Computational Advertising: Techniques for Targeting Relevant Ads

Kushal Dave
LTRC
International Institute of Information Technology
Hyderabad, India
kushal.dave@research.iiit.ac.in

Vasudeva Varma
LTRC
International Institute of Information Technology
Hyderabad, India
vv@iiit.ac.in

the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

K. Dave and V. Varma. *Computational Advertising: Techniques for Targeting Relevant Ads*. Foundations and Trends® in Information Retrieval, vol. 8, no. 4-5, pp. 263–418, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-60198-833-1

Foundations and Trends® in Information Retrieval
Volume 8, Issue 4-5, 2014

Editorial Scope

Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation, and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2014, Volume 8, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Foundations and Trends® in Information Retrieval
Vol. 8, No. 4-5 (2014) 263–418

DOI: 10.1561/1500000045

Computational Advertising: Techniques for Targeting Relevant Ads

Kushal Dave
LTRC
International Institute of Information Technology
Hyderabad, India
kushal.dave@research.iiit.ac.in

Vasudeva Varma
LTRC
International Institute of Information Technology
Hyderabad, India
vv@iiit.ac.in

Contents

Abstract

Computational Advertising, popularly known as online advertising or Web advertising, refers to finding the most relevant ads matching a particular *context* on the Web. The context depends on the type of advertising and could mean – content where the ad is shown, the user who is viewing the ad or the social network of the user. Computational Advertising (CA) is a scientific sub-discipline at the intersection of information retrieval, statistical modeling, machine learning, optimization, large scale search and text analysis. The core problem addressed in Computational Advertising is of match-making between the ads and the context.

CA is prevalent in three major forms on the Web. One of the forms involves showing textual ads relevant to a query on the search page, known as Sponsored Search. On the other hand, showing textual ads relevant to a third party webpage content is known as Contextual Advertising. The third form of advertising also deals with the placement of ads on third party webpages, but the ads in this form are rich multimedia ads – image, video, audio, flash. The business model with rich media ads is slightly different from the ones with textual ads. These ads are also called banner ads, and this form of advertising is known as Display Advertising.

Both Sponsored Search and Contextual Advertising involve retrieving relevant ads for different types of content (query and Web page). As ads are short and are mainly written to attract the user, retrieval of ads pose challenges like vocabulary mismatch between the query/content and the ad. Also, as the user's probability of examining an ad decreases with the position of the ad in the ranked list, it is imperative to keep the best ads at the top positions. Display Advertising poses several challenges including modeling user behaviour and noisy page content and bid optimization on the advertiser's side. Additionally, online advertising faces challenges like false bidding, click spam and ad spam. These challenges are prevalent in all forms of advertising. There has been a lot of research work published in different areas of CA in the last one and a half decade. The focus of this survey is to discuss the problems and solutions pertaining to the information retrieval, machine learning and

statistics domain of CA. This survey covers techniques and approaches that deal with several issues mentioned above.

Research in Computational Advertising has evolved over time and currently continues both in traditional areas (vocabulary mismatch, query rewriting, click prediction) and recently identified areas (user targeting, mobile advertising, social advertising). In this study, we predominantly focus on the problems and solutions proposed in traditional areas in detail and briefly cover the emerging areas in the latter half of the survey. To facilitate future research, a discussion of available resources, list of public benchmark datasets and future directions of work is also provided in the end.

K. Dave and V. Varma. *Computational Advertising: Techniques for Targeting Relevant Ads*. Foundations and Trends® in Information Retrieval, vol. 8, no. 4-5, pp. 263–418, 2014.

DOI: 10.1561/1500000045.

1

Introduction

Advertising plays a vital role in supporting free websites and smartphone apps. Most of the popular websites like Google, Bing, YouTube, Yahoo!, Facebook, LinkedIn have a major share of their revenue coming through some form of advertising. Even small sites like blogs, home pages, forums are mostly supported by ads. The recent surge of interest in the research communities (industry and academia) is a testimonial of the huge promise the science of CA has on offer.

Computational Advertising, a term recently coined, is about using various computational methodologies to do contextually targeted advertising Broder [2008]. The central problem addressed in CA is: targeting ads that best match the context. The context involves content (query, Web page content), user information and location information. Instances of content based targeting include Sponsored Search and Contextual Advertising. *Sponsored Search* (SS) refers to the placement of ads on search results page. In SS, the context is the query issued by the user and the problem is to retrieve top relevant ads that semantically matches the query. *Contextual Advertising* (ConAd) deals with the placement of ads on third-party Web pages. It is similar to SS, with the ads being matched to the complete Web page text as opposed to a query. *Display Advertising* involves showing rich media ads (image, flash, video and audio) based on the page context, user information and/or location.

Placing contextually relevant ads has a two-fold advantage. First, the user's immediate interest in the topic can be exploited, which in turn increases the chance of users exploring the ads. More relevant the ads, higher are the chances of user viewing/clicking the ad and better are the chances of increase in the revenue generated Kirmani and Yi [1991], YI [1990], Wang et al. [2002]. Second, it leads to a better user experience. On the other hand, randomly placing ads may lead to a poor user experience Wang et al. [2002].

1.1 Introduction to Computational Advertising

The core problem addressed in Computational Advertising is to find the best matching ads for a given context. Based on the targeting scheme, the context involves a combination of the content (Web content/query), user profile, demographics and other contextual aspects. Based on the form of advertising, one or more of the contextual factors may be leveraged to get relevant ads. Ad targeting in Sponsored Search and Contextual Advertising is different on many levels than Display Advertising. One of the primary differences is that Display Advertising deals with rich media ads (also known as banner/display ads) as compared to the other two forms which deal with textual ads. Also, the underlying business model for display ads is different from the textual ads. The challenges faced in textual ads and banner ads however are similar as all three forms of ads look at putting the best ads matching the context. In this survey, we mostly look at the challenges from the information retrieval and modeling perspective. Hence, most part of the survey is focussed on presenting techniques dealing with textual ads. Having said that, some of the techniques presented in this study also apply to Display Advertising as the science involved is similar. In the latter part of the second half, we discuss the business model and the recently evolved Real-Time bidding process in Display Advertising Wang and Yuan [2013], Pandey [2013], iPinYou [2014], Yuan et al.

[2013], Chen et al. [2011], Weinan Zhang [2014]. In the first half of the survey, we discuss techniques from the perspective of textual ads. Also, we refer to textual ads as ads unless otherwise mentioned. Displaying

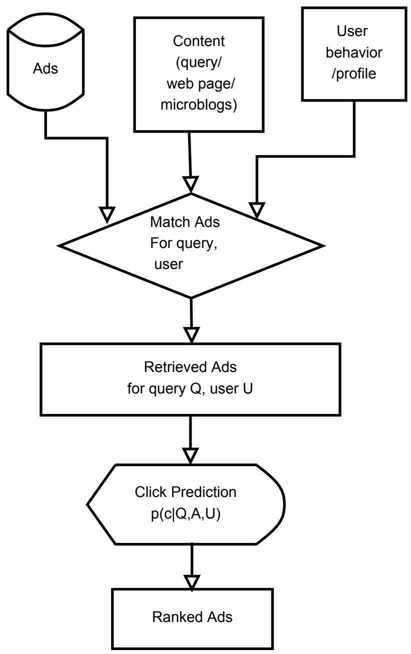

Figure 1.1: A typical ad system for Sponsored Search and Contextual Advertising: Once the ads are retrieved, they are ranked based on the probability of a click given the query, ad and the user

textual ads is typically done through a two-step process. The first step is to retrieve the relevant ads, as shown in Figure 1.1. The retrieved ads are then ranked based on the relevance and the ad value (bid amount). The retrieval and ranking of ads are separate stages in the overall ad placement process for the following reasons: 1. The retrieval of ads is done based on relevance only while the ranking needs to be done based on the value of the ad (bid amount) along with the relevance of the ad to the context. Hence, the criteria for both the processes are different. 2. Ad engines typically have billions of ads registered with it, and it is infeasible to rank a billion ads for a given query/content. Instead, first retrieving top-k relevant ads and ranking them based on the monetary

Figure 1.2: Structure of a typical textual Ad

value and relevance is more feasible and reasonable. Figure 1.1 shows a typical ad retrieval process. First, the top-k ads are retrieved. Next, they are ranked based on the click-through rate of the ad and the bid amount for the ad.

As content targeting deals with textual ads, we start with the description of a typical textual ad. Next, we describe how different types of content targeted advertising work. A sample ad is as shown in Figure 1.2.

1.1.1 Anatomy of a Textual Ad

A typical textual ad contains following fields Bendersky et al. [2010]:

- **Bid term/phrase**: The term bid by the advertiser for the ad. This is invisible to the user, and it is used to indicate what content the ad should be shown against. For each bid term bid by the advertiser, they have to pay the bid amount.
- **Bid amount**: The amount bid by the advertiser for the bid phrase. This too is invisible to the user.
- **Title**: This is the title of the ad.
- **Description/Creative**: The description is the text displayed below the title. It typically consists of a short description of the ad and is usually written to attract the user. It is also known as *creative*.

