Saracevic,.T. (2009). Information science. In: Marcia J. Bates and Mary Niles Maack (Eds.) *Encyclopedia of Library and Information Science.* New York: Taylor & Francis. pp. 2570-2586.

Final galleys are below:

Information Science

Tefko Saracevic
School of Communication and Information, Rutgers University, New Brunswick New Jersey, U.S.A.

Abstract

The purpose of this entry is to provide an overview of information science as a field or discipline, including a historical perspective to illustrate the events and forces that shaped it. Information science is a field of professional practice and scientific inquiry dealing with effective communication of information and information objects, particularly knowledge records, among humans in the context of social, organizational, and individual need for and use of information. Information science emerged in the aftermath of the Second World War, as did a number of other fields, addressing the problem of information explosion and using technology as a solution. Presently, information science deals with the same problems in the Web and digital environments. This entry covers problems addressed by information science, the intellectual structure of the field, and the description of main areas—information retrieval, human information behavior, metric studies, and digital libraries. This entry also includes an account of education related to information science and conclusions about major characteristics.

INTRODUCTION

The purpose of this entry is to provide an overview of information science as a field or discipline, including a historical perspective to illustrate the events and forces that shaped it.

Information science is the science and practice dealing with the effective collection, storage, retrieval, and use of information. It is concerned with recordable information and knowledge, and the technologies and related services that facilitate their management and use. More specifically, information science is a field of professional practice and scientific inquiry addressing the effective communication of information and information objects, particularly knowledge records, among humans in the context of social, organizational, and individual need for and use of information.[1] The domain of information science is the transmission of the universe of human knowledge in recorded form, centering on manipulation (representation, organization, and retrieval) of information, rather than knowing information.[2]

There are two key orientations: toward the human and social need for and use of information pertaining to knowledge records, on the one hand, and toward specific information techniques, systems, and technologies (covered under the name of *information retrieval*) to satisfy that need and provide for effective organization and retrieval of information, on the other hand. From the outset, information science had these two orientations: one that deals with information need, or more broadly human information behavior, and the other that deals with information retrieval techniques and systems.

Information science is a field that emerged in the aftermath of the Second World War, along with a number of new fields, with computer science being but one example. While developments and activities associated with information science already started by the end of 1940s, the very term "information science" came into full use only at the start of the 1960s. A significant impetus for the coalescence of the field was the *International Conference on Scientific Information*, held in Washington, D.C., November 16–21, 1958, sponsored by the (U.S.) National Science Foundation, the National Academy of Sciences—National Research Council, and the American Documentation Institute, and attended by some 1000 delegates from 25 countries. The conference was meticulously planned for over 3 years and attracted wide international attention. The 75 papers and lively discussions that followed, all recorded in the *Proceedings* of over 1600 pages, affected the direction of research, development, and professional practice in the field for at least a decade if not longer.[3] It also affected the internationalization of the field and the approaches used. They became global.

This entry covers problems addressed by information science, the intellectual structure of the field, and the further description of main areas—information retrieval, human information behavior studies, metric studies, and digital libraries. At the end, the entry includes an account of education related to information science and conclusions about major trends.

PROBLEMS ADDRESSED

To understand information science, as with any other field, a description of problems addressed and methods used in their solution is crucial. Generally, information

Encyclopedia of Library and Information Sciences, Third Edition DOI: 10.1081/E-ELIS3-120043704

science addressed the problem of information explosion and used information technology as a solution.

The rapid pace of scientific and technical advances that were accumulating since the start of the twentieth century produced by mid-century a scientific and technical revolution. A most visible manifestation of this revolution was the phenomenon of "information explosion," referring to the unabated, exponential growth of scientific and technical publications and information records of all kinds. The term "information explosion" is a metaphor (as is "population explosion") because nothing really exploded but just grew at a high rate, even exponentially at times. Simply put, information explosion is information and information objects piling up at a high rate. The problem this presents is getting to right information as needed at a given time.

A number of scientists documented this growth, but none better and more vividly than Derek de Solla Price (1922–1983, British and American physicist, historian of science and information scientist), recognized as the father of scientometrics. In his seminal works, S*cience since Babylon* followed by *Little Science, Big Science*, Price documented the exponential and logistical growth of scientific publications linking them with the growth of the number of scientists. The logistical growth started slow right after the appearance of the first scientific journals in the seventeenth century, accelerated by the start of the twentieth century and became explosive after the Second World War.[4,5]

The impetus for the development of information science, and even for its very origin and agenda, can be traced to a 1945 article, "As We May Think" by Vannevar Bush (1890–1974), a respected MIT scientist and, even more importantly, the head of the U.S. scientific effort during World War II.[6] In this influential article, Bush did two things: a) he succinctly defined the critical and strategic problem of information explosion in science and technology that was on the minds of many, and b) proposed a solution that was a "technological fix," and thus in tune with the spirit of the time. Both had wide appeal. Bush was neither the first nor the only one that addressed these issues, but he was listened to because of his stature. He defined the problem in almost poetic terms as "The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships." In other words, Bush addressed the problem of information explosion and the associated methods for finding relevant information.

As a solution, Bush proposed a machine, dubbed Memex, incorporating (in his words) a capability for "association of ideas," and the duplication of "mental processes artificially." A prescient anticipation of information science and artificial intelligence is evident. Memex, needless to say, was never built, but to this day is considered an ideal, a wish list, an agenda, and, some think, a utopia. Information science is still challenged by the ever-worsening problem of information explosion, now universal and in a variety of digital formats, and the field is still trying to fix things technologically.

A number of scientists and professionals in many fields around the globe listened and took up Bush's challenge. Most importantly, governments listened, as well, and provided funding. The reasoning went something like this: Because science and technology are strategically important for society, efforts that help them, information activities in particular, are also important and need support. In the United States, the United Kingdom, and other countries, this led to the support of research and development related to information problems and solutions. By the end of the 1940s information science was well on its way.

Bush also participated in the establishment of the National Science Foundation (NSF) in the United States. The National Science Foundation Act of 1950 (P.L. 81-507) provided a number of mandates, among them "to foster the interchange of scientific information among scientists in the U.S. and foreign countries" [Section 3(a)3] and "to further the full dissemination of [scientific and technical] information of scientific value consistent with the national interest" [Section 11(g)]. The 1958 National Defense Education Act (P.L 85-864) (the "Sputnik act") enlarged the mandate: "The National Science Foundation shall [among others].undertake programs to develop new or improved methods, including mechanized systems, for making scientific information available" (Title IX, Section 901). By those mandates, an NSF division, which after a number of name and direction changes is now called the Division of Information and Intelligent Systems (IIS), has supported research in these areas since the 1950s. Information science evolution, at least in the United States., was greatly affected by the support of the U.S. government. In this respect it was not an exception. For instance, artificial intelligence, among others, was for decades supported by the U.S. government starting in the 1950s and ending by the 1990s.

Historically, one force affecting government support of information science, as of many other fields in the United States and a number of European countries, had to do with the cold war. Among others, one impetus was the establishment in 1952 of the All-Union Scientific and Technical Information Institute of the Academy of Sciences of the USSR (Russian acronym: VINITI). VINITI implemented a massive gathering and bibliographic control of scientific and technical information from around the world, eventually covering some 130 countries in 66 languages; it employed thousands of scientists and engineers full- and part-time. In the framework of the Cold War, VINITI was repeatedly brought up as a challenge needing a response.

At the start, information science was directed solely toward information explosion in science and technology.

However soon it expanded to other areas, including business, humanities, law, and eventually any area of human endeavor. In all areas, the phenomenon of information explosion is continuing and even accelerating to this day, particularly in the digital and Web environments. Addressing the problems of dealing with information explosion in any human area where information and knowledge records are overbearing is at the heart of information science. The approach to these problems involves a number of disciplines; in other words, information science, as many other modern fields, is interdisciplinary in nature.

