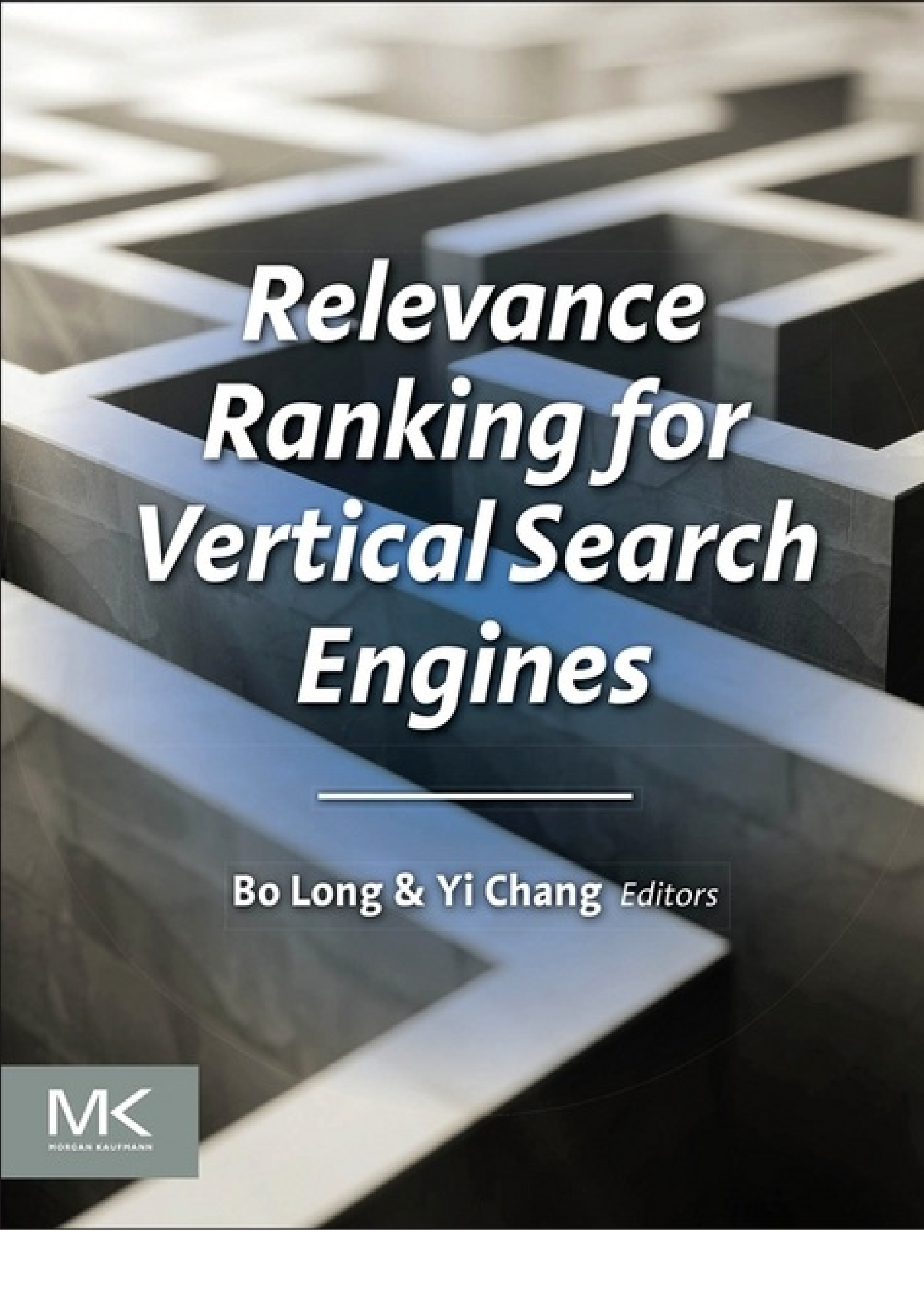
Relevance Ranking for Vertical Search Engines
Bo Long & Yi Chang Editors
MK
MORGAN KAUFMANN

Relevance Ranking for Vertical Search Engines

FIRST EDITION

Bo Long and Yi Chang

LinkedIn Inc., Mountain View, CA, USA
Yahoo! Labs, Sunnyvale, CA, USA

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Morgan Kaufmann is an imprint of Elsevier