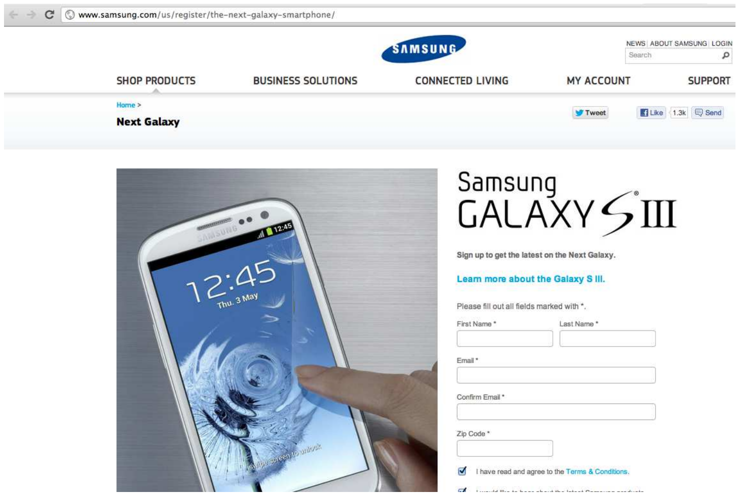

Figure 1.3: Landing page for the ad shown in Figure 1.2 (Notice the original/landing page URL is different from the display URL)

- **Display URL**: The URL displayed in the ad. To improve the presentation of ads and to reduce the space, the display URL is usually different from that of the original/landing page URL. The *landing page* for an ad is the page where a user lands after clicking on an ad as shown in Figure 1.3.

1.1.2 Matching Strategies and Pricing

Typically ad placement engines allow two different matching strategies for advertisers – *Exact match* and *Broad/Advance match* Choi et al. [2010]. Regardless of the matching strategy, every advertiser has to bid some amount on their bid phrase as shown in Figure 1.2. Next, the advertiser needs to choose the matching strategy. In the case of exact match, an ad is retrieved only if there is an exact match between the bid phrase of the ad and the text (query or Web page). In this scenario, the advertiser has knowledge of the keywords that are relevant to their business and makes a bid accordingly. Traditional information retrieval algorithms[1] like vector space model are usually employed for exact match systems.

Broad match allow advertisers to choose initial bid phrase, and the ad placement engine takes care of finding relevant content for the ad even if there is no exact match. This relaxes the constraint of coming up with all relevant bid phrases for the exact match involved in the previous case. Advertisers still have to bid on their ad. This bidding is real time, and as we will see later on in Chapter 6, the bid amount plays an important role in the position at which the ad is shown. With broad/advance match the ad placement engines employ sophisticated techniques to retrieve ads that are outside the syntax of the bid phrase of an ad. Due to the ease and the coverage involved with the broad/advance match, a majority of advertisers opt for advance match.

In an online advertising ecosystem, one of the following pricing schemes is adopted: Pay-per-Click (PPC), Pay-per-Impression (PPI), and Pay-per-Transaction (PPT) Broder et al. [2007]. In PPC model, the advertiser pays some amount each time a user clicks their ad. In PPI model, the advertiser pays every time their ad is displayed against the content. While in a PPT model the advertiser has to pay only when a user does a transaction after clicking on the ad. Sponsored Search and Contextual Advertising typically follow PPC model Broder et al. [2007, 2008b], Radlinski et al. [2008]. Display Advertising follows the PPI model Shen [2002], Li and Jhang-Li [2009].

Earlier, ad engines used to rank ads solely based on the amount bid by the advertiser. This, intuitively, was the most obvious way of maximizing revenue. However, ad engines soon realized that not all the top bid ads are relevant to the content. Irrelevant ads can result in user dissatisfaction Wang et al. [2002]. Hence, ad engines started ranking ads as a function of both relevance and expected revenue Richardson et al. [2007]. Displaying ads against the content is typically done through a two-step process. First, the top k ads are fetched from the ad database based on the extent to which they match the content. Fetching the ini-

[1] For a timeline on IR techniques, readers are advised to refer to Sanderson and Croft [2012]

tial top-k ads based on the content ensures that the ads to be displayed are relevant to the content. Once these top ads are retrieved they are ranked so as to maximize the overall expected revenue. Ranking in such a two-step fashion caters to the need of all the four parties involved – User, Advertiser, Publisher and Ad engine.

1.1.3 Scenarios in Online Advertising

In this section, we present the three most prevalent advertising scenarios in online advertising – Contextual Advertising, Sponsored Search and Display Advertising.

Contextual Advertising

A typical Contextual Advertising scenario is as shown in Figure 1.4. Today, many of the non-transactional websites rely at least to some extent on advertising revenue. Content targeting involves targeting websites ranging from blogs, forums, news pages, home pages to products sites and beyond. A user's visit on a page typically indicates their implicit interest in Web page's topic Broder et al. [2007]. This implicit interest can be exploited by placing relevant ads next to the content as there is a higher chance of user visiting the ad if it is relevant to the content. As shown in Figure 1.4, the content is about 'Fishing tips' and hence the relevant ads on fishing equipments and places for fishing.

Contextual Advertising can be seen as an interaction between the *publisher*, *advertiser*, *ad placement engine*, and the *user*. The publisher is the owner of content/Web page being targeted. The advertiser seeks to place their ad on the Web page. The ad placement engine acts as a mediator between the publisher and the advertiser. The ad placement engine decides which advertisement to be shown to which user. The user visits a Web page and is served the advertisements. Many research papers discuss work on Contextual Advertising Ribeiro Neto et al. [2005], Broder et al. [2007], Yih et al. [2006], Chakrabarti et al. [2008].

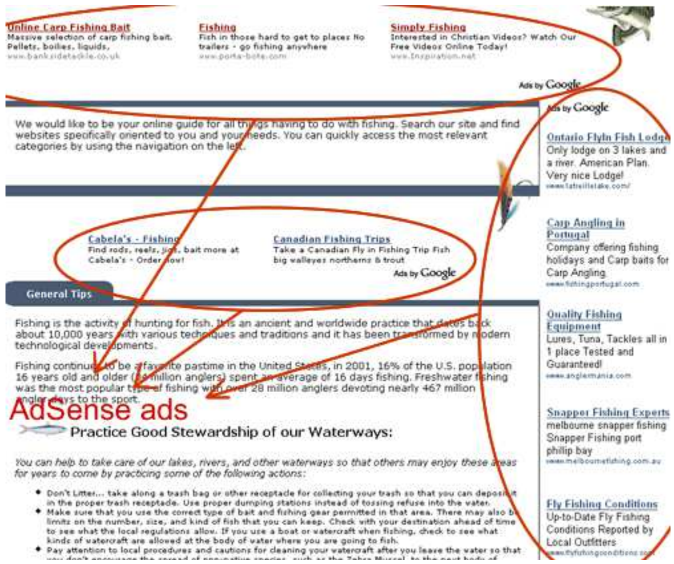

Figure 1.4: A typical Contextual Advertising scenario. Permission to use the image taken from the source: `http://www.ezmoneyon.net/wp-content/uploads/2008/01`.

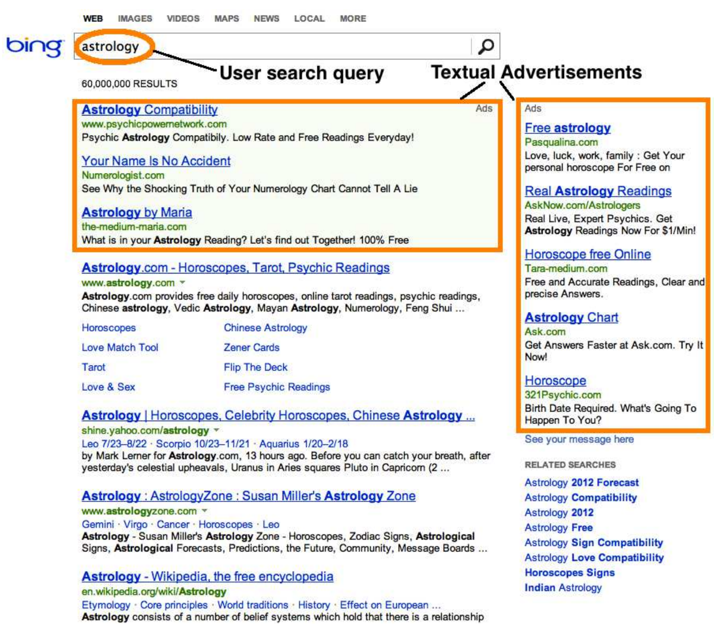

Figure 1.5: A typical Sponsored Search scenario

Figure 1.6: Showing Display Advertising scenario

Sponsored Search

In Sponsored Search, relevant advertisements are shown in response to a search query. A typical Sponsored Search scenario is illustrated in Figure 1.5. As can be seen, various relevant ads are shown for the query 'Astrology'. With Sponsored Search, user explicitly mentions their interest in the topic by issuing a query related to the topic. This explicit interest is exploited in Sponsored Search.

