research Laboratory. The objective of MRAS is to provide facilities for students to record notes, questions, and comments while they are watching web-based video lectures (Bargeron *et al.*, 1999).

MRAS allows users to create annotations on any time range of the target video file. The annotations can be text or audio. They can be shared with others either via the system or email.

In addition, all shared annotations are organized as an annotation set drop-down list. It helps users access shared annotations. Users can also retrieve existing annotations from a MRAS server.

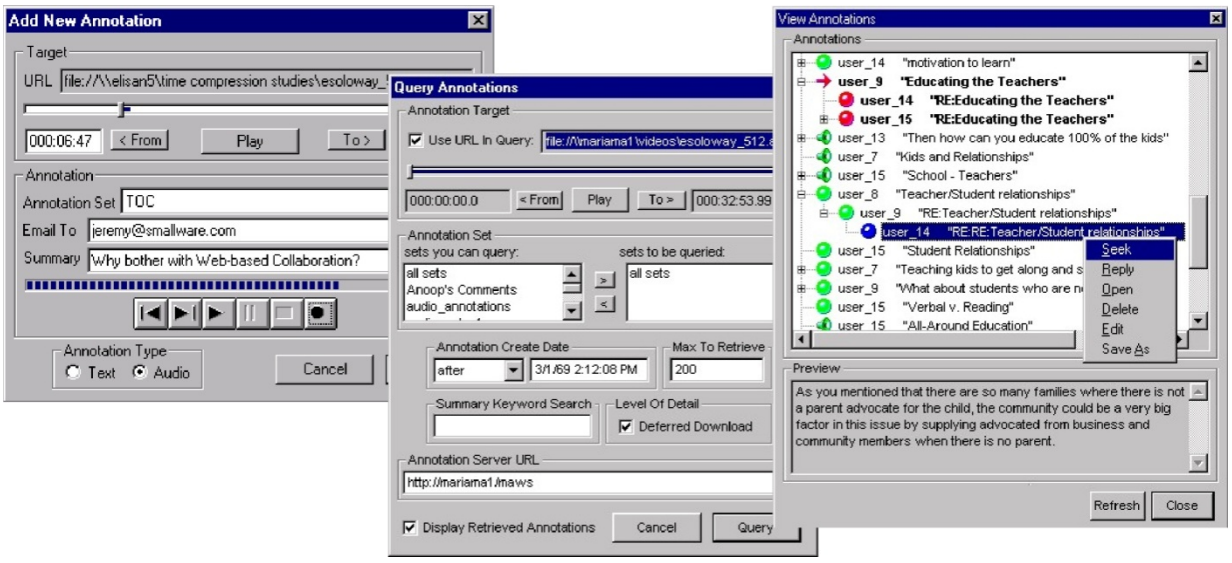


Figure 19. MRAS screen shot [19]

2.2.3 Systems that Used Annotations for Description Purposes

2.2.3.1 YAWAS

YAWAS (Yet Another Web Annotation System) is a web annotation prototype developed by Denoue and Vignollet from Université de Savoie in France (Denoue and Vignollet, 2000). It is written in Java and JavaScript. The purpose of this system is to study how annotations help improve the search function on bookmarks.

This system allows users to create personal annotations and store them locally for privacy. Users can highlight and annotate a whole web page or specific pieces of text in a web page. Each annotation contains several attributes that includes topic and type of the web document, type of the selected or highlighted text, comment, and author of annotation. Users can view their annotations attached to a given web page. The system provides a list of all annotations

[19] From (Bargeron *et al*., 1999)

in a pop-up window to help navigating to the anchor of an annotation on the web page. If a web page changes, the anchor of the annotation may be lost. The system displays annotations that have lost their anchor in gray to allow users to retrieve an annotation even if the anchor has been lost.

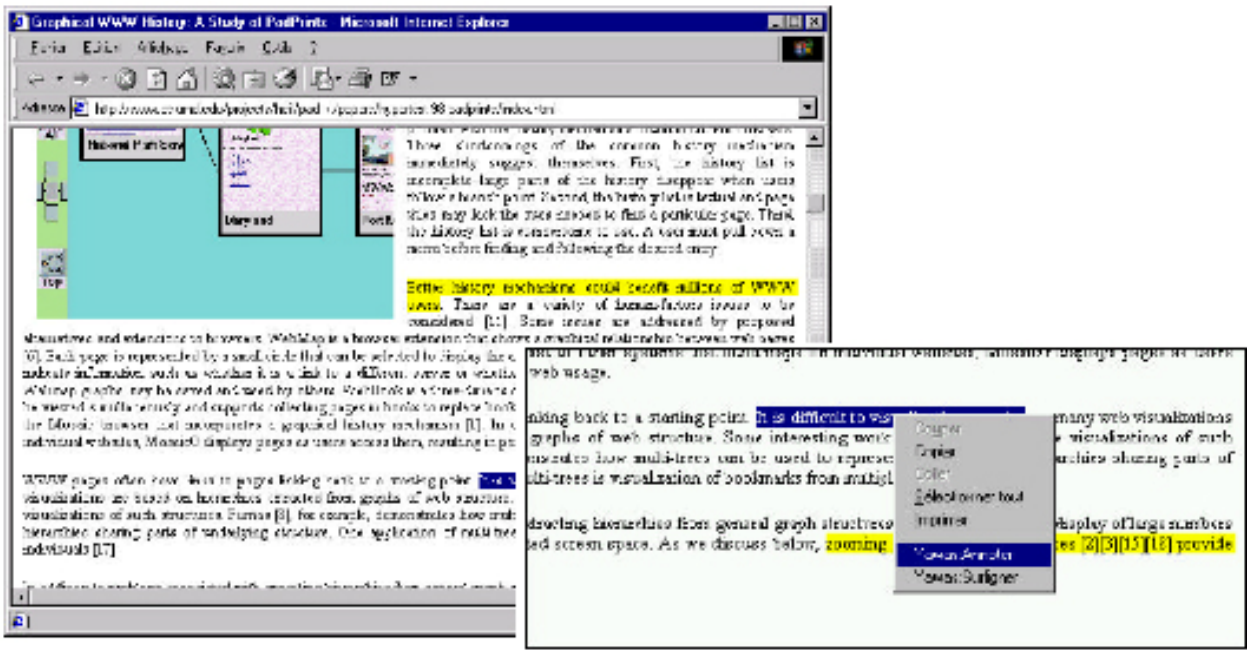


Figure 20. Yawas screen shot : Adding annotation [20]

YAWAS also provides a search function to retrieve the annotations. User can specify search criteria according to the attributes of annotation. The following figure shows the interface for the search function.

[20] From (Denoue and Vignollet, 2000)

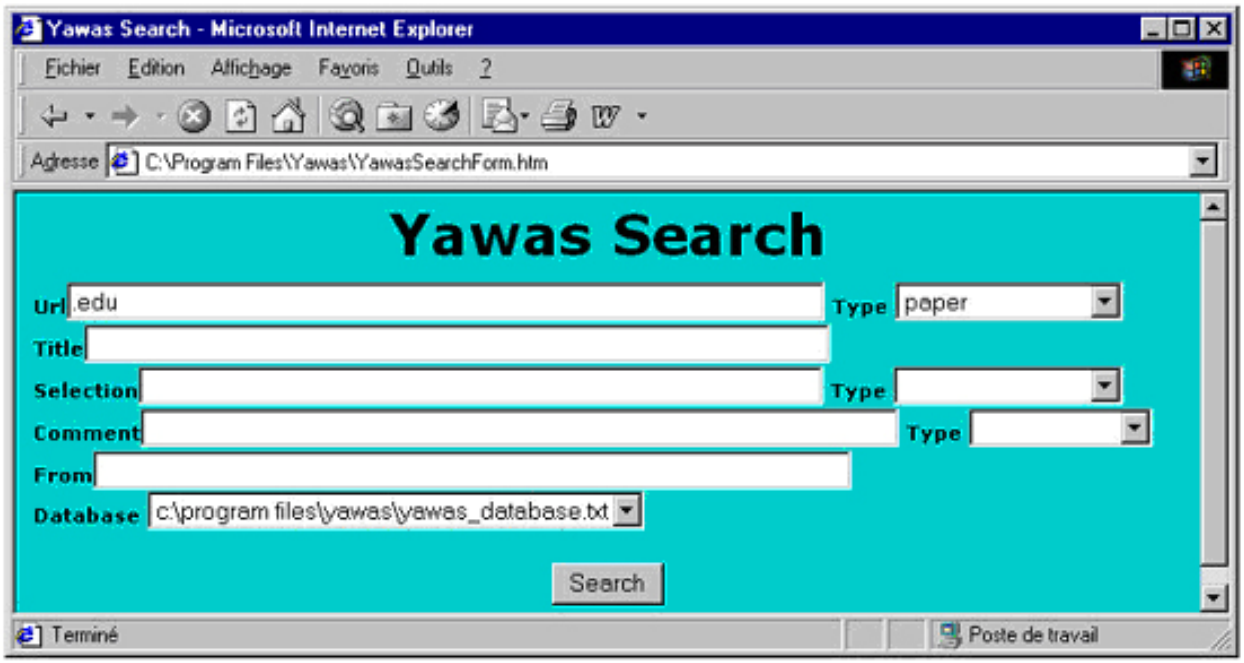


Figure 21. Yawas screen shot : Searching annotation [21]

Support for sharing annotations in Yawas is limited. Users must export all the annotation into text file and sent it via email. Users can then import the annotation file back into the system.

Denoue and Vignollet evaluated the user interface of Yawas, to determine how well annotations improve document clustering and to determine how well annotations improve information access and retrieval. In the evaluation, seven researchers and eighteen masters' students in computer science used the system. They concluded that it reduces cognitive overload when they retrieve a document previously annotated. They also reported that the navigation function was helpful.

In determining how well annotation improves document clustering, they asked a human classifier to manually cluster a set of 350 documents highlighted by others. The clustering produced by the human classifier was compared to the clustering made by the creator of the

[21] From (Denoue and Vignollet, 2000)

annotations. They conclude that annotation helps improve document clustering and suggest that highlighted text may be important when performing automatic clustering of Web pages.

In determining how well annotations improve information access and retrieval, eighteen masters' students were divided into two groups and asked to write a summary report, which required them to read at least twenty documents. They were not permitted to speak to each others directly and could not print the document and annotate them. They were allowed to use bookmarks, email and a shared word processor using NetMeeting. Only one group used YAWAS and the shared annotation server which allowed them to cooperatively annotate the documents. The annotations created by this group were extensively used as a way to summarize the documents when they wrote the report. They could easily create summaries of all annotated documents because the search engine provided by YAWAS helped concatenate all highlighted text in each document. The group who used a traditional bookmark was obviously unable to produce such summaries. Denoue and Vignollet observed that students did not need to access the full text of the document to retrieve what they were looking for. They concluded that annotation improved information access and retrieval.

2.2.3.2 Delicious

Delicious[22], pronounced as "delicious", is a web based bookmark sharing system developed by Joshua Schachter. The main objective of the system is to allow users to store, share, and discover web bookmarks. When users add bookmarks to the system, they have the option to share them. Users annotate a given URL in three categories -- descriptions, notes and tags.

[22] From http://del.icio.us

The system recommends possible tags for the URL that other users have used. These annotations will be used by the Delicious search engine. Users can bookmark html pages, audio files, video files, image files, pdf files and doc files -- any resource identified by a URL.

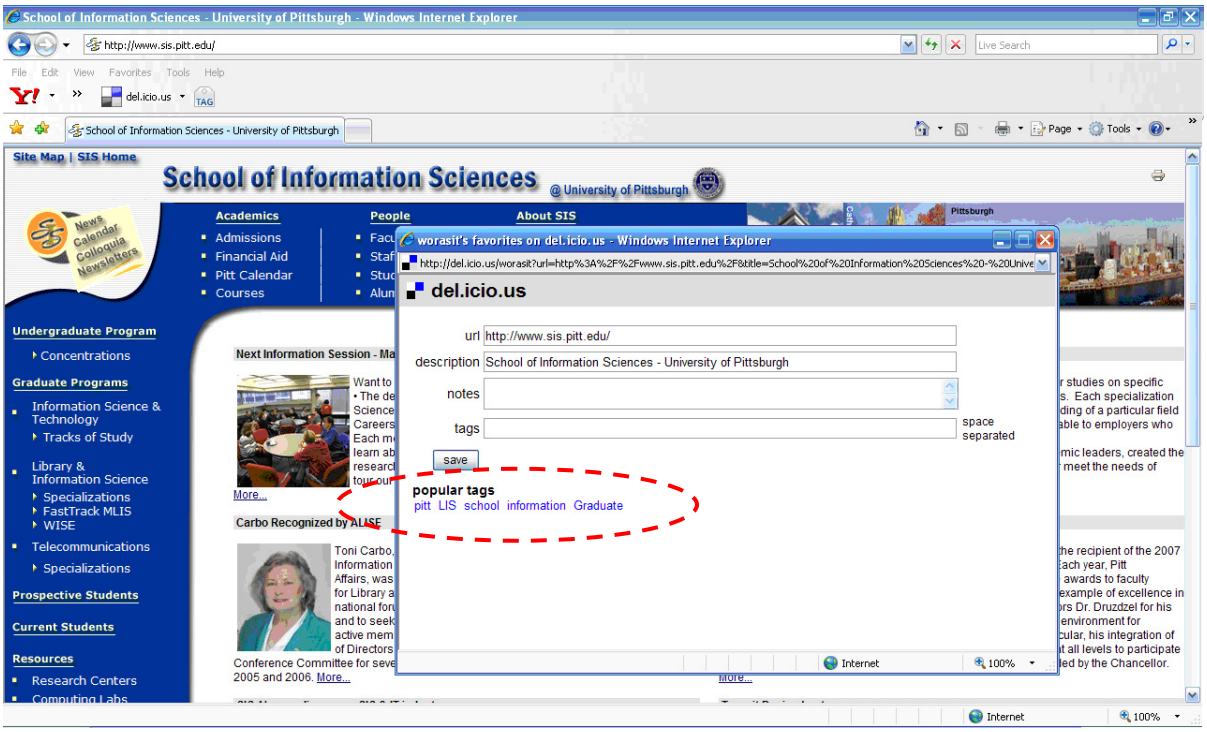


Figure 22. Screen shot of adding SIS webpage to Delicious

2.2.3.3 OntoAnnotate

OntoAnnotate (Staab *et al.*, 2001) is a tool for creating metadata by annotating web documents. It allows users to annotate HTML documents to define the semantic meaning of the objects and text passages on that page. It provides semi-automatic annotation mechanisms including *Wrapper Generation, Pattern Matching and Information Extraction*. Figure 23 provides a screen shot of the system.

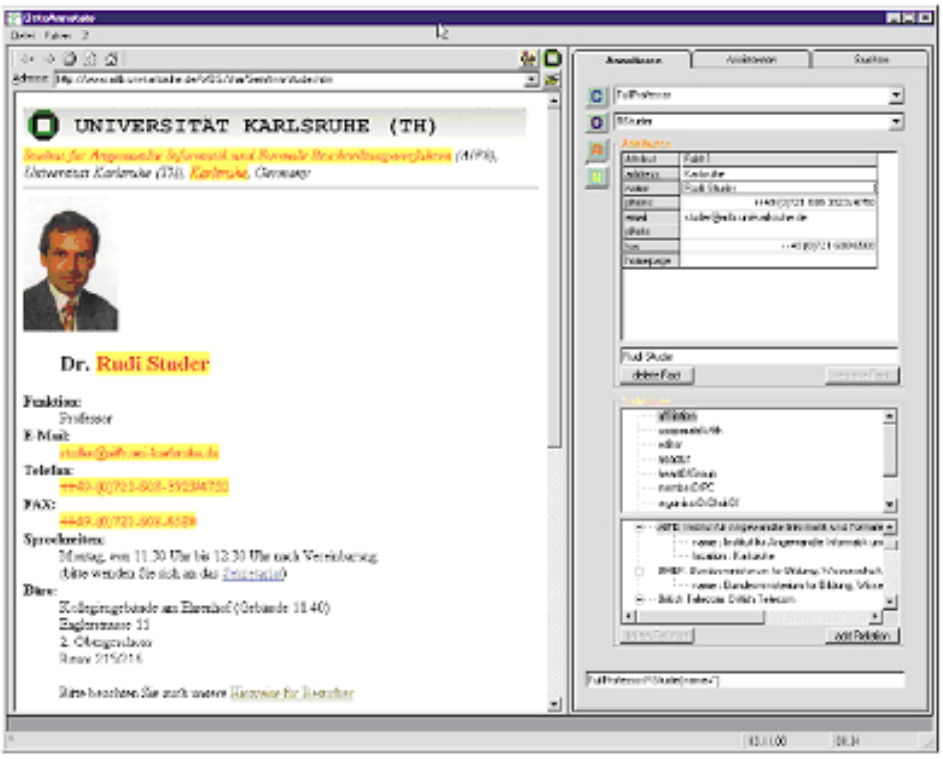


Figure 23. Screen shot of the OntoAnnotate [23]

2.2.3.4 Ont-O-Mat

Ont-O-Mat (Handschuh and Staab, 2002) is a component-based ontology-driven Web page authoring and annotation tool developed by Handschuh and Stabb from Institute of Applied Informatics and Formal Description Methods (AIFB) at University of Karlsruhe, Germany. It is based of their CREAM framework, which allows for creation of relational metadata for a web page. The following figure illustrates the architecture of CREAM.

[23] From (Staab *et al.*, 2001)

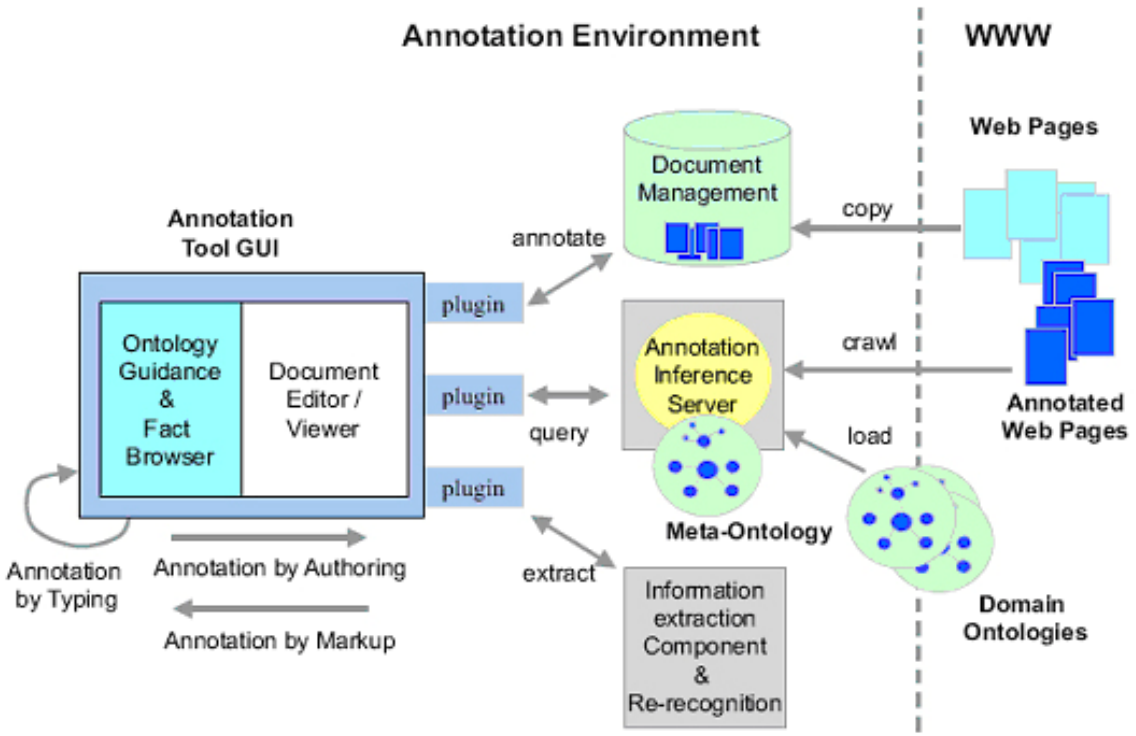


Figure 24. Architecture of CREAM [24]

The design of CREAM consists of several modules. *Document Editor/Viewer* and *Content generation* visualize the document content. *Ontology Guidance and Fact Browser* helps users create annotations that are consistent with a community's ontology. *Annotation Inference Server* reasons on the ontology and allows query for existing classes, instances and properties. *Document Management* keeps track of annotation and changes of annotation. *Metadata Re-recognition & Information Extraction* facilitates the metadata creation task. *Meta Ontology* describes how classes, attributes and relationships from the ontology should be used by the CREAM environment.

[24] From (Handschuh and Staab, 2002)

Ont-O-Mat allows users to combine authoring of a Web page and creation of relational metadata describing its content. That means users can create metadata while they are authoring their web page.

2.2.3.5 Flickr

Flickr is a web-based shared image system. The difference between Flickr and other web-based shared image systems such as shutterfly.com is that Flickr provides an image annotation feature. Users can add an annotation for a whole image or a selected part of an image. These annotations associated with images are searchable making it easier to find images associated with a particular concept.

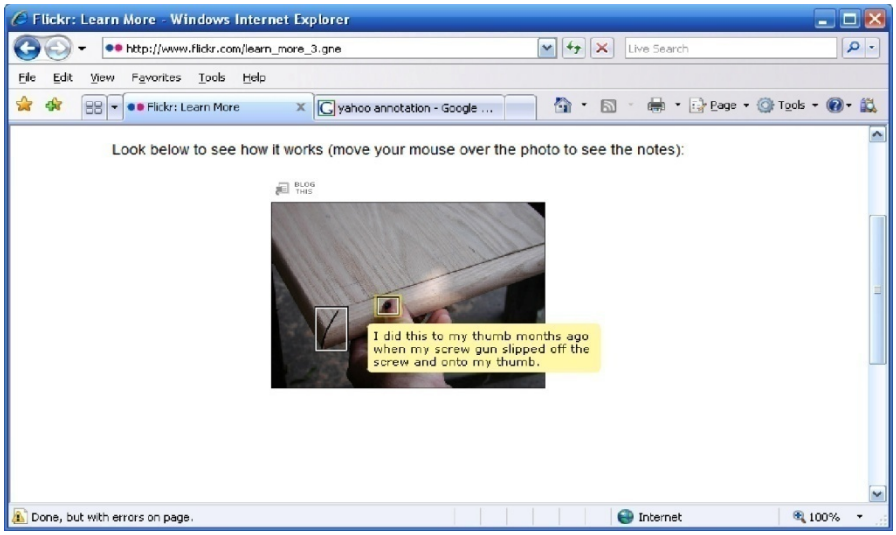


Figure 25. Flickr image annotation screen example[25]

[25] From http://www.flickr.com

2.2.3.6 PhotoFinder

PhotoFinder is an image annotation system developed by Sheiderman and Kang from University of Maryland, College Park. It is a part of their project on storage and retrieval of a personal photo library. It focuses on annotating images of people with the person's name for retrieval purposes (Sheiderman and Kang, 2000). With this system, users are allowed to add descriptions of an image and to annotate an image. Users are also allowed to construct boolean query using image description and annotation to search for images.

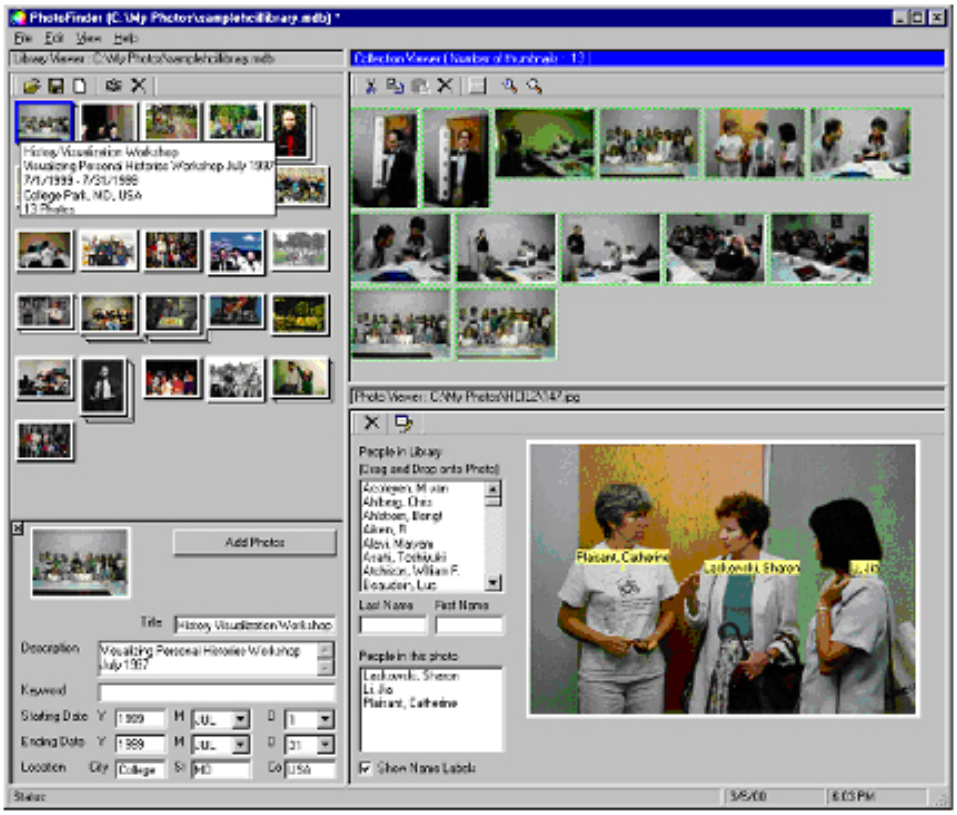


Figure 26. Photo finder screen shot [26]

They develop the concept of direct annotation. When users annotate an image, they can just select, drag, and place a label directly on the image. A label name and an X-Y location,

[26] From (Shneiderman and Kang, 2000)

based on an origin in the upper left hand corner of the photo are stored in the database. The evaluation of the system was focused on how easily the direct annotations could be made. No data was gathered on search and retrieval effectiveness.

2.2.3.7 HBP Image Annotation Service

Human Brain Project Image Annotation Service is a tool for annotating scientific images developed by Gertz and his colleagues from University of California at Davis and University of Magdeburg, Germany. The system allows scientists to annotate human brain MRI images. They can select any region of interest on the image and associate an annotation. They employed an ontology like structure to form an annotation concept structure (Gertz *et al.*, 2002). The concept-based annotations are stored in a relational database. They also developed set of procedures to determine whether two annotated regions of interest refer to the same region since regions in an image can be free drawing (circles, rectangles, or polygons).

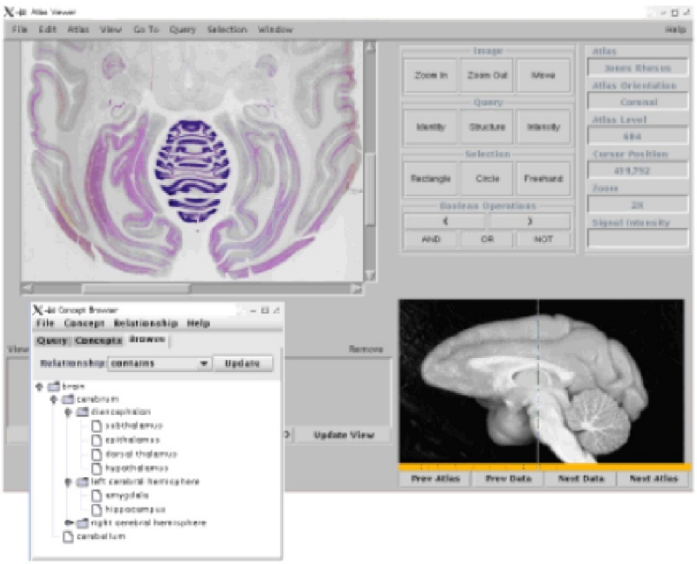


Figure 27. Screen shot of annotating human brain image [27]

[27] From (Gertz *et al.*, 2002)

The system also has an image retrieval function. Any queries users enter are compared to an annotation concept structure. Then, the search results are sent back to the users. This helps improve the retrieval function performance since the related concepts are also compared against the queries. The results from the related concepts are also sent back to the users.

2.2.3.8 MADCOW

MADCOW, a multimedia digital annotation system, is developed by Bottoni and his colleagues from Università di Roma La Sapienza in Italy (Bottoni *et al.*, 2004). It allows users to annotate multimedia documents. User can annotate whole documents or selected objects/parts of the documents. Users can create, read modify, save, search for and filter private and public annotations. Annotations in MADCOW consist of two main components *metadata* and *content*. Metadata refers to a set of attributes, for example, author, title, creation date, modification date, location (a reference to the position of the annotated object in the document), URL, and type of annotation. Content refers to content of the annotation that can contain textual information, video, image or audio files.

MADCOW is based on a client-server architecture. The server contains repositories of annotations, while the client is a plug-in for a standard web browser, which allows production of new annotations and display of existing annotations. The MADCOW annotation server provides storage for newly created annotations, updating of existing annotations, retrieval of all annotations related to a specific document, retrieval of all annotations from searching, retrieving of all URLs of all annotated documents.

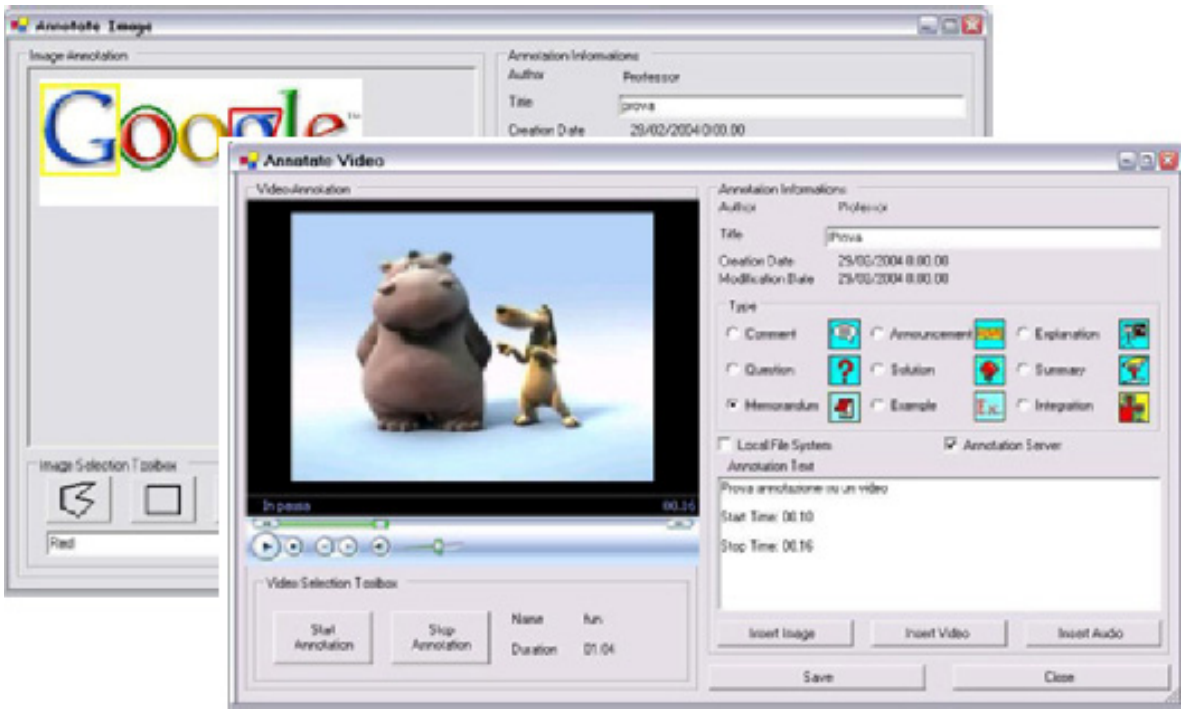


Figure 28. Annotation dialog windows of MADCOW [28]

2.2.3.9 AntV

Annotations in Video (AntV) is a video annotation tool developed by Nuno Correia and Teresa Chambel (Correia and Chambel, 1999). It allows users to add and edit annotations on video streams in specific points in space and time. Users can place an anchor of annotation on the screen while playing the video and can specify a time interval for that annotation. Unlike other video annotation systems, the annotations can be text, image or video.

[28] From (Bottoni , *et al.*, 2004)

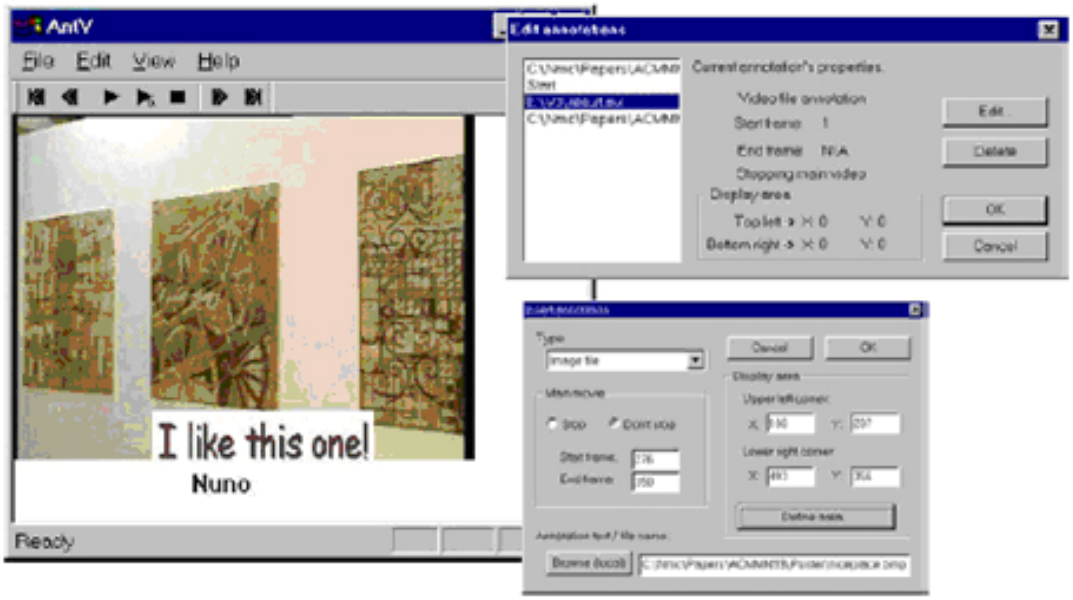


Figure 29. Screen shot of AntV [29]

2.2.3.10 VAnnotator

VAnnotator is a video annotation system developed by Miguel Costa, Nuno Correia, and Nuno Guimarães (Costa *et al.*, 2002). It is one of the three modules of the Vizard Project. The Vizard Project consists of VEplorer, a video collection search, organization and management tool; VPublisher, a new-generation storyboard, video editing, and video publishing tool; and VAnnotator, a flexible and intuitive video annotation tool.

Unlike other video annotation system, it provides a timeline model, which represent time-related multimedia content. Thus, users are allowed to annotate video content. A portion of video content can be selected and annotated. Users can specify the start time and ending time.

[29] From (Correia and Cabral , 2005)

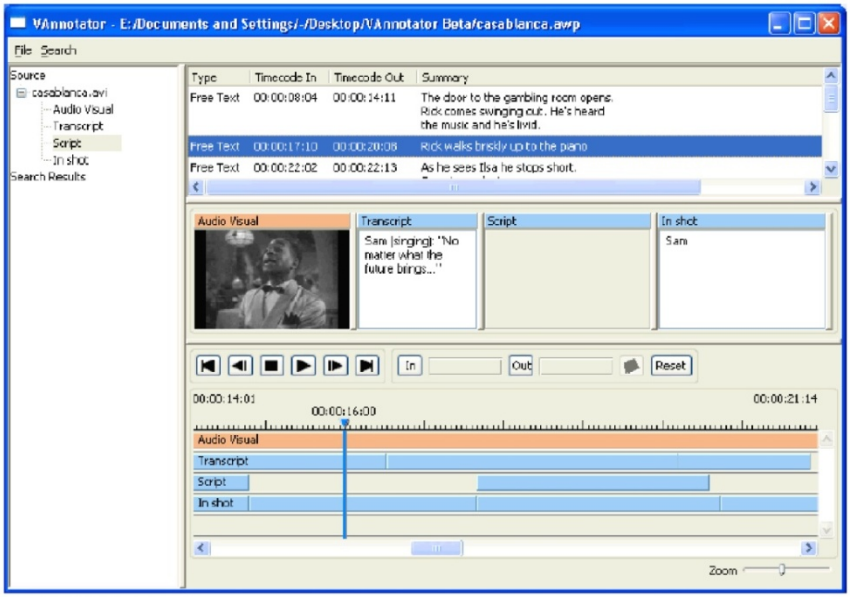


Figure 30. The VAnnotator application screen shot [30]

Annotations in VAnnotator can be viewed as a description of the video content over a time interval. MPEG-7 standard is used as a format to store and exchange of annotation.

2.3 USING ANNOTATIONS TO IMPROVE SEARCH RESULTS

Annotations have been used for many years to analyze and describe documents. They have also been used to support collaboration and communication in group work. Recent research has explored these uses as well as annotation for classification and retrieval.

Annotations have been used to improve image and video retrieval. As mentioned previously, many image and video annotation systems have been developed for the retrieval purposes (Bargeron *et al.*, 1999, Sheiderman and Kang, 2000, Gertz *et al.*, 2002, Costa *et al.*,

[30] From (Costa *et al.*, 2002)

2002, Bottoni *et al.*, 2004). Commercial systems such as Google video, YouTube, Yahoo Video, Yahoo Photo, Flickr and Delicious, provide their users an ability to add annotations, which they call "tags", on images, multimedia and URL. They use these tags as indexes. The process is referred to as *social tagging*, *collaborative tagging*, *social book marking* or *mob indexing*. The information is used to retrieve images and multimedia content.

On March 29, 2007, Delicious had more than one million registered participants[31]. The tags and annotations created by participants provide information that can be used to improve web searching. Marlow *et al.* (2006) point out that the social annotations have the potential to improve the search for resources. However, published research on using social annotation to improve web search is sparse. Bao *et al.* (2007) is one significant paper.

Annotation has also been used to improve document retrieval. Denoue and Vignollet (2000) developed YAWAS to show how annotations improve document access and retrieval. Each annotation is composed of topic and type of the document, URL, type of selection text, comment and its sense (agree, disagree, …) and identification of author as shown in the following figure. These five attributes are used as search criteria in YAWAS search engine.

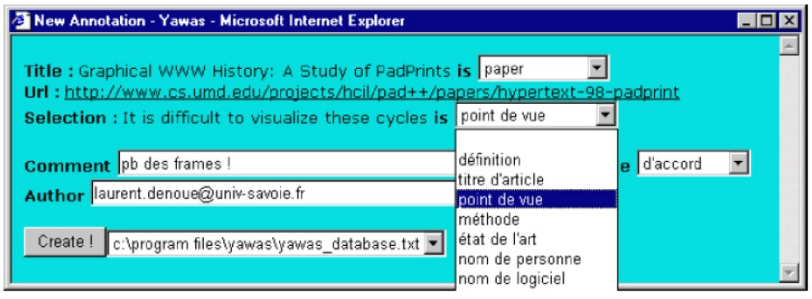


Figure 31. YAWAS annotation form[32]

[31] http://blog.del.icio.us
[32] From (Denoue and Vignollet, 2000)

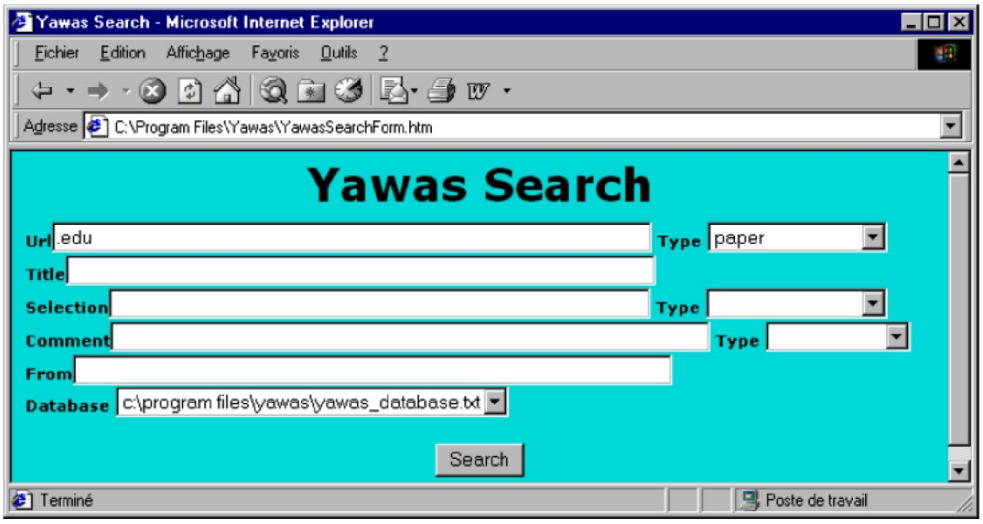


Figure 32. YAWAS search function form[33]

In determining how well annotations improve information access and retrieval, eighteen masters' students were divided into two groups and asked to write a summary report, which required them to read at least twenty documents. They were not permitted to speak to each other directly and could not print documents and annotate them. They were allowed to use bookmarks, email and a shared word processor (using NetMeeting). Only one group used YAWAS and the shared annotation server which allowed them to cooperatively annotate the documents. The annotations created by this group were used extensively as a way to summarize the documents when they wrote the report. They could easily create summaries of all annotated documents because the search engine provided by YAWAS help concatenate all highlighted text in each document. The group that used traditional bookmarks was unable to produce such summaries. Denoue and Vignollet observed that students did not need to access the full text of the document

[33] From (Denoue and Vignollet, 2000)

to retrieve what they were looking for. They concluded that annotation could be used to improve information access and retrieval.