Table of Contents

Copyright

Acquiring Editor: Steve Elliot
Editorial Project Manager: Lindsay Lawrence
Project Manager: Punithavathy Govindaradjane
Designer: Maria Inês Cruz
Morgan Kaufmann is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
Relevance ranking for vertical search engines / Bo Long, Yi Chang (Editors).
pages cm
Includes bibliographical references and index.
ISBN 978-0-12-407171-1
1. Text processing (Computer science) 2. Sorting (Electronic computers) 3. Relevance. 4. Database searching. 5. Search engines–Programming. I. Long, Bo, editor of compilation.
II. Chang, Yi (Computer expert)
QA76.9.T48R455 2014
025.04–dc23
2013039777

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library
ISBN: 978-0-12-407171-1

Printed and bound in the United States of America
14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

For information on all MK publications visit our website at *www.mkp.com*

List of Tables

List of Figures

About the Editors

Bo Long is currently a staff applied researcher at LinkedIn Inc., and was formerly a senior researc scientist at Yahoo! Labs. His research interests lie in data mining and machine learning wit applications to web search, recommendation, and social network analysis. He holds eight innovatio and has published peer-reviewed papers in top conferences and journals including ICML, KD ICDM, AAAI, SDM, CIKM, and KAIS. He has served as reviewer, workshop co-organizer, conferenc organizer, committee member, and area chair for multiple conferences, including KDD, NIPS, SIGI ICML, SDM, CIKM, JSM, etc.

Yi Chang is a principal scientist and sciences director in Yahoo Labs, where he leads the search an anti-abuse science group. His research interests include web search, applied machine learning, natur language processing, and social computing. Yi has published more than 60 conference/journal paper and has served as workshop coorganizer, conference organizer, committee member, and area chair fo multiple conferences, including WWW, SIGIR, ICML, KDD, CIKM, etc.

List of Contributors

Jaime Arguello University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Jiang Bian Microsoft Inc., Beijing, P.R. China

Yi Chang Yahoo! Labs, Sunnyvale, CA, USA

Fernando Diaz Microsoft Inc., New York, NY, USA

Anlei Dong Yahoo! Labs, Sunnyvale, CA, USA

David Hanauer University of Michigan, Ann Arbor, MI, USA

Yoshiyuki Inagaki Yahoo! Labs, Sunnyvale, CA, USA

Changsung Kang Yahoo! Labs, Sunnyvale, CA, USA

Yuan Liu Microsoft Inc., Beijing, P.R. China

Bo Long LinkedIn Inc., Mountain View, CA, USA

Yuanhua Lv Microsoft Inc., Mountain View, CA, USA

Qiaozhu Mei University of Michigan, Ann Arbor, MI, USA

Tao Mei Microsoft Inc., Beijing, P.R. China

Belle Tseng Apple Inc., Cupertino, CA, USA

Srinivas Vadrevu Microsoft Inc., Mountain View, CA, USA

Hongning Wang University of Illinois at Urbana-Champaign, Urbana, IL, USA

Xuanhui Wang Facebook Inc., Menlo Park, CA, USA

Kai Zheng University of Michigan, Ann Arbor, MI, USA

Foreword

As the information available on the Internet continues to grow, Web searchers increasingly run into critical but not broadly discussed challenge: finding relevant, rich results for highly targeted ar specialized queries.

General Web searching has been with us for a long time, has spawned international powerhous companies like Yahoo! and Google, and is a staple of everyday life for hundreds of millions of peopl around the world. It is probable that literally billions of generic searches—for general informatio celebrities, sports scores, common products, and other items of interest—are well satisfied by th commonly known major search engines whose names are so familiar as to have become part of th vernacular. But since the Internet has become the business- and daily-life critical worldwide resourc it is today, increasingly diverse groups of people are relying on it to look for more and more divers things—things not so easily found by generic Web search engines.

For example, searches focused on travel planning may often have very specific but implic assumptions about results, such as the expectation of itineraries listed in order of departure or cos They might also benefit from additional nonobvious information that could be critically helpful, suc as changes in checked baggage policies, road construction near relevant airports, or State Departme travel warnings. General search engines have no more clue about these things than they do abo dosages of medications, celebrity divorces, slugging percentages, and other narrow, domain-specif information; they don't have access to site-specific signals. As a result, the role of so-called "vertica search engines, which focus on specific segments of online content and deep site-specific informatio has quietly increased to become essential to most people looking for key items online.

