Digital Advertising: An Information Scientist's Perspective

James G. Shanahan[1] , Goutham Kurra[2]

[1] Church and Duncan Group, 541 Duncan Street, San Francisco, CA 94131, USA
[2]Turn Inc., 835 Main St, Redwood City, CA 94063, USA
James.Shanahan@gmail.com, Goutham.Kurra@turn.com

Abstract. Digital online advertising is a form of promotion that uses the Internet and World Wide Web for the express purpose of delivering marketing messages to attract customers. Examples of online advertising include text ads that appear on search engine results pages, banner ads, in-text ads, or Rich Media ads that appear on regular web pages, portals, or applications. Over the past 15 years online advertising, a $65 billion industry worldwide in 2009, has been pivotal to the success of the World Wide Web. That being said, the field of advertising has been equally revolutionized by the Internet, World Wide Web, and more recently, by the emergence of the social web, and mobile devices. This success has arisen largely from the transformation of the advertising industry from a low-tech, human intensive, "Mad Men" way of doing work to highly optimized, quantitative, mathematical, computer- and data-centric processes that enable highly targeted, personalized, performance based advertising. This chapter provides a clear and detailed overview of the technologies and business models that are transforming the field of online advertising primarily from statistical machine learning and information science perspectives.

Keywords: Online advertising, learning to rank ads, machine learning, sponsored search, contextual advertising, display advertising, behavioral targeting, ad networks, ad exchanges, online learning.

1 Introduction

Online advertising is a form of promotion that uses the Internet and World Wide Web for the express purpose of delivering marketing messages to attract customers. Since its fledgling beginning in 1994, online advertising has become a $65 billion dollar industry worldwide (in 2009) resulting in a double digit annual growth on average. It makes up almost 10% of the overall spending on advertising (across all media types such TV, radio, press, outdoor etc.). This success has arisen largely from the transformation of the advertising industry from a low-tech, human intensive, "Mad

Men"[1] way of doing work, that was common place for much of the 20th century and the early days of online advertising, to highly optimized, quantitative, mathematical, computer and data-centric processes (some of which have been adapted from Wall Street) that form the backbone of many current online advertising systems. More concretely, a modern day publisher and advertising system include modules for targeting, pricing, prediction, forecasting, and large scale storage and analytics. These components build on ideas from information retrieval, machine learning, statistics, economic models, operations research, and distributed systems. This chapter focuses primarily on the information science aspects of online advertising. This is supplemented with background material on most aspects of a modern day online advertising system. Aspects of media planning and scheduling that mainly draw from operations research are beyond the scope of this chapter.

This chapter is structured as follows: Section 2 summarizes the key business concepts of online advertising; Section 3 presents sponsored search, reviewing key aspects of organic and sponsored search which highlighting active areas of research; Section 4 follows a similar structure focusing primarily on contextual advertising and display advertising; Section 5 reviews briefly auction models; Section 6 overviews new directions and issues, while Section 7 concludes the chapter.

2 Online Advertising Background

Traditionally, online advertising has been a formal relationship between the advertiser and the publisher. Each party in this relationship has a different objective. Advertisers want to convey a message to the consumer about a product or service to convince them to purchase or use that product or service; the more consumers that sign up the more revenue the advertiser makes. Often, advertisers are equally interested in latent effects, such as a positive branding experience. In short, advertisers see advertising as an investment for the growth of their sales and their brand, and wish to maximize their return on investment. On the other hand, publishers wish to generate revenue and, ultimately, a profit, be it from a news report written by a professional journalist, or from a blog entry or video created for free by a member of the public. This revenue can be offset against costs such as reporter salaries or online publishing fees and to potentially generating profits for the owners.

2.1 Purchase Funnel

From a marketing perspective, online advertising can be crudely categorized based on the primary objective of the advertising campaign: branding or direct marketing. This is more naturally framed in the field of marketing within the *advertising funnel* (also known as a *purchase funnel*). This funnel is divided into a sequence of phases organized around the following marketing objectives (ordered chronologically within

[1] "Mad men", as an expression, was coined in the late 1950s and refers to the people working on Madison Avenue, New York City in the advertising industry. It is also the name of a US AMC TV series that was first broadcast in 2007.

a marketing campaign): category awareness, brand awareness, brand consideration, brand preference, purchase intent, purchase, customer retention, and customer advocacy. A marketing manager will select the parameters of an advertising campaign depending on the stage of marketing. These parameters will include the format of the ad (text, graphic, video), the message (purely informational or a call to action like a purchase or signup), and the desired reach (the reach refers to the total number of different people (unique users) or households exposed, at least once, to a medium during a given period of time). For example, for a product launch, a marketing manager may want to create product appeal through a broad reach ad campaign; this could be accomplished by reaching millions of people (with adequate frequency) through display advertising. Here, all online media sources can be leveraged, including contextual advertising around web pages (such as blogs and newswire), or online video, and sponsored search. Moving down through the funnel, ad formats such as rich media and online video can use their storytelling power to build favor toward the brand. Likewise, online sponsorships of unique content or events can push the middle funnel metrics. In addition, the advertiser website (specifically the landing page) will likely do great things for branding, provided the advertiser can get people there; search plays a critical role in driving traffic (be it organic search or sponsored search). The next stage in the marketing funnel is for the advertiser's site to do one or more of the following — close the sale there and then, push visitors to a retail dealer, or acquire an email address to continue the digital dialogue.

