TREC10 Web and Interactive Tracks at CSIRO

Nick Craswell   David Hawking   Ross Wilkinson   Mingfang Wu

Technologies for Electronic Documents
Division of Mathematical and Information Sciences
CSIRO, Australia
{Nick.Craswell; David.Hawking; Ross.Wilkinson; Mingfang.Wu}@csiro.au

1. Overview

For the 2001 round of TREC, the TED group of CSIRO participated and completed runs in two tracks: web and interactive.

Our primary goals in the Web track participation were two-fold: A) to confirm our earlier finding [1] that anchor text is useful in a homepage finding task, and B) to provide an interactive search engine style interface to searching the WT10g data. In addition, three title-only runs were submitted, comparing two different implementations of stemming to unstemmed processing of the raw query. None of these runs used pseudo relevance feedback.

In the interactive track, our investigation was focused on whether there exists any correlation between delivery (searching/presentation) mechanisms and searching tasks. Our experiment involved three delivery mechanisms and two types of searching tasks. The three delivery mechanisms are: a ranked list interface, a clustering interface, and an integrated interface with ranked list, clustering structure, and Expert Links. The two searching tasks are searching for an individual document and searching for a set of documents. Our experiment result shows that subjects usually used only one delivery mechanism regardless of the searching task. No delivery mechanism was found to be superior for any particular task, the only difference was the time used to complete a search, that favored the ranked list interface.

2. Web Track

For the topic relevance runs, an index was built in a similar way to TREC-9. The PADRE99 retrieval system was used. Stemming and stopword elimination were not applied and sequences of letters and or digits were considered as indexable words. Words occurring in titles, metadata fields (if any), URL strings and in referring anchor text were distinguished in the index both from each other and from the normal document text. In order to keep the (compressed) index file under the 2gB filesize limit, only the first 3500 words in each document were indexed. Indexing took just under 2 hours elapsed time on a Dell Latitude C600 laptop with an 850 MHz processor and 512MB of RAM.

When processing queries, the Okapi BM25 relevance function used in TREC-9 was employed. Query-time stemming was employed in title-only automatic runs csiro0awa1 amd csiro0awa3 but not in csiro0awa2.

Performance of the search-engine style interactive interface to WT10g searching was quite acceptable despite the computation involved in extracting documents for display. Hawking and

Craswell took half of the topics each and interacted with the interface for an average of 5-6 minutes per topic while attempting to develop a set of good queries. Each query was saved for later batch processing and submission.

On the homepage finding task, run csiro0awh1 used exactly the same index and query processing machinery as did run csiro0awa2 in the topic relevance task. Results in the homepage finding task were much better than in the topic relevance task.

Homepage finding run csiro0awh2 corresponded to the machinery used in runs described in [1]. A perl script was used to extract anchor text from all the WT10g documents and collect it in pseudo documents named after the target URL. Later, pseudo documents corresponding to documents outside WT10g were removed and a PADRE99 index was built. The unstemmed queries were then processed against this unstemmed index using straightforward Okapi BM25 scoring(taking no account of URLs, or document structure.)

3. Interactive Track

3.1. Introduction

People search information for various purposes/tasks, and information retrieval systems are targeting their searching technologies to specific tasks. For example, some search mechanisms are good at content finding, and some others are good at homepage finding or online service finding[2,3]. An interesting question is whether a user can recognize the special merits of a search mechanism and take advantage of them for his/her searching tasks accordingly.

In this experiment, we conducted a user study in an attempt to gain a better understanding of users' mental model of searching mechanisms and users' searching tasks. Particularly, we investigated the following three research questions:

- If a user has a set of available searching mechanisms and a set of searching tasks, would the user be able to select a suitable mechanism that is optimal for that searching task?
- If a user has a set of available searching mechanisms and is aware of the advantages of each mechanism for certain types of searching tasks, would the user select the one that is optimal for that searching task?
- Can we improve a user's searching performance by guiding the user to use a suitable searching mechanism for a specific task?

