A model of cognitive load for IR: implications for user relevance feedback interaction

Jonathan Back and **Charles Oppenheim**
Department of Information Science
Loughborough University, UK

Abstract

Preliminary experimentation demonstrates that although a system that allows for context specific spontaneity is likely to improve the acquisition of user relevance feedback, this is by no means a solution. We believe that the problems associated with a lack of willingness to modify queries or to provide relevance feedback within the Web environment, are indicative of a high state of cognitive load. A number of task and cognitive variables exist and need to be identified before a model of cognitive load for IR can be developed.

Introduction

Web search engine queries, on average, comprise less than three search terms. Comprehensive query formulation is a rarity. Query modification is not a typical occurrence. Jansen, Spink, and Saracevic (2000) found that a considerable majority of Excite search engine users (67%) did not submit more than one query, and only 5% of users utilized User Relevance Feedback (URF). A number of search engines have attempted to incorporate URF to enable query modification. URF is an interactive process in which the users are encouraged to utilise their domain knowledge to allow the generation of more comprehensive queries. The effectiveness of an information retrieval (IR) system can be measured in terms of how long it takes for a user to find sufficient relevant information, or discover that no relevant information exists. A typical Web query can retrieve hundreds of thousands of results. Document relevancy ranking is therefore important in minimising the time spent by an individual searching. Within the laboratory environment, experimentation demonstrates that URF can be used to provide improved document rankings (Harman, 1992).

However, implementation of an on-line URF mechanism is not straightforward, as every interaction with a document changes a user's state of knowledge. The view of relevance might, as a consequence also change; URF should reflect the changing needs of a user and be captured within the context of a user's relevancy determination process. Search engines do not support browsing behaviour. Viewing a document, navigating back to the search engine page and then submitting a dichotomous relevance judgement is clearly not an ideal method of obtaining URF. It is not surprising, therefore, that very few users are prepared to do this. Users must be able to revise their query and relevance judgements to reflect their fluctuating understanding while they browse documents./P>

Preliminary findings

Ranking error is the degree to which the actual document rankings, performed by a Web search engine, deviate from the optimal rankings, identified by experiment participants. Preliminary experimentation performed by Back, Oppenheim, and Summers (2000) utilised URF for automatic query expansion.

The figure below shows the extent ranking error can be minimised by the use of different URF capturing techniques.

Query expansion

Ranking error (%)

30
25
20
15
10
5
0

0 1 2 3 4 5 6

Number of terms used

URF (a)　URF (b)　URF (c)
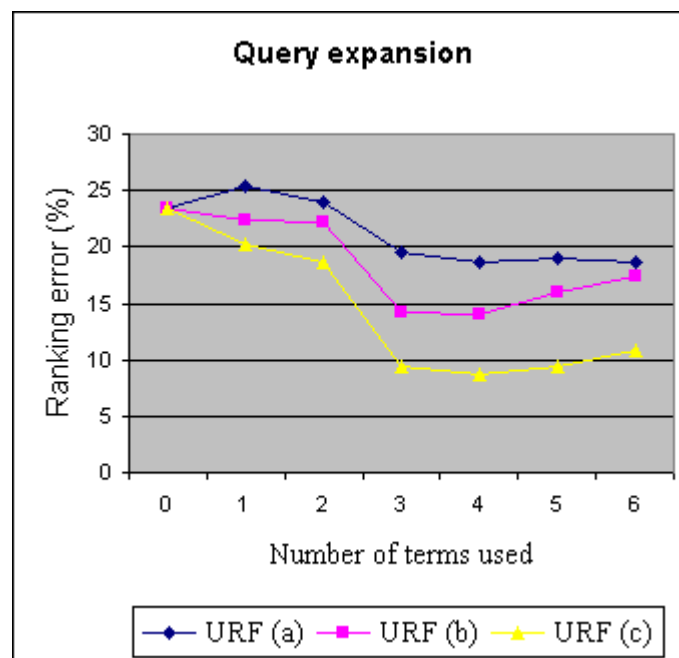
URF (a): The 'check-box' approach to obtaining relevance judgements. This approach is the most commonly adopted and is also the least successful.

Relevance

☑

URF (b): Unsurprisingly performs better than URF (a) because the relevancy associated with a document has been specified to a greater degree. Outside the laboratory environment, specifying the degree of relevancy can be complicated. A user would be required to spend a considerable amount of time cross-evaluating documents.

Relevance

○　Highly Relevant
◉　Partially Relevant
○　Not Relevant

URF (c): A contextually specific and spontaneous method of obtaining relevance judgements. While users browse documents, they are asked to mark what they consider to be the most relevant passages using a highlighting tool. This was the most successful approach.