Sponsored Search can be seen as an interaction between three parties - *search engine, user* and the *advertiser*. The user issues a query to search engine related to the topic on which he/she seeks information. Advertisers and search engines try to exploit the immediate interest of the user in the topic by displaying ads relevant to the query topic. In a typical setting, advertisers bid on certain keywords known as bid terms and choose either advance or broad match. The advertiser's ad may get displayed based on the match between ad's bid term and the search query and the amount bid by the advertiser. Search engines try to rank the ads in a way that maximizes their revenue. For an excellent history of Sponsored Search please refer to Fain and Pedersen [2006], Jansen and Mullen [2008].

Display Advertising

Figure 1.6 shows the Yahoo! page with two display advertisements. Display Advertising is different from Contextual Advertising and Sponsored Search in many ways. Display ads (also called banner ads) usually come in a rich multimedia form – image, video, flash and audio. Li and Jhang-Li [2009], Barford et al. [2014]. In addition to direct response, display ads are also used for brand building Li and Jhang-Li [2009]. Also unlike Sponsored Search and Contextual Advertising, display ads are charged on a per impression basis Ghosh et al. [2009]. Almost 90% of the ads are billed on PPI basis in Display Advertising Shen [2002]. Publishers allot some space on their pages to show ads (could be a text ad or a banner ad). Display ads are usually targeted based on page content and user information. Barford et al. [2014] show that around 80% of display ads are targeted on profiles. Barford et al. [2014] give an excellent overview of the whole display ad landscape – they study different types of display ads prevalent in online advertising, analyse the dynamics of display ads.

As in the PPI model, bidding in Display Advertising happens on a per impression basis. Predominantly, the sale of the impression slots on the publisherâĂŹs page can happen in two ways – (a) Bulk sale of impressions and (b) Auction individual impressions in real time. In the case of bulk sale of impressions, the advertiser buys n number of impressions on the publisherâĂŹs page. The ad is shown on the page until the advertiserâĂŹs budget is exhausted. In a bulk sale, all the impressions are bought at a flat price. In the second type, the impressions are auctioned similar to a share market. For each impression, a separate auction takes place where a variety of advertisers bid for the impression slot. This entire process of auction happens in real time – the user visits a site, the publisher raises a bid request for the ad slot, the advertisers bid for the impression and the winner of the auction is allowed to display their ad on the page. This real time auction of impressions is commonly known as Real-Time Bidding (RTB). More details on RTB are given in Chapter 9.

1.2 Issues and Challenges

Content level targeting, at heart, is a combination of retrieval and ranking problem. However, unlike document retrieval, the ads are short and noisy. Hence, apart from the challenges faced in organic search, ad retrieval involve additional challenges. Based on the content to be targeted, following are the impediments and challenges in CA:

- **Short Ad text:**
  Ad text is short and is intended to attract the user, hence it contains short non-grammatical English phrases. This poses a lot of challenges in content level targeting Choi et al. [2010], Ribeiro Neto et al. [2005], Broder et al. [2007]. Traditional retrieval algorithms are not mainly designed to handle short text.

- **Sparse queries (Vocabulary mismatch):**
  In case of Sponsored Search, the query is issued by the user and ads are submitted by the advertiser and both are short, this often induces a problem called *vocabulary mis-match* Ribeiro Neto et al. [2005] between the ads and the queries Radlinski et al. [2008], Raghavan and Iyer [2010], Jones et al. [2006]. As the name suggests, vocabulary mismatch implies that the ad and query are semantically related but there is no syntactic similarity (word overlap) between them. For example, a query 'Camera' should also retrieve ads bidding on terms like 'Sony Cyber-shot' or 'Sony EOS'.

- **Noisy Web content:**
  Web pages usually contain noisy data. The application of traditional information retrieval algorithm to retrieve ads from such noisy pages may lead to irrelevant ads. Therefore, the noisy content of the Web page needs to be dealt with in a more sophisticated manner Yih et al. [2006], Dave and Varma [2010a], Wu and Bolivar [2008].

- **No Page Rank!:** Unlike Web search, there is no link structure among the documents (ads) that can be exploited to apply algorithms like Page Rank or HITS to serve authoritative and relevant ads.

- **Ad Spam and Click Spam:** Advertisers bid on false keywords or highly frequent keywords that are not related to their business. Identifying such spam ads is one of the biggest challenges. Click spam is the fraudulent spam by the user with no real intention of exploring the ad. If such clicks are not detected, advertisers can get falsely billed for such clicks Dave et al. [2012b].

- **Opinionated Content:**
  Some of the Web page content like forums and in particular microblogs are highly opinionated. Targeting ads on opinionated posts involves dealing with negative sentiments. Negative sentiments demand a separate treatment. Intuitively, targeting ads on negative sentiments may defeat the intended purpose of advertising. Imagine an ad for a fast food product, on a Web page talking about health concerns caused by fast food Fan and Chang [2009], Liu et al. [2008].

- **Dealing with new Ads in Ranking:**
  In order to maximize the expected revenue, the search engine must predict the probability of a click on an ad, more commonly known as click-through rate (CTR) of an ad. Historical click-through log is the most obvious proxy for estimating the CTR of the ads. However, for new ads entering into the system and infrequent/rare ads, it is very difficult to estimate the CTR as there is a very little or no information available through the click-through logs Dave and Varma [2010b], Richardson et al. [2007], Shaparenko et al. [2009], Regelson and Fain [2006], Ashkan et al. [2009], Debmbsczynski et al. [2008].

- **How much can behavioral targeting help online advertising? :** One big question in the case of content level targeting is whether user behavior can also be incorporated to retrieve more relevant ads. If incorporating user behavior helps, it evokes second-order questions, what kind of data should be used to profile the user behavior and what should be the time frame from which the data is considered for user modeling Yan et al. [2009], Cheng and Cantú-Paz [2010b], Ahmed et al. [2011]. Display Advertising leverage user behavioral information for showing their ads. Hence modeling user information is critical to Display Advertising.

- **What to consider while targeting users?:**
  In the case of user level targeting, one of the challenges is to profile the user for targeting them. Advertisers gather information about the user from the cookies. User modeling is more challenging than content modeling, as unlike the content, the user behavior changes with time. In the case of user targeting based on their social circle, formulating a user's influence on their contacts for various actions (like clicking on ads) is a big challenge Cheng and Cantú-Paz [2010b], Dave et al. [2011], Kempe et al. [2003], Hartline et al. [2008].

1.3 Scope of the Survey

Computational Advertising is a vast area encompassing different sciences in itself. It requires borrowing methodologies from information retrieval, machine learning, statistical modeling, microeconomics and game theory. Specifically, one needs information retrieval techniques to efficiently retrieve ads in real time and semantic matching of ads with the text. Machine learning techniques are used for tasks such as learning the ranking of ads and prediction of parameters. Tasks like modeling the user, recommending ads based on history and finding similar ads require statistical expertise, while microeconomics and game theory are involved in ad auctioning and bid economics.

In this study, we restrict ourselves to problems and techniques from the field of information retrieval, machine learning and statistical modeling. Modeling the auction process and the various problems and the solutions pertaining to bid optimization during auctions is outside the scope of this survey.

For a good read on various bid algorithms and the auction theory associated with them, readers are encouraged to refer to the Computational Advertising course mentioned in Section Section §11.3.

1.4 Organization of the Survey

In the coming chapters, we look at various research work done to overcome the issues and challenges mentioned in Section §1.2. In the first part till Chapter 5, we look at retrieving ads for different content types – Webpage content and search queries. In Chapter 2, we look at the problem of reducing noise from the Web page content to facilitate the matching of ads to the content. As ads are short, retrieving them requires certain preprocessing to overcome the shortness, like expanding the ad content or transforming the ads to other dimension. This is explained in Chapter 3. Queries are shorter than ads, and they need to be expanded before retrieving ads for them. Chapter 4 looks at the query treatment problem with respect to retrieving relevant ads in Sponsored Search. Click spam and false bidding are significant challenges in the retrieval of ads. Chapter 5 explains the work on determining the ad quality. Once ads are retrieved they need to be ranked based on the probability of a click. Chapter 6 and Chapter 7 describe the work on ranking ads in Sponsored Search and Contextual Advertising respectively. Chapter 8 describes work on user behavioral modeling and targeting part. Chapter 9 discusses Display Advertising and the recently evolved Real-Time Bidding process that lets advertisers micromanage their budget. We discuss some of the emerging advertising trends like Mobile Advertising, Advertising in Social news-feed. in Chapter 10. To facilitate future research work in CA, we enlist some publicly available datasets and mention some of the relevant conferences/journals and workshops to publish and/or find further relevant work in Chapter 11. We conclude in Chapter 12.