3.2. Experimental setting

3.2.1. Topic

We selected eight searching topics from the TREC-10 interactive track. The eight search topics are:

1. Tell me three categories of people who should or should not get a flu shot and why.
2. Find a website likely to contain reliable information on the effect of second-hand smoke.
3. Find three articles that a high school student could use in writing a report on the Titanic.
4. Find three different information resources that may be useful to a high school student in writing a biography of Sr. Ernest Shackleton.
5. Find a website where I can find material on global warming.
6. I want to visit Antarctica. Find a website with information on organized tours/trips there.

7. Identify three interesting places to visit in Perth.
8. Find two websites that will let me buy a teddy bear online.

The above eight searching topics are of two types:

- Type I topic: Searching for a single document (- the information need can be satisfied by a single web document), the Topics 1, 2, 5 and 6 are of this type.
- Type II topic: Searching for a collection of documents (- the information need can be satisfied by a set of web documents), the Topics 3, 4, 7 and 8 are of this type. (It turns out that Topic 7 can also be satisfied by a document.)

3.2.2. Searching mechanism

In this experiment, we used Teoma (http://www.teoma.com) search engine for backend information retrieving. We chose it because the three types of search results returned from Teoma meet our requirements on two selected types of tasks.

Teoma provides the following three searching mechanisms:

- **Web page search (ranked list)**
  This searching mechanism is similar to other web search engines, which return the retrieved documents as a ranked list.

- **Experts' links**
  When a user wants to collect information about a certain topic, the user may not be the only one in this world who is interested in this topic. Very likely, someone else may already build his/her own portal for the topic and make that information available on the internet. If the user can get this portal directly, he/she will save time by avoiding searching/selecting the information piece by piece as from the ranked list.

- **Web pages by topic (clusters)**
  With this mechanism, the top ranked documents are grouped into topic/theme related clusters based on their topic keywords. For example, if a user wants to collect information on "global warming", the top retrieved documents will be clustered dynamically into the categories like "Institute", "Science", "Climate Change, Warming Climate" etc. The user can either drill down to a cluster to get information about a certain topic in depth, or browse a few clusters to collect information in width. This clustering structure can guide the user to collect the needed information purposely and avoid selecting/viewing the duplicated documents.

Intuitively, we think the web page search mechanism is suitable for the single document finding task, while Experts' links and clustering mechanisms are suitable for the information collection task.

Our experiment focused on users' mental model of their searching tasks and assigned searching/presentation mechanisms, instead of on the query formulation/reformulation, we chose to let subjects to perform their searching tasks by using predetermined (canned) queries. Therefore all subjects of the same searching topic got the same set of retrieved documents. We expected this would reduce the effect of query variation.