Dmitriev *et al.* (2006) explored the use of user annotation to improve the quality of intranet search. They defined an annotation as a short description of the contents of a web page. In their study, both explicit annotation and implicit annotations were used. The explicit annotations were entered for each web page by users while implicit annotations were the query log, which recorded the queries users submitted and the results they clicked on. The basic idea was to treat the query as an annotation for pages relevant to the query. The figure 34 shows the format of the log file. They included both explicit and implicit annotation in the search engine index.

```
LogRecord ::= <Query> | <Click>
Query ::= <Time>\t<QueryString>\t<UserID>
Click ::= <Time>\t<QueryString>\t<URL>\t<UserID>
```

Figure 33. Format of Trevi log file [34]

They proposed several strategies to determine which pages were relevant to the query. The first strategy was that when user clicked on a page in the search results, the system assumed that this page was relevant to the query. However, it is possible to attach an annotation to an irrelevant page because user can click on a page that is not relevant to the query. The second strategy is based on the notion of a session. A session is a time-ordered sequence of clicks on search results that the user makes for a given query. It is assumed to be short when the page is not relevant to the query. The third and the fourth are based on the fact that users often reformulate their original query. They are similar to the previous two but they use query chains

[34] From (Dmitriev *et al.*, 2006)

instead of individual queries. The query chains are extracted from the log file based on the time stamps of the log records.

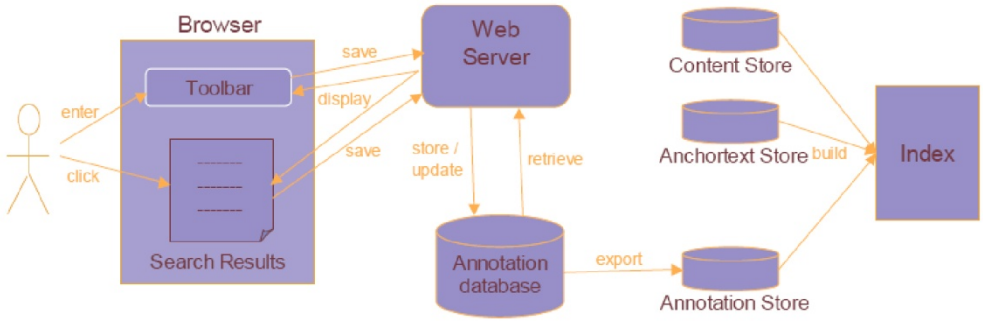


Figure 34. Flow of Annotation through the system[35]

As shown in figure 34 the annotations are stored a database for later display back to the user. Periodically, they are exported into an annotation store. The annotation store is a special format document repository used by the indexing system. Data in the annotation store, the content store, and anchor text store was used to produce the index. This is done by sequentially scanning these three stores in batch mode and using a disk-based sort merge algorithm for building the index.

In their experiment, the explicit annotation dataset consisted of 67 pages with a total of 158 annotations. The implicit annotation dataset consists of annotations extracted from the log files for the period of approximately 3 months. The table 5 shows the number of annotated pages with each strategy.

[35] From (Dmitriev *et al.*, 2006)

Table 5. The number of annotated pages with different implicit annotation strategies

Strategy	1	2	3	4
Number of Annotated Page	12,433	8,563	12,433	4,126

They used explicit annotations to generate 158 test queries. The annotated pages containing those explicit annotations were assumed to be the correct answer. They used the performance of the search engine without annotations as a baseline for their experiments. The table 6 shows the performance of explicit and implicit annotations in terms of the percentage of queries for which the correct answer was returned in the top 10 results.

Table 6. Summary of the results measured by the percentage of queries for which the correct answer was returned in the top 10

Baseline	Explicit Annotation	Implicit Annotation 1	Implicit Annotation 2	Implicit Annotation 3
8.9%	13.9%	8.9%	8.9%	9.5%

The improvement in search results with explicit annotations was statistically significant at the .05 level compared to the base line even though the results are rather low. One explanation for the low percentage of relevant results was that many annotations were attached to dynamically generated pages, which were not indexed by the search engine. Implicit annotations did not show any significant improvements. They planned to conduct more experiments to evaluate the implicit annotations and the differences among the four strategies. This suggests that explicit annotations have a greater potential to improve searching for intranet documents than do implicit annotations.

Bao *et al.*, (2007) explored the use of social annotation to improve web search. They observed that the social annotations can benefit web search in two aspects: 1) the annotations are

usually good summaries of web pages and, 2) the number of annotations indicates the popularity of web pages. They proposed two novel algorithms to use social annotations to rank web pages: SocialSimRank (SSR) and SocialPageRank (SPR).

SocailSimRank (SSR) calculates the similarity between queries and social annotations. Bao *et al.* (2007) express the idea behind this algorithm as follows:

> "*Similar (semantically-related) annotations are usually assigned to similar (semantically-related) web pages by users with common interests. In the Social annotation environment, the similarity among annotations in various forms can further be identified by the common web pages they annotated (p.503).*"

To explore the semantically related annotation, they built a bipartite-graph between social annotations and web pages with its edges indicating the user count. Assume that there are N_A annotations, N_P web pages and N_U users. M_{AP} is N_A x N_P association matrix between annotations and pages. $M_{AP}(a_x,p_y)$ is the number of users who assign annotation a_x to page p_y. S_A is the N_A x N_A matrix whose elements $S_A(a_i,a_j)$ indicates the similarity score between annotation a_i and a_j and S_P is the N_P x N_P matrix whose elements indicate the similarity between two web pages. The SSR algorithm is illustrated in the figure 35.

Algorithm 1: SocialSimRank (SSR)

Step 1 ***Init:*** ***Let*** $S_A^0(a_i, a_j) = 1$ for each $a_i = a_j$ otherwise 0

$S_P^0(p_i, p_j) = 1$ for each $p_i = p_j$ otherwise 0

Step 2 ***Do*** {

For each annotation pair (a_i, a_j) ***do***

$$S_A^{k+1}(a_i, a_j) = \frac{C_A}{|P(a_i)||P(a_j)|} \sum_{m=1}^{|P(a_i)|} \sum_{n=1}^{|P(a_j)|} \frac{\min(M_{AP}(a_i, p_m), M_{AP}(a_j, p_n))}{\max(M_{AP}(a_i, p_m), M_{AP}(a_j, p_n))} S_P^k(P_m(a_i), P_n(a_j)) \quad (2)$$

For each page pair (p_i, p_j) ***do***

$$S_P^{k+1}(p_i, p_j) = \frac{C_P}{|A(p_i)||A(p_j)|} \sum_{m=1}^{|A(p_i)|} \sum_{n=1}^{|A(p_j)|} \frac{\min(M_{AP}(a_m, p_i), M_{AP}(a_n, p_j))}{\max(M_{AP}(a_m, p_i), M_{AP}(a_n, p_j))} S_A^{k+1}(A_m(p_i), A_n(p_j)) \quad (3)$$

}***Until*** $S_A(a_i, a_j)$ converges.

Step 3 ***Output:*** $S_A(a_i, a_j)$

Figure 35. SocialSimRank algorithm

C_A and C_P are the damping factors of similarity propagation for annotation and web pages. $P(a_i)$ is the set of web pages annotated with annotation a_i and $A(p_j)$ is the set of annotations given to page p_j. $P_m(a_i)$ is the mth page annotated by a_i and $A_m(p_i)$ is the mth page annotation assigned to page p_i . In their experiment, both C_A and C_P are set to 0.7.

SocialPageRank (SPR) measures the quality of web pages from users' perspective. Bao *et al.* (2007) express the idea behind this algorithm as follows:

> "*High quality web pages are usually popularly annotated and popular web pages, up-to-date web users and hot social annotations have the following relations: popular web pages are bookmarked by many up-to-date users and annotated by hot annotations; up-to-date user like to bookmark popular pages and use hot annotations; hot annotations are used to annotate popular web pages and used by up-to-date users (p.504).*"

Assume that there are N_A annotations, N_P web pages and N_U users. M_{PU} is N_P x N_U association matrix between pages and users, M_{AP} is N_A x N_P association matrix between annotations and pages and M_{UA} is N_U x N_A association matrix between users and annotations. Element $M_{PU}(p_x,u_y)$ is assigned the count of annotations used by user u_y to annotate page p_x. Element M_{AP} and Element M_{UA} are initialized similarly. The SPR algorithm is illustrated in the figure 36.

Algorithm 2: SocialPageRank (SPR)

Step 1

Input:

Association matrices M_{PU}, M_{AP}, and M_{UA} and the random initial SocialPageRank score P_0

Step 2

Do:

$$
\begin{aligned}
U_i &= M_{PU}^{T} \cdot P_i && (5.1)\\
A_i &= M_{UA}^{T} \cdot U_i && (5.2)\\
P_i' &= M_{AP}^{T} \cdot A_i && (5.3)\\
A_i' &= M_{AP} \cdot P_i' && (5.4)\\
U_i' &= M_{UA} \cdot A_i' && (5.5)\\
P_{i+1} &= M_{PU} \cdot U_i' && (5.6)
\end{aligned}
\qquad (5)
$$

Until P_i converges.

Step 3:

Output:

P^*: the converged SocialPageRank score.

Figure 36. SocialPageRank algorithm

To evaluate both SocialSimRank and SocialPageRank, they used social annotation data crawled from Delicious during May 2006, which consists of 1,736,268 web pages and 269,566 different annotations. Fifty manual queries (MQ) were created by a group of CS students. The

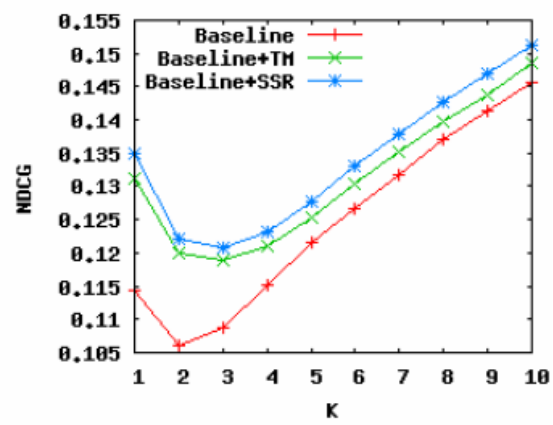
ground truth of each query was built by browsing the top 100 documents returned by Lucene search engine. Three thousand automatic queries were automatically extracted from the Open Directory Project.

In their experiment, Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) were used as evaluation metrics. They used BM25 formula to calculate the similarity between document and page content as the baseline.

They concluded that the SocialSimRank was able to find semantically related annotations and find more semantically related web pages. Figure 37 show the top four semantically related annotations by SSR. Figure 38 shows the comparison between NDCG of the term-matching and SocialSimRank on the AQ set and the comparison of MAP on both AQ and MQ.

Technology related:	
dublin	metadata, semantic, standard, owl
debian	distribution, distro, ubuntu, linux
Economy related:	
adsense	sense, advertise, entrepreneur, money
800	number, directory, phone, business
Entertainment related:	
album	gallery, photography, panorama, photo
chat	messenger, jabber, im, macosx
Entity related:	
einstein	science, skeptic, evolution, quantum
christian	devote, faith, religion, god

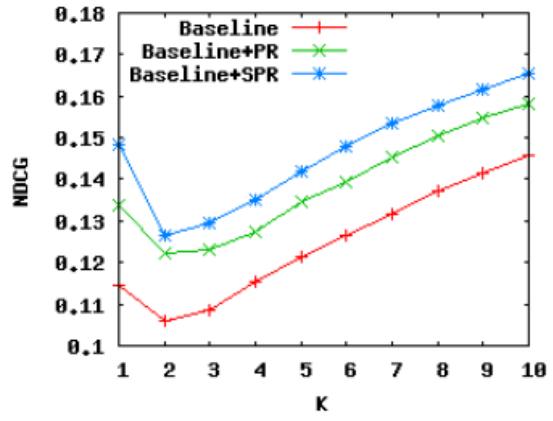
Figure 37. The top 4 semantically related annotations by SSR



Method	MQ50	AQ3000
Baseline	0.4115	0.1091
Baseline +TM	0.4341	0.1128
Baseline +SSR	**0.4697**	**0.1147**

Figure 38. The comparison between NDCG of the term-matching and SocialSimRank on the AQ set and the comparison of MAP on both AQ and MQ.

They point out that the SocialPageRank provides benefits to web search. Figure 39 shows the comparison between NDCG of PageRank and SocialPageRank on the AQ set and the comparison of MAP on both AQ and MQ.



Method	MQ50	AQ3000
Baseline	0.4115	0.1091
Baseline +PR	0.4141	0.1166
Baseline +SPR	**0.4278**	**0.1225**

Figure 39. The comparison between NDCG of the PageRank and SocialPageRank on the AQ set and the comparison of MAP on both AQ and MQ.

Combining both SocialSimRank and SocialPageRank, the best search result was achieved as shown in figure 40. T-test on MAP shows statically significant improvement.

Method	MQ50	AQ3000
Baseline	0.4115	0.1091
Baseline+SSR,SPR	**0.4724 (+14.80%)**	**0.1364 (+25.02%)**

Figure 40. Ranking with both SSR and SPR

Yanbe *et al.* (2007) also used social annotation to enhance web search. They proposed to combine Google PageRank with the one derived using social annotation, called Social Bookmarking Rank(SBRank). They use social annotation to support various query types. Their search engine does support not only content query but also other type of query e.g. temporal query, and sentiment query. Temporal queries can be constructed by exploiting timestamps associated with bookmarks. They found out that about 10% of tags used in social bookmarking systems are sentiment type tags, which are user feeling about that resource, such as funny, useful, and inspirational. These kinds of tags are used to implement the sentiment based search.

In their search engine, a tag and its frequency is converted to be a tag vector that represent a page's content. The vector model is used to measure the similarity between query and tags. Figure 41 and 42 show their proposed ranking formula and ranking algorithm.

$$Rank(p_i) = (1 + B(p_i)) \cdot (1 + F(p_i)) \cdot (1 + V(p_i)) \cdot (1 + C(p_i)) \cdot (1 + T(p_i, q)) \cdot (1 + T^{sen}(p_i, q)) \cdot (1 + S(p_i, t_{beg}, t_{end}))$$

where:

$$B(p_i) = \alpha \cdot SBRank(p_i) + (1 - \alpha) \cdot SearchRank(p_i)$$

$$F(p_i) = \beta \cdot \frac{FirstDate(p_i) - \min_{1 \le j \le n}(FirstDate(p_j))}{\max_{1 \le j \le n}(LastDate(p_j)) - \min_{1 \le j \le n}(FirstDate(p_j))}$$

$$V(p_i) = \gamma \cdot \frac{Var(p_i, FirstDate(p_i), LastDate(p_i))}{\max_{1 \le j \le n}(Var(p_j, FirstDate(p_j), LastDate(p_j)))}$$

$$C(p_i) = \delta \cdot \frac{N(comment_i)}{N(comment)}$$

$$T(p_i, q) = \alpha \cdot sim(tag_i, tag_q)$$

$$T^{sen}(p_i, q) = \alpha \cdot sim(tag^{sen}{}_i, tag_q)$$

$$S(p_i, t_{beg}, t_{end}) = \frac{AddBook(p_i, t_{beg}, t_{end})}{AddBook(p_i, FirstDate(p_i), LastDate(p_i))}$$

$SearchRank(p_i)$ is a rank of p_i

$SBRank(p_i)$ is the number of bookmarks of p_i

$FirstDate(p_i)$ is the first date when a bookmark was made to p_i

$LastDate(p_i)$ is the last date when a bookmark was made to p_i

tag_i is a tag vector of p_i

$tag^{sen}{}_i$ is a sentiment tag vector of p_i

tag_q is a tag vector of q

$AddBook(p_i, t_{left}, t_{right})$ is the number of bookmarks made to p_i in (t_{left}, t_{right})

$sim(tag^{sen}{}_i, tag_q)$ is similarity between tag vector of p_i and tag vector of q

Figure 41. Yanbe *et al.* proposed ranking formula

1. Obtain top n pages from search results returned by a search engine $P=\{p_1, p_2, \ldots, p_n\}$ for query q
2. Obtain *SBRank* values for each p_i where $p_i \in P$
3. Obtain every bookmark and its associated data for each p_i that has $SBRank > 0$ (i.e., the page has at least one social bookmark)
4. Count the number of occurrences of users and tags to be used for providing "Related Tags" and "Related Bookmarks" capabilities (described in Section 6.1)

Figure 42. Yanbe *et al.* proposed ranking algorithm

In their experiment, the Hatena Bookmark, a Japanese bookmarking system, was used as the data source. Manually created queries were issued to the system. Figure 43 shows example queries and the top 3 results. They concluded from the preliminary experiment that popularity of the page can indicate its quality and tag content can be used to filter the pages by user impression, sentiment characteristics or controversy levels.

Query	Top 3 Results	Google
search:digital library SBRank:0.5 lang:all	Internet Archive	77th
	CiteSeer: The NEC Research Institute Scientific Literature Digital Library	19th
	The Online Books Page	2nd
search:Vancouver SBRank:0.5 time: new lang: ja	World Peace Forum 2006@ Vancouver report Blog	98th
	Mapletown Vancouver information	9th
	Tourism Vancouver	1st
search:wii SBRank:0.5 emo:useful lang: ja	Wii-Tube: Let's watch YouTube with wii	63rd
	Yahoo News: a man confirmed effect of diet by playing wii	80th
	Itmedia Biz ID: Can we control PowerPoint with Wii remote?	33rd
search:iphone SBRank:0.5 from:20040101 to:20061201 lang:ja	Itmedia News:Is Apple iPhone appear in next a half year ?	4th
	A Fake iPhone CM which made too better : Gizmodo Japan	19th
	Various expectation of Apple iPhone Design - GIGAZINE	6th
search:gap-widening society SBRank:0.5 res: buzz lang: ja	Japanese gap-widening society from the point of view of India, a country which existing the caste	23rd
	Daily report from mad boy - three gap-widening societies	5th
	A thing desired by underdog sort of accept gap-widening society	42nd
search:sns SBRank:0.5 tag:compilation lang:ja	SNSLinK	47th
	SNS list | SNS portal site "SNS Navi"	13th
	SNS Navi: SNS information portal site about SNS, building SNS, etc	12th
search:web design SBRank:1 freq:cont	Stylegala - Web Design Publication	95th
	www.welie.com -- patterns in Interaction Design	77th
	Web Design Library — One-stop resource for web designers	28th
search:windows vista SBRank:1 lang:ja	Windows Vista Encyclopedia - From install to settings, application	15th
	Irregular column by Kazuhisa Nishikawa "Some reasons I can't like Windows Vista"	11th
	FrontPage – Windows Vista Wiki	13th
search:Kyoto sightseeing SBRank:1	Let's go Kyoto – Hot Kyoto sightseeing information	46th
	For Kyoto information, e-Kyoto net – whole Kyoto portal site	5th
	Kyoto sightseeing taxi: sightseeing in Kyoto, autumn tint guide and cherry blossom information	13th

Figure 43. Example queries and their top 3 results

Recently Google and Yahoo introduced new services using annotation to improve their search engine. *Google Co-Op* allows users to contribute information that helps Google to improve search results for everyone by allowing users to annotate or label URLs. Yahoo offered the similar kind of service, called *My Web*, which allows users to annotate URLs.

3.0 PRELIMINARY ANALYSIS AND RESEARCH DESIGN

This research demonstrates that social annotations contain useful information that can be used to improve a web search. Integrating social annotations into web searching improves users' satisfaction with the search results. The primary challenge is how to integrate the social annotation into the web search. This chapter describes the problem and research questions, shows the results of a preliminary analysis and outlines the research design.

3.1 PROBLEM AND RESEARCH QUESTIONS

As mentioned previously, published research on using social annotations to improve web search is sparse. Bao *et al.*, (2007) proposed two novel algorithms to use social annotations to rank search results. However, these algorithms are computational intensive. The algorithm may be too slow if the scale of social annotations keeps growing exponentially. In addition, Yanbe *et al*., (2007) converted social annotations and their frequency to be vector to represent a web document. Lacking experimental proof, it is still a question whether social annotation can be resource indexes using this approach. This work seeks to further investigate how to use social annotations to provide a practical ranking method and on how to use social annotations to improve resource indexes. The following are the research questions that need to be addressed.

1) How should the search results be ranked using social annotation?
2) How can the similarity between query and social annotations be measured?
3) How can social annotations be used as resource indexes?

3.2 PRELIMINARY ANALYSIS

In order to prepare to carry out the research described in the next chapter, it was necessary to clarify several concerns about using social annotations to improve web search. The first concern was how to obtain the social annotations. As mentioned in the first chapter, Delicious, is a web based bookmark sharing system that allows participants to store, share, and discover web resources. The participants can create tags for the resources with a number of freely chosen keywords, which are called, social tags. These tags will be viewed as the social annotations in this research study. We wrote programs that were able to successfully mine and store URLs and the tags associated with them on Delicious.

The next concern was the difference in the size of the resource collection in the search engine database and in the social bookmarking system. The following table shows the reported database size of commercial search engines.

Table 7. The search engine database size[36]

Search Engine	Reported Size	Page Depth
Google	8.1 billion	101K
MSN	5.0 billion	150K
Yahoo	4.2 billion (estimate)	500K
Ask Jeeves	2.5 billion	101K+

There is no publicly available report on the database size of the Delicious. Delicious has more than 1 million registered users. The social bookmarking systems do not have any crawler or spider to read all the web pages on the internet, registered users can enter them to the system database if they are useful resources. Thus, it is likely that the number of the resources in the social bookmarking systems is smaller than the number of the resources in search engine databases. This leads to the next question, which is whether or not social annotation on the limited resources can be used to improve web search. For this question, a small study was conducted to determine how many of the results returned from a search engine, Google, have been bookmarked in the social bookmarking system, Delicious. Ten technical related and non-technical related queries were manually created. The following table shows the number of results returned from Google that have been bookmarked in Delicious.

[36] http://blog.searchenginewatch.com/blog/041111-084221

Table 8. The number of URL returned from Google that have been bookmarked in Delicious

Queries	Number pf URL				
	Top 5	Top 10	Top 25	Top 50	Top 100
Q1(calories chinese food)	3	4	8	9	19
Q2(love poem greeting)	2	4	7	18	26
Q3(american food history)	5	7	12	20	31
Q4(thai food recipe)	5	10	20	34	49
Q5(auto buying guide)	4	7	17	25	46
Q6(java database example)	4	9	17	24	39
Q7(web design usability)	5	10	24	43	72
Q8(php database)	5	9	22	44	74
Q9(javascript ajax example)	5	10	25	49	89
Q10(java example)	5	9	17	30	47
% Average (Q1-Q10)	86.00%	79.00%	67.60%	59.20%	49.20%
% Average (Q1-Q5) : Non-Technical Queries	76.00%	64.00%	51.20%	42.40%	34.20%
% Average (Q6-Q10) : Technical Queries	96.00%	94.00%	84.00%	76.00%	64.20%

From Table 8, it can be concluded that technically oriented web resources tend to be bookmarked more than non-technical web resources. To explore this conclusion further, a snap shot of a tag cloud from Delicious as show in the following figure is taken into consideration. The similar conclusion can be made. Many popular tags are technical and computer related.

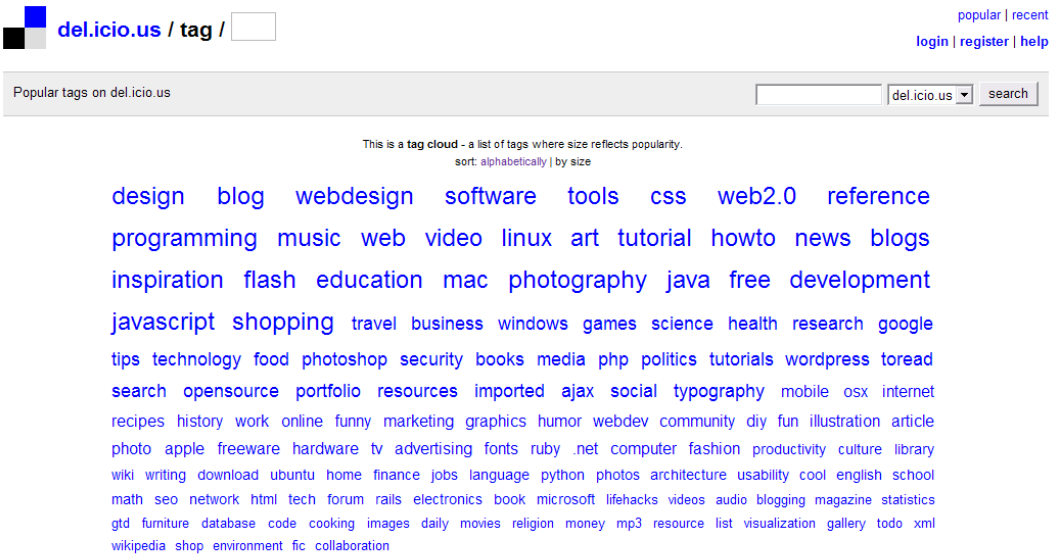


Figure 44. The tag cloud – listed by size

Given that technical content bookmarks are more prevalent in Delicious the questions and queries used in this research will be limited to technically related queries. At the end of this research study, if there is evidence that using social annotations on the technical related web resources help improve web searching for the technical resources, this conclusion may be generalizable to the general case as Delicious use becomes more widespread.

In this study, the search results returned from Google will be used as a base line when evaluating how relevant the search results are. How Google retrieves and ranks the search results is not reported in the public literature. As mentioned in an earlier chapter, most of the current commercial search engines use similarity ranking, also known as query-dependent ranking, to measure the match between query and the content of web resources. Beyond this, the systems use various other factors known as query-independent ranking, to measure the quality of the web resources. These query independent ranks, also known as static ranks, help improve the ranking of the search results obtained from the similarity ranking. Google uses a variation of PageRank

to improve the ranking of search results. There is no publicly available article explaining how Google combines PageRank with its similarity ranking. It was important to determine how much PageRank influences the search result ranking.

Another small study was conducted. In this study, thirty five queries were created. These queries were submitted to Google and AltaVista. The search results returned from both search engines were compared to find the number of matched results. The results returned from Google represent the PageRank influenced results, while the results returned from AltaVista represent the content based similarity results – assuming that Altavista does not use a similar static ranking method. The table 9 shows the overlap between Google and AltaVista for Top 20 and Top 100. An average number of overlapped results is 5.26 for Top 20 and 20.54 for Top 100. The Wilcoxon's Matched-Pairs Signed-Ranks Test was performed on the overlapped results to test whether the rank returned from the search engines was significantly different. Only one of the 35 queries rejected the null hypothesis. While this indicates an influence, it is difficult to draw any specific conclusion about the nature of PageRank's influence on the search result ranking. Further investigation was performed. Query number 1 was resubmitted to both search engines. This time all the results returned from Google and AltaVista were grabbed, 864 URLs returned from Google and 1000 URL returned from AltaVista. The overlap is only 228 URLs. When performing the Wilcoxon's Matched-Pairs Signed-Ranks Test, the null hypothesis was again rejected, which means the ranking of the results returned from Google is different from the AltaVista. Due to the low number of the overlapped results, it is still difficult to draw any firm conclusion.

Table 9. The number of the results overlapped between Google and AltaVista for top 20 and top 100

	Qurey String	Top 100	Top 20
1	java	41	12
2	"iowa straw poll results"	33	4
3	Harry Potter	33	6
4	"perseid meteor shower"	30	6
5	hurricane watch	27	7
6	pad thai recipe	26	7
7	pittsburgh weather	26	9
8	thai food	26	9
9	iMac	25	10
10	php ajax tutorial	25	4
11	yao ming wedding	25	3
12	central pacific hurricane center	23	6
13	"san jose jazz festival"	22	3
14	lisa lampanelli	22	5
15	wireless network security	22	6
16	cgi perl example	20	4
17	servlet ajax example	20	1
18	"big brother winner"	19	3
19	c socket programming	19	5
20	computer buying guide	19	0
21	regression analysis excel	19	8
22	"hawaii earthquake"	18	5
23	thai grocery store	18	8
24	"lauren conrad sex tape"	16	5
25	nfl score	16	7
26	pittsburgh light up night	16	7
27	tropical storm dean	16	4
28	"you'll never walk alone"	15	5
29	minnesota bridge collapse	15	3
30	brian crush adams	12	6
31	hawaii hurricane	12	5
32	java mysql	12	4
33	steelers mascot	11	1
34	java webservice tutorial	10	4
35	"perl how to program"	10	2

Finally, the site http://ranking.thumbshots.com/ provides a live picture of the overlap and ranking differences between any numbers of search engines. A few samples are shown below. While we still conclude that PageRank does influence both the top results returned by Google and the rank of those results, the lack of overlap between all the search engines makes it difficult

to draw a firm conclusion. It is likely that most of commercial search engines use their own complex algorithms in ranking their search result.

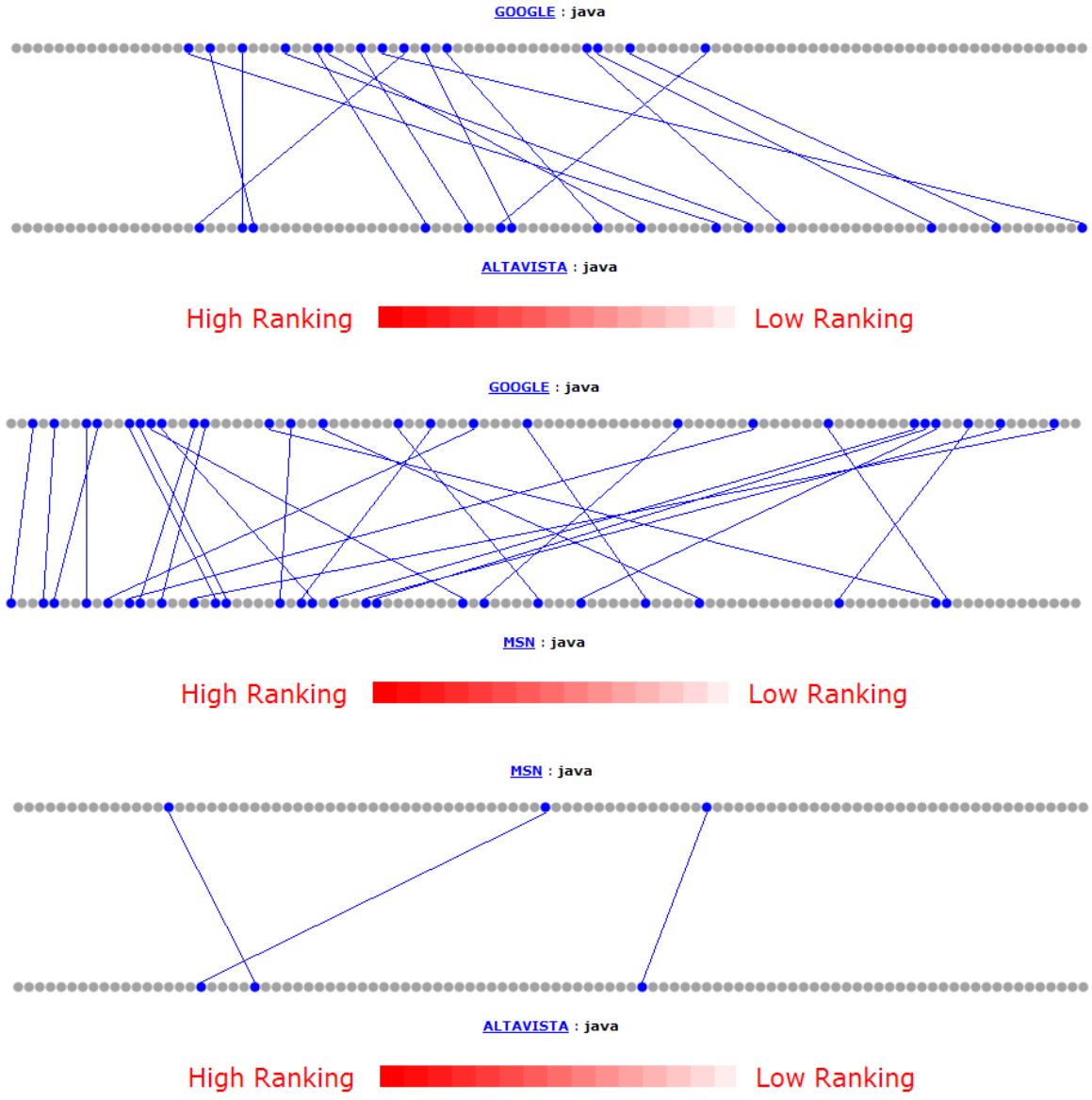


Figure 45. The search results overlapped between difference search engines

Due to the strong preference of people in using Google when they are searching for information on the internet, this research study will use the PageRank influenced Google result set as the base or current gold standard. We will explore whether social annotations can improve

web searching both as a supplement to Google and as an independent mechanism. The annotation information will be obtained from Delicious.

The last question is whether the social annotations help improve the web resource retrieval. To answer this question, another small experiment has been conducted. The objective of this experiment is to find out whether the web resources returned from Delicious are viewed as more relevant than the results returned from Google. In this experiment, ten subjects were recruited. They were asked to find web resources about how to develop web applications using servlets with a MySQL database connection. They were asked to formulate their own query and submitted it to Google and Delicious. Then, they judged the first 20 result URLs returned from both Google and Delicious in terms of whether it was relevant to their query. Table 10 shows the number of relevant web resources returned from Google and Delicious in the first 20 results.

Table 10. The number of relevant web resources returned from Google and Delicious in the first 20 results

	Google	del.icio.us
Subject 1	6	7
Subject 2	5	6
Subject 3	6	6
Subject 4	4	5
Subject 5	5	7
Subject 6	5	7
Subject 7	5	5
Subject 8	5	7
Subject 9	6	7
Subject 10	5	5
Average	5.2	6.2

The average number of relevant web resources return from Delicious is a little bit higher than from Google. As mentioned previously, the number of web resources in Delicious is likely to be much smaller than Google. The web resources in Delicious can be viewed as classified

resources by a group of Delicious users. They tend to be viewed as more relevant by the users. Thus, it seems reasonable to explore whether using social annotations obtained from Delicious might improve web search.

3.3 RESEARCH DESIGN

Three related experiments are proposed. The goal of the experiments is to determine if social annotations can be used to improve the user experience by providing a better set of resources and a ranking of search results compared to that provided by a commercial search engine, which is Google in this particular case.

3.3.1 Evaluation metric

Precision and recall are the most common metrics used to measure the retrieval performance of a system (van Rijsbergen, 1979). In traditional information retrieval research, test collections consisting of a set of documents, a set of queries, and expert relevance judgments for each document-query pair, are used. While these measures make sense for controlled retrieval usage, for the model proposed here, each subject, who will be a student in a programming course, will be considered an expert on whether a given retrieved document provides information that matches the query. The student decision about relevance is considered perfect. Agichtein *et al*., (2006) proposed a modified Discounted Cumulative Gain (DCG) for retrieval result rating, called Normalized Discount Cumulative Gain at K (NDCG at K). It was originally proposed by Jarvelin and Kekalainen (2000). This metric is based on human judgments. Human judges rate

how relevant each retrieval result is on an n-point scale. For a given query q, the ranked results are evaluated from the top ranked down, and the NDCG is computed as shown below.

$$NDCG_q = M_q \sum_{j=1}^{K} \frac{(2^{r(j)} - 1)}{\log(1 + j)}$$

Where Mq is a normalization constant calculated so that the perfect ordering would obtain NDCG of 1; each r(j) is an integer representing the relevancy rated by human judges (0 = "Not relevant at all" and 4="Perfectly Relevant" at position j).

NDCG rewards relevant documents in the top ranked results more heavily than those ranked lower and punishes irrelevant documents by reducing their contributions to NDCG (Agichtein et al., 2006). The NDCG will be used in this study to measure the performance of the search system.

3.3.2 Experiment 1: Re-ranking search results

In this experiment, social annotations will be used to adjust the ranking of search results. Given a set of documents that match a query with roughly equal closeness, it may be that a resource that has social annotations that match with the query string and has a high number of social annotations should be ranked higher than those with a low number of social annotations.

3.3.2.1 Variables and Expected Results

The independent variable is the ranking method, Google rank versus social annotation based rank. In social annotation based ranking, the number of people tagging a resource can be used to determine how interesting and informative they are.

The Delicious site allows users to provide tags for a resource. For any given resource, it is possible to determine how many users bookmarked a given resource, how many tags were used to describe a URL, and how many times each tag was used.

The Popularity Count (PC) is the simplest method. The ranking of search results can be obtained by ordering the number of people bookmarking web resources. The Propagated Popularity Count (PPC) is a Popularity Count of a web resource added to a Popularity Count of a domain page[37] of that web resource. The Query Weighted Popularity Count (QWPC) and Query Weighted Propagated Popularity Count (QWPPC) would simply be a PC and PPC of a given resource weighted by the count of shared terms between query string and tag of the web resource. Thus PC, PPC, QWPC, and QWPPC score would vary from 0 to unbounded number. When the PC is 0, it means no one bookmarked that resource. When PPC is 0, it means no one bookmarked either resource or its domain page. When the QWPC and QWPPC are 0, it means either the resource does not have a tag that matches a term in the query string or no one has bookmarked that resource. The maximum QWPC score will be the count of users who bookmarked the resource. Considering only the popularity of resources sounds reasonable, however, it may be better to consider how many times each tag was used as well.

The Matched Tag Count (MTC) would be an unbounded number that would sum the total number of users that used tags that matched terms in the query string. While, the Normalized Matched Tag Count (NMTC) takes the total count of all tag for a given resource into consideration. The NMTC can vary between 0-1. Each social annotation based ranking method has pros and cons. The experiment is designed to evaluate the performance of the proposed

[37] A domain page is the main index page of any web site (e.g. www.sis.pitt.edu) including index.html, index.htm, default.html, default.htm, and index.php.

ranking methods and compare them with Google's ranking. The social annotation based rankings can be obtained by the formulas in the table 11.

Table 11. Social annotation based ranking formulas

Social annotation based ranking method	**Formula**
Popularity Count	$NoOfUsersBookmarkedResource$
Propagate Popularity Count	$(10 \times NoOfUsersBookmarkedResource) + NoOfUsersBookmarkedDomainPageOfThisResource$
Query Weighted Popularity Count	$\frac{NoOfMatchedTag}{NoOfQueryTerm} \times \text{Popularity Count}$
Query Weighted Propagate Popularity Count	$\frac{NoOfMatchedTag}{NoOfQueryTerm} \times \text{Propagate Popularity Count}$
Matched Tag Count	$\sum_{i=1}^{n} UserCountOfTermMatched_i$
Normalized Matched Tag Count	$\sum_{i=1}^{n} \frac{UserCountOfTermMatched_i}{TotalCountOfAllTag}$
where n is the number of matched tag term with query string	

The dependent variable is the NDCG at K=10. Using ranking based on social annotation is expected to provide better NDCG than using ranking obtained from Google. The ranking formulas represent a progression, where PC and PPC is very much like PageRank. Other methods add query dependent ranking and assess the relationship between the tags and the query. QWPC and QWPPC is the most basic, simply asking how many of the query terms were found in any tags. MTC uses a metric that provides a higher ranking when more bookmarks use the query tag. NMTC is basically the same, but normalizes the rank value.

