

# **Information Retrieval**

## **Data Structures & Algorithms**

Edited by

**William B. Frakes**  
*Software Engineering Guild*

**Ricardo Baeza-Yates**  
*University of Chile*

# Contents

---

---

**FOREWORD** v

**PREFACE** vii

## **INTRODUCTION**

- 1** Introduction to Information Storage and Retrieval Systems 1  
*W. B. Frakes*
- 2** Introduction to Data Structures and Algorithms Related  
to Information Retrieval 13  
*R. S. Baeza-Yates*

## **FILE STRUCTURES**

- 3** Inverted Files 28  
*D. Harman, E. Fox, R. A. Baeza-Yates, W. Lee*
- 4** Signature Files 44  
*C. Faloutsos*
- 5** New Indices for Text: PAT trees and PAT arrays 66  
*G. H. Gonnet, R. A. Baeza-Yates, T. Snider*

- 6** File Organizations for Optical Disks    **83**  
*D. A. Ford, S. Christodoulakis*

## **TERM AND QUERY OPERATIONS**

- 7** Lexical Analysis and Stoplists    **102**  
*C. Fox*
- 8** Stemming Algorithms    **131**  
*W. B. Frakes*
- 9** Thesaurus Construction    **161**  
*P. Srinivasan*
- 10** String Searching Algorithms    **219**  
*R. A. Baeza-Yates*
- 11** Relevance Feedback and Other Query Modification Techniques    **241**  
*D. Harman*

## **DOCUMENT OPERATIONS**

- 12** Boolean Operations    **264**  
*S. Wartik*
- 13** Hashing Algorithms    **293**  
*S. Wartik, E. Fox, L. Heath, Q. Chen*
- 14** Ranking Algorithms    **363**  
*D. Harman*
- 15** Extended Boolean Models    **393**  
*E. Fox, S. Betrabet, M. Koushik, W. Lee*
- 16** Clustering Algorithms    **419**  
*E. Rasmussen*

## **HARDWARE**

- 17** Special-Purpose Hardware for Information Retrieval    **443**  
*L. Hollaar*
- 18** Parallel Information Retrieval Algorithms    **459**  
*C. Stanfill*

**Index**    **498**