Somewhat hidden from view but no less important than general Web search technology, vertica search algorithms have been key to helping users find household products they care about, movi they want to see, potential dating partners, their perfect automobiles, well-matched insurance policie and thousands of other things for which generalized text processing algorithms are not well tuned ar cannot make the right judgments of relevance for results.

In this context, the term *vertical* is usually taken to connote in-depth treatment of fairly narro domains, e.g., medical information, rather than broad ranges of information that meet a very wid range of needs (as you will see, for the purposes of this book, *vertical* can also refer to a limited rang of search result types, such as entities or measurements or dates, or specific types of informatio access modalities, such as mobile search). What sets vertical search apart from more general, broa based search is the fact that relatively specific domain knowledge can be leveraged to find the rig pieces of information. Further, an understanding of a more limited set of information-seeking tasl (for example, looking for specific kinds of football statistics for your fantasy team) can also play a important role in satisfying narrower information needs. With fewer but very common tasks carrie out by users, it may be easier to infer a user's intent for a particular vertical search, which cou dramatically improve the quality and value of the results for the user.

In addition to the opportunity to provide highly relevant and richer results to information seeker

vertical search engines that focus on specific segments of online content have shown great potential offer advertisers more contextually relevant, better-targeted audiences for their ads. Given th dependence of the Internet search industry on advertising, this makes vertical search an economical central part of the Internet's future. There is no doubt that vertical search is starting to play a role which the significance was probably never imagined in the early days of Internet searching.

As with other forms of search, the heart of successful vertical search is relevance rankin Specialized understanding of the domain and sophisticated ranking algorithms is critical. Algorithn that work hard to infer a user's intention when doing a search are the ones that are successful. Th ability to use the right signals and successfully compare various aspects of a query and potenti retrieved results will make or break a search engine. And that is the focus of this book: introducin and evaluating the critical ranking technology needed to make vertical searching successful.

Although there exist many books on general Web search technology, this new volume is a uniqu resource, dedicated to vertical search technologies and the relevance ranking technology that mak them successful. The book takes a comprehensive view of this area and aims to become a authoritative source of information for search scientists, engineers, and other interested readers with technical bent. Despite many years of research on algorithms and methods of general Web searc vertical search deserves its own dedicated study and in-depth treatment because of the unique natu of its structures and applications. This volume provides that focused treatment, covering key issu such as cross-vertical searching, vertical selection and aggregation, news searches, object searche image searches, and medical domain searches.

The authors represented in this book are active researchers who cover many different aspects vertical search technology and who have made tangible contributions to the progress of what is clear a dynamic research frontier. This ensures that the book is authoritative and reflects the current state the art. Nevertheless—and importantly—the book gives a balanced treatment of a wide spectrum topics, well beyond the individual authors' own methodologies and research specialties.

The book presents in-depth and systematic discussions of theories and practices for vertical searc ranking. It covers the obvious major fields as well as recently emerging areas for vertical searc including news search ranking, local search ranking, object search ranking, image search rankin medical domain search ranking, cross-vertical ranking, and vertical selection and aggregation. F each field, the book provides state-of-the-art algorithms with detailed discussions, includin background, derivation, and comparisons. The book also presents extensive experimental results various real application datasets to demonstrate the performance of various algorithms as well guidelines for practical use of those algorithms. It introduces ranking algorithms for various vertic search ranking applications and teaches readers how to manipulate ranking algorithms to achiev better results in real applications. Finally, the book provides thorough theoretical analysis of variou algorithms and problems to lay a solid foundation for future advances in the field.

Vertical search is still a fairly young and dynamic research field. This volume offers researcher and application developers a comprehensive overview of the general concepts, techniques, an applications of vertical search and helps them explore this exciting field and develop new metho and applications. It may also serve graduate students and other interested readers with a gener introduction to the state of the art of this promising research area. It uses plain language with detaile examples, including case studies and real-world, hands-on examples to explain the key concept models, and algorithms used in vertical search ranking. I think that overall you will find it qui readable and highly informative.

Although not widely known, vertical search is an essential part of our everyday lives on th

Internet. It is increasingly critical to users' satisfaction and increasing reliance on online data source and it provides extraordinary new opportunities for advertisers. Given recent growth in the applicatio of vertical search and our increasing daily reliance on it, you hold in your hands the first guidebook the next generation of information access on the Internet. I hope you enjoy it.