3.3.2.2 The 1st Hypothesis

H_0: There is no statistical difference between the means of the NDCG at K=10 of the Google ranking and the social annotation based rankings.

$(\mu_{google} = \mu_{pc} = \mu_{ppc} = \mu_{qwpc} = \mu_{qwppc} = \mu_{mtc} = \mu_{nmtc})$

H_1: Not all approaches are equal $(\mu_{google} \neq \mu_{pc} \neq \mu_{ppc} \neq \mu_{qwpc} \neq \mu_{qwppc} \neq \mu_{mtc} \neq \mu_{nmtc})$

3.3.2.3 Subjects, Evaluation, and Analysis Procedure

Twenty students from the IS and CS department were recruited as subjects for the experiment. They were individuals who were taking or had taken a programming course in Java or had experiences in using Java to develop an application. They were given a brief training session to ensure that they knew how to use the system. Each subject was given six questions. They were asked to find web pages that helped them answer each question. They formulated a query for each given question. They were asked to rate the relevancy of each resource in the retrieval result set on five-point scale; where '0' means not relevant at all, '1' means probably not relevant, '2' means less relevant, '3' means probably relevant, and '4' means extremely relevant. The subjects in this experiment were considered experts. Their relevancy ratings for each query were considered perfect.

To obtain the search result set for each query, the system sent the query to Google and got the search results back. The top 20 resources were presented in search result set. The search result set was then displayed in a randomized order to the subject for the relevancy rating. Before rating the relevancy, subjects were informed that the results would be displayed in a random order.

The relevancy ratings of each resource in the result set were used to rank the resources for the perfect ordering of the result set which were used as the normalization constant for NDCG. The relevancy ratings of each resource in the result set were recorded. For each query, the ranking of the search result set from Google and the ranking based on social annotations could be obtained. The value of NDCG at K=10 for each query and for each method were calculated.

To measure the consistency of the relevance judgments of the subjects, a modified Fleiss' kappa was used. One-way Analysis of Variance (ANOVA) was applied to test the hypothesis. The null hypothesis would be rejected if the results from the F-test indicate a significant difference at the 0.05 level. When the null hypothesis was rejected, all pairwise differences would be examined with the Scheffe procedure.

3.3.3 Experiment 2: Resource Indexing Augmentation

Full text indexing is the basis of information retrieval. Numerous variations on indexing have been tried over the years. Modern search engines use several methods to find additional metadata information to improve a resource indexing for enhancing the performance of the similarity ranking. As examples, document title (Hu *et al.*, 2005), anchor text (Brin and Page, 1998, Caswel *et al*., 2001, Westveld *et al*., 2002, Eiron and McCurley, 2003) and user query log (Xue *et al*., 2005) have all been used.

Craswell *et al.*, (2001) and Westerveld *et al.*, (2002) pointed out that anchor text helps improve the quality of search results significantly. The anchor text can be viewed as web page creator annotation. This experiment investigates how well social annotations, which are viewed as user annotations, contribute to the search results when they are used as resource indexes.

3.3.3.1 Variables and Expected Results

The independent variable is the method used to index the web resource. There will be three indexing approaches, full text indexing using the content of the resources, annotation indexing using only the annotations, and indexing using both annotations and the full text resource. The Term Frequency/Inverse Document Frequency (TF/IDF) will be used to compute the term weight. In this experiment, cosine similarity will be used as a query dependent ranking.

The dependent variables is the NDCG at K=10. The search results returned from the annotation indexing approach and annotation with full text indexing approach are expected to provide better NDCG than the search results returned from the full text indexing approach..

3.3.3.2 The 2nd Hypothesis

H_0: There is no statistical difference between the means of the NDCG at K=10 of from the full text indexing, annotation indexing and annotation with full text indexing approach $(\mu_{Full-Text} = \mu_{Annotation} = \mu_{Full-TextWithAnnotationt})$

H_1: Not all approaches are equal $(\mu_{Full-Text} \neq \mu_{Annotation} \neq \mu_{Full-TextWithAnnotationt})$

3.3.3.3 Subjects, Evaluation, and Analysis Procedure

Twenty students from the IS and CS department were recruited as subjects for the experiment. They were individuals who were taking or had taken a programming course in Java or had experiences in using Java to develop an application. They were given a brief training session to ensure that they know how to use the system. Each subject was given six questions. They were asked to find web pages that helped them answer each question. They formulated a query for each given question. They were asked to rate the relevancy of each resource in the retrieval result set on five-point scale; where '0' means not relevant at all, '1' means probably not relevant, '2'

means less relevant, '3' means probably relevant, and '4' means extremely relevant. The subjects in this experiment were considered experts. Their relevancy ratings for each query were considered perfect.

To obtain the search result set, the search engine was queried three times, once against each of the indexes created. The first top 15 resources of each approach were combined to form the search result set. This was done to make relevancy rating task easier for the subjects – it eliminated asking the user to rate any duplicated resources in the set more than once. This combined set was obtained by selecting one entry at a time from each result set in a round robin fashion until a set of twenty resources was obtained. It was assumed that there would be significant but not perfect overlap in the results sets. The search result set was then displayed in randomized order to the subject for the relevancy rating. Before rating the relevancy, subjects were informed that the results would be displayed in a random order.

The relevancy ratings of each resource in the result set were used to rank the resource for the perfect ordering of the result set which were used as the normalization constant for NDCG. The relevancy ratings of each resource in the result set were recorded. For each query, the ranking of the search result set from each indexing approach could be obtained. The value of NDCG at K=10 for each query and for each approach were calculated.

To measure the consistency of the relevance judgments of the subjects, modified Fleiss' kappa were used. One-way Analysis of Variance (ANOVA) was applied to test the hypothesis. The null hypothesis would be rejected if the results from the F-test indicate a significant difference at the 0.05 level. When the null hypothesis was rejected, all pairwise differences would be examined with the Scheffe procedure.

3.3.4 Experiment 3: Resource Retrieval and Ranking of Search Results

The search engine Google combines similarity ranking and static ranking together to rank the search results. Similarity ranking, also known as query-dependent ranking, measures the match between query and the content of the web document. Static ranking, also known as query-independent ranking, measures the quality of a web document, e.g. Google PageRank. Ranking search results can improve user satisfaction with information retrieval systems (Witten *et al.*, 1994). Resources, which are popularly annotated with consistency by many users, may indicate how interesting or informative they are. This experiment investigates how well social annotations contribute to the search results when they are used both as resource indexes and as means of ranking results.

A resource indexing approach that shows the best performance from the second experiment will be selected as the indexing approach for this experiment. A cosine similarity measurement will be used as a query dependent ranking. An annotation based ranking that shows the best performance from the first experiment will be selected as the ranking method for this experiment.

To exploit the social annotations for web searching, integrating both similarity and static feature to rank the resources may provide better ranking. The following is the formula for combining both the similarity ranking score and the static ranking score.

$$SimRank(Q,p) = \alpha\left(\frac{Sim(Q,p)}{MAX\{Sim(Q,p_1),...,Sim(Q,p_n)\}}\right) + (1-\alpha)\left(\frac{PR_{Ann}(p)}{MAX\{PR_{Ann}(p_1),...,PR_{Ann}(p_n)\}}\right)$$

where α is the weight of ranking technique

3.3.4.1 Variables and Expected Results

The independent variable is the ranking method, the Google ranking and social annotations based ranking with different weight ($\alpha = \left\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\right\}$).

The dependent variables is the NDCG at K=10. The ranking returned from the social annotation based ranking is expected to provide better NDCG than the Google ranking.

3.3.4.2 The 3rd Hypothesis

H_0: There is no statistical difference between the means of the NDCG at K=10 of the ranking returned from Google and the combination both similarity ranking score and static ranking score with $\alpha = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. $(\mu_{google} = \mu_{\alpha=0} = \mu_{\alpha=\frac{1}{4}} = \mu_{\alpha=\frac{1}{2}} = \mu_{\alpha=\frac{3}{4}} = \mu_{\alpha=1})$

H_1: Not all approaches are equal. $(\mu_{google} \neq \mu_{\alpha=0} \neq \mu_{\alpha=\frac{1}{4}} \neq \mu_{\alpha=\frac{1}{2}} \neq \mu_{\alpha=\frac{3}{4}} \neq \mu_{\alpha=1})$

3.3.4.3 Subjects, Evaluation, and Analysis Procedure

Twenty students from the IS and CS department were recruited as subjects for the experiment. They were individuals who are taking or had taken a programming course in Java or had experiences in using Java to develop an application. They were given a brief training session to ensure that they know how to use the system. Each subject was given six questions. They were asked to find web pages that help them answer each question. They formulated a query for each given question. They were asked to rate the relevancy of each document in the retrieval result set on five-point scale; where '0' means not relevant at all, '1' means probably not relevant, '2' means less relevant, '3' means probably relevant, and '4' means extremely relevant. The subjects

in this experiment were considered experts. Their relevancy ratings for each query were considered perfect.

To obtain the search result set for each query, the system sent the query to Google and got the search results back. At the same time, the query was sent to social annotation based search engine. The first top 15 resources of each approach were combined to form the search result set to make relevancy rating task easier for the subjects. This set was obtained by selecting alternating entries from each result set until all of the resources from the result set of all approaches have been selected – it was assumed that there would be significant but not perfect overlaped in the results sets. The search result set were then displayed in the randomized order to the subject for the relevancy rating. Before rating the relevancy, subjects were informed that the results were displayed in a random order.

The relevancy ratings of each resource in the result set were used to rank the resources for the perfect ordering of the result set which were used as the normalization constant for NDCG. The relevancy ratings of each resource in the result set were recorded. For each query, the ranking of the search result set from Google and social annotation based search engine with different weights of ranking techniques could be obtained. The value of NDCG at K=10 for each query and for each method were calculated.

To measure the consistency of the relevance judgments of the subjects, a modified Fleiss' kappa was used. One-way Analysis of Variance (ANOVA) was applied to test the hypothesis. The null hypothesis would be rejected if the results from the F-test indicated a significant difference at the 0.05 level. When the null hypothesis were rejected, all pairwise differences would be examined with the Scheffe procedure.

4.0 RESULTS AND DISCUSSIONS

This chapter describes the results of the experiments. First, the data that was used for the experiment and the development process for the questions used in the experiments are described. Second, participant demographics are described. Third, the results of each experiment are reviewed. Fourth, the consistency of the relevance judgments of the subject is reviewed. The chapter concludes with a discussion of the research results.

4.1 DELICIOUS DATA

Data was crawled from Delicious between November 2007 and January 2008. Beginning with a given tag, a set of URL's that have been tagged with that tag can be retrieved. Given those URL's, more tags can be identified and for each URL, the number of individuals who bookmarked the URL can be obtained along with the count of the number of times a given tag was used to describe that resource. Over the three months that Delicious was crawled, the crawlers looked at approximately 500,000 URLs and about 1,000,000 different tags. In all, the tagging and bookmarking of about 1,200,000 users was examined. Because many of the URLs collected were related to non technical resources – dining, vacations, business, etc. The crawlers were provided, by the researcher, with tag instances that should be used to terminate crawling into non-technical areas – to keep the data set and indexing issues manageable. The final set

consisted of 233,727 web pages and 544,467 unique tag annotations. Given the heavier use of delicious by technically oriented individuals and the subsequent design of the experiment based on technical issues related to programming, the resources and annotations collected were focused in the areas of programming. As indicated these non-technical resources and associated tags were discarded. Figure 46 shows the top 100 annotations with their frequencies in the document corpus.

Rank	Tag	Frequency
1	css	1,669,049
2	javascript	1,606,096
3	webdesign	1,598,007
4	design	1,535,883
5	programming	1,507,997
6	web	1,396,820
7	ajax	1,294,603
8	tools	1,227,746
9	reference	1,126,328
10	web2.0	1,074,260
11	software	915,092
12	java	867,874
13	development	766,403
14	tutorial	718,161
15	blog	583,958
16	opensource	520,523
17	html	512,831
18	free	506,899
19	firefox	503,806
20	howto	458,092
21	webdev	442,468
22	php	434,338
23	google	403,344
24	flash	321,036
25	rss	305,312
26	tips	282,116
27	search	281,942
28	linux	275,181
29	news	273,877
30	tutorials	272,159
31	library	269,112
32	graphics	264,775
33	code	260,222
34	inspiration	258,246
35	internet	251,548
36	ruby	224,717
37	xml	223,899
38	video	217,612
39	freeware	204,054
40	technology	202,532
41	resources	201,451
42	blogs	198,698
43	framework	189,820
44	art	184,052
45	cool	181,953
46	community	174,146
47	windows	172,825
48	database	171,270
49	mac	168,631
50	music	167,190

Rank	Tag	Frequency
51	online	166,713
52	social	166,535
53	tool	165,484
54	photography	163,539
55	api	158,357
56	mysql	157,394
57	security	155,323
58	color	155,220
59	prototype	151,473
60	rails	149,749
61	productivity	145,924
62	books	145,529
63	browser	143,969
64	tech	138,395
65	imported	138,248
66	xhtml	137,388
67	gallery	137,267
68	business	136,681
69	extension	136,486
70	images	133,749
71	wiki	133,658
72	fonts	129,047
73	usability	124,237
74	python	123,670
75	extensions	121,830
76	visualization	121,715
77	documentation	119,935
78	fun	119,264
79	photoshop	118,450
80	download	118,059
81	osx	117,055
82	layout	116,315
83	photo	115,684
84	icons	114,640
85	computer	113,832
86	generator	112,106
87	collaboration	110,214
88	typography	108,313
89	education	107,820
90	dhtml	107,329
91	eclipse	105,664
92	article	104,765
93	photos	104,650
94	maps	103,947
95	wordpress	102,369
96	cms	100,827
97	sql	99,082
98	games	94,447
99	templates	94,295
100	research	92,634

Figure 46. The top 100 tags with their frequency

To prepare the data before the experiments, three different crawlers and two small data processing programs were developed. Each crawler had different responsibilities. The first crawler, called Delicious Resource Crawler, was responsible for retrieving web resources bookmarked on the Delicious. These web resources (URLs) were stored in a database. The database were used to store web search collection that used by different social annotation based search engines in all experiments.

The second crawler, called Delicious Annotation Crawler, was responsible for retrieving social annotation information, that is information about who bookmarked a resource, when it was bookmarked, what tag(s) were used, and what note were attached. This crawler read URLs from the database and downloaded social annotation information from Delicious. Delicious provided a RSS file that contained all social annotation information of the web resource. Delicious limits downloads of RSS file -- it only allows 100 RSS downloads per hour. If the crawler downloads more than 100 RSS files in one hour, the crawler will be blocked and will not be able to download RSS files from the Delicious for at least two hours. To make sure that the crawler was not blocked, it was configured to download 95 RSS files every hour. Due to the short period of time in preparing the data for the experiments, seven machines were running this crawler. The following figure shows an example of partial downloaded RSS file for http://www.sis.pitt.edu/.

The third crawler, called Web Resource Crawler, was responsible for downloading web documents. The downloaded web documents were used during indexing processes for the second experiment

```
<?xml version="1.0" encoding="UTF-8"?>

<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns="http://purl.org/rss/1.0/"
 xmlns:content="http://purl.org/rss/1.0/modules/content/"
 xmlns:taxo="http://purl.org/rss/1.0/modules/taxonomy/"
 xmlns:dc="http://purl.org/dc/elements/1.1/"
 xmlns:syn="http://purl.org/rss/1.0/modules/syndication/"
 xmlns:admin="http://webns.net/mvcb/"
>

<channel rdf:about="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6">
<title>del.icio.us bookmarks for http://www.sis.pitt.edu/</title>
<link>http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6</link>
<description></description>
<items>
 <rdf:Seq>
  <rdf:li rdf:resource="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#ucpl1978" />
  <rdf:li rdf:resource="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#rcp_delicious" />
  <rdf:li rdf:resource="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#fontgoddess" />
  <rdf:li rdf:resource="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#zelditch" />
  <rdf:li rdf:resource="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#tylerstjohn" />
  <rdf:li rdf:resource="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#creativejuices" />
</rdf:Seq>
</items>
</channel>

<item rdf:about="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#ucpl1978">
<title>[from ucpl1978] School of Information Sciences - University of Pittsburgh</title>
<link>http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#ucpl1978</link>
<description></description>
<dc:creator>ucpl1978</dc:creator>
<dc:date>2008-02-28T18:25:50Z</dc:date>
<dc:subject></dc:subject>
</item>

<item rdf:about="http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#rcp_delicious">
<title>[from rcp_delicious] U of Pitt: School of Information Sciences</title>
<link>http://del.icio.us/url/ef6e880d3544af2748aff7f09d6981e6#rcp_delicious</link>
<description></description>
<dc:creator>rcp_delicious</dc:creator>
<dc:date>2008-01-26T18:03:01Z</dc:date>
<dc:subject>school sis university_of_pittsburgh</dc:subject>
<taxo:topics>
 <rdf:Bag>
  <rdf:li resource="http://del.icio.us/tag/school" />
  <rdf:li resource="http://del.icio.us/tag/university_of_pittsburgh" />
  <rdf:li resource="http://del.icio.us/tag/sis" />
 </rdf:Bag>
</taxo:topics>
</item>
```

Figure 47. Partial RSS file for http://www.sis.pitt.edu/

After the social annotation information in the form of an RSS file was obtained, a program, called RSS Processor, was activated to extract social annotation information and store it in the database. For each web resource, username, tags used, notes added and bookmark dates were stored in the database. Another small program, called Social Annotation Processor, was used to prepare social annotation information for indexing and ranking purposes. The count of number of people who bookmarked web resources, the count of the tag used for each web resource was stored in the database.

4.2 QUESTIONS USED IN THE EXPERIMENTS

As mentioned in the previous chapter, subjects in all three experiments were given six questions and were asked to find web resources that helped them answer those questions. Developing the questions were one of the critical success factors for this study. Human judgements on the relevancy of search results are important. They were used for evaluation of each ranking method and each indexing approach in all three experiments. The given questions should be at the same level as the subjects' experience and knowledge.

In all three experiments, subjects could be individuals, who were taking or had taken a programming course in java or had experience in using Java to develop an application. They could be in any level of education.

Course descriptions and syllabi for Java programming courses were reviewed. Ten questions were created based on the course materials. They were divided into two groups, specific questions and exploratory questions. The specific questions asked for specific information e.g. Find an example of how to format a date object using Java. On the other hand,

the exploratory questions asked for general information related to the topic in the question e.g. Find information about sorting algorithms and explain them. This kind of questions allowed subjects to judge the relevancy according to their understanding of the question and their preference.

These ten questions were first tested with a native English speaking doctoral student to correct the wording in the question. After that an undergraduate student, who had taken an intermediate Java programming course was asked to find web resources that helped him answer those questions. He faced difficulties with some questions, e.g. database connection with Java and GUI in Java. The questions that the undergraduate student faced difficulties were dropped.

The final set contained 8 questions as shown below. Question A-D are specific questions while question E-H are the exploratory questions. Each subject was given three specific questions and three exploratory questions. The questions were manually preselected and assigned to the Subjects to make sure that there were 15 subjects who got each question.

- Question A: *Find an example that shows how to write text to a file (output to a file) using Java*
- Question B: *Find a Java example that shows how to format a Date object e.g. dd-MMM-yy*
- Question C: *Find an example that shows how to use ArrayList in Java*
- Question D: *Find an example that shows how to write a recursive program in Java*
- Question E: *There are many different sorting algorithms e.g. Bubble sort, Quick sort, Heap sort. Find information about sorting algorithms and explain them.*

- Question F: *Java method modifiers can be public, private, or protected. Find information about these modifiers and explain the differences.*
- Question G: *Find information about features of the various Integrate Development Environment (IDE) for Java that currently in the market e.g. NetBeans, Eclipse, etc.*
- Question H: *Explain how try-catch blocks work to allow you to account for multiple exceptions (I/O exception, FileNotFound exception, EOF exception, and etc.) in a section of Java code*

4.3 PARTICIPANT DEMOGRAPHICS

The experiments were carried out between January and March 2008. Sixty five students from both University of Pittsburgh and Carnegie Mellon University applied to be participants. Five participants withdrew from the experiments. The First-Come-First-Serve basis was applied to assign the participants to the experiments. First twenty participants were assigned to be the subjects in the first experiment. Next twenty participants were assigned to be the subjects in the second experiment. The last twenty participants were assigned to be the subject in the last experiment.

All participants completed the entry questionnaires which solicited demographic data as well as their experience in Java and in searching for information on the web. Forty two participants were male. Seventeen participants were between the ages of 18 to 22 years old, eighteen between the ages of 23 to 27 years old, sixteen between the ages of 28 to 32 years old, and nine were 33 or older.

The educational levels of the participants were as follows: 22% undergraduate, 40% graduate and 38% doctoral. Sixty percent of participants are studying in a information science and information system management major, twenty percent are in a computer science major, and twenty percent are in other majors e.g. computer engineering and business administration.

Twenty-five percent of participants considered themselves to be experts in Java, fifty-three percent considered themselves to be intermediate in Java, and twenty-two percent considered themselves to be beginner in Java. Regarding how long the participants have been using Java, 27% using for more than 3 years, 42% using for 1 to 3 years, and 31% using for less than 1 year.

Among the participants, 57% reported searching for information on the internet more than fifteen times a day, 15% eleven to fifteen times a day, 25% six to ten time a day, and 3% less than six times a day. Ninety-five percent of participants reported that most of the time they find what they want. Table 12 shows a summary of the participants' demographics.

Table 12. A summary of the participants' demographics

	Experiment 1	Experiment 2	Experiment 3	Total
Gender				
Female	4	9	5	18
Male	16	11	15	42
Age				
18-22	7	3	7	17
23-27	5	8	5	18
28-32	4	6	6	16
>=33	4	3	2	9
Level of Study				
Undergraduate	4	4	5	13
Graduate (Master)	11	6	7	24
Doctoral	5	10	8	23
Major of Study				
Computer Science	2	3	7	12
Information Science and Information System Management	14	14	8	36
Others	4	3	5	12
Java Knowledge				
Expert	5	6	4	15
Intermediate	9	7	16	32
Beginner	6	7	-	13
Learn Java				
Self-Study	5	5	4	14
Take Class	15	14	16	45
On the Job Training	-	1	-	1
Others	-	-	-	-
Using Java				
<1 years	9	4	6	19
1-3 years	6	9	10	25
>3 years	5	7	4	16
Searching for Information on the Internet				
<=5 times per day	1	-	1	2
6-10 times per day	5	7	3	15
11-15 times per day	3	2	4	9
>15 times per day	11	11	12	34
Found Information				
Always found what looking for	-	-	1	1
Most of the time found what looking for	19	20	18	57
Half of the time found what looking for	1	-	1	2
Never found what looking for	-	-	-	-

4.4 THE CONSISTENCY OF THE RELEVANT JUDGMENTS OF THE SUBJECTS

In the previous chapter, Fleiss' kappa was proposed to measure the consistency of the relevance judgments of the subjects. The Fleiss' kappa is a statistical measure of inter-rater reliability. It is suitable for any number of raters rating a fixed number of items. Figure 48 shows the equations of Fleiss's kappa.

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

$$\bar{P} = \frac{1}{N}\sum_{i=1}^{N} p_i$$

$$\bar{P}_e = \sum_{j=1}^{k} p_j^2$$

$$p_i = \frac{1}{n(n-1)}\left(\sum_{j=1}^{k} n_{ij}^2 - n\right)$$

$$p_j = \frac{1}{Nn}\sum_{i=1}^{N} n_{ij}$$

κ is the Kappa

$\bar{P}$ is the mean of the extent to which rater agree for the i-th item p_i

$\bar{P}_e$ is the sum of square of the proportion of all assignments which were to the j-th category p_j

N is the total number of items

n is the number of ratings per item

k is the number of categories

n_{ij} is the number of rater who assigned the i-th item to the j-th categories

Figure 48. Equations of Fleiss's kappa

Subjects were asked to find web pages that helped them answer a given set of questions. As much as possible, the research setting imitated the real environment where people search for information on the web. For each question, subjects had to formulate their own query and submit it to search engines.

Subjects were expected to use similar queries for a particular question. There were, however, differences in queries used for questions. As a result, the search results returned for each subject were different. Each resource had a different number of raters. As mentioned in the previous section, the maximum number of raters for earch resource related to a query was 15. Because the overlap of search results returned for each subject was low, there were differences in number of raters for each resource. The following figures show number of raters of each resource for experiment 1, 2, and 3 respectively. In Experiment 1, only 2.15% of the URLs were rated by 8 or more subjects. In experiments 2 and 3, that percentage grew to 7.65% and 7.68% respectively.

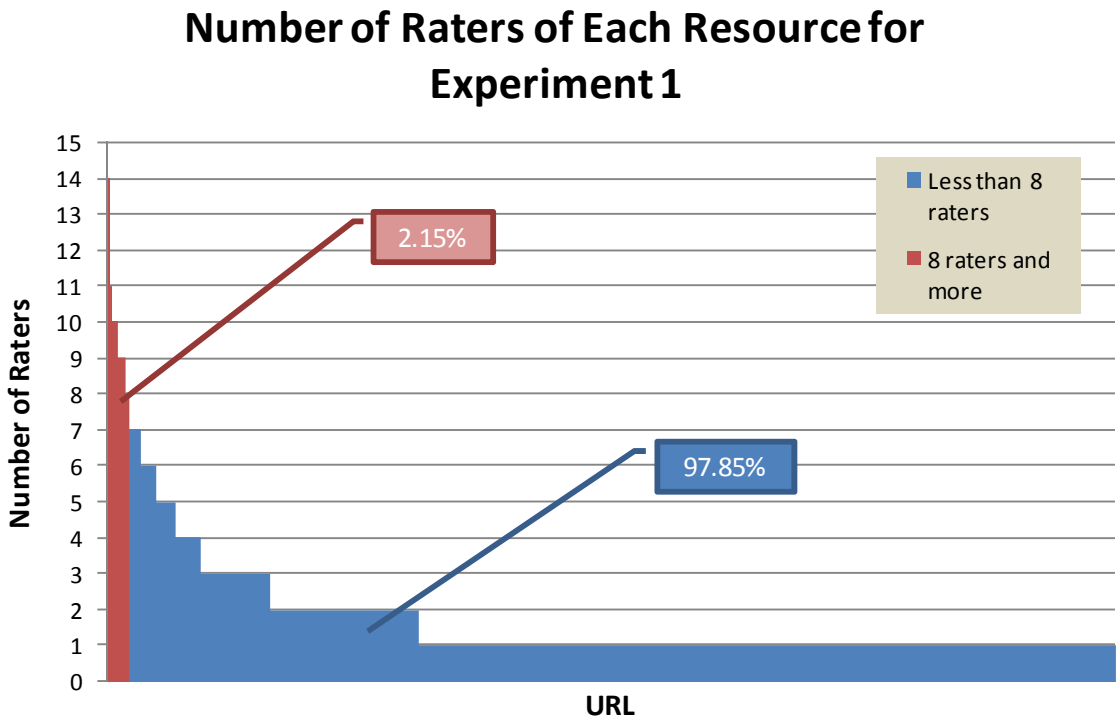


Figure 49. Number of rater of each resource for experiment 1

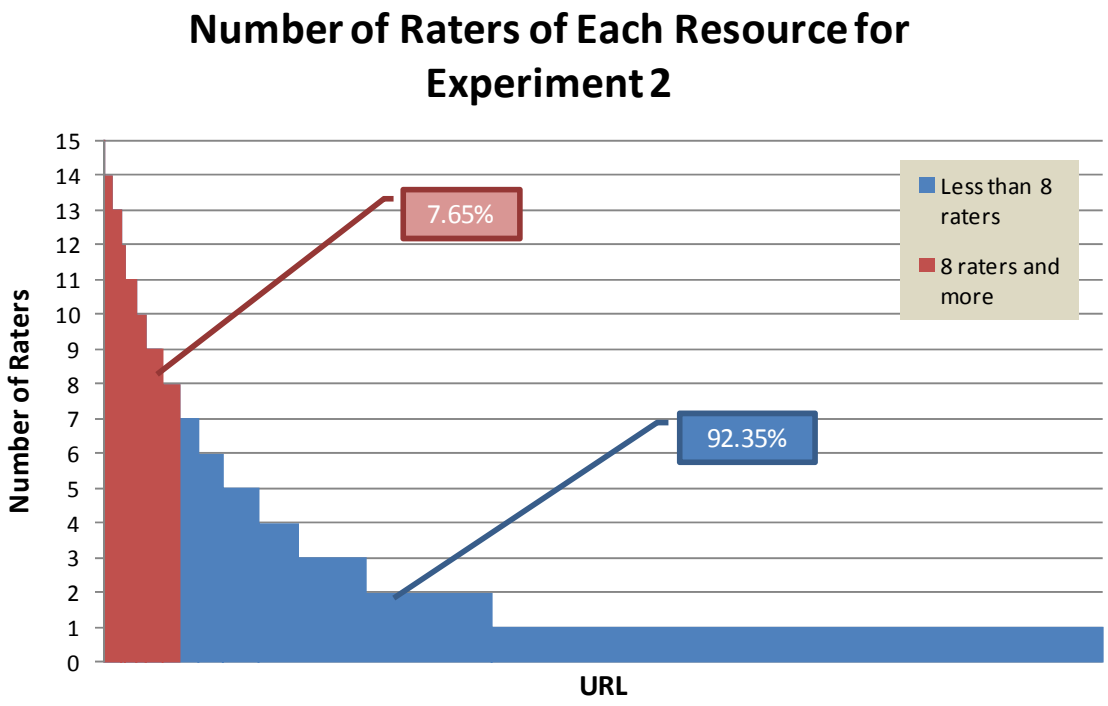


Figure 50. Number of rater of each resource for experiment 2

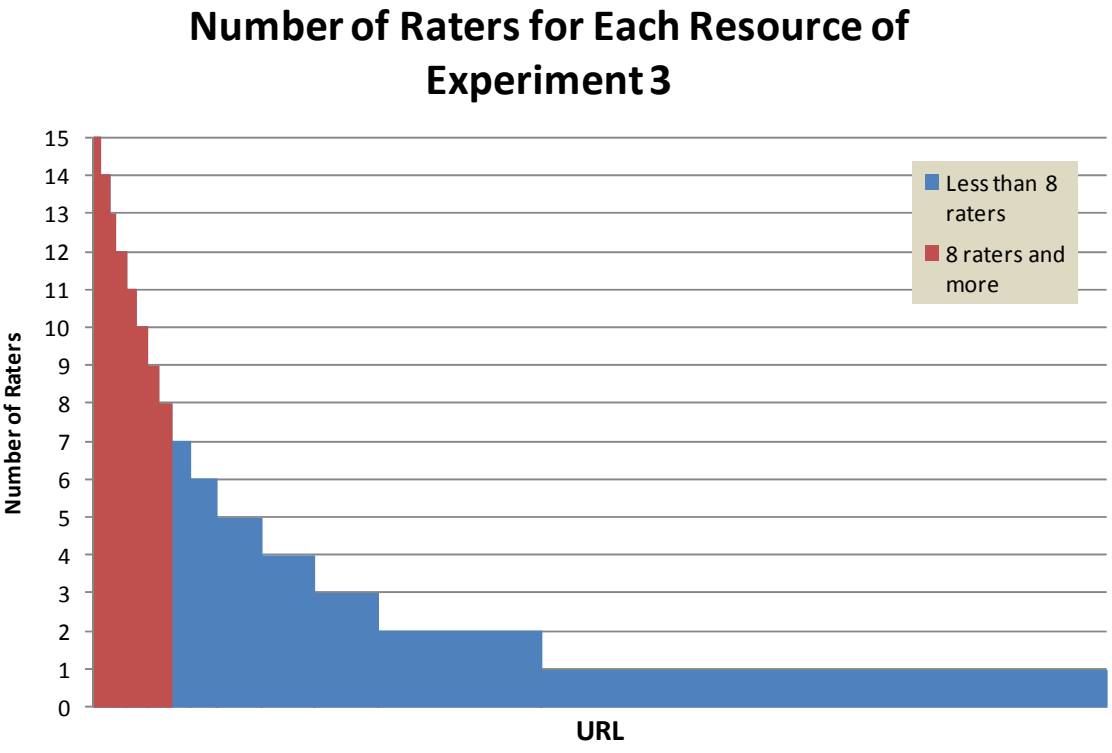


Figure 51. Number of rater of each resource for experiment 3

Because the Fleiss' kappa requires each rater rate each resource, it could not be used in totality to measure the consistency of the relevance judgments of the subjects. However, the row formulated equations can be used to provide an average rating agreement across the resources.

$$\bar{P} = \frac{1}{N}\sum_{i=1}^{N} p_i$$

The problem of using a part of Fleiss's kappa formula was the interpretation of the $\bar{P}$ obtained. To make some assessment of the value obtained in the experimental results, the value of $\bar{P}$ when ratings are randomly generated was calculated. A simulation was set up by assuming that there were 10 subjects rating 10 resources on a scale 0 to 4. The simulation was run over 1000 trials and $\bar{P}$ was calculated. The average of $\bar{P}$ for 1000 trails was 0.200958. $\bar{P}$ when the ratings are in perfect agreement is 1.0. These values were used to compare with the $\bar{P}$ for all three experiments.

For resources that have more than eight subject relevancy judgments, the average rating agreement was computed using a part of Fleiss's kappa equations. (For each resource, the maximum number of subjects that could rate the resource related to a particular query is 15, so we are selecting those resources rated by at least 50% of the subjects. Figures 64-66 show means ($\bar{P}$) of the extent to which rater agree for the i-th item (p_i)in all three experiments. The means ($\bar{P}$)for all three experiments are 0.49616, 0.46434 and 0.50739 respectively. Table 13-15 summarize the average for each question on each experiment. Compared with the value from the simulation, this suggests that the ratings from subjects in all three experiments showed moderate agreement. Appendix B shows the overlap of the search results in all three experiments with subjects' relevance judgment.

Table 13. The means of the extent to which raters agree about a resource in experiment 1

Experiment 1								
Resource	**Question**	**Number of Subjects Rating**					**Total Number of Raters**	**Pi**
		0	**1**	**2**	**3**	**4**		
http://www.javacoffeebreak.com/faq/faq0004.html	A	0	0	2	3	5	10	0.31111
http://abbeyworkshop.com/howto/java/writeText/index.htr	A	1	0	1	1	7	10	0.46667
http://www.javapractices.com/topic/TopicAction.do?Id=42	A	0	1	1	3	5	10	0.28889
http://searchdomino.techtarget.com/tip/0,289483,sid4_gci1	A	7	1	1	0	0	9	0.58333
http://www.leepoint.net/notes-java/io/10file/10readfile.htn	A	2	3	1	1	1	8	0.14286
http://javatechniques.com/blog/dateformat-and-simpledate	B	0	0	1	1	7	9	0.58333
http://www.wellho.net/resources/ex.php4?item=j714/Arlist	C	1	0	2	3	8	14	0.35165
http://kickjava.com/220.htm	C	1	0	0	0	13	14	0.85714
http://www.anyexample.com/programming/java/java_array	C	0	1	0	0	10	11	0.81818
http://users.cs.dal.ca/~sedgwick/ArrayList.html	C	0	0	0	4	6	10	0.46667
http://www.javafaq.nu/java-example-code-classes-2288.htm	C	1	1	1	2	4	9	0.19444
http://www.idevelopment.info/data/Programming/java/PRC	C	0	3	4	2	0	9	0.27778
http://www.javafaq.nu/java-example-code-89.html	C	0	0	0	0	9	9	1.00000
http://www.idevelopment.info/data/Programming/java/coll	C	0	0	0	2	7	9	0.61111
http://www.javadeveloper.co.in/java-example/java-arraylist	C	0	0	0	0	9	9	1.00000
http://danzig.jct.ac.il/java_class/recursion.html	D	0	0	0	1	14	15	0.86667
http://weblogs.java.net/blog/mortazavi/archive/2005/08/re	D	1	1	0	1	8	11	0.50909
http://www.tech-recipes.com/java_programming_tips1266.l	D	0	0	1	0	9	10	0.80000
http://www.ibm.com/developerworks/java/library/j-diag8.h	D	1	3	1	1	4	10	0.20000
http://www.ib-computing.com/java/datastructures/recursio	D	0	0	0	2	8	10	0.64444
http://www.ahmadsoft.org/articles/recursion/index.html	D	2	0	0	2	4	8	0.28571
http://www.hostitwise.com/java/java_recursion.html	D	0	0	1	1	6	8	0.53571
http://www.softpanorama.org/Algorithms/sorting.shtml	E	1	1	0	3	7	12	0.36364
http://en.wikipedia.org/wiki/Sorting_algorithm	E	0	0	0	0	10	10	1.00000
http://www.allapplabs.com/interview_questions/java_interv	F	1	0	0	3	5	9	0.36111
http://www.ibm.com/developerworks/java/library/j-emacs/	G	1	0	2	4	2	9	0.22222
http://developers.sun.com/jsenterprise/	G	1	2	1	2	2	8	0.10714
http://www.dreamincode.net/forums/showtopic22661.htm	H	0	0	1	3	5	9	0.36111
http://www.ociweb.com/jnb/archive/jnbMay2000.html	H	0	1	2	3	2	8	0.17857
							Average	**0.49616**

Table 14. The means of the extent to which raters agree about a resource in experiment 2

Experiment 2								
Resource	Question	Number of Subjects Rating					Total Number of Raters	Pi
		0	1	2	3	4		
http://java.sun.com/developer/onlineTraining/Programming	A	1	2	1	3	7	14	0.27473
http://saloon.javaranch.com/forums/forum-038.html	A	4	3	1	4	1	13	0.19231
http://www.javacoffeebreak.com/java103/java103.html	A	0	0	0	2	9	11	0.67273
http://www.javapractices.com/Topic42.cjp	A	0	0	1	1	8	10	0.62222
http://www.tizag.com/phpT/fileread.php	A	5	3	0	0	0	8	0.46429
http://www.dan.co.uk/mysql-date-format/	B	10	1	0	1	0	12	0.68182
http://www.joda.org/	B	6	1	2	1	1	11	0.29091
http://www.svendtofte.com/code/date_format/	B	6	0	1	1	2	10	0.35556
http://www.mattkruse.com/javascript/date/	B	7	1	1	0	1	10	0.46667
http://dev.mysql.com/doc/refman/5.0/en/date-and-time-fu	B	8	1	0	0	1	10	0.62222
http://en.wikipedia.org/wiki/ISO_8601	B	8	2	0	0	0	10	0.64444
http://msdn2.microsoft.com/en-us/library/8kb3ddd4.aspx	B	7	2	1	0	0	10	0.48889
http://www.methods.co.nz/rails_date_kit/rails_date_kit.htm	B	8	1	1	0	0	10	0.62222
http://www.databasejournal.com/features/mssql/article.ph	B	5	3	0	1	0	9	0.36111
http://www.daniweb.com/code/snippet515.html	B	0	0	0	1	8	9	0.77778
http://www.w3.org/TR/NOTE-datetime	B	5	2	1	1	0	9	0.30556
http://java.sun.com/j2se/1.4.2/docs/api/java/text/SimpleDa	B	0	1	2	2	4	9	0.22222
http://javatechniques.com/public/java/docs/basics/dateforr	B	0	0	0	1	8	9	0.77778
http://javafaq.nu/java-example-code-list.html	B	0	0	2	2	4	8	0.28571
http://www.mattkruse.com/javascript/date/source.html	B	4	1	0	2	1	8	0.25000
http://www.codefetch.com/search?qy=date&lang	B	8	0	0	0	0	8	1.00000
http://javatechniques.com/blog/dateformat-and-simpledate	B	0	0	0	0	8	8	1.00000
http://www.java-examples.com/java-collections-and-data-st	C	1	0	1	1	11	14	0.60440
http://www.scribd.com/doc/259808/Collections-and-Generi	C	3	3	2	4	2	14	0.15385
http://java.sun.com/j2se/1.4.2/docs/api/java/util/ArrayList.h	C	1	1	3	5	4	14	0.20879
http://www.idevelopment.info/data/Programming/java/coll	C	0	1	0	1	11	13	0.70513
http://blogs.msdn.com/joshwil/archive/2004/04/13/112598	C	4	2	0	3	4	13	0.20513
http://www.onjava.com/pub/a/onjava/2003/03/12/java_co	C	1	1	2	3	6	13	0.24359
http://www.javaworld.com/javaworld/javaqa/2001-06/03-q	C	3	4	5	0	1	13	0.24359
http://www.kickjava.com/220.htm	C	0	0	0	2	11	13	0.71795
http://www.pankaj-k.net/archives/2004/06/arraylist_versu.h	C	1	6	3	1	2	13	0.24359
http://www.idevelopment.info/data/Programming/java/PRC	C	0	1	2	0	10	13	0.58974
http://blogs.worldnomads.com.au/matthewb/articles/187.a	C	13	0	0	0	0	13	1.00000
http://www.scribd.com/doc/271835/Memory-Leaks-in-Java-	C	7	4	0	1	0	12	0.40909
http://www.javadeveloper.co.in/java-example/java-arraylist	C	0	0	0	1	10	11	0.81818
http://www.onjava.com/pub/a/onjava/2001/05/30/optimiz	C	2	3	1	1	3	10	0.15556
http://www.javainthebox.net/laboratory/J2SE1.5/LangSpec/	C	3	1	1	0	4	9	0.25000
http://www.nextindex.net/java/collection/ArrayList.html	C	3	1	1	0	4	9	0.25000
http://eclipsetutorial.sourceforge.net/Total_Beginner_Comp	C	1	1	3	1	3	9	0.16667
http://java.sun.com/developer/JDCTechTips/2002/tt0910.ht	C	0	0	1	1	7	9	0.58333
http://www.rgagnon.com/javadetails/java-0521.html	C	0	0	2	1	5	8	0.39286
http://www.precisejava.com/javaperf/j2se/Collections.htm	C	1	1	2	0	4	8	0.25000
http://www.beginner-java-tutorial.com/java-arraylist.html	C	0	0	1	1	6	8	0.53571
http://www-128.ibm.com/developerworks/linux/library/l-re	D	3	4	3	0	4	14	0.19780
http://www.webinade.com/web-development/creating-recu	D	10	3	0	1	0	14	0.52747
http://thedailywtf.com/forums/thread/89324.aspx	D	2	5	3	2	1	13	0.19231
http://www.oreillynet.com/onjava/blog/2006/03/recursive_	D	7	2	2	0	0	11	0.41818
http://leepoint.net/notes-java/	D	0	2	0	3	6	11	0.34545
http://www.theserverside.net/tt/articles/showarticle.tss?id=	D	10	1	0	0	0	11	0.81818
http://msdn2.microsoft.com/en-us/library/ms186243.aspx	D	10	1	0	0	0	11	0.81818
http://joseph.randomnetworks.com/archives/2005/08/18/p	D	10	1	0	0	0	11	0.81818
http://en.wikipedia.org/wiki/Recursive_descent_parser	D	6	4	1	0	0	11	0.38182

Table14. The means of the extent to which raters agree about a resource in experiment 2 (Cont.)

