

A Research Program for User-Based Information System Design: Towards Resolution of the Representation Problem in Information Retrieval

Gregory B. Newby
School of Information Studies
Syracuse University

The quantity of information available in the world has grown tremendously over the history of the field of information science. It is paradoxical that information system design continues to follow essentially the same recipe as it has from the beginning -- hampering our ability to effectively tap the information resources potentially available to us. Because of the lack of a successful paradigm for research in information system design, ability to produce information has far exceeded the ability to store and retrieve the information. It is proposed here that an approach to information system design which relies on the system's users for design specifications is the approach that will yield the most success.

Traditional information systems have been designed on one of two logics. The engineering approach involves building a bridge between what the system can do and mnemonic terms for the user to employ to access system functions. The enhanced engineering approach goes beyond the engineering approach by producing an interface which is user friendly. For both approaches, system design specifications, including interface specifications, are generated by the system designers themselves. A third approach to system design is the user-based approach. The user-based approach starts with the philosophy that the users of the information system are the best starting place for system design. By identifying the processes that users go through, the problems they have, and the terminology that they use, it is possible to generate an information system that not only meets the user's needs, but also requires minimal training. In the current context, information systems are not limited to those which are computer-based. Other information systems include books, user manuals, brochures, and so forth.

Effective methodologies for eliciting user requirements for information system design have been demonstrated for on-line help systems, training systems, and several other computer-related and non-computer-related activities. It is anticipated that the vendors who produce software based on user requirements will necessarily be more successful than vendors using the engineering approach, for the complimentary reasons that the software provides what the user needs, when the user needs it, and without the learning curve associated with traditional software.

Information retrieval may be the nemesis of information science. There are few computer applications which can claim the low performance that information retrieval has demonstrated, and with very little improvement over time. It is suggested that the implementation of a user-based approach to information retrieval (IR) system design may yield the increases in performance which have been sought for so long using engineering approaches. Information retrieval systems have a more difficult task than many other computer applications. The variety of users, the size of the database, and the inexhaustible supply of different user needs combine to make for a class of information systems which are difficult to design by any method.

Perhaps the most difficult task of IR systems is "matching." The general purpose of IR systems is to match a user's expression of an information need to the most appropriate document representations in the system's database. Matching may be seen as composed of two tasks. The first task is translation of the user query to terms appropriate to the system document representation (this is most often

accomplished by the user). The second task is to select the set of document representations which most closely match the translated version of the user query. In more advanced systems, especially experimental systems, there is an attempt to have the system perform the translation of the user query to system terms.

An important reason for the difficulty of both components of matching is that IR systems' document representations have very little basis in human experience. Both boolean and orthogonal vector systems make use of a set of keywords used to describe the documents. No matter which of the variety of variations are included in these systems -- thesauri, user modeling, relevance feedback, etc. -- the document representations are limited to a set of keywords (usually without even a set of relations among keywords). Research in cognitive science has generated a wide variety of representation schemes which people use. These psychological schemes offer little basis for reliance on those schemes used in information retrieval systems.

Document representation has been chosen as a starting place for the development of user-based specifications for information retrieval system design. A multi-method approach to the elicitation of user representations will be taken. The goal of the research is to identify useful representation schemes that users use for sets of documents, and investigate the appropriateness of these schemes for IR systems.

It is anticipated that the representation of documents by user-derived schemes will drastically reduce or eliminate the need for the translation component of the matching function. It may also change nature of the matching component by introducing alternative representation schemes for the system database. The production of an isomorphic relationship between user representations and system representations will enable users' direct access to the documents a system holds, without concern for insuring that the users' expressions of information needs are in terms appropriate to the system's database representation.

Research on user representations for IR systems will begin with a series of unstructured interviews, with a focus on identifying potential "dimensions" or criteria by which documents may be represented. More structured interviews will allow a focus on identifying the processes that users go through when forming representations of documents. Additional methodologies will be chosen based on their appropriateness to the criteria and processes which are identified, and may include field studies of user groups, or creation of spatial representations of documents using multidimensional scaling or other methodologies relying on spatial metaphor. Theory and research supports models of human representation based on mental "maps." Insight into differences among user groups' representations, and the appropriateness for the incorporation of these differences into IR system design specifications, will be sought.

There is little doubt that the successful information systems of the future will be designed to user-based specifications. The force of market pressure, and the necessity of shortening or eliminating the learning curve, provides an environment ripe for user-based approaches to system design. Information retrieval is a stronghold of the engineering approach to information system design. Traditional IR systems accept but ignore the implications of inconsistencies between IR system design and human approaches to information seeking behavior: key terms are not independent, users do not think in boolean algebra, all users are not alike, and user models should incorporate a focus on process, not just demographics (to name a few). The user-based approach to IR system design is a step towards the incorporation of user-based design for all information systems, along a road that represents the future of information science.

----- Notes:

1. Adapted from Michael S. Nilan, Gregory B. Newby, Woojin Paik, and Kevin Lopatin (1989) "User-oriented interfaces for computer systems: A user-defined online help system for desktop publishing." in Jeffrey Katzer and Gregory B. Newby, eds. Proceedings of the 52nd Annual Meeting of the American Society for Information Science. Medford, NJ: Learned Information.

2. See: Joseph Woelfel and Edward L. Fink (1980) The measurement of communication processes: Galileo theory and method. New York: Academic Press.

For an introduction to cognitive science with a focus on the role of mental maps, see: Philip N. Johnson-Laird (1988) The computer and the mind: An introduction to cognitive science. Cambridge: Harvard University Press.