Explicit verses Implicit:
An Analysis of a Multiple Search Result Visualization

Edward Suvanaphen and Jonathan C. Roberts
University of Kent at Canterbury
es45@kent.ac.uk, j.c.roberts@kent.ac.uk

Abstract

When searching on the web, users often reformulate their queries after viewing the results and viewing some of the pages. After one or two reformulations the user may implicitly realize patterns and relationships between the multiple search results. We believe that these patterns can be used to identify interesting results. We have developed the prototype Search Engine Similarity (SES) tool which explicitly visualizes the similarity between multiple searches.

In this paper we describe an experiment to determine whether explicitly visualizing the relationships between multiple searches will let users browse more effectively. Our results show that explicit difference visualizations can enhance the search process for some tasks.

Keywords—**Multiple Search Result Visualization, Information Visualization, Evaluation, Comparison visualization**

1   Introduction

Web searching is the process of locating information on the Internet. It is understood that web searching is an information seeking process. Many attempts have been made to model this process (see [1], [9], [11], [16]). In its simplest form it can be described as a set of stages.

Stage 1, the user forms a query appropriate for the information retrieval, and inputs this into a web search engine interface such as Google [7] or Yahoo [18].

Stage 2, the results are returned and displayed to the user in a rank ordered list. The user views the meta-data provided, (e.g. the title, snippet of text, web address and size) and from this decides whether to view any pages.

Stage 3, The user selects the result they wish to view and the web page is returned.

Stage 4, After the user has viewed a sufficient number of results, the user will either quit the search or reformulate their query using new terms discovered during their search session. This may result in a comparison of two or more web searches, in which users will unconsciously begin to realize similarities and differences between the data sets.

We believe that by explicitly visualizing the associations between the results sets, rather than relying on the
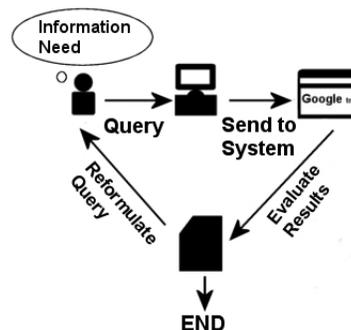

Figure 1: A model of the Information Seeking Process

user to implicitly learn such differences, the user will gain a better understanding of the searched information and will be able to more effectively browse. Such associations between multiple data sets can be pre-attentively visualized, through the use of visual cues. Traditionally comparison can only be achieved by displaying two web searches side by side in separate windows and inspecting each result in turn. This can make locating similarities and differences between multiple searches difficult.

Furthermore, each stage in the information seeking process takes a certain amount of time [16] and thus the time spent processing increases as the user views more and more results. Consequently, the comparison of several reformulations can increase the search process time dramatically.

Our comparison tool (SES) visualizes the associations between multiple search lists using a textual difference visualization. We believe that by drawing the users attention to significant results, we will create a more efficient browsing process. Explicitly visualizing the associations between the searches will help the user identify interesting results more accurately, early in the search process, thus decreasing the search process time. Visualizing these associations will also reduce the cognitive overhead gained when switching between two different result sets. We will describe the tool in more detail in section 3.

In this paper we will examine the usefulness of this explicit visualization and analyze its effectiveness in identifying relevant results in comparison with traditional methods of search result browsing.

2 Related Work

There are a number of search result visualizations that compare multiple sets of data, examples include SQWID [13], Sparkler [8], and Vr-vibe [15]. Both Sparkler and SQWID plot their data onto a two dimensional vector space, and use distance to determine the relevancy of results.

Sparkler uses search result lists from several different search engines, each list being displayed as a single dotted line emanating from the center of the screen. Each result in the list is represented by a dot, with each dot placed in ranked order (highest closest to the center). Results which exist in more than of the lists, are linked through similar colour shading.

SQWID on the other hand, obtains a list of interesting words from the resultant search result list. Three of these interesting words form the points of a triangle, and search results (represented by rectangular nodes) are positioned within the triangle based on their relevancy to each of the 3 words.

