

# Tangible Interface for Collaborative Information Retrieval

Alan F. Blackwell, Mark Stringer, Eleanor F. Toye and Jennifer A. Rode

Computer Laboratory, University of Cambridge

J. J. Thomson Avenue, Cambridge CB3 0FD, UK

{afb21, ms508, eft20, jar46}@cl.cam.ac.uk

## ABSTRACT

Most information retrieval (IR) interfaces are designed for a single user working with a dedicated interface. We present a system in which the IR interface has been fully integrated into a collaborative context of discussion or debate relating to the query topic. By using a tangible user interface, we support multiple users interacting simultaneously to refine the query. Integration with more powerful back-end query processing is still in progress, but we have already been able to evaluate the prototype interface in a real context of use, and confirmed that it can improve relevance rankings compared to single-user dedicated search engines such as Google.

## Author Keywords

Tangible Interfaces, Computer Supported Collaborative Argumentation, Probabilistic Information Retrieval.

## ACM Classification Keywords

H.5.3 [Group and Organization Interfaces]: Asynchronous interaction, Collaborative computing. H.3.3 [Information Search and Retrieval]: Relevance feedback, Search process.

## INTRODUCTION

Information retrieval systems, especially query languages and search engines, are notable as one of the few areas of user interface research that has concentrated almost exclusively on single users at individual workstations. We are not aware of any examples of information retrieval systems that are designed for collaborative use in a group setting. This is increasingly problematic in contexts where online information or local digital archives are used in collaborative situations such as an investigatory, planning or design meetings.

We have developed a collaborative query technique, suitable for use in this type of context, and validated it in preliminary trials in an educational context. We had a

number of goals going into our project, towards which progress is being made at this stage:

- to allow multiple users to contribute to query construction. This is being achieved by applying probabilistic information retrieval (IR) approaches.
- to allow several people at once to interact with a shared interface. This is being achieved by use of a tangible user interface (TUI) whose elements can be handled simultaneously and collaboratively by several members of a group.
- to support this activity as a secondary task rather than forcing users to focus on query construction. This is being achieved by conceptualizing the collaborative situation as a semi-structured process of reasoned rhetorical argument.

This combination of techniques, which we call Query By Argument (QBA), has undergone initial evaluation in an educational context.

The rest of this paper describes the general approach of query by argument, then discusses the specific technical approaches applied to the design of the TUI and probabilistic IR. We also give a brief summary of the evaluation that we have conducted so far.

## RHETORIC AND COLLABORATION

Many systems have been developed that use computer-supported collaborative argumentation (CSCA) as a technique for facilitating meetings. These systems impose logical structure on discussion so that it can be encoded, manipulated and visualized (often diagrammatically). A well known example in the technical domain is the Questions Options Criteria system that was developed for the capture of design rationale [3]. Other systems have applied first order predicate logic, for example, to structure legal arguments in industrial negotiation [9].

Unfortunately, there are many domains of collaboration where individual contributions to the discussion cannot easily be classified according to their propositional content. Furthermore, when a large amount of information is being introduced into a discussion (for example relevant documents that have been gathered through web queries), it is likely to be too laborious to precisely encode every piece of relevant evidence.

We have therefore taken a semi-structured approach, in which contributions are classified according to their

rhetorical function, rather than their propositional content. Participants in the discussion collect relevant pieces of evidence from the query system, selecting relevant passages from documents, and providing a few words to summarise how each piece of evidence relates to the discussion. These summary statements can then be arranged into a rhetorical structure, based on outline argument “scaffolding” that has been agreed by participants earlier. Contributed statements and selected evidence are gradually arranged into a coherent summary of the available evidence, which can then be presented or debated.

We use the juxtaposition of earlier search results to provide relevance feedback for a probabilistic IR system. The summary statements provide additional ranking keywords, and selected portions of documents provide positive and negative (unselected) example texts. Statistical information gathered during the discussion can then be used for query refinement, so that when further evidence on a specific topic is needed, query sets are ranked according to probable relevance to critical aspects of the discussion – those where several sources of evidence are combined, or where particular documents have been referred to more often.