Experiment 2								
Resource	Question	Number of Subjects Rating					Total Number of Raters	Pi
		0	1	2	3	4		
http://www.cs.princeton.edu/introcs/27recursion/	D	0	0	0	2	8	10	0.64444
http://www.cs.princeton.edu/introcs/23recursion/	D	0	0	1	1	8	10	0.62222
http://www.mssqltips.com/tip.asp?tip=938	D	10	0	0	0	0	10	1.00000
http://www.ibm.com/developerworks/java/library/j-diag8.h	D	0	3	2	2	3	10	0.17778
http://www.cs.dartmouth.edu/farid/teaching/cs15/cs5/lectu	D	2	0	1	2	4	9	0.22222
http://lists.w3.org/Archives/Public/www-dom/1998OctDec/	D	6	1	1	0	1	9	0.41667
http://www.setfocus.com/TechnicalArticles/sql-server-2005-	D	7	1	0	0	0	8	0.75000
http://thedailywtf.com/forums/89353/ShowThread.aspx	D	1	1	3	2	1	8	0.14286
http://www.sitepoint.com/article/hierarchical-data-database	D	6	2	0	0	0	8	0.57143
http://www.cs.ubc.ca/spider/harrison/Java/sorting-demo.ht	E	0	0	0	5	10	15	0.52381
http://www.cs.ubc.ca/~harrison/Java/sorting-demo.html	E	0	0	1	4	10	15	0.48571
http://cg.scs.carleton.ca/~morin/misc/sortalg/	E	0	1	3	1	9	14	0.42857
http://en.wikipedia.org/wiki/Sorting_algorithm	E	0	0	0	2	12	14	0.73626
http://www.cs.rit.edu/~atk/Java/Sorting/sorting.html	E	0	0	2	2	10	14	0.51648
http://critticall.com/	E	10	1	0	1	1	13	0.57692
http://www.answers.com/topic/sorting-algorithm	E	0	1	0	1	11	13	0.70513
http://www.softpanorama.org/Algorithms/sorting.shtml	E	0	1	1	0	11	13	0.70513
http://www.algosort.com/	E	1	2	0	5	4	12	0.25758
http://www.cs.bu.edu/teaching/alg/sort/demo/	E	0	0	2	3	7	12	0.37879
http://www.davekoelle.com/alphanum.html	E	3	0	2	3	3	11	0.18182
http://www.datastructures.info/	E	0	0	0	1	10	11	0.81818
http://www.inf.fh-flensburg.de/lang/algorithmen/sortieren/	E	1	0	0	3	6	10	0.40000
http://www.geocities.com/wezam/sort22.html	E	0	0	0	2	8	10	0.64444
http://dmoz.org/Computers/Algorithms/Sorting_and_Search	E	0	3	0	1	5	9	0.36111
http://www.google.com/Top/Computers/Algorithms/Sorting	E	0	1	2	3	3	9	0.19444
http://www.codinghorror.com/blog/archives/001015.html	E	6	0	2	1	0	9	0.44444
http://www.google.com/alpha/Top/Computers/Algorithms/	E	0	2	1	4	2	9	0.22222
http://www.cs.ubc.ca/spider/harrison/Java/	E	0	0	1	3	4	8	0.32143
http://www.sysarch.com/Perl/sort_paper.html	E	0	0	1	2	5	8	0.39286
http://www2.hig.no/~algmet/animate.html	E	1	0	0	6	1	8	0.53571
http://www.idiotworld.com/story/258/5_algorithms_you_m	E	0	1	0	1	6	8	0.53571
http://java.sun.com/docs/books/tutorial/java/javaOO/acces	F	0	0	0	2	12	14	0.73626
http://www.htmlgoodies.com/primers/jsp/article.php/3600	F	6	2	0	1	0	9	0.44444
http://blog.zerosum.org/2007/11/22/ruby-method-visibility	F	1	1	1	2	4	9	0.19444
http://blog.csdn.net/ladofwind/archive/2006/06/05/774072	F	6	1	0	0	2	9	0.44444
http://members.tripod.com/~MoisesRBB/java3.htm	F	8	0	0	0	1	9	0.77778
http://www.extreme-java.de/junitx	F	7	2	0	0	0	9	0.61111
http://mindprod.com/jgloss/privatescope.html	F	1	0	3	0	5	9	0.36111
http://blog.jonudell.net/2007/03/27/authenticated-rss-feed	F	8	0	0	0	0	8	1.00000
http://www.crockford.com/javascript/private.html	F	4	3	0	1	0	8	0.32143
http://www.uni-bonn.de/~manfear/javaprotection.php	F	1	0	1	0	6	8	0.53571
http://www.jchq.net/tutorial/01_02Tut.htm	F	0	0	0	2	6	8	0.57143
http://www.netbeans.org/	G	0	1	0	3	10	14	0.52747
http://www.jetbrains.com/idea/	G	0	2	2	0	10	14	0.51648
http://springide.org/project	G	0	3	1	3	5	12	0.24242
http://www.eclipse.org/home/categories/languages.php	G	0	1	2	2	6	11	0.30909
http://www.bluej.org/	G	2	4	2	2	1	11	0.16364
http://www.netbeans.org/kb/kb.html	G	2	2	3	0	4	11	0.20000
http://www.myeclipseide.com/	G	0	1	1	2	7	11	0.40000
http://www.netbeans.org/kb/articles/books.html	G	2	4	0	0	3	9	0.27778

Table14. The means of the extent to which raters agree about a resource in experiment 2 (Cont.)

Experiment 2								
Resource	Question	Number of Subjects Rating					Total Number of Raters	Pi
		0	1	2	3	4		
http://www.jetbrains.com/	G	0	0	3	1	5	9	0.36111
http://akamai.infoworld.com/pdf/special_report/2007/13SR	G	0	0	1	1	6	8	0.53571
http://developers.sun.com/prodtech/javatools/jscreator/ind	G	0	1	3	0	4	8	0.32143
http://www.easyeclipse.org/site/home/	G	1	0	0	1	6	8	0.53571
http://today.java.net/pub/a/today/2006/04/06/exception-h	H	0	0	2	4	7	13	0.35897
http://dev2dev.bea.com/pub/a/2006/11/effective-exceptior	H	0	2	3	3	3	11	0.18182
http://www.onjava.com/pub/a/onjava/2003/11/19/exceptic	H	1	0	1	3	5	10	0.28889
http://www.faqs.org/docs/think_java/TIJ311.htm	H	0	0	0	2	7	9	0.61111
http://today.java.net/pub/a/today/2003/12/04/exceptions.h	H	0	0	1	3	5	9	0.36111
http://www.mindview.net/Etc/Discussions/CheckedExceptio	H	0	0	2	3	4	9	0.27778
http://www.javaolympus.com/J2SE/Exceptions/JavaExceptic	H	0	1	1	3	3	8	0.21429
http://www.adtmag.com/java/articleold.aspx?id=1242	H	0	1	0	3	4	8	0.32143
http://ww2.cis.temple.edu/sorkin/ExceptionHandlingJava.ht	H	0	0	1	0	7	8	0.75000
http://java.sun.com/docs/books/tutorial/essential/exceptior	H	0	1	0	2	5	8	0.39286
http://www.artima.com/designtechniques/exceptionsP.html	H	0	0	1	1	6	8	0.53571
http://www.idi.ntnu.no/grupper/su/fordypningsprosjekt-200	H	1	0	0	0	7	8	0.75000
http://www.andreashalter.ch/phpug/20040115/	H	2	4	0	2	0	8	0.28571
							Average	**0.46434**

Table 15. The means of the extent to which raters agree about a resource in experiment 3

Experiment 3								
Resource	Question	Number of Subjects Rating					Total Number of Raters	Pi
		0	1	2	3	4		
http://www.javacoffeebreak.com/java103/java103.html	A	0	0	1	1	10	12	0.68182
http://java.sun.com/docs/books/tutorial/essential/io/	A	6	2	2	0	1	11	0.30909
http://www.idevelopment.info/data/Programming/java/PRC	A	0	2	1	6	2	11	0.30909
http://java.sun.com/developer/onlineTraining/Programming	A	0	1	1	2	6	10	0.35556
http://java.sun.com/j2se/1.4.2/docs/api/java/io/FileOutputS	A	0	0	1	1	7	9	0.58333
http://www.myjavatools.com/projects/v.6.0/lib/doc/com/m	A	1	1	1	4	1	8	0.21429
http://java.sun.com/j2se/1.4.2/docs/api/java/text/SimpleDa	B	0	0	1	1	13	15	0.74286
http://java.sun.com/j2se/1.5.0/docs/api/java/text/SimpleDa	B	0	0	1	1	13	15	0.74286
http://java.sun.com/j2se/1.3/docs/api/java/text/DateForma	B	0	0	1	5	9	15	0.43810
http://java.sun.com/javase/6/docs/api/java/text/SimpleDate	B	0	0	1	3	11	15	0.55238
http://www.kickjava.com/492.htm	B	0	0	1	3	10	14	0.52747
http://java.sun.com/j2se/1.4.2/docs/api/java/text/DateForm	B	0	0	0	5	9	14	0.50549
http://java.boot.by/scjp-tiger/ch03s04.html	B	1	0	0	2	11	14	0.61538
http://www.exampledepot.com/egs/java.text/FormatDate.h	B	0	0	0	2	12	14	0.73626
http://java.sun.com/j2se/1.5.0/docs/api/java/text/DateForm	B	0	0	0	4	10	14	0.56044
http://www.javatechniques.com/public/java/docs/basics/da	B	0	0	0	0	13	13	1.00000
http://www.javatechniques.com/blog/dateformat-and-simpl	B	0	0	0	0	13	13	1.00000
http://java.sun.com/j2se/1.3/docs/api/java/text/SimpleDate	B	0	0	0	2	9	11	0.67273
http://www.java-forums.org/java-tutorials/2775-java-date.h	B	0	0	1	1	7	9	0.58333
http://www.kickjava.com/524.htm	B	0	0	0	0	8	8	1.00000
http://www.beginner-java-tutorial.com/java-date.html	B	0	0	0	0	8	8	1.00000
http://www.daniweb.com/code/snippet515.html	B	0	0	0	0	8	8	1.00000
http://www.unix.org.ua/orelly/java-ent/jnut/ch22_01.htm	B	0	0	2	6	0	8	0.57143
http://www.javaworld.com/jw-12-2000/jw-1229-dates.html	B	0	0	1	1	6	8	0.53571
http://www.java-examples.com/java-collections-and-data-st	C	0	0	0	2	13	15	0.75238
http://www.javadeveloper.co.in/java-example/java-arraylist	C	0	0	0	0	14	14	1.00000
http://www.kickjava.com/220.htm	C	0	0	1	3	10	14	0.52747

Table15. The means of the extent to which raters agree about a resource in experiment 3 (Cont.)

Experiment 3								
Resource	Question	Number of Subjects Rating					Total Number of Raters	Pi
		0	1	2	3	4		
http://www.javafaq.nu/java-example-code-89.html	C	1	0	1	0	11	13	0.70513
http://www.anyexample.com/programming/java/java_array	C	0	0	2	1	10	13	0.58974
http://www.javabeginner.com/java-arraylist.htm	C	0	0	0	2	10	12	0.69697
http://www.java-samples.com/showtutorial.php?tutorialid=	C	0	0	0	2	10	12	0.69697
http://www.beginner-java-tutorial.com/java-arraylist.html	C	0	0	0	3	9	12	0.59091
http://java.sun.com/developer/JDCTechTips/2002/tt0910.ht	C	1	1	2	4	4	12	0.19697
http://eclipsetutorial.sourceforge.net/Total_Beginner_Comp	C	0	0	2	6	4	12	0.33333
http://www.java2s.com/Code/JavaAPI/java.util/ArrayListiter	C	0	0	0	3	9	12	0.59091
http://www.wellho.net/resources/ex.php4?item=j714/Arlist	C	1	0	1	3	7	12	0.36364
http://java.sun.com/developer/technicalArticles/releases/j2	C	2	5	3	0	2	12	0.22727
http://www.developerzone.biz/index.php?option=com_cont	C	0	0	1	1	8	10	0.62222
http://java.sun.com/j2se/1.4.2/docs/api/java/util/ArrayList.h	C	1	0	1	3	4	9	0.25000
http://www.idevelopment.info/data/Programming/java/PRO	C	0	0	1	1	7	9	0.58333
http://users.cs.dal.ca/~sedgwick/ArrayList.html	C	0	0	1	2	6	9	0.44444
http://www.javafaq.nu/java-example-code-list.html	C	0	0	0	3	6	9	0.50000
http://www.leepoint.net/notes-java/data/collections/lists/a	C	0	0	1	4	4	9	0.33333
http://www.idevelopment.info/data/Programming/java/coll	C	0	0	1	1	6	8	0.53571
http://www.leepoint.net/notes-java/	C	0	0	0	3	5	8	0.46429
http://danzig.jct.ac.il/java_class/recursion.html	D	1	0	0	4	10	15	0.48571
http://www.thedailywtf.com/forums/thread/89324.aspx	D	3	9	2	1	0	15	0.38095
http://www.thedailywtf.com/forums/89353/ShowThread.as	D	5	6	3	1	0	15	0.26667
http://www.ibm.com/developerworks/java/library/j-diag8.h	D	2	1	3	2	5	13	0.19231
http://www.ahmadsoft.org/articles/recursion/	D	1	0	1	2	7	11	0.40000
http://weblogs.java.net/blog/mortazavi/archive/2005/08/re	D	0	0	2	3	6	11	0.34545
http://chortle.ccsu.edu/CS151/cs151java.html	D	1	1	0	4	5	11	0.29091
http://www.tech-recipes.com/java_programming_tips1266.l	D	0	0	0	3	7	10	0.53333
http://www.cs.princeton.edu/introcs/23recursion/	D	0	0	0	0	10	10	1.00000
http://www-128.ibm.com/developerworks/java/library/j-dia	D	1	1	0	3	5	10	0.28889
http://www.cs.princeton.edu/introcs/27recursion/	D	1	0	0	0	9	10	0.80000
http://en.wikipedia.org/wiki/Recursion_(computer_science)	D	0	2	1	1	4	8	0.25000
http://www-128.ibm.com/developerworks/linux/library/l-re	D	1	0	1	1	5	8	0.35714
http://www.cs.dartmouth.edu/farid/teaching/cs15/cs5/lectu	D	1	0	0	1	6	8	0.53571
http://www.cs.ubc.ca/spider/harrison/Java/sorting-demo.ht	E	0	0	1	5	9	15	0.43810
http://www.cs.rit.edu/~atk/Java/Sorting/sorting.html	E	0	0	2	7	6	15	0.35238
http://cg.scs.carleton.ca/~morin/misc/sortalg/	E	0	0	1	4	9	14	0.46154
http://www.cs.ubc.ca/~harrison/Java/sorting-demo.html	E	0	1	0	6	7	14	0.39560
http://en.wikipedia.org/wiki/Sorting_algorithm	E	0	0	0	0	14	14	1.00000
http://www.softpanorama.org/Algorithms/sorting.shtml	E	0	0	3	5	6	14	0.30769
http://www.answers.com/topic/sorting-algorithm	E	1	0	2	0	11	14	0.61538
http://www.davekoelle.com/alphanum.html	E	2	3	1	7	0	13	0.32051
http://www.datastructures.info/	E	0	0	0	7	6	13	0.46154
http://www.algosort.com/	E	0	1	0	6	5	12	0.37879
http://www.google.com/Top/Computers/Algorithms/Sorting	E	0	0	0	3	9	12	0.59091
http://www.google.com/alpha/Top/Computers/Algorithms/	E	0	1	1	0	10	12	0.68182
http://www.dmoz.org/Computers/Algorithms/Sorting_and_	E	0	0	2	4	6	12	0.33333
http://linux.wku.edu/~lamonml/algor/sort/sort.html	E	0	0	1	1	10	12	0.68182
http://www.ddj.com/dept/cpp/184402000	E	2	2	1	5	1	11	0.21818
http://www.cs.ubc.ca/spider/harrison/Java/	E	0	0	1	4	6	11	0.38182
http://www.geocities.com/wezam/sort22.html	E	2	1	1	6	1	11	0.29091

Table15. The means of the extent to which raters agree about a resource in experiment 3 (Cont.)

Experiment 3								
Resource	Question	Number of Subjects Rating					Total Number of Raters	Pi
		0	1	2	3	4		
http://www.coyotesong.com/sort/	E	0	0	1	2	8	11	0.52727
http://maven.smith.edu/~thiebaut/java/sort/demo.html	E	0	1	5	2	2	10	0.26667
http://linux.wku.edu/~lamonml/algor/sort/	E	0	0	0	0	9	9	1.00000
http://max.cs.kzoo.edu/~abrady/java/sorting/	E	0	1	2	1	5	9	0.30556
http://www.idiotworld.com/story/258/5_algorithms_you_m	E	0	0	0	7	2	9	0.61111
http://www2.hig.no/~algmet/animate.html	E	0	1	1	5	2	9	0.30556
http://www.cs.bu.edu/teaching/alg/sort/demo/	E	0	0	2	5	2	9	0.33333
http://www.cs.hope.edu/alganim/ccaa/sorting.html	E	0	2	0	3	3	8	0.25000
http://atschool.eduweb.co.uk/mbaker/sorts.html	E	0	0	0	2	6	8	0.57143
http://math.hws.edu/TMCM/java/labs/xSortLabLab.html	E	0	0	1	0	7	8	0.75000
http://www.codestyle.org/java/faq-Inheritance.shtml	F	0	1	2	5	7	15	0.30476
http://www.allapplabs.com/interview_questions/java_interv	F	0	0	1	2	11	14	0.61538
http://java.sun.com/j2se/1.5.0/docs/api/java/lang/reflect/M	F	1	3	1	3	4	12	0.18182
http://www.jchq.net/tutorial/01_02Tut.htm	F	0	1	2	2	6	11	0.30909
http://java.sun.com/docs/books/tutorial/java/javaOO/acces	F	0	0	0	2	8	10	0.64444
http://www.daimi.au.dk/dRegAut/JavaBNF.html	F	5	2	1	0	0	8	0.39286
http://www.unf.edu/~rzucker/cop3540dir/modifiers.html	F	0	0	0	0	8	8	1.00000
http://www.eclipse.org/home/categories/languages.php	G	0	2	1	7	4	14	0.30769
http://www.netbeans.org/	G	0	2	0	5	6	13	0.33333
http://www.bluej.org/	G	3	2	1	2	5	13	0.19231
http://www.easyeclipse.org/site/home/	G	0	1	2	8	1	12	0.43939
http://java.sun.com/developer/technicalArticles/tools/intro.	G	0	0	2	3	6	11	0.34545
http://www.jcreator.com/	G	0	2	2	3	4	11	0.20000
http://www.jetbrains.com/idea/	G	0	2	1	3	5	11	0.25455
http://www.jetbrains.com/	G	0	0	3	1	6	10	0.40000
http://hossamahmed.wordpress.com/2006/09/13/java-idee	G	0	0	0	0	10	10	1.00000
http://www.netbeans.org/kb/articles/import-jbuilder.html	G	1	1	1	4	3	10	0.20000
http://www.netbeans.org/kb/articles/books.html	G	2	5	0	2	1	10	0.26667
http://javaboutique.internet.com/demoIDEs/	G	0	0	0	0	10	10	1.00000
http://developers.sun.com/prodtech/javatools/jscreator/	G	0	1	1	4	4	10	0.26667
http://www.myeclipseide.com/	G	0	3	2	1	4	10	0.22222
http://www.apl.jhu.edu/~hall/java/IDEs.html	G	0	0	0	1	9	10	0.80000
http://www.springide.org/project	G	0	3	1	1	4	9	0.25000
http://www.stylusstudio.com/java_ide.html	G	0	0	2	4	3	9	0.27778
http://www.easyeclipse.org/site/distributions/	G	0	4	1	2	2	9	0.22222
http://www.netbeans.org/kb/kb.html	G	2	1	2	3	1	9	0.13889
http://today.java.net/pub/a/today/2006/04/06/exception-h	H	1	1	1	8	3	14	0.34066
http://ww2.cis.temple.edu/sorkin/ExceptionHandlingJava.ht	H	0	0	0	1	11	12	0.83333
http://pages.cs.wisc.edu/~cs302/io/Exceptions.html	H	0	0	0	1	11	12	0.83333
http://www.artima.com/designtechniques/exceptionsP.html	H	0	0	0	2	9	11	0.67273
http://neptune.netcomp.monash.edu.au/JavaHelp/howto/tr	H	0	0	0	2	9	11	0.67273
http://java.sun.com/docs/books/tutorial/essential/exceptior	H	0	0	1	1	8	10	0.62222
http://sharat.wordpress.com/2007/05/16/exception-drill/	H	0	0	1	4	4	9	0.33333
http://www.smartdataprocessing.com/lessons/l5.htm	H	0	1	0	0	8	9	0.77778
http://www.faqs.org/docs/think_java/TIJ311.htm	H	0	0	0	2	6	8	0.57143
http://www.janeg.ca/scjp/flow/try.html	H	0	1	1	1	5	8	0.35714
http://www.sys-con.com/story/?storyid=38160	H	1	0	0	1	6	8	0.53571
http://today.java.net/pub/a/today/2003/12/04/exceptions.h	H	0	0	2	0	6	8	0.57143
http://www.javalobby.org/java/forums/t105307.html	H	1	0	1	0	6	8	0.53571
							Average	**0.50739**

Table 16. The means of the extent to which raters agree about a resource for each question in all three experiments

Question	Means of the extent to which raters agree about a resource		
	Experiment 1	Experiment 2	Experiment 3
A	0.35857	0.44525	0.40886
B	0.58333	0.53841	0.71025
C	0.61966	0.41558	0.52405
D	0.54880	0.50362	0.43765
E	0.68182	0.47122	0.47523
F	0.36111	0.54529	0.49262
G	0.16468	0.36590	0.37459
H	0.26984	0.40998	0.58904

From table 16, the extent to which raters agree about resource seems to be moderate for all questions except question G. Question G asked the subjects to find web pages about features of the various Integrated Development Environment (IDE) for Java that currently in the market. During the experiments, the subjects were observed and asked to explain the criteria they used to judge the search results. Some subjects explained that they were satisfied with search results that contained content about features of a IDE. On the other hand, some subjects were looking for search results that provided information about features of difference Java IDEs. As a result, the ratings from subjects in all three experiments for question G showed low agreement.

4.5 RESULT OF EXPERIMENTS

4.5.1 Experiment 1: Re-ranking Search Results

As mentioned in the previous chapter, social annotations were used to adjust the ranking of search results returned from Google. For any given resource, it is possible to determine how

many users bookmarked a given resource, how many tags were used to describe a given resource, and how many times each tag was used.

Six social annotation based ranking methods have been examined and compared with Google. The Popularity Count (PC) is the simplest method. The ranking of search results can be done by ordering results in terms of the number of people bookmarking the web resources. The Propagated Popularity Count (PPC) is a Popularity Count of a web resource determined from the Popularity Count of the main page of the domain of that web resource. This method was used when a given web page retrieved in a search was not found in the mined data. The Query Weighted Popularity Count (QWPC) and Query Weighted Propagated Popularity Count (QWPPC) are the PC or PPC of a given resource weighted by the count of shared terms between query string and the tags on the web resource. The QWPC and QWPPC consider the number of tag that matched query terms as well as the number of people bookmarking the web resource. . The Matched Tag Count (MTC) would be an unbounded number that would sum the total number of users that used tags that matched terms in the query string. The Normalized Matched Tag Count (NMTC) takes the total count of all tag for a given resource into consideration. The following figure shows the comparison of NDCG for Google and social annotation based ranking methods.

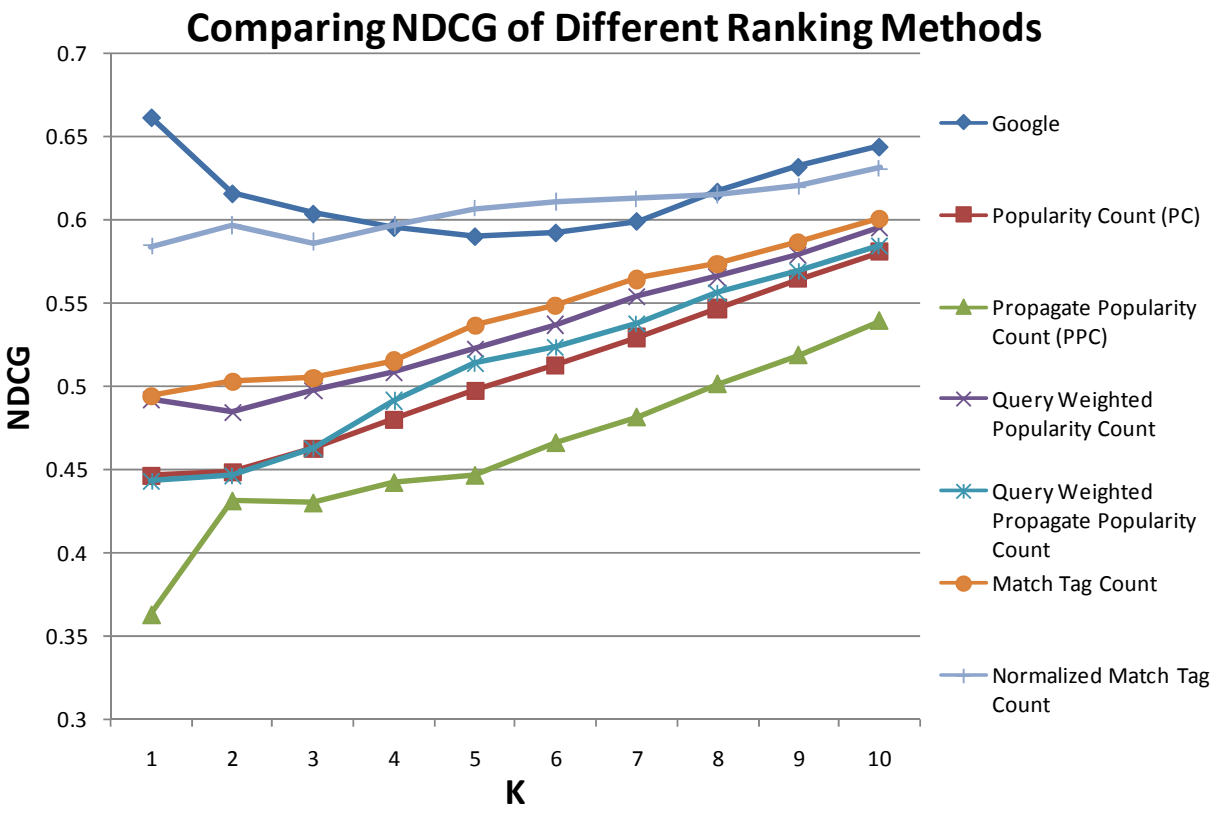


Figure 52. Comparing NDCG of different ranking methods

H_0: There is no statistical difference between the means of the NDCG at K=10 of the Google ranking and the social annotation based rankings.

$(\mu_{google} = \mu_{pc} = \mu_{ppc} = \mu_{qwpc} = \mu_{qwppc} = \mu_{mtc} = \mu_{nmtc})$

H_1: Not all approaches are equal $(\mu_{google} \neq \mu_{pc} \neq \mu_{ppc} \neq \mu_{qwpc} \neq \mu_{qwppc} \neq \mu_{mtc} \neq \mu_{nmtc})$

One-way Analysis of Variance (ANOVA) was applied to test the hypothesis. The following figure shows the results.

Descriptives

NDCG at K=10

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Google	120	.6437169	.14730868	.01344738	.6170898	.6703441	.27128	1.00000
PC	120	.5800970	.14431960	.01317452	.5540101	.6061839	.21049	.93953
PPC	120	.5389828	.15353045	.01401535	.5112310	.5667346	.12433	.91153
QWPC	120	.5950730	.15518442	.01416633	.5670223	.6231238	.12003	.96442
QWPPC	120	.5841603	.16340203	.01491650	.5546242	.6136965	.15165	.96442
MTC	120	.6003971	.15425413	.01408141	.5725145	.6282797	.11052	.96442
NMTC	120	.6306589	.16700292	.01524521	.6004719	.6608460	.13502	.98975
Total	840	.5961552	.15789309	.00544783	.5854622	.6068481	.11052	1.00000

Test of Homogeneity of Variances

NDCG at K=10

Levene Statistic	df1	df2	Sig.
.859	6	833	.525

ANOVA

NDCG at K=10

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.857	6	.143	5.932	.000
Within Groups	20.059	833	.024		
Total	20.916	839			

Figure 53. Result of Hypothesis # 1 Testing

From the above figure, the null hypothesis is rejected. There is an evidence that not all of the means of the NDCG at K=10 of ranking methods are equal at $\alpha = .05$ level of significant. In other word, the rankings of search results obtained from Google and Social Annotation Based Ranking are statistically significant different. Then, the multiple comparisons were performed to find which rankings are different.

Figure 54 and Figure 55 show the results of multiple comparisons examined with Turkey, and Scheffe procedure respectively. Figure 56 shows homogeneous subsets of all pairwise differences of Newman-Keuls, Tukey, and Scheffe procedure.

Multiple Comparisons

NDCG at K=10
Tukey HSD

(I) Ranking Method	(J) Ranking Method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Google	PC	.06361992*	.02003368	.026	.0044092	.1228306
	PPC	.10473414*	.02003368	.000	.0455234	.1639449
	QWPC	.04864386	.02003368	.188	-.0105669	.1078546
	QWPPC	.05955656*	.02003368	.048	.0003458	.1187673
	MTC	.04331979	.02003368	.317	-.0158909	.1025305
	NMTC	.01305797	.02003368	.995	-.0461528	.0722687
PC	Google	-.06361992*	.02003368	.026	-.1228306	-.0044092
	PPC	.04111423	.02003368	.383	-.0180965	.1003250
	QWPC	-.01497605	.02003368	.989	-.0741868	.0442347
	QWPPC	-.00406335	.02003368	1.000	-.0632741	.0551474
	MTC	-.02030013	.02003368	.951	-.0795109	.0389106
	NMTC	-.05056195	.02003368	.152	-.1097727	.0086488
PPC	Google	-.10473414*	.02003368	.000	-.1639449	-.0455234
	PC	-.04111423	.02003368	.383	-.1003250	.0180965
	QWPC	-.05609028	.02003368	.077	-.1153010	.0031204
	QWPPC	-.04517758	.02003368	.267	-.1043883	.0140331
	MTC	-.06141435*	.02003368	.036	-.1206251	-.0022036
	NMTC	-.09167617*	.02003368	.000	-.1508869	-.0324654
QWPC	Google	-.04864386	.02003368	.188	-.1078546	.0105669
	PC	.01497605	.02003368	.989	-.0442347	.0741868
	PPC	.05609028	.02003368	.077	-.0031204	.1153010
	QWPPC	.01091270	.02003368	.998	-.0482980	.0701234
	MTC	-.00532408	.02003368	1.000	-.0645348	.0538867
	NMTC	-.03558589	.02003368	.565	-.0947966	.0236248
QWPPC	Google	-.05955656*	.02003368	.048	-.1187673	-.0003458
	PC	.00406335	.02003368	1.000	-.0551474	.0632741
	PPC	.04517758	.02003368	.267	-.0140331	.1043883
	QWPC	-.01091270	.02003368	.998	-.0701234	.0482980
	MTC	-.01623677	.02003368	.984	-.0754475	.0429740
	NMTC	-.04649859	.02003368	.235	-.1057093	.0127121
MTC	Google	-.04331979	.02003368	.317	-.1025305	.0158909
	PC	.02030013	.02003368	.951	-.0389106	.0795109
	PPC	.06141435*	.02003368	.036	.0022036	.1206251
	QWPC	.00532408	.02003368	1.000	-.0538867	.0645348
	QWPPC	.01623677	.02003368	.984	-.0429740	.0754475
	NMTC	-.03026182	.02003368	.739	-.0894725	.0289489
NMTC	Google	-.01305797	.02003368	.995	-.0722687	.0461528
	PC	.05056195	.02003368	.152	-.0086488	.1097727
	PPC	.09167617*	.02003368	.000	.0324654	.1508869
	QWPC	.03558589	.02003368	.565	-.0236248	.0947966
	QWPPC	.04649859	.02003368	.235	-.0127121	.1057093
	MTC	.03026182	.02003368	.739	-.0289489	.0894725

*. The mean difference is significant at the 0.05 level.

Figure 54. Multiple comparisons of all ranking methods by Tukey procedure

Multiple Comparisons

NDCG at K=10
Scheffe

(I) Ranking Method	(J) Ranking Method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Google	PC	.06361992	.02003368	.123	-.0076524	.1348922
	PPC	.10473414*	.02003368	.000	.0334619	.1760064
	QWPC	.04864386	.02003368	.436	-.0226284	.1199161
	QWPPC	.05955656	.02003368	.184	-.0117157	.1308288
	MTC	.04331979	.02003368	.586	-.0279525	.1145921
	NMTC	.01305797	.02003368	.999	-.0582143	.0843303
PC	Google	-.06361992	.02003368	.123	-.1348922	.0076524
	PPC	.04111423	.02003368	.648	-.0301581	.1123865
	QWPC	-.01497605	.02003368	.997	-.0862483	.0562962
	QWPPC	-.00406335	.02003368	1.000	-.0753356	.0672089
	MTC	-.02030013	.02003368	.985	-.0915724	.0509722
	NMTC	-.05056195	.02003368	.384	-.1218342	.0207103
PPC	Google	-.10473414*	.02003368	.000	-.1760064	-.0334619
	PC	-.04111423	.02003368	.648	-.1123865	.0301581
	QWPC	-.05609028	.02003368	.252	-.1273626	.0151820
	QWPPC	-.04517758	.02003368	.533	-.1164499	.0260947
	MTC	-.06141435	.02003368	.154	-.1326866	.0098579
	NMTC	-.09167617*	.02003368	.002	-.1629485	-.0204039
QWPC	Google	-.04864386	.02003368	.436	-.1199161	.0226284
	PC	.01497605	.02003368	.997	-.0562962	.0862483
	PPC	.05609028	.02003368	.252	-.0151820	.1273626
	QWPPC	.01091270	.02003368	1.000	-.0603596	.0821850
	MTC	-.00532408	.02003368	1.000	-.0765964	.0659482
	NMTC	-.03558589	.02003368	.789	-.1068582	.0356864
QWPPC	Google	-.05955656	.02003368	.184	-.1308288	.0117157
	PC	.00406335	.02003368	1.000	-.0672089	.0753356
	PPC	.04517758	.02003368	.533	-.0260947	.1164499
	QWPC	-.01091270	.02003368	1.000	-.0821850	.0603596
	MTC	-.01623677	.02003368	.995	-.0875091	.0550355
	NMTC	-.04649859	.02003368	.496	-.1177709	.0247737
MTC	Google	-.04331979	.02003368	.586	-.1145921	.0279525
	PC	.02030013	.02003368	.985	-.0509722	.0915724
	PPC	.06141435	.02003368	.154	-.0098579	.1326866
	QWPC	.00532408	.02003368	1.000	-.0659482	.0765964
	QWPPC	.01623677	.02003368	.995	-.0550355	.0875091
	NMTC	-.03026182	.02003368	.892	-.1015341	.0410105
NMTC	Google	-.01305797	.02003368	.999	-.0843303	.0582143
	PC	.05056195	.02003368	.384	-.0207103	.1218342
	PPC	.09167617*	.02003368	.002	.0204039	.1629485
	QWPC	.03558589	.02003368	.789	-.0356864	.1068582
	QWPPC	.04649859	.02003368	.496	-.0247737	.1177709
	MTC	.03026182	.02003368	.892	-.0410105	.1015341

*. The mean difference is significant at the 0.05 level.

Figure 55. Multiple comparisons of all ranking methods by Scheffe procedure

Homogeneous Subsets

NDCG at K=10

	Ranking Method	N	Subset for alpha = 0.05		
			1	2	3
Student-Newman-Keuls[a]	PPC	120	.5389828		
	PC	120	.5800970	.5800970	
	QWPPC	120	.5841603	.5841603	
	QWPC	120		.5950730	.5950730
	MTC	120		.6003971	.6003971
	NMTC	120		.6306589	.6306589
	Google	120			.6437169
	Sig.		.063	.086	.073
Tukey HSD[a]	PPC	120	.5389828		
	PC	120	.5800970	.5800970	
	QWPPC	120	.5841603	.5841603	
	QWPC	120	.5950730	.5950730	.5950730
	MTC	120		.6003971	.6003971
	NMTC	120		.6306589	.6306589
	Google	120			.6437169
	Sig.		.077	.152	.188
Scheffe[a]	PPC	120	.5389828		
	PC	120	.5800970	.5800970	
	QWPPC	120	.5841603	.5841603	
	QWPC	120	.5950730	.5950730	
	MTC	120	.6003971	.6003971	
	NMTC	120		.6306589	
	Google	120		.6437169	
	Sig.		.154	.123	

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 120.000.

Figure 56 Homogeneous subsets of all pairwise differences of Newman-Keuls, Tukey, and Scheffe procedure

Figure 52 suggests that the Normalized Match Tag Count provided the best ranking among the annotation based ranking methods. It also provided a similar quality ranking when compared with Google ranking. The results from the multiple comparison confirmed that the ranking provided by Normalized Match Tag Count is not significantly different from the ranking provided by Google. The results also suggested that the ranking provided by Match Tag Count and Query Weighted Popularity Count are not significantly different from the ranking provided by Google.