In its goals and activities, information science established early, and maintains prominently, a social and human function—not only a technological one. On the social level, it participates actively, with many other fields, in the evolution of information society around the globe. Yet information science also has an individual human function. It relates to searching for and use of information as done by (or on behalf of) individuals. People individually search for and use relevant information. For information science, managing information is a global, social function, while providing and using information is an intense individual function.

INTELLECTUAL STRUCTURE

Information science, like any other field, has a dynamic intellectual structure; the objects of study and practice appear, change, disappear or are emphasized, realized, and interwoven in different ways over time. A general framework for the intellectual structure for the field can be derived from the Three Big Questions for information science as identified by Bates[2]:

1. The physical question: What are the features and laws of the recorded information universe?
2. The social question: How do people relate to, seek, and use information?
3. The design question: How can access to recorded information be made most rapid and effective?

Indeed, when looking at the literature of information science since its emergence to this day, the general structure can be discerned from these questions in both research and practice reported. While they can be approached individually, the three questions are not independent but interdependent. Effective design is highly dependent on the consideration of social and physical features. Over time, details in the answers differed greatly. But, as is seen from three examples below, the general structure stands.

Three examples illustrating the intellectual structure of information science spanning some five decades are presented here. The first one is the enumeration of topics in the proceedings of the mentioned 1959 *International Conference on Scientific Information*.[3] The second one is an author cocitation analysis mapping information science for the years 1972–1995.[7] And the third one is a similar analysis, using the same methods, mapping information science for the years 1996–2005.[8] Author cocitation analysis is a statistical and visualization method developed in information science that allows for mapping of connections between authors in a given domain and identifying clusters or oeuvres of work in that domain. The raw data reflects the number of times selected author pairs are cited together in papers, regardless of which of their work is cited.

The 1959 Proceedings had seven areas covering the research, practice, and interests of information science at the time and illustrating the intellectual structure of the field by the end of 1950s. These were

1. *Literature and reference needs of scientists*. An example of a title of a paper in the area: An Operations Research Study of the Dissemination of Scientific Information.
2. *The function and effectiveness of abstracting and indexing services*. A paper example: All-Union Institute for Scientific and Technical Information (VINITI).
3. *Effectiveness of monographs, compendia, and specialized centers. Present trends and new and proposed techniques and types of services*. A paper example: Scientific, Technical, and Economic Information in a Research Organization.
4. *Organization of information for storage and search. Comparative characteristics of existing systems*. A paper example: The Evaluation of Systems Used in Information Retrieval.
5. *Organization of information for storage and retrospective search. Intellectual problems and equipment considerations in the design of new systems*. A paper example: Linguistic Transformations for Information Retrieval.
6. *Organization of information for storage and retrospective search. Possibility for a general theory*. A paper example: The Structure of Information Retrieval Systems.
7. *Responsibilities of government, professional societies, universities, and industry for improved information services and research*. A paper example: Differences in International Arrangements for Financial Support of Information Services.

Results from the next two studies are comparable—they used the same set of basic data (major journals in information science) and the same method (author cocitation analysis and mapping).[7,8] The authors of both studies mapped clusters of authors, classifying their areas of publications in a number of categories—they labeled the categories—and showing the relation or lack thereof between categories. The categories reflecting clusters of work in the two studies, as labeled by the authors, are shown in Table 1.

Comp. by: WOMAT Stage : Revises1 ChapterID: 0000962777 Date:30/6/09 Time:09:33:26 Filepath://
spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000962777.3D

Table 1 Intellectual structure of information science as presented in studies of two time periods (labels provided by authors of respective studies)

1972–1995	1996–2006
1. Experimental retrieval (design and evaluation of IR systems)	1. User studies (information seeking/searching behavior, user-centered approach to IR, users and use)
2. Citation analysis (interconnectedness of scientific and scholarly literatures)	2. Citation analysis (scientometrics; evaluative bibliometrics)
3. Practical retrieval (applications in "real world")	3. Experimental retrieval (algorithms, models, systems, evaluation of IR)
4. Bibliometrics (statistical distributions of texts and mathematical modeling)	4. Webometrics
5. General library systems (library automation, library operations research, services)	5. Visualization of knowledge domains (author cocitation analysis)
6. Science communication (including social sciences)	6. Science communication
7. User theory (information needs and users)	7. Users' judgment of relevance (situational relevance)
8. Online Public Access Catalogs (OPACs) (design, subject searching)	8. Information seeking and context
9. Imported ideas (information theory, cognitive science, etc.)	9. Children's information searching behavior (usability, interface design)
10. Indexing theory	10. Metadata and digital resources
11. Citation theory	11. Bibliometric models and distributions
12. Communication theory	12. Structured abstracts (academic writing)

Some of the areas in the three examples remain the same over time, showing an overall stability of general interests and foci of information science from its emergence to this day. The three areas of major and continuing interest are *information retrieval, user and use studies, and metric studies*. They correspond to the Three Big Questions for information science listed at the start of this section. Naturally, the variety and type of work in these three areas has changed and evolved over time, as elaborated below, but the general thrust and emphasis stayed stable.

Some areas have disappeared. The interest in the functioning of abstracting and indexing services, specialized information centers, and the responsibilities of different agencies for improved information services, so prominent in the 1959 Proceedings, are not prominent at all in later periods. Online Public Access Catalogs (OPACs) were prominent as an area cluster in the period 1971–1995 but did not appear in the 1996–2005 period; research in this area waned. The same holds for general library systems, covering library automation; the area was prominent during 1971–1995, but not anymore. The field had a prominent area of imported ideas between 1971–1995, covering the deliberations of adaptation and the application of various theories from information theory (Shannon), sociology (Merton), and other fields, but not anymore. Theory importing is not a major area any longer in information science. However, there is a significant exception. A major trend is evident in the incorporation of ideas, theories, and methods from the social sciences into many studies related to human information behavior to such an extent that they are not considered as imported any more.

In the Web age, covering the period 1996–2005, new areas have appeared. Not surprisingly, one of them is webometrics, extending the metric studies to the Web. Another new area is the visualization of knowledge domains, providing a new method of presenting retrieval processes and results and also extending citation and metric analyses.

The intellectual structure of information science also covers two camps of authors concentrating in different areas. White and McCain called them "retrieval people" and "literature people."[7] The first group congregates in the area of information retrieval; the second in the area of human information behavior and metric studies. They represent two broad branches of information science, one system-oriented and the other user-oriented. They are relatively isolated from each other. In the words of White and McCain again: "As it turns out, information science looks rather like Australia: Heavily coastal in its development, with a sparsely settled interior." The relative isolation is conceived as unproductive for all areas. There were a number of calls for collaboration, some quite impatient, and a few efforts at actually bridging the gap, but the gap has yet to be effectively bridged.

INFORMATION RETRIEVAL

Considering the Three Big Questions for information science, stated above, this section addresses the design question: *How can access to recorded information be made most rapid and effective?* The area is concentrated on systems and technology.

Right after the Second World War a variety of projects started applying a variety of technologies to the problem of controlling information explosion, particularly in science and technology. In the beginning the technologies were punched cards and microfilm, but soon after computers became available the technology shifted to, and stayed with, computers. Originally, many activities involved specific fields of application, such as chemistry. By the mid-1960s computer science joined the efforts in a big way.