The Vr-vibe visualization utilizes a three dimensional space to display the results of several search queries of corpus document data. The user specifies keywords that they wish to use to generate the visualization which are positioned in three dimensional space. Representations of each document are then positioned in the space according to how relevant each document is to each of the keywords. Another three dimensional visualization is the three dimensional scatter plot which formed part of the NIRVE project [3]. Three dimensional glyphs were plotted in three dimensional space where each axes represented a term used in the search. This allowed users to compare the relevancy of each result to each search term used.

Each of the techniques described so far use graphical mappings. While there are advantages to graphical presentations (e.g. they can show more results in a smaller screen space, and more parameters can be incidentally depicted), there are also a number of disadvantages. For example graphical displays are often more abstract (requiring users to learn the various associations), they are less precise because descriptive information is lost, and users are less familiar with graphical tools.

There are also a number of tools used for textual comparison and differencing, in such application areas as program code analysis, literature comparison and plagiarism detection.

The majority of tools operate on two files side-by-side, annotating the differences (the 'side-by-side with cues' cat-

egory of [5]). Usually these visualize the differences between datasets as opposed to the similar results, this is because the difference is usually a smaller set which is easier and clearer to visualize.

The majority of file-difference tools are classified as 'side-by-side with cues' techniques [5]. Usually these visualize the differences between datasets as opposed to the similar results, because these are usually a smaller set, and thus easier to visualize in a way that will draw the users attention.

Vdiff [4] and WinDiff [17] are both examples of file difference tools. Vdiff places two files side-by-side and uses a combination of colours and bounding boxes to identify areas of similar and different code. Boxes of matching code are joined by lines so as to visually associate them, when the windows are scrolled, associated parts of code stay aligned keeping identical parts adjacent. WinDiff merges all the data into one view, and uses colours to represent different and similar pieces of text. SeeSoft [6] is a tool that visualizes multiple text files simultaneously. Each file is depicted in a column, and every line of code in that file is graphically represented by a greeked line. Greeking exchanges the characters of the text with straight lines and hence is useful for providing summary information about a document. Robertson and Mackinlay use greeking as part of their Document Lens[14].

3 The SES Tool: Search Engine Similarity

In this section we briefly describe the SES visualization tool, a more detailed description of the tool can be found in [5]. When designing the tool our aim was to explicitly visualize the textual difference of various search results. A screenshot of the system can be seen in Figure 2.

The user begins their search session by inputting either two or three sets of search terms which share a common theme, subject or are related in some way. These results are then submitted to Google via the Google API [7]. The results are then returned and displayed on three coordinated views (the summary view, overview panel and bracketed view). Interaction with the visualization can be achieved through the use of scrollbars, or by clicking on results within the bracketed view. Users can open webpages by clicking on the titles of the results.

A problem inherent in text-based interfaces is the lack of screen space. Because of the large amounts of textual data we needed to visualize, we used both detail in context and multiple views techniques to display the data.