References

A Simplex Method for Function Minimization. *The Computer Journal*, 7(4): 308–313, January 1965. ISSN 1460-2067. . URL http://dx.doi.org/10.1093/comjnl/7.4.308.

ADKDD '09: Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-671-7.

ADKDD '10: Proceedings of the Fourth International Workshop on Data Mining for Online Advertising, New York, NY, USA, 2010. .

ADKDD '11: Proceedings of the Fifth International Workshop on Data Mining for Online Advertising, New York, NY, USA, 2011. .

. *ADKDD '12: Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1545-6.

ADKDD '13: Proceedings of the Seventh International Workshop on Data Mining for Online Advertising, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2323-9.

Deepak Agarwal, Rahul Agrawal, Rajiv Khanna, and Nagaraj Kota. Estimating rates of rare events with multiple hierarchies through scalable log-linear models. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 213–222, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0055-1. . URL http://doi.acm.org/10.1145/1835804.1835834.

Deepak Agarwal, Bo Long, Jonathan Traupman, Doris Xin, and Liang Zhang. Laser: A scalable response prediction platform for online advertising. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 173–182, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2351-2. . URL `http://doi.acm.org/10.1145/2556195.2556252`.

Rahul Agrawal, Archit Gupta, Yashoteja Prabhu, and Manik Varma. Multi-label learning with millions of labels: recommending advertiser bid phrases for web pages. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, pages 13–24, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2035-1. URL `http://dl.acm.org/citation.cfm?id=2488388.2488391`.

Amr Ahmed, Yucheng Low, Mohamed Aly, Vanja Josifovski, and Alexander J. Smola. Scalable distributed inference of dynamic user interests for behavioral targeting. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 114–122, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0813-7. . URL `http://doi.acm.org/10.1145/2020408.2020433`.

Aris Anagnostopoulos, Andrei Z. Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Lance Riedel. Just-in-time contextual advertising. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, CIKM '07, pages 331–340, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-803-9. . URL `http://doi.acm.org/10.1145/1321440.1321488`.

Tasos Anastasakos, Dustin Hillard, Sanjay Kshetramade, and Hema Raghavan. A collaborative filtering approach to ad recommendation using the query-ad click graph. In *Proceeding of the 18th ACM conference on Information and knowledge management*, CIKM '09, pages 1927–1930, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-512-3. . URL `http://doi.acm.org/10.1145/1645953.1646267`.

Peter Anick. Using terminological feedback for web search refinement: A log-based study. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, pages 88–95, New York, NY, USA, 2003. ACM. ISBN 1-58113-646-3. . URL `http://doi.acm.org/10.1145/860435.860453`.

Azin Ashkan and Charles Clarke. Impact of query intent and search context on click through behavior in sponsored search. *Knowledge and Information Systems*, pages 1–28, 2012. ISSN 0219-1377. URL `http://dx.doi.org/10.1007/s10115-012-0485-x`. 10.1007/s10115-012-0485-x.

Azin Ashkan, Charles L. A. Clarke, Eugene Agichtein, and Qi Guo. Estimating ad clickthrough rate through query intent analysis. In *WI-IAT '09: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pages 222–229, Washington, DC, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3801-3. .

Martin Azizyan, Ionut Constandache, and Romit Roy Choudhury. Surroundsense: Mobile phone localization via ambience fingerprinting. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, MobiCom '09, pages 261–272, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-702-8. . URL `http://doi.acm.org/10.1145/1614320.1614350`.

Ricardo Baeza-Yates and Felipe Saint-Jean. A three level search engine index based in query log distribution. In MarioA. Nascimento, EdlenoS. Moura, and ArlindoL. Oliveira, editors, *String Processing and Information Retrieval*, volume 2857 of *Lecture Notes in Computer Science*, pages 56–65. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-20177-9.

Ricardo A. Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999. ISBN 020139829X.

Somnath Banerjee, Krishnan Ramanathan, and Ajay Gupta. Clustering short texts using wikipedia. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 787–788, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-597-7. . URL `http://doi.acm.org/10.1145/1277741.1277909`.

Paul Barford, Igor Canadi, Darja Krushevskaja, Qiang Ma, and S. Muthukrishnan. Adscape: Harvesting and analyzing online display ads. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 597–608, Republic and Canton of Geneva, Switzerland, 2014. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2744-2. . URL `http://dx.doi.org/10.1145/2566486.2567992`.

Hila Becker, Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. Context transfer in search advertising. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 656–657, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. . URL `http://doi.acm.org/10.1145/1571941.1572063`.

Ben. Ssp to dsp cookie-synching explained. *http://www.adopsinsider.com/ad-exchanges/ssp-to-dsp-cookie-synching-explained/*, 2011.

Michael Bendersky, Evgeniy Gabrilovich, Vanja Josifovski, and Donald Metzler. The anatomy of an ad: Structured indexing and retrieval for sponsored search. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 101–110, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-799-8. . URL `http://doi.acm.org/10.1145/1772690.1772702`.

Adam Berger and John Lafferty. Information retrieval as statistical translation. In *In Proceedings of the 1999 ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 222–229, 1999.

Hemant K. Bhargava and Juan Feng. Paid placement strategies for internet search engines. In *Proceedings of the 11th international conference on World Wide Web*, WWW '02, pages 117–123, New York, NY, USA, 2002. ACM. ISBN 1-58113-449-5. . URL `http://doi.acm.org/10.1145/511446.511462`.

Rushi Bhatt, Vineet Chaoji, and Rajesh Parekh. Predicting product adoption in large-scale social networks. In *CIKM '10*, 2010.

Mikhail Bilenko, Evgeniy Gabrilovich, Matthew Richardson, and Yi Zhang. Information retrieval and advertising. *SIGIR Forum*, 43(2):29–33, December 2009. ISSN 0163-5840. . URL `http://doi.acm.org/10.1145/1670564.1670569`.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003. ISSN 1532-4435.

Google Developers Blog. Cookie matching. *https://developers.google.com/ad-exchange/rtb/cookie-guide#what-is*, 2014.

Burton H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM*, 13(7):422–426, July 1970. ISSN 0001-0782. . URL `http://doi.acm.org/10.1145/362686.362692`.

Gerlof Bouma. Normalized (pointwise) mutual information in collocation extraction. *Proc. GSCL Conf.*, 2009.

L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.

Andrei Broder. Computational advertising. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '08, pages 992–992, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics. URL `http://dl.acm.org/citation.cfm?id=1347082.1347190`.

Andrei Broder, Marcus Fontoura, Vanja Josifovski, and Lance Riedel. A semantic approach to contextual advertising. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 559–566, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-597-7. . URL http://doi.acm.org/10.1145/1277741.1277837.

Andrei Broder, Massimiliano Ciaramita, Marcus Fontoura, Ev-geniy Gabrilovich, Vanja Josifovski, Donald Metzler, Vanessa Murdock, and Vassilis Plachouras. To swing or not to swing: learning when (not) to advertise. In *Proceeding of the 17th ACM conference on Information and knowledge management*, CIKM '08, pages 1003–1012, New York, NY, USA, 2008a. ACM. ISBN 978-1-59593-991-3. . URL http://doi.acm.org/10.1145/1458082.1458216.

Andrei Broder, Peter Ciccolo, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Lance Riedel, and Jeffrey Yuan. Online expansion of rare queries for sponsored search. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 511–520, New York, NY, USA, 2009a. ACM. ISBN 978-1-60558-487-4. . URL http://doi.acm.org/10.1145/1526709.1526778.

Andrei Broder, Peter Ciccolo, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Lance Riedel, and Jeffrey Yuan. Online expansion of rare queries for sponsored search. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 511–520, New York, NY, USA, 2009b. ACM. ISBN 978-1-60558-487-4. . URL http://doi.acm.org/10.1145/1526709.1526778.

Andrei Broder, Evgeniy Gabrilovich, and Vanja Josifovski. Information retrieval challenges in computational advertising. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 908–908, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0153-4. . URL http://doi.acm.org/10.1145/1835449.1835680.

Andrei Z. Broder, Peter Ciccolo, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, and Lance Riedel. Search advertising using web relevance feedback. In *Proceeding of the 17th ACM conference on Information and knowledge management*, CIKM '08, pages 1013–1022, New York, NY, USA, 2008b. ACM. ISBN 978-1-59593-991-3. . URL http://doi.acm.org/10.1145/1458082.1458217.

Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Comput. Linguist.*, 19(2):263–311, June 1993. ISSN 0891-2017. URL `http://dl.acm.org/citation.cfm?id=972470.972474`.

Internet Advertising Bureau. Internet advertising revenues hit all-time first quarter high. 2014. .

Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 89–96, New York, NY, USA, 2005. ACM. ISBN 1-59593-180-5. . URL `http://doi.acm.org/10.1145/1102351.1102363`.

Christopher J. C. Burges, Robert Ragno, and Quoc V. Le. Learning to Rank with Nonsmooth Cost Functions. In Bernhard Schölkopf, John C. Platt, Thomas Hoffman, Bernhard Schölkopf, John C. Platt, and Thomas Hoffman, editors, *NIPS*, pages 193–200. MIT Press, 2006. ISBN 0-262-19568-2.

Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 129–136, New York, NY, USA, 2007a. ACM. ISBN 978-1-59593-793-3. . URL `http://doi.acm.org/10.1145/1273496.1273513`.

Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 129–136, New York, NY, USA, 2007b. ACM. ISBN 978-1-59593-793-3. . URL `http://doi.acm.org/10.1145/1273496.1273513`.

John Joseph Carrasco, John Joseph, Carrasco Daniel, C. Fain, Kevin J. Lang, and Leonid Zhukov. Clustering of bipartite advertiser-keyword graph. Yahoo technical report, http://labs.corp.yahoo.com/publications/17.pdf, 2003.

Deepayan Chakrabarti, Deepak Agarwal, and Vanja Josifovski. Contextual advertising by combining relevance with click feedback. WWW '08, pages 417–426. ACM, 2008. ISBN 978-1-60558-085-2. . URL `http://doi.acm.org/10.1145/1367497.1367554`.

Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In John Shawe-Taylor, Richard S. Zemel, Peter L. Bartlett, Fernando C. N. Pereira, and Kilian Q. Weinberger, editors, *NIPS*, pages 2249–2257, 2011. URL `http://dblp.uni-trier.de/db/conf/nips/nips2011.html#ChapelleL11`.

Olivier Chapelle and Ya Zhang. A dynamic bayesian network click model for web search ranking. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 1–10, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-487-4. .

Bowei Chen, Shuai Yuan, and Jun Wang. A dynamic pricing model for unifying programmatic guarantee and real-time bidding in display advertising. In *Proceedings of 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ADKDD'14, pages 1:1–1:9, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2999-6. . URL `http://doi.acm.org/10.1145/2648584.2648585`.

Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 199–208. ACM, 2009. ISBN 978-1-60558-495-9. .

Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 1029–1038, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0055-1. . URL `http://doi.acm.org/10.1145/1835804.1835934`.

Ye Chen, Pavel Berkhin, Bo Anderson, and Nikhil R. Devanur. Real-time bidding algorithms for performance-based display ad allocation. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 1307–1315, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0813-7. . URL `http://doi.acm.org/10.1145/2020408.2020604`.

Haibin Cheng and Erick Cantú-Paz. Personalized click prediction in sponsored search. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 351–360, New York, NY, USA, 2010a. ACM. ISBN 978-1-60558-889-6. . URL `http://doi.acm.org/10.1145/1718487.1718531`.

Haibin Cheng and Erick Cantú-Paz. Personalized click prediction in sponsored search. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 351–360, New York, NY, USA, 2010b. ACM. ISBN 978-1-60558-889-6. . URL `http://doi.acm.org/10.1145/1718487.1718531`.

Haibin Cheng, Roelof van Zwol, Javad Azimi, Eren Manavoglu, Ruofei Zhang, Yang Zhou, and Vidhya Navalpakkam. Multimedia features for click prediction of new ads in display advertising. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 777–785, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1462-6. . URL `http://doi.acm.org/10.1145/2339530.2339652`.

Yejin Choi, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Mauricio Mediano, and Bo Pang. Using landing pages for sponsored search ad selection. WWW '10, pages 251–260. ACM, 2010. ISBN 978-1-60558-799-8.

Massimiliano Ciaramita, Vanessa Murdock, and Vassilis Plachouras. Online learning from click data for sponsored search. In *Proceeding of the 17th international conference on World Wide Web*, WWW '08, pages 227–236, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2. . URL `http://doi.acm.org/10.1145/1367497.1367529`.

William W. Cohen and Yoram Singer. A simple, fast, and effective rule learner. In *In Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pages 335–342. AAAI Press, 1999.

David Cossock and Tong Zhang. Subset ranking using regression. In *Proceedings of the 19th Annual Conference on Learning Theory*, COLT'06, pages 605–619, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3-540-35294-5, 978-3-540-35294-5. . URL `http://dx.doi.org/10.1007/11776420_44`.

Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 87–94, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-927-9. .

Kushal S. Dave and Vasudeva Varma. Pattern based keyword extraction for contextual advertising. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 1885–1888. ACM, 2010a. ISBN 978-1-4503-0099-5. . URL `http://doi.acm.org/10.1145/1871437.1871754`.

Kushal S. Dave and Vasudeva Varma. Learning the click-through rate for rare/new ads from similar ads. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 897–898, New York, NY, USA, 2010b. ACM. ISBN 978-1-4503-0153-4. . URL `http://doi.acm.org/10.1145/1835449.1835671`.

Kushal S. Dave, Ankit Patil, and Vasudeva Varma. Topic models for retrieving relevant ads. In *Proceedings of workshop on Machine Learning for Online Advertising*, AdML @ ICML 2012, 2012a.

Kushal Shailesh Dave and Vasudeva Varma. Identifying microblogs for targeted contextual advertising. In John G. Breslin, Nicole B. Ellison, James G. Shanahan, and Zeynep Tufekci, editors, *ICWSM*. The AAAI Press, 2012. URL `http://dblp.uni-trier.de/db/conf/icwsm/icwsm2012.html#DaveV12`.

Kushal Shailesh Dave, Rushi Bhatt, and Vasudeva Varma. Modelling action cascades in social networks. In Lada A. Adamic, Ricardo A. Baeza-Yates, and Scott Counts, editors, *ICWSM*. The AAAI Press, 2011. URL `http://dblp.uni-trier.de/db/conf/icwsm/icwsm2011.html`.

Vacha Dave, Saikat Guha, and Yin Zhang. Measuring and fingerprinting click-spam in ad networks. *SIGCOMM Comput. Commun. Rev.*, 42(4):175–186, August 2012b. ISSN 0146-4833. . URL `http://doi.acm.org/10.1145/2377677.2377715`.

K. Debmbsczynski, W. Kotlowski, and D. Weiss. Predicting ads clickthrough rate with decision rules. In *Workshop on Target and Ranking for Online Advertising, WWW 08*, 2008.

Krzysztof Dembczynski, Wojciech Kotlowski, and Roman Slowinski. Solving regression by learning an ensemble of decision rules. In *Proceedings of the 9th International Conference on Artificial Intelligence and Soft Computing*, ICAISC '08, pages 533–544, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-69572-1.

Hongbo Deng, Irwin King, and Michael R. Lyu. Entropy-biased models for query representation on the click graph. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 339–346, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. .

Pedro Domingos. A few useful things to know about machine learning. *Commun. ACM*, 55(10):78–87, October 2012. ISSN 0001-0782. . URL `http://doi.acm.org/10.1145/2347736.2347755`.

Pedro Domingos and Matt Richardson. Mining the network value of customers. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66. ACM, 2001. ISBN 1-58113-391-X. .

Quang Duong and SÃľbastien Lahaie. Discrete choice models of bidder behavior in sponsored search. In Ning Chen, Edith Elkind, and Elias Koutsoupias, editors, *Internet and Network Economics*, volume 7090 of *Lecture Notes in Computer Science*, pages 134–145. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-25509-0. . URL `http://dx.doi.org/10.1007/978-3-642-25510-6_12`.

Georges E. Dupret and Benjamin Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 331–338, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-164-4. . URL `http://doi.acm.org/10.1145/1390334.1390392`.

Benjamin Edelman, Michael Ostrovsky, Michael Schwarz, Thank Drew Fudenberg, Louis Kaplow, Robin Lee, Paul Milgrom, Muriel Niederle, and Ariel Pakes. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97, 2005.

Benjamin Edelman, Michael Ostrovsky, Michael Schwarz, Thank Drew Fudenberg, Louis Kaplow, Robin Lee, Paul Milgrom, Muriel Niederle, and Ariel Pakes. A structural model of sponsored search advertising auctions. *Technical report, Microsoft*, 2010.

D. C. Fain and J. O. Pedersen. Sponsored search: a brief history. In *Second Workshop on Sponsored Search Auctions in ACM Conference on Electronic Commerce*. ACM, 2006.

Teng-Kai Fan and Chia-Hui Chang. Sentiment-oriented contextual advertising. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, ECIR '09, pages 202–215, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-00957-0.

Paolo Ferragina and Ugo Scaiella. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In Jimmy Huang, Nick Koudas, Gareth J. F. Jones, Xindong Wu, Kevyn Collins-Thompson, and Aijun An, editors, *CIKM*, pages 1625–1628. ACM, 2010. ISBN 978-1-4503-0099-5. URL `http://dblp.uni-trier.de/db/conf/cikm/cikm2010.html`.

Eibe Frank, Gordon W. Paynter, Ian H. Witten, Carl Gutwin, and Craig G. Nevill-Manning. Domain-specific keyphrase extraction. In *IJCAI '99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 668–673, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc. ISBN 1-55860-613-0.

Matthew Fredrikson and Benjamin Livshits. Repriv: Re-envisioning in-browser privacy.