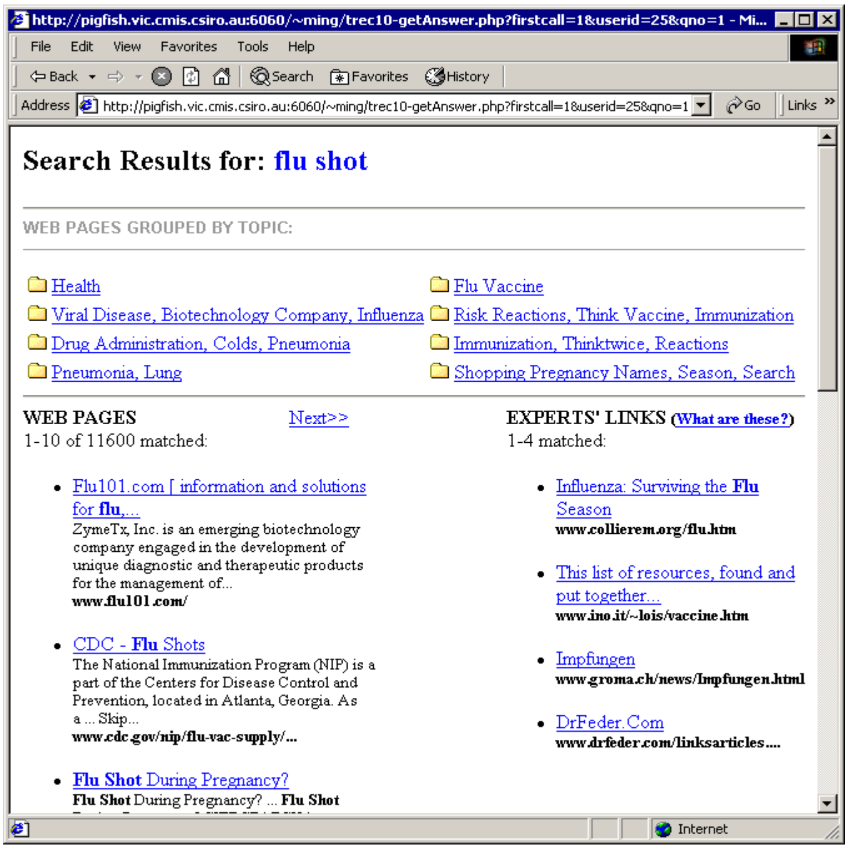
3.2.3. Experimental design

We recruited 24 subjects and divided them evenly into three groups. The experimental designs for each group are as following:

- Group 1
  Subjects were told only about the characteristics of each searching mechanism (as introduced in Teoma's help page). The aim was to observe whether subjects can recognize their task difference and choose a suitable searching mechanism accordingly.
- Group 2
  Subjects were told about the advantages of each searching mechanism and how they are related to the type of tasks, but subjects were still free to choose any mechanism for the search. The aim was to observe whether subjects in this group would select a suitable searching mechanism for a specific task when they clearly recognize the difference in searching topics and searching mechanisms.

In Groups 1 & 2, each searching topic is rotated in each searching position. The interface is similar to that from Teoma (see Figure 1). We developed an interface on the top of Teoma to keep just three necessary searching mechanisms, so that subjects would not know which search engine we are using and therefore concentrate on these three testing mechanisms.

- Group 3
  For this group of subjects, we developed two interfaces: one interface for supporting the web page search only (see Figure 2), and the other for supporting the clustering only (see Figure 3). The searching topics were blocked according to their types. The Latin-square experiment design was used here – each subject used two interfaces to search a block of four topics according to a predetermined order. With this experimental design, we try to compare whether a subject's searching performance would be improved by using a suitable interface (that we think) for that task.



http://pigfish.vic.cmis.csiro.au:6060/~ming/trec10-getAnswer.php?firstcall=1&userid=25&qno=1 - Mi...
File Edit View Favorites Tools Help
Back Search Favorites History
Address http://pigfish.vic.cmis.csiro.au:6060/~ming/trec10-getAnswer.php?firstcall=1&userid=25&qno=1
Go Links
Internet

Search Results for: flu shot

WEB PAGES GROUPED BY TOPIC:

- Health
- Viral Disease, Biotechnology Company, Influenza
- Drug Administration, Colds, Pneumonia
- Pneumonia, Lung
- Flu Vaccine
- Risk Reactions, Think Vaccine, Immunization
- Immunization, Thinktwice, Reactions
- Shopping Pregnancy Names, Season, Search

