Dagobert Soergel
College of Library and Information Services
University of Maryland

An Information Science Manifesto

American Society for Information Science Award of Merit Acceptance Speech

Washington, DC, November 5, 1997

Good evening. I am gratified that a jury of my peers has found my thinking important enough to honor it with this award. I thank the society, which has been my professional home for 35 years, the committee, and all others involved in the process.

This honor brings with it the task of explaining the essence of information science in 10 minutes or less.

Before I do that, I want to acknowledge an intellectual debt by paying homage to one of the great thinkers and visionaries of our field, Calvin Mooers. I met Calvin for the first time at an IFIPS meeting in Munich in 1962. After his paper on the nature of descriptors and the structure of classification, a small group of us had an hour-long discussion energized by his intellectual spark. This and my subsequent study of his work was the starting point of a long journey that brought me here and that has not come to an end yet. Many of the ideas in his 1959 paper *The next twenty years in information retrieval* could still be published today under the same title. Calvin Mooers pioneered user-oriented indexing 40 years before the term user-centered became fashionable.

I also thank my students who, through their questions and discussions, have contributed much to sharpening my thinking.

I will begin my remarks with my definition of information science.

Information science is concerned with both research and design. It conducts research into the nature of information, its creation, organization, use, and impact. It studies information needs and the interaction between people and information. It combines conceptual structures with appropriate technology in the design of systems for information sharing, retrieval, and access, as well as information assimilation, processing, and learning. Information science is concerned with conceptual foundations: with the structure of knowledge and the structure of problems; with problem solving and decision making and their synergy with external information; with information processing, learning, and reasoning by people and machines; and with the role of information in social groups and organizations. As information is pervasive in much of human activity, information science knows no boundaries; it overlaps with many fields and draws on their ideas, methods, and results: mathematics and statistics; computer science, artificial intelligence, and human-computer interaction; cognitive science with its constituents cognitive psychology, linguistics, epistemology, and philosophy of knowledge; communication and journalism; education; economics; anthropology, sociology, and political science; and, finally, administration and management. As it incorporates principles and results from these disciplines, information science analyzes them for commonalities, derives generalizations, advances the thinking, and thus creates a new   synthesis.

At the heart of information science are the twin concerns of

(1)   understanding users in their quest for meaning and problem solutions and

(2)   representing knowledge structures that support the construction of meaning and the solution of problems.

User models and knowledge representation, like a binary star, revolve around each other and depend on each other. To know what is important in knowledge representation, we need to know user requirements and user thought patterns. To understand users and their backgrounds, we need to represent their knowledge and their information needs.

The tension between the two poles of information need and stored information representations defines relevance, perhaps the central concept of the field. To understand relevance, we must understand the user model and the representation of knowledge and the complexities of their interrelationships.

Against this background let me outline some important tasks ahead. I will speak about two ingredients of success, knowledgeable systems and knowledgeable users working together. I will speak about the next revolution, tight integration of information seeking and use with learning and work.

Everything we do must ultimately revolve around serving users well. The term *user-centered* has come into vogue, but I prefer *user-oriented*. The interactions with an information system and with the information provided must take the user beyond his or her present state. How can the system do that if is centered where the user is now. The term *user-centered* might also be read to imply that system's conceptual structures should mirror the users' conceptual structures. But that may not serve the users best. Based on our knowledge of conceptual structure principles and on an analysis of users' problems and users' thinking, we can come up with structures that take the user a step further

Truly user-oriented service requires substantial knowledge, knowledge of problem-solving strategies, information organization, information systems, and search strategies. Skilled human intermediaries have that knowledge. Their most important contribution is helping users clarify their problem and their information need — more than half the battle in searching. Mediated access is useful and will persist. But with the revolutionary expansion of access through the Internet, the World Wide Web, and digital libraries, an increasing number of users do their own searching. So now <u>systems</u> must possess the knowledge to take over the intermediary function and, through their interface and behind-the-scenes intelligence, enable the user to easily find the information needed. What we have presently on the Web is a far cry from that.

A truly user-oriented system must empower the user to take full advantage of information resources and search capabilities. To this end, the system must present its information structure — its conceptual schema and its hierarchy of concepts — in a way the user can understand. It must enable the user to do tightly focussed searches often stated best in a Boolean query. We should not use the alleged inability of users to understand Boolean queries as an excuse for our own inability to design interfaces that make Boolean queries as clear, obvious, and immediately intuitive as they should be.