An important aspect of the visualization is the Coordination between the three views. Colour is a good visual cue to use to link similar results (Gestalt principle of similarity [12]), and as such is used in our visualization. Colour highlighting is used to link the focused result in the bracketed view, to results with the same web address in the other
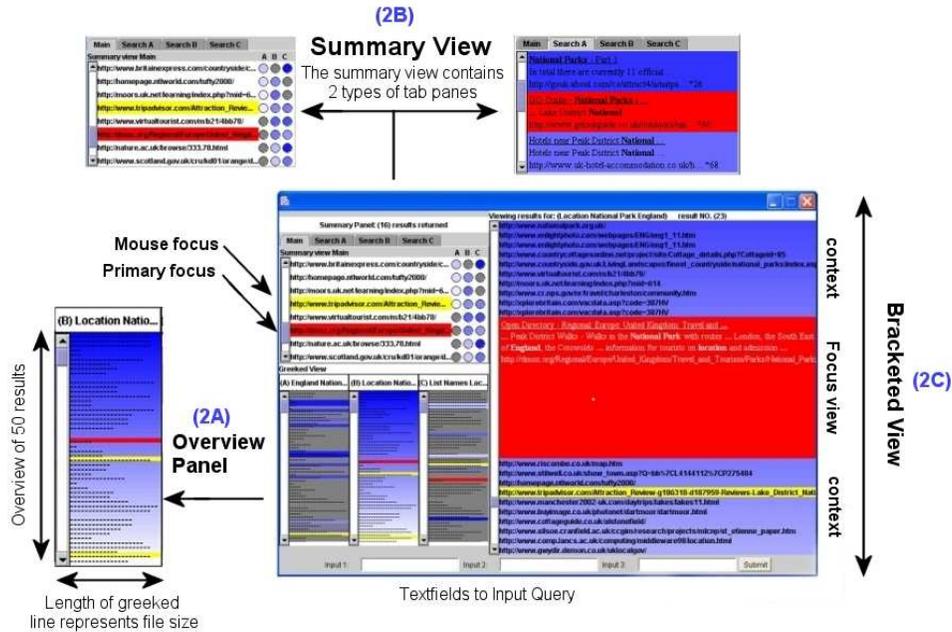
Figure 2: The SES (Search Engine Similarity) comparison tool. There are three main views, the overview panel (2A), the summary panel (2B), and the bracketed view (2C). The summary view consists of two tabs, the main tab(left) contains all the results which appear in more than one search. The Search Tab (right) containts the results from a particular list, that exist in multiple lists. (These are displayed as full detail SRE's)

views. Similarly the mouse focus uses colour highlighting to link the result the mouse pointer is currently over, with its other instances in other views. Colour is also used in the overview panel to identify results which exist in more than one list.

3.1 The summary view

The summary view (Figure 2A) is located in the top left of the main window. This displays results which exist in at least two of the search result lists. The results are displayed on a tabbed pane. The main tab displays a summary of the data, with the results displayed as a list of URLs, each result also is followed by three circles (one for each search) which are shaded to a value equivalent to its ranking in its list (or gray if they do not exist in that list). Each of the other three tabbed lists show detailed information for each specific search.

3.2 The overview panel

The overview panel (Figure 2B) is located in the bottom left of the main window and provides a context for the other views, enabling users to identify the position of their results. The panel uses greeking in order to display the whole list in one view, the length of the line is proportional to the size of the file that the URL represents. The position of the results in each of the columns is displayed in rank order. Whereas the colour will change according to the current selected column, such that when the user selects

one column (e.g. column A) the results in this column are highlighted in a colour proportional to its rank, the intersections with the other lists (columns B and C) are shown in a shade equivalent to the rank of the selected column (column A).

3.3 The bracketed view

The bracketed view (Figure 2C) is located on the right-hand side of the main window. This view is based on the visual bracketing principles in [10]. The view shows the currently focused result in full detail, with results either side of it displayed as URLs. The bracketing principle is a focus-in-context technique that allows the user to view on one element in detail, while still being aware of his surrounding context.

4 Experiment

In this section we will discuss an experiment to test the effectiveness of the SES tool compared to that of a standard ranked ordered list visualization (Google). It was our hypothesis that users would benefit from a tool that explicitly displayed the similarities between multiple sets of results. Indeed, that users would be more efficient at choosing relevant results (stage 2 as detailed in the introduction). Thus to effectively test this hypothesis and because our aim was to concentrate on the second stage of the search process, all other elements in the environment had to be controlled.

This meant that certain steps in the information seeking process had to be carefully choreographed so as to eliminate any influences from these phases. Consequently, the users were given pre-determined search terms, they were restricted from viewing webpages, and restricted from using advanced search options.

The information retrieval task given to the user was presented in the form of a question. Appropriate questions had to be chosen carefully. As was shown by Dempsey et al. [2] it is often difficult to find a neutral question when designing a study. They discovered that certain key variables (such as proper names) were excellent discriminators and greatly affected the results.