Yoav Freund and Robert E. Schapire. *A decision-theoretic generalization of on-line learning and an application to boosting*. 1995.

Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. An efficient boosting algorithm for combining preferences. *J. Mach. Learn. Res.*, 4: 933–969, December 2003. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=945365.964285`.

Hillel Fuld. The most amazing viral campaign in the history of the world-wide web. *http://blog.inner-active.com/2012/04/the-most-amazing-viral-campaign-in-the-history-of-the-worldwide-web-no-really/*, 2014.

Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. Introduction to computational advertising. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Tutorial Abstracts*, HLT-Tutorials '08, pages 1–1, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics. URL `http://dl.acm.org/citation.cfm?id=1564169.1564170`.

Arpita Ghosh and Aaron Roth. Selling privacy at auction. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC '11, pages 199–208, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0261-6. . URL `http://doi.acm.org/10.1145/1993574.1993605`.

Arpita Ghosh, Benjamin I.P. Rubinstein, Sergei Vassilvitskii, and Martin Zinkevich. Adaptive bidding for display advertising. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 251–260, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-487-4. . URL `http://doi.acm.org/10.1145/1526709.1526744`.

Google-Adwords. Enhance your ad with extensions. *https://support.google.com/adwords/answer/2375499?hl=en*, 2014.

Amit Goyal, Francesco Bonchi, and Laks V.S. Lakshmanan. Learning influence probabilities in social networks. In *WSDM '10: Proceedings of the third ACM international conference on Web search and data mining*, pages 241–250. ACM, 2010. ISBN 978-1-60558-889-6. .

Thore Graepel, Joaquin QuiÃśonero Candela, Thomas Borchert, and Ralf Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In Johannes FÃijrnkranz and Thorsten Joachims, editors, *ICML*, pages 13–20. Omnipress, 2010. ISBN 978-1-60558-907-7. URL `http://dblp.uni-trier.de/db/conf/icml/icml2010.html`.

Laura A. Granka, Thorsten Joachims, and Geri Gay. Eye-tracking analysis of user behavior in www search. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 478–479, New York, NY, USA, 2004. ACM. ISBN 1-58113-881-4. . URL `http://doi.acm.org/10.1145/1008992.1009079`.

Saikat Guha, Alexey Reznichenko, Kevin Tang, Hamed Haddadi, and Paul Francis. Serving ads from localhost for performance, privacy, and profit. In *In Proceedings of the 8th Workshop on Hot Topics in Networks (HotNets âĂŹ09*, 2009.

Hamed Haddadi, Pan Hui, and Ian Brown. Mobiad: Private and scalable mobile advertising. In *Proceedings of the Fifth ACM International Workshop on Mobility in the Evolving Internet Architecture*, MobiArch '10, pages 33–38, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0143-5. . URL `http://doi.acm.org/10.1145/1859983.1859993`.

Parissa Haghirian and Akihiro Inoue. An advanced model of consumer attitudes toward advertising on the mobile internet. *Int. J. Mob. Commun.*, 5(1):48–67, December 2007. ISSN 1470-949X. . URL `http://dx.doi.org/10.1504/IJMC.2007.011489`.

Michaela Hardt and Suman Nath. Privacy-aware personalization for mobile advertising. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 662–673, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1651-4. . URL `http://doi.acm.org/10.1145/2382196.2382266`.

Jason Hartline, Vahab Mirrokni, and Mukund Sundararajan. Optimal marketing strategies over social networks. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 189–198. ACM, 2008. ISBN 978-1-60558-085-2. .

T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2001. ISBN 9780387952840. URL `http://books.google.co.in/books?id=VRzITwgNV2UC`.

R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. In A.J. Smola, P.L. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 115–132, Cambridge, MA, 2000. MIT Press. URL `http://stat.cs.tu-berlin.de/publications/papers/herobergrae99.ps.gz`.

Dustin Hillard, Stefan Schroedl, Eren Manavoglu, Hema Raghavan, and Chirs Leggetter. Improving ad relevance in sponsored search. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 361–370, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-889-6. . URL http://doi.acm.org/10.1145/1718487.1718532.

Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '99, pages 50–57, New York, NY, USA, 1999. ACM. ISBN 1-58113-096-1. . URL http://doi.acm.org/10.1145/312624.312649.

Anette Hulth. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 216–223, Morristown, NJ, USA, 2003. Association for Computational Linguistics. .

IAB. *IAB internet advertising revenue report 2013 full year results*. 2014. URL http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_FY_2013.pdf.

Samuel Ieong, Mohammad Mahdian, and Sergei Vassilvitskii. Advertising in a stream. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 29–38, Republic and Canton of Geneva, Switzerland, 2014. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2744-2. . URL http://dx.doi.org/10.1145/2566486.2568030.

iPinYou. Rtb101. *http://contest.ipinyou.com/manual.shtml*, 2014.

Bernard J. Jansen and Tracy Mullen. Sponsored search: an overview of the concept, history, and technology. *International Journal of Electronic Business (IJEB)*, 6(2):114–131, 2008.

Kalervo Järvelin and Jaana Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00, pages 41–48, New York, NY, USA, 2000. ACM. ISBN 1-58113-226-3. . URL http://doi.acm.org/10.1145/345508.345545.

Joanna Jaworska and Marcin Sydow. Behavioural targeting in on-line advertising: An empirical study. In *Proceedings of the 9th International Conference on Web Information Systems Engineering*, WISE '08, pages 62–76, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-85480-7. . URL http://dx.doi.org/10.1007/978-3-540-85481-4_7.

Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 133–142, New York, NY, USA, 2002. ACM. ISBN 1-58113-567-X. . URL `http://doi.acm.org/10.1145/775047.775067`.

Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. Accurately interpreting clickthrough data as implicit feedback. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 154–161, New York, NY, USA, 2005. ACM. ISBN 1-59593-034-5. .

Rosie Jones, Benjamin Rey, Omid Madani, and Wiley Greiner. Generating query substitutions. In *Proceedings of the 15th international conference on World Wide Web*, WWW '06, pages 387–396, New York, NY, USA, 2006. ACM. ISBN 1-59593-323-9. . URL `http://doi.acm.org/10.1145/1135777.1135835`.

Maryam Karimzadehgan, Wei Li, Ruofei Zhang, and Jianchang Mao. A stochastic learning-to-rank algorithm and its application to contextual advertising. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 377–386, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0632-4. . URL `http://doi.acm.org/10.1145/1963405.1963460`.

Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *Proceedings of the 23rd International Conference on Algorithmic Learning Theory*, ALT'12, pages 199–213, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-34105-2. . URL `http://dx.doi.org/10.1007/978-3-642-34106-9_18`.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003. ISBN 1-58113-737-0. .

Amna Kirmani and Youjae Yi. The effects of advertising context on consumer responses. In *Association for Consumer Research, Advances in Consumer Research*, Volume 18, pages 414–416, 1991.

Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. In *SODA '98: Proceedings of the 9th ACM-SIAM symposium on Discrete algorithms*, pages 668–677, Philadelphia, PA, USA, 1998. ISBN 0-89871-410-9.

Aleksander Kolcz, Vidya Prabakarmurthi, and Jugal Kalita. Summarization as feature selection for text categorization. In *Proceedings of the Tenth International Conference on Information and Knowledge Management*, CIKM '01, pages 365–370, New York, NY, USA, 2001. ACM. ISBN 1-58113-436-3. . URL http://doi.acm.org/10.1145/502585.502647.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. Adaptive computation and machine learning. MIT Press, 2009. ISBN 9780262013192. URL http://books.google.co.in/books?id=7dzpHCHzNQ4C.

Yehuda Koren. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 426–434, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-193-4. . URL http://doi.acm.org/10.1145/1401890.1401944.

Aleksandra Korolova. Privacy violations using microtargeted ads: A case study. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops*, ICDMW '10, pages 474–482, Washington, DC, USA, 2010. IEEE Computer Society. ISBN 978-0-7695-4257-7. . URL http://dx.doi.org/10.1109/ICDMW.2010.137.

Anísio Lacerda, Marco Cristo, Marcos André Gonçalves, Weiguo Fan, Nivio Ziviani, and Berthier Ribeiro-Neto. Learning to advertise. SIGIR '06, pages 549–556. ACM, 2006. ISBN 1-59593-369-7. . URL http://doi.acm.org/10.1145/1148170.1148265.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. ISBN 1-55860-778-1. URL http://dl.acm.org/citation.cfm?id=645530.655813.

Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, EC '06, pages 228–237, New York, NY, USA, 2006. ACM. ISBN 1-59593-236-4. . URL http://doi.acm.org/10.1145/1134707.1134732.

P. Li, C. Burges, Q. Wu, J. C. Platt, D. Koller, Y. Singer, and S. Roweis. McRank: Learning to Rank Using Multiple Classification and Gradient Boosting. *Advances in Neural Information Processing Systems*, 2007a.

