Evaluating a Task-specific Information Retrieval Interface

Russell Swan James Allan Don Byrd
Center for Intelligent Information Retrieval
Computer Science Department
University of Massachusetts
Amherst, MA 01003
{swan, allan, dbyrd}@cs.umass.edu

ABSTRACT

We present an evaluation of an information retrieval system designed for the 1997 TREC-6 Interactive Track; that is, Aspect Oriented Retrieval, or finding documents that cover all aspects of relevance to a given topic. Our system includes a basic search system, a task-specific "aspect window", and a 3-D visualization of document and aspect relationships. We compare two versions of our system against ZPRISE, a baseline system provided by NIST. A study of 20 searchers shows significant differences between two classes of searchers, and supports several hypotheses about the design of an aspect oriented system. An interesting result is a likely correlation between structural visualization ability and facility with a 3-D visualization.

KEYWORDS

Interactive Information Retrieval, User Interfaces, 3-D User Interfaces, Information Visualization, User Studies, Aspect Oriented Retrieval

INTRODUCTION

An information analyst needs tools that help him or her rapidly retrieve data, assess their meaning and impact, and distribute the results appropriately. In this study, we examine two complementary tools that we built for a typical problem that an analyst might face: so-called "aspect oriented retrieval". The problem was defined by the Interactive Retrieval track of TREC, the Text REtrieval Conference sponsored by NIST[7], and our system was built and tested in that context.

Our interest was in crafting a user interface for this problem and then evaluating how well it meets the needs of that task. Evaluations of the quality of an Information Retrieval system most commonly focus on retrieval effectiveness: measuring the precision or recall of a retrieved set–i.e., how accurately does the system get the useful material–and sidestepping the question of what a user does with the information that is returned. In this study we focus on the quality of the system, evaluating its potential for assisting users with aspect oriented retrieval. We rely upon a high-quality retrieval system, but are less concerned with its effectiveness than with what we can do with its results.

In the next section we describe aspect oriented retrieval in detail and discuss our hypotheses about important components of the task-specific system and user interface that we built. We follow that by a description of the two tools we built, the system that combines them, and a discussion of how we used the TREC Interactive Track framework to evaluate the quality of our system. We then analyze the results of the experiments and show how our hypotheses were or were not supported.

ASPECT ORIENTED RETRIEVAL

The specifications for the TREC Interactive Track define an aspect to be "roughly one of many possible answers to a question which the topic in effect poses."[13] For example, consider the sample query from TREC-5, "What are the latest developments in the production of electric automobiles?" The NIST judges looked at documents and came up with several aspects of relevance for the question. Among the 11 aspects are:

- government funding of electric car development programs
- industrial development of high energy batteries
- setbacks (planned developments dropped, difficulties)
- increased use of aluminum bodies

A searcher and system combination is evaluated based on how well the user can retrieve documents that cover *every one* of those 11 aspects. The searchers were given these instructions:

> Please save at least one document that identifies *each different* recent development in the field of electric automobiles. If one document discusses several developments, then you need not save other documents that repeat those developments, since your goal is to identify the different ones that have been discussed.

Note that unlike in a traditional retrieval setting, there is no expectation that the searcher find *all* relevant documents. Instead, the searcher needs to have a means for deciding whether all aspects of the topic have been covered.

Finding every aspect is likely to be a difficult and open-ended task, so the search length is capped at 20 minutes. Any searcher still working after that time was told to stop. The system is therefore being evaluated on its ability to help the user find all aspects as rapidly as possible.

HYPOTHESES

We designed our systems with several hypotheses in mind about how best a system can help a user with aspect retrieval. The hypotheses that we were able to test in this experiment were:

1. A tool that is designed specifically to help the searcher keep track of aspects will be helpful.

2. A searcher will be likely to submit several queries on the same topic, so will benefit from a means for indicating which documents were seen on earlier queries.

3. Extracting the few most significant terms from a set of grouped documents (that represent an aspect) will help the searcher meaningfully label that aspect.