The task was to "List the names and depths of the Worlds Deepest Seas". The terms given were (1) "World's Deepest Oceans", (2) "Seas Deepest Depth, and Oceans", (3) "What are the World's Deepest Seas".

The test subjects comprised of a mixed gender group of 20 people from various backgrounds. At least half of the group were non-computer professionals but all had a working knowledge of web searching using search engines. One half of the group (the control group) used Google, the other half (the test group) were given the SES tool.

Each subject was monitored in three ways, (1) through their interactions with the keyboard and mouse, (2) through the data submitted via the user interface, and (3) by human observation. Monitoring software was used to capture screen shots of the users interactions on the screen using the mouse and keyboard. Each of these capture sessions was stored as both a set of thumbnails and an animation. All the users interactions with the Internet were passed through an Apache server which had been setup with the proxy module enabled, ensuring that all requests were filtered through a script that both controlled and automatically monitored the information returned to the user.

Before starting each test each user was given a demonstration as to how they were to conduct their tasks. They were also given an opportunity to familiarize themselves with the controls of their interface (the Google search engine, for the control group, and the SES tool for the test group). A webpage interface was used to guide the users step-by-step through the process.

Users in both the control and test groups were given a search scenario and three sets of search terms (as stated previously).

In the control study each of the sets of search terms was a hyperlink which opened up a modified Google web page (via our proxy), where links to all of Google's facilities (e.g. image search, news search) outside of its web search were disabled. Users were then told to browse the results and select what they thought were the five most relevant results to the search. To mark a result as relevant,

users clicked on the title of the result, which would open up a ranking page (as stated previously we deliberately restricted users from browsing web pages) on the rating page the users could rate how relevant they thought the result was based on the meta-information alone and compared to the scenario given. The results were rated from one to three (one being least relevant), See table 1. Users were required to locate five relevant results for each of the three sets of search terms.

In the test study, the three sets of search terms were entered into SES and the comparison between the result sets was displayed. As with the control group, the test group users had to select five results which they deemed relevant to the scenario, and rate each of these. Finally Users filled in a short questionnaire at the end of the test.

Rating	Criteria
1	The Web page doesn't contain the data you are seeking, but the information can be found on another page in the website, or via a link on the website.
2	The Web page contains some of the data required but not all
3	The Web page contains all the relevant data

Table 1: Criteria for selecting relevant web sites

4.1 Method of analysis

We analyzed the results using two weightings strategies. First, the users ratings were compared with those determined by a small study group (R1). Weightings were assigned proportionally to the difference between the two results (table 2). Second, users were allocated a point for each relevant result they selected, irrespective of any rating given (R2).

		Users ratings		
		1	2	3
	1	+2	+1	0
Expert	2	+1	+2	+1
Ratings	3	0	+1	+2

Table 2: Additional points given based on a comparison of users ratings and expert ratings

We performed a statistical analysis of the results, using the T-test method; this method assesses whether the mean of two data sets are statistically different from each other. This is important because a comparison on averages alone does not take into account the spread or variability of the scores. For example the difference between two

means may be the same for two experiments, but one may have a high distribution across the data set, and the other a lower distribution. The higher distribution will hence have a larger overlap than the lower distribution, and therefore will be a more similar set of results (and hence less significant) than the lower distribution data.

$$\frac{X_t - X_c}{\sqrt{\dfrac{\sigma^2{}_t}{N_t} + \dfrac{\sigma^2{}_c}{N_c}}} = \frac{Difference\ between\ group\ means}{Variability\ of\ groups} \qquad (1)$$

$X_t =$ Mean of the test group	$X_c =$ Mean of the control group
$\sigma^2{}_t =$ Standard Deviation of test group	$\sigma^2{}_c =$ Standard Deviation of Control Group
$N_t =$ Number of users in the test group	$N_c =$ Number of users in the control group

Once a result has been obtained from the T-test formula (see equation 1), the number coupled with an alpha/risk level, and degree of freedom is compared against a table of results. The Alpha level is usually set to 0.05 (meaning there is less than a 5% possibility that the difference between the means is due to chance) and the degree of freedom is the sum of the persons in both groups minus 2. If the result is greater than the arbitrary value in the table then the value is generally accepted as significant. The T-test formula was used to analyse the results using both weightings (R1) and (R2).