Various names were applied to these efforts, such as "*machine literature searching*," or "*mechanical organization of knowledge*" but by the mid-1950s "*information retrieval*" prevailed. Actually, the term "*information retrieval*" (IR) was coined by mathematician and physicist Calvin N. Mooers (1919–1994), a computing and IR pioneer, just as the activity started to expand from its beginnings after Second World War. He posited that

> Information retrieval is . . . the finding or discovery process with respect to stored information . . . useful to [a user]. Information retrieval embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, technique, or machines that are employed to carry out the operation.[9]

Over the next half century, information retrieval evolved and expanded widely. In the beginning IR was static. Now it is highly interactive. Earlier it dealt only with representations—indexes, abstracts—now it deals with full texts as well. It concentrated on print only, now it covers every medium, . . ., and so on. Advances in the field are impressive, now covering the Web, and still go on. Contemporary search engines are about information retrieval. But in a basic sense, IR still continues to concentrate on the same fundamental things Mooers described. Searching was and still is about retrieval of relevant (useful) information or information objects.

It is of interest to note what made IR different, compared to many other techniques, applied to the control of information records over a long period of time. The key difference between IR and related methods and systems that long preceded it—such as classifications, subject headings, various indexing methods, or bibliographic descriptions, including the contemporary Functional Requirements for Bibliographic Records—is that IR specifically included "specification for search." The others did not. Since the days of the pioneers in bibliographic organization in the United States, Charles Ammi Cutter (1837–1903) and Melvil Dewey (1851–1931), the emphasis was on the creation of systems for bibliographic representation and control. In these long-standing techniques, what users' needs are and should be fulfilled was specified in detail. Following that, the representation of information objects was also prescribed in detail. In other words, data about information objects (books, articles, etc.) in bibliographic records are organized in a way to fulfill the specified needs. However, how the search ought to be done was not specified or addressed at all. Searching was assumed and left to itself—it just happens. In IR, users' needs are assumed as well, but the search process is specified in algorithmic detail and data is organized to enable the search. Search engines are about searching to start with; everything else is subsumed to that function.

Relevance

The fundamental notion used in bibliographic description and in all types of classification or categorization, including those used in contemporary databases, is *aboutness.* Cataloging and classification describe what the documents were all about with an implied idea that *about*, among others, may facilitate searching. Machine Readable Cataloging (MARC) that was developed by Henriette Avram (1919–2006) at the Library of Congress beginning in the 1960s follows the same principle. Online Public Access Catalogs (OPACs) that emerged more than a decade later are based on MARC. They include various search mechanisms, but are relatively difficult and ineffective as search tools. While implying searching *aboutness* does not really facilitate it.

The fundamental notion used in IR is *relevance*. Retrieval is not about any kind of information, and there are a great many, but about *relevant* information (or as Mooers called it *useful to a user* or Bush *momentarily important*). Basically, relevant information is that which pertains to the matter or problem at hand. Fundamentally, bibliographic description and classification concentrate on describing and categorizing information objects. IR is also about that, but in addition IR is about searching, and searching is about relevance. Very often, the differences between databases and IR are discussed in terms of differences between structured and unstructured data, which is OK, but the fundamental difference is in the basic notion used: *aboutness* in the former and *relevance* in the latter. The two notions are not at all equivalent. Relevance entered as a basic notion through the specific concentration on searching. True, searching for relevant information precedes IR by centuries, if not millennia, but in IR the process was adapted as primary orientation.

By choosing relevance as a basic, underlying notion of IR, related information systems, services, and activities—and with it, the whole field of information science—went in a direction that differed from approaches taken in librarianship, documentation, and related information services, and even in expert systems and contemporary databases in computer science.

In this sense, information science is on the one hand connected to relevance and on the other hand to technologies and techniques that enhance probability of the retrieval of relevant and the suppression of nonrelevant information. Relevance, as a basic notion in information science, is a human notion, widely understood in similar ways from one end of the globe to the other. This affected the widespread acceptance of information retrieval techniques globally. However, relevance, and with it information retrieval, involves a number of complexities: linguistic,

Comp. by: WOMAT Stage : Revises1 ChapterID: 0000962777 Date:30/6/09 Time:09:33:26 Filepath://
spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000962777.3D

cognitive, psychological, social, and technological, requiring different solutions. But the basic idea that searching is for relevant information does not.

As mentioned, relevance is a human notion. In human applications, relevance judgments exhibit inconsistencies, situational and dynamic changes, differences in cognitive interpretations and criteria, and other untidy properties common to human notions. This stimulated theoretical and experimental investigations about the notion and applications of relevance in information science. The experiments, mostly connected to relevance judgments and clues (what affected the judgments, what are people using in judgments) started already in the 1960s and continue to this day. The idea was and still is that findings may affect development of more effective retrieval algorithms. This is still more of a goal; actual translations from research results to development and practical applications were meager, if attempted at all.

Algorithms

IR systems and techniques, no matter in what form and including contemporary search engines, are geared toward retrieval of relevant information. To achieve that they use algorithms—logical step-by-step procedures—for the organization, searching, and retrieval of information and information objects. Contemporary algorithms are complex and in a never-ending process of improvement, but they started simple and still incorporate those simple roots.

The first and simple algorithm (although at the time it was not called that) applied in the 1940s and early 1950s was aimed at searching and retrieving from edge-notched punch cards using the operation of Boolean algebra. In the early 1950s Mortimer Taube (1910–1965), another IR pioneer and entrepreneur, founded a company named Documentation Inc. devoted to the development and operation of systems for the organization and retrieval of scientific and technical information. Taube broke away from the then-standard methods of subject headings and classification, by developing Uniterms and coordinate indexing. Uniterms were keywords extracted from documents; a card for a given Uniterm listed the documents that were indexed by that Uniterm. Coordinate indexing was actually a search and retrieval method for comparing (coordinating) document numbers appearing on different Uniterm cards by using a logical AND, OR, or, NOT operation. Although at the time the algorithm was not recognized as Boolean algebra by name, the operation was in effect the first application of a Boolean algorithm for information retrieval. Uniterms and coordinate indexing were controversial for a time but soon it was recognized that the technique was a natural for use as a base for computerized search and retrieval. All IR systems built in the next few decades incorporated Boolean algebra as a search algorithm and most have it under the hood today, along with other algorithms. All search engines offer, among others, Boolean search capabilities.

At the start of IR, and for a long time to come, the input—indexes and abstracts in particular—was constructed manually. Professionals indexed, abstracted, classified, and assigned other identifiers to information objects in a variety of fields. Input was manual; output—searching—was automated. Big online systems and databases, such as Medline and Dialog, which came about respectively in 1971 and 1972 and operate to this day, were based on that paradigm. Efforts to automate input, as well, commenced in the 1950s by the development of various algorithms for handling of texts. They took much longer to be developed and adopted more operationally than searching algorithms—the problem was and still is much tougher.

Hans Peter Luhn (1896–1964) a prodigious inventor with a broad range of patents joined IBM in 1941 and became a pioneer in development of computerized methods for handling texts and other IR methods in the 1950s. Luhn pioneered many of the basic techniques now common to IR in general. Among others, he invented the automatic production of indexes from titles and texts—Key Words in Context or KWIC indexing that lead to automatic indexing from full texts; automatic abstracting that lead to summarization efforts; and Selective Dissemination of Information (SDI) to provide current awareness services that led to a number of variations, including today's RSS (Really Simple Syndication). The demonstration of automatic KWIC indexing was the sensation at the aforementioned 1959 International Conference on Scientific Information.

Luhn's basic idea to use various properties of texts, including statistical ones, was critical in opening the handling of input by computers for IR. Automatic input joined the already automated output. Of course, Luhn was not the only one who addressed the problems of deriving representations from full texts. In the same period of the 1950s for instance, Phyllis Baxendale developed methods of linguistic analysis for automatic phrase detection and syntactic manipulations and Eugene Garfield was among the first, if not even the first, to join automated input and output in an operational system, that of citation indexing and searching.