4. Ability to use a 3-D visualization will correlate with testable structural visualization ability.

The next sections detail the experimental design and then shows how the results support or fail to support our hypotheses.

SYSTEM

We used three systems for the experiments discussed in this study:

1. ZPRISE is a basic GUI information retrieval system acquired from NIST.
2. AspInquery is a GUI implementation of Inquery that includes an "aspect window" to help with the task. The core of AspInquery is a basic GUI similar to ZPRISE.
3. AspInquery Plus is an extension of (2) that includes a 3-D visualization of document relations.

We discuss the systems in more detail below.

The baseline system for our experiments was ZPRISE, NIST's publicly available search system, modified slightly for the aspect oriented retrieval task (some advanced functionality was removed by NIST). ZPRISE uses a straightforward user interface much like that used by most Internet search engines: it has an area for typing in a query, a window for displaying a ranked list of documents, and a window for viewing a document of interest. For each document in the ranked list, ZPRISE displays the date, the document number, the headline, and a list of terms from the query that were found in that document. When the full text of a document is viewed, query terms contained in the document are highlighted. There is a button for each document on both the ranked list and in the document window; clicking on the button marks the document as being relevant.

Inquery

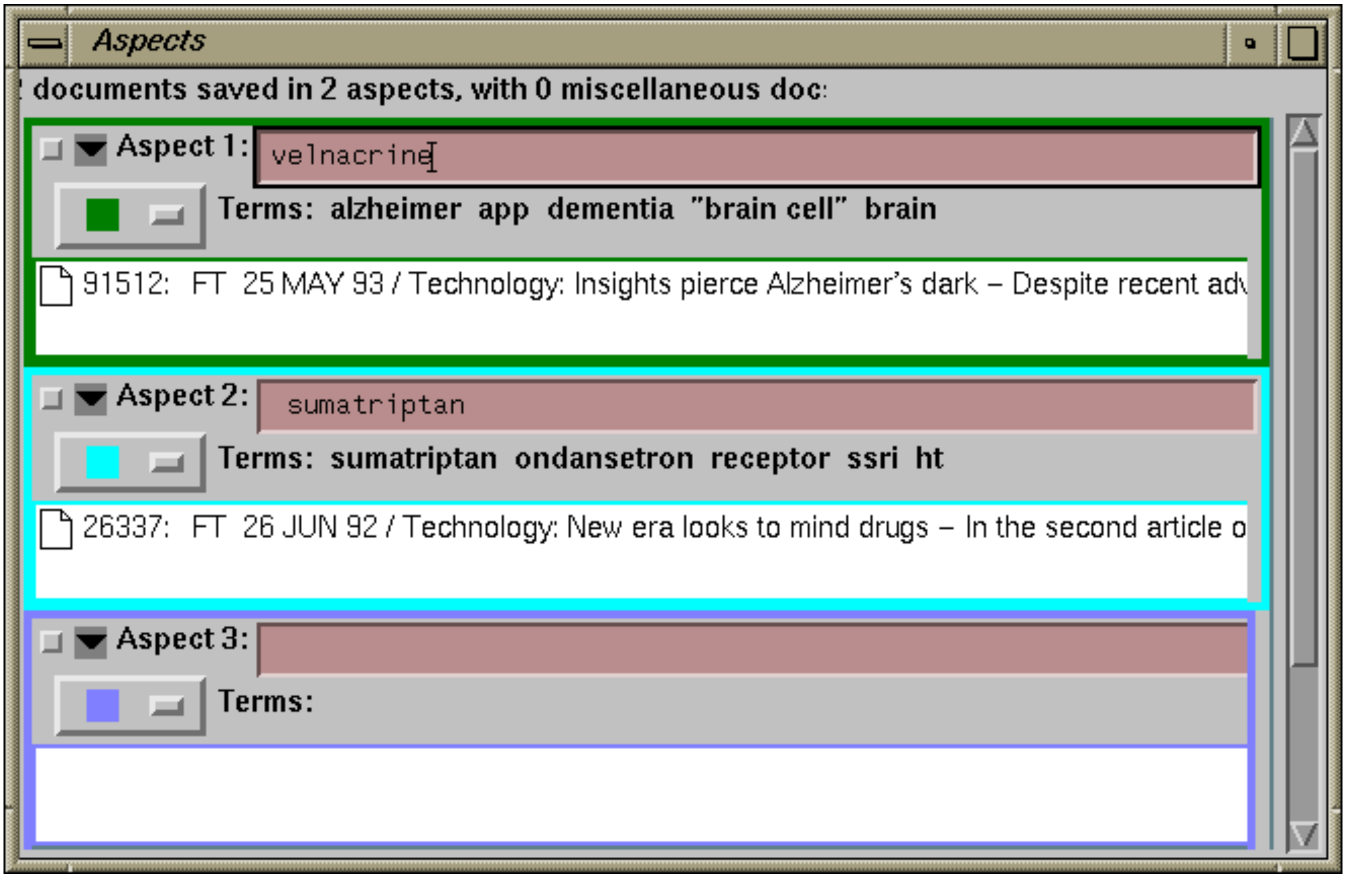
Our system consisted of the InQuery search engine[3] with a new interface. Our basic user interface has much in common with the ZPRISE interface, differing in two significant ways: ZPRISE displays the query terms contained in a document after the headline but our system does not, and our system color codes whether a document has been viewed but ZPRISE does not. Specifically we write the headline information of a document in blue if it has not been viewed before, and purple if it has been seen. (This scheme was modeled after the default color scheme Web browsers use to show if a hypertext link has been followed or not.)