5 Results

The results of both datasets (R1) and (R2) (seen in Figure 3) show that the test group (those using SES) out performed the control group (those using Google) when choosing and rating relevant results. However when the T-test was applied to the data set R1 (relevant results with weightings), a value of 1.726 was returned. This was lower than the designated value in the table of 1.734 which showed that the variance between the two sets was not enough to make the results significantly different. But, when the T-test was applied to the data set R2 (relevance results solely), a value of 2.307 was returned showing that the variance between the results was significantly different.

We conjecture that the reason for the discrepancy in the results of R1, is due to the process of rating pages. In the post-study questionnaire over 50% of the users, expressed their unease about having to rate a page purely based on the meta-information returned by Google. They could use the meta-information to locate what they thought was a relevant page, but could not determine the exact relevancy of the page. We believe that by asking users to rate the pages

we forced a change in the users search behaviour and thus the results became skewed.

	Pages of web results		
	Page 1	Page 2	Page 3 and over
Search 1	46%	44%	10%
Search 2	28%	30%	42%
Search 3	60%	16%	2%

Table 3: The percentage of relevant results chosen at each page. Each page contains ten results.

Another measure is to evaluate how the users interacted with Google. We discovered that on average nearly all the users of the Google interface viewed results beyond the the first ten results, this is shown in table 3. Previous experience had suggested that users of the traditional search result tools would rarely view beyond the first ten or twenty results, which is contrary to the results shown. We believe that one reason for this is that, by controlling the users environment (not allowing reformulation, or viewing of web pages) we unintentionally changed the users search process. The user could not reformulate their query after the first ten results, and was forced to view results further down the list.

However an important and interesting discovery was that 72% of the users of the SES interface used the Summary view to search for relevant results from which 83% were correct. The Summary view in the SES interface was created with the aim of providing the users with a quick and easy view from which similar results could be located. Our results reinforce this theory, and suggest that search difference visualization is useful and enables users to more quickly drill down to interesting results.

6 Conclusions

We have conducted an experiment to determine the effectiveness of a Multiple Search Result visualization compared with traditional search result visualization. Our results show that by explicitly visualizing the relationships between multiple results lists, users can better identify relevant results.

Our experiment also shows that the summary view, (figure 2 - where only the items which exist in more than one list are shown) is an effective way of locating relevant results. This is supported by both data obtained from the questionnaires and user observation during the experiment.

However our study has also demonstrated that the search result meta-data (title, text snippet, size and web address) does not provide enough information to accurately determine the relevancy of individual results. This means that by restricting the user from viewing web pages during the
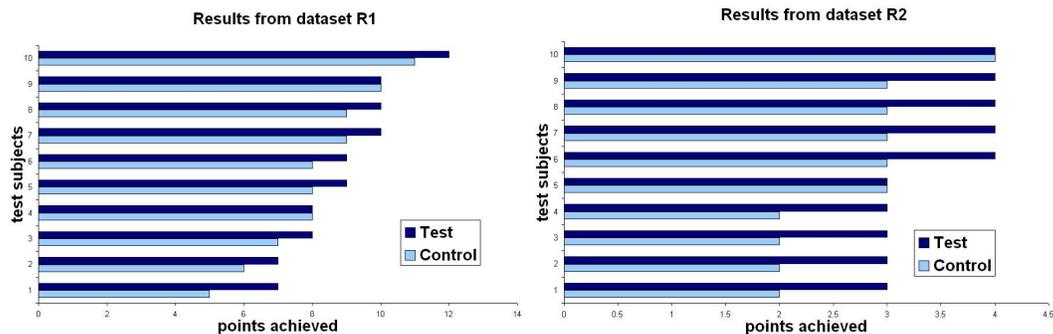
Figure 3: Each bar represents a test subject (totalling 20 people), subjects from the test group (SES) are coloured in dark blue, while those from the control group (Google) are coloured in light blue. The two graphs show the results using different weightings (R1) and (R2) (left and right respectively).

experiment their search process will be affected. On the contrary, by allowing the user freedom to explore different pages, the emphasis of 'what is being tested' would be changed to testing the whole information seeking process, rather than evaluating the visualization effectiveness.