Further advances that eventually defined modern IR came about in the 1960s. Statistical properties of texts—frequency and distribution of words in individual documents and in a corpus or collection of documents—were expressed in terms of probabilities that allowed for a variety of algorithms not only to extract index terms, but also to indicate term relations, distances, and clusters. The relations are inferred by probability or degree of certainty. They are inductive not deductive. The assumption, traced to Luhn, was that frequency data can be used to extract significant words to represent the content of a document and the relation among words. The goal was to find a match between queries and potentially relevant documents, based on a probability of documents being relevant. Once expressed in terms of probabilities, documents can be ranked from those that have a higher probability to those that have a lower probability of a match. There are many methods for doing this. The basic plan was to search for underlying mathematical structures to guide computation. These were powerful ideas that led to an ever-expanding array of new and improved algorithms for indexing and other information organization methods, along with the associated search and retrieval. Moreover, they lend themselves to experimentation.

A towering figure in advancing experimentation with algorithms for IR was Gerard (Gerry) Salton (1927–1995), a computer scientist and academic (Harvard and Cornell Universities) who firmly connected IR with computer science. Within a framework of a laboratory he established, (entitled the SMART project) Salton and collaborators, mostly his students, ran IR experiments from the mid-1960s to the time of his death in 1995. Many new IR algorithms and approaches were developed and tested; they inspired practical IR developments and further IR research in many countries around the world. Many of his students became leaders in the IR community. Salton was very active nationally and internationally in the promotion of IR; he is the founder of the Special Interest Group on Information Retrieval (SIGIR) of the Association of Computing Machinery (ACM). SIGIR became the preeminent international organization in IR with annual conferences that are the main event for reporting advances in IR research. As a result of global interest in IR, these conferences now alternate between continents. While Salton's research group started in the United States, today many similar groups operate in academic and commercial environments around the globe.

Contemporary IR has spread to many domains. Originally, IR concentrated on texts. This has expanded to any and all other media. Now there are research and pragmatic efforts devoted to IR in music, spoken words, video, still and moving images, and multimedia. While originally IR was monolingual, now many efforts are devoted to cross-lingual IR (CLIR). Other efforts include IR connected with Extensible Markup Language (XML), software reuse, restriction to novelty, adversarial conditions, social tagging, and a number of special applications.

With the appearance and rapid growth of the Web starting in the mid-1990s many new applications or adaptations of IR sprouted, as well. The most prominent are search engines. While a few large search engines dominate the scene globally, practically, there is no nation that does not have its own versions tailored to its own populace and interests. While practical IR was always connected with commercial concerns and information industry, the appearance, massive deployment and use of search engines pushed IR into a major role commercially, politically, and socially. It produced another effect, as well. Most, if not all, search engines use many well-known IR algorithms and techniques. But many search engines, particularly the major ones, in addition have developed and deployed their own IR algorithms and techniques, not known in detail and not shared with the IR community. They support aggressive efforts in IR research and development, mostly in-house. Contemporary IR also includes a proprietary branch, like many other industries.

Testing

Very soon after IR systems appeared, a number of claims and counterclaims were made about the superiority of various IR methods and systems, without supporting evidence. In response, the perennial questions asked of all systems were raised: *What is the effectiveness and performance of given IR approaches? How do they compare?* It is not surprising that these questions were raised in IR. At the time; most developers, funders, and users associated with IR were engineers, scientists, or worked in related areas where the question of testing was natural, even obligatory.

By the mid-1950s suggestions for two measures for evaluation of effectiveness of IR systems were made; they were precision and recall. Precision measures how many of *retrieved* items (let's say documents) were relevant or conversely how many were noise. Recall measures how many of the *potentially relevant items in a given file or system* were actually retrieved, or conversely how many were not retrieved even though they were relevant. The measures were widely adopted and used in most evaluation efforts since. Even today, the two measures, with some variation, are at the base for evaluation of the effectiveness of output using given retrieval algorithms and systems. It is significant to note that the two measures are based on the comparison of human (user or user surrogate) judgments of relevance with IR algorithms' or systems' retrieval of what it considered as relevant, where human judgment is the gold standard.

A pioneer in IR testing was Cyril Cleverdon (1914–1997), a librarian at the Cranfield Institute of Technology (now Cranfield University) in the United Kingdom. From the late-1950s until the mid-1970s Cleverdon conducted a series of IR tests under the name "Cranfield tests." Most famous were the tests sponsored by the (U.S.) National Science Foundation from 1961 to 1966 that established a model of IR systems (the so-called traditional model that concentrates on query on the one end and matched with static retrieval from an IR system or algorithm on the other end), and a methodology for testing that is still in use. One of the significant and surprising finding from Cranfield tests was that uncontrolled vocabularies based on natural language (such as keywords picked by a computer algorithm) achieve retrieval effectiveness comparable to vocabularies with elaborate controls (such as those using thesaurus, descriptors, or classification assigned by indexers). The findings, as expected, drew skepticism and strong critique, but were confirmed later by Salton and others. Not surprisingly these conclusions caused a huge controversy. But they also provided recognition of automatic indexing as an effective approach to IR.

Salton coupled development of IR algorithms and approaches with testing; he enlarged on Cranfield approaches and reaches. Everything that Salton and his group proposed and developed was mandatorily tested. The norm was established: No new algorithms or approaches were accepted without testing. In other words, testing became mandatory for any and all efforts that propose new algorithms and methods. It became synonymous with experimentation in IR.

After Salton, contemporary IR tests and experiments are conducted under the umbrella of the Text REtrieval Conference (TREC). TREC, started in 1992 and continuing to date, is a long-term effort at the (U.S.) National Institute for Standards and Technology (NIST), that brings various IR teams together annually to compare results from different IR approaches under laboratory conditions. Over the years, hundreds of teams from dozens of countries participated in TREC covering a large number of topics. TREC is dynamic: As areas of IR research change, so do the topics in TREC. Results are at the forefront of IR research.[10]

In many respects, IR is the main activity in information science. It has proved to be a dynamic and ever-growing area of research, development, and practice, with strong commercial interest and global use. Rigorous adherence to testing contributed to the maturing of information retrieval.

HUMAN INFORMATION BEHAVIOR

Considering the Three Big Questions for information science, stated above, this section addresses the social and individual question: *How do people relate to, seek and use information?* While often connected with systems, the emphasis in this area of information science is on people rather than systems.

Human information behavior refers to a wide range of processes which people employ when engaged with information and to related cognitive and social states and effects. In his book that comprehensively covers research on information behavior (with over 1100 documents cited, most since 1980), Case defines that information behavior:

> "encompasses information seeking as well as the totality of other *unintentional* or *passive* behaviors (such as glimpsing or encountering information), as well as purposive behaviors that do not involve seeking, such as actively *avoiding* information. [11, p.5]. (emphasis in the original).

As can be imagined, human information behavior, as with many other human behaviors, is complex, not fully understood, and of interest in a number of fields. A great many studies and a number of theories address various aspects related to human information behavior in psychology, cognitive science, brain sciences, communication, sociology, philosophy and related fields, at times using different terminology and classifications. Under various names, scholarly curiosity about human information behavior is longstanding, going back to antiquity.

Of particular interest in information science are processes, states, and effects that involve *information needs and use* and *information seeking and searching*. The order in which these two major areas of human information behavior studies are listed represents their historic emergence and emphasis over time.

Historically, the study of information needs and use preceded information science. Many relevant studies were done during the 1930s and 1940s in librarianship, communication, and specific fields, such as chemistry, concentrating on use of sources, media, systems, and channels. Already by the 1950s this area of study was well developed in information science—for instance, the aforementioned 1959 *Proceedings of the International Conference on Scientific Information*[3] had a whole area with a number of papers devoted to the topic. The *Annual Review of Information Science and Technology* had regular annual chapters on "information needs and use" starting with the first volume in 1966 and ongoing through 1978. Thereafter, chapters covering this area were broadened to cover in addition various aspects or contexts of information behavior, including information seeking. This change illustrates how the emphasis in topics studied significantly changed over time. Studies in human information behavior are evolving and slowly maturing.

