

Preface

Information Retrieval has radically changed over the last 25 years. When I first started teaching Information Retrieval and developing large Information Retrieval systems in the 1980s it was easy to cover the area in a single semester course. Most of the discussion was theoretical with testing done on small databases and only a small subset of the theory was able to be implemented in commercial systems. There were not massive amounts of data in the right digital format for search. Since 2000, the field of Information retrieval has undergone a major transformation driven by massive amounts of new data (e.g., Internet, Facebook, etc.) that needs to be searched, new hardware technologies that makes the storage and processing of data feasible along with software architecture changes that provides the scalability to handle massive data sets. In addition, the area of information retrieval of multimedia, in particular images, audio and video, are part of everyone's information world and users are looking for information retrieval of them as well as the traditional text. In the textual domain, languages other than English are becoming far more prevalent on the Internet.

To understand how to solve the information retrieval problems is no longer focused on search algorithm improvements. Now that Information Retrieval Systems are commercially available, like the area of Data Base Management Systems, an Information Retrieval System approach is needed to understand how to provide the search and retrieval capabilities needed by users. To understand modern information retrieval it's necessary to understand search and retrieval for both text and multimedia formats. Although search algorithms are important, other aspects of the total system such as pre-processing on ingest of data and how to display the search results can contribute as much to the user finding the needed information as the search algorithms.

This book provides a theoretical and practical explanation of the latest advancements in information retrieval and their application to existing systems. It takes a system approach, discussing all aspects of an Information Retrieval System. The system approach to information retrieval starts with a functional discussion of what is needed for an information system allowing the reader to understand the scope of the information retrieval problem and the challenges in providing the needed functions. The book, starting with the Chap. 1, stresses that information retrieval

has migrated from textual to multimedia. This theme is carried throughout the book with multimedia search, retrieval and display being discussed as well as all the classic and new textual techniques. Taking a system view of Information Retrieval explores every functional processing step in a system showing how decisions on implementation at each step can add to the goal of information retrieval; providing the user with the information they need minimizing their resources in getting the information (i.e., time it takes). This is not limited to search speed but also how search results are presented can influence how fast a user can locate the information they need. The information retrieval system can be defined as four major processing steps. It starts with “ingestion” of information to be indexed, the indexing process, the search process and finally the information presentation process. Every processing step has algorithms associated with it and provides the opportunity to make searching and retrieval more precise. In addition the changes in hardware and more importantly search architectures, such as those introduced by GOOGLE, are discussed as ways of approaching the scalability issues. The last chapter focuses on how to evaluate an information retrieval system and the data sets and forums that are available. Given the continuing introduction of new search technologies, ways of evaluating which are most useful to a particular information domain become important.

The primary goal of writing this book is to provide a college text on Information Retrieval Systems. But in addition to the theoretical aspects, the book maintains a theme of practicality that puts into perspective the importance and utilization of the theory in systems that are being used by anyone on the Internet. The student will gain an understanding of what is achievable using existing technologies and the deficient areas that warrant additional research. What used to be able to be covered in a one semester course now requires at least three different courses to provide adequate background. The first course provides a complete overview of the Information Retrieval System theory and architecture as provided by this book. But additional courses are needed to go in more depth on the algorithms and theoretical options for the different search, classification, clustering and other related technologies whose basics are provided in this book. Another course is needed to focus in depth on the theory and implementation on the new growing area of Multimedia Information Retrieval and also Information Presentation technologies.

Gerald Kowalski



<http://www.springer.com/978-1-4419-7715-1>

Information Retrieval Architecture and Algorithms

Kowalski, G.

2011, XII, 305 p., Hardcover

ISBN: 978-1-4419-7715-1