Both observed and the feedback from the questionnaires have inspired ideas for future experiments and visualizations. For instance, suggestions were given on how to improve the usability of the SES interface. One idea is to create an interface that allows users to compare any number of multiple search results at any time (even historically).

Acknowledgements

References

[1] Ben Shneiderman, Donald Byrd, D W Bruce Croft. Sorting out searching: a user-interface framework for text searches. *Communications of the ACM*, 41(4):95–98, 1998.

[2] B.J.Dempsey, R. C. Vreeland, R. G. Sumner, and K. Yang. Design and empirical evaluation of search software for legal professionals on the www. *Inf. Process. Manage.*, 36(2):253–273, 2000.

[3] J. Cugini, S. Laskowski, and M. Sebrechts. Design of 3d visualization of search results: Evolution and evaluation. In *Proceedings of IST/SPIE*, Visual Data Exploration and Analysis, pages 198–210, January 2000.

[4] D. J. Barnes and M. T. Russell and M. C. Wheadon. Developing and adapting UNIX tools for workstations. In *Proc of EUUG*, pages 321–333, 1988.

[5] E. Suvanaphen and J. C. Roberts. Textual Difference Visualization of Multiple Search Results utilizing Detail in Context. In *Theory and Practice of Computer Graphics*. IEEE Computer Society Press, June 2004. To appear.

[6] S. G. Eick, J. L. Steffen, and E. E. Sumner. Seesoft - a tool for visualizing line oriented software statistics. *IEEE Trans on Software Engineering*, 18(11):957–968, 1992.

[7] Google Web API. www.google.com/api, March 2004.

[8] S. Havre, E. Hetzler, K. Perrine, E. Jurrus, and N. Miller. Interactive visualization of multiple query results. In *2001 IEEE Symposium on Information Visualization (InfoVis '01)*, pages 105–112. IEEE, Oct. 2001.

[9] M. A. Hearst. *User interfaces and visualization: in Modern Information retrieval bu Baeza-Yates and Ribeiro-Neto*. Addison Wesley Longman, 1999.

[10] Jonathan C. Roberts and Edward Suvanaphen. Visual bracketing for web search result visualization. In Ebad Banissi et al, editor, *Proceedings Information Visualization (IV03)*, pages 264–269. IEEE Computer Society, July 2003.

[11] G. Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Pres, 1995.

[12] Max Wertheimer. Laws of Organization in Perceptual forms (Untersuchungen zur Lehre von der Gestalt II). *Psycologische Forschung* , 4:301–350, 1923.

[13] S. McCrickard and C. Kehoe. Visualizing search results using SQWID. In *Proceedings of the Sixth International World Wide Web Conference*, 1997.

[14] G. G. Robertson and J. D. Mackinlay. The document lens. In *Proceedings of the 6th Annual Symposium on User Interface Software and Technology*, pages 101–108. ACM Press, Nov. 1993.

[15] Steve D. Benford and Dave N. Snowdonand and Chris M. Greenhalgh and Rob J. Ingram and Ian Knox and Chris C. Brown. VR-VIBE: A virtual environment for cooperative information retrieval. *Computer Graphics Forum*, 14(3):349–360, 1995.

[16] A. Veerasamy and R. Heikes. Effectiveness of a graphical display of retrieval results. In *SIGIR '97: Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 27-31, 1997, Philadelphia, PA, USA*, pages 236–245. ACM, 1997.

[17] WinDiff Tool, microsoft sdk tools, March 2004.

[18] Yahoo Search engine. www.yahoo.com, March 2004.