Preface

Information retrieval (IR) is concerned with the effective and efficient retrieval of information based on its semantic content. The central problem in IR is the quest to find the set of relevant documents, among a large collection, containing the information sought, thereby satisfying a user's information need usually expressed in a natural language query. Documents may be objects or items in any medium: text, image, audio, or indeed a mixture of all three.

This book contains the proceedings of the *Third European Summer School in Information Retrieval (ESSIR 2000)*, held on 11–15 September 2000, in Villa Monastero, Varenna, Italy.

The event was jointly organised by the Institute of Multimedia Technologies of the CNR (National Council of Research) based in Milan (Italy), the Department of Electronics and Computer Science of the University of Padova (Italy), and the Department of Computer Science of the University of Strathclyde, Glasgow (UK). Administrative support was provided by Milano Ricerche, a consortium of industries, research institutions and the University of Milano, whose purpose is to provide administrative and technical support for the research and development activities of its members.

This third edition of the European Summer School in Information Retrieval is part of the ESSIR series which began in 1990. The first was organised by Maristella Agosti of the University of Padova and was held in Bressanone (Italy) in 1990. The second ESSIR was organised by Keith van Rijsbergen of the University of Glasgow (UK) and held in Glasgow in 1995, in the context of the IR Festival.

At the time of the first ESSIR, the Internet did not exist, so there is no website available for this event, but from its second edition a web presentation has been made available: the URL for ESSIR'95 is: `http://www.dcs.gla.ac.uk/essir/`, and the URL for ESSIR 2000 is: `http://www.itim.mi.cnr.it/Eventi/ essir2000/index.htm`. These websites contain useful material. In particular, the ESSIR 2000 website contains copies of the material distributed at the school (presentation, notes, etc.).

The aim of ESSIR 2000 was to give participants a grounding in the core subjects of IR, including methods and techniques for designing and developing IR systems, web search engines, and tools for information storing and querying in digital libraries. To achieve these aims, the program of ESSIR 2000 was organised into a series of lectures divided into foundations and advanced parts as reported in the next section. The lecturers were leading European researchers (with only one non-European exception), their course subjects strongly reflecting the research work for which they are all well known.

ESSIR 2000 was intended for researchers starting out in IR, for industrialists who wish to know more about this increasingly important topic and for people

working on topics related to the management of information on the Internet. This book, distributed at the school in draft form to incorporate in the final version useful participants' comments, contains 12 chapters written by the school's lecturers, providing surveys of the state of the art of IR and related areas.

Book Structure

The ESSIR 2000 programme of lectures and this book are divided into in two parts: one part on the foundations of IR and related areas (e.g. digital libraries), and one on advanced topics.

The part on foundations contains seven papers/chapters. In Chap. 1, Keith van Rijsbergen introduces some underlying concepts and ideas essential for understanding IR research and techniques. He also highlights some related hot areas of research, emphasising the role of IR in each. In Chap. 2, Norbert Fuhr presents the main mathematical models of IR. This paper provides the theoretical basis for representing the informative content of documents and for estimating the relevance of a document to a query. In Chap. 3, Páraic Sheridan and Carol Peters detail the issues and proposed solutions for multilingual information access in digital archives. Chapter 4, by Stephen Robertson, addresses the topic of evaluation, a very important aspect of IR. In Chap. 5 and 6, Alan Smeaton and John Eakins address issues and techniques related to indexing, browsing and searching multimedia information (audio, image, or digital video). Finally, in Chap. 7 Ingeborg Solvberg covers the basics and the challenges of digital libraries.

The part on advanced topics contains five papers/chapters. In Chap. 8, Peter Ingwersen concentrates on user issues and the usability of interactive IR. Chap. 9, by Fabio Crestani and Mounia Lalmas addresses the use of logic and uncertainty theories in IR. Closely related is Chap. 10, by Gabriella Pasi and Gloria Bordogna, which presents the area of research that aims at modelling the vagueness and imprecision involved in the IR process. In Chap. 11, Maristella Agosti and Massimo Melucci address the use of IR techniques on the Web for searching and browsing. Finally, in Chap. 12, Yves Chiaramella addresses the issues related to indexing and retrieval of structured documents.

Acknowledgements

The editors would like to thank all the participants of ESSIR 2000 for making the event a success. ESSIR 2000 was a success not just for the quality of the lectures, the authority of the lecturers, and the beautiful surroundings, it was a success because it was informal and interactive. For the best part of a week, more than 60 participants and 12 lecturers exchanged ideas and inspirations on where IR is at and where it should go. Many attendants (not just school participants, but some of the lecturers too) returned home with renewed encouragement and motivation.

We thank the sponsoring and supporting institutions for making it possible, financially, to hold the event. Also, we thank the Local Organising Committee,

the student volunteers and the personnel of Villa Monastero (Rino Venturini) for their invaluable help.

A special thanks to all the lecturers for their contributions, encouragement, and support. The quality of this book is mostly due to their work.

Finally, we would like to thank the Board of the Special Interest Network on Information Retrieval of the Council of European Professional Informatics Societies (CEPIS-IR), which includes Keith van Rijsbergen, Norbert Fuhr and Alan Smeaton, for their scientific support and invaluable advice on the school content and program.

September 2000   Maristella Agosti
                                                                    Fabio Crestani
                                                                   Gabriella Pasi

Organisation and Support

Scientific Program and Organising Committee

ESSIR 2000 was jointly organised by:

- Maristella Agosti, Department of Electronics and Computer Science, University of Padova, Padova, Italy;
- Fabio Crestani, Department of Computer Science, University of Strathclyde, Glasgow, UK;
- Gabriella Pasi, Institute of Multimedia Technologies, National Council of Research (CNR), Milan, Italy.

Local Organising Committee

ESSIR 2000 was locally organised by the Institute of Multimedia Technologies of CNR in Milan, Italy. In particular by: Gabriella Pasi, Gloria Bordogna, Paola Carrara, Alba L'Astorina, Luciana Onorato and Bruna Zonta.

Sponsoring Institutions

The main sponsoring and supporting organisation was the Special Interest Network on Information Retrieval of the Council of European Professional Informatics Societies (CEPIS-IR). CEPIS-IR provided a running grant, which made it possible to award a number of bursaries to support young students and researchers to attend the school. CEPIS-IR also provided invaluable advice on the school program.

The other sponsors were:

- Arnoldo Mondadori Editore, Verona, Italy;
- Microsoft Italia, Milan, Italy;
- Oracle Italia, Milan, Italy;
- Sharp Laboratories of Europe, Oxford, UK;
- 3D Informatica, San Lazzaro di Savena (Bologna), Italy.

Supporting Institutions

ESSIR 2000 benefited from the support of the following organisations:

- CEPIS-IR (Special Interest Network on Information Retrieval of the Council of European Professional Informatics Societies);
- AEI (Gruppo Specialistico Tecnologie e Applicazioni Informatiche);
- EUREL (Convention of National Societies of Electrical Engineers of Europe).