

## Panel: User Modeling and User Interfaces

**Susan T. Dumais**

Bellcore  
445 South St.  
Morristown, NJ 07960  
std@bellcore.com

### The User Knows Best

I will consider the problem of helping user's retrieve information from large diverse collections of textual materials such as bibliographic databases, electronic bulletin boards and online manuals, although I believe that many of my arguments apply to other interactive environments as well. Information retrieval systems like this must handle a heterogeneous population of users, and even a single user will have a range of information needs including very specific requests as well as more general and often ill-defined topical searches.

The idea of employing user models to enhance human-computer interaction in IR applications is very appealing. The hope is that some form of encapsulated knowledge about the user will mediate between user input and system responses resulting in more effective, efficient and enjoyable interactions (Rich, 1983; Sleeman, 1985). There are, however, few success stories, especially in the sense of working systems that are demonstrably better than similar systems that don't have a user model. One reason for this, I believe, is that researchers have underestimated the difficulties involved in discovering what user characteristics are related to performance differences and, more importantly, how the system should respond to different users (i.e. what to do about it). At a more pragmatic level, obtaining information that can be used to reliably characterize different users is difficult except in the simplest cases. A system which responds differently to users who speak different languages (e.g. English vs. French) might be described as having a simple but effective 'user model'. In this example: it is clear that language differences pose a major barrier to retrieval; detecting what language the query is in is relatively straightforward; and tailoring the system response to be in the appropriate language is easy. Handling expert-novice differences is quite difficult by comparison, but is still a relatively easy user modeling problem. How can a user's expert or novice status regarding a particular query be detected on the basis of implicit or explicit information? How should the information presented to experts and novices differ - in amount, kind, level of detail, etc.?

When does it matter? The general case of information retrieval is especially difficult because of the variety and dynamic nature of users, tasks and materials.

Most existing text retrieval systems work by matching words in (or assigned to) objects against words in users' queries. While these systems are quite useful, they are far from perfect. Many retrieval problems can be traced to the fact that individual words or phrases provide unreliable evidence about the meaning or topic of a text object. There are many places in the retrieval process where improvements could be targeted - indexing or describing text objects and user requests; matching requests and objects; modeling inter-object relationships; and supporting the search process. Proponents of user modeling have argued that better information requests could be generated by exploiting models that understand user's goals, problem contexts, knowledge, etc. There is little evidence that such models can be successfully constructed or exploited to enhance retrieval.

I agree that user's queries are seriously deficient, but argue that the most effective tools for flexible and customizable query specification are those that give users better ways to speak for themselves. The user knows more about his/her interests, goals, and state of knowledge than could be communicated to a search intermediary, or defined, abstracted and exploited by most modeling mechanisms. Given existing query methods (e.g. generating natural language queries or complicated Boolean expressions), users find it quite difficult to describe their information needs. With appropriate tools for users to iteratively and interactively specify queries, and to view the results of searches, many of these barriers can be removed. These ideas are not new (e.g. Oddy, 1977), but there is now adequate computer power to implement and to evaluate them with large collections.

Egan et al. (1989), for example, describe an online text browser, SuperBook, that was designed to enhance user's abilities to search in and navigate through electronic texts. In behavioral studies, they found that users who searched for information using SuperBook could find answers more

quickly and accurately than their counterparts who used a paper version of the same book. The advantages of SuperBook come from: full-text indexing which increases the likelihood that users will find the information they need; the posting of matched query terms against a dynamic hierarchical view of the document which enables user to quickly find the most relevant passages; and automatic page reformatting and highlighting to accentuate the users' current search interests. The dynamic hierarchical display is a very simple kind of domain model; but it is one that users understand and can exploit.

In bibliographic retrieval situations where it is more difficult to capitalize on inter- or intra-document structure other search enhancements are often quite effective. Retrieval performance can be improved by a number of automatic means that incorporate simple statistical information about the collection of objects and terms - e.g. differentially term weighting or post-coordination to rank objects from most to least similar to the query. Larger improvements (averaging about 60%) can be obtained by automatically constructing a new query based on user feedback about which of the objects returned in response to an initial query are relevant (Salton & McGill, 1983). Since user's initial queries are usually quiet impoverished, the feedback process can be viewed as an easy and reliable method for generating rich, context-appropriate information requests.

In the two cases described above, the ability of users to efficiently and effectively retrieve information was greatly enhanced by providing them with flexible methods for crafting queries and understanding the system's responses.

## References

- Egan, D. E. et al. 1989. Formative design evaluation of SuperBook. *ACM TOIS*, 7(1), 30-57.
- Oddy, R. N. 1977. Information retrieval through man-machine dialogue. *Journal of Documentation*, 33, 1-14.
- Rich, E. 1983. Users are individuals: Individualizing user models. *International Journal of Man Machine Studies*, 18, 199-214.
- Salton, G. and McGill, M. J. 1983. *Introduction to Modern Information Retrieval*, McGraw-Hill, N.Y.
- Sleeman, D. H. 1985. UMFE: A user modelling front-end sub-system. *International Journal of Man Machine Studies*, 23, 71-88.