Both ZPRISE and our system accept plain text input for queries. Our system also supports a phrase operator, invoked by placing terms together within double quotes (e.g., "balanced budget"). The phrase operator increases the ranking assigned to documents where all terms in the phrase are found in close proximity.

Aspect Window

With a basic IR system, an analyst may be able to find the documents containing various aspects, but he or she has to use another window or a piece of paper to keep track of what has been found already. We implemented an "aspect window" tool to help with this task. The idea is to provide an area where documents on a particular aspect can be stored. To help label the information, statistical analysis of word and phrase occurrences is used to decide what terms and phrases are most distinctive about a document or set of documents in an aspect. We provided an area for the user to manually assign additional keywords or labels if needed.

Each area of the aspect window has a colored border, a text field at the top for entering a descriptive label, and an automatically generated list of the five noun phrases that most distinguish the group of documents assigned to this aspect from the remainder of the collection. The description field is solely for the user’s convenience and need not be filled. If the user wants a description they can type or paste into it, or drag automatically generated phrases into it. Figure 1 shows an example of the aspect window. The system shows two groups of documents (two aspects) already identified and a third area waiting for the next aspect. The first aspect contains one documents, 91512. The user entered this document into the aspect by dragging a listed document from the ranked list display (part of the basic interface) into the aspect's document list. The system then analyzed the selected document and found five phrases that describe the aspect: they are the terms “alzheimer”, “app”, “dementia”, "brain", and the phrase “brain cell”. The analyst did not find those phrases descriptive enough, so he or she manually inserted the term "velnacrine". For the second aspect the automatically identified term “sumatriptan” is an adequate descriptor.



Aspects

documents saved in 2 aspects, with 0 miscellaneous doc

Aspect 1: velnacrine

🟩 Terms: alzheimer app dementia "brain cell" brain

91512: FT 25 MAY 93 / Technology: Insights pierce Alzheimer’s dark – Despite recent adv

Aspect 2: sumatriptan

🟦 Terms: sumatriptan ondansetron receptor ssri ht

26337: FT 26 JUN 92 / Technology: New era looks to mind drugs – In the second article o

Aspect 3:

🟪 Terms:

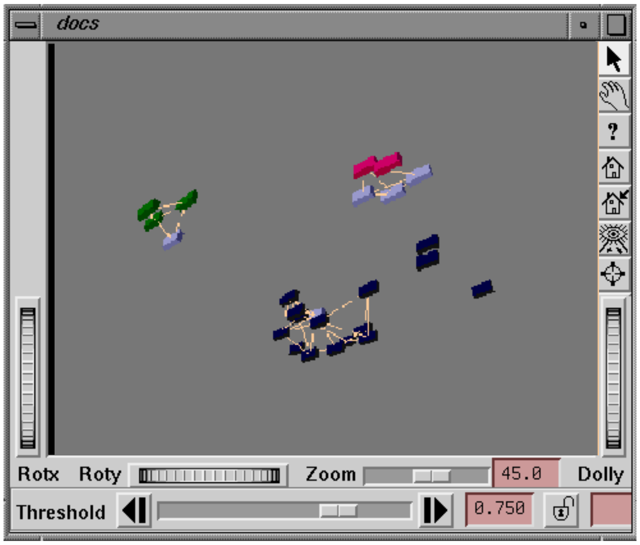
Figure 1. The Aspect Window

The purpose of the aspect window is to assist the user in categorizing the information as it is discovered, and to keep an overview of the information discovered so far. In an aspect oriented or briefing type setting this step is required for the task to be completed properly, but to our knowledge no systems have been built so far which provide any assistance for this task.

Visualization: AspInquery Plus

Another important step in the aspect oriented retrieval task is deciding (repeatedly) which document to look at next. In a ranked retrieval system the documents are presented in the order of probability of relevance, so the user is more likely to encounter relevant documents at the top of the list than further down. Nearly always the headline is used to decide if the full text is worth reviewing or not. Some systems [8, 15], ZPRISE among them, give information about the query terms that appear in the document, expecting that they can be used to help decide whether to investigate further. But for an aspect retrieval task, the deciding point of whether to investigate a document further is not the information content, but the marginal information content–i.e., the information content in the context of what has already been seen. We believe that documents that are similar in terms of information content will also be lexically similar. The Cluster Hypothesis[14] states that relevant documents tend to cluster, and it has been shown to be valid in top-ranked documents[5, 9]. Aspects represent different forms of relevance, and we believe that they will group together within the set of relevant documents.

AspInquery Plus compares documents in an extremely high-dimensional space (approximately 400,000 for this collection) where each dimension corresponds to a feature in the collection and the distance was measured by the sine of the angle between the vectors. That space was collapsed to 3 dimensions for visualization using a spring embedding algorithm[12]. The interface included a slider for adjusting the threshold that determined the tightness of the generated clusters. The resulting visualization is similar in style to BEAD[4], differing in a few key aspects: BEAD was used on an entire (though small) corpus, and this display is used only on the retrieved set; the vectors used by BEAD were based on document abstracts but the vectors used by this display are based on full text.



Zoom 45.0

Threshold 0.750

Figure 2. The 3-D Window

Documents that are nearby in 3-space are generally nearby in the high dimensional space also (though the spring embedding dimensional reduction occasionally forces unrelated documents to be near one another), meaning that they share information content to a considerable degree. For that reason, the 3-D display provides the user with information about whether the document is worth investigating further, helping the user to sort through documents more quickly. Documents that have not been assigned to any aspect have the same blue/purple (read/unread) color scheme that is used in the main window. Documents in the 3-D window are persistent between queries: when new documents are retrieved they are colored light blue (light purple when read) and are placed in the 3-D window by the forces exerted from already placed documents. Figure 2 shows five newly retrieved documents in light gray. It is easy to see that three of these documents fall into a group of two previously seen documents (upper right of figure) and the other new documents fall into the small group in the upper left and the large group. An analyst who is under time pressure could use the 3-D display to decide that the unjudged document near that aspect is probably on the same aspect and so not worth examining. A retrieved document that is far from any already-marked aspect is more likely to be useful. (We have been investigating variations on the visualization that enhance the ability for a user to find new and interesting material. [2])

The three windows—result list, aspect, and 3-D—were tightly integrated. If a document is selected by a mouse click in any of the three windows, that document is highlighted in all windows in which it is visible. A document can be opened for viewing by double clicking in any of the three windows. The colors were coordinated between the windows: if a document has been saved to an aspect, that aspect's color is assigned to the document in the 3-D window and also placed before the document in the list.