Information Needs and Use

Over the years "information needs and use" was used as a phrase. However, while related information need and information use are distinct concepts. *Information need* refers to a cognitive or even a social state and *information use* to a process.

For decades, *information need* was used as a primitive concept on two levels: on an individual level it signified a cognitive state which underlies questions posed to information systems and requests for information in general; on a social level it signified information required for functioning and keeping abreast of a whole group, such as chemists. On the first, or cognitive, level it was assumed that individuals ask questions and request information because of a recognition that the knowledge one has is inadequate for a given problem or situation; it is subjective as represented by individuals; it is in the head of a user. On the second, or social, level it was assumed that a social group with common characteristics, goals, or tasks shares common information requirements that may be satisfied by specific information sources; it is more objective as determined by a group of individuals on the basis of some consensus or by experts based on experience. In general, information need was considered as instrumental in reaching a desired informational goal.

The concept of *information need* was entrenched until the start of the 1980s. Slowly, critiques of the concept gained ground by pointing out that it is nebulous, as are most other "need" concepts in every field where they are used; that it is often substituted for "information demand," which is a very different process and not a state; that it is associated with behaviorism, which in itself fell out of favor; that it is a subjective experience in the mind of a person and therefore not accessible for observation; and that it ignores wider social aspects and realities. Moreover, underlying assumptions were challenged. By the end of the decade information need was largely abandoned as a subject of study or explanation of underlying information processes. Instead, studies of information seeking and other aspects of information behavior gained ground. However, information need is still represented in the traditional IR model (mentioned above) as the source of questions that are submitted to retrieval systems. It is not further elaborated in that framework, just listed as a primitive concept.

The concept of *information use* is more precise and it is operationally observable. Studies of information use were done for a long time and in many fields. For instance, use of libraries or use of literature in a given area was investigated long before information science emerged and before information use became one of the major topics of information science research. In information science, information use refers to a process in which information, information objects, or information channels are drawn on by information users for whatever informational purpose. The process is goal-directed. Questions are asked: *Who are the users of a given information system or resource? What information objects do they use? What information channels are used to gather information?* Or in other words: *Who uses what? How? For what purpose?*

The studies addressing these questions were, and still are, pragmatic, retrospective, and descriptive. Historically, as they emerged in the early 1950s, they were directed toward fields and users in science and technology. This is not surprising. As mentioned, information science emerged as a response to the problem of information explosion in science and technology thus the use studies were in those areas. Regarding topics, many early studies addressed users' distribution of time and resources over different kinds of documents: scientific journals, books, patents, abstracting and indexing services, and so on. As the realm of information science expanded to cover other areas and populations, use studies expanded their coverage as well. By the 1990s, studies emerged that also covered information use in many populations and activities, including the small worlds of everyday living.

The early motivation for user studies was pragmatic: to discover guidelines for the improvement of practice. This was of great concern to practitioners, and consequently most such studies were done by practitioners. By 1970 or so there was a move toward academic studies of information use motivated by a desire to understand the process better and provide models and theories. By 2008 there are still two worlds of user studies: one more pragmatic, but now with the goal of providing the basis for designing more effective and usable contemporary IR and Web systems, including search engines, and the other more academic, still with the goal of expanding understanding and providing more plausible theories and models. The two worlds do not interact well.

Information Seeking and Searching

Information seeking refers to a set of processes and strategies dynamically employed by people in their quest for and pursuit of information. Information seeking also refers to the progression of stages in those processes. In the majority of theories and investigations about information seeking, the processes are assumed to be goal directed. In his aforementioned book, Case defines information seeking as

> "a conscious effort to acquire information in response to a need or gap in your knowledge." [11, p.5]

Not surprisingly, information seeking is of interest in a number of fields from psychology, sociology, and political science to specific disciplines and professions, often under different names and classifications, such as information gathering or information foraging. The literature on the theme is large, spanning many decades. Historically, information-seeking concerns and studies in information science emerged by the late 1970s in academic rather than pragmatic environments. Only lately have they turned toward pragmatic concerns, as well. It was recognized that information use was the end process, preceded by quite different, elaborate, and most importantly, dynamic behavior and processes not well understood. The studies began in large part by trying to observe and explain what people do when they search and retrieve information from various retrieval systems, to expand fast to involving a number of different contexts, sources—formal and informal—and situations or tasks. The dynamic nature of information-seeking became the prime focus in observations, experiments, models, and theories. Questions are asked: *What do people actually do when they are in a quest for and pursuit of information? How are they going about and how are they changing paths as they go about? What are they going through on a personal level? What information channels are used to gather information? How?*

Information seeking, as is the case with most human information behavior, is highly dependent on context. While context may be everything, the very concept of context is ill defined, or taken as primitive and not defined. The contexts may involve various motivations for information seeking, various cognitive and affective states, various social, cultural, or organizational environments, various demographic characteristics, values, ways of life, and so on. A number of information-seeking studies were indeed directed toward various contexts. Thus, there is a wide range of such studies regarding context, accompanied by difficulties toward generalization.

To deal with more defined contexts, and enable specific observation, task-oriented information-seeking studies emerged in the 1990s. And they are going strong up to this day. Task studies deal with specific goals, mostly related to assignments in defined circumstances, time periods, or degree of difficulty. They represent a step in the ongoing evolution, not only of information-seeking studies in particular but also in information behavior research in general. By the 2000s we also see the emergence of studies in collaborating behaviors, also related to given tasks.

Information searching is a subset of information seeking, and in the context of information science, it refers to processes used for interrogating different information systems and channels in order to retrieve information. It is the most empirical and pragmatic part of information-seeking studies. Originally, search studies concentrated on observation and modeling of processes in the interrogation of IR systems. With the advent of digital environments, the focus shifted toward Web searching by Web users. New observational and experimental methods emerged, becoming a part of exploding Web research. Such search studies have a strong pragmatic orientation in that many are oriented toward improving search engines and interfaces, and enhancing human–computer interactions.

Models and Theories

The research area and accompanying literature of information behavior in information science is strong on models and theories. It follows a tradition and direction of such research in many other disciplines, particularly psychology, communication, and philosophy. Being primarily pragmatic and retrospective, information use studies were not a great source for models and theories. In contrast, broader studies of information behavior, and particularly of information seeking, are brimming with them. Numerous models and theories emerged, some with more, others with less staying power. The extent of this work is exemplified in a compilation *Theories of Information Behavior*,"[12] where some 70 different (or differing) theories and models are synthesized. To illustrate, we should sample three well-known theories, each in one of the three areas of human information behavior described above. Each of them is widely accepted and cited, and tested, as well.

What is behind an information need? Why do people seek information in the first place? Starting in late 1970s and for the next two decades or so, Nicholas Belkin and his colleagues addressed this question by considering that the basic motivation for seeking information is what they called "anomalous state of knowledge" (ASK), thus the "ASK theory," or as they called it, "ASK hypothesis" (described among others in Ref. [13]). Explicitly following a cognitive viewpoint, they suggest that the reason for initiating an information-seeking process could be best understood at the cognitive level, as a user (information seeker) recognizes that the state of his/her knowledge is in some way inadequate (anomalous) with respect to the ability to resolve a problematic situation and achieve some goal. Anomaly was used explicitly, not only to indicate inadequacy due to lack of knowledge, but also due to other problems, such as uncertainty of application to a given problem or situation. ASK theory is an attempt to provide an explicit cognitive explanation of information need or gap by proposing specific reasons why people engage in information seeking. It also suggests that anomalous states could be of different types. One of the strengths of ASK theory is that, unlike many other similar theories, it was successfully tested in a few experiments. One of the weaknesses is that it rests solely on a cognitive basis, using the problem or situation toward which the whole process is oriented as a primitive term.