Relevance

This passage is not relevant the user has decided not to highlight it.

This passage is relevant the user has therefore decided to highlight it.

This passage is not relevant the user has decided not to highlight it.

Problems with on-line URF implementation

Although a system that allows for context specific spontaneity is likely to improve the acquisition of URF, this is by no means a solution. The majority of Web search engines favour automated feedback techniques. Those that have implemented URF have found very little evidence of improved search effectiveness. Many researchers and practitioners suggest that information visualisation that incorporates URF is the future for IR on the Web. Information visualisation techniques aim to limit the number of results returned to a user by clustering. This can

accelerate the user's relevancy determination process considerably. However, it can be argued that this activity does not support the cognitive model of a user because domain knowledge is only gathered at an abstract level. The user may as a consequence discard relevant clusters during the problem definition stage. Actual domain knowledge is essential in order to make an accurate relevancy judgement. Furthermore, the value associated with irrelevant or partially relevant information when acquiring domain knowledge is not appreciated by visualisation techniques.

Cognitive load

The term "cognitive load" has been used loosely within IR research, mostly in reference to Human Computer Interaction (HCI) issues. Although the term has been utilized frequently, the only IR study that has attempted to define the concept was performed by Hu, Ma, and Chau (1999). They examined the effectiveness of designs (graphical or list-based) that best supported the communication of an object's relevance. Cognitive load was used as a measure of information processing effort a user must expend to take notice of the visual stimuli contained in an interface and comprehend its significance. It was assumed that users would prefer an interface design that requires a relatively low cognitive load and at the same time, can result in high user satisfaction. A self-reporting method was used to obtain individual users' assessments of the cognitive load associated with a particular interface. The focus of this study was interface design, so the use of the term "cognitive load" was valid from a HCI viewpoint. However, as we will attempt to demonstrate, the concept of cognitive load during IR can be extended far beyond interface design.

In IR, the concept of cognitive load rarely extends beyond the ideas presented by Miller (1956). In Miller's famous paper "The magical number seven plus or minus two", a human's capacity for processing information was explored. It was concluded that short-term memory (working memory) has a limited retention. The study by Hu, Ma, and Chau is typical of much research in IR that advocates attempts to minimise cognitive load during interface design by recognising the limitations of working memory.

Some of the more insightful studies in IR have shown that recognising the limitations of working memory may not be the only method of minimising cognitive load. Beaulieu (1997) suggested that there is a need to consider cognitive load not just in terms of the number and presentation of options, but more importantly to take account of the integration and interaction between them. Beaulieu, however, was not the first, as Chang and Rice (1993) had proposed that interactivity could reduce cognitive load. Although these studies point to 'interaction' as being an important factor, they do not explain why or how.

Many IR researchers use the term "cognitive load" with a limited understanding of Cognitive Load Theory. This theory has been developed by educational psychologists and is documented by Sweller (1988; 1994). Learning structures (schemas) are used during problem solving. IR can be viewed as a problem solving process (Kuhlthau, Spink, and Cool, 1992). The psychologist Cooper (1998) explains that Cognitive Load Theory can be used to describe learning structures. Intrinsic cognitive load is linked to task difficulty, while extraneous cognitive load is linked to task presentation. If intrinsic cognitive load is high, and extraneous cognitive load is also high, then problem solving may fail to occur. When intrinsic load is low, then sufficient mental resources may remain to enable problem solving from any type of task presentation, even if a high level of extraneous cognitive load is imposed. Modifying the task presentation to a lower level of extraneous cognitive load will facilitate problem solving if the resulting total cognitive load falls to a level within the bounds of mental resources.

Clearly, there is more to the concept of cognitive load than careful interface design. In IR it would be tempting to associate intrinsic cognitive load (task difficulty) with query difficulty, and extraneous cognitive load (task presentation) with interactivity. However, this would be too simplistic. A greater number of task and cognitive variables exist and need to be identified before cognitive load in IR can be defined.

> We believe that the problems associated with a lack of willingness to modify queries or to provide RF within the Web environment, are indicative of a high state of cognitive load. **Cognitive load can be measured by the difficulty associated with providing a relevancy judgement.** If the cognitive load is high, then providing the system with URF is unlikely.

Theoretical model

Kuhlthau (1993) explained that uncertainty is a cognitive state which commonly causes affective symptoms of

anxiety and lack of confidence. Uncertainty due to a lack of understanding, a gap in meaning, or a limited construct, initiates the process of information seeking. She suggested that six corollaries exist. We have simplified and re-interpreted these corollaries as follows.

Process corollary
- Understanding information results in a shift from uncertainty to clarity.

Formulation corollary
- Individual interpretations of information may result in uncertainty or clarity.