EXPERIMENT

The purpose of the experiments described in this study is to evaluate how two different systems affected user performance and user perception of the problem. In accordance with the TREC-6 interactive track experimental design [13] the experiment was performed with six queries. The corpus was newspaper articles from the Financial Times 1991-1994, approximately 200,000 articles total. This corpus is a subset of the TREC collection[7].

The basic experimental design called for four participants, each running all six queries. The order of topics was held the same for all participants, but the systems were alternated. The participant was given a tutorial to learn the system, then ran three queries. After a break the participant was given a tutorial on the second system, then ran the other three queries. The basic design leads to six 2x2 Latin squares to estimate the difference between the control and experimental systems. We extended the experiment by adding more participants in groups of four, replicating the 4x6 design, and by adding a second system and using the same 4x6 design for it. (See Table 1.)

Group	Type	Control	Experimental	Size
1	General	ZP	AI	4
2	Librarian	ZP	AI	4
3	General	ZP	AI+	4
4	Librarian	ZP	AI+	4
5	General	AI	AI+	4

Table 1. Experimental Setup

Before the searches, each participant filled out a questionnaire to determine age, education, gender and computer experience, and two psychometric tests[6], a test of verbal fluency (Controlled Associations, test FA-1) and a test for structural visualization (Paper Folding, test VZ-2). We gave each participant a piece of scratch paper before each search, and a short questionnaire after each. After all the searches were finished the participant was given a final questionnaire, and then "debriefed". The study was conducted single blind: the participants were not told until the debriefing which system was the control and which was the experimental system. Each search had a 20 minute time limit, and the participant was instructed to stop the search if they had not finished in 20 minutes.

Population

We were interested in how librarians perform search tasks as compared to a more general user population. Eight university librarians were recruited for the study. Four were placed in one experimental group and used AspInquery, and the other four were placed in a separate group and used AspInquery Plus. All eight of the librarians had MLS degrees, and several had an additional Masters degree. One had a JD. Seven of the eight librarians were over forty (the other was in her twenties). Six of the librarians were women and two were men.

The general population was recruited by flyers distributed on campus. This group was primarily students (10 of 12 participants). In most ways this was a very diverse group, ranging from undergraduates to a post doctoral student. However, these people were much younger than the librarians: one participant was in her forties; other than her, the oldest participant was in his thirties. Five were women and seven were men. Table 1 shows the systems used by the different groups in the experiment.

RESULTS

A full analysis of all the data from the experiment is beyond the scope of this paper and is available elsewhere [1].

Population Comparison

The preferred systems for members of the different groups are summarized in Tables 2. For the 16 participants that used the control system, nine preferred the control system, three preferred AspInquery, and four preferred AspInquery Plus. Seven of eight librarians preferred the control system. The librarian who preferred the experimental system (she used AspInquery Plus) was in her twenties. Three of four users from the general population preferred AspInquery to ZPRISE, and three of four preferred AspInquery Plus to ZPRISE. The data presented in the last two rows of Table 2 has a χ^2 value of 6.349, which is significant at the $p < 0.025$ level.

The response of the two groups (librarians and students) was quite different. However, except for one individual, all the librarians were older than all the students. Table 3 presents the preferred systems of participants with participants divided by age. This breakdown has a χ^2 value of 9.679 which is significant at the $p < 0.005$ level, indicating that age may be a more relevant predictor of system preference than occupation.