This approach satisfies our third primary design goal – an information retrieval system in which we do not force users to focus on query construction as a dedicated task, but instead retrieve relevant information as a by-product of activity that is in itself of value to the users’ goals.

### TANGIBLE USER INTERFACE

Our early design work was based on an observational study of users building a structured argument based on material that they had found on the Web. This alerted us to the value of tangible materials in facilitating discussion. Users worked with print-outs of web pages, displaying them to the group, sharing them with neighbours, annotating them, or pointing to relevant passages. As noted by Sellen and Harper, paper is an extremely valuable and versatile tangible information device [6]. Furthermore, we observed all of these things being done in parallel by multiple users. While two people at one end of a table are leaning together over a piece of paper, someone at the other end might be annotating a different document. This kind of real-time simultaneous contribution satisfies our second primary design goal for a shared interface.

We therefore concentrated on preserving the advantages of paper for collaborative discussion, but in a manner that would provide Query By Argument through electronic interaction with a back-end IR system. Some of the main components are shown in use in figure 1.

The QBA TUI is based on radio-frequency ID (RFID) technology. RFID tags are embedded in “statement tokens”, which are used as place-holders for contributions to the discussion. When a user selects a relevant piece of evidence, or proposes a novel contribution, the summary

statement is printed on a card which is inserted into the statement token.



Figure 1. Tangible User Interface, showing statement tokens and rhetorical activities

The agreed rhetorical structure is printed onto larger labels attached to racks that can hold a number of statement tokens. These racks contain RFID reader aerials, and are connected to a central workstation which constantly monitors the position of tokens in the racks. As users collect evidence, and arrange material into the agreed structure, the system records every manipulation as the basis for relevance feedback. There is no need for users to coordinate their activities – they can work in small groups or individually, with the TUI becoming a focus of attention, a site of deictic reference, or simply a background aide-memoire as the discussion flows.

As the structure of the argument is built up, the whole group receives feedback through a graphical overview of the argument structure that is displayed on a monitor or (as in our evaluation trials) projected on a wall of the meeting room. This means that users receive immediate feedback with respect to their primary task, while the system continues to collect relevance feedback unobtrusively without interrupting the discussion. There are several other elements of the TUI that can be used to display specific parts of a document when a completed argument is presented or reviewed. We do not have space in this paper to discuss these additional functions (see [8] for details).

### BACK-END INFORMATION RETRIEVAL

Our initial approach to creating a multi-user query language was based on the assumption that information relevant to the discussion would be stored in some kind of local repository, and that the query would be a deterministic query against that repository. In deterministic IR, the goal is to separate a set of documents into those that are relevant and those that are irrelevant. When the initial set of candidate documents is quite small (a few thousand), a good query might return three or four documents that the user can then review.

However, when we considered the Web as a source of discussion material, we realised that even a good query can result in thousands of relevant documents. We therefore focused on probabilistic IR, in which the goal is not to partition a set of documents, but rather to rank them according to probability of relevance. In our QBA system, the probabilities are derived from collaborators' interaction with the TUI.

We employed the approach of Robertson and Sparck Jones to implement our IR back-end [5, 7]. At the start of a QBA session, the topic of discussion is established either by defining a small set of keywords as for a conventional search engine, or by providing URLs of online resources known to relate to the discussion. Either type of initial input can be used as the starting point from which to produce a *collection* of documents. Our partners in the European WEBKIT project have developed a web harvester robot that collects initial material by following links from a set of seed URLs. Alternatively, we have also used the QBA system as a front end to a conventional search engine, by calling the Google API [1] with the argument topic keywords as initial search terms.

