

Preface

While information retrieval was developed within the librarians' community well before the use of computers, its importance boosted at the turn of the century, with the diffusion of the World Wide Web. Big players in the computer industry, such as Google and Yahoo!, were the primary contributors of a technology for fast access to Web information. Searching capabilities are now integrated in most information systems, ranging from business management software and customer relationship systems to social networks and mobile phone applications. The technology for searching the Web is thus an important ingredient of computer science education that should be offered at both the bachelor and master levels, and is a topic of great interest for the wide community of computer science researchers and practitioners who wish to continuously educate themselves.

Contents

This book consists of three parts.

- The first part addresses the *principles of information retrieval*. It describes the classic metrics of information retrieval (such as precision and relevance), and then the methods for processing and indexing textual information, the models for answering queries (such as the binary, vector space, and probabilistic models), the classification and clustering of documents, and finally the processing of natural language for search. The purpose of Part I is to provide a systematic and condensed description of information retrieval before focusing on its application to the Web.
- The second part addresses the *foundational aspects of Web information retrieval*. It discusses the general architecture of search engines, focusing on the crawling and indexing processes, and then describes link analysis methods (and specifically PageRank and HITS). It then addresses recommendation and diversification as two important aspects of search results presentation and finally discusses advertising in search, the main fuel of search industry, as it contributes to most of the revenues of search engine companies.

- The third part of the book describes *advanced aspects of Web search*. Each chapter provides an up-to-date survey on current Web research directions, can be read autonomously, and reflects research activities performed by some of the authors in the last five years. We describe how data is published on the Web in a way to provide usable information for search engines. We then address meta-search and multi-domain search, two approaches for search engine integration; semantic search, an important direction for improved query understanding and result presentation which is becoming very popular; and search in the context of multimedia data, including audio and video files. We then illustrate the various ways for building expressive search interfaces, and finally we address human computation and crowdsearching, which consist of complementing search results with human interactions, as an important direction of development.

Educational Use

This book covers the needs of a short (3–5 credit) course on information retrieval. It is focused on the Web, but it starts with Web-independent foundational aspects that should be known as required background; therefore, the book is self-contained and does not require the student to have prior background. It can also be used in the context of classic (5–10 credit) courses on database management, thus allowing the instructor to cover not only structured data, but also unstructured data, whose importance is growing. This trend should be reflected in computer science education and curricula.

When we first offered a class on Web information retrieval five years ago, we could not find a textbook to match our needs. Many textbooks address information retrieval in the pre-Web era, so they are focused on general information retrieval methods rather than Web-specific aspects. Other books include some of the content that we focus on, however dispersed in a much broader text and as such difficult to use in the context of a short course. Thus, we believe that this book will satisfy the requirements of many of our colleagues.

The book is complemented by a set of author slides that instructors will be able to download from the Search Computing website, www.search-computing.org.

Milan, Italy

Stefano Ceri
Alessandro Bozzon
Marco Brambilla
Emanuele Della Valle
Piero Fraternali
Silvia Quarteroni

Web Information Retrieval

Ceri, S.; Bozzon, A.; Brambilla, M.; Della Valle, E.;

Fraternali, P.; Quarteroni, S.

2013, XIV, 284 p., Hardcover

ISBN: 978-3-642-39313-6