What is behind the information search process? How is it constructed? Carol Collier Kuhlthau addressed these questions in a series of empirically grounded studies through a period of some 20 years starting in the early 1980's.[14] Her model and theory, called the Kuhlthau Information Search Model, provides a conceptual and relatively detailed framework of the information-seeking and search process. It is based on the personal construct theory in psychology that views learning as a process of testing constructs; consequently it views the search as a dynamic process of progressive construction. The model describes common patterns in the process of information seeking for complex tasks that have a discrete beginning and ending over time and that require construction and learning. The innovative part of the model is that it integrates thoughts, feelings, and actions in a set of stages from initiation to presentation of the search process. Not only cognitive, but also affective aspects, such as uncertainty connected with anxiety, are brought in the explanation of the process. The work started within learning context in schools, continued with a series of longitudinal studies, and moved on to a series of case studies in a number of fields. The strength of the model is that it incorporates affective factors that play a great role not only in searching but in human information behavior at large; furthermore it was extensively verified and revised over time. The weakness is that its educational roots are still recognizable—many search processes have different goals and contexts, thus the model may not fit.

What types of activities are involved in information seeking in general and information retrieval searching in particular? What is the relation between different activities? Starting in the mid-1980s and continuing for close to two decades, David Ellis and his colleagues addressed these questions in a series of empirical studies that led to the formulation and continuing refinement of a model known as Ellis's Model of Information-Seeking Behavior, primarily oriented toward behavior in information retrieval.[15] The model is based on a theoretical premise that the study of behavior presents a more tractable and observable focus for study than cognitive approaches. Consequently, its base is behavioral rather than cognitive. The model incorporates a premise that the complex process of information seeking, particularly as related to information retrieval, rests on a relatively small and finite number of different types of interacting activities, these include starting, chaining, browsing, differentiating, monitoring, and extracting. The explicit goal of studies associated with Ellis' model was pragmatic: to inform design and operations of IR systems. The strength of the model is in the reduction of a complex process to a relatively small set of distinct and dynamically interacting processes. The weakness is that it does not address cognitive and affective aspects, shown to be of importance.

The three models can be considered also as theories of information behavior. In turn, each of them is based on a different approach and theory. The first one is related to cognition as treated in cognitive science, the second to personal construct theory in psychology, and the third to behaviorism in psychology. This illustrates different approaches and multidisciplinary connections of human information behavior studies in information science. As yet, they have not found a common ground.

METRICS

Considering the Three Big Questions for information science, stated above, this section addresses the physical question: *What are the features and laws of the recorded information universe?* While often connected with systems, the emphasis in this area of information science is on information objects or artifacts rather than systems; these are the content of the systems. It is about characterizing content objects.

Metrics, such as econometrics, biometrics, sociometrics, etc., are important components in many fields; they deal with statistical properties, relations, and principles of a variety of entities in their domain. Metric studies in information science follow these by concentrating on statistical properties and the discovery of associated relations and principles of information objects, structures, and processes. The goals of metric studies in information science, as in other fields, are to characterize statistically entities under study and more ambitiously to discover regularities and relations in their distributions and dynamics in order to observe predictive regularities and formulate laws.

The metric studies in information science concentrate on a number of different entities. To denote a given entity under study over time, these studies were labeled by different names. The oldest and most widely used is *bibliometrics*—the quantitative study of the properties of literature, or more specifically of documents, and document-related processes. Bibliometric studies in information science emerged in the 1950s right after the start of the field. *Scientometrics*, which came about in the 1960s, refers to bibliometric and other metric studies specifically concentrating on science. *Informetrics,* emerging in the 1990s, refers to the quantitative study of properties of all kinds of information entities in addition to documents, subsuming bibliometrics. *Webometrics,* which came about at the end of the 1990s, concentrates, as the name implies, on Web-related entities. *e-Metrics*, which emerged around 2000, are measures of electronic resources, particularly in libraries.

Studies that preceded bibliometrics in information science emerged in the 1920s and 1930s; they were related to authors and literature in science and technology. A number of studies went beyond reporting statistical distributions, concentrating on relations between a quantity and the related yield of entities under study. Here are two significant studies that subsequently greatly affected development of bibliometrics. In the 1920s, Alfred Lotka (1880–1949, American mathematician, chemist, and statistician) reported on the distribution of productivity of authors in chemistry and physics in terms of articles published. He found a regular pattern where a large proportion of the total literature is actually produced by a small proportion of the total number of authors, falling down in a regular pattern, where the majority of authors produce but one paper—after generalization this became known as Lotka's law. In the 1930s, Samuel Bradford (1878–1948, British mathematician and librarian), using relatively complete subject bibliographies, studied the scatter of articles relevant to a subject among journals. He found that a small number of journals produce a large proportion of articles on the subject and that the distribution falls regularly to a point where a large number of journals produce but one article on the same subject—after generalization this became known as Bradford's law or Bradford's distribution. Similar quantity-yield patterns were found in a number of fields and are generally known as Pareto distributions (after Italian economist Vilfredo Pareto, 1848–1923). Lotka's and Bradford's distributions were confirmed many times over in subsequent bibliometric studies starting in the 1950s. They inspired further study and moreover set a general approach in bibliometric studies that was followed for decades.

Comp. by: WOMAT Stage : Revises1 ChapterID: 0000962777 Date:30/6/09 Time:09:33:27 Filepath://
spiina1003z/Womat/Production/PRODENV/0000000020/0000011876/0000000005/0000962777.3D

Data Sources

All metric studies start from and depend on data sources from which statistics can be extracted. Originally, Lotka used, among others, Chemical Abstracts, and Bradford used bibliographies in applied geophysics and in lubrication. These were printed sources and analysis was manual. For a great many years, the same kind of print sources and manual analysis methods were used.

The advent of digital technology vastly changed the range of sources, as well as significantly enlarged the type and method of analysis in bibliometrics, or as Thelwall put it, in a historical synthesis of the topic, "bibliometrics has changed out of all recognition since 1958."[16] This is primarily because sources of data for bibliometric analyses proliferated (and keep proliferating), inviting new analysis methods and uses of results.

In 1960 Eugene Garfield (U.S. chemist, information scientist, and entrepreneur) established the Institute for Scientific Information (ISI), which became a major innovative company in the creation of a number of information tools and in bibliometric research. In 1964, ISI started publishing the *Science Citation Index,* created by use of computers. Citation indexes in social sciences and in art and humanities followed. While citation indexes in various subjects, law in particular, existed long before Garfield applied them in science, the way they were produced and used was innovative. Besides being a commercial product, citation indexes became a major data source for bibliometric research. They revolutionized bibliometrics.

In addition to publication sources—journal articles and citations—de Solla Price pioneered the use of a range of statistics from science records, economics, social sciences, history, international reports, and other sources to derive generalizations about the growth of science and the factors that affected information explosion.[5] Use of diverse sources became a trademark of scientometrics.

As the Web became the fastest growing and spreading technology in history it also became a new source of data for ever-growing types of bibliometric-like analyses, organized under the common name of webometrics. The Web has a number of unique entities that can be statistically analyzed, such as links, which have dynamic distributions and behavior. Thus, webometrics started covering quite different grounds.

As more and more publications, particularly journals and more recently books, became digital they also became a rich source for bibliometric analyses. Libraries and other institutions are incorporating these digital resources in their collections, providing a way for various analyses of their use and other aspects. Most recently, digital libraries became a new source of analysis for they are producing massive evidence of the usage patterns of library contents, such as journal articles, for the first time. Thus, the emergence of e-metrics.

[From now on all the metric studies in information science (bibliometrics, scientometrics, informetrics, webometrics, and e-metrics) for brevity will be collectively referred to as *bibliometrics*.]