—Ron Brachman
Chief Scientist and Head, Yahoo! Labs
Yahoo!, Inc.

Introduction

Abstract

This book aims to a present systematic study of practices and theories for vertical search ranking. The studies in this book can categorized into two major classes. One class is single-domain-related ranking that focuses on ranking for a specific vertic such as news search ranking, medical domain search ranking, visual search ranking, mobile search ranking, and entity sear ranking. Another class is multidomain-related ranking, which focuses on ranking that involves multiple verticals, such multiaspect ranking, aggregating vertical search ranking, and cross-vertical ranking. This chapter discusses organizatio audience, and further reading for this book.

Keywords

Vertical search ranking
news search ranking
medical domain search ranking
visual search ranking
mobile search ranking
multiaspect relevance ranking
entity ranking
aggregated vertical search
cross-vertical search ranking

1.1 Defining the Area

In the past decade, the impact of general Web search capabilities has been stunning. However, wit exponential information growth on the Internet, it becomes more and more difficult for a general We search engine to address the particular informational and research needs of niche users. As a respons to the great need for deeper, more specific, more relevant search results, vertical search engines hav emerged in various domains. By leveraging domain knowledge and focusing on specific user task vertical search has great potential to serve users highly relevant search results from specific domains

The core component of vertical search is *relevance ranking*, which has attracted more and mo attention from both industry and academia during the past few years. This book aims to prese systematic study of practices and theories for vertical search ranking. The studies in this book can b categorized into to two major classes. One class is single-domain-related ranking that focuses o ranking for a specific vertical, such as news search ranking and medical domain search rankin However, in this book the term *vertical* has a more general meaning than topic. It refers to specif topics such as news and medical information, specific result types such as entities, and specific searc

interfaces such as mobile search. The second class of vertical search study covered in this book cla is multidomain-related ranking, which focuses on ranking involving multiple verticals, such multiaspect ranking, aggregating vertical search ranking, and cross-vertical ranking.

1.2 The Content and Organization of This Book

This book aims to present an in-depth and systematic study of practices and theories related to vertic search ranking. The organization of this book is as follows.

Chapter 2 covers news vertical search ranking. News is one of the most important of Internet user online activities. For a commercial news search engine, it is critical to provide users with the mo relevant and fresh ranking results. Furthermore, it is necessary to group the related news articles that users can browse search results in terms of news stories rather than individual news articles. Th chapter describes a few algorithms for news search engines, including ranking algorithms an clustering algorithms. For the ranking problem, the main challenge is achieving appropriate balan between topical relevance and freshness. For the clustering problem, the main challenge is how group related news articles into clusters in a scalable mode. Chapter 2 introduces a few news searc ranking approaches, including a learning-to-rank approach and a joint learning approach fro clickthroughs. The chapter then describes a scalable clustering approach to group news search results

Chapter 3 studies another important vertical search, the medical domain search. With th exponential growth of electronic health records (EHRs), it is imperative to identify effective means help medical clinicians as well as administrators and researchers retrieve information from EHR Recent research advances in natural language processing (NLP) have provided improved capabiliti for automatically extracting concepts from narrative clinical documents. However, before these NL based tools become widely available and versatile enough to handle vaguely defined informatio retrieval needs by EHR users, a convenient and cost-effective solution continues to be in gre demand. In this chapter, we introduce the concept of medical information retrieval, which provid medical professionals a handy tool to search among unstructured clinical narratives via an interfa similar to that of general-purpose Web search engines, e.g., Google. In the latter part of the chapte we also introduce several advanced features, such as intelligent, ontology-driven medical search que recommendation services and a collaborative search feature that encourages sharing of medical searc knowledge among end users of EHR search tools.

Chapter 4 is intended to introduce some fundamental and practical technologies as well as som major emerging trends in visual search ranking. The chapter first describes the generic visual searc system, in which three categories of visual search are presented: i.e., *text-based*, *query example-base* and *concept-based* visual search ranking. Then we describe the three categories in detail, including review of various popular algorithms. To further improve the performance of initial search result visual search re-ranking of four paradigms will be presented: 1) *self-reranking*, which focuses o detecting relevant patterns from initial search results without any external knowledge; 2) *exampl based reranking*, in which the query examples are provided by users so that the relevant patterns ca be discovered from these examples; 3) *crowd-reranking*, which mines relevant patterns from crow sourcing information available on the Web; and 4) *interactive reranking*, which utilizes us interaction to guide the reranking process. In addition, we also discuss the relationship betwe learning and visual search, since most recent visual search ranking frameworks are developed base on machine learning technologies. Last, we conclude with several promising directions for futu research.

