# Editorial – The Context of Information Retrieval

Information Retrieval (usually abbreviated IR) has been around for at least 40 years. By this we mean that the mechanisation of information storage and retrieval using computer technology has a history about as long as that of the modern computer. Probably the first person to think about computer-based solutions for information retrieval problems was Robert Fairthorne. In the early fifties he investigated the use of Hollerith punched-card equipment for simple retrieval of bibliographical references (the report of this work was only published in 1958). There was also the early visionary paper by Vanevar Bush in the *Atlantic Monthly* in 1945 which revealed 'Memex: a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.' Bush's ideas were recognised by Nelson, who traced back his ideas for Hypertext to at least this article by Bush, thus establishing a direct link between current work on hypermedia and early ideas in IR.

In the early sixties much exciting theoretical work was done in information retrieval. At this stage there was no clear idea of the limitations on computer power and storage, and so people enthusiastically proposed techniques, systems and models under the illusion that the computer would magically solve the storage and computation problems. Soon it was realised by some that to advance IR, experimental or practical demonstrations of new ideas were needed on realistic data. Substantial hard scientific work was done in the late sixties and early seventies establishing the viability and soundness of some simple underlying IR models.

In the mid-seventies a number of mathematically sophisticated IR models were defined such as the vector space and probabilistic models for retrieval. There was at this stage a significant experimental methodology for IR, and so these models were subjected to thorough experimental investigation. One of the concrete results to emerge from all this was that relevance feedback was indeed an effective way of enhancing retrieval. This technique has since then filtered into a number of commercial systems, albeit in a somewhat simplified form. The developments both theoretical and experimental continued into the eighties. During this latter phase interest shifted back to examining natural-language-processing techniques as a way of improving retrieval performance when the stored information is largely textual. This is after a long period during which it had been conceded that linguistic tools could not be used to enhance retrieval. Also, because of developments in work stations and more generally in human–computer interaction, interest has been rekindled in studying interface issues for IR systems, especially for multi-media environments.

We think that in the nineties IR research is more confident and has established strong links with AI, DBMS, HCI and Cognitive Science. To a certain extent this strength is reflected in the spread of papers presented in this issue. We do not attempt to cover all the active research areas in IR; for example, the interface area is hardly represented (but see the forthcoming special issue on multi-media).

In this issue we deal with a number of models for both linear (conventional) and non-linear (hypermedia) systems. It describes a number of formal, e.g. probabilistic, logical and network approaches to the underlying inference methods used to support retrieval. There is a long history associated with these, so there have been a number of revisions. Nevertheless, it is clear that modelling the retrieval process through a well-defined paradigm has led to a much deeper understanding of the problems of IR, opening up the prospect for a unified theory.

IR has always had a close relationship with database research; already in the mid-sixties Maron and Levien proposed a relational data model for information retrieval. Ever since that time researchers in both disciplines have made frequent attempts to develop models and to build systems that will deal efficiently and effectively with both structured and unstructured data. In this issue we present a recent attempt to continue the debate.

Undoubtedly one of the most vexing questions in IR is how to extract automatically the content from a textual document. There is now a vast literature on automatic indexing, and more recently, researchers have turned to the language tools studied in the AI context. The demands made of these tools by IR are considerable; for a tool to be acceptable it must scale to very large collections of textual objects. Here we contribute to both the philosophy and implementation of natural-language processing.

The roots of many of the ideas in IR lie in the earlier traditional work of librarians; after all, they were the experts in information storage and retrieval of old! To this day research work benefits from a study of the information-seeking activities of librarians and the users of library services. In many ways modern IR systems attempt to automate the expert decisions of a librarian. Thus it makes sense to model the librarian in interaction with a literature-seeking user; we present one such attempt here.

KEITH VAN RIJSBERGEN
MARISTELLA AGOSTI