Redundancy corollary
- Balance of redundant (known) and unique (unknown) information is critical. Too much either way results in uncertainty.

Mood corollary
- An invitational mood is more appropriate for the early stages of the search (definition) and an indicative mood is more appropriate for the later stages (resolution). Uncertainty may arise if this is not the case.

Prediction corollary
- If expectations are not met then uncertainty may increase.

Interest corollary
- Interest may lead to a reduction of uncertainty due to intellectual engagement.

Kuhlthau concluded by suggesting that uncertainty can result in a user being less prepared to interact with a system. We believe that uncertainty can be considered as one of the components that contribute to cognitive load.

Wilson et al. (2000) showed that it is possible to measure the level of uncertainty experiment participants have at each stage of the problem-solving process in which they are involved. Wilson et al. speculated that two different ideas of uncertainty exist. Affective uncertainty is associated with affective dimensions such as pessimism/optimism. Cognitive state uncertainty is associated with more rational judgements about the problem stages.

Complex data collection during an empirical investigation of cognitive load will enable a data-driven model to be derived instead of a purely theoretical one. A wider range of task and cognitive data will be collected, enabling new components of cognitive load to be uncovered. *Cognitive load is not, as currently assumed, based on only the following three components:*

Domain knowledge: User's knowledge of the information need under investigation. Cognitive load reduces as more domain knowledge is captured.

Cognitive state uncertainty: User's overall level of doubt associated with the search process. Cognitive load reduces as the user becomes more confident that their information need can be addressed.

Retrieval performance: Cognitive load increases as the number of potentially relevant documents identified by the IR system increase.

The figure below outlines our novel theoretical model of cognitive load for IR during the search process. This model may provide an insight into why different types of URF are required during the search process, and why users are sometimes unprepared to provide a system with URF.
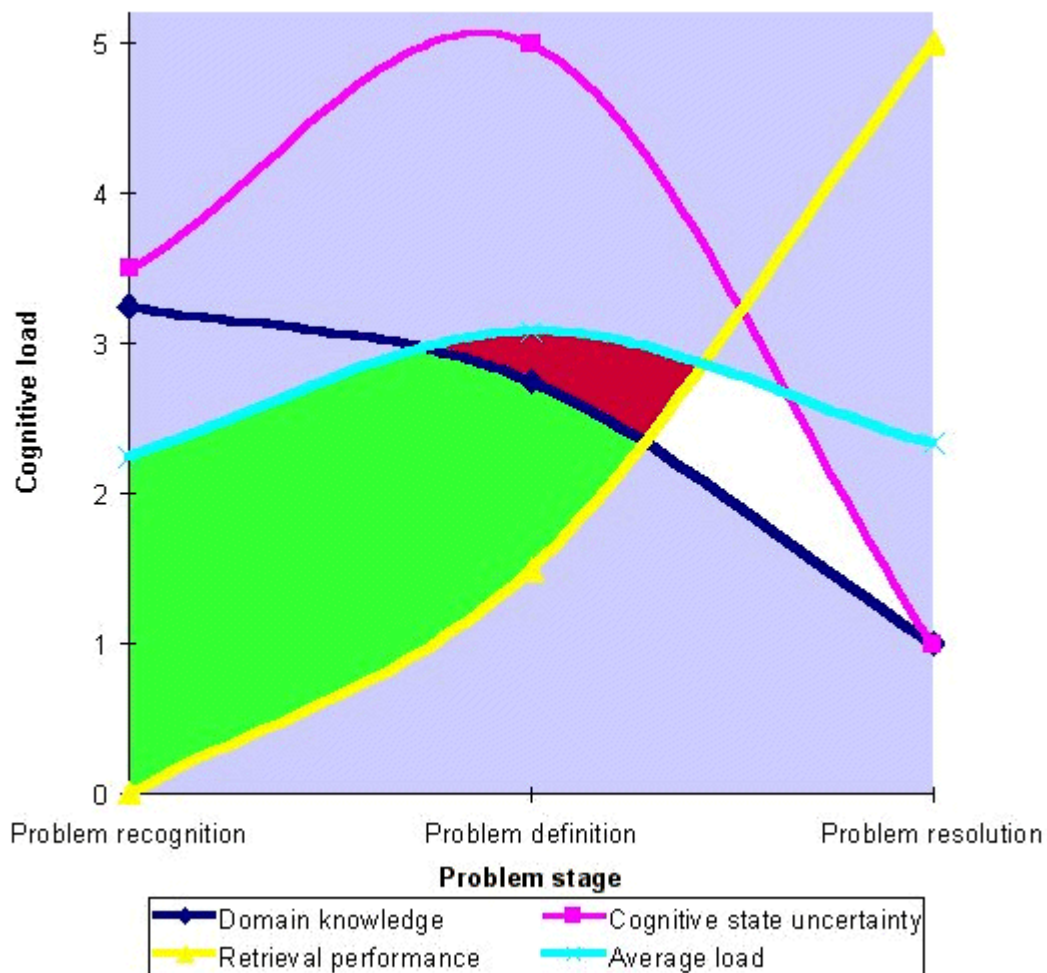
Key to URF types:

Term-suggestion RF [green]
Judgement RF [red]
Visualisation RF [white]

All URF techniques:

Usability is limited by average load. URF techniques that place too much cognitive load on the user are unlikely to be utilised.