Chapter 5 introduces *mobile search ranking*. The wide availability of Internet access on mobi devices, such as phones and personal media players, has allowed users to search and access We information while on the go. The availability of continuous fine-grained location information on the devices has enabled mobile local search, which employs user location as a key factor to search f local entities (e.g., a restaurant, store, gas station, or attraction) to overtake a significant part of th query volume. This is also evident by the rising popularity of location-based search engines on mobi devices, such as Bing Local, Google Local, Yahoo! Local, and Yelp. The quality of any mobile loca search engine is mainly determined by its ranking function, which formally specifies how we retriev and rank local entities in response to a user's query. Acquiring effective ranking signals and heuristic to develop an effective ranking function is arguably the single most important research problem mobile local search. This chapter first overviews the ranking signals in mobile local search (e.g distance and customer rating score of a business), which have been recognized to be quite differe from general Web search. We next present a recent data analysis that studies the behavior of mobil local search ranking signals using a large-scale query log, which reveals interesting heuristics that ca be used to guide the exploitation of different signals to develop effective ranking features. Finally, w also discuss several interesting future research directions.

Chapter 6 is about *entity ranking*, which is a recent paradigm that refers to retrieving and rankin related objects and entities from different structured sources in various scenarios. Entities typical have associated categories and relationships with other entities. In this chapter, we introduce how t build a Web-scale entity ranking system based on machine = learned ranking models. Specifically, th entity ranking system usually takes advantage of structured knowledge bases, entity relationsh graphs, and user data to derive useful features for facilitating semantic search with entities direct within the learning-to-rank framework. Similar to generic Web search ranking, entity pairwis preference can be leveraged to form the objective function of entity ranking. More than that, th chapter introduces ways to incorporate the categorization information and preference of relate entities into the objective function for learning. This chapter further discusses how entity ranking different from regular Web search in terms of presentation bias and the interaction of categories q query entities and result facets.

Chapter 7 presents learning to rank with multiaspect relevance for vertical searches. Many vertic searches, such as local search, focus on specific domains. The meaning of relevance in these vertica is domain-specific and usually consists of multiple well-defined aspects. For example, in local searc text matching and distance are two important aspects to assess relevance. Usually, the overa relevance between a query and a document is a tradeoff among multiple aspect relevancies. Given single vertical, such a tradeoff can vary for different types of queries or in different contexts. In th chapter, we explore these vertical-specific aspects in the learning-to-rank setting. We propose a nove formulation in which the relevance between a query and a document is assessed with respect to eac aspect, forming the multiaspect relevance. To compute a ranking function, we study two types o learning-based approaches to estimate the tradeoff among these aspect relevancies: a lab aggregation method and a model aggregation method. Since there are only a few aspects, a minim amount of training data is needed to learn the tradeoff. We conduct both offline and online bucket-te experiments on a local vertical search engine, and the experimental results show that our propose multiaspect relevance formulation is very promising. The two types of aggregation methods perfor more effectively than a set of baseline methods including a conventional learning-to-rank method.

Chapter 8 focuses on *aggregated vertical search*. Commercial information access provide increasingly incorporate content from a large number of specialized services created for particul

information-seeking tasks. For example, an aggregated Web search page may include results fro image databases and news collections in addition to the traditional Web search results; a new provider may dynamically arrange related articles, photos, comments, or videos on a given artic page. These auxiliary services, known as *verticals*, include search engines that focus on a particul domain (e.g., news, travel, sports), search engines that focus on a particular type of media (e.g images, video, audio), and application programming interfaces (APIs) to highly targeted informatio (e.g., weather forecasts, map directions, or stock prices). The goal of *aggregated search* is to provi integrated access to all verticals within a single information context. Although aggregated search related to classic work in distributed information retrieval, it has unique signals, techniques, an evaluation methods in the context of the Web and other production information access systems. In th chapter, we present the core problems associated with aggregated search, which include sources predictive evidence, relevance modeling, and evaluation.