In the digital age, sources for bibliometric analyses are becoming more diversified, complex, and richer. They have become a challenge for developing new methods and refining existing methods and types of analysis.

Types and Application of Results

Lotka showed distribution of publication regarding authors and Bradford distribution of articles regarding journals. In seeking generalization, both formulated respective numerical distributions in a mathematical form. The generalizations sought a scientific law-like predictive power, with full realization that social science laws are not at all like natural science laws. In turn, mathematical expressions of Lotka's and Bradford's laws were refined, enlarged, and corrected in numerous subsequent mathematical papers; the process is still going on. This set the stage for the development of a branch of bibliometrics that is heavily mathematical and theoretical; it is still growing and continuously encompassing new entities and relations as data becomes available. Bradford also illustrated the results graphically. This set the stage for the development of visualization methods for showing distributions and relations; the efforts evolved to become quite sophisticated using the latest methods and tools for data visualization to show patterns and structures.

Over the years bibliometric studies showed many features of the ever-growing number of entities related to information. Some were already mentioned, here is a sample of others: frequency and distribution analysis of words; cowords; citations; cocitations; emails; links; etc., and quite a few others.

Until the appearance of citation indexes, bibliometric studies in information science were geared to analysis of relations; many present studies continue with the same purpose and are geared toward relational applications. But with the appearance of citation data, a second application emerged: evaluative.[16]

Relational applications seek to explicate relationships that are results of research. Examples are emergence of research fronts; institutional, national, and international authorship productivity and patterns; intellectual structure of research fields or domains; and the like.

Evaluative applications seek to assess or evaluate the impact of research, or more broadly, scholarly work in general. Examples are use of citations in promotion and tenure deliberations; ranking or comparison of scholarly productivity; relative contribution of individuals, groups, institutions, or nations; relative standing of journals; and the like.

Evaluative indicators were developed to numerically express the impact of given entities. Here are two of the most widely used indicators. The first deals with journals, the second with authors. *Journal Impact Factor,* devised in the 1960s by Garfield and his colleagues, provides a numerical value to how often a given journal is included in citations in all journals over a given period of time, normalized for the number of articles appearing in a journal. Originally, it was developed as a tool to help selection of journals in *Science Citation Index* but it morphed into a widely used tool for ranking and comparing the impact of journals. The second indicator deals with authors. A most influential new indicator of impact is the *h-index* (proposed in 2005 by Jorge Hirsh, a U.S. physicist). It quantifies and unifies both an author's scientific productivity (number of papers published by an author) and the apparent scientific impact of a scientist (number of citations received)—it unifies how much was published with how much was cited. Both of the indices are continuously discussed, mathematically elaborated, and criticized.

Evaluative studies are controversial at times. By and large, evaluative applications rest on citations. The central assumption here is that citation counts can be used as an indicator of value because the most influential works are most frequently cited. This assumption is questioned at times, thus it is at the heart of controversies and skepticism about evaluative approaches.

Evaluative applications are used at times in support of decisions related to tenure and promotion processes; academic performance evaluations of individuals and units in universities; periodic national research evaluations; grant applications; direction of research funding; support for journals; setting science policies; and other decisions involving science. Several countries have procedures in place that mandate bibliometric indicators for the evaluation of scientific activities, education, and institutions. They are also used in the search of factors influencing excellence.

The current and widening range of bibliometric studies are furthering understanding of a number of scholarly activities, structures, and communication processes. They are involved in the measuring and mapping of science. In addition, they have a serious impact on evaluation, policy formulation, and decision making in a number of areas outside of information science.

DIGITAL LIBRARIES

Long before digital libraries emerged in the mid-1990s, J. C. R. Licklider (1915–1990, U.S. computer scientist) in a prescient 1965 book *Libraries of the Future* envisioned many of the features of present digital libraries, with some still to come.[17] While Licklider was a technology enthusiast and formulated his vision of the library in a technological context, he also foresaw the handling of content in cognitive, semantic, and interactive ways.

Many of the components were in place quite some time before they were shaped and unified operationally into digital libraries. For instance, online searching of abstracting and indexing databases; a number of network information services; library automation systems; document structuring and manipulation procedures based on metadata; digitized documents; human computer interfaces; and others. With the advent of the Web, many of these older components were refined as needed and amalgamated with a number of new ones to form digital libraries as we know them today.

From the outset, people from a number of fields and backgrounds got involved in the development of digital libraries. Thus various conceptions were derived. Two viewpoints crystallized, one more technological the other more organizational. From the first point of view, a digital library is a managed collection of digital information with associated services, accessible over a network. From the second point of view, a digital library is that, but in addition it involves organizations that provide resources to select, structure, and offer intellectual access to collections of digital works for use by defined communities, and to preserve integrity and ensure persistence of collections and services. The first viewpoint comes mostly from computer science and the second from libraries and other organizations that house and provide digital library services. Digital libraries continue this dual orientation, technological and organizational, because, yes, they are indeed completely dependent on technology but by their purpose and functions they are social systems in the first place.

Many organizations other than libraries enthusiastically started developing and operating digital libraries—museums, historical societies, academic departments, governments, professional organizations, publishers, nonprofit organizations, and so on. As a result, digital libraries take many shapes and forms. They involve a variety of contexts, media, and contents. Many are oriented toward a specific subject. Most importantly, they are used by a variety of users and for a variety of uses. Digital libraries are a highly diverse lot.

The wide and constantly increasing diversity of digital libraries and related collections and portals suggest several issues: traditional libraries are not traditional any more, but hybrid and coming in many digital library forms; many new players have entered the arena, particularly in subject areas; and many new types of uses have emerged in addition to the traditional use of libraries. Digital libraries are truly interdisciplinary. Information science was one of the fields that actively participated in digital library formation, development, and research.

Through NSF and other agencies, the U.S. government funded research in digital libraries through Digital Library Initiatives; European Union and other governments funded similar research and development programs. Governmental funding started around 1995 and lasted about a decade. Most of the funding went toward technological aspects and demonstrations. An important by-product of this funding was the creation of a strong international community of digital library researchers from a number of fields, information science included. Here is another by-product often mentioned: Google was initially developed at Stanford University under an NSF grant in the Digital Library Initiatives program.

From the outset, information science was involved with digital libraries in a number of ways. Professionally, many information scientists work in digital libraries, particularly in relation to their architecture, systems operations, and services. A diverse number of topics were addressed in research covering the whole life-cycle of digital libraries as reflected in numerous reports, journals, proceedings, and books. Here is a sample: development and testing of digital library architecture; development of appropriate metadata; digitization of a variety of media; preservation of digital objects; searching of digital library contents; evaluation of digital libraries; access to digital libraries; security and privacy issues; study of digital libraries as a place and space; study of users, use, and interactions in digital libraries; effect of digital libraries on educational and other social institutions; impact of digital libraries on scholarship and other endeavors; and policy issues. New research topics are coming along at a brisk pace.

The rapid development and widespread deployment of digital libraries became a force that is determining not only the future of libraries but also of many other organizations as social, cultural, and community institutions. It is instrumental in the development of e-science. It is also affecting the direction of information science in that the domain of problems addressed has been significantly enlarged.

EDUCATION

The fact that education is critical for any field is a truism that hardly needs to be stated. Information science education began slowly in the 1950s and 1960s. Two educational models evolved over time and were followed for decades to come: For brevity, they should be referred to as the Shera and Salton models, after those that pioneered them. Both have strengths and weaknesses. A third model is presently emerging, under the label of i-Schools.