Xiao Li, Ye-Yi Wang, and Alex Acero. Extracting structured information from user queries with semi-supervised conditional random fields. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, pages 572–579, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. . URL `http://doi.acm.org/10.1145/1571941.1572039`.

Ying Li, Arun C. Surendran, and Dou Shen. Data mining and audience intelligence for advertising. *SIGKDD Explor. Newsl.*, 9(2):96–99, December 2007b. ISSN 1931-0145. . URL `http://doi.acm.org/10.1145/1345448.1345470`.

Yung-Ming Li and Jhih-Hua Jhang-Li. Pricing display ads and contextual ads: Competition, acquisition, and investment. *Electron. Commer. Rec. Appl.*, 8(1):16–27, January 2009. ISSN 1567-4223. . URL `http://dx.doi.org/10.1016/j.elerap.2008.06.001`.

Kangmiao Liu, Quang Qiu, Can Wang, Jiajun Bu, Feng Zhang, and Chun Chen. Incorporate sentiment analysis in contextual advertising. In *Proceedings of the workshop on Targeting and Ranking for Online Advertising at WWW*, TROA, Beijing, China, 2008.

Tie-Yan Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331, March 2009. ISSN 1554-0669. . URL `http://dx.doi.org/10.1561/1500000016`.

Yandong Liu, Sandeep Pandey, Deepak Agarwal, and Vanja Josifovski. Finding the right consumer: optimizing for conversion in display advertising campaigns. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 473–482, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-0747-5. . URL `http://doi.acm.org/10.1145/2124295.2124353`.

Hong Lu, Wei Pan, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. Soundsense: Scalable sound sensing for people-centric applications on mobile phones. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, MobiSys '09, pages 165–178, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-566-6. . URL `http://doi.acm.org/10.1145/1555816.1555834`.

H. Brendan McMahan, Gary Holt, D. Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. Ad click prediction: A view from the trenches. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 1222–1230, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2174-7. . URL http://doi.acm.org/10.1145/2487575.2488200.

Aditya Krishna Menon, Krishna-Prasad Chitrapura, Sachin Garg, Deepak Agarwal, and Nagaraj Kota. Response prediction using collaborative filtering with hierarchies and side-information. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 141–149, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0813-7. . URL http://doi.acm.org/10.1145/2020408.2020436.

Rada Mihalcea and Andras Csomai. Wikify!: linking documents to encyclopedic knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, CIKM '07, pages 233–242, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-803-9. . URL http://dx.doi.org/10.1145/1321440.1321475.

David Milne and Ian H. Witten. An effective, low-cost measure of semantic relatedness obtained from wikipedia links. In *In Proceedings of AAAI 2008*, 2008a.

David Milne and Ian H. Witten. Learning to link with wikipedia. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, pages 509–518, New York, NY, USA, 2008b. ACM. ISBN 978-1-59593-991-3. . URL http://dx.doi.org/10.1145/1458082.1458150.

Emiliano Miluzzo, Cory T. Cornelius, Ashwin Ramaswamy, Tanzeem Choudhury, Zhigang Liu, and Andrew T. Campbell. Darwin phones: The evolution of sensing and inference on mobile phones. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, pages 5–20, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-985-5. . URL http://doi.acm.org/10.1145/1814433.1814437.

Mark Montague and Javed A. Aslam. Relevance score normalization for metasearch. In *Proceedings of the tenth international conference on Information and knowledge management*, CIKM '01, pages 427–433. ACM, 2001. ISBN 1-58113-436-3. . URL http://doi.acm.org/10.1145/502585.502657.

Abhirup Nath, Shibnath Mukherjee, Prateek Jain, Navin Goyal, and Srivatsan Laxman. Ad impression forecasting for sponsored search. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, pages 943–952, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2035-1. URL http://dl.acm.org/citation.cfm?id=2488388.2488470.

Richard J. Oentaryo, Ee-Peng Lim, Jia-Wei Low, David Lo, and Michael Finegold. Predicting response in mobile advertising with hierarchical importance-aware factorization machine. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 123–132, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2351-2. . URL http://doi.acm.org/10.1145/2556195.2556240.

OneUpWeb. *Relevance Feedback in Information Retrieval*. 2005. URL http://www.oneupweb.com/landing/keywordstudylanding.htm.

Larry Page, Sergey Brin, R. Motwani, and T. Winograd. *The PageRank Citation Ranking: Bringing Order to the Web*, 1998.

Vivek Pandey. How do ad exchanges and real-time bidding platforms work? *http://www.quora.com/How-do-ad-exchanges-and-real-time-bidding-platforms-work/answer/Vivek-Pandey-11?srid=3eDR&share=1*, 2013.

Christos H. Papadimitriou, Hisao Tamaki, Prabhakar Raghavan, and Santosh Vempala. Latent semantic indexing: A probabilistic analysis. In *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, PODS '98, pages 159–168, New York, NY, USA, 1998. ACM. ISBN 0-89791-996-3. . URL http://doi.acm.org/10.1145/275487.275505.

Panagiotis Papadimitriou, Hector Garcia-Molina, Prabhakar Krishnamurthy, Randall A. Lewis, and David H. Reiley. Display advertising impact: search lift and social influence. In Chid AptÃĨ, Joydeep Ghosh, and Padhraic Smyth, editors, *KDD*, pages 1019–1027. ACM, 2011. ISBN 978-1-4503-0813-7. URL http://dblp.uni-trier.de/db/conf/kdd/kdd2011.html#PapadimitriouGKLR11.

Ankit Patil, Kushal Dave, and Vasudeva Varma. Leveraging latent concepts for retrieving relevant ads for short text. In *Proceedings of the 35th European conference on Advances in Information Retrieval*, ECIR'13, pages 780–783, Berlin, Heidelberg, 2013. Springer-Verlag. ISBN 978-3-642-36972-8.

Xuan-Hieu Phan, Le-Minh Nguyen, and Susumu Horiguchi. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 91–100, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2. . URL `http://doi.acm.org/10.1145/1367497.1367510`.

Furcy Pin and Peter Key. Stochastic variability in sponsored search auctions: observations and models. In Yoav Shoham, Yan Chen, and Tim Roughgarden, editors, *ACM Conference on Electronic Commerce*, pages 61–70. ACM, 2011. ISBN 978-1-4503-0261-6. URL `http://dblp.uni-trier.de/db/conf/sigecom/sigecom2011.html#PinK11`.

Xiaojun Quan, Gang Liu, Zhi Lu, Xingliang Ni, and Liu Wenyin. Short text similarity based on probabilistic topics. *Knowledge and Information Systems*, 25(3), December 2010. ISSN 0219-1377. . URL `http://dx.doi.org/10.1007/s10115-009-0250-y`.

Filip Radlinski, Andrei Broder, Peter Ciccolo, Evgeniy Gabrilovich, Vanja Josifovski, and Lance Riedel. Optimizing relevance and revenue in ad search: a query substitution approach. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 403–410, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-164-4. . URL `http://doi.acm.org/10.1145/1390334.1390404`.

Hema Raghavan and Dustin Hillard. A relevance model based filter for improving ad quality. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 762–763, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. .

Hema Raghavan and Rukmini Iyer. Probabilistic first pass retrieval for search advertising: from theory to practice. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 1019–1028, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0099-5.

Colin R. Reeves, editor. *Modern Heuristic Techniques for Combinatorial Problems*. John Wiley & Sons, Inc., New York, NY, USA, 1993. ISBN 0-470-22079-1.

Moira Regelson and Daniel C. Fain. Predicting click-through rate using keyword clusters. In *Electronic Commerce (EC)*. ACM, 2006.

Berthier Ribeiro Neto, Marco Cristo, Paulo B. Golgher, and Edleno Silva de Moura. Impedance coupling in content-targeted advertising. SIGIR '05, pages 496–503. ACM, 2005. ISBN 1-59593-034-5. . URL `http://doi.acm.org/10.1145/1076034.1076119`.

Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: estimating the click-through rate for new ads. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 521–530, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-654-7. .

S. E. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pages 232–241, New York, NY, USA, 1994. Springer-Verlag New York, Inc. ISBN 0-387-19889-X. URL `http://dl.acm.org/citation.cfm?id=188490.188561`.

S.E. Robertson, S. Walker, S. Jones, M.M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. pages 109–126, 1996.

J. Rocchio. *Relevance Feedback in Information Retrieval*, pages 313–323. 1971. URL `http://scholar.google.com/scholar?hl=en38;client=firefox-a38;q=relevance+feedback+in+information+retrievalbtnG=Search`.

G. Salton and C. S. Yang. On the specification of term values in automatic indexing. *Department of Computer Science, Cornell University, Ithaca, New York*, 28(1):73–173, 1973.

G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620, November 1975. ISSN 0001-0782. . URL `http://doi.acm.org/10.1145/361219.361220`.

Mark Sanderson and W. Bruce Croft. The history of information retrieval research. *Proceedings of the IEEE*, 100(Centennial-Issue):1444–1451, 2012. . URL `http://dx.doi.org/10.1109/JPROC.2012.2189916`.