Chapter 9 presents recent advances in *cross-vertical ranking*. A traditional Web search engin conducts ranking mainly in a single domain, i.e., it focuses on one type of data source, and effecti modeling relies on a sufficiently large number of labeled examples, which require an expensive an time-consuming labeling process. On the other side, it is very common for a vertical search engine conduct ranking tasks in various verticals, which presents a more challenging ranking problem, that *cross-domain ranking*. Although in this book our focus is on cross-vertical ranking, the propose approaches can be applied to more general cases, such as cross-language ranking. Therefore, we use more general term, cross-domain ranking, in this book. For cross-domain ranking, in some domai we may have a relatively large amount of training data, whereas in other domains we can only colle very little. Theretofore, finding a way to leverage labeled information from related heterogeneou domain to improve ranking in a target domain has become a problem of great interest. In this chapte we propose a novel probabilistic model, pairwise cross-domain factor (PCDF) model, to address th problem. The proposed model learns latent factors (features) for multidomain data in partial overlapped heterogeneous feature spaces. It is capable of learning homogeneous feature correlatio heterogeneous feature correlation, and pairwise preference correlation for cross-domain knowledg transfer. We also derive two PCDF variations to address two important special cases. Under the PCD model, we derive a stochastic gradient-based algorithm, which facilitates distributed optimization an is flexible to adopt various loss functions and regularization functions to accommodate different da distributions. The extensive experiments on real-world data sets demonstrate the effectiveness of th proposed model and algorithm.

1.3 The Audience for This Book

The book covers major fields as well as recently emerging fields for vertical search. Therefore, th expected readership of this book includes all the researchers and systems development enginee working in these areas, including, but not limited to, Web search, information retrieval, data minin and specific application areas related to vertical search, such as various specific vertical searc engines. Since this book is self-contained in its presentation of the material, it also serves as an ide reference book for people who are new to the topic of vertical search ranking. Consequently, addition, the audience also includes anyone with interest or who works in a field requiring th reference book. Finally, this book can be used as a reference for a graduate course on advanced topi of information retrieval or data mining, since it provides a systematic introduction to this boomin new subarea of information technology.

1.4 Further Reading

As a newly emerging area of information retrieval and data mining, vertical search ranking is still its infant stage; currently there is no dedicated, premier venue for the publication of research in th area. Consequently, related work in this area, as the supplementary information to this book f further readings, may be found in the literature of the two parent areas.

In information retrieval, related work may be found in the premier conferences, such as the annu Association for Computing Machinery (ACM) Special Interest Group on Information Retriev (SIGIR) conference, the International World Wide Web Conference (WWW), and the ACM, th International Conference on Information and Knowledge Management (ACM CIKM). For journal the premier journals in the information retrieval area, including *Information Retrieval*, *Foundatio and Trends in Information Retrieval* (FTIR), may contain related work in vertical search ranking.

In data mining, related work may be found in the premier conferences, such as the AC International Conference on Knowledge Discovery and Data Mining (KDD), the Institute of Electric and Electronics Engineers (IEEE), International Conference on Data Mining (ICDM), and the Socie for Industrial and Applied Mathematics (SIAM) International Conference on Data Mining (SDM). particular, related work may be found in the workshop dedicated to the area of relational learnin such as the Statistical Relational Learning workshop. The premier journals in the data mining are including *IEEE Transactions on Knowledge and Data Engineering* (TKDE), *ACM Transactions o Data Mining* (TDM), and *Knowledge and Information Systems* (KAIS), may contain related work relational data clustering.

News Search Ranking

Abstract

News search is one of the most important Internet user activities. For a commercial news search engine, it is critical to provi users with the most relevant and fresh ranking results. Furthermore, it is necessary to group the related news articles so that us can browse search results in terms of news stories rather than individual news articles. This chapter describes a few algorith for news search engines, including ranking algorithms and clustering algorithms. For the ranking problem, the main challenge achieving appropriate balance between topical relevance and freshness. For the clustering problem, the main challenge is grouping related news articles into clusters in a scalable mode. We begin by introducing a few news search ranking approach including a learning-to-rank approach (Section 2.1) and a joint learning approach from clickthroughs (Section 2.2). We th describe a scalable clustering approach to group news search results (Section 2.3).

