

## ENTREVISTA

### Tefko Saracevic

**Tefko Saracevic** é Professor Emeritus na Rutgers University (NJ) e desempenhou as funções de Reitor de 2003 a 2006. Ao longo dos anos, realizou pesquisas e publicou diversos trabalhos sobre: teste e avaliação de sistemas de recuperação de informação e bibliotecas digitais; noção de relevância na ciência da informação; aspectos humanos na interação humano-computador na recuperação da informação; estudos de usuários e uso em ciência da informação e biblioteconomia; análise de desempenho dos motores de busca na Web, e vários aspectos das bibliotecas digitais. Realizou seminários, palestras, cursos, consultorias e apresentou trabalhos em encontros internacionais em mais de 46 países, e foi conferencista convidado em dezenas de conferências internacionais. Lecionou no IBICT e Universidade de Brasília em 1970 e 1980 e ao longo dos anos fez apresentações em muitas outras instituições brasileiras e conferências no país. Foi o presidente da American Society for Information Science e recebeu o Society's Award of Merit (o maior prêmio dado pela Sociedade). Entre outros prêmios, recebeu também o Gerard Salton Award for Excellence in Research, dado pelo Special Interest Group on Information Retrieval da Association for Computing Machinery (também o maior prêmio dado pelo Grupo). Recebeu 1296 citações de seus trabalhos na SciSearch Social (Social Sciences Citation Index, 1972 -), e SciSearch (Science Citation Index, 1974 -), excluindo auto-citações. Por ocasião do seu 80º aniversário a School of Communication and Information, da Rutgers University organizou uma conferência intitulada TEFKO 2010 - <http://comminfo.rutgers.edu/events/tefko-2010.html> e <https://sites.google.com/site/tefko2010/home>

**InCID: Professor Saracevic, you were in the USA, back in the 60's, one of the pioneers in Information Science and you are worldwide recognized for your research on notion of relevance in information science. However, here in Brazil, you are better known through your paper published in Belo Horizonte (1996) about the interdisciplinary nature of Information Science, specially examining the origin, evolution and relationships of Information Science with other disciplines (Ciência da informação: origem, evolução e relações. Perspectivas em Ciência da Informação, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996). This allows us to ask some questions:**

**InCID: The areas of interface with the Information Science that you have listed in 1996: library**

**science, computer science, cognitive science and communication remain the most important ones? Has something new appeared over the past 20 years?**

**Tefko Saracevic:** These areas are still in strong interaction with information science. But let me elaborate emphasizing changes over 30 years or so<sup>1</sup>. There are two key orientations in information science: 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 - IR) 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. In both of these areas great many new things happened over the past 20 years. Here are a few.

- **Information retrieval:** Searching for information became ubiquitous – everybody is a searcher now. Among others, advances in IR made this possible. In the process, IR became highly commercial as well. Big contemporary search engines are a part of this trend. The field of computer science is very much involved. 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). However, the big question for information retrieval remains constant: *How can access to recorded information be made most rapid and effective?*
- **Human information behavior:** Studies in this area expanded to cover a wide range of processes which people employ when engaged with information and to related cognitive and social states and effects. But information science is not the only field interested in this area of research. 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. Of particular interest in information science are processes, states and effects that involve *information needs*

---

<sup>1</sup> For answers to this question, I will use some excerpts from my article *Information Science* in the *Encyclopedia of Library and Information Science*, 2010; could be seen on my site, under Selected articles <http://comminfo.rutgers.edu/~tefko/articles.htm>

*and use and information seeking and searching.* Of course, here is also a close connection with interests in librarianship, particularly in the area of information use. Contemporary studies on information seeking and searching address questions such as: *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?* The big question for human information behavior area is: *How do people relate to, seek and use information?*

- **Metrics (bibliometrics):** This is another area with massive changes. Bibliometrics expanded to various areas as scientometrics, webometrics, and most general infometrics. (I will call all these here under a common name bibliometrics). Bibliometrics has changed out of all recognition since 1950s when it started to be an important part of information science. This is primarily because sources of data for bibliometric analyses proliferated (and keep proliferating), inviting new analysis methods and uses of results. A most important area that emerged is evaluative bibliometrics: applications that seek to assess or evaluate the impact of research, or more broadly, scholarly work in general. Examples: use of citations in promotion and tenure deliberations; ranking or comparison of scholarly productivity; relative contribution of individuals, groups, institutions, nations; relative standing of journals; and the like. Evaluative indicators were developed to numerically express the impact of given entities. They have a serious impact on evaluation, policy formulation, and decision-making in a number of areas outside of information science. The big question for bibliometrics is: *What are the features and laws of the recorded information universe?*
- **Digital libraries:** 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. Information science is involved in both of these orientations, with connections to computer science and librarianship. The big question for digital libraries area is: *How can access to and use of recorded information be made most rapid and effective using digital resources and libraries?*

**InCID: How do you understand today, the development of research on evaluation of information systems in the digital world? What theme deemed more important or urgent: usability, quality of metadata in digital libraries human behavior, or something else?**

**Tefko Saracevic:** Evaluation in information science encompasses two very different worlds: that of information retrieval and that of digital libraries. 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 [US] 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. The basic criterion for evaluation is relevance, and the basic measures precision and recall, all established in the 1950s – so nothing changed there. 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; policy issues. New research topics are coming along at a brisk pace.