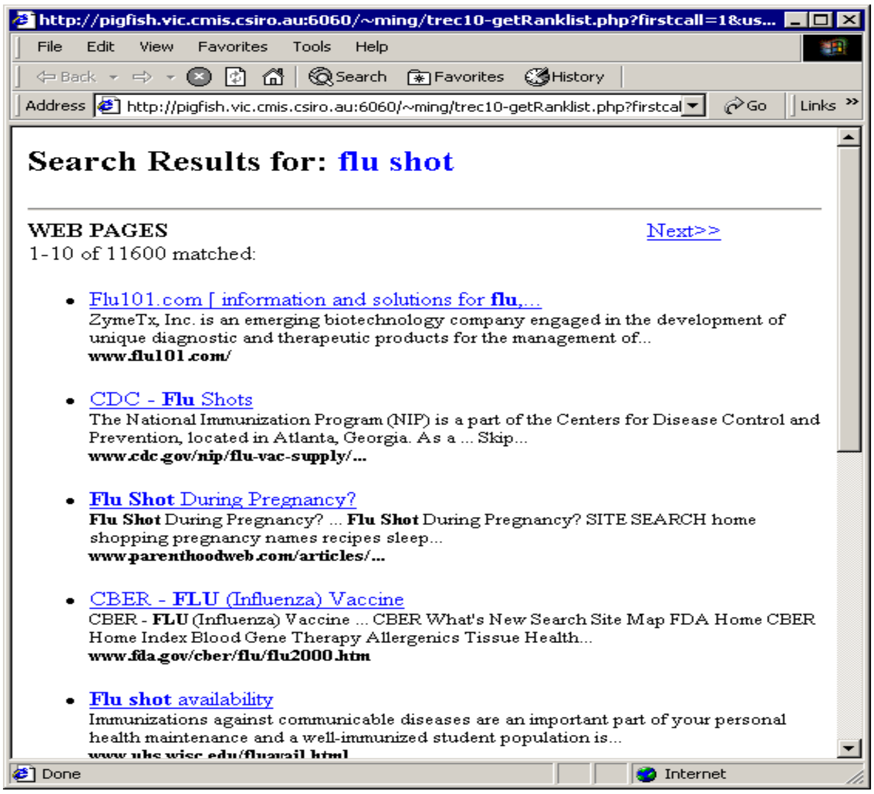
WEB PAGES Next>>
1-10 of 11600 matched:

- Flu101.com [information and solutions for **flu**,...
  ZymeTx, Inc. is an emerging biotechnology company engaged in the development of unique diagnostic and therapeutic products for the management of...
  www.flu101.com/
- CDC - **Flu** Shots
  The National Immunization Program (NIP) is a part of the Centers for Disease Control and Prevention, located in Atlanta, Georgia. As a ... Skip...
  www.cdc.gov/nip/flu-vac-supply/...
- **Flu Shot** During Pregnancy?
  Flu Shot During Pregnancy? ... **Flu Shot**

EXPERTS' LINKS (**What are these?**)
1-4 matched:

- Influenza: Surviving the **Flu** Season
  www.collierem.org/flu.htm
- This list of resources, found and put together...
  www.ino.it/~lois/vaccine.htm
- Impfungen
  www.groma.ch/news/Impfungen.html
- DrFeder.Com
  www.drfeder.com/linksarticles....

Figure 1. The integrated interface


http://pigfish.vic.cmis.csiro.au:6060/~ming/trec10-getRanklist.php?firstcall=1&us...
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
History
Address
http://pigfish.vic.cmis.csiro.au:6060/~ming/trec10-getRanklist.php?firstcal
Go
Links
Done
Internet