Keywords

News search

freshness

relevance

clustering

temporal features

2.1 The Learning-to-Rank Approach

The main challenge for ranking in news search is how to make appropriate balance between tw factors:Relevance and freshness. Here relevance includes both topical relevance as well as new source authority.

A widely adopted approach in practice is to use a simple formula to combine relevance an freshness. For example, the final ranking score for a news article can be computed as

$$\mathbf{score}_{rank} = \mathbf{score}_{relevance} e^{-\beta t} \quad (2.1$$

where $\mathbf{score}_{relevance}$ is the value representing the relevance between query and news article, t isnew article age and $e^{-\beta t}$ is a time decay term, for which the older a news article is, the more penalty th article will receive for its final ranking. The parameter β is used to control the relative importance freshness in the final ranking result. In the literature of information retrieval, *document* is usually use to refer to a candidate item in ranking tasks. In this chapter, we use the terms *document* and *new article* equally because the application here is to rank news articles in a search.

The advantage of such a heuristic approach to a relevance and freshness combination is i efficiency in real practice, for which only the value of the parameter β needs to be tuned by usin

some ranking examples. Furthermore, the appropriate β value often leads to good ranking results f many queries, which also makes this approach effective.

The drawback of this approach is that it is incapable of further improving ranking performanc because such a heuristic rule is too naive to handle more complicated ranking cases. For example, (2.1), time decay is represented by the term $e^{-\beta t}$, which is totally dependent on the document age. fact, an appropriate time decay factor should also rely on the nature of the query, since differe queries have different time sensitivities: If a query is related to breaking news, such as an earthquak that has just happened and has extensive media reports on casualty and rescue, then freshness shou be very important because even a document published only one hour ago could be outdated. On th other hand, if a query is for an event that happened weeks ago, then relevance is more important ranking because the user would like to find the most relevant and comprehensive reports in the searc results.

2.1.1 Related Works

Many prior works have exploited the temporal dimension in searches. For example, Baeza-Yates *et a* [22] studied the relation among Web dynamics, structure, and page quality and demonstrated th PageRank is biased against new pages. In T-Rank Light and T-Rank algorithms [25], both activi (i.e., update rates) and freshness (i.e., timestamps of most recent updates) of pages and links are take into account in link analysis. Cho *et al.* [66] proposed a page quality ranking function in order alleviate the problem of popularity-based ranking, and they used the derivatives of PageRank forecast future PageRank values for new pages. Nunes [269] proposed to improve Web informatio retrieval in the temporal dimension by combining the temporal features extracted from bo individual documents and the whole Web. Pandey *et al.* [276] studied the tradeoff between new pag exploration and high-quality page exploitation, which is based on a ranking method to random promote some new pages so that they can accumulate links quickly.

Temporal dimension is also considered in other information retrieval applications. Del Corso *et a* [94] proposed the ranking framework to model news article generation, topic clustering, and sto evolution over time, and this ranking algorithm takes publication time and linkage time in consideration as well as news source authority. Li *et al.* [221] proposed a TS-Rank algorithm, whic considers page freshness in the stationary probability distribution of Markov chains, since th dynamics of Web pages are also important for ranking. This method proves effective in th application of publication search. Pasca [277] used temporal expressions to improve questio answering results for time-related questions. Answers are obtained by aggregating matching pieces information and the temporal expressions they contain. Furthermore, Arikan *et al.* [20] incorporate temporal expressions into a language model and demonstrated experimental improvement in retriev effectiveness.

Recency query classification plays an important role in recency ranking. Diaz [98] determined th newsworthiness of a query by predicting the probability of a user clicking on the news display of query. König *et al.* [204] estimated the clickthrough rate for dedicated news search results with supervised model, which is to satisfy the requirement of adapting quickly to emerging news event.

2.1.2 Combine Relevance and Freshness

Learning-to-rank algorithms have shown significant and consistent success in various applicatio

[226,184,406,54]. Such machine-learned ranking algorithms learn a ranking mechanism by optimizi particular loss functions based on editorial annotations. An important assumption in those learni methods is that document relevance for a given query is generally stationary over time, so that, long as the coverage of the labeled data is broad enough, the learned ranking functions wou generalize well to future unseen data. Such an assumption is often true in Web searches, but it is le likely to hold in news searches because of the dynamic nature of news events and the lack of time annotations.