Systems must assist users in making sense. Sense and meaning come from structure, structure that assists both in framing a problem and posing a query and in interpreting the information found. We must provide well-structured conceptual frameworks and classifications and represent them well for ease of assimilation and incorporation into the user's own mental structure.

Beyond the Internet and the Web, there is a second, less glamorous but ultimately more important, revolution: The creation of tightly integrated systems that support all facets of individual and collaborative work. In the business world the two sides of this revolution are known as business process reengineering and knowledge management. Of course, this revolution will make full use of the Internet and intranets

Information is central to people's work. We must assist users with harnessing information to do their work better and increase productivity. Therefore, the provision of information should be integrated seamlessly into the users' work. Users' work, problem solving, and information seeking form one integrated process. This integrated process must be supported by an equally integrated system for all of a user's tasks with optimal division of labor between user and system. Such a system is proactive; it gets the information needed before the user asks. Consider processing insurance claims, a laborious and information-intensive process. A new-generation system knows types of claims, their processing steps and decision rules. It sends queries for the information needed to the appropriate in-house and external databases, assembles the information in a neat package, and suggests a resolution. The claim adjuster follows the process and intervenes as special circumstances require. When a search must be initiated by the user, the user should be able to do so from within any application she is working on. Consider a chemist who just ran a program to generate a tentative structure from computer-captured mass spectrography data. By a simple click on the structure diagram she starts a Chemical Abstracts search for substances with similar structure. An integrated system should include common tools for multiple functions, such as a thesaurus and dictionary for search support, for support of writing, and for spell checking. The system should provide tools for processing information after retrieval, such as efficient searching and perusal of texts, statistical analysis, and modelling and simulation.

The system must relieve the user of as much work as possible. It should deliver pre-digested information as a means to cope with the sheer mass. A search should return not just pointers to data sources, such as documents, but the actual substantive data themselves, either found directly or extracted and compiled from texts. Even further, the system should provide not just data but conclusions from data, problem solutions derived from data. The system should prepare reports based on predefined templates rather than just delivering bits and pieces of information for the report. This requires increasingly sophisticated methods for data extraction, compilation, and presentation and for deriving problem-oriented answers from vast amounts of data through statistical analysis, machine learning, and automated reasoning.

All of these system functions require use of knowledge: Knowledge-based retrieval; knowledge-based data analysis; and knowledge-based document design for optimal display of information.

But how to get the huge reservoirs of knowledge needed, how to break the knowledge acquisition bottleneck? There are four strategies for knowledge acquisition that make large-scale use of knowledge-based systems possible:

1   Make better use of what is already there by providing unified access to multiple heterogeneous knowledge bases.

2   In the production of handbooks, encyclopedias, dictionaries and the like, that require lots of intellectual work anyway, assemble the content in a structured format. It would then be available for preparing not one but multiple printed formats, for customized retrieval and presentation online, and for machine knowledge processing.

3   Use knowledge-based extraction of data from text.

4   Use machine learning, especially systems that learn incrementally, such as a parser that learns lexical and ontological information on words as it analyzes text, or a retrieval system that learns semantic relations from user queries.

All of these strategies do in turn require knowledge, so this is a bootstrap process. The more knowledge is available in structured form, the more knowledge can be acquired automatically.

So far I have talked about what systems should do for the user. But user-system interaction is not a one-way street. Users must know how to exploit system possibilities. Perhaps more importantly, people play an essential role in communicating information to other people. So people need information competency: competency in finding, evaluating, analyzing, and synthesizing information and competency in expressing information in writing or images. We must insert ourselves into the educational process to make sure students at all levels acquire this information competency.

Students graduating in the 21st century must be able to move with ease in the world of information. Information competency, developed in symbiotic interaction with thinking skills, is intrinsically and critically linked with all of a student's intellectual endeavors. For example, learning about frame structures and semantic networks helps students not only to think and to structure problems, but also to structure compositions. Learning and information seeking become one integrated process, particularly if students have access to learner-directed inquiry systems. Students can then carry this integration forward to lifelong learning and to their work. Today undergraduate education is woefully incomplete unless it provides both the conceptual foundation and the practical skills that make up information competency. What is needed, then, is an expansion of the customary one-semester "Introduction to writing" into a two-semester integrated course "Thinking and communicating in a world of information."

Always remember that the focus must be on information, not on technology. Technology is wonderful, information is essential; computer literacy is important, information competency is essential. Technology is a means, information and problem solving are the end. Making the most of technology requires information concepts, requires conceptual structures. It is the responsibility of our field and the society that represents it to keep that simple principle in the forefront as decisions on education, funding, and information policy are made.

Thank you for your attention.