We then derive an *index*, listing all unique words that occur in the collection. We reduce words to their stems using Lovins algorithm [2], and delete stopwords (“the”, “and” etc.) that are likely to appear in every document. The result is a set of indexing *terms* that can be used for calculating relevance. The terms are *weighted* according to how well they discriminate between documents in the collection. Terms that occur in only a few documents are more valuable than common ones, while a term that occurs many times in a single document is highly related to that document. If the document is short, then the term is even more closely related than if it occurred sporadically in a long document. These three weighting factors:  $CFW_{(i)}$  (frequency of a term  $i$  in the collection),  $TF_{(i,j)}$  (frequency of term  $i$  within document  $j$ ) and  $NDL_{(j)}$  (normalised length of document  $j$  relative to other documents), are combined to give an array of weights as follows:

$$CFW_{(i,j)} = CFW_{(i)} * TF_{(i,j)} * (K_1 + 1) / K_1 * ((1-b) + b(NDL_{(j)})) + TF_{(i,j)}$$

Equation 1: Index Weighting. The tuning constants  $K_1$  and  $b$  can be improved by experiment. Robertson and Sparck-Jones, based on their studies of the TREC benchmark for information retrieval, recommend  $K_1=2$  and  $b=0.75$ , and we currently use these values [5].

It is now possible to use this weighted index for interactive query refinement using the TUI. As an initial phase, participants review part of the collection (as they would do after any web search), each choosing one or two documents that relate to their own individual concerns. Participants then summarise their concerns by providing summary statements that are printed onto statement tokens for use during discussion. The vocabulary used in these statements is processed in the same way as for other documents in the collection: stemming, removing stopwords and weighting the terms.

When the TUI tokens are manipulated, grouped and presented during discussion, the available collection of documents can be re-ranked according to evidence of cumulative relevance. The collection can also be ranked relative to the statement on a specific token, so that any user can request further information relevant to a particular topic by placing one of their tokens on a reader to see the most closely related documents.

While reviewing the documents in the collection, participants can also highlight parts of a web page that they consider particularly important – these passages of text can be used to modify relevance weightings ( $RW_{(i)}$ ) for each term in the selection. Finally, all the terms taken from the initial selection of relevant documents, the position statements, and the selected highlights are used to choose the best (most selective) terms in the index according to their *offer weight* ( $OW_{(i)}$ ). The terms with the best offer weights characterise those parts of the collection that are most relevant to the overall argument. These terms can then be added to the keywords that were provided in the initial query, running a further Google API or harvester search that is more specific to the concerns that have arisen during discussion. At this point new documents may be added to the collection, refining it in response to the emerging concerns of the group.

Almost all of this processing occurs in the background, as a by-product of natural discussion processes, rather than requiring individual users to spend time composing and refining search engine queries. It is not necessarily the most efficient approach to information retrieval in its use of computational resources, but it is efficient in its use of human attention. Furthermore, the eventual query is a collaborative product, reflecting the concerns of whole group, rather than those of any individual member. This satisfies our first goal, of allowing multiple users to collaborate in query construction.

## EVALUATION

The interaction techniques that we have employed (choosing an initial set of relevant material, summarising discussion points in position statements, highlighting important text, and using position statements as tokens) have been developed on the basis of our observation of actual activities during classroom discussions directed by a schoolteacher. We have also built a series of prototype TUIs in order to evaluate their effectiveness in actual discussion situations.

### QBA as a Tool for Argument

We have conducted a series of eight preliminary evaluation trials using incremental generations of TUI prototype. Participants are 11-year old children working on Key Stage 2 and Key Stage 3 of the English National Curriculum [4]. The process of researching and developing an argument is relevant across the curriculum in several subject areas, where children are required to justify what they think based on listening to others (“Speaking and Listening”), use

persuasive techniques and rhetorical devices (“Writing”) and justify orally and in writing a personal opinion about issues, problems or events (“Citizenship”).