A typical procedure is as follows:

- Collect query-URL pairs.
- Ask editors to label the query-URL pairs with relevance grades.
- Apply a learning-to-rank algorithm to the train ranking model.

Traditionally, in learning-to-rank, editors label query-URL pairs with relevance grades, whic usually have four or five values, including *perfect*, *excellent*, *good*, *fair*, or *bad*. Editorial labeli information is used for ranking model training and ranking model evaluation. For training, the relevance grades are directly mapped to numeric values as learning targets.

For evaluation, we desire an evaluation metric that supports graded judgments and penalizes erro near the beginning of the ranked list. In this work, we useDCG [175],

$$\mathrm{DCG}_n = \sum_{i=1}^{n} \frac{G_i}{\log_2(i+1)}, \tag{2.2}$$

where i is the position in the document list, and G_i is the function of relevance grade. Because t range of DCG values is not consistent across queries, we adopt the NDCG as our primary rankin metric,

$$\mathrm{NDCG}_n = Z_n \sum_{i=1}^{n} \frac{G_i}{\log_2(i+1)}, \tag{2.3}$$

where Z_n is a normalization factor, which is used to make the NDCG of the ideal list be 1. We can us NDCG_1 and NDCG_5 to evaluate the ranking results.

We extend the learning-to-rank algorithm in news searches, for which we mainly make tw modifications due to the dynamic nature of the news search: (1) training sample collection and (editorial labeling guideline.

2.1.2.1 Training Sample Collection

The training sample collection has to be near real time for news searches by the following steps:

1. Sample the *latest* queries from the news search query log.
2. *Immediately* get the candidate URLs for the sampled queries.
3. *Immediately* ask editors to do judgments on the query-URL pairs with relevance and freshness grades.

We can see that all the steps need to be accomplished in a short period. Therefore, the trainin sample collection has to be well planned in advance; otherwise, any delay during this procedure wou

affect the reliability of the collected data. If queries are sampled from an outdated query log or if a of the selected candidate URLs are outdated, they cannot represent the real data distribution. If edito do not label query-URL pairs on time, it will be difficult for them to provide accurate judgment because editors' judgments rely on their good understanding of the related news events, whic becomes more difficult as time elapses.

2.1.2.2 Editorial Labeling

In a news search, editors should provide query-URL grades on both traditional relevance an freshness. Although document age is usually available in news searches, it is impossible to determin a document's freshness based solely on document age. A news article published one day ago coul either be very fresh or very old, depending on the nature of the related news event. So, we ask edito to provide subjective judgments on document freshness, which can have a few different grades, suc as the following:

- Very fresh: latest documents (promote grade: 1)
- Fresh (promote grade: 0)
- A little bit outdated (demote grade: −1)
- Totally outdated and useless (demote grade: −2)

For ranking model training, we combine relevance grade and recency grade for a new grade learning target. An example of such a grade combination is shown in the list. If the document is ver fresh, we promote one grade from its original relevance grade. For example, if a document has a *goo* grade on relevance and a *fresh* grade on freshness, its final grade is *excellent*, which is one grad higher than it is relevance grade, *good*. If the document is fresh, we neither promote nor demote. If th document is a little bit outdated, we demote one grade from its original relevance grade. If th document is totally outdated and useless, we demote two grades.

For ranking model evaluation, we can compute DCG values based on either the combined grades the relevance grade and freshness grade separately.

To evaluate freshness in isolation, we also include a freshness metric based on DCG,called DCF:

$$\mathrm{DCF}_n = \sum_{i=1}^{n} \frac{F_i}{\log_2(i+1)}, \tag{2.4}$$

where i is the position in the document list, and F_i is the freshness label (1 or 0). A query may hav multiple very fresh documents—for example, when multiple news sources simultaneously publis updates to some ongoing news story. Note that DCF is a recency measurement that is independent overall relevance. Therefore, when we evaluate a ranking, we should first consider demoted NDC which represents the overall relevance, then inspect the value of the DCF. We also define normalize discounted cumulative freshness (NDCF) in the same way as for NDCG in Eq. 2.3.

2.2 Joint Learning Approach from Clickthroughs

Given the highly dynamic nature of news events and the sheer scale of news reported around th globe, it is often impractical for human editors to constantly keep track of the news and provid timely relevance and freshness judgments. To validate this conjecture, we asked editors to annota