Term-suggestion URF:

Located above *the retrieval performance line* indicating that it is a recall tactic. Useful when domain knowledge is limited and prevents the user from generating their own terms for query expansion.

Judgement URF:

Located above *the retrieval performance line* indicating that it is a recall tactic. Could be used as a precision tactic but negative relevance judgements are too discriminating within the Web search engine exact match environment. Judgement URF is useful when domain knowledge is sufficient to make accurate document relevancy judgements.

Visualisation URF:

Located below *the retrieval performance line* indicating that it is a precision tactic. Useful when domain knowledge is sufficient to make accurate cluster relevancy judgements. Unlikely to be used when the *cognitive state uncertainty line* approaches *the domain knowledge line* because the user's information need has possibly been sufficiently addressed.

Research status

Now that a theoretical model has been proposed, it needs to be tested. The primary objective of our experimental methodology is to justify the proposed components of cognitive load and to identify new ones. Collection of task and cognitive data will be extensive, enabling a data-driven model to be established.

The first stage of experimentation will evaluate a range of URF techniques within the Web environment. Simulated work tasks will be assigned to experiment participants. Queries submitted to the system will be pre-defined; participants will only be able to modify the query using URF. Pre-defined queries will be short, representative of a typical Web search engine query. Participants who have had extensive IR experience will be able to select the URF technique that they consider to be the most appropriate for a particular problem stage of the search process. Participants who have had limited IR experience will be restricted to utilising a specific URF technique. This first stage of experimentation will test the hypothesis that the effectiveness of a specific URF technique is dependent on user domain knowledge, user cognitive state uncertainty, and the problem stage, i.e. URF effectiveness is limited by cognitive load.

The second stage of experimentation will be an evaluation of prototype software that we will develop. The software will attempt to predict the most appropriate URF technique at a particular problem stage. This prediction will be based upon both the system state and the user state. The acquisition of cognitive variables will be required to enable a prediction. The difficulty associated with acquiring these variables without increasing cognitive load will be investigated. This second stage of experimentation will test the hypothesis that the acquisition and utility of URF needs to be optimised by considering cognitive load, thereby allowing significant IR performance improvements.

References

- Back, Oppenheim, & Summers. (2000). Relevance statistics: how important is spontaneous relevance feedback during Web search engine usage? Paper submitted to the *Journal of the American Society for Information Science*.
- Beaulieu, M. (1997). Experiments on interfaces to support query expansion. *Journal of Documentation, 53*, (1), 8-19.
- Chang, S. & Rice, R.E. (1993). Browsing a multidimensional framework. *Annual Review of Information Science & Technology, 28*, 619-627.
- Cooper, G. (1998, December). Research into cognitive load theory and instructional design at UNSW. University of New South Wales, Sydney, Australia. http://www.arts.unsw.edu.au/eductaion/CLT_NET_Aug_97.html
- Harman, D. (1992). Relevance feedback and other query modification techniques. In: Frakes, W. B. and Baeza-Yates, R. (Ed.). *Information Retrieval Data Structures and Algorithms*. (pp. 241-263). Englewood Cliffs, New Jersey: Prentice Hall.
- Hu, P.J.H., Ma, P.C., & Chau, P.Y.K. (1999). Evaluation of user interface designs for information retrieval systems: a computer-based experiment. *Decision Support Systems, 27*, (1-2), 125-143.
- Jansen, B.J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the Web. *Information Processing and Management, 36*, 207-227.
- Kuhlthau, C. (1993). A principle of uncertainty for information seeking. *Journal of Documentation, 49*, 4, 339-355.
- Kuhlthau, C., Spink, A., & Cool, C. (1992). Exploration into stages in the information search process in online information retrieval. *Proceedings of the ASIS annual meeting, 29*, 67-71.
- Miller, G.A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.
- Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295-312.
- Wilson, T.D., Ford, N.J., Ellis, D., Foster, A.E., & Spink, A. (2000, August). Uncertainty and its correlates. Paper presented at *Information Seeking in Context, 2000*, Gothenburg, Sweden.