We conducted our evaluation trials during actual classroom sessions, when students were focusing on these subject areas. Trials were conducted by a researcher who was a qualified teacher, and observed by two further researchers who were present in the classroom, and made video recordings of participants’ interaction with the interface. Each trial extended over more than one class session that was devoted to a particular discussion topic. Children first made use of material retrieved from the web to select and formulate summary statements. They then created an argument structure based on the rhetorical activities.

We found that children were enthusiastic about the TUI, and required little instruction from the facilitator as to how to use it. Together the group discovered all of the functionality of the TUI and GUI by exploration. Direct benefits were seen in three respects: firstly, in the organization of material into a structured argument, secondly, in recognizing where further evidence for the argument is required, and finally in the transformation of structured material into prose.

#### **QBA as a Tool for Query Refinement**

The TUI has proven successful as a multi-user interface, allowing several participants to interact with the application simultaneously, and avoiding a common problem with shared GUI interfaces in which one individual dominates the relevance judgements. The final system resulting from the WEBKIT project will use a novel web harvester and repository as a back-end for this interface. The retrieval performance of that system will be evaluated once the back-end implementation is complete, but in the meantime, we have been able to conduct a preliminary assessment in terms of improvement in relevance ranking relative to the standard behaviour of Google.

As a concrete example, in a research discussion on the topic of Ubiquitous Computing that we have used to demonstrate our current prototype to colleagues, we compared the ranking results of a real Google query to the ranked list produced by our refined query. A core publication in ubiquitous computing is Weiser’s seminal 1991 paper [9]. In our original Google result set, a link to this paper was ranked number 68. Newcomers investigating the topic without knowing of this paper would have to browse 6 pages of results before encountering it. The refined query based on relevance feedback from our interface ranked this paper at the top of the list.

#### **CONCLUSION**

These preliminary results indicate that we have succeeded in creating a novel IR interface that achieves our three original design goals. It allows multiple users to contribute to query construction, it allows several people at once to interact with a shared query interface, and it supports this activity as a secondary task rather than forcing users to focus on query construction. Although this interface is yet to be integrated with the complex back-end system to be provided by our partners on the WEBKIT project, we believe these early results indicate a good chance of success in achieving a new paradigm for user interaction with information retrieval systems.

#### **ACKNOWLEDGMENTS**

Amanda Simpson conducted the classroom observation and evaluation work. Chris Vernall of CRL built the physical prototypes. Stephen Robertson and Karen Sparck-Jones provided advice on probabilistic IR. This research is funded by European Union grant IST-2001-34171 “WEBKIT”. This paper does not represent the opinion of the EC, which is not responsible for any use of this data.

#### **REFERENCES**

1. Google Web APIs. <http://www.google.com/apis/>
2. Lovins, J.B.. Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics* 11 (1968), 22-31.
3. MacLean, A., Young, R.M. Belotti V. & Moran T.. Questions, Options, Criteria: Elements of Design Space Analysis. *Human-Computer Interaction* 6(1991), 201-50
4. National Curriculum. <http://www.nc.uk.net/home.html>.
5. Robertson, S.E. & Sparck Jones, K. Simple, proven approaches to text retrieval. University of Cambridge Computer Laboratory Technical Report (1994). no. 356.
6. Sellen, A.J. & Harper, R.H.R. *The Myth of the Paperless Office*. MIT Press (2001).
7. Sparck Jones, K., Walker S. & Robertson, S.E. A probabilistic model of information retrieval: development and comparative experiments. *Information Processing and Management* 36 (2000), 779-840.
8. Stringer, M., Rode J.A., Blackwell A.F, and Toye E.F. "Facilitating Argument in Physical Space." *UbiComp* 2003. 123-6.
9. van Gelder, T. Enhancing deliberation through computer supported argument visualization. In *Visualizing Argumentation; Software tools for collaborative and educational sense-making*. Springer Verlag (2003) 97-115.
10. Weiser, M. The computer for the 21st century. *Scientific American* 265 (1991), 94-104.