In this experiment, the social annotation based ranking methods were applied to a set of search results obtained from Google. Given the possibility that there might not be any social annotation on some search results obtained from Google, these results would be ordered at the bottom of the rank. If there was more than one search result for which there was not a social annotation ranking, the order obtained from Google was preserved.

Although the results from the multiple comparisons Normalized Match Tag Count and Query Weighted Popularity Count are not statistically different from the ranking provided by Google, by considering only the comparison of NDCG as shown in Figure 52, it suggested that the Normalized Match Tag Count provided a similar quality of ranking obtained from Google.

4.5.2 Experiment 2: Resource Indexing Augmentation

In this experiment, social annotations were used to augment resource indexing. The experiment investigated how well social annotations contribute to obtaining relevant search results when they are used as resource indexes. Three indexing approaches were examined. The base case was full text indexing using only the content of the resources. Annotation indexing used only social annotations and their frequency as resource indexes. Full text with annotations indexing used both content and social annotations and their frequency as resource indexes.

Indri, a language model search engine, was used to create resource indexes and compute similarity between submitted queries and resource indexes. The following figure shows the comparison of NDCG of the different indexing approaches. It should be noted that this experiment only compared retrieval sets from the mined resources that had been collected from Delicious

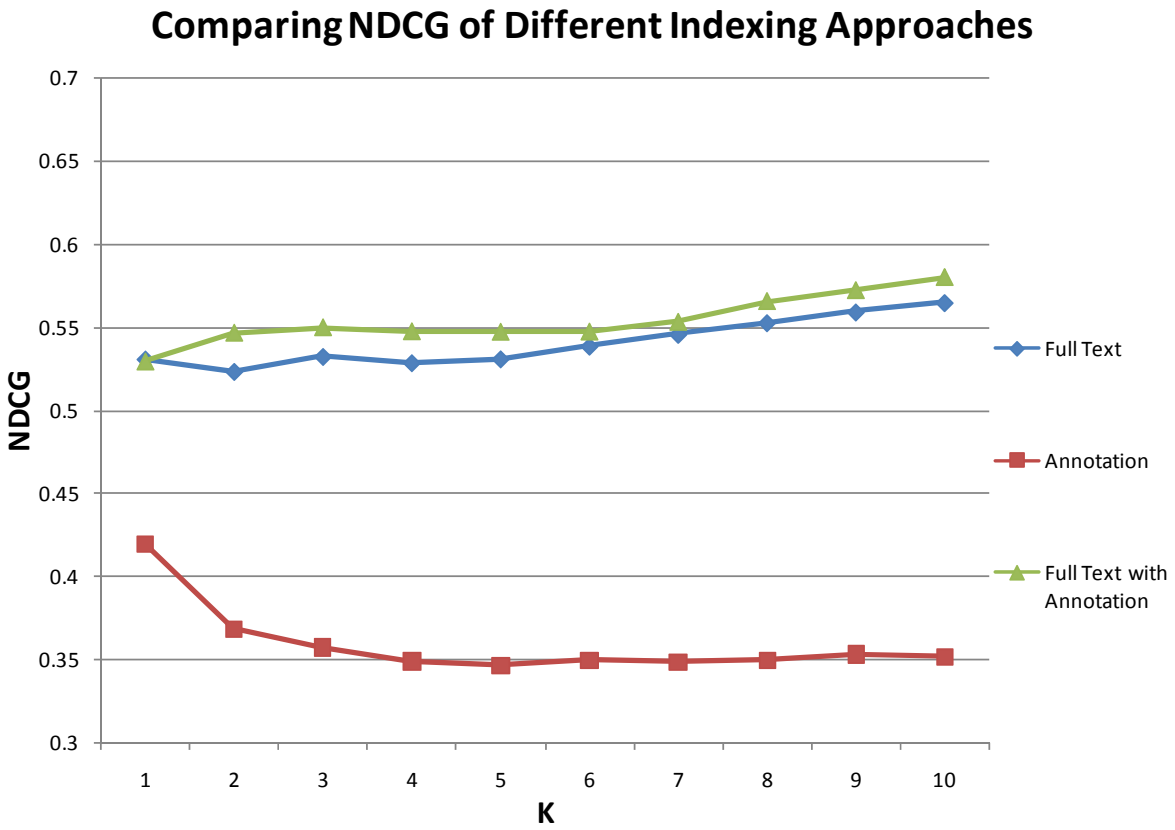


Figure 57. Comparing NDCG of different indexing approaches

H_0: There is no statistical difference between the means of the NDCG at K=10 of from the full text indexing, annotation indexing and annotation with full text indexing approach $(\mu_{Full-Text} = \mu_{Annotation} = \mu_{Full-TextWithAnnotationt})$

H_1: Not all approaches are equal $(\mu_{Full-Text} \neq \mu_{Annotation} \neq \mu_{Full-TextWithAnnotationt})$

One-way Analysis of Variance (ANOVA) was applied to test the hypothesis. The following figure shows the result of hypothesis testing.

Descriptives

NDCG K=10

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Full-Text Indexing	120	.5644773	.23840840	.02176361	.5213832	.6075714	.00000	1.00000
Annotation Indexing	120	.3515879	.25706965	.02346714	.3051206	.3980552	.00000	.95459
Full-Text with Annotation Indexing	120	.5798930	.24181944	.02207499	.5361824	.6236037	.01452	1.00000
Total	360	.4986528	.26648392	.01404494	.4710321	.5262734	.00000	1.00000

Test of Homogeneity of Variances

NDCG K=10

Levene Statistic	df1	df2	Sig.
.625	2	357	.536

ANOVA

NDCG K=10

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.907	2	1.954	32.310	.000
Within Groups	21.587	357	.060		
Total	25.494	359			

Figure 58. Result of Hypothesis # 2 Testing

From the above figure, the null hypothesis is rejected. There is evidence that not all of the means of the NDCG at K=10 of indexing approaches are equal at $\alpha = .05$ level of significance. In others words, the difference in the set of search results returned from three indexing approaches are statistically significant. The multiple comparisons were then performed to find which indexing approaches are different.

Figure 59 shows the results of multiple comparisons examined with Turkey, and Scheffe procedure. Figure 60 shows homogeneous subsets of all pairwise differences of Newman-Keuls, Tukey, and Scheffe procedure.

Post Hoc Tests

Multiple Comparisons

Dependent Variable:NDCG K=10

	(I) Indexing Method	(J) Indexing Method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Full-Text Indexing	Annotation Indexing	.21288934*	.03174552	.000	.1381753	.2876034
		Full-Text with Annotation Indexing	-.01541577	.03174552	.878	-.0901298	.0592983
	Annotation Indexing	Full-Text Indexing	-.21288934*	.03174552	.000	-.2876034	-.1381753
		Full-Text with Annotation Indexing	-.22830511*	.03174552	.000	-.3030192	-.1535911
	Full-Text with Annotation Indexing	Full-Text Indexing	.01541577	.03174552	.878	-.0592983	.0901298
		Annotation Indexing	.22830511*	.03174552	.000	.1535911	.3030192
Scheffe	Full-Text Indexing	Annotation Indexing	.21288934*	.03174552	.000	.1348572	.2909215
		Full-Text with Annotation Indexing	-.01541577	.03174552	.889	-.0934479	.0626164
	Annotation Indexing	Full-Text Indexing	-.21288934*	.03174552	.000	-.2909215	-.1348572
		Full-Text with Annotation Indexing	-.22830511*	.03174552	.000	-.3063373	-.1502730
	Full-Text with Annotation Indexing	Full-Text Indexing	.01541577	.03174552	.889	-.0626164	.0934479
		Annotation Indexing	.22830511*	.03174552	.000	.1502730	.3063373

*. The mean difference is significant at the 0.05 level.

Figure 59. Multiple comparisons of all indexing approaches by Tukey and Scheffe procedure

Homogeneous Subsets

NDCG K=10

	Indexing Method	N	Subset for alpha = 0.05	
			1	2
Student-Newman-Keuls[a]	Annotation Indexing	120	.3515879	
	Full-Text Indexing	120		.5644773
	Full-Text with Annotation Indexing	120		.5798930
	Sig.		1.000	.628
Tukey HSD[a]	Annotation Indexing	120	.3515879	
	Full-Text Indexing	120		.5644773
	Full-Text with Annotation Indexing	120		.5798930
	Sig.		1.000	.878
Scheffe[a]	Annotation Indexing	120	.3515879	
	Full-Text Indexing	120		.5644773
	Full-Text with Annotation Indexing	120		.5798930
	Sig.		1.000	.889

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 120.000.

Figure 60. Homogeneous subsets of all indexing methods by Newman-Keuls, Tukey, and Scheffe procedure

Figure 57 suggests that the Full-Text with Annotation indexing approach provided a better set of search results compared with the Full-Test indexing approach and the Annotation

indexing approach. However, the results from the multiple comparisons showed that a set of search results provided by the Full-Text with Annotation indexing approach is not statistically different from a set of search results provided by the Full-Text indexing approach.

4.5.3 Experiment 3: Resource Retrieval and Ranking of Search Results

In this experiment, the resource indexing approach that showed the best performance from the second experiment was selected as the indexing approach. The Annotation based ranking method that showed the best performance from the first experiment is selected as the ranking method.

Although the results from the multiple comparisons showed that a set of search results provided by the Full-Text with Annotation indexing approach is not statistically significantly different from a set of search results provided by the Full-Text indexing approach, the Full-Text with Annotation indexing approach was selected to be an indexing approach in this experiment.

The results from the first experiment suggested that the ranking provided by Normalized Math Tag Count, Match Tag Count and Query Weighted Popularity Count are not significantly different from the ranking provided by Google. In this experiment, they were applied to the set of search results obtained from the Full-Text with Annotation indexing approach. The following figure shows the comparison of NDCG of Google and Full-Text Indexing with different annotation based ranking method.

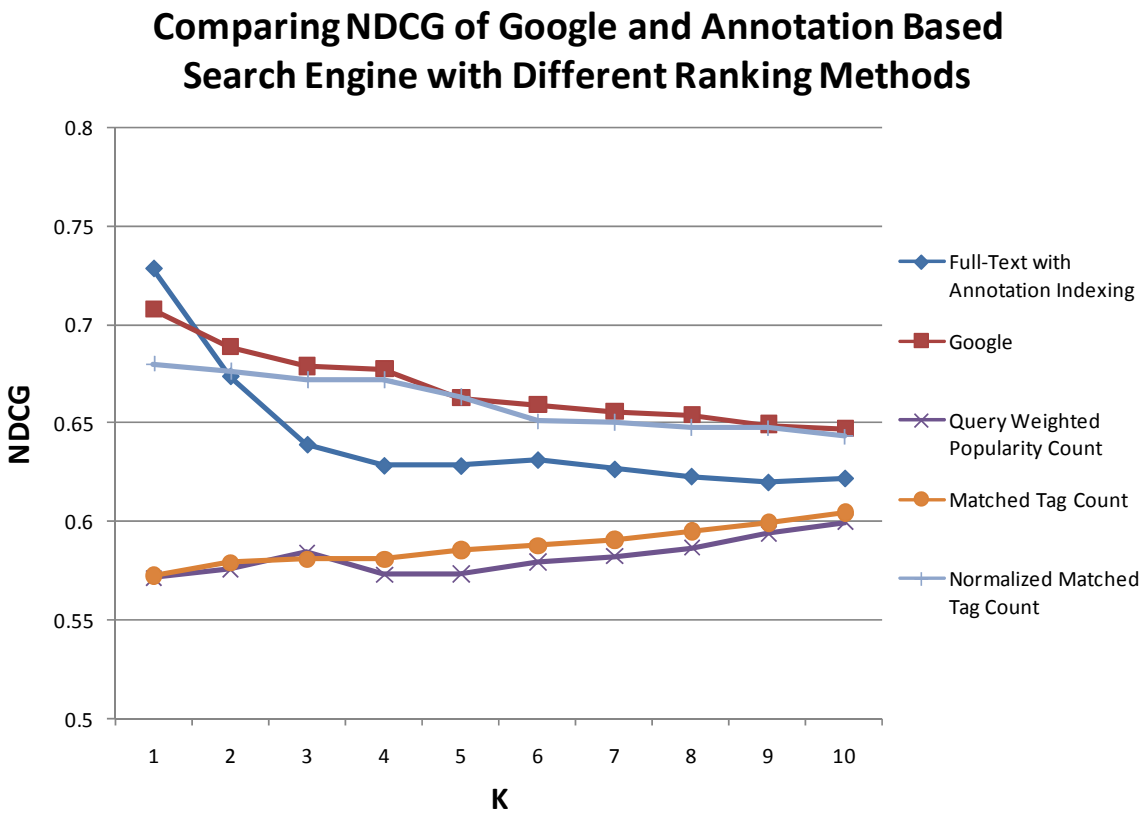


Figure 61. Comparing NDCG of Google and annotation bases search engines with different ranking methods

Figure 61 showed that the Full-Text with Annotation indexing, together with Normalized Match Tag Count provided a similar quality ranking compared with Google.

To exploit the social annotations for web searching, it is interesting to integrate both Full-Text with Annotation indexing and Normalized Matched Tag Count to rank the resources. They may provide better ranking. The following is the formula for combining ranking score from both the similarity ranking score and the annotation based ranking score.

$$SimRank(Q,p) = \alpha\left(\frac{Sim(Q,p)}{MAX\{Sim(Q,p_1),...,Sim(Q,p_n)\}}\right) + (1-\alpha)\left(\frac{PR_{Ann}(p)}{MAX\{PR_{Ann}(p_1),...,PR_{Ann}(p_n)\}}\right)$$

where α is the weight of ranking technique

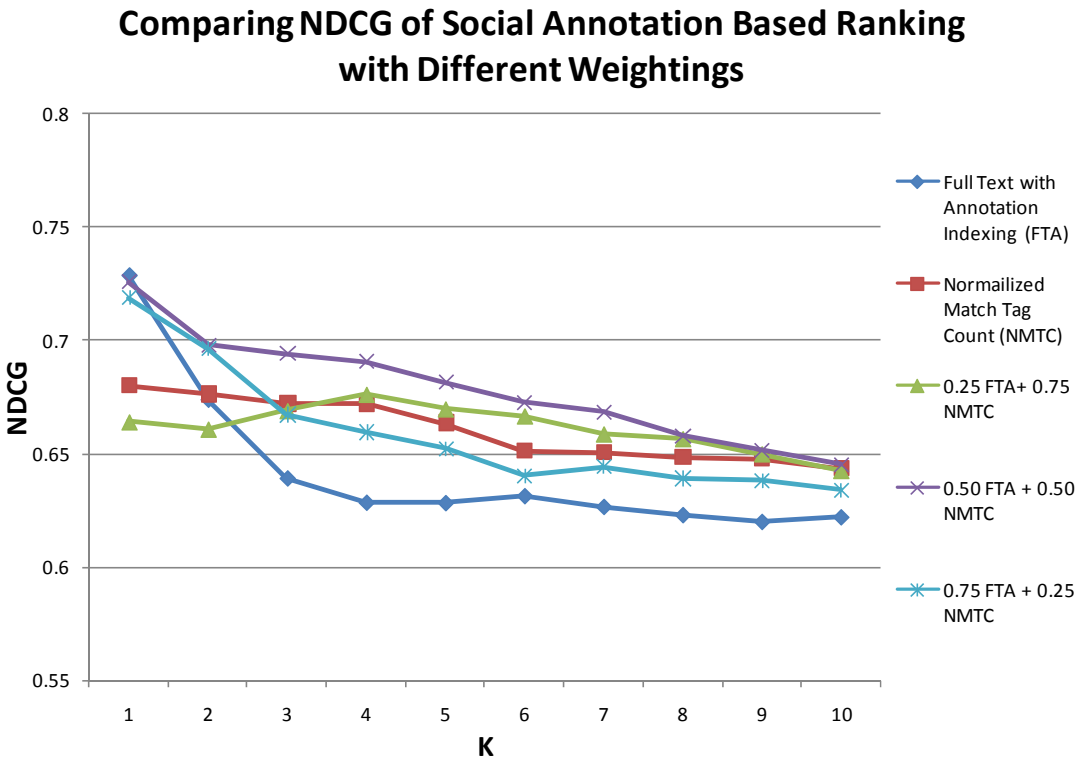


Figure 62. Comparing NDCG of social annotation based ranking with different weightings

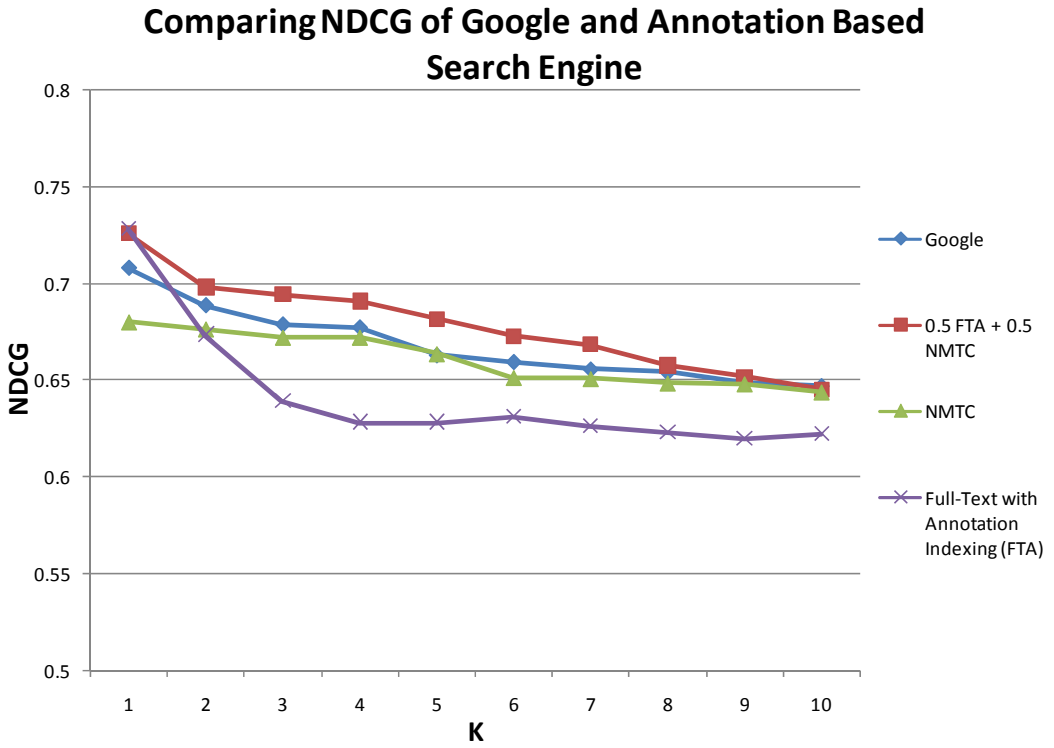


Figure 63. Comparing NDCG of Google and annotation based search engine

H_0: There is no statistical difference between the means of the NDCG at K=10 of the ranking returned from Google and the combination both similarity score and annotation based ranking score with $\alpha = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. $(\mu_{google} = \mu_{\alpha=0} = \mu_{\alpha=\frac{1}{4}} = \mu_{\alpha=\frac{1}{2}} = \mu_{\alpha=\frac{3}{4}} = \mu_{\alpha=1})$

H_1: Not all approaches are equal. $(\mu_{google} \neq \mu_{\alpha=0} \neq \mu_{\alpha=\frac{1}{4}} \neq \mu_{\alpha=\frac{1}{2}} \neq \mu_{\alpha=\frac{3}{4}} \neq \mu_{\alpha=1})$

One-way Analysis of Variance (ANOVA) was applied to test the hypothesis. The following figure shows the result of hypothesis testing.

Descriptives

NDCG at K=10

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Google	120	.6471553	.20292008	.01852398	.6104760	.6838347	.17801	1.00000
Full-Text with Annotation (FTA)	120	.6220980	.23552161	.02150008	.5795257	.6646703	.05621	1.00000
NMTC	120	.6436176	.21670053	.01978196	.6044473	.6827878	.08655	1.00000
0.25 FTA + 0.75 NMTC	120	.6423094	.22811862	.02082429	.6010752	.6835435	.05389	1.00000
0.50 FTA + 0.50 NMTC	120	.6453728	.22649381	.02067596	.6044324	.6863133	.03683	1.00000
0.75 FTA + 0.25 NMTC	120	.6339198	.23132667	.02111714	.5921057	.6757338	.04929	1.00000
Total	720	.6390788	.22316577	.00831690	.6227505	.6554071	.03683	1.00000

Test of Homogeneity of Variances

NDCG at K=10

Levene Statistic	df1	df2	Sig.
.209	5	714	.959

ANOVA

NDCG at K=10

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.054	5	.011	.216	.956
Within Groups	35.754	714	.050		
Total	35.808	719			

Figure 64. Result of Hypothesis # 3 Testing

From the above figure, the null hypothesis is accepted. There is no evidence that the means of the NDCG at K=10 of Google and Annotation Based Search Engine with different weightings are not equal at $\alpha = .05$ level of significance. In other words, both Google and Annotation Based Search Engine with different weighting provided a similar quality of the top 10 of the search results.

It is interesting to do further analysis on the difference between Google and Annotation Based Search Engine. Instead of considering only NDCG at K=10, this time the means of the NDCG at K=1 to 10 are considered.

H_0: There is no statistical difference between the means of the NDCG at K=1 to 10 of the ranking returned from Google, Full-Text with Annotation, NMTC and 0.5 FTA + 0.5 NMTC

$(\mu_{google} = \mu_{FTA} = \mu_{NMTC} = \mu_{0.5FTA+0.5NMTC})$

H_1: Not all approaches are equal. $(\mu_{google} \neq \mu_{FTA} \neq \mu_{NMTC} \neq \mu_{0.5FTA+0.5NMTC})$

One-way Analysis of Variance (ANOVA) was applied to test the hypothesis. The following figure shows the result of hypothesis testing.

Descriptives

NDCG at K=1 to 10

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Google	1200	.6680931	.26554951	.00766575	.6530533	.6831329	.00000	1.00000
0.5FTA+0.5NMTC	1200	.6785608	.27487105	.00793484	.6629931	.6941285	.00000	1.00000
FTA	1200	.6421317	.28230969	.00814958	.6261427	.6581207	.00000	1.00000
NMTC	1200	.6605199	.27237595	.00786282	.6450935	.6759463	.00000	1.00000
Total	4800	.6623264	.27407963	.00395600	.6545708	.6700820	.00000	1.00000

Test of Homogeneity of Variances

NDCG at K=1 to 10

Levene Statistic	df1	df2	Sig.
.691	3	4796	.557

ANOVA

NDCG at K=1 to 10

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.849	3	.283	3.776	.010
Within Groups	359.650	4796	.075		
Total	360.499	4799			

Figure 65. Result of Hypothesis # 4 Testing

From the above figure, the null hypothesis is rejected. There is evidence that not all of the means of the NDCG at K=1 to 10 of Google and Social Annotation Based Search Engine are equal at $\alpha = .05$ level of significance. In others word, the rankings of the search results obtained from Google and Social Annotation Based Search Engine are different.

Figure 66 shows the results of multiple comparisons examined with Turkey, and Scheffe procedure. Figure 67 shows homogeneous subsets of all pairwise differences of Newman-Keuls, Tukey, and Scheffe procedure.

Multiple Comparisons

Dependent Variable:NDCG at K=1 to 10

	(I) Ranking Method	(J) Ranking Method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Google	0.5FTA+0.5NMTC	-.01046767	.01117956	.785	-.0391984	.0182630
		FTA	.02596137	.01117956	.093	-.0027693	.0546921
		NMTC	.00757319	.01117956	.906	-.0211575	.0363039
	0.5FTA+0.5NMTC	Google	.01046767	.01117956	.785	-.0182630	.0391984
		FTA	.03642905*	.01117956	.006	.0076983	.0651598
		NMTC	.01804086	.01117956	.371	-.0106899	.0467716
	FTA	Google	-.02596137	.01117956	.093	-.0546921	.0027693
		0.5FTA+0.5NMTC	-.03642905*	.01117956	.006	-.0651598	-.0076983
		NMTC	-.01838819	.01117956	.354	-.0471189	.0103425
	NMTC	Google	-.00757319	.01117956	.906	-.0363039	.0211575
		0.5FTA+0.5NMTC	-.01804086	.01117956	.371	-.0467716	.0106899
		FTA	.01838819	.01117956	.354	-.0103425	.0471189
Scheffe	Google	0.5FTA+0.5NMTC	-.01046767	.01117956	.831	-.0417310	.0207957
		FTA	.02596137	.01117956	.145	-.0053020	.0572247
		NMTC	.00757319	.01117956	.928	-.0236902	.0388366
	0.5FTA+0.5NMTC	Google	.01046767	.01117956	.831	-.0207957	.0417310
		FTA	.03642905*	.01117956	.014	.0051657	.0676924
		NMTC	.01804086	.01117956	.457	-.0132225	.0493042
	FTA	Google	-.02596137	.01117956	.145	-.0572247	.0053020
		0.5FTA+0.5NMTC	-.03642905*	.01117956	.014	-.0676924	-.0051657
		NMTC	-.01838819	.01117956	.439	-.0496516	.0128752
	NMTC	Google	-.00757319	.01117956	.928	-.0388366	.0236902
		0.5FTA+0.5NMTC	-.01804086	.01117956	.457	-.0493042	.0132225
		FTA	.01838819	.01117956	.439	-.0128752	.0496516

*. The mean difference is significant at the 0.05 level.

Figure 66. Multiple comparisons of Google and social annotation based search engines

Homogeneous Subsets

NDCG at K=1 to 10

	Ranking Method	N	Subset for alpha = 0.05	
			1	2
Student-Newman-Keuls[a]	FTA	1200	.6421317	
	NMTC	1200	.6605199	.6605199
	Google	1200	.6680931	.6680931
	0.5FTA+0.5NMTC	1200		.6785608
	Sig.		.053	.240
Tukey HSD[a]	FTA	1200	.6421317	
	NMTC	1200	.6605199	.6605199
	Google	1200	.6680931	.6680931
	0.5FTA+0.5NMTC	1200		.6785608
	Sig.		.093	.371
Scheffe[a]	FTA	1200	.6421317	
	NMTC	1200	.6605199	.6605199
	Google	1200	.6680931	.6680931
	0.5FTA+0.5NMTC	1200		.6785608
	Sig.		.145	.457

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 1200.000.

Figure 67. Homogeneous subsets of Google and social annotation based search engines

Figure 63 suggests that Social Annotation Based Search Engine with 0.50 weighting on the Full-Text with Annotation indexing and the Normalized Match Tag Count provided a better ranking of search results than Google. However, the results from the multiple comparisons showed that the ranking of search results provided by the Social Annotation Based Search Engine with 0.50 weighting on the Full-Text with Annotation indexing and the Normalized Match Tag Count is not statistically different from the ranking of search results provided by Google and Social Annotation Based Search Engine with the Normalized Match Tag Count as a ranking method.

4.6 DISCUSSION OF RESEARCH RESULTS

4.6.1 Using Social Annotations for Re-ranking Search Results

Given a set of documents that match a query with roughly equal closeness, it may be that a resource that has social annotation that match with query string and has a high number of social annotation should be ranked higher than those with a low number of social annotations.

In this research, six different annotation based ranking methods have been proposed. Each method examined different aspects of using social annotation to rank search results. The Popularity Count (PC), a simplest method focused only static feature, considered only the number of people bookmarking web resource. Due to the fact that people bookmarking an index page of a web resource rather than each individual page of the web site, the Propagated Popularity Count (PPC) considered both the number of people bookmarking of index page and

an individual page of web resource. The Query Weighted Popularity Count (QWPC) and the Query Weighted Propagated Popularity Count (QWPPC), more complex methods focused both static feature and similarity feature, are a PC and PPC of a given resource weighted by the count of shared terms between query string and tag of the web resource. Instead of considering the number of people bookmarking a web resource, the last two methods considered the number of people that used tags that matched terms in the query string. The Match Tag Count (MTC) only sum the total number of people that used tags that matched terms in the query string, while the Normalized Match Tag Count took the total count of all matched tags for a given resource into consideration.

While other annotation based ranking methods considered both static and similarity feature in ranking of search results, the PC and PPC focused only static They were expected to be the best comparing with other annotation based ranking method for ranking the search result return from Google. However, from the result of the first experiment, the NMTC is considered to be the best social annotation based ranking method. It considers the count of bookmarks using the query term, together with normalization of rank by taking the count of all tags into consideration. When a resource has a small number of different tags, it suggests that a content of the resource is more focused on a single concept. On the other hand, when a resource has a large number of different tags, it suggests that a content of the resource is broader potentially addressing many concepts. Two resources with a similar matched tag count will be ranked differently if they have different number of different tags. The one with a small number of different tags tends to be ranked higher than the one with a large number of different tags. In addition, the terms used as social annotations on each resource are most likely to be the terms that will be used as a query. This provides one explanation of why the Normalized Match Tag

Count provided the best performance when compared with other social annotation based ranking methods.

The result from the first experiment suggested that both static feature and similarity feature should be considered when using social annotations to re-rank search result. The count of the number of people that used tags that matched terms in the query string normalized by the total count of all tags for a given resource put more interesting resources on the top rank and less interesting resources on the lower rank. This method can be applied not only to improve web resource searching but also to improve image retrieval in online photo sharing system e.g. Flickr.

4.6.2 Using Social Annotations for Resource Indexing Augmentation

Resource indexing is one of the critical components of information retrieval system. The better the indexes that represent a resource, the better the set of search results will be.

In this research, two annotation based indexing approaches have been proposed. They examined the contribution of social annotations when they were used as resource indexes and as resource indexes augmentation. In the annotation indexing approach, social annotations and their frequencies were converted to a vector to represents a web resource. On the other hand, the content of a web resource together with social annotations and their frequencies were converted to a vector to represent a web resource in the full-text with annotation indexing approach.

Social annotations could be viewed as a user-defined taxonomy of web resources. Using them as indexes of the web resources should have provided a better representation of the web resource.

The result of the second experiment showed that using only annotation as an index of resources may not be a good idea. Since social Annotations could be viewed as a high level

concept of the content, combining them to the content of resource could add some more important concepts to the resources.

4.6.3 Using Social Annotations for Resource Retrieval and Ranking of Search Result

From the literature, it is clear that a popular search engine, such as Google, combines query dependent ranking and query independent ranking together to rank the search results. Query dependent ranking measures the match between query and content of resource, while query independent ranking measures the quality of resources. The Full-Text with Annotation indexing approach, showed the best performance in the second experiment, was selected to be an indexing approached for this experiment. The similarity measurement provided by Indri was used as a query dependent ranking. The result from the first experiment suggested that the ranking of search results provided by Normalized Match Tag Count, Match Tag Count and Query Weighted Popularity Count are not statistically different from the ranking provided by Google. All three social annotation based ranking methods were selected to be a ranking method of the third experiment. Only Match Tag Count is considered to be a query independent ranking method. Normalized Match Tag Count and Query Weighted Popularity Count are considered to be the combination of both query dependent ranking and query independent ranking.

The result of the third experiment is interesting. Using Full-Text with Annotation Indexing and Normalized Match Tag Count provided a better ranking when compared with Google as shown in Figure 63. Although the ranking of the search result was not statistically different from the ranking obtained by Google, It suggested that the proposed simple indexing approach together with the proposed social based annotation ranking method could provide the

similar quality ranking as the more complex indexing approach and ranking method used by Google.

This result confirmed that the combination of using social annotations to rank the search result and using social annotations as resource index augmentation provided a promising rank of search results. It showed that social annotations could benefit web search.

4.6.4 Google Document Collections VS Delicious Document Collections

As mentioned in the previous chapter, the number of the resources in Delicious is smaller than the number of the resources in a search engine database. To show the different in size of document collections of both Delicious and Google, top 20 tags from Delicious were used. They were submitted to Google and Delicious. Table 17 shows the total pages returned from both Google and Delicious.

Table 17. The estimated total page returned from both Google and Delicious

Tag	Estimated total pages returned from Google	Estimated total pages returned from Delicious	% of Estimated total pages returned from Google
Design	1,460,000,000	1,429,553	0.09791%
Blog	2,650,000,000	2,483,221	0.09371%
Software	1,690,000,000	1,208,182	0.07149%
Music	2,180,000,000	1,398,368	0.06415%
Webdesign	208,000,000	371,898	0.17880%
Programming	226,000,000	690,051	0.30533%
Video	2,870,000,000	1,454,182	0.05067%
Art	1,390,000,000	1,045,665	0.07523%
Reference	664,000,000	727,661	0.10959%
web2.0	23,900,000	433,771	1.81494%
Tools	1,070,000,000	682,972	0.06383%
Web	3,760,000,000	1,640,875	0.04364%
Inspiration	86,000,000	282,735	0.32876%
News	3,450,000,000	1,601,268	0.04641%
Blogs	698,000,000	691,129	0.09902%
Photography	323,000,000	543,768	0.16835%
Education	785,000,000	585,074	0.07453%
Linux	484,000,000	656,363	0.13561%
Css	450,000,000	216,146	0.04803%
shopping	1,550,000,000	592,685	0.03824%
		Average	0.19541%

As shown in Table 17, the size of Delicious document collections is much smaller than the size of Google document collections. The social bookmarking systems do not have any crawler or spider to read all the web resources. Registered users can enter interested web resources to the system database. Thus, it stands to reason that the number or resources in social bookmarking system is smaller than the number of resources in search engine document collections.

Due to the fact that people tend to bookmark interesting web resources, frequently used web resources or useful web resources. With people filtering web resources, the Delicious

document collections might result in a better quality document corpus. Searching on other people bookmark is like searching on a set of useful web resource. It is easier to discover what a searcher are looking for. The small experiment in preliminary analysis section in chapter 3 did confirm this claim. The subjects in that experiment rated the relevancy of the search results returned from the Delicious higher than the search results returned from Google. In addition, the results from the second and third experiment confirmed that the social bookmark contains useful web resources and social annotations are useful information

5.0 CONCLUSIONS

This chapter describes the contributions and implications of this research. Then, future work is discussed.

5.1 CONTRIBUTIONS AND IMPLICATIONS

Annotations have been used for many years to analyze and describe documents. They have also been used to support collaboration and communication in group work. Recent research has explored the use of annotations for classification and retrieval. Social annotations have the potential to improve searching for, and ranking of resources. The published research on using social annotations to improve web search is sparse.

This research examined two major issues in integrating social annotations into web search to improve users' satisfaction with the search results: web resource index augmentation and search result ranking. Resource indexing and search result ranking are critical components of a search engine. The better the indexes that represent a resource, the better the set of search results will be. The better the ranking method, the more satisfaction a search engine user will have.

Controlled experiments were designed and conducted. Subjects were asked to find web pages that helped them ask answer a set of given questions. As much as possible, the research

setting imitated the real environment of people searching for information on the web. For each questions, subjects had to formulate their own query and submit it to search engines. Behind the scenes multiple searches were done and presented to the user as if they were the result of a single simple search.

Social annotations do benefit web search. When using social annotations to rank search results, both static features and similarity features should be considered. For this particular study, the count of the number of people that used tags that matched terms in the query string normalized by the total count of all tags for a given resource ranked useful web resources higher and less useful resources lower. In addition, social annotations can provide high level concepts about web resources useful in indexing. The combination of social annotations and content of web resources can provide a better representation of web resources. Last but not least, using social annotation for resource index augmentation and using social annotations to rank the search results provides a promising improvement in search results.

5.2 FUTURE WORK

This research shows that social annotations do benefit web resource searching. However, there are still more research questions to be studied to exploit the use of social annotation to improve web searches. These include:

- Applying both similarity features and static features to rank search results
- Web resource collection extension
- More robust web resource indexing approaches

- Effect of specific vs exploratory queries on search result relevancy judgment

5.2.1 Applying both similarity features and static features to rank search results

In the third experiment, only top 15 resources from Google and social annotation based search engine with full text with annotation indexing were combined to form a search result set that was displayed to a subject for relevancy judgment. While the search results from Google were obtained by both similarity ranking and static ranking, the search results from Social annotation based search engine were obtained using only similarity ranking. The social annotation ranking methods then were applied to rank these results after subject rated relevancy judgment. The reason for this was that there were four different ranking methods. If they had been applied to retrieve search results, the final set of search results would have been larger. It was not desireable to ask subjects to make relevancy judgments on a very large set. However, applying social ranking to the set of fifteen resources might not be fair to social annotation based search engine. If they had been applied to retrieve search results, some useful resource might be in the final set of search results.

The result from the third experiment showed that the combination of similarity score from the Full Text with Annotation indexing and ranking score from the Normalized Match Tag Count with 0.50 weighting of ranking technique provided the best search result ranking. It is interesting to see whether the ranking of the search results will be improved when both similarity ranking and static ranking are applied to retrieve and rank the search result.

In this experiment, after subjects submit their query, the social annotation based search engine will first measure the similarity between query and resource index. A ranking score from the Normalize Match Tag Count will, then, be applied to similarity score to compute the final

ranking of the search result. The final set of search result will depend on both similarity score and ranking score. It is now fair to compare the ranking returned from the social annotation based search engine with the ranking returned from Google.

5.2.2 Web resource collection extension

As mentioned in the Chapter 3, the availability of technical content bookmarking in the Delicious is greater than non-technical content. The questions in all three experiments were limited to technically related question. The web resource collection used was also limited to technically related resources. Other resources and associated tag were discarded. At the end of the research, there is evidence that using social annotation on the technically related web resources did benefit web search.

It is interesting to extend this research to non-technical related web resources as the scope of annotated resources expands. This will prove that social annotation not only benefits technical related web search but will also improve more general web searching.

5.2.3 More robust web resource indexing approaches

In the second experiment, one of the proposed social annotation based indexing approaches combined the content of a web resource and social annotations and their frequencies and converted it to a vector to represent a web resource. Since social annotations could be viewed as a high level concept of the content defined by users, combining them with the content of resource could add some more important concepts to the resources. However, the result of experiment 2 showed that a set of search results provided by the Full-Text with Annotation

indexing approach is not statistically different from a set of search results provided by the Full-Text indexing approach.

It would interesting to exploreif there is an alternative indexing approach that exploits the use of social annotations to augment resource indexing. One possible way is to apply weighting when combining the content of a web resource together with social annotations and their frequencies.

5.2.4 Effect of specific vs exploratory queries on search result relevancy judgment

The questions used in all three experiments were divided into two group, specific questions and exploratory questions. The specific questions sought specific information e.g. Find an example of how to format a date object using Java. On the other hand, the exploratory questions asked for general information related to the topic in the question e.g. Find information about sorting algorithms and explain them. This kind of questions allowed subjects to judge the relevancy according to their understanding of the question and their preference.

It is interesting to do further investigation whether the result of all experiments are still the same when specific question and exploratory question analyst separately.

APPENDIX A

ENTRY QUESTIONAIRE

UNIVERSITY OF PITTSBURGH
STUDY OF USING SOCIAL ANNOTATION TO IMPROVE WEB SEARCHING
ENTRY QUESTIONNAIRE

1. What is your gender?
   __Female __Male
2. How old are you?
   __18-22 __ 23-27 __ 28-32 __ 33 and above
3. You are
   __ Freshman __Sophomore __Junior __Senior
   __Graduate(Master) __Graduate(Ph.D.)
4. What is your major of study?
   __Computer Science
   __Information Science
   __Other (Please specify ……)
5. Would you consider yourself as
   __Expert in Java
   __Intermediate in Java
   __Novice in Java
6. How did you learn Java?
   __Self-Study
   __Take Java Class
   __On the Job Training
   __Other (Please specify …….)
7. How long did you use Java?
   __less than 1 year
   __1-3 years
   __4 years or more

8. How many projects did you use Java as a development tool?
   __3 or less
   __4 – 6 projects
   __7 – 9 projects
   __10 or more
9. Describe those projects
   __All academic assignments
   __Some academic assignments and some commercial system projects
   __All commercial system projects
   __Other (Please specify ……..)
10. Other than Java, what programming languages do you use?
    ……
    ……
11. How many hours do you use a computer?
    __3 hours or less per day
    __4 – 6 hours per day
    __7 – 9 hours per day
    __10-12 hours per day
    __13 hours or more
12. How many hours per day do you spend your time on site such as
    12.1 Myspace/Facebook
    ……
    12.2 Instant Messaging (MSN, Yahoo, AOL and etc)
    ……
    12.3 Social Bookmarking System (del.cio.us, connotea, and etc)
    ……
    12.4 Search Engine (Google, Yahoo, Live search and etc)
    ……
13. How many times do you search for information on the web?
    __ 5 or less per day
    __ 6 – 10 time per day
    __11- 15 time per day
    __ 15 or more per day
14. What statement better describe the way do you search information on the web?