		ZP	AI, AI+
AI vs ZP	Group 1	1	3
	Group 2	4	0
AI+ vs ZP	Group 3	1	3
	Group 4	3	1
Groups 1+3 (General)		2	6
Groups 2+4 (Librarian)		7	1

Table 2. Preferred Systems by Group

	ZP	AI, AI+
Under 40	2	7
Over 40	7	0

Table 3. Preferred Systems by Age

What could account for this difference? Several participants stated that they found the mouse movements required in a drag and drop interface to be quite awkward. All the librarians had extensive experience with computers but most of them had limited experience with mouse based direct manipulation interfaces, as compared to the general group. When asked to score their familiarity with mouse based interfaces (1 =

low, 5 = high) the 8 librarians scored $\overline{x} = 3.625$, s = 1.061 and the general users scored $\overline{x} = 4.417$, s = 0.669 for a t statistic of 2.057, which is significant at the level of $p < 0.05$. The seven participants over forty scored $\overline{x} = 3.429$, s = 0.976 and the thirteen under forty scored $\overline{x} = 4.462$, s = 0.660 for a t statistic of 2.826 which is significant at $p < 0.025$. This indicates that experience with drag and drop may be a very strong factor in a person's like or dislike of a system.

Support for Hypotheses

Aspect window will be useful to the user for organizing information

All participants were given a scratch sheet at the start of each search. The two sets in the AI/ZP comparison did 24 searches with ZP and 24 searches with AI. Scratch sheets were used on 12 of the ZP searches and only 7 of the AI searches. For the groups doing AIP/ZP, 10 of 24 used scratch sheets on ZP searches and only 7 of 24 used then on AIP searches. For all the experimental systems, scratch sheets were used on 14 out of 48 searches, and for all the controls, scratch sheets were used on 22 of 48 searches. These proportions lead to a χ^2 statistic of 2.844. $\chi^2_{0.05,1} = 3.841$ so these results are not significant at $p < 0.05$. We noticed that the use of scratch sheets varied greatly between participants, with one subject using a sheet on all six searches and seven subjects never using a sheet. The eight participants that sometimes used scratch sheets used them on 19 out of 24 searches with ZPRISE but on only 11 out of 24 searches with the experimental systems. These ratios have a χ^2 value of 5.689 which is significant at $p < 0.025$.

Of the 36 searches where the participant used scratch sheets, the participant was a librarian 26 times. This is a significant difference in the behavior of the two populations ($\chi^2 = 11.378$). Librarians used scratch sheets on 14 of 24 ZPRISE searches and on 12 of 24 InQuery searches for an insignificant difference ($\chi^2 = 0.336$). Students used scratch sheets on 8 of 24 ZPRISE searches and only 2 of 24 InQuery searches, for a χ^2 value of 4.547 (significant at $p < 0.05$). Only three students ever used a scratch sheet. For them the χ^2 value is 8.1. The hypothesis is supported.

Providing visual cues about which documents have been seen is useful.

After performing multiple searches and reviewing a large number of documents it is easy to forget the headlines of non relevant documents read early in the search. We colored the headlines blue (for unread documents) and purple (for read documents) so that the user would not accidentally reread something previously read and deemed non relevant. We decided to compare the difference between the number of documents read during a search and the number of unique documents read. Sometimes it is necessary to reread a document to check on a piece of information. However, a large number of documents reread might be symptomatic of wasted effort.

Of the 48 searches using ZPRISE, 3.771 document rereadings occurred on average, with a standard deviation of 6.821. For the experimental systems, 2.875 rereadings occurred with a standard deviation of 3.517. The differences between these groups has a t-statistic of 0.808 which is not statistically significant. Dividing the sample into librarians and students does not change the result. The results suggest the hypothesis is possible, but do not support it otherwise.

Terms that are automatically extracted from a group of documents will be useful in describing those documents.

Participants in the study identified 396 aspects, and 368 of these were labelled. Of the 368 labelled aspects, 168 of the labels were one or more of the automatically generated terms, and an additional 20 labels used one or more of the terms with additional user-added terms. For over half of the labelled aspects the user did not feel that any of the presented terms described the content satisfactorily. Presenting the words to the user may save them some time, and is often useful, but allowing the user to choose their own labels is clearly important.

Successful usage of 3-D interfaces will correlate with structural visualization.