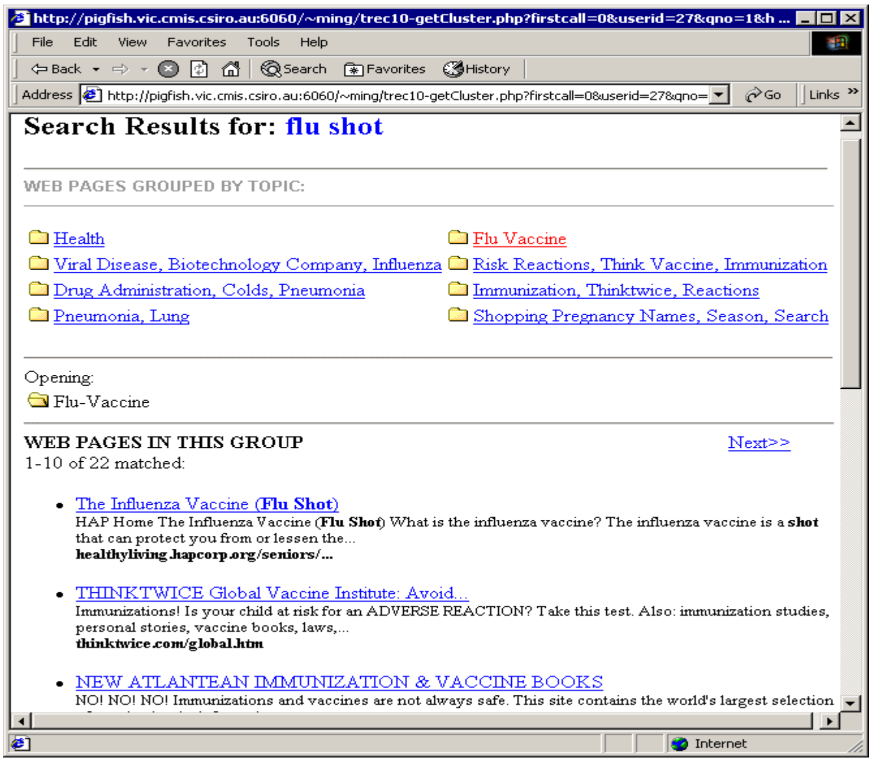
Search Results for: flu shot

WEB PAGES Next>>

1-10 of 11600 matched:

- Flu101.com [information and solutions for **flu**,...
  ZymeTx, Inc. is an emerging biotechnology company engaged in the development of unique diagnostic and therapeutic products for the management of...
  www.flu101.com/
- CDC - **Flu** Shots
  The National Immunization Program (NIP) is a part of the Centers for Disease Control and Prevention, located in Atlanta, Georgia. As a ... Skip...
  www.cdc.gov/nip/flu-vac-supply/...
- **Flu Shot** During Pregnancy?
  Flu Shot During Pregnancy? ... **Flu Shot** During Pregnancy? SITE SEARCH home shopping pregnancy names recipes sleep...
  www.parenthoodweb.com/articles/...
- CBER - **FLU** (Influenza) Vaccine
  CBER - **FLU** (Influenza) Vaccine ... CBER What's New Search Site Map FDA Home CBER Home Index Blood Gene Therapy Allergenics Tissue Health...
  www.fda.gov/cber/flu/flu2000.htm
- **Flu shot** availability
  Immunizations against communicable diseases are an important part of your personal health maintenance and a well-immunized student population is...
  www.uhs.wisc.edu/fluavail.html

Figure 2. The ranked list interface


http://pigfish.vic.cmis.csiro.au:6060/~ming/trec10-getCluster.php?firstcall=0&userid=27&qno=1&h ...
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
History
Address
http://pigfish.vic.cmis.csiro.au:6060/~ming/trec10-getCluster.php?firstcall=0&userid=27&qno=
Go
Links
Internet

Search Results for: flu shot

WEB PAGES GROUPED BY TOPIC:

- Health
- Viral Disease, Biotechnology Company, Influenza
- Drug Administration, Colds, Pneumonia
- Pneumonia, Lung
- Flu Vaccine
- Risk Reactions, Think Vaccine, Immunization
- Immunization, Thinktwice, Reactions
- Shopping Pregnancy Names, Season, Search

Opening:
Flu-Vaccine

WEB PAGES IN THIS GROUP Next>>

1-10 of 22 matched:

- The Influenza Vaccine (**Flu Shot**)
  HAP Home The Influenza Vaccine (**Flu Shot**) What is the influenza vaccine? The influenza vaccine is a **shot** that can protect you from or lessen the...
  healthyliving.hapcorp.org/seniors/...
- THINKTWICE Global Vaccine Institute: Avoid...
  Immunizations! Is your child at risk for an ADVERSE REACTION? Take this test. Also: immunization studies, personal stories, vaccine books, laws,...
  thinktwice.com/global.htm
- NEW ATLANTEAN IMMUNIZATION & VACCINE BOOKS
  NO! NO! NO! Immunizations and vaccines are not always safe. This site contains the world's largest selection

Figure 3: The clustering Interface

3.2.4. Experimental procedure

All experiments followed the following procedure:

1. Pre-search questionnaire
2. Training session
3. A search session (maximal 8 min)
4. Post-search questionnaire
   Repeat 3 & 4 until all topics are searched.
5. Exit questionnaire
   (The whole procedure took about 1.5 hours.)

3.3. Experimental result

3.3.1. Search Performance

All subjects successfully searched all topics within assigned time period. It seems that this year's searching topics are relatively easy. For each topic, the needed information can usually be found from the top 5 retrieved documents, though more within-a-site browsing/searching is needed for Topics 6, 7, and 8.

We can see from Table 1 that there is no significant difference between each group in terms the number of document read, either within the same type of topics or across all eight topics. (The mean across all eight topics for each group is: M(Group1) = 4.1, M(Group2) = 4.1, M(Group3-Clus) = 3.9, and M(Group3-List) = 4.1.)

Table 1. The number of documents read by each group

		Type I topic					Type II topic				
Topic		1	2	5	6	Mean	3	4	7	8	Mean
Group1		5.5	3	3	4	3.9	5.4	4.6	3.4	4.4	4.5
Group2		5.8	3.0	2.2	3.6	3.7	4.9	4.7	3.9	4.8	4.6
Group3	Clus	5.9	3.0	3.0	3.3	3.8	4.8	4.7	3.1	3.7	4.1
	List	5.9	3.3	2.3	3.3	3.7	4.9	4.7	4.0	4.3	4.5

Table 2 shows the time taken to finish each search session. Overall, subjects from Group3 used the least time by using the ranked list interface, this is probably because subjects of this interface got less distraction, they concentrated on one interface, and the quality of this list is very high. In this table, significant difference only exists between the Group1 and the list interface of the Group3.

Table 2. Time to finish each search session (in second)

		Type I topic					Type II topic				
Topic		1	2	5	6	Mean	3	4	7	8	Mean
Group1		230.3	184.9	119.6	232.4	191.8	303.1	306.0	146.5	356.1	277.9
Group2		131.4	132.5	116.9	230.5	152.8	315.6	228.6	243.3	208.0	248.9
Group3	Clus	248.0	162.5	154.5	196.0	190.1	349.8	253.0	192.3	271.5	266.7
	List	93.0	125.0	91.3	178.5	122.0	245.0	248.3	125.5	334.3	238.3

To answer our research question 3, we compare subject's performance with the Group 3 by using either clustering interface or the ranked list interface, we do not see any evidence to show that the subjects of the clustering interface would finish their searching sessions faster than other interfaces, and take less effort.

3.3.2. Search Behavior

During the experiment, we observed the following searching behaviors:

- There were generally two browsing strategies used by Group1 and Group2 subjects. One type of subjects read the search result page from top to bottom, and picked up a possibly relevant document to read along the way. Another type of subjects read and browsed the search result first, then selected the (possibly most) relevant documents to read. The distribution of each type is relatively even (52 : 76)[1]. This may imply that to the first type of subjects, the ranking is important, while to the second type of subjects, the site summary is more important.
- To answer our research question 1 and 2, we went through the recorded screen actions of subjects from Group1 and Group2. Generally, subjects of Group1 and Group2 checked the ranked list first, when they could not find satisfactory document(s), then they would switch to clusters or expert links. Only for a small number of sessions (17 in Group 1 vs. 24 in Group 2) from each group, a searching session was started by using clustering or expert links. Subjects from Group 2 used more clustering organization and expert links (63 sessions in Group 2 vs. 47 sessions in Group 1). This may indicate that if the subjects understand more about the clustering organization and expert links, they would use these two mechanisms more. A further experiment with more subjects is needed to verify this claim.
- The interaction with the searching system is mixed up with the usability of a website. For example, for tourism and shopping topics (topics 6, 7 and 8), the needed information is usually not brought on the first page. A within-a-site browsing is needed. The searching success depends on the design of the site, subject's searching habit and luck. For example, for Topic 8, ten subjects (from Group 1 & 2) read the first ranked document, only half of them found the needed information; the quickest one took 40 seconds while the slowest one took 3 minutes. For another example, the first ranked document of the topic 6 has nearly 10 screens of text, the first screen has a lot of links for browsing, and the needed information is on the fourth screen. If subjects scroll a lot and read the whole text (or up to the fourth screen) of the page, they would be able to find the needed information easily. However some subjects just read the first screen, they then followed whatever links on the first screen that they were interested in, as a result, they usually got lost within this site.
- In our post-experimental questionnaire/interview, we asked subjects to describe their daily searching habits. Generally, subjects recognized that they search for various purposes, but they usually stick to one search engine. They switch to other search engines only when they can not get satisfactory results. Only two subjects claimed that they choose search engines depending on what they are searching. For example, they would select Google to search someone's homepage, NorthernLight for research documents, and Yahoo for shopping stuff. (It is not clear if an interface is designed to encourage users to switch from one service to another, will more users do so?)

3.3.3. Subjects' feedback

Subjects from Group 1 and Group 2 gave similar comments on each mechanism. Here are some typical examples what they like about each mechanisms and what they dislike each mechanisms.

Pros:

Cluster

- Make it easier to find information related to the search topic

[1] Group 1 and Group 2 together had 128 searching sessions (8 topics x 16 subjects) in total.

- Easy to find specific information
- If you can find a useful topic, then you get a good list of useful relevant sites/links
- When the groups are accurate they are very useful for finding multiple sites with similar content.
- It gave me the opportunity to learn more about the topic.

Web pages

- Highlight the match words
- More detail description of each link, so it helps to search quickly.
- Perhaps contains the "most relevant sites, but also contains many irrelevant sites, but if you can sift them out, you usually get good results.

Expert Links

- Credible, often lots of links
- Have a page with many good, specific links

Cons:

Cluster

- Can be hard to find a completely relevant topic

Web pages

- Sometimes requires experience to be able to look at the briefs/summaries and quickly find the useful relevant ones
- Many pages, have to scroll more.

Expert links

- Show only the address, it is better to show a short description of the page.
- Often disorganized, and occasionally just too much information, overwhelming
- Sometimes too specific, not general enough

Most subjects said the searching topics are close to their daily search but these selected topics are relatively easy. One subject commented: "Overall, most searches were fairly easy and quick."; and another subject wrote: "I hope I didn't finish it too quickly! It might seem like I was rushing, but it was just that I found results quickly."

3.4. Conclusion

Based on the experimental results, we can't conclude that users would select a particular mechanism for a certain type of search tasks, neither we can assume that user's search behavior is task driven at this stage. Further analysis and research are needed.

4. References

1. Nick Craswell, David Hawking and Stephen Roberson. Effctive site finding using link anchor information. Proceedings of ACM SIGIR 2001, September 2001, pp.250-257

2. David Hawking, Nick Craswell and Kathleen Griffiths. *Which search engine is best at finding online services?.* WWW-10 Poster Proceedings, 2001.

3. Nick Craswell, David Hawking and Kathleen Griffiths. *Which search engine is best at finding airline site home pages?.* CSIRO Mathematical and Information Sciences TR01/45, 2001.