	Totally Agree									Totally Disagree
14.1 I used 2 – 3 term in my query	10	9	8	7	6	5	4	3	2	1
14.2 I used + and – in my query	10	9	8	7	6	5	4	3	2	1
14.3 I use phrase in my query (without “”)	10	9	8	7	6	5	4	3	2	1
14.4 I used “” to in my query to find specify phrases.	10	9	8	7	6	5	4	3	2	1
14.5 I used +, -, and “” to in my query.	10	9	8	7	6	5	4	3	2	1

15. Searching information on the web, what characteristic is best describe you?
__ Always find what I want
__ Most of the time I find what I want
__ Half of the time I find what I want
__ Never find what I want

APPENDIX B

THE OVERLAP OF THE SEARCH RESULTS IN ALL THREE EXPERIMENTS WITH SUBJECTS' RELEVANCE JUDGMENT

The resources that have more than 5 subject relevancy judgments are shown in the table below.

B.1 EXPERIMENT 1

B.1.1 Question A

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://abbeyworkshop.com/howto/java/writeText/ir	-	-	-	4	4	-	4	4	4	-	-	4	-	-	-	0	2	-	3	4
http://www.javacoffeebreak.com/faq/faq0004.html	-	-	-	3	4	-	2	4	4	-	-	4	-	-	-	2	3	-	4	3
http://www.javapractices.com/topic/TopicAction.do	-	-	-	3	4	-	2	4	3	-	-	3	-	-	-	1	4	-	4	4
http://searchdomino.techtarget.com/tip/0,289483,si	-	-	-	0	1	-	-	0	0	-	-	0	-	-	-	0	0	-	0	2
http://www.leepoint.net/notes-java/io/10file/10reac	-	-	-	1	3	-	-	0	2	-	-	-	-	-	-	1	1	-	0	4
http://www.exampledepot.com/egs/java.io/ReadLin	-	-	-	0	2	-	-	0	0	-	-	-	-	-	-	0	1	-	-	2
http://www.javacoffeebreak.com/java103/java103.h	-	-	-	4	-	-	2	-	-	-	-	4	-	-	-	4	4	-	4	-
http://java.sun.com/docs/books/tutorial/essential/ic	-	-	-	1	-	-	0	-	-	-	-	4	-	-	-	0	1	-	-	-
http://www.exampledepot.com/egs/java.io/WriteTo	-	-	-	4	-	-	0	-	-	-	-	3	-	-	-	2	3	-	-	-
http://www.tech-recipes.com/java_programming_tip	-	-	-	4	4	-	-	0	1	-	-	-	-	-	-	-	-	-	-	4

B.1.2 Question B

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://javatechniques.com/blog/dateformat-and-sim	-	3	-	2	-	4	-	4	4	-	-	4	-	4	-	-	4	-	-	4
http://articles.techrepublic.com.com/5100-22-04641	-	3	-	-	-	4	-	4	-	-	-	4	-	4	-	-	-	-	-	4
http://www.exampledepot.com/egs/java.text/Forma	-	4	-	-	4	-	-	-	4	-	-	4	-	-	-	-	-	4	-	4

B.1.3 Question C

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://kickjava.com/220.htm	4	4	4	-	4	4	4	-	4	0	4	-	-	4	4	-	4	4	4	-
http://www.wellho.net/resources/ex.php4?item=j71	4	2	0	-	4	3	4	-	4	3	3	-	-	4	4	-	4	2	4	-
http://www.anyexample.com/programming/java/jav	4	-	4	-	4	-	4	-	4	1	4	-	-	4	4	-	4	4	-	-
http://users.cs.dal.ca/~sedgwick/ArrayList.html	3	-	4	-	4	-	4	-	3	-	3	-	4	4	-	-	3	4	-	-
http://www.idevelopment.info/data/Programming/j	4	-	4	-	4	-	4	-	4	-	3	-	-	3	-	-	4	4	-	-
http://www.idevelopment.info/data/Programming/j	2	-	4	-	2	-	1	-	2	-	3	-	-	1	-	-	3	4	-	-
http://www.javadeveloper.co.in/java-example/java-a	4	-	4	-	4	-	4	-	4	-	4	-	-	4	-	-	4	4	-	-
http://www.javafaq.nu/java-example-code-89.html	4	-	4	-	4	-	4	-	4	-	4	-	-	4	-	-	4	4	-	-
http://www.javafaq.nu/java-example-code-classes-2	3	-	4	-	0	-	4	-	2	-	3	-	-	1	-	-	4	4	-	-
http://java.sun.com/j2se/1.4.2/docs/api/java/util/Ar	-	1	1	-	-	4	-	-	-	4	-	-	2	-	4	-	-	-	1	-
http://www.java2s.com/Code/Java/Language-Basics/	-	-	3	-	4	-	4	-	4	-	-	-	-	3	-	-	2	4	-	-
http://www.roseindia.net/java/beginners/array_list_	-	-	4	-	4	-	4	-	4	-	4	-	-	-	-	-	4	4	-	-
http://www.roseindia.net/javatutorials/linkedlistvsar	3	-	-	-	4	-	4	-	3	-	3	-	-	4	-	-	0	-	-	-
http://java.sun.com/docs/books/tutorial/collections/	3	-	-	-	0	-	2	-	3	-	2	-	-	-	-	-	3	-	-	-
http://ltiwww.epfl.ch/sJava/version2/Introduction/C	4	-	-	-	4	-	1	-	2	-	3	-	-	-	-	-	0	-	-	-
http://www.csd.abdn.ac.uk/~etadjoud/teaching/CS5	4	4	2	-	-	-	4	-	-	-	4	-	-	-	-	-	-	4	-	-
http://www.csd.abdn.ac.uk/~spt/teaching/CS1012/ir	-	-	-	-	4	4	-	-	4	-	-	-	-	4	-	-	1	-	4	-
http://www.java-samples.com/showtutorial.php?tut	-	4	-	-	-	3	-	-	-	4	-	-	-	4	3	-	-	-	3	-
http://www.java-tips.org/java-se-tips/java.lang/use-c	-	4	-	-	-	2	-	-	-	3	-	-	-	4	4	-	-	-	4	-
http://forum.java.sun.com/thread.jspa?threadID=75	-	-	-	-	1	-	3	-	-	-	2	-	-	0	-	-	1	-	-	-
http://java.sun.com/docs/books/tutorial/collections/	-	-	-	-	1	-	3	-	-	-	3	-	-	2	-	-	0	-	-	-
http://mindprod.com/jgloss/arraylist.html	-	4	-	-	-	4	-	-	-	4	-	-	-	-	4	-	-	-	3	-
http://www.deitel.com/articles/java_tutorials/20051	-	-	-	-	2	-	2	-	-	-	2	-	-	4	-	-	2	-	-	-
http://www.onjava.com/pub/a/onjava/2003/03/12/j	-	-	2	-	4	-	3	-	-	-	-	-	-	-	-	-	2	1	-	-

B.1.4 Question D

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://danzig.jct.ac.il/java_class/recursion.html	-	4	4	4	-	3	4	4	-	4	4	4	-	4	4	4	-	4	4	4
http://weblogs.java.net/blog/mortazavi/archive/200	-	-	4	-	-	0	4	4	-	1	4	3	-	-	-	4	-	4	4	4
http://www.ib-computing.com/java/datastructures/r	-	3	4	-	-	4	4	-	-	-	-	4	-	4	1	3	-	4	4	-
http://www.ibm.com/developerworks/java/library/j-	-	4	4	1	-	-	1	-	-	-	2	-	-	3	1	0	-	4	4	-
http://www.tech-recipes.com/java_programming_tip	-	4	4	-	-	-	4	4	-	2	4	-	-	-	-	4	-	4	4	4
http://www.ahmadsoft.org/articles/recursion/index.	-	0	3	4	-	-	-	0	-	3	4	-	-	-	-	-	-	4	-	4
http://www.hostitwise.com/java/java_recursion.htm	-	-	-	-	-	2	-	4	-	3	-	4	-	-	-	4	-	4	4	4
http://answers.yahoo.com/question/index?qid=2007	-	-	-	-	-	0	0	-	-	-	-	4	-	4	4	2	-	-	4	-
http://chortle.ccsu.edu/CS151/cs151java.html	-	1	-	-	-	-	-	4	-	-	-	-	-	2	-	3	-	4	1	1
http://www.faqs.org/docs/javap/c11/s1.html	-	4	1	-	-	-	-	0	-	-	-	-	-	-	4	4	-	4	4	-
http://www.joelonsoftware.com/articles/ThePerilsof	-	0	-	0	-	0	0	-	-	-	-	1	-	0	0	-	-	-	-	-
http://www.juniata.edu/faculty/kruse/cs2java/recurs	-	-	-	-	-	0	-	0	-	-	3	0	-	-	-	0	-	4	4	-
http://www.cafeaulait.org/javatutorial.html	-	-	-	-	-	-	-	0	-	-	1	-	-	3	-	0	-	4	4	-
http://www.cs.may.ie/~pgibson/Teaching/Schools/Ja	-	-	-	-	-	2	-	4	-	-	3	4	-	-	-	3	-	4	-	-
http://www.go4expert.com/forums/showthread.php	-	4	4	-	-	-	4	-	-	1	-	-	-	-	-	-	-	4	-	4
http://www.leepoint.net/notes-java/data/numbers/6	-	-	-	-	-	-	1	4	-	-	-	-	-	4	-	2	-	4	4	-
http://java.sun.com/new2java/supplements/2003/Ju	-	-	4	-	-	3	4	-	-	-	-	4	-	4	-	-	-	-	-	-
http://www.amazon.com/Thinking-Recursively-Java-I	-	-	1	-	-	-	-	-	-	-	-	-	-	-	2	0	-	0	0	-
http://www.cs.cornell.edu/courses/cs211/2007sp/Le	-	-	-	-	-	3	1	-	-	-	-	3	-	4	1	-	-	-	-	-
http://www.cs.utexas.edu/~scottm/cs307/codingSan	-	-	-	-	-	-	-	4	-	-	4	-	-	-	-	4	-	4	4	-

B.1.5 Question E

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.softpanorama.org/Algorithms/sorting.sh	4	-	4	0	4	-	1	3	3	-	3	4	4	-	4	-	-	-	-	4
http://en.wikipedia.org/wiki/Sorting_algorithm	-	-	4	4	4	-	4	4	-	-	-	4	4	-	4	4	-	-	-	4
http://www.geocities.com/siliconvalley/network/185	-	-	1	2	-	-	3	0	-	-	-	2	1	-	-	3	-	-	-	-
http://www.cs.ubc.ca/~harrison/Java/sorting-demo.h	-	-	-	-	3	-	-	2	-	-	-	3	-	-	4	4	-	-	-	3
http://cg.scs.carleton.ca/~morin/misc/sortalg/	-	-	-	-	3	-	-	2	-	-	-	3	-	-	4	-	-	-	-	2
http://cs.smith.edu/~thiebaut/java/sort	-	-	-	-	4	-	-	1	-	-	-	3	-	-	4	-	-	-	-	4
http://digg.com/programming/Amazing_Sorting_Alg	-	-	-	0	-	-	1	-	0	-	2	-	1	-	-	-	-	-	-	-
http://math.hws.edu/TMCM/java/xSortLab/	-	-	1	-	4	-	-	-	-	-	-	-	-	-	3	0	-	-	-	3
http://www.coyotesong.com/sort/index.html	-	-	3	4	4	-	3	-	-	-	-	-	-	-	-	-	-	-	2	-
http://www.cs.rit.edu/~atk/Java/Sorting/sorting.htm	-	-	-	-	3	-	-	2	-	-	-	3	-	-	4	-	-	-	-	2

B.1.6 Question F

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.allapplabs.com/interview_questions/jav	-	-	-	0	4	4	-	4	-	-	-	-	-	3	-	3	3	4	-	4
http://en.wikibooks.org/wiki/Ruby_Programming/Sy	-	-	-	0	-	-	-	4	-	-	-	-	-	0	-	0	3	0	-	1
http://java.sun.com/docs/books/tutorial/java/javaO(	4	-	-	4	-	-	-	4	-	-	-	-	-	-	-	2	3	4	-	4
http://www.codestyle.org/java/faq-Inheritance.shtm	-	1	-	-	4	-	-	-	-	-	-	-	4	3	-	-	-	4	-	-
http://xahlee.org/java-a-day/access_specifiers.html	-	-	-	-	-	-	-	4	-	-	-	-	-	4	-	-	4	4	-	4

B.1.7 Question G

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.ibm.com/developerworks/java/library/j-	3	2	3	-	3	3	-	-	4	-	-	-	-	0	-	-	-	4	2	-
http://developers.sun.com/jsenterprise/	-	0	4	-	1	2	3	-	3	-	-	-	-	1	-	-	-	4	-	-
http://www.objectcentral.com/vide.htm	3	4	-	-	3	4	-	-	4	-	-	-	-	-	-	-	-	4	1	-
http://javaboutique.internet.com/reviews/netbeans5	3	-	-	-	3	4	-	-	4	-	-	-	-	-	4	-	-	2	-	-
http://www.jetbrains.com/idea/	-	4	-	-	3	4	-	-	3	-	-	-	3	-	-	-	-	4	-	-
http://www.stylusstudio.com/java_ide.html	3	4	-	-	3	2	-	-	3	-	-	-	-	-	-	-	-	4	-	-
http://www.codegear.com/products/jbuilder	-	1	-	-	2	3	-	-	2	-	-	-	-	-	-	-	-	4	-	-
http://www.javalobby.org/java/forums/t106117.htm	-	-	-	-	4	3	-	-	3	-	-	-	-	-	-	-	-	4	2	-
http://www.sdtimes.com/article/special-20070501-0	-	-	-	-	-	3	4	-	-	-	-	-	-	3	-	-	3	3	-	-

B.1.8 Question H

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.dreamincode.net/forums/showtopic226	-	-	-	3	-	2	4	-	-	-	4	3	-	4	4	-	-	-	3	4
http://www.ociweb.com/jnb/archive/jnbMay2000.ht	-	-	-	3	-	1	3	-	-	-	3	2	-	4	-	-	-	2	-	4
http://neptune.netcomp.monash.edu.au/JavaHelp/h	-	4	-	-	-	4	4	-	-	-	3	3	-	4	4	-	-	-	-	-
http://www.smartdataprocessing.com/lessons/l5.htr	-	-	-	4	-	3	4	-	-	-	3	-	-	4	-	-	-	-	2	4
http://java.about.com/od/tutorials/ss/JavaException	-	4	-	-	-	-	-	-	-	-	-	-	-	3	4	-	-	2	2	1
http://java.sun.com/docs/books/tutorial/essential/e	-	3	-	-	-	-	-	4	-	-	-	-	-	2	4	-	-	4	-	1
http://www.akgupta.com/Java/Notes/section2-3.htm	-	-	-	1	-	1	4	-	-	-	3	3	-	-	-	-	-	4	-	-
http://pages.cs.wisc.edu/~cs302/io/Exceptions.html	-	-	-	3	-	-	-	4	-	-	2	3	-	-	-	-	-	-	-	4
http://www.patentstorm.us/patents/6412109-descri	-	-	-	-	-	-	1	-	-	-	1	1	-	2	3	-	-	-	-	-
http://www.tech-recipes.com/java_programming_tip	-	3	-	-	-	-	-	-	-	-	-	-	-	4	4	-	-	-	4	4

B.2 EXPERIMENT 2

B.2.1 Question A

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://java.sun.com/developer/onlineTraining/Progr	1	-	4	1	4	-	4	4	4	-	3	4	-	-	3	4	0	-	2	3
http://saloon.javaranch.com/forums/forum-038.htm	0	-	4	0	3	-	3	-	3	-	3	0	0	-	1	2	-	-	1	1
http://www.javacoffeebreak.com/java103/java103.h	4	-	-	3	3	-	4	4	4	-	4	4	-	-	-	4	-	-	4	4
http://www.javapractices.com/Topic42.cjp	4	-	-	2	4	-	-	3	4	-	4	4	-	-	4	-	-	-	4	4
http://www.tizag.com/phpT/fileread.php	0	-	-	-	0	-	-	1	0	-	1	0	-	-	-	-	-	-	1	0
http://www.acm.org/crossroads/xrds6-3/ovp63.html	3	-	-	3	4	-	-	3	-	-	4	-	-	-	-	-	-	-	3	1
http://freetts.sourceforge.net/docs/index.php	-	-	0	-	-	-	-	-	0	-	-	0	0	-	0	-	-	-	-	0
http://howtowriteastory.wordpress.com/	-	-	-	-	-	-	-	0	-	-	-	0	0	-	0	-	-	-	0	0
http://java.sun.com/docs/books/tutorial/reallybiginc	-	-	-	-	3	-	-	-	3	-	-	0	2	-	1	-	-	-	-	3
http://javafaq.nu/java-example-code-list.html	4	-	-	-	-	-	-	1	-	-	-	1	-	-	-	1	3	-	-	4
http://kb.mozillazine.org/File_IO	0	-	-	-	0	-	-	4	1	-	-	-	-	-	-	0	-	-	3	-
http://www.captain.at/programming/xul/	-	-	-	0	0	-	-	1	-	-	1	-	-	-	-	-	-	-	2	0
http://www.codealchemists.com/jdarkroom/	0	-	-	-	0	-	-	-	-	-	1	0	-	-	0	-	-	-	-	0
http://www.idevelopment.info/data/Programming/j	4	-	-	-	-	-	-	4	-	-	-	2	-	-	-	3	3	-	-	4
http://www.javafaq.nu/java-example-code-list.html	4	-	-	-	-	-	-	2	-	-	-	1	-	-	-	1	4	-	-	3
http://java.sun.com/developer/technicalArticles/scri	0	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	4	-	-	0
http://myjavatools.com/projects/v.6.0/lib/doc/com/	-	-	-	1	-	-	4	3	-	-	-	-	-	-	-	3	-	-	2	-
http://www.cplusplus.com/doc/tutorial/files.html	0	-	-	0	-	-	-	-	0	-	0	-	-	-	-	1	-	-	-	-
http://www.developer.com/java/other/article.php/3	0	-	-	0	-	-	-	-	0	-	1	-	-	-	-	2	-	-	-	-
http://www.idevelopment.info/data/Oracle/DBA_tip	-	-	-	-	2	-	-	-	-	-	0	3	-	-	0	-	-	-	-	0

B.2.2 Question B

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.dan.co.uk/mysql-date-format/	0	3	-	-	0	0	-	1	-	0	-	0	0	0	-	0	0	0	-	-
http://www.joda.org/	0	2	-	-	4	-	-	2	-	0	-	0	0	1	-	0	0	3	-	-
http://dev.mysql.com/doc/refman/5.0/en/date-and-	0	4	-	-	-	0	-	1	0	-	-	0	-	0	-	0	0	0	-	-
http://en.wikipedia.org/wiki/ISO_8601	0	1	-	-	-	0	-	1	-	-	-	0	0	0	-	0	0	0	-	-
http://msdn2.microsoft.com/en-us/library/8kb3ddd4	0	2	-	-	-	0	-	1	-	-	-	0	0	1	-	0	0	0	-	-
http://www.mattkruse.com/javascript/date/	0	4	-	-	-	0	-	1	-	-	-	0	0	2	-	0	0	0	-	-
http://www.methods.co.nz/rails_date_kit/rails_date	0	1	-	-	0	0	-	0	-	-	-	0	-	2	-	0	0	0	-	-
http://www.svendtofte.com/code/date_format/	0	3	-	-	-	0	-	4	-	-	-	0	2	4	-	0	0	0	-	-
http://java.sun.com/j2se/1.4.2/docs/api/java/text/Si	4	4	-	-	-	2	-	4	-	-	-	4	-	3	-	3	2	1	-	-
http://javatechniques.com/public/java/docs/basics/c	4	3	-	-	-	4	-	4	4	-	-	4	-	-	-	4	4	4	-	-
http://www.daniweb.com/code/snippet515.html	4	4	-	-	-	4	-	4	-	-	-	4	-	-	-	3	4	4	-	4
http://www.databasejournal.com/features/mssql/ar	0	1	-	-	-	0	-	1	-	-	-	0	-	3	-	1	0	0	-	-
http://www.w3.org/TR/NOTE-datetime	0	3	-	-	-	0	-	1	-	-	-	0	-	2	-	1	0	0	-	-
http://javafaq.nu/java-example-code-list.html	4	-	-	-	-	4	-	4	3	-	-	4	-	-	-	3	2	2	-	-
http://javatechniques.com/blog/dateformat-and-sim	4	4	-	-	-	4	-	4	-	-	-	4	-	-	-	4	4	4	-	-
http://www.codefetch.com/search?qy=date&ar	0	0	-	-	-	-	-	0	-	-	-	0	-	0	-	0	0	0	-	-
http://www.javafaq.nu/java-example-code-list.html	4	-	-	-	-	4	-	4	3	-	-	4	-	-	-	3	2	2	-	-
http://www.mattkruse.com/javascript/date/source.h	4	3	-	-	-	0	-	1	-	-	-	0	-	3	-	0	-	0	-	-
http://delete.me.uk/2005/03/iso8601.html	0	-	-	-	-	0	-	1	-	-	-	0	-	-	-	0	0	0	-	-
http://www.java-forums.org/java-tutorials/2775-java	-	-	-	-	4	-	-	4	4	-	-	-	-	-	-	4	4	4	-	4
http://www.javafaq.nu/java-example-code-96.html	4	-	-	-	-	3	-	4	-	-	-	4	-	-	-	2	3	3	-	-
http://www.roseindia.net/struts/struts2/date/struts-	4	3	-	-	-	-	-	3	-	-	-	2	-	-	-	2	0	0	-	-
http://www.sql-server-helper.com/tips/date-formats	0	2	-	-	-	-	-	1	-	-	-	0	-	0	-	0	-	0	-	-
http://blogs.techrepublic.com.com/howdoi/?p=116	4	-	-	-	4	4	-	-	-	-	-	4	4	-	-	-	-	-	-	4
http://www.zdnetasia.com/techguide/java/0,390448	4	-	-	-	4	4	-	-	-	-	-	4	4	-	-	-	-	-	-	4
http://blogs.techrepublic.com.com/howdoi/?p=1168	4	-	-	-	4	4	-	-	-	-	-	4	4	-	-	-	-	-	-	-
http://forum.java.sun.com/thread.jspa?threadID=54	4	3	-	-	4	4	-	-	-	-	-	4	-	-	-	-	-	-	-	-
http://hi.baidu.com/xkyue23/blog/item/caffaf38595	-	0	-	-	-	-	-	1	-	-	-	-	-	2	-	0	-	4	-	-
http://java.boot.by/scjp-tiger/ch03s04.html	-	4	-	-	-	-	-	3	-	-	-	-	-	-	-	4	4	3	-	-
http://java.sun.com/docs/books/tutorial/java/data/r	0	-	-	4	-	0	-	-	-	-	-	4	0	-	-	-	-	-	-	-
http://java.sun.com/j2se/1.3/docs/api/java/text/Dat	-	4	-	-	-	-	-	4	-	-	-	-	-	-	-	4	1	3	-	-
http://java.sun.com/j2se/1.4.2/docs/api/java/util/cla	1	-	-	-	4	1	-	-	-	0	-	1	-	-	-	-	-	-	-	-
http://www.builderau.com.au/program/java/soa/Ma	4	-	-	-	4	0	-	-	-	-	-	4	-	4	-	-	-	-	-	-
http://www.builderau.com.au/program/java/soa/Per	4	-	-	-	4	4	-	-	-	-	-	4	4	-	-	-	-	-	-	-

B.2.3 Question C

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://java.sun.com/j2se/1.4.2/docs/api/java/util/Ar	4	3	3	-	4	0	4	-	3	1	3	-	2	3	-	-	2	2	4	-
http://www.java-examples.com/java-collections-and	4	2	4	-	4	4	4	-	3	0	4	-	4	4	-	-	4	4	4	-
http://www.scribd.com/doc/259808/Collections-and	0	0	4	-	4	1	3	-	2	0	3	-	3	3	-	-	2	1	1	-
http://blogs.msdn.com/joshwil/archive/2004/04/13/	4	4	3	-	4	0	3	-	0	1	4	-	0	-	-	-	0	1	3	-
http://blogs.worldnomads.com.au/matthewb/article	0	0	0	-	0	0	0	-	0	0	0	-	0	-	-	-	0	0	0	-
http://www.idevelopment.info/data/Programming/j	4	3	4	-	4	4	4	-	4	1	4	-	4	-	-	-	4	4	4	-
http://www.idevelopment.info/data/Programming/j	4	2	4	-	4	4	4	-	3	4	4	-	4	-	-	-	4	4	3	-
http://www.javaworld.com/javaworld/javaqa/2001-(	2	1	2	-	2	0	4	-	1	0	2	-	1	-	-	-	0	1	2	-
http://www.kickjava.com/220.htm	4	3	4	-	4	4	4	-	4	3	4	-	4	-	-	-	4	4	4	-
http://www.onjava.com/pub/a/onjava/2003/03/12/j	4	4	3	-	4	4	3	-	2	0	4	-	4	-	-	-	3	1	2	-
http://www.pankaj-k.net/archives/2004/06/arraylist	2	1	1	-	4	0	4	-	3	1	2	-	1	-	-	-	1	2	1	-
http://www.scribd.com/doc/271835/Memory-Leaks-	0	1	-	-	1	0	1	-	0	0	3	-	0	-	-	-	0	0	1	-
http://kickjava.com/220.htm	4	3	4	-	4	4	4	-	-	2	4	-	-	-	-	-	4	4	4	-
http://www.javadeveloper.co.in/java-example/java-a	4	4	4	-	4	3	4	-	-	-	4	-	4	-	-	-	4	4	4	-
http://www.onjava.com/pub/a/onjava/2001/05/30/(	-	1	1	-	4	-	4	-	3	1	4	-	0	-	-	-	-	0	2	-
http://eclipsetutorial.sourceforge.net/Total_Beginne	-	2	4	-	4	1	0	-	4	2	3	-	-	-	-	-	-	-	2	-
http://java.sun.com/developer/JDCTechTips/2002/tt	4	2	4	-	-	3	4	-	-	-	4	-	-	4	-	-	4	-	4	-
http://www.javainthebox.net/laboratory/J2SE1.5/Lar	-	0	4	-	4	0	0	-	-	2	4	-	-	-	-	-	-	4	1	-
http://www.nextindex.net/java/collection/ArrayList.h	4	0	-	-	4	0	0	-	-	2	4	-	-	-	-	-	-	4	1	-
http://www.beginner-java-tutorial.com/java-arraylist	-	3	-	-	4	-	4	-	4	2	4	-	-	-	-	-	-	4	4	-
http://www.precisejava.com/javaperf/j2se/Collectio	-	2	4	-	4	0	4	-	-	2	4	-	-	-	-	-	-	-	1	-
http://www.rgagnon.com/javadetails/java-0521.htm	-	3	4	-	4	-	4	-	-	2	4	-	-	-	-	-	-	2	4	-
http://javafaq.nu/java-article1111.html	4	-	0	-	-	1	-	-	2	-	-	-	0	-	-	-	0	3	-	-
http://www.codeproject.com/csharp/sortingarraylist	-	4	4	-	4	-	0	-	-	1	0	-	-	-	-	-	-	-	2	-
http://www.javabeginner.com/java-arraylist.htm	-	3	-	-	4	-	4	-	4	-	4	-	-	-	-	-	-	4	4	-
http://www.leepoint.net/notes-java/data/collections	-	4	3	-	4	-	4	-	-	3	4	-	-	-	-	-	-	-	4	-
http://www.wellho.net/resources/ex.php4?item=j71	4	-	4	-	-	4	-	-	2	-	-	-	3	-	-	-	4	2	-	-
http://javafaq.nu/java-example-code-list.html	4	-	-	-	-	4	-	-	4	-	-	-	4	-	-	-	3	1	-	-
http://www.asahi-net.or.jp/~dp8t-asm/java/articles/	-	0	-	-	4	-	0	-	-	0	4	-	-	-	-	-	-	-	0	-
http://www.idevelopment.info/data/Programming/j	-	-	-	-	4	4	-	-	3	3	-	-	4	-	-	-	4	-	-	-
http://www.java2s.com/Code/JavaAPI/java.util/Arra	-	1	-	-	4	-	4	-	3	-	4	-	-	-	-	-	-	-	4	-
http://www.java2s.com/Code/JavaAPI/java.util/Arra	-	4	-	-	4	-	4	-	4	-	4	-	-	-	-	-	-	-	3	-
http://www.javafaq.nu/java-example-code-87.html	4	-	-	-	-	4	-	-	2	-	-	-	1	-	-	-	4	0	-	-
http://www.javafaq.nu/java-example-code-list.html	4	-	-	-	-	4	-	-	3	-	-	-	4	-	-	-	3	2	-	-
http://japan.internet.com/developer/20060314/prin	-	0	-	-	4	-	0	-	-	-	4	-	-	-	-	-	-	-	0	-
http://java.sun.com/developer/TechTips/1999/tt080	4	-	-	-	-	3	-	-	3	-	-	-	2	-	-	-	-	2	-	-
http://www.j2mepolish.org/	-	0	0	-	-	-	0	-	-	-	0	-	-	-	-	-	-	-	0	-
http://www.javafaq.nu/java-example-code-89.html	4	-	4	-	-	4	-	-	-	-	-	-	-	-	-	-	4	4	-	-
http://www.leepoint.net/notes-java/	4	-	4	-	-	-	-	-	3	-	-	-	-	-	-	-	4	1	-	-
http://www.theserverside.de/java-generics-generisch	-	0	-	-	4	-	0	-	-	-	4	-	-	-	-	-	-	-	0	-

B.2.4 Question D

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www-28.ibm.com/developerworks/linux/libra	-	4	4	0	-	0	1	-	-	4	4	0	-	2	1	2	-	1	2	1
http://www.webinade.com/web-development/creat	-	3	0	0	-	0	0	-	-	1	1	0	-	0	0	0	-	0	1	0
http://thedailywtf.com/forums/thread/89324.aspx	-	1	4	0	-	0	1	-	-	1	1	3	-	3	2	2	-	1	2	-
http://en.wikipedia.org/wiki/Recursive_descent_pars	-	1	1	0	-	0	0	-	-	-	1	0	-	0	2	-	-	1	-	0
http://joseph.randomnetworks.com/archives/2005/(	-	1	0	0	-	0	0	-	-	-	0	0	-	0	0	-	-	0	-	0
http://leepoint.net/notes-java/	-	3	3	-	-	4	1	4	-	-	4	4	-	4	3	-	-	1	-	4
http://msdn2.microsoft.com/en-us/library/ms18624	-	1	0	0	-	0	0	-	-	-	0	0	-	0	0	-	-	0	-	0
http://www.leepoint.net/notes-java/	-	3	3	-	-	4	1	4	-	-	4	4	-	4	3	-	-	1	-	4
http://www.oreillynet.com/onjava/blog/2006/03/rec	-	1	0	0	-	0	0	-	-	-	1	0	-	2	0	-	-	2	-	0
http://www.theserverside.net/tt/articles/showarticle	-	1	0	0	-	0	0	-	-	-	0	0	-	0	0	-	-	0	-	0
http://www-28.ibm.com/developerworks/java/librar	-	3	3	2	-	0	3	-	-	-	4	4	-	1	1	-	-	1	-	-
http://www.cs.princeton.edu/introcs/23recursion/	-	4	4	3	-	4	4	-	-	-	4	4	-	2	4	-	-	4	-	-
http://www.cs.princeton.edu/introcs/27recursion/	-	4	4	3	-	4	3	-	-	-	4	4	-	4	4	-	-	4	-	-
http://www.ibm.com/developerworks/java/library/j-	-	1	3	2	-	3	4	-	-	-	4	4	-	1	2	-	-	1	-	-
http://www.mssqltips.com/tip.asp?tip=938	-	0	0	0	-	0	0	-	-	-	0	0	-	0	-	-	-	0	-	0
http://lists.w3.org/Archives/Public/www-dom/1998(	-	0	2	0	-	-	0	-	-	-	4	0	-	1	-	-	-	0	-	0
http://www.cs.dartmouth.edu/farid/teaching/cs15/c	-	2	4	0	-	4	4	-	-	-	3	3	-	0	-	-	-	4	-	-
http://thedailywtf.com/forums/89353/ShowThread.a	-	1	4	-	-	-	2	-	-	-	0	-	-	3	-	2	-	3	2	-
http://www.setfocus.com/TechnicalArticles/sql-serve	-	1	0	0	-	-	0	-	-	-	0	0	-	-	-	-	-	0	-	0
http://www.sitepoint.com/article/hierarchical-data-c	-	0	-	-	-	-	-	-	-	1	1	0	-	0	-	0	-	0	-	0
http://en.wikipedia.org/wiki/Recursion_(computer_s	-	1	-	0	-	-	-	-	-	-	1	0	-	-	-	-	-	1	4	1
http://etutorials.org/Programming/Java+performanc	-	3	-	3	-	-	-	-	-	-	4	4	-	-	-	-	-	4	-	1
http://www.behindthesite.com/blog/C1931765677/I	-	-	-	0	-	0	1	-	-	-	-	0	-	-	0	-	-	-	-	0
http://www.ibm.com/developerworks/linux/library/l	-	-	4	3	-	0	3	-	-	-	-	-	-	1	4	-	-	-	-	-
http://www.wwwcoder.com/main/parentid/191/site	-	1	-	0	-	-	-	-	-	-	0	0	-	-	-	-	-	0	-	0
http://leepoint.net/notes-java/index.html	-	-	-	-	-	4	1	-	-	-	-	4	-	-	3	-	-	-	-	4
http://remus.rutgers.edu/cs111/2007/summer/texts	-	4	3	-	-	-	-	-	-	-	4	-	-	0	-	-	-	1	-	-
http://thedailywtf.com/forums/89353/ShowThread.a	-	-	-	0	-	0	-	-	-	0	-	3	-	-	2	-	-	-	-	-
http://www.behindthesite.com/blog/C1931765677/I	-	0	0	-	-	-	-	-	-	-	0	-	-	0	-	-	-	1	-	-
http://www.gnu.org/software/qexo/	-	0	0	-	-	-	-	-	-	-	1	-	-	1	-	-	-	0	-	-
http://www.idevelopment.info/data/Programming/ja	-	-	1	-	-	0	0	-	-	-	-	-	-	-	1	-	-	-	-	4
http://www.java2s.com/	-	-	0	-	-	0	0	4	-	-	-	-	-	-	-	-	-	-	-	0
http://www.leepoint.net/notes-java/index.html	-	-	-	-	-	4	1	-	-	-	-	4	-	-	3	-	-	-	-	4
http://www.ozonehouse.com/ContextFree/	-	-	0	-	-	0	0	-	-	-	-	-	-	0	0	-	-	-	-	-

B.2.5 Question E

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.cs.ubc.ca/spider/harrison/Java/sorting-d	4	-	4	3	4	-	4	4	4	-	3	4	4	-	4	3	3	-	4	3
http://www.cs.ubc.ca/~harrison/Java/sorting-demo.h	4	-	4	3	4	-	4	4	4	-	3	4	4	-	4	2	4	-	3	3
http://cg.scs.carleton.ca/~morin/misc/sortalg/	4	-	1	2	4	-	4	4	4	-	4	4	-	-	4	2	2	-	3	4
http://en.wikipedia.org/wiki/Sorting_algorithm	4	-	4	3	4	-	4	4	4	-	4	4	-	-	4	4	4	-	4	3
http://www.cs.rit.edu/~atk/Java/Sorting/sorting.htm	4	-	4	2	4	-	4	4	4	-	4	4	-	-	4	2	4	-	3	3
http://critticall.com/	-	-	0	0	1	-	0	4	0	-	0	0	-	-	3	0	0	-	0	0
http://www.answers.com/topic/sorting-algorithm	4	-	4	1	4	-	4	-	4	-	4	4	-	-	3	4	4	-	4	4
http://www.softpanorama.org/Algorithms/sorting.sh	2	-	4	-	4	-	4	4	4	-	4	4	-	-	4	4	4	-	4	1
http://www.algosort.com/	3	-	-	-	4	-	4	3	3	-	-	1	0	-	4	3	4	-	3	1
http://www.cs.bu.edu/teaching/alg/sort/demo/	4	-	4	-	4	-	4	4	4	-	2	4	-	-	-	2	3	-	3	3
http://www.datastructures.info/	4	-	-	3	4	-	4	4	4	-	-	4	-	-	4	4	4	-	4	-
http://www.davekoelle.com/alphanum.html	-	-	-	0	3	-	3	-	4	-	0	4	4	-	2	0	3	-	2	-
http://www.geocities.com/wezam/sort22.html	-	-	-	-	4	-	4	-	4	-	-	4	4	-	4	3	4	-	3	4
http://www.inf.fh-flensburg.de/lang/algorithmen/so	-	-	-	0	4	-	4	-	3	-	4	3	-	-	4	3	4	-	4	-
http://dmoz.org/Computers/Algorithms/Sorting_and	4	-	-	-	4	-	4	-	3	-	-	1	-	-	4	4	1	-	1	-
http://www.codinghorror.com/blog/archives/00101	-	-	-	0	2	-	3	-	-	-	0	2	-	-	0	0	0	-	0	-
http://www.google.com/alpha/Top/Computers/Algo	2	-	-	-	4	-	3	-	3	-	-	1	-	-	4	3	1	-	3	-
http://www.google.com/Top/Computers/Algorithms	2	-	-	-	4	-	4	-	3	-	-	1	-	-	4	3	2	-	3	-
http://www.cs.ubc.ca/spider/harrison/Java/	-	-	-	3	4	-	4	-	4	-	-	4	-	-	-	2	3	-	3	-
http://www.idiotworld.com/story/258/5_algorithms	-	-	-	-	4	-	4	-	4	-	-	4	1	-	4	4	-	-	3	-
http://www.sysarch.com/Perl/sort_paper.html	-	-	-	-	4	-	3	-	4	-	-	4	-	-	3	4	4	-	2	-
http://www2.hig.no/~algmet/animate.html	0	-	-	-	3	-	3	-	3	-	4	3	-	-	-	3	-	-	3	-
http://math.hws.edu/TMCM/java/labs/xSortLabLab.h	-	-	-	-	4	-	4	-	-	-	4	4	-	-	-	4	4	-	4	-
http://www.awprofessional.com/bookstore/product	1	-	-	-	4	-	4	-	1	-	-	1	-	-	-	0	-	-	2	-
http://www.coyotesong.com/sort/	-	-	2	-	4	-	-	-	4	-	4	-	-	-	4	4	-	-	4	-
http://www.ddj.com/dept/cpp/184402000	-	-	-	-	4	-	4	-	-	-	-	4	-	-	4	2	-	-	3	-
http://www.awprofessional.com/bookstore/product	-	-	-	-	3	-	3	-	-	-	-	1	-	-	-	0	-	-	2	-
http://www.awprofessional.com/titles/0-201-31452-	-	-	-	-	3	-	4	-	-	-	-	1	-	-	-	0	-	-	2	-
http://www.educypedia.be/education/mathematicsj	-	-	4	-	-	-	-	-	-	-	4	-	-	-	4	-	4	-	-	4