Structural visualization is the ability to mentally manipulate three dimensional objects. It is one of the oldest known psychometric aptitudes[11] having been discovered as one of the defining traits of successful engineers. It correlates strongly with success in careers that deal with designing and manipulating solid structures, such as engineering, science, architecture and surgery. The Paper Folding Test (test VZ-2) is a standard test for measuring this aptitude [10].

Three dimensional interfaces are inherently more complicated than two dimensional interfaces, and have a great number of methods of interacting with them. During usage of the system features will often become obscured and the user must mentally keep track of these features. We instrumented the different windows in the system to measure how much interaction took place in each one, recording all mouse clicks and all interactions with controls. There was a large variation in the amount of interaction with the 3-D window, with two participants never using it at all, and two participants interacting with it over 250 times. There were equally extreme qualitative reactions to the window, with the term "worthless" being used more than once, to several participants asking why this window was not available on every search engine.

VZ-2

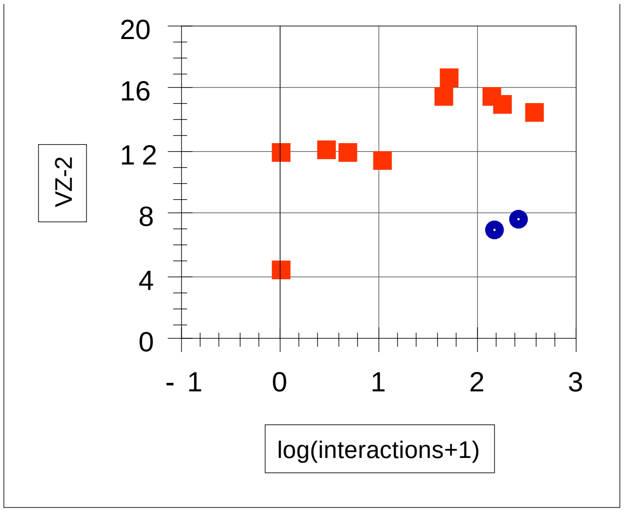


Figure 3. Visualization Score vs. 3-D Interaction

coefficient (r_s) between interaction levels and the score on VZ-2 is 0.75. For the four librarians in group 4 the rank correlation is 0.95. Combining these two groups to obtain a larger sample we get a rank correlation of 0.9018, significant at $p < 0.01$.

Adding in the four students in the fifth group the correlation drops to 0.2413. This is a very surprising result. The last two participants used the 3-D window a tremendous amount, but they both scored very poorly on VZ-2. We tend to be suspicious of this result for two reasons: the last two participants in the experiment were unfortunately tested by a different administrator, and they both had an educational background that implies they should have had high scores on VZ-2. (One had just completed an MS in Chemistry, and the other has a BS in Physics and is working towards an MS in Mechanical Engineering.) Figure 3 shows a scatterplot of the scores on VZ-2 plotted against the log of the number of 3-D interactions plus one. The data being questioned is represented by circles. Calculating a rank correlation on the ten remaining points gives a value of r_s of 0.8818 which is significant at the level of $p < 0.005$. Usage of the 3-D window correlated with ratings of usefulness of the window with $r_s = 0.8916$, $p < 0.002$. Usage of the 3-D window correlated with reported experience with 3-D with $r_s = 0.5$.

We still believe that a strong correlation exists between usefulness to an individual of a 3-D interface and their structural visualization ability, but it will require more testing, with extra care to avoid data contamination.

CONCLUSIONS

We have presented the results of a user study that evaluates how easily a searcher can use a task-specific Information Retrieval system to accomplish that task. In this work we analyzed and discussed issues related to the usability of the interface: were the provided tools used readily and accurately by searchers. We did not address issues of effectiveness (were the searchers successful in finding aspects?), though they are covered elsewhere. [1]

Our hypotheses were generally supported, though usually not with statistical significance. Specifically, we found that:

1. Participants tended to use the task-specific "aspect window" when it was available, indicating that it was helpful. However, these results were not significant.

2. There was apparent advantage to showing the user that he or she had previously seen a document (e.g., the blue/purple highlighting), but the results were not significant.