Jesse H. Shera (1903—1982, librarian and library educator) was a library school dean at Western Reserve University (later Case Western Reserve) from 1952 to 1970. Among others, he was instrumental in starting the Center for Documentation and Communication Research at the library school there in 1955. The Center was oriented toward research and development in IR. Shortly thereafter, the library school curriculum started to include courses such as “machine literature searching” (later to become “information retrieval”), and a few other more advanced courses and laboratories on the topics of research in the Center. The basic approach was to append those courses, mostly as electives, to the existing library school curriculum, without modifications of the curriculum as a whole, and particularly not the required core courses. Information science (or information retrieval) became one of the specialty areas of library science. The base or core courses that students were taking rested in the traditional library curriculum. Information science education was an appendage to library science. Library schools in the United States and in many other countries imitated Shera’s model. They used the same approach and started incorporating information science courses in their existing curriculum as a specialty.

The strength of the Shera model is that it posits education within a service framework, connects the education to professional practice and a broader and user-oriented frame of a number of other information services, and relates it to a great diversity of information resources. The weakness is a lack of a broader theoretical framework, and a lack of teaching of formalism related to systems, such as the development and understanding of algorithms. A majority of researchers in the human information behavior and user-centered approach are associated with this educational environment. Out of this was born the current and widely used designation *library and information science*.

Shera’s model, with contemporary modifications is still the prevalent approach in a majority of schools of library and information science. Some schools evolved to include a major in information science, or reoriented the curriculum toward some of the aspects of information science, or even provided a separate degree. The changes in curricula are accelerating. Dissatisfaction with the model as not in synch with contemporary developments related to information-spurred development of i-Schools discussed below.

Gerard Salton (already mentioned above) was first and foremost a scientist, and a computer scientist at that. As such, he pioneered the incorporation into IR research a whole array of formal and experimental methods from science, as modified for algorithmic and other approaches used so successfully in computer science. His primary orientation was research. For education, he took the time-honored approach of a close involvement with research. The Salton model was a laboratory and research approach to education related to IR. As Shera’s model resulted in information science education being an appendage to library science education, Salton’s model of IR education resulted in being a specialty of and an appendage to computer science education. Computer science students that were already well-grounded in the discipline got involved in SMART and other projects directed by Salton, worked and did research in the laboratory, completed their theses in areas related to IR, and participated in the legendary IR seminars. They also published widely with Salton and with each other and participated with high visibility in national and international conferences. From Harvard and Cornell, his students went to a number of computer science departments where they replicated Salton’s model. Many other computer science departments in the United States and abroad took the same approach. The strength of Salton's model is that it: i) starts from a base of a firm grounding in formal mathematical and other methods; and ii) relates directly to research. The weakness is in that it: i) ignores the broader aspects of information science, as well as any other disciplines and approaches dealing with the human aspects, that have great relevance to both outcomes of IR research and research itself; and ii) does not incorporate professional practice where these systems are realized and used. It loses users. Consequently, this is a successful, but narrowly concentrated education in IR as a specialty of computer science, rather than in information science. Not surprisingly, the researchers in the systems-centered approach came out of this tradition.

The two educational approaches are completely independent of each other. Neither reflects fully what is going on in the field. While in each model there is an increase in cognizance of the other, there is no educational integration of the systems- and user-centered approaches. The evident strengths that are provided by Shera's and Salton's model are not put together.

The late 1990s and early 2000s saw a movement to broaden and reorient information science education, spearheaded by a number of deans of schools with strong information science educations. Some library and information science schools were renamed into Information Schools or i-Schools. An informal i-School Caucus was formed in 2005. By 2008, the Caucus included over 20 schools quite diverse in origin. They include schools of: information; library and information science; information systems; informatics; public policy and management; information and computer sciences; and computing. The i-Schools are primarily interested in educational and research programs addressing the relationship between information, technology, and people and understanding the role of information in human endeavors. While the i-School movement was originally restricted to the United States, some schools outside the United States are joining. The movement is attracting wide international interest.

The i-Schools represent an innovative, new approach to information science education, with some true interdisciplinary connections. As the millennial decade draws toward an end, it is also signifying a new direction to information science education.

CONCLUSIONS

It was mentioned that information science has two orientations: one that deals with information retrieval techniques and systems and the other that deals with information needs and uses, or more broadly with human information behavior. One is technical and system-oriented, the other individual and social and user-oriented. In pursuing these orientations certain characteristics of the field emerged.

Information science has several general characteristics that are the leitmotif of its evolution and existence. These are shared with many modern fields.

- First, information science is interdisciplinary in nature. However, with various advances, relations with various disciplines are changing over time. The interdisciplinary evolution is far from over.
- Second, information science is inexorably connected to information technology. A technological imperative is compelling and encouraging the evolution of information science, as is the evolution of a number of other fields, and moreover, of the information society as a whole.
- Third, information science is, with many other fields, an active participant in the evolution of the information society. Information science has a strong social and human dimension, above and beyond technology.
- Fourth, while information science has a strong research component that drives advances in the field, it also has an equally strong, if not an even stronger, professional component oriented toward information services in a number of environments. Many innovations come from professionals in the field.
- Fifth, information science is also connected with information industry, a vital, highly diversified, and global branch of the economy.

With accelerating changes in all these characteristics, information science is a field in a constant flux. So are many other fields. The steady aspect is in its general orientation toward information, people, and technology.

REFERENCES

1. Saracevic, T. Information science. J. Am. Soc. Info. Sci. **1999**, *50* (12), 1051–1063.
2. Bates, M.J. The invisible substrata of information science. J. Am. Soc. Info. Sci. **1999**, *50* (12), 1043–1050.
3. National Science Foundation, National Academy of Sciences, American Documentation Institute, National Research Council. *Proceedings of the International Conference on Scientific Information*; The National Academies Press: Washington, DC, 1959; 2 volumes. http://books.nap.edu/openbook.php?isbn=NI000518&page=R19 (accessed April 15, 2008).
4. Price, D.J.; de, S. *Science Since Babylon*; Yale University Press: New Haven, CT, 1961.
5. Price, D.J.; de, S. *Little Science Big Science*; Columbia University Press: New York, 1963.
6. Bush, V. As we may think. Atlantic Mon. **1945**, *176* (11), 101–108. http://www.theatlantic.com/doc/194507/bush (accessed April, 14, 2008).
7. White, H.D.; McCain, K.W. Visualizing a discipline: An author cocitation analysis of information science. 1972–1995. J. Am. Soc. Info. Sci. **1998**, *49* (4), 327–355.

8. Zhao, D.; Strotmann, A. Information science during the first decade of the Web: An enriched cocitation analysis. J. Am. Soc. Info. Sci. Technol. **2008**, *59* (6), 916–937.
9. Mooers, C.N. Zatocoding applied to mechanical organization of knowledge. Am. Doc. **1951**, *2* (1), 20–32.
10. Voorhees, E.M.; Harman, D.K., Eds. *TREC. Experiment and Evaluation in Information Retrieval*; MIT Press: Cambridge, MA, 2005.
11. Case, D.O. *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior,* 2nd Ed.; Academic Press, Elsevier: New York, 2007.
12. Fisher, K.E.; Erdelez, S.; McKechnie, L.E.F. *Theories of Information Behavior*; American Society for Information Science and Technology: Washington DC, 2005.
13. Belkin, N.J.; Oddy, R.N.; Brroks, H.M. ASK for information retrieval. Parts 1 and 2. J. Doc. **1986**, *28* (2), 61–71, 145–164.
14. Kuhlthau, C.C. *Seeking Meaning: A Process Approach to Library and Information Services,* 2nd Ed.; Libraries Unlimited: Westport, CT. 2004.
15. Ellis, D. A behavioral model for information retrieval system design. J. Doc. **1989**, *45*, 171–212.
16. Thelwall, M. Bibliometrics to webometrics. J. Info. Sci. **2008**, *34* (4), 605–621.
17. Licklider, J.C.R. *Libraries of the Future*; The MIT Press: Cambridge, MA, 1965.