B.2.6 Question F

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://java.sun.com/docs/books/tutorial/java/javaO(	4	-	-	4	4	4	-	4	4	3	-	4	4	4	-	4	4	4	-	3
http://blog.csdn.net/ladofwind/archive/2006/06/05/	1	-	-	0	4	-	-	-	0	-	-	0	-	0	-	-	0	4	-	0
http://blog.zerosum.org/2007/11/22/ruby-method-v	4	-	-	0	4	-	-	-	4	-	-	3	-	3	-	-	4	2	-	1
http://members.tripod.com/~MoisesRBB/java3.htm	0	-	-	0	4	-	-	-	0	-	-	0	-	0	-	-	0	0	-	0
http://mindprod.com/jgloss/privatescope.html	4	-	-	2	4	-	-	-	4	-	-	4	-	2	-	-	0	2	-	4
http://www.extreme-java.de/junitx	1	-	-	0	0	-	-	-	0	-	-	0	-	1	-	-	0	0	-	0
http://www.htmlgoodies.com/primers/jsp/article.ph	0	-	-	0	0	-	-	-	1	-	-	1	-	3	-	-	0	0	-	0
http://blog.jonudell.net/2007/03/27/authenticated-r	0	-	-	0	0	-	-	-	0	-	-	0	-	-	-	-	0	0	-	0
http://www.crockford.com/javascript/private.html	0	-	-	0	0	-	-	-	1	-	-	1	-	-	-	-	1	0	-	3
http://www.jchq.net/tutorial/01_02Tut.htm	4	-	-	3	4	-	-	-	-	-	-	4	-	-	-	3	4	4	-	4
http://www.uni-bonn.de/~manfear/javaprotection.p	-	-	-	-	-	0	-	4	-	2	-	-	4	4	-	4	-	4	-	4
http://javascript.crockford.com/private.html	0	-	-	0	0	-	-	-	1	-	-	-	-	-	-	-	2	0	-	2
http://www.dustindiaz.com/javascript-private-public	0	-	-	0	0	-	-	-	1	-	-	-	-	-	-	-	1	0	-	0
http://www.litotes.demon.co.uk/js_info/private_stat	0	-	-	0	-	-	-	-	1	-	-	1	-	2	-	-	-	0	-	0
http://www.unix.org.ua/orelly/java-ent/jnut/ch26_0	0	1	-	0	0	-	-	-	1	-	-	0	-	-	-	-	0	-	-	-
http://java.sun.com/j2se/1.5.0/docs/api/java/lang/re	-	-	-	-	-	1	-	4	-	-	-	-	1	1	-	-	-	3	-	0
http://uk.php.net/manual/en/language.oop5.visibilit	4	-	-	0	0	-	-	-	1	-	-	0	-	-	-	-	2	-	-	-
http://www.faqts.com/knowledge_base/view.phtml/	-	-	-	0	-	-	-	-	4	-	-	-	4	4	-	-	-	4	-	3
http://www.uni-bonn.de/~manfear/javamodifiers.ph	-	-	-	-	-	2	-	4	-	-	-	-	4	-	-	4	-	4	-	4
http://www.whalin.com/memcached/javadocs/com/	0	-	-	0	0	-	-	-	1	-	-	0	-	-	-	-	0	-	-	-
http://uic.rsu.ru/doc/programming/java/TIJE.ru/Chap	0	-	-	0	-	-	-	-	0	-	-	0	-	-	-	-	0	-	-	-
http://www.litotes.demon.co.uk/js_info/private_stat	-	-	-	0	0	-	-	-	1	-	-	1	-	-	-	-	0	-	-	-
http://www.tech-recipes.com/_tips1135.html	-	-	-	0	0	-	-	-	0	-	-	-	-	-	-	-	-	0	-	0
http://www.uic.rsu.ru/doc/programming/java/TIJE.ru	0	-	-	0	-	-	-	-	0	-	-	0	-	-	-	-	0	-	-	-

B.2.7 Question G

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.jetbrains.com/idea/	4	4	4	-	4	2	4	-	4	2	4	-	4	1	4	-	1	-	4	-
http://www.netbeans.org/	4	4	4	-	4	4	4	-	4	-	4	-	4	3	4	-	1	3	3	-
http://springide.org/project	-	1	4	-	4	2	-	-	3	1	3	-	4	4	4	-	-	1	3	-
http://www.bluej.org/	1	-	1	-	3	0	-	-	-	1	0	-	2	1	4	-	2	-	3	-
http://www.eclipse.org/home/categories/languages.	4	3	4	-	4	-	-	-	3	2	1	-	-	4	4	-	2	-	4	-
http://www.myeclipseide.com/	4	-	4	-	4	3	-	-	-	1	4	-	4	2	4	-	-	3	4	-
http://www.netbeans.org/kb/kb.html	4	-	0	-	4	-	-	-	2	1	0	-	4	1	4	-	-	2	2	-
http://www.jetbrains.com/	4	-	4	-	4	3	-	-	-	-	2	-	4	2	4	-	-	-	2	-
http://www.netbeans.org/kb/articles/books.html	-	-	1	-	4	1	-	-	-	1	4	-	-	0	4	-	-	0	1	-
http://akamai.infoworld.com/pdf/special_report/200	-	4	4	-	-	-	-	-	4	3	4	-	-	4	4	-	-	2	-	-
http://developers.sun.com/prodtech/javatools/jscrea	-	-	4	-	4	1	-	-	-	-	2	-	-	4	2	-	2	-	4	-
http://www.easyeclipse.org/site/home/	4	-	4	-	4	3	-	-	-	-	4	-	-	-	-	-	0	4	4	-
http://beust.com/weblog/archives/000369.html	-	-	0	-	0	0	-	-	1	-	0	-	0	-	-	-	-	-	0	-
http://blogs.sun.com/dannycoward/entry/java_se_6	-	0	0	-	-	-	-	-	0	0	2	-	-	0	1	-	-	-	-	-
http://java.sun.com/j2se/1.5.0/docs/relnotes/featur	-	3	0	-	-	-	-	-	0	0	0	-	-	2	1	-	-	-	-	-
http://java.sun.com/javaone/sf/javauniversity.jsp	-	1	0	-	-	-	-	-	-	2	1	-	-	3	0	-	1	-	-	-
http://tech.puredanger.com/java7	-	4	2	-	-	-	-	-	0	1	1	-	-	3	1	-	-	-	-	-
http://tnlessone.wordpress.com/2007/02/28/ruby-ra	-	-	4	-	-	-	-	-	3	4	2	-	-	3	4	-	-	2	-	-
http://www.javapassion.com/netbeans/	-	-	0	-	0	-	-	-	-	-	0	-	-	0	0	-	-	0	0	-
http://java.sun.com/developer/technicalArticles/too	-	-	4	-	-	-	-	-	-	3	4	-	-	2	4	-	4	-	-	-
http://java.sun.com/features/2003/05/bloch_qa.htm	-	1	0	-	-	-	-	-	-	2	1	-	-	0	1	-	-	-	-	-
http://lifeonrails.org/2007/8/27/netbeans-the-best-r	-	-	0	-	-	-	-	-	-	-	0	-	0	0	0	-	-	0	-	-
http://www.eclipse.org/	4	-	-	-	4	3	4	-	-	-	-	-	-	-	-	-	1	-	3	-
http://www.eclipse.org/downloads/	4	-	-	-	4	3	-	-	-	-	-	-	4	-	-	-	2	-	2	-
http://www.eclipseplugincentral.com/	4	-	-	-	4	0	4	-	-	-	-	-	-	-	-	-	-	3	2	-
http://www.netbeans.org/kb/articles/import-jbuilde	-	-	4	-	4	1	-	-	-	-	4	-	-	4	-	-	-	-	2	-
http://cafe.elharo.com/java/type-inference-another-	-	-	0	-	-	-	-	-	0	-	0	-	-	0	0	-	-	-	-	-
http://java.sun.com/j2se/1.5.0/docs/guide/language	-	0	0	-	-	-	-	-	-	-	0	-	-	1	1	-	-	-	-	-
http://mashable.com/2007/11/17/ide-toolbox	-	-	4	-	-	-	-	-	-	4	4	-	-	3	4	-	-	-	-	-
http://today.java.net/pub/a/today/2007/08/09/look	-	-	0	-	-	-	-	-	1	-	0	-	-	2	1	-	-	-	-	-
http://wiki.netbeans.org/wiki/view/Ruby	-	-	0	-	-	-	-	-	-	-	4	-	-	4	4	-	-	4	-	-
http://www.easyeclipse.org/site/distributions/	-	-	4	-	4	-	-	-	-	-	2	-	-	-	4	-	-	-	3	-
http://www.ibm.com/developerworks/opensource/li	-	-	0	-	-	-	-	-	2	-	4	-	4	-	4	-	-	-	-	-
http://www.netbeans.org/index.html	4	-	-	-	-	3	-	-	-	2	-	-	4	-	-	-	0	-	-	-
http://www.netbeans.org/kb/50/flash.html	-	-	4	-	-	-	-	-	-	2	4	-	-	4	4	-	-	-	-	-
http://www.onjava.com/pub/a/onjava/2006/06/28/	-	-	4	-	-	-	-	-	-	-	3	-	-	3	4	-	-	4	-	-

B.2.8 Question H

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://today.java.net/pub/a/today/2006/04/06/exce	-	4	4	2	-	3	4	3	-	-	4	4	-	2	4	3	-	4	3	-
http://devDev.bea.com/pub/a/2006/11/effective-ex	-	1	4	1	-	3	4	-	-	-	-	2	-	3	4	3	-	2	2	-
http://www.onjava.com/pub/a/onjava/2003/11/19/	-	3	4	2	-	0	4	-	-	-	-	4	-	3	-	4	-	4	3	-
http://today.java.net/pub/a/today/2003/12/04/exce	-	3	-	2	-	3	4	-	-	-	-	-	-	3	4	4	-	4	4	-
http://www.faqs.org/docs/think_java/TIJ311.htm	-	4	-	3	-	-	4	4	-	-	4	4	-	-	-	4	-	-	4	3
http://www.mindview.net/Etc/Discussions/CheckedE	-	2	-	3	-	4	4	-	-	-	-	-	-	4	4	3	-	3	2	-
http://java.sun.com/docs/books/tutorial/essential/e	-	1	4	-	-	-	-	3	-	-	4	-	-	3	-	4	-	4	4	-
http://ww2.cis.temple.edu/sorkin/ExceptionHandling	-	4	4	4	-	-	4	4	-	-	4	4	-	-	-	-	-	-	-	2
http://www.adtmag.com/java/articleold.aspx?id=124	-	-	4	3	-	-	4	-	-	4	-	3	-	4	-	3	-	-	1	-
http://www.andreashalter.ch/phpug/20040115/	-	1	0	0	-	-	3	1	-	-	3	1	-	-	-	-	-	-	-	1
http://www.artima.com/designtechniques/exceptior	-	3	4	-	-	-	-	4	-	-	4	4	-	4	4	-	-	-	-	2
http://www.idi.ntnu.no/grupper/su/fordypningspros	-	-	-	4	-	-	4	4	-	-	4	-	-	-	4	4	-	-	4	0
http://www.javaolympus.com/J2SE/Exceptions/JavaE	-	-	4	1	-	2	4	-	-	-	-	-	-	-	-	3	-	3	3	4
http://devDev.bea.com/pub/a/2006/11/effective-ex	-	1	-	-	-	0	-	-	-	-	-	-	-	2	4	3	-	4	2	-
http://www.janeg.ca/scjp/flow/try.html	-	-	4	2	-	-	4	4	-	-	4	4	-	-	-	-	-	-	-	4
http://netevil.org/blog/2004/may/structured-errors-	-	-	0	-	-	-	0	1	-	-	0	0	-	-	-	-	-	-	-	0
http://pages.cs.wisc.edu/~hasti/cs368/JavaTutorial/N	-	-	4	4	-	-	-	4	-	-	4	4	-	-	-	-	-	-	-	3
http://www.faqs.org/docs/javap/c9/s3.html	-	-	4	-	-	-	4	4	-	-	4	4	-	-	-	-	-	-	-	4
http://www.symfony-project.com/snippets/snippets/	-	-	0	-	-	-	0	2	-	-	0	0	-	-	-	-	-	-	-	0
http://www.topxml.com/javascript/javascript_error.a	-	-	0	-	-	-	0	1	-	-	3	1	-	-	-	-	-	-	-	1
http://developer.mozilla.org/en/docs/Core_JavaScrip	-	-	0	-	-	-	0	1	-	-	4	1	-	-	-	-	-	-	-	-
http://java.sun.com/docs/books/tutorial/essential/e	-	-	-	1	-	-	4	-	-	-	-	3	-	-	4	-	-	-	-	4
http://java.sun.com/docs/books/tutorial/essential/e	-	-	4	-	-	-	-	4	-	-	4	4	-	-	-	-	-	-	-	4
http://pages.cs.wisc.edu/~cs302/io/Exceptions.html	-	-	4	-	-	-	4	4	-	-	4	-	-	-	-	-	-	-	-	4
http://weblogs.asp.net/fmarguerie/archive/2007/12/	-	-	0	-	-	-	-	2	-	-	2	0	-	-	-	-	-	-	-	1
http://www.developer.com/java/article.php/10922_	-	-	-	-	-	-	4	4	-	-	4	4	-	4	-	-	-	-	-	-
http://www.javascriptkit.com/javatutors/conditional	-	-	0	-	-	-	-	1	-	-	1	1	-	-	-	-	-	-	-	1
http://www.javaworld.com/javaworld/jw-07-2005/jv	-	-	-	-	-	-	4	-	-	-	-	-	-	2	-	2	-	4	1	-
http://www.manageability.org/blog/stuff/exceptiona	-	-	-	-	-	-	4	-	-	-	-	-	-	4	-	3	-	4	2	-
http://www.programfan.com/article/showarticle.asp	-	-	0	-	-	-	0	0	-	-	4	-	-	-	-	-	-	-	-	0
http://www.rooftopsolutions.nl/article/126	-	-	0	-	-	-	-	0	-	-	0	0	-	-	-	-	-	-	-	0
http://www.w3schools.com/js/js_try_catch.asp	-	-	0	-	-	-	-	1	-	-	1	1	-	-	-	-	-	-	-	1

B.3 EXPERIMENT 3

B.3.1 Question A

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.javacoffeebreak.com/java103/java103.h	4	-	2	4	-	-	3	4	4	-	4	4	4	-	-	4	-	-	4	4
http://java.sun.com/docs/books/tutorial/essential/ic	0	-	1	0	-	-	-	2	-	-	2	0	0	-	0	0	-	-	4	1
http://www.idevelopment.info/data/Programming/j	3	-	1	4	-	-	1	3	4	-	-	-	3	-	3	3	-	-	3	2
http://java.sun.com/developer/onlineTraining/Progr	4	-	0	4	2	-	1	4	4	-	-	3	-	-	-	4	4	-	-	-
http://java.sun.com/j2se/1.4.2/docs/api/java/io/File	4	-	-	4	-	-	2	4	3	-	4	4	4	-	-	-	-	-	-	4
https://java.sun.com/j2se/1.4.2/docs/api/java/io/File	4	-	-	4	-	-	2	4	4	-	4	4	4	-	-	-	-	-	-	4
http://www.myjavatools.com/projects/v.6.0/lib/doc/	-	-	1	-	-	-	-	-	-	-	3	3	3	-	3	-	2	-	0	4
http://developers.sun.com/mobility/javacard/article	-	-	0	-	-	-	-	-	-	-	0	1	1	-	1	-	-	-	0	0
http://developers.sun.com/techtopics/mobility/javac	-	-	0	-	-	-	-	-	-	-	0	1	1	-	1	-	-	-	0	0
http://java.sun.com/docs/books/tutorial/	2	-	-	2	3	-	1	2	2	-	-	-	-	-	-	2	-	-	-	-
http://www.acm.org/crossroads/xrds6-3/ovp63.html	4	-	-	4	2	-	2	4	4	-	-	-	-	-	-	4	-	-	-	-
http://www.cs.usfca.edu/~parrt/course/601/lectures	4	-	4	4	-	-	3	4	4	-	-	4	-	-	-	-	-	-	-	-
http://www.dickbaldwin.com/java/Java829.htm	3	-	-	3	-	-	0	3	3	-	-	3	-	-	-	3	-	-	-	-
http://www.javapractices.com/Topic42.cjp	4	-	4	4	-	-	3	4	4	-	-	4	-	-	-	-	-	-	-	-
http://developers.sun.com/jscreator/learning/tutoria	1	-	-	2	-	-	0	1	1	-	-	-	-	-	-	2	-	-	-	-
http://developers.sun.com/prodtech/javatools/jscrea	2	-	-	2	-	-	0	2	2	-	-	-	-	-	-	2	-	-	-	-
http://forum.java.sun.com/thread.jspa?messageID=3	3	-	-	4	-	-	3	4	4	-	-	-	-	-	-	3	-	-	-	-
http://forum.java.sun.com/thread.jspa?threadID=67	2	-	-	2	-	-	2	2	2	-	-	-	-	-	-	2	-	-	-	-
http://search400.techtarget.com/tip/0,289483,sid3_	1	-	-	1	-	-	0	1	1	-	-	-	-	-	-	1	-	-	-	-
http://www.cookienest.com/content/javabasics-writ	1	-	-	1	-	-	0	1	1	-	-	-	-	-	-	1	-	-	-	-
http://www.exampledepot.com/egs/java.io/WriteTo	4	-	-	4	-	-	3	2	3	-	-	-	-	-	-	-	-	-	4	-
http://www.javabeat.net/tips/java/2007/08/recursiv	1	-	-	1	-	-	0	2	1	-	-	-	-	-	-	2	-	-	-	-
http://www.javapractices.com/topic/TopicAction.do	4	-	-	4	-	-	3	2	4	-	-	-	-	-	-	4	-	-	-	-
http://cs.middlesexcc.edu/~schatz/csc211/handouts/	3	-	-	4	-	-	3	4	4	-	-	-	-	-	-	-	-	-	-	-
http://e-docs.bea.com/wls/docs92/webserv/use_cas	-	-	0	-	-	-	-	-	-	-	0	-	1	-	-	-	0	-	-	0
http://edndoc.esri.com/arcobjects/9.2/java/api/arco	-	-	1	-	-	-	-	-	-	-	0	1	1	-	-	-	-	-	-	1
http://edocs.bea.com/wls/docs100/webserv/use_cas	-	-	0	-	-	-	-	-	-	-	0	-	0	-	-	-	0	-	-	0
http://home.cogeco.ca/~ve3ll/jatutor9.htm	4	-	-	4	-	-	3	4	4	-	-	-	-	-	-	-	-	-	-	-
http://java.sun.com/developer/JDCTechTips/2003/tt	-	-	0	-	1	-	-	-	-	-	0	-	2	-	-	-	-	-	-	1
http://java.sun.com/j2se/1.4.2/docs/api/java/io/pac	-	-	3	-	-	-	-	-	-	-	4	-	2	-	-	-	3	-	-	3
http://pages.cs.wisc.edu/~cs302/io/JavaIO.html	-	-	2	-	-	-	-	-	-	-	0	4	4	-	-	-	-	-	-	4
http://searchdomino.techtarget.com/tip/0,289483,si	0	-	-	0	-	-	1	1	0	-	-	-	-	-	-	-	-	-	-	-
http://www.abbeyworkshop.com/howto/java/writeT	4	-	-	4	-	-	4	4	4	-	-	-	-	-	-	-	-	-	-	-
http://www.devdaily.com/blog/post/java/java-faq-h	4	-	-	-	-	-	0	4	4	-	-	-	-	-	-	4	-	-	-	-
http://www.enete.com/noel/nuggets_java/	-	-	3	-	-	-	-	-	-	-	4	3	3	-	-	-	-	-	-	1
http://www.exampledepot.com/egs/java.io/ReadLin	3	-	-	-	-	-	2	4	4	-	-	-	-	-	3	-	-	-	-	-
http://www.java2s.com/Code/Php/File-Directory/Te	0	-	-	0	-	-	0	0	0	-	-	-	-	-	-	-	-	-	-	-
http://www.javacoffeebreak.com/faq/faq0004.html	4	-	-	4	-	-	-	4	4	-	-	4	-	-	-	-	-	-	-	-
http://www.javafaq.nu/java-example-code-983.html	3	-	-	4	-	-	0	4	3	-	-	-	-	-	-	-	-	-	-	-
http://www.mrx.net/java/program.html	4	-	-	4	-	-	0	4	4	-	-	-	-	-	-	-	-	-	-	-
http://www.nabble.com/Java-J2EE-Developers-f1969	0	-	-	0	-	-	0	1	0	-	-	-	-	-	-	-	-	-	-	-
http://www.nabble.com/Lucene-f44.html	0	-	-	0	-	-	2	0	0	-	-	-	-	-	-	-	-	-	-	-
http://www.stanford.edu/class/cs108/106a-java-han	4	-	-	4	-	-	3	4	4	-	-	-	-	-	-	-	-	-	-	-
http://www.theserverside.com/discussions/thread.ts	4	-	-	4	-	-	0	4	4	-	-	-	-	-	-	-	-	-	-	-
http://www.vipan.com/htdocs/log4jhelp.html	-	-	-	-	-	-	-	-	-	-	0	-	1	-	1	-	-	-	0	0

B.3.2 Question B

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://java.sun.com/j2se/1.3/docs/api/java/text/Dat	4	3	-	4	4	2	-	3	4	4	-	3	4	4	-	4	3	3	-	4
http://java.sun.com/j2se/1.4.2/docs/api/java/text/Si	4	3	-	4	4	2	-	4	4	4	-	4	4	4	-	4	4	4	-	4
http://java.sun.com/j2se/1.5.0/docs/api/java/text/Si	4	3	-	4	4	2	-	4	4	4	-	4	4	4	-	4	3	4	-	4
http://java.sun.com/javase/6/docs/api/java/text/Sim	4	3	-	3	4	2	-	3	4	4	-	4	4	4	-	4	4	4	-	4
http://java.boot.by/scjp-tiger/ch03s04.html	4	3	-	4	0	4	-	4	4	4	-	4	4	4	-	4	4	-	-	3
http://java.sun.com/j2se/1.4.2/docs/api/java/text/D	4	3	-	4	3	-	-	3	4	4	-	4	4	4	-	4	3	3	-	4
http://java.sun.com/j2se/1.5.0/docs/api/java/text/D	4	3	-	4	4	-	-	3	4	4	-	4	4	4	-	4	3	3	-	4
http://www.exampledepot.com/egs/java.text/Forma	4	3	-	4	3	-	-	4	4	4	-	4	4	4	-	4	4	4	-	4
http://www.kickjava.com/492.htm	4	3	-	4	4	3	-	4	4	2	-	4	4	3	-	4	4	4	-	-
http://www.javatechniques.com/blog/dateformat-ar	4	4	-	4	4	-	-	4	4	4	-	4	4	4	-	4	4	4	-	-
http://www.javatechniques.com/public/java/docs/ba	4	4	-	4	4	-	-	4	4	4	-	4	4	4	-	4	4	4	-	-
http://java.sun.com/j2se/1.3/docs/api/java/text/Sim	4	-	-	4	4	-	-	-	4	-	-	4	4	4	-	4	3	3	-	4
http://www.java-forums.org/java-tutorials/2775-java	4	-	-	4	4	2	-	-	4	-	-	4	-	-	-	4	4	-	-	3
http://www.beginner-java-tutorial.com/java-date.htr	-	4	-	-	-	4	-	4	-	4	-	4	4	4	-	-	4	-	-	-
http://www.daniweb.com/code/snippet515.html	-	4	-	4	-	-	-	4	-	4	-	4	4	4	-	-	4	-	-	-
http://www.javaworld.com/jw-2-2000/jw-229-dates.	4	-	-	4	-	4	-	-	3	-	-	2	-	-	-	-	4	4	-	4
http://www.kickjava.com/524.htm	4	4	-	4	4	-	-	-	4	4	-	-	-	-	-	4	-	4	-	-
http://www.unix.org.ua/orelly/java-ent/jnut/ch22_0	3	-	-	-	3	2	-	-	3	-	-	3	-	-	-	3	3	-	-	2
http://forum.java.sun.com/thread.jspa?threadID=54	-	3	-	-	-	-	-	4	-	3	-	4	4	3	-	-	4	-	-	-
http://forum.java.sun.com/thread.jspa?threadID=52	1	-	-	-	1	-	-	-	1	-	-	-	-	-	-	3	0	0	-	-
http://www.exampledepot.com/egs/java.text/ParseI	3	-	-	4	3	-	-	-	4	-	-	-	-	-	-	4	-	3	-	-
http://www.javaworld.com/javaworld/javaqa/2001-(	-	1	-	-	-	-	-	4	-	0	-	-	4	4	-	-	4	-	-	-
http://www.jguru.com/faq/view.jsp?EID=422110	3	-	-	-	0	-	-	-	4	-	-	-	-	-	-	2	3	1	-	-
http://www.kodejava.org/examples/19.html	4	-	-	4	3	-	-	-	4	-	-	-	-	-	-	4	-	4	-	-
http://www.roseindia.net/struts/struts2/date/struts-	-	3	-	3	-	-	-	3	-	2	-	-	3	1	-	-	-	-	-	-
http://xml.apache.org/xalan-j/extensions.html	1	-	-	-	0	3	-	-	1	-	-	-	-	-	-	1	-	0	-	-
http://articles.techrepublic.com.com/5100-22-04641	-	2	-	4	-	-	-	4	-	4	-	-	-	-	-	-	-	-	-	4
http://coding.moris.org/archives/2003/08/23/forma	4	-	-	-	2	-	-	-	3	-	-	-	-	-	-	4	4	-	-	-
http://java.sun.com/j2se/1.4.2/docs/api/java/sql/Tin	-	-	-	-	-	1	-	-	-	-	-	3	-	-	-	-	3	3	-	1
http://java.sun.com/j2se/1.5.0/docs/api/java/sql/Tin	-	-	-	-	-	2	-	-	-	-	-	2	-	-	-	-	2	2	-	1
http://www.dreamincode.net/forums/showtopic148	-	2	-	-	-	-	-	1	-	3	-	-	1	4	-	-	-	-	-	-
http://www.java2s.com/Code/Java/Development-Cla	-	1	-	-	-	-	-	2	-	2	-	-	1	3	-	-	-	-	-	-
http://www.javafaq.nu/java-example-code-list.html	3	-	-	4	4	-	-	-	3	-	-	-	-	-	-	3	-	-	-	-

B.3.3 Question C

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.java-examples.com/java-collections-and	4	3	4	-	4	4	3	-	4	4	4	-	4	4	4	-	4	4	4	-
http://www.javadeveloper.co.in/java-example/java-a	4	4	4	-	4	4	4	-	4	4	4	-	4	4	-	-	4	4	4	-
http://www.kickjava.com/220.htm	4	3	3	-	4	2	3	-	4	4	4	-	4	4	-	-	4	4	4	-
http://www.anyexample.com/programming/java/jav	4	2	4	-	4	4	4	-	4	-	2	-	3	4	-	-	4	4	4	-
http://www.javafaq.nu/java-example-code-89.html	4	2	4	-	4	4	-	-	4	-	0	-	4	4	4	-	4	4	4	-
http://eclipsetutorial.sourceforge.net/Total_Beginne	3	4	3	-	-	3	4	-	3	4	2	-	3	4	-	-	3	-	2	-
http://java.sun.com/developer/JDCTechTips/2002/tt	3	2	3	-	2	1	4	-	4	-	3	-	3	4	-	-	4	-	0	-
http://java.sun.com/developer/technicalArticles/rele	2	1	0	-	1	1	-	-	1	4	2	-	1	4	-	-	2	-	0	-
http://www.beginner-java-tutorial.com/java-arraylist	4	3	3	-	4	-	3	-	4	4	4	-	4	4	-	-	4	-	4	-
http://www.java-samples.com/showtutorial.php?tut	4	3	4	-	4	3	-	-	4	4	4	-	4	4	-	-	4	-	4	-
http://www.java2s.com/Code/JavaAPI/java.util/Arra	4	3	4	-	-	-	3	-	4	4	4	-	4	4	-	-	3	4	4	-
http://www.javabeginner.com/java-arraylist.htm	4	4	4	-	4	-	3	-	4	4	3	-	4	4	-	-	4	-	4	-
http://www.wellho.net/resources/ex.php4?item=j71	4	0	4	-	4	-	3	-	4	-	2	-	3	4	-	-	4	4	3	-
http://www.developerzone.biz/index.php?option=co	4	4	4	-	-	-	-	-	4	4	4	-	2	4	-	-	4	-	3	-
http://java.sun.com/j2se/1.4.2/docs/api/java/util/Ar	-	3	4	-	-	-	3	-	4	-	4	-	4	-	2	-	-	3	0	-
http://users.cs.dal.ca/~sedgwick/ArrayList.html	4	2	4	-	4	-	-	-	4	-	-	-	3	4	-	-	3	-	4	-
http://www.idevelopment.info/data/Programming/j	4	-	2	-	4	3	-	-	-	4	-	-	-	4	-	-	4	4	4	-
http://www.javafaq.nu/java-example-code-list.html	3	-	4	-	4	3	-	-	-	-	-	-	-	4	3	-	4	4	4	-
http://www.leepoint.net/notes-java/data/collections	4	2	4	-	3	-	3	-	4	-	3	-	3	-	-	-	4	-	-	-
http://www.idevelopment.info/data/Programming/j	4	-	3	-	4	2	-	-	-	-	-	-	-	4	-	-	4	4	4	-
http://www.leepoint.net/notes-java/	4	-	4	-	-	3	-	-	-	-	-	-	-	4	3	-	4	3	4	-
http://java.sun.com/developer/technicalArticles/Coll	3	-	2	-	4	2	-	-	-	-	-	-	-	4	-	-	3	-	4	-
http://www.onjava.com/lpt/a/6014	4	-	0	-	3	2	-	-	-	-	-	-	-	4	-	-	4	-	0	-
http://blowed.serveusers.com/example-of-arraylist.h	0	-	0	-	-	-	-	-	-	-	-	-	-	0	-	-	0	0	0	-
http://j-integra.intrinsyc.com/support/espresso/doc/	-	0	-	-	-	-	0	-	0	-	0	-	0	-	-	-	-	1	-	-
http://verify.stanford.edu/uli/java_cpp.html	-	0	-	-	4	-	-	-	0	-	1	-	0	3	-	-	-	-	-	-
http://www.java2s.com/Code/JavaAPI/java.util/Arra	-	3	-	-	-	2	-	-	4	4	4	-	4	-	-	-	-	-	-	-
http://www.steaua.com/store/products_pictures/th	-	-	0	-	0	0	-	-	-	-	-	-	-	0	-	-	2	-	0	-
http://java.sun.com/developer/onlineTraining/new2j	-	4	-	-	-	-	3	-	2	-	2	-	1	-	-	-	-	-	-	-
http://java.sun.com/javase/6/docs/api/java/util/Arra	-	3	4	-	-	-	-	-	3	-	4	-	4	-	-	-	-	-	-	-
http://www.actionhotdoggo.com/forums2/attachme	-	-	0	-	0	0	-	-	-	-	-	-	-	0	-	-	-	-	0	-
http://www.hznavi.com/news/archives/2004/cat17/	1	-	-	-	-	0	-	-	-	-	-	-	-	0	-	-	2	-	0	-
http://www.java-tips.org/java-se-tips/java.lang/use-c	-	3	-	-	-	-	4	-	3	-	3	-	3	-	-	-	-	-	-	-
http://www.javaworld.com/javaworld/javaqa/2001-(	-	0	-	-	-	-	0	-	1	-	1	-	1	-	-	-	-	-	-	-
http://www.lifestyletoolbox.com/admin/sales/snaps	1	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	3	-	0	-
http://www.mindprod.com/jgloss/arraylist.html	-	4	-	-	-	-	3	-	4	-	4	-	4	-	-	-	-	-	-	-

B.3.4 Question D

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://danzig.jct.ac.il/java_class/recursion.html	-	3	0	4	-	4	3	4	-	3	3	4	-	4	4	4	-	4	3	4
http://www.thedailywtf.com/forums/89353/ShowTh	-	2	0	1	-	0	3	1	-	2	0	1	-	1	1	2	-	1	0	0
http://www.thedailywtf.com/forums/thread/89324.a	-	1	0	1	-	1	3	1	-	0	2	1	-	1	1	2	-	1	0	1
http://www.ibm.com/developerworks/java/library/j-	-	0	0	4	-	2	1	-	-	2	-	4	-	4	4	3	-	2	3	4
http://chortle.ccsu.edu/CS151/cs151java.html	-	0	1	-	-	3	-	4	-	4	4	3	-	4	-	-	-	4	3	3
http://weblogs.java.net/blog/mortazavi/archive/200	-	-	3	4	-	-	-	3	-	4	2	4	-	4	-	4	-	3	4	2
http://www.ahmadsoft.org/articles/recursion/	-	-	0	4	-	-	-	4	-	4	2	4	-	4	4	-	-	3	4	3
http://www-28.ibm.com/developerworks/java/librar	-	0	3	4	-	3	1	-	-	3	-	-	-	4	4	4	-	-	-	4
http://www.cs.princeton.edu/introcs/23recursion/	-	4	4	4	-	4	4	-	-	4	-	-	-	4	4	4	-	-	-	4
http://www.cs.princeton.edu/introcs/27recursion/	-	4	4	4	-	4	0	-	-	4	-	-	-	4	4	4	-	-	-	4
http://www.tech-recipes.com/java_programming_tip	-	-	4	4	-	-	-	4	-	4	4	3	-	4	-	-	-	3	3	4
http://en.wikipedia.org/wiki/Recursion_(computer_s	-	-	-	4	-	-	-	-	-	1	-	4	-	2	4	-	-	4	1	3
http://www-28.ibm.com/developerworks/linux/libra	-	2	4	-	-	4	0	-	-	4	-	-	-	-	4	3	-	-	-	4
http://www.cs.dartmouth.edu/farid/teaching/cs15/c	-	0	4	4	-	3	4	-	-	-	-	-	-	4	4	4	-	-	-	-
http://www.ibm.com/developerworks/xml/library/x-	-	-	-	-	-	-	-	4	-	-	1	4	-	4	4	-	-	3	0	-
http://www.javaworld.com/javaworld/javatips/jw-ja	-	-	-	4	-	3	-	2	-	4	1	-	-	-	4	-	-	-	-	0
http://www.leepoint.net/notes-java/	-	-	3	4	-	4	-	-	-	4	-	-	-	4	-	3	-	-	-	2
http://www.leepoint.net/notes-java/data/numbers/(	-	-	3	2	-	-	-	-	-	-	-	3	-	4	-	4	-	3	0	-
http://xanedu.proquest.com/originalworks/Prevac	-	-	-	1	-	-	-	0	-	-	0	0	-	0	-	-	-	1	0	-
http://en.wikipedia.org/wiki/Recursive_descent_pars	-	-	2	1	-	1	-	-	-	0	-	-	-	1	-	-	-	-	-	0
http://www-28.ibm.com/developerworks/java/librar	-	0	-	-	-	-	-	2	-	-	1	2	-	-	-	-	-	2	3	-
http://www.cafeaulait.org/javatutorial.html	-	-	3	-	-	3	-	-	-	-	-	1	-	4	-	-	-	4	4	-
http://www.google.com/search?q=recursion+in+java	-	0	-	-	-	-	-	4	-	-	4	3	-	-	-	-	-	3	2	-
http://www.greenleecds.com/javaap.html	-	1	-	-	-	-	-	3	-	-	3	3	-	-	-	-	-	3	4	-
http://www.hostitwise.com/java/java_recursion.htm	-	-	-	4	-	-	-	4	-	-	3	4	-	4	-	-	-	-	4	-
http://www.ibm.com/developerworks/linux/library/l	-	-	3	-	-	3	1	-	-	4	-	-	-	-	-	1	-	-	-	4
http://chortle.ccsu.edu/CS151/	-	-	-	-	-	-	-	4	-	-	4	3	-	-	-	-	-	4	4	-
http://chortle.ccsu.edu/java5/cs151java.html	-	-	-	-	-	-	-	3	-	-	4	3	-	-	-	-	-	4	3	-
http://www.brpreiss.com/books/opus5/programs/	-	2	3	-	-	3	3	-	-	-	-	-	-	-	-	4	-	-	-	-
http://www.cs.may.ie/~pgibson/Teaching/Schools/Ja	-	-	3	-	-	-	-	-	-	-	-	2	-	4	-	-	-	3	4	-
http://www.cs.umd.edu/class/fall2002/cmsc214/Tut	-	-	-	4	-	-	4	-	-	-	-	-	-	4	4	4	-	-	-	-
http://www.etutorials.org/Programming/Java+perfor	-	-	-	4	-	-	-	-	-	4	-	-	-	4	4	-	-	-	-	4
http://www.faqs.org/docs/javap/c11/s1.html	-	1	4	-	-	-	-	-	-	-	-	-	-	4	-	-	-	4	-	4
http://www.javabat.com/	-	-	-	-	-	-	-	3	-	-	3	4	-	-	-	-	-	4	0	-
http://www.leepoint.net/notes-java/io/10file/20recu	-	-	-	4	-	-	-	-	-	-	-	1	-	4	-	-	-	2	4	-

B.3.5 Question E

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.cs.rit.edu/~atk/Java/Sorting/sorting.htm	3	-	3	4	4	-	4	3	4	-	2	3	3	-	3	4	3	-	4	2
http://www.cs.ubc.ca/spider/harrison/Java/sorting-d	4	-	4	4	3	-	4	3	4	-	2	3	4	-	4	3	3	-	4	4
http://cg.scs.carleton.ca/~morin/misc/sortalg/	4	-	3	4	3	-	3	4	4	-	2	4	3	-	-	4	4	-	4	4
http://en.wikipedia.org/wiki/Sorting_algorithm	4	-	4	-	4	-	4	4	4	-	4	4	4	-	4	4	4	-	4	4
http://www.answers.com/topic/sorting-algorithm	4	-	4	4	4	-	3	4	4	-	4	4	4	-	3	-	4	-	4	0
http://www.cs.ubc.ca/~harrison/Java/sorting-demo.h	3	-	4	4	4	-	4	3	4	-	1	3	3	-	-	4	3	-	3	4
http://www.softpanorama.org/Algorithms/sorting.sh	4	-	4	-	3	-	2	4	2	-	4	3	2	-	3	4	4	-	3	3
http://www.datastructures.info/	3	-	4	-	4	-	4	3	3	-	4	3	4	-	4	-	3	-	3	3
http://www.davekoelle.com/alphanum.html	3	-	1	3	0	-	1	3	3	-	1	3	-	-	-	3	3	-	2	0
http://linux.wku.edu/~lamonml/algor/sort/sort.html	4	-	4	4	4	-	3	4	4	-	-	-	2	-	-	4	4	-	4	4
http://www.algosort.com/	3	-	4	-	4	-	3	3	3	-	4	3	4	-	-	-	4	-	3	1
http://www.dmoz.org/Computers/Algorithms/Sortin	4	-	4	-	2	-	4	3	4	-	4	3	-	-	-	3	2	-	4	3
http://www.google.com/alpha/Top/Computers/Algo	4	-	4	-	1	-	4	4	4	-	4	4	-	-	-	4	4	-	4	2
http://www.google.com/Top/Computers/Algorithms	4	-	4	-	3	-	3	4	4	-	4	4	-	-	-	4	4	-	4	3
http://www.coyotesong.com/sort/	4	-	3	-	3	-	4	4	4	-	2	-	-	-	4	-	4	-	4	4
http://www.cs.ubc.ca/spider/harrison/Java/	3	-	4	-	4	-	4	3	4	-	2	3	-	-	-	4	3	-	-	4
http://www.ddj.com/dept/cpp/184402000	3	-	1	3	1	-	2	3	3	-	0	-	-	-	-	4	3	-	-	0
http://www.geocities.com/wezam/sort22.html	3	-	0	-	0	-	2	3	3	-	-	3	-	-	3	-	3	-	1	4
http://maven.smith.edu/~thiebaut/java/sort/demo.h	2	-	3	-	2	-	4	2	1	-	2	-	-	-	-	-	3	-	2	4
http://linux.wku.edu/~lamonml/algor/sort/	4	-	4	-	4	-	4	4	4	-	4	-	-	-	-	-	4	-	-	4
http://max.cs.kzoo.edu/~abrady/java/sorting/	2	-	4	-	4	-	4	1	4	-	2	-	-	-	-	-	3	-	-	4
http://www.cs.bu.edu/teaching/alg/sort/demo/	3	-	-	-	2	-	4	3	3	-	2	3	-	-	-	4	3	-	-	-
http://www.cs.ubc.ca/~harrison/Java/	2	-	4	-	4	-	4	2	4	-	2	-	-	-	-	-	3	-	-	4
http://www.idiotworld.com/story/258/5_algorithms	4	-	-	-	3	-	3	3	3	-	4	-	3	-	-	-	3	-	3	-
http://www2.hig.no/~algmet/animate.html	3	-	-	-	4	-	1	3	3	-	4	-	-	-	3	-	3	-	2	-
http://atschool.eduweb.co.uk/mbaker/sorts.html	4	-	3	-	3	-	4	4	4	-	-	-	-	-	-	-	4	-	-	4
http://math.hws.edu/TMCM/java/labs/xSortLabLab.h	4	-	-	4	4	-	2	4	4	-	-	-	-	-	-	4	4	-	-	-
http://www.cs.hope.edu/alganim/ccaa/sorting.html	3	-	1	-	1	-	4	4	4	-	-	-	-	-	-	-	3	-	-	3
http://www.awprofessional.com/bookstore/product	1	-	-	-	0	-	0	0	1	-	-	-	0	-	-	-	1	-	-	-
http://www.awprofessional.com/bookstore/product	1	-	-	-	0	-	0	0	0	-	-	-	0	-	-	-	1	-	-	-
http://www.awprofessional.com/titles/0-201-31452-	1	-	-	-	0	-	0	0	0	-	-	-	0	-	-	-	1	-	-	-
http://www.brian-borowski.com/Sorting/	2	-	-	-	0	-	-	3	0	-	3	-	-	-	-	-	1	-	-	-
http://www.educypedia.be/education/mathematicsj	-	-	3	3	-	-	-	-	-	-	-	-	3	-	3	-	-	-	4	4
http://cs.smith.edu/~thiebaut/java/sort	2	-	4	-	-	-	-	2	2	-	-	-	-	-	-	-	3	-	-	-