3. Automatically providing key terms and phrases for the participants' use was helpful, but not significantly so. Our evaluation of this feature was indirect: we know whether participants used the terms, not whether they felt they were correct.

4. There is strong evidence of a correlation between structural visualization ability and facility with a 3-D interface.

We also found the unexpected (though not unreasonable) result that GUI interfaces are less acceptable to participants who are less familiar with mouse-driven displays.

Most of the results in this study are preliminary because of the nature of the problem. The TREC Interactive Track required that we evaluate our system against the ZPRISE system. Unfortunately, the systems are so different that we feel the abrupt and dramatic change from one system to another was a likely cause for some user confusion. For example, a 3-D display is likely to be overwhelming to someone who is also being introduced to the "aspect window". Further studies will be needed to draw more accurate conclusions. The fifth group of participants was the start of that process for us: we were comparing AspInquery against an enhanced version of AspInquery (by adding the 3-D window). We anticipate continuing with experiments such as that one, that are more incremental in nature and therefore allow for better control of the variables.

ACKNOWLEDGEMENTS

We thank Victor Lavrenko, Anton Leouski, Darren Mas, and the experiment participants for their assistance. This material is based on work supported in part by the National Science Foundation under grant number IRI-9619117, and in part by the National Science Foundation, Library of

Congress and Department of Commerce under cooperative agreement number EEC-9209623, and supported in part by Defense Advanced Research Projects Agency/ITO under ARPA order number D468, issued by ESC/AXS contract number F19628-95-C-0235.

Any opinions, findings and conclusions or recommendations expressed in this material are the author's and do not necessarily reflect those of the sponsor.

REFERENCES

[1]Allan, Callan, Croft, Ballesteros, Byrd, Swan, Xu "INQUERY does battle with TREC-6", *TREC-6 Conference Proceedings*, 1998 (to appear).

[2]Allan, J., Leouski, A., and Swan, R. "Interactive Cluster Visualization for Information Retrieval", Tech Report IR-116, CIIR, University of Massachusetts Computer Science Department,1997

[3]Callan, J., Croft, W. B. and Harding, S. "The INQUERY Text Retrieval System" in *DEXA-3: Third International Conference on Database and Expert Systems Applications* (1992)

[4]Chalmers, M. and Chitson, P. "Bead: Explorations in Information Visualization" in *Proceedings of the 15th International ACM/SIGIR Conference*, pp 330-337, Copenhagen, Denmark, 1992

[5]Croft, W. B. *Organizing and Searching Large Document Collections*, PhD Thesis, University of Cambridge, 1979

[6]Ekstrom, R.B., French, J. W., Harman, H. H., and Dermen, D. *Manual for Kit of Factor-Referenced Cognitive Tests 1976 Educational Testing Service*, Princeton, New Jersey 1976

[7]Harman, Donna *The Fifth Text REtrieval Conference*, 1997 (Forthcoming)

[8]Hearst, Marti A. "Visualization of Term Distribution Information in Full Text Information Access" in *CHI '95 Conference Proceedings*, ACM Press, pp 59 - 66

[9]Hearst, Marti A. and Pedersen, Jan O. "Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results" in *Proceedings of SIGIR '96* (Zurich, Switzerland, August 1996) ACM Press

[10]McGee, Mark G. *Human Spatial Abilities* Praeger Publishers, 1979

[11]O'Connor, Johnson *Structural Visualization* Human Engineering Laboratory Inc. 1943

[12]Swan, Russell C and Allan, James "Improving Interactive Information Retrieval Effectiveness with 3-D Graphics" Tech Report TR-96, IR-100, University of Massachusetts Computer Science Department,1996

[13]TREC Interactive Specification at http://www-nlpir.nist.gov/~over/t6i August 1997

[14]van Rijsbergen, C. J *Information Retrieval*, second edition, Butterworths, 1979

[15]Veerasamy, Aravindan and Belkin, Nicholas J. "Evaluation of a Tool for Visualization of Information Retrieval Results", *Proc of SIGIR '96*, pp. 85-92.