As mentioned, evaluation is one of the topics addressed. The issue of what criteria to use as a

base for evaluation is not yet settled. There are no standardized criteria yet, many were used. We can summarize these criteria in six classes with accompanying questions:

- **Content:** how well are digital collections selected, developed; objects created, organized, represented, presented?
- **Technology:** how well do hardware & software support library functions?
- **Interface:** what is available for users to interact & how much is interaction supported or hindered?
- **Process/service:** what processes & assistance are provided; what range of services is available; how well are they functioning; (carrying out tasks as: *search, browse, navigate, find, evaluate or obtain a resource*) ?
- **User:** what are the outcomes of DL use – changes in human information behavior, cognitive state, decision-making, problem-solving; impact on accomplishing tasks; broader impact/benefit in research, professional work?
- **Context:** how well does a DL fit into, respond to, follow larger context – institutional, economic, legal, social, cultural; effects on context?

As can be seen, usability is not mentioned separately and specifically. As important as usability is it proved to be a very broad, even slippery, criterion that encompasses some of these listed.

**InCID:** You have always emphasized that Information Science was not about technology, neither was a science dedicated on studying technologies, but rather, a user of technology, as a tool for solving problems, like any other modern science. And the problem in the 60's was the explosion of the literature. What, in your opinion, is the problem today?

**Tefko Saracevic:** The basic definition and notion of the field still stands: 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. The domain of information science was and still 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. The rapid pace of scientific and technical advances that were accumulating since the start of the 20<sup>th</sup> 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.

The same phenomenon of information explosion is continuing to date, even accelerating. But the nature and manifestation of knowledge records – the artifacts – have changed and dramatically. The variety of these knowledge artifacts is much greater. Technologies have changed – the Web and digitization are prime examples. The notion of what constitutes a “document” has also changed accordingly. Information explosion continues but is encompassing a much larger and diverse universe of knowledge records.

At the start of the millennium we have an additional phenomenon in many sciences: data explosion. The amount and diversity of data on many phenomena, primarily in natural sciences, is massive and growing at a high rate. Data is being produced at high speed – often real time, at an unprecedented scale, and with many new sensors. As yet, information science is not involved with data explosion. It should be. Possibly lessons learned from handling information explosion should be adapted and applied.

**InCID: Professor, you were, for over 20 years, editor of *Information Processing & Management Journal*. In which extension, this magazine has helped to consolidate Information Science in its intersection with managing and processing information.**

**Tefko Saracevic:** Here are some statistics that will illustrate changes better than any discussion. In 1985 (volume 21) when I assumed editorship of *Information Processing & Management* we published 552 pages with 34 articles from 57 authors residing in 13 countries. In 2007 (I stepped down at the end of volume 43) we published 1834 pages with 112 articles from 273 authors residing in 26 countries. In vol. 44, issue 2 (March 2008) we published an “Honor List of Reviewers for *Information Processing & Management*, June 2006 – December 2007.” For this year and a half the list had 398 reviewers residing in 43 countries. The growth of the journal was a reflection of the growth of the subjects we covered, information science in general and information retrieval and related areas in particular. The ever widening international spread of authors and reviewers is also a sign of global interest in these areas. Information science is indeed global. By the way, the original name of the

journals was Information Processing & Retrieval; the present name, Information Processing & Management was adopted with volume 11 in 1975.

**InCID: Forty years after the consolidation of Information Science, what do you think about some emergent criticisms named neo-documentation represented by authors such as Buckland, Frohmann or Rayward - among others - in which the central notion of Information Science - information - is being replaced by the notion of document?**

**Tefko Saracevic:** A field is defined by problems addressed and solutions used. Every field in history over time has concentrated on different (or differing) problems using a variety of solutions. The notion of document comes from the field of documentation, as the name implies. Documentation, as an area of inquiry and professional practice, was one such phase in addressing the overall problems of information explosion. The solution was organization of information in documents – various methods of indexing and classification with concentration on records in sciences. It began in Europe at the close of the 19<sup>th</sup> century. Many good and successful approaches were developed. Even bibliometrics emerged as an area of study with laws based on data analyses. Documentation was never strong in the United States. It waned after the Second World War worldwide. It is a historical phase. Information science addressed the same problem of information explosion but with very different solutions and approaches. The notion of information is central and still remains so, despite a number of academic discussions. The notion of document is neither central nor particularly useful.

Submetido em 19 nov. 2010