B.3.6 Question F

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.codestyle.org/java/faq-Inheritance.shtm	3	2	-	4	4	1	-	3	4	3	-	4	3	4	-	3	4	4	-	2
http://www.allapplabs.com/interview_questions/jav	4	3	-	4	4	3	-	4	4	-	-	4	4	4	-	4	4	4	-	2
http://java.sun.com/j2se/1.5.0/docs/api/java/lang/re	-	3	-	-	4	1	-	2	3	4	-	1	3	4	-	-	0	1	-	4
http://www.jchq.net/tutorial/01_02Tut.htm	-	4	-	-	2	2	-	4	-	4	-	3	4	4	-	-	4	3	-	1
http://java.sun.com/docs/books/tutorial/java/javaOO	4	-	-	-	4	3	-	4	4	4	-	-	4	4	-	-	4	-	-	3
http://www.daimi.au.dk/dRegAut/JavaBNF.html	-	0	-	-	2	0	-	-	-	-	-	1	0	-	-	-	0	1	-	0
http://www.unf.edu/~rzucker/cop3540dir/modifiers.	4	4	-	-	4	-	-	-	-	-	-	4	-	4	-	4	-	4	-	4
http://www.faqts.com/knowledge_base/view.phtml/	-	-	-	-	3	1	-	-	-	4	-	-	4	4	-	-	4	-	-	2
http://java.sun.com/j2se/1.5.0/docs/api/java/lang/re	-	2	-	-	2	-	-	-	-	-	-	0	-	-	-	-	1	1	-	1
http://www.d116.com/hacks/emacs/java-flock.el	-	1	-	-	0	-	-	-	-	-	-	0	0	-	-	-	-	1	-	0
http://www.javacamp.org/javaI/modifier.html	-	3	-	-	4	-	-	-	-	-	-	3	4	-	-	-	-	3	-	4
http://www.landofcode.com/java/java-oop1.php	-	4	-	-	-	-	-	-	-	-	-	3	4	4	-	-	4	4	-	-
http://www.uni-bonn.de/~manfear/javamodifiers.ph	-	0	-	-	3	-	-	-	-	-	-	3	-	-	-	-	4	3	-	4
http://en.wikipedia.org/wiki/Java_syntax	-	3	-	-	2	-	-	-	-	-	-	3	-	-	-	-	-	3	-	3
http://interviewjava.blogspot.com/2007/04/what-ar	-	3	-	-	0	-	-	-	-	-	-	-	0	-	-	4	-	-	-	0
http://java.sun.com/docs/books/jls/second_edition/	-	3	-	-	-	-	-	-	-	-	-	3	-	4	-	-	4	3	-	-
http://java.sun.com/docs/books/jls/second_edition/	-	3	-	-	3	-	-	-	-	-	-	3	-	-	-	-	-	3	-	2
http://java.sun.com/docs/books/jls/third_edition/ht	-	3	-	-	3	-	-	-	-	-	-	3	-	-	-	-	-	3	-	2
http://java.sun.com/docs/books/jvms/second_editio	-	3	-	-	-	-	-	-	-	-	-	3	-	3	-	2	-	3	-	-
http://java.sun.com/docs/books/jvms/second_editio	-	2	-	-	3	-	-	-	-	-	-	3	-	-	-	-	-	3	-	2
http://java.sun.com/docs/books/vmspec/2nd-editior	-	1	-	-	3	-	-	-	-	-	-	3	-	-	-	-	-	3	-	2
http://www.csci.csusb.edu/dick/samples/java.glossa	-	-	-	-	-	2	-	3	-	4	-	-	3	-	-	-	4	-	-	-
http://www.geekinterview.com/question_details/56	4	-	-	4	-	-	-	-	4	-	-	-	-	4	-	4	-	-	-	-
http://www.meshplex.org/wiki/Java/Modifiers	-	0	-	-	1	-	-	-	-	-	-	0	-	-	-	-	-	1	-	0
http://www.unix.org.ua/orelly/java/langref/ch05_04	-	4	-	-	-	-	-	-	-	-	-	4	4	-	-	-	4	4	-	-

B.3.7 Question G

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://www.eclipse.org/home/categories/languages.	3	2	3	-	3	1	4	-	-	3	1	-	4	4	3	-	3	4	3	-
http://www.bluej.org/	4	1	0	-	3	1	0	-	-	3	0	-	4	4	-	-	2	4	4	-
http://www.netbeans.org/	3	1	-	-	4	1	3	-	4	4	3	-	4	4	3	-	-	3	4	-
http://www.easyeclipse.org/site/home/	3	2	2	-	3	1	3	-	-	3	3	-	3	4	-	-	-	3	3	-
http://java.sun.com/developer/technicalArticles/too	4	3	2	-	3	-	3	-	-	-	2	-	4	4	-	-	4	4	4	-
http://www.jcreator.com/	3	1	-	-	4	-	1	-	-	4	2	-	3	4	-	-	2	4	3	-
http://www.jetbrains.com/idea/	-	1	-	-	3	1	4	-	4	4	2	-	3	4	-	-	-	4	3	-
http://developers.sun.com/prodtech/javatools/jscrea	3	1	-	-	2	-	4	-	-	-	4	-	3	4	-	-	3	3	4	-
http://hossamahmed.wordpress.com/2006/09/13/ja	-	4	-	-	4	-	4	-	-	4	4	-	4	4	4	-	-	4	4	-
http://javaboutique.internet.com/demoIDEs/	4	4	4	-	-	4	4	-	-	4	4	-	-	-	4	-	4	-	4	-
http://www.apl.jhu.edu/~hall/java/IDEs.html	4	4	4	-	-	4	4	-	4	-	4	-	-	-	3	-	4	-	4	-
http://www.jetbrains.com/	-	2	-	-	2	2	3	-	-	4	4	-	4	4	-	-	-	4	4	-
http://www.myeclipseide.com/	-	1	-	-	3	1	4	-	-	2	2	-	4	4	-	-	-	4	1	-
http://www.netbeans.org/kb/articles/books.html	-	1	-	-	0	1	0	-	-	1	1	-	3	4	-	-	-	3	1	-
http://www.netbeans.org/kb/articles/import-jbuilde	-	1	-	-	2	-	4	-	4	3	0	-	3	4	-	-	-	3	3	-
http://www.easyeclipse.org/site/distributions/	-	2	-	-	1	1	3	-	-	-	1	-	3	4	-	-	-	4	1	-
http://www.netbeans.org/kb/kb.html	3	2	-	-	1	-	0	-	-	-	0	-	3	4	-	-	-	3	2	-
http://www.springide.org/project	-	1	-	-	3	-	2	-	-	4	1	-	4	4	-	-	-	4	1	-
http://www.stylusstudio.com/java_ide.html	-	2	-	-	4	-	3	-	-	2	3	-	4	4	-	-	-	3	3	-
http://www.eclipse.org/	3	2	-	-	-	1	4	-	-	3	1	-	-	-	-	-	-	-	4	-
http://www.eclipse.org/downloads/	-	2	3	-	-	1	4	-	-	-	2	-	-	-	-	-	3	-	2	-
http://www.netbeans.org/features/	3	-	-	-	4	-	-	-	-	4	-	-	3	4	-	-	3	4	-	-
http://www.thefreecountry.com/programming/javai	3	4	-	-	-	3	4	-	-	4	-	-	-	-	-	-	4	-	4	-
http://developers.sun.com/jsenterprise/features/	3	-	-	-	3	-	-	-	-	4	-	-	1	4	-	-	-	4	-	-
http://java.sun.com/javaone/sf/javauniversity.jsp	1	-	-	-	2	-	-	-	-	-	-	-	0	0	-	-	1	1	-	-
http://www.eclipseplugincentral.com/	-	1	-	-	-	1	3	-	-	2	1	-	-	-	-	-	-	-	2	-
http://www.mashable.com/2007/11/17/ide-toolbox	-	-	4	-	4	-	-	-	-	-	-	-	3	4	-	-	3	3	-	-
http://www.netbeans.org/features/java/	-	-	-	-	3	-	4	-	-	4	-	-	0	4	-	-	-	4	-	-
http://www.netbeans.org/kb/articles/learn-java.htm	-	1	-	-	-	1	4	-	3	-	0	-	-	-	-	-	-	-	3	-
http://eclipse-plugins.2y.net/eclipse/	-	2	-	-	-	-	3	-	-	3	0	-	-	-	-	-	-	-	2	-
http://forums.aptana.com/viewtopic.php?t=4028	-	1	-	-	-	-	0	-	-	1	0	-	-	-	-	-	-	-	0	-
http://javapowertools.wikidot.com/	-	2	-	-	-	1	1	-	-	-	0	-	-	-	-	-	-	-	1	-
http://jdee.sunsite.dk/	-	1	-	-	-	-	0	-	-	3	0	-	-	-	-	-	-	-	4	-
http://wiki.netbeans.org/wiki/view/NB6L10nKit	-	1	-	-	-	-	3	-	3	-	1	-	-	-	-	-	-	-	3	-
http://www.codegear.com/products/jbuilder	-	-	-	-	3	-	-	-	-	4	-	-	0	4	-	-	-	1	-	-
http://www.ibm.com/developerworks/opensource/li	-	-	-	-	4	-	-	-	-	4	-	-	1	3	-	-	-	4	-	-
http://www.javaworld.com/javaworld/jw-0-2002/jw-	4	-	-	-	4	-	-	-	-	-	-	-	3	4	-	-	-	3	-	-
http://www.netbeans.org/features/web/java-ee.htm	-	-	-	-	4	-	-	-	-	4	-	-	0	4	-	-	-	4	-	-
http://www.netbeans.org/kb/trails/java-se.html	-	2	-	-	-	-	3	-	3	-	0	-	-	-	-	-	-	-	2	-
https://supernova.dev.java.net/	-	-	-	-	4	-	-	-	-	4	-	-	3	4	-	-	-	4	-	-

B.3.8 Question H

Resource	Subject																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
http://today.java.net/pub/a/today/2006/04/06/exce	-	3	1	3	-	2	0	3	-	4	3	3	-	3	4	4	-	3	-	3
http://pages.cs.wisc.edu/~cs302/io/Exceptions.html	-	3	-	4	-	4	4	4	-	-	4	4	-	4	4	4	-	4	-	4
http://ww2.cis.temple.edu/sorkin/ExceptionHandling	-	4	4	4	-	4	3	-	-	4	4	4	-	-	4	4	-	4	-	4
http://neptune.netcomp.monash.edu.au/JavaHelp/h	-	3	4	4	-	3	4	-	-	4	4	-	-	-	4	-	-	4	4	4
http://www.artima.com/designtechniques/exceptior	-	4	4	4	-	4	3	-	-	4	4	4	-	-	4	-	-	3	-	4
http://java.sun.com/docs/books/tutorial/essential/e	-	4	4	2	-	4	3	-	-	4	-	-	-	-	-	4	-	4	4	4
http://sharat.wordpress.com/2007/05/16/exception	-	3	3	4	-	3	2	-	-	4	4	-	-	-	-	4	-	3	-	-
http://www.smartdataprocessing.com/lessons/l5.htr	-	-	4	4	-	4	4	-	-	4	4	1	-	-	4	-	-	4	-	-
http://today.java.net/pub/a/today/2003/12/04/exce	-	2	4	-	-	-	2	4	-	-	4	4	-	4	-	4	-	-	-	-
http://www.faqs.org/docs/think_java/TIJ311.htm	-	3	-	4	-	4	3	-	-	4	4	-	-	-	-	-	-	4	-	4
http://www.janeg.ca/scjp/flow/try.html	-	-	-	-	-	-	2	3	-	-	4	1	-	4	-	-	-	4	4	4
http://www.javalobby.org/java/forums/t105307.htm	-	4	0	4	-	2	4	-	-	-	-	4	-	-	-	4	-	4	-	-
http://www.sys-con.com/story/?storyid=38160	-	4	4	3	-	4	4	-	-	-	4	-	-	-	0	-	-	4	-	-
http://java.sun.com/docs/books/tutorial/essential/e	-	-	-	-	-	-	-	3	-	4	-	4	-	4	-	4	-	-	3	4
http://java.sys-con.com/read/38160.htm	-	-	4	4	-	4	3	-	-	-	4	-	-	-	3	-	-	4	-	-
http://www.cs.wisc.edu/~cs302/io/Exceptions.html	-	4	-	4	-	4	4	-	-	-	-	-	-	-	4	-	-	4	-	4
http://www.dreamincode.net/forums/showtopic226	-	-	4	3	-	2	-	4	-	-	-	4	-	-	4	-	-	4	-	-
http://www.javaworld.com/javaworld/jw-07-998/jw-	-	-	-	4	-	4	4	-	-	4	-	-	-	-	4	-	-	4	-	4
http://www.tech-recipes.com/java_programming_tip	-	-	-	-	-	2	-	2	-	4	4	-	-	4	-	-	-	-	3	4
http://fishbowl.pastiche.org/2003/03/12/wanted_m	-	0	2	3	-	1	3	-	-	-	-	-	-	-	-	-	-	2	-	-
http://www.tutorialhero.com/tutorial-73-catch_mult	-	-	-	2	-	3	4	-	-	-	4	3	-	-	-	-	-	4	-	-
http://dev2dev.bea.com/pub/a/2006/11/effective-ex	-	-	1	-	-	-	-	4	-	-	-	3	-	3	-	4	-	-	-	-
http://scv.bu.edu/Doc/Java/tutorial/java/exceptions/	-	3	-	4	-	4	3	-	-	-	-	-	-	-	-	-	-	4	-	-
http://www.adtmag.com/java/articleold.aspx?id=124	-	-	2	4	-	3	-	-	-	4	-	-	-	-	-	-	-	-	-	3
http://www.aspalliance.com/147	-	0	1	0	-	0	-	-	-	-	-	-	-	-	-	-	-	0	-	-
http://www.javaolympus.com/J2SE/Exceptions/Javal	-	-	-	-	-	-	0	4	-	-	-	3	-	0	-	4	-	-	-	-

BIBLIOGRAPHY

Adler, M. How to Mark a Book. Saturday Review of Literature, July 6, 1941 [Online]. Available: http://www.maebrussell.com/Articles%20and%20Notes/How%20To%20Mark%20A%20Book.html

Agichtein, E., Brill, E., and Dumais, S. *Improving Web Search Ranking by Incorporating User Behavior Information.* In Proceedings of the 29th Annual international ACM SIGIR Conference on Research and Development in information Retrieval. August 06-11, 2006, Seattle, Washington, United States.

Bao, S., Wu, X., Fei, B., Xue, G., Su, Z., and Yu, Y. *Optimizing Web Search Using Social Annotatins.* In Proceedings of the International World Wide Web Conference (WWW2007). May 8-12, 2007, Banff, Alberta Canada.

Baecker, R., Nastos, D., Posner, I., and Mawby, K. *The User-centerd Iterative Design of Collaborative Writing Software*. In Proceedings of the 1993 ACM conference on Human Factors in Computing Systems (INTERCHI 93).

Bargeron, D., Gupta, A., and Brush, B. *A Common Annotation Framework.* Microsoft Tech. Report MSR-TR-2001-108. Available: http://research.microsoft.com/research/pubs/view.aspx?tr_id=524

Bargeron, D., Gupta, A., Grudin, J. and Sanocki, E. *Annotations for streaming video on the Web: system design and usage studies.* Computer Networks (Netherlands), Elsevier Science, 17 May 1999, Vol. 31, No. 11-16, pp. 1139-1153.

Bateman S., Farzan R., Brusilovsky P., and McCalla G. *OATS: The Open Annotation and Tagging System.* In Proceedings 3rd annual e-learning conference on Intelligent Interactive Learning Object Repositories. November 9 - 11, 2006, Montreal, Quebec, Canada.

Bird, S. and Liberman, M. *A formal framework for linguistic annotation*. Speech Commun. 33, 1-2 (Jan. 2001), 23-60.

Bottoni, P., Civica, R., Levialdi, S., Orso, L., Panizzi, E., and Trinchese, R. *MADCOW: a Multimedia Digital Annotation System*. In AVI'04 Advance Visual Interfaces. May 25-28, 2004, Gallipoli, Italy.

Bottoni, P., Labella, A., Trinchese, R., and Gigli, L. *MADCOW: a visual interface for annotating web pages.* In the Proceedings of AVI'06 Advance Visual Interfaces. May 23-26, 2006, Venezia, Italy.

Brin, S., and Page, L. *The Anatomy of a Large-Scale Hyupertextual Web Search Engine.* In Computer Networks and ISDN Systems, 30(1-7) pp. 107-117, 1998.

Brush, B. *Annotating digital documents: anchoring, educational use, and notification*. In CHI '02 Extended Abstracts on Human Factors in Computing Systems. April 20 - 25, 2002, Minneapolis, Minnesota, United States.

Brush, B. *Annotating Digital Documents for Asynchronous Collaboration.* Technical Report 02-09-02. Available: http://citeseer.ist.psu.edu/556584.html

Brush, B, Bargeron, D., Grudin, J., Borning, A., and Gupta, A., *Supporting Interaction Outside of Class: Anchored Discussions vs. Discussion Boards.* In Proceedings of Computer Support for Collaborative Learning (CSCL'02), January, 7-11, 2002, Boulder, Colorado, United States.

Brush, B, Bargeron, D., Grudin, J., Gupta, A. and Cadiz, J., *Robust Annotation Position in Digital Documents.*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01), 2001, Seattle, Washington, United States.

Bush, V. *As We May Think*. In The Atlantic Monthly, July 1945, 641--649.

Cadiz, J., Gupta, A., and Grudin, J. *Using Web annotations for asynchronous collaboration around documents.* In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00). Philadelphia, Pennsylvania, United States. 309-318.

Catlin, T., Bush, P., and Yankelovich, N. *InterNote: extending a hypermedia framework to support annotative collaboration.* In Proceedings of the Second Annual ACM Conference on Hypertext (HYPERTEXT '89). Pittsburgh, Pennsylvania, United States.

Chatti, M., Sodhi, T., Specht, M., Klamma, R., and Klemke, R. *u-Annotate: An Application for User-Driven Freeform Digital Ink Annotation of E-Learning Content*. In Proceedings of the 6th International Conference on Advanced Learning Technologies (ICALT 2006), July 5-7, Kerkrade, Netherlands.

Chiueh, T., and Katz, R. *An Annotation System for VLSI Design.* In Proc of Intl. Phoenix Conf. on Computers and Communications, February, 1991, Phoenix, Arizona, United States.

Conklin, J. and Begeman, M. *gIBIS: a hypertext tool for exploratory policy discussion*. ACM Trans. Inf. Syst. 6, 4 (Oct. 1988), 303-331.

Correia, N. and Cabral, D. *VideoStore: A system to store, annotate and share video based content*. In Proceedings of the third international Conference on Multimedia and Information and Communication Technologies in Education (m-ICTE2005). June 7 - 10, 2005, Caceres, Spain.

Correia, N. and Chambel, T. *Active video watching using annotation*. In Proceedings of the Seventh ACM international Conference on Multimedia. October 30 - November 05, 1999, Orlando, Florida, United States.

Costa, M., Correia, N., and Guimarães, N. *Annotations as multiple perspectives of video content*. In Proceedings of the Tenth ACM international Conference on Multimedia, December 01 - 06, 2002, Juan-les-Pins, France.

Craswell, N., Hawking, D., and Robertson, S. *Effective site finding using link anchor information.* In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. September, 2001, New Orleans, Louisiana, United States.

Cross, J., Baber, C., and Woolley, I. Layered Annotations of Digital Images for Data Collection in the Field. In Proceedings of the Seventh IEEE International Symposium on Wearable Computer (ISWC'03), October 21-23, 2003, White Plains, New York, United States.

Davis, J., and Huttenlocher, P. *Shared Annotation for cooperative Learning System.* In Proceedings of CSCL'95, October 17-20, 1995, Bloomington, Indiana, United States.

Decurtins, C., Norrie, M. C., and Signer, B. *Digital annotation of printed documents*. In Proceedings of CIKM'03, New Orleans, LA, USA, November, 2003.

Denoue, L.,and Vignollet, L. *New ways of using web annotations.*. In Proceedings of 9th International World Wide Web Conference, 2000, Amsterdam, Netherland.

Denoue, L.,and Vignollet, L. *An Annotation tool for Web browsers and its applications to information retrieval*. In Proceedings of RAI2000, April, 2000, Paris.

Desmoulins, C.,and Mille, D. *Pattern-Based Annotations on E-Books: From Personal to Shared Didactic Content*. In Proceedings of the IEEE International Workshop on Wireless and Mobile Techonologies in Education (WMTE'02), 2002.

Dmitriev, P., Eiron, N., Fontoura, M., and Shekita, E. *Using Annotations in Enterprise Search*. In Proceedings of the 15th International World Wide Web Conference. May 23-26, 2006, Edinburgh, Scotland,

Eiron, N., and McCurley, K. *Analysis of Anchor Text for Web Search.* In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. July 28 – August 01, 2003, Toronto, Canada.

Euzenat, J. *Eight Questions about Semantic Web Annotations.* IEEE Intelligent Systems, vol. 17, no. 2, pp. 55-62, March/April, 2002.

Farzan, R. and Brusilovsky, P. *Social navigation support through annotation-based group modeling*. In Proceedings of 10th International User Modeling Conference (Edinburgh, UK, July 24-29, 2005). Lecture Notes in Artificial Intelligence, vol. 3538. Berlin: Springer Verlag, pp. 463-472.

Farzan R., and Brusilovsky P. *AnnotatEd: A Social Navigation and Annotation Service for Web-based Educational Resources*. In Proceedings of E-Learn 2006--World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education. October 13-17, 2006, Honolulu, Hawaii, United States.

Fleiss, J. *Measuring nominal scale agreement among many raters.* In Psychological Bulletin, Vol. 76, No. 5, 1971 pp. 378--382

Franke, K., Guyon, I., Schomaker, L., and Vuurpijl, L. *The WandaML markup language for digital document annotation*. In Proceeding of 9th IWFHR, Japan, Los Alamitos: IEEE Computer Society, pp. 563-568.

Freyne J., Farzan R., Brusilovsky P., Smyth B., and Coyle M. *Collecting Community Wisdom: Integrating Social Search & Social Navigation*. In Proceedings of International Conference on Intelligent User Interfaces. January 28-31, 2007, Honolulu, Hawaii, United States.

Fogli, D., Fresta, G., and Mussio, P. *On electronic annotation and its implementation*. In Proceedings of AVI'04, May, 2004, Gallipoli, Italy.

Gertz, M., Sattler, K., Gorin, F., Hogarth, M., and Stone, J. *Annotating Scientific Images: A Concept-Based Approach*. In Proceedings of the 14th international Conference on Scientific and Statistical Database Management (July 24 - 26, 2002). SSDBM. IEEE Computer Society, Washington DC, United States, 59-68.

Ginsburg, M. *Annotate! A Tool for Collaborative Information Retrieval*. In Proceedings of the Seventh International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, 1998. 75-80.

Golovchinsky, G., Price, M., and Schilit, B. *From reading to retrieval: Freeform ink annotations as queries*. In Proceedings of the 24 Annual ACM Conference on Research and Development in Information Retrieval (SIGIR '99), United States, 19-25.

Halasz, F., Moran, T., and Trigg, R. *NoteCards in a Nutshell*. In Proceedings of ACM CHI+GI'87 Conference on Human Factors in Computing Systems and Graphics Interface. New York: ACM Press, 45-52.

Handschuh, S. and Staab, S. *Authoring and annotation of web pages in CREAM.* In Proceedings of the 11th international Conference on World Wide Web (WWW '02). May 07 - 11, 2002, Honolulu, Hawaii, United States.

Harmon, R., Patterson, W., Ribarsky, W., and Bolter, J. *The Virtual Annotation System.* In Proceedings of the 1996 Virtual Reality Annual international Symposium (VRAIS 96) March 30 - April 03, 1996.

Hong, L., Chi, E. H., and Card, S. K. *Annotating 3D electronic books*. In CHI '05 Extended Abstracts on Human Factors in Computing Systems. April 02 - 07, 2005, Portland, Oregon, United States.

Hotho, A., Jäschke, R., Christoph, S., and Stumme, G. *FolkRank: A Rankiong Algorithm for Folksonomies.* In Proceedings of FGIR 2006, Germany.

Hotho, A., Jäschke, R., Christoph, S., and Stumme, G. *FolkRank: Information Retrieval in Folksonomies: Search and Ranking.* In York Sure and John Domingue, editor(s), The Semantic Web: Research and Applications, (4011):411-426, Springer,Heidelberg,2006.

Hu, B., Dasmahapatra, S., Lewis, P., and Shadbolt, N. *Ontology-Based Medical Image Annotation with Description Logics.* In Proceedings of 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'03), 2003.

Hu, Y., Xin, G., Song, R., Hu, G., Shi, S., Cao, Y., and Li, H. *Title Extraction from Bodies of HTML Docuements and Its; Application to Web Page Retrieval..* In Proceedings of SIGIR 2005, pp. 250-257.

Hua, Z., Wang, X., Liu, Q., and Lu, H. *Semantic knowledge extraction and annotation for web images.* In Proceedings of the 13th Annual ACM international Conference on Multimedia, November 06 - 11, 2005, Hilton, Singapore.

Jackson, H. *Marginalia: Readers writing in Books*, New Haven: Yale University Press, 2001

Jarvelin, K., and Kekalainen., J. *IR evaluation methods for retrieving highly relevant documents.* In Proceedings of the 23rd annual international ACM SIGIR conference on Research and Development on Information Retrieval, July 24-28, 2000, Athens, Greece.

Jung, T., Gross, M. D., and Do, E. Y. *Annotating and sketching on 3D web models*. In Proceedings of the 7th international Conference on intelligent User interfaces, January 13 - 16, 2002, San Francisco, California, United States.

Kahan, J. and Koivunen, M. *Annotea: an open RDF infrastructure for shared Web annotations*. In Proceedings of the 10th International Conference on World Wide Web. May 01 - 05, 2001, Hong Kong.

Kang, H., and Shneiderman, B. *Visualization Methods for Personal Photo Collections: Browsing and Searching in the PhotoFinder.* In Proceeding of the IEEE international Conference on Mutimedia and Expo (ICME2000). New York. United States.

Kurhila, J., Miettinen, M., Nokelainen, P., Floren, P. and Tirri, H. *Joint Annotation and Knowledge Building in Collaborative E-Learning.* In Proceeding of World Conference of E-Learning in Corporate, Government, Healthcare and Higher Education (E-Learn 2003), pages 2249--2252.

Lee, S., and Yong, H. *TagPlus: ARetrieval System using Synonym Tag in Foklsonomy.* In Proceeding of International Conference on Mutimedia and Ubiquitous Engineering (MUE'07), April 26-28, 2007, Seoul, Korea.

Leland, M., Fish, R., and Kraut, R. *Collaborative document production using quilt*. In Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work CSCW '88, September 26 - 28, 1988, Portland, Oregon, United States, 206-215.

Luz, S. *Y-notes: unobtrusive devices for hypermedia annotation.* In Proceedings of the Twelfth ACM Conference on Hypertext and Hypermedia, August 14 - 18, 2001, Århus, Denmark.

Marlow, C., Naaman, M., Boyd, D., and Davis, A. *Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead.*. In Proceeding of the 17th ACM Conference on Hypertext and Hypermedia (HT'06), August, 23-25 2006, Odense, Denmark..

Marshell, C. *Annotation: From Paper Books to Digital Library*. In Proceedings of DL'97, Jul 1997, Philadelphia, United States.

Marshall, C. *The Future of Annotation in a Digital (Paper) World*. In Successes and Failures of Digital Libraries (Harum and Twidale, eds). Urbana-Champaign: University of Illinois, 2000, pp. 97-117

Marshell, C. *Toward and Ecology of Hypertext Annotation*. In Proceeding of ACM Hypertext'98, Pittsburgh, USA, Jun 1998.

Marshell, C., and Brush, B. *From Personal to Shared Annotations*. In Proceeding of CHI2002, April 20-25, 2002, Minneapolis, Minnesota, United States, 812-813.

Marshell, C., and Brush, B. *Exploring the Relationship between Personal and Public Annotations*. In Proceedings Joint ACM IEEE Conference on Digital Libraries (JCDL'04), June 7-11, 2004, Tucson, Arizona, USA, 349-357.

Marshall, C., Price, M., Golovchinsky, G., and Schilit, B. *Collaborating over portable reading appliances*. Personal Technologies 3, 43--53. 1999. Available http://citeseer.nj.nec.com/ marshall99collaborating.html

Mashayekhi, V., Feulner, C., and Riedl, J. *CAIS: collaborative asynchronous inspection of software*. In Proceedings of the 2nd ACM SIGSOFT Symposium on Foundations of Software Engineering, December 06 - 09, 1994, New Orleans, Louisiana, United States.

Miettinen, M., Nokelainen, P., Floréen, P., Tirri, H., and Kurhila, J. *EDUCOSM -Personalized Writable Web for Learning Communities*. In Proceedings of the international Conference on information Technology: Computers and Communications. IEEE Computer Society, April 28 - 30, 2003, Washington DC, United States.

Nagao, K. *Digital Content Annotation and Transcoding*. Artech House Publishers, 2003.

Nagao, K., Ohira, S., and Yoneoka, M. *Annotation-based multimedia summarization and translation*. In Proceedings of the 19th international Conference on Computational Linguistics - Volume 1, Taipei, Taiwan, August 24 - September 01, 2002.

Neuwirth, C., Kaufer, D., Chandhok, R., and Morris, J. *Issues in the Design of Computer Support for Co-authoring and Commenting*. In Proceedings the 1990 ACM Conference on Human Factors in Computing System(CHI 97).

Noël, S., Robert, J.-M. *How the Web is used to support collaborative writing*. Behaviors & Information Technology, 22 (4), 2003, 245-262.

Nokelainen, P., Miettinen, M., Kurhila, J., Floréen, P., and Tirri, H. *A Shared Documents-Based Annotation Tool To Support Learner-Centered Collaorative Learning.* HIIT Technical Reports 2004-8, Helsinki Institute for Information Technology HITT, Finland [Online]. Available : http://cosco.hiit.fi/Articles/hiit-2004-8.pdf

Nokelainen P., Kurhila J., Miettinen M., Floreen P., and Tirri H., *Evaluating the role of a shared document-based annotation tool in learner-centered collaborative learning*, In Proceedings of IEEE ICALT 2003 conference, p.p.200-203

O'Hara, K, and Sellen, A. *A Comparison of Reading Paper and On-Line Document*. In Proceedings of ACM CHI'97, Mar 1997.

Olsen, D., Taufer, T., and Fails, J. *ScreenCrayons: annotating anything*. In Proceedings of the 17th Annual ACM Symposium on User interface Software and Technology. Santa Fe, NM, USA, October 24 - 27, 2004 (UIST '04). ACM Press, New York, United States, 165-174.

Page, L., Brin, S., Motwani., R, and Winograd, T. *The PageRank citation ranking: Bringing order to the web.* Technical report, Stanford Digital Library Technologies Project, 1998.

Perry, M. *Use Scenarios for Shared Editing in Scientific Collaborations*. In Proceedings of the fourth International Workshop on Collaborative Editing, CSCW2002 Conference, November 16-20, 2002, New Orleans, Louisiana, United States.

Petkovic, D., Lank, E., Ramirez, F., Raghavendra, S., Chen, F., Pathuri, S., Pekiner, C., Fergoso, J., Hsieh, A., and Marquez, A. *Asynchronous Multimedia Annotations for Web-Base Collaboration in Biology Education*. In Proc of SPIE Conference on Multimedia Storage and Retrieval, January, 2005, San Jose, California, United States.

Priest, R. and Plimmer, B. *RCA: experiences with an IDE annotation tool.* In Proceedings of the 6th ACM SIGCHI New Zealand Chapter's international Conference on Computer-Human interaction: Design Centered HCI, July 06 - 07, 2006, Christchurch, New Zealand.

Pryde, D. *The Highways of Literature; or, What to Read and How to Read*. Edinburgh: Nimimo, 1882.

Phelps, T., and Wilensky R. *Robust Intra-Document Locations*, In Proceedings of the 9th WWW Conference, May, 2000, Amsterdam, Natherland.

Ramachandran, S., and Kashi, R. *An Architecture for Ink Annotations on Web Documents*. In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03), Aug 3-6, 2003. 256 – 260.

Richardson, M., Prakash, A., and Brill, E. *Beyond PageRank: Machine Learning for Static Ranking*. In Proceedings of the WWW2006, May 23-26, 2006.

Rodríguez, H. *Using the WWW as infrastructure for collaborative production of documents*. Licentiate thesis. TRITA-NA-0117, Stockholm: The Royal Institute of Technology of Sweden, 2001.

Rosenthal, B. *The Rosenthal Collection of Printed Books with Manuscript Annotations*. New Haven: Yale University Press, 1997.

Sanderson, M. *Retrieval with good sense.* Information Retrieval(2), pp 47-67, 2000.

Sannomiya, T., Amagosa, T., Yoshikawa, M., and Uemura, S. *A Framework for Sharing Personal Annotations on Web Resources Using XML.* In Proceedings of Workshop on Information Technology for Virtual Enterprises, 2001. 40-48.

Schilit, N., Golovchinsky, G., and Price, M. *Beyond paper: supporting active reading with freee-form digital ink annotations.* In Proceedings of CHI'98, p. 249-256, ACM Press, 1998.

Shirkey, C *Ontology is Overrated: Categories, Links, and Tags.* April 2005, [Online] Available: http://www.shirky.com/writings/ontology_overrated.html

Shneiderman, B., and Kang, H. *Direct Annotation: A Drag-and-Drop Strategy for Labeling Photo.* In Proceedings of the international Conference on information Visualisation. July 19 - 21, 2000. Los Alamitos, California, United States, 88-95.

Silva, L., Mastella, L., Abel, M., Galante, R., and De Ros, F. *An Ontology-Based Approach for Visual Knowledge: Image Annotation and Interpretation.* In Proceedings of Workshop on Ontologies and their Applications, in XVII Brazilian Symposium on Artificial Intelligence (SBIA), September 20 – October 1, 2004, Brasil.

Spring, M., Fritsch, R., Lautenbacher, G., Lenox, T., Morse, E., Sapsomboon, B. Stewart, D. and Vathanophas, V., *Embodying Social Capital Facilitators in a Collaborative Authoring System*, In Proceedings of Association for Information Systems 1997, Americas Conference, August 15-17, 1997, Indannapolis, Indiana, United States.

Spring, M., Andriati, R. and Vathanophas, V., *Usability of a Collaborative Authoring System for Standards Development: Preferences, Problems, and Prognosis*, In Proceedings of the First IEEE Conference on Standardization and Innovation in Information Technology, Aachen, September 15-17, 1999, Germany.

Staab, S., Mdche, A., Handschuh, S. *An Annotation Framework for the Semantic Web*. In Proceedings of The First International Workshop on MultiMedia Annotation. January, 30 - 31, 2001. Tokyo, Japan.

Supsomboon, B. *Shared Defect Detection: The Effects of Annotation in Asynchronous Software Inspection.* Ph.D. dissertation, University of Pittsburgh, Pennsylvania, United States, 2000.

Takeda, T., Suthers, D. *Online Workspaces for Annotation and Discussion of Documents*. In Proceedings of International Conference on Computers in Education. December 3 - 6, 2002. IEEE Computer Society, Washington, DC, United States.

Teylingen, R., Ribarsky, E., Mast, C. *Virtual Data Visualizer*. IEEE Transactions on Visualization and Computer Graphics, vol. 03, no. 1, pp. 65-74, Jan-Mar, 1997.

Trigg, R., Suchman, L., and Halasz, F. *Supporting collaboration in notecards*. In Proceedings of the 1986 ACM Conference on Computer-Supported Cooperative Work (CSCW '86). December 03 - 05, 1986. Austin, Texas, United States.

Van Rijsbergen, C. *Information Retrieval*. Butterworth's, London, 1979.

Vathanophas, V., & Spring, M. *The Use of Peripheral Awareness Tools in Collaborative Systems*. WITS'01, 11th Workshop on Information Technologies and Systems. December 15-16, 2001, New Orleans, Louisiana, USA.

Vatton, I., Quint, V., Kahan, J., Cramer, K., Nylander, K., Rosen, K., Spinella, M., and LeDoux, L. *Amaya User Manual*. October, 2004 [On-Line]. Available: http://www.w3.org/Amaya/User/Manual.html

Wang, L., Khan, L., Liu, L., and Wu, W. *Automatic Image Annotation and Retrieval Using Weighted Feature Selection.* In Proceedings of the IEEE Sixth international Symposium on Multimedia Software Engineering (Ismse'04). December 13 - 15, 2004. Washington DC, United States, 435-442.

Wang, X., Zhang, L., Jing, F., and Ma, W. *Image annotation using search and mining technologies.* In Proceedings of the 15th international Conference on World Wide Web (WWW '06). May 23 - 26, 2006. Edinburgh, Scotland.

Weng, C. *Annotations and Asynchronous Collaborative Writing*. In Proceedings of CSCW'04 Doctoral Consortium, November 2004, Chicago, Illinois, United States.

Weng, C., Gennari, J., and McDonald, D. *A Collaborative Clinical Trial Protocol Writing System*, In Proceedings of 11th World Congress on Medical Informatics (MedInfo'04), November 2004, San Francisco, California, United States.

Weng, C., and Gennari, J. *Asynchronous Collaborative Writing through Annotations*, Notes, In Proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW'04), November 2004, Chicago, Illinois, United States, 578-81.

Wilhelm, A., Takhteyev, Y., Sarvas, R., Van House, N., and Davis, M. *Photo annotation on a camera phone*. In Proceedings of CHI '04 Extended Abstracts on Human Factors in Computing Systems, April 24 - 29, 2004, Vienna, Austria.

Westerveld, T., Kraaij, W., and Hiemstra, D. *Retrieving web pages using content, links, URLs and anchors*. In E. M. Voorhees and D. K. Harman, editors, NIST Special Publication 500-250: The Tenth Text Retrieval Conference, 2002

Willms, S., Morse, E., and Spring, M. *System Design for an Integrated Document System and its Impact on Standards Development Efficiency*. In Proceedings of IEEE International Engineering Management Conference '96, August 18-20, 1996, Vancouver, Canada.

Wolfe, J. *Effects of Annotations on Student Readers and Writers*, In Proceedings of the fifth ACM conference on Digital libraries, 2000, San Antonio, Texas, United States, 19-26.

Xue, G., Zeng, H., Chen, Z., Yu, T., Ma, W., Xi, W., and Fan, W. *Optimizing Web Search Using Web Click through Data*, In Proceedings of the CIKM'04, November 8-13, 2004 Washington DC, United States.

Yanbe, Y., Jatowt, A., Nakamura, S., and Tanaka, K. *Can Social Bookmarking Enhance Search in the Web?*. In Proceedings of the ACM IEEE Joint Conference on Digital Libraries, June 18-23, 2007, Vancouver, Brithish Columbia, Canada.

Yang, S., Du, V., Shao, N., and Chen, I. *Applying Personalized Annotation Mechanism to e-Document*. In Proceedings of the IEEE International Conference on E-Commerce Technology for Dynamic E-Business(CEC-East'04). 142-145.

Yanagisawa, M., Ito, K., and Akahori, K. *Web Memo: A WWW Browser for Learning with Support for Writing Notes in Web Material*. In Proceeding of International Conference on Computers in Education, 2004, Melbourne, Australia.

Yee, K. *CritLink: Advanced Hyperlinks Enable Public Annotation on the Web*. In Proceedings of Conference on Computer Supported Cooperative Work, CSCW2002. November 16-20, 2002, New Orleans, Louisiana, United States.