D. Sculley, Robert G. Malkin, Sugato Basu, and Roberto J. Bayardo. Predicting bounce rates in sponsored search advertisements. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 1325–1334, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-495-9. . URL `http://doi.acm.org/10.1145/1557019.1557161`.

D. Sculley, Matthew Eric Otey, Michael Pohl, Bridget Spitznagel, John Hainsworth, and Yunkai Zhou. Detecting adversarial advertisements in the wild. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 274–282, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0813-7. . URL http://doi.acm.org/10.1145/2020408.2020455.

Benyah Shaparenko, Özgür Çetin, and Rukmini Iyer. Data-driven text features for sponsored search click prediction. In *ADKDD '09: Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising*, pages 46–54, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-671-7. .

Dou Shen, Zheng Chen, Qiang Yang, Hua-Jun Zeng, Benyu Zhang, Yuchang Lu, and Wei-Ying Ma. Web-page classification through summarization. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 242–249, New York, NY, USA, 2004. ACM. ISBN 1-58113-881-4. . URL http://doi.acm.org/10.1145/1008992.1009035.

Dou Shen, Rong Pan, Jian T. Sun, Jeffrey J. Pan, Kangheng Wu, Jie Yin, and Qiang Yang. Q2C@UST: our winning solution to query classification in KDDCUP 2005. *SIGKDD Explor. Newsl.*, 7(2):100–110, December 2005. ISSN 1931-0145. . URL http://dx.doi.org/10.1145/1117454.1117467.

Dou Shen, Arun C. Surendran, and Ying Li. Report on the second kdd workshop on data mining for advertising. *SIGKDD Explor. Newsl.*, 10(2): 47–50, December 2008. ISSN 1931-0145. . URL http://doi.acm.org/10.1145/1540276.1540291.

Fuyuan Shen. Banner advertisement pricing, measurement, and pretesting practices: Perspectives from interactive agencies. *Journal of Advertising*, 31(3):59–67, 2002. . URL http://dx.doi.org/10.1080/00913367.2002.10673676.

Libin Shen and Aravind Joshi. Ranking and Reranking with Perceptron. *Machine Learning*, 60(1-3):73–96, 2005. . URL http://dx.doi.org/10.1007/s10994-005-0918-9.

Arun Sundararajana Sinan Aral, Lev Muchnika. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. 2009.

Lara Sinclair. Click fraud rampant in online ads, says bing. *http://www.theaustralian.com.au/media/click-fraud-rampant-in-online-ads-says-bing/story-e6frg996-1226056349034?nk=8da0e2fdd40d687781c53af5de0bfdfd*, 2011.

Parag Singla and Matthew Richardson. Yes, there is a correlation: - from social networks to personal behavior on the web. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 655–664. ACM, 2008. ISBN 978-1-60558-085-2. .

Alexander Smola and Shravan Narayanamurthy. An architecture for parallel topic models. *Proc. VLDB Endow.*, pages 703–710, September 2010. ISSN 2150-8097. URL `http://dl.acm.org/citation.cfm?id=1920841.1920931`.

Karen Spärck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21, 1972.

Amanda Spink, Dietmar Wolfram, Major B. J. Jansen, and Tefko Saracevic. Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3):226–234, 2001. ISSN 1532-2890. . URL `http://dx.doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1591>3.0.CO;2-R`.

Valentin I. Spitkovsky and Angel X. Chang. A Cross-Lingual Dictionary for English Wikipedia Concepts. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, 2012.

Yukihiro Tagami, Shingo Ono, Koji Yamamoto, Koji Tsukamoto, and Akira Tajima. Ctr prediction for contextual advertising: Learning-to-rank approach. In *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising*, ADKDD '13, pages 4:1–4:8, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2323-9. . URL `http://doi.acm.org/10.1145/2501040.2501978`.

Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 807–816. ACM, 2009. ISBN 978-1-60558-495-9. .

W. R. Thompson. On the Likelihood that one Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika*, 25:285–294, 1933.

Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas. Adnostic: Privacy preserving targeted advertising. In *NDSS*. The Internet Society, 2010. URL `http://dblp.uni-trier.de/db/conf/ndss/ndss2010.html#ToubianaNBNB10`.

Melody M. Tsang, Shu-Chun Ho, and Ting-Peng Liang. Consumer attitudes toward mobile advertising: An empirical study. *Int. J. Electron. Commerce*, 8(3):65–78, April 2004. ISSN 1086-4415. URL `http://dl.acm.org/citation.cfm?id=1278124.1278129`.

Peter D. Turney. Learning algorithms for keyphrase extraction. *Inf. Retr.*, 2 (4):303–336, 2000. ISSN 1386-4564. .

Alexander Tuzhilin. Click fraud rampant in online ads, says bing. *http://googleblog.blogspot.in/pdf/Tuzhilin_Report.pdf*, 2008.

Vivek Vaidya. Cookie synching. *http://www.admonsters.com/blog/cookie-synching*, 2010.

Chingning Wang, Ping Zhang, Risook Choi, and Michael D'Eredita. Understanding consumers attitude toward advertising. In *In: Eighth Americas Conference on Information Systems. (2002)*, pages 1143–1148, 2002.

Haofen Wang, Yan Liang, Linyun Fu, Gui-Rong Xue, and Yong Yu. Efficient query expansion for advertisement search. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 51–58, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. . URL `http://doi.acm.org/10.1145/1571941.1571953`.

Jun Wang and Shuai Yuan. Real-time bidding: A new frontier of computational advertising research. In *CIKM 2013: Proceeding of the 23rd ACM conference on Information and knowledge management*, New York, NY, USA, 2013. ACM.

Xing Wei and W. Bruce Croft. Lda-based document models for ad-hoc retrieval. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '06, pages 178–185, New York, NY, USA, 2006. ACM. ISBN 1-59593-369-7. . URL `http://doi.acm.org/10.1145/1148170.1148204`.

Jun Wang Weinan Zhang, Shuai Yuan. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, New York, NY, USA, 2014. ACM. .

Sholom Weiss. Lightweight rule induction. Number 6523020, February 2003. URL `http://www.freepatentsonline.com/6523020.html`.

Dominic Widdows and Trevor Cohen. The semantic vectors package: New algorithms and public tools for distributional semantics. In *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing*, ICSC '10, pages 9–15, Washington, DC, USA, 2010. IEEE Computer Society. ISBN 978-0-7695-4154-9. . URL `http://dx.doi.org/10.1109/ICSC.2010.94`.

Xiaoyuan Wu and Alvaro Bolivar. Keyword extraction for contextual advertisement. In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 1195–1196, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2. . URL `http://doi.acm.org/10.1145/1367497.1367723`.

Jun Xu and Hang Li. Adarank: A boosting algorithm for information retrieval. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 391–398, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-597-7. . URL `http://doi.acm.org/10.1145/1277741.1277809`.

Wanhong Xu, Eren Manavoglu, and Erick Cantu-Paz. Temporal click model for sponsored search. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 106–113, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0153-4. . URL `http://doi.acm.org/10.1145/1835449.1835470`.

Jun Yan, Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, and Zheng Chen. How much can behavioral targeting help online advertising? In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 261–270, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-487-4. .

Y. YI. *COGNITIVE AND AFFECTIVE PRIIMING EFFECTS OF THE CONTEXT FOR PRINT ADVERTISEMENTS*. 1990. URL `http://books.google.co.in/books?id=2OiyVLxENhYC`.

Wen tau Yih, Joshua Goodman, and Vitor R. Carvalho. Finding advertising keywords on web pages. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 213–222, New York, NY, USA, 2006. ACM. ISBN 1-59593-323-9. .

Dawei Yin, Shike Mei, Bin Cao, Jian-Tao Sun, and Brian D. Davison. Exploiting contextual factors for click modeling in sponsored search. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 113–122, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2351-2. . URL `http://doi.acm.org/10.1145/2556195.2556237`.

Kai Yu, Jinbo Bi, and Volker Tresp. Active learning via transductive experimental design. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 1081–1088, New York, NY, USA, 2006. ACM. ISBN 1-59593-383-2. . URL `http://doi.acm.org/10.1145/1143844.1143980`.

Xiaohui Yu and Huxia Shi. Query segmentation using conditional random fields. In *Proceedings of the First International Workshop on Keyword Search on Structured Data*, KEYS '09, pages 21–26, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-570-3. . URL `http://doi.acm.org/10.1145/1557670.1557680`.

Shuai Yuan, Jun Wang, and Xiaoxue Zhao. Real-time bidding for online advertising: Measurement and analysis. In *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising*, ADKDD '13, pages 3:1–3:8, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2323-9. . URL `http://doi.acm.org/10.1145/2501040.2501980`.

Wei Vivian Zhang, Xiaofei He, Benjamin Rey, and Rosie Jones. Query rewriting using active learning for sponsored search. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 853–854, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-597-7. . URL `http://doi.acm.org/10.1145/1277741.1277942`.