2.2 Types of Advertising

The advertiser message is generally embedded within an ad creative (be it text or graphic), which is subsequently embedded within a webpage that a consumer views via their web browser or application program. This page is commonly referred as the target page. Ad creatives are commonly hyperlinked to a landing page – a page on the advertiser's website that provides more details on the product or services advertised in the ad creative and how to obtain them. The landing page is rendered when the viewer clicks on the ad creative. Online creatives come in a huge variety of sizes and formats, ranging from text ads, to medium rectangle graphic ads (300 pixels wide by 250 pixels high), to skyscraper banners (728 pixels by 90 pixels), and from embedded ads (graphic, video) to popups to overlays and interstitials. In the US, the Interactive Advertising Bureau (IAB), comprised of more than 460 leading media and technology companies who are responsible for selling 86% of online advertising in the United States, works with its member companies, evaluates and recommends standards and practices and fields critical research on interactive advertising [1].

Embedding text ads within a search engine's results page (SERP) is referred to as *sponsored search*, while embedding text ads within publisher's online media such as a web pages, portals and online/mobile/offline applications is commonly referred to as *contextual advertising.* Embedding graphical/video ads within web pages or apps is referred to as *display advertising*. Other categories of advertising that take the form of text-ads or graphical ads include: classifieds and auctions (on newspaper sites, for example); local search; e-mail-based advertising; and sponsorship.

Other Types of Advertising. There are many other types of advertising that are summarized here. One of the bigger but less known categories is that of classifieds and auctions; this refers to advertisers who pay Internet companies to list specific products or services (e.g., online job boards and employment listings, real estate listings, automotive listings, auction-based listings, yellow pages). A good example is YellowPages.com, who provides a local search engine, where users can submit geographically constrained searches against a structured database of local business listings. Like sponsored, search some of the results page is sponsored, consisting of business/service listings that are paid for by the advertiser on a per impression-basis, or on a pay per transaction basis such as pay-per-call-basis (if a consumer calls a dedicated phone number then AT&T gets paid a negotiated fee in the case of YP.com). Other forms of advertising include lead generation, and email-based advertising.

2.3 Payment Models

From a business model perspective the field has adapted the traditional offline cost model (CPM), and also developed other custom built models for the online field. The CPM business model, *cost per mille* (corresponding to a thousand impressions), formed the core business model in the early days of online banner advertising (being directly adapted from traditional offline media such as newspaper). In this model, advertisers negotiate (or bid) a rate that will be paid for every thousand times the ad is displayed. This model is most often used when the placement of the ads (the sites and audience) is predetermined and when the value of the impressions in fairly uniform and known. The CPM model imposes a large return on investment (ROI) risk on the advertiser as the ad impressions may not result in user engagement or purchases. The CPC model (*cost per click*) also known as PPC (pay per click) was first rolled out in 1997 by Goto.com (later acquired by Overture which was subsequently acquired by Yahoo!); CPC was originally created for advertising on search engine result pages. Google tweaked the CPC model in 2002 adding in an ad quality component [2]. CPC reduces the risk to the advertiser in that the advertiser doesn't pay for impressions that are not creating value. On the other hand, the publisher or intermediary now shares some of the risk. By using clicks as a proxy for user interest and engagement with the ads, the advertiser has some form of measuring and controlling the success of the campaign. However, not all clicks are valuable to advertisers as not all clicks convert to useful actions such as purchases. Furthermore, fraudulent clicks are easy to generate and pose an ROI risk to the advertiser. In the CPA model, *cost per action*, (sometimes known as Pay Per Action or PPA), the advertiser pays for each specified action (e.g., a purchase of a product, a lead form submission for a loan) linked to the advertisement. Since an action is usually the very thing that generates revenue for the advertiser, this is similar to sharing revenue with the publisher or intermediary. Actions can be defined as post-click or post-view and may have different values depending on the type. The CPA model is most common with Ad Networks which are intermediaries between publishers and advertisers. While CPA removes most of the advertiser's risk, it transfers all the business risk to the intermediary ad network or the publisher since users may not actually buy the advertiser's services or products despite a lot of advertising. Hybrids of these models have also been constructed such as dCPM (dynamic CPM), where dCPM pricing optimizes an ad towards the users, sites and site sections that perform best for the advertiser, dynamically paying the most efficient CPM for the value of the inventory to the advertiser. dCPM ads are driven by two parameters, a maximum average CPM and a performance goal such as a CPA target. dCPM allows the risk of a campaign to be shared between the advertiser and the intermediary, typically, an ad network or demand-side platform (DSP) or ad exchange (defined subsequently).

2.4 Market Numbers and Trends

The field of online advertising was about $65 billion worldwide in 2009. This is broken down regionally as follows: Western Europe ad spending was $18 billion US Dollars (€14.7 billion Euros) in 2009 or 27% of global online ad spending; the US revenue was $22.7 billion; Latin America ad revenue was $2 billion; China had a similar revenue of $2 billion; $2 billion in Latin America; and Russia accounted for $720 million [3]. Looking more closely at these numbers trends are starting to emerge based on 2009 revenue numbers in the United States by the IAB 2009 [4], which are largely echoed worldwide:

- Sponsored search accounts for 47 percent of 2009 full year revenues ($10.7 billion);
- Display-related advertising revenues totaled $8.0 billion;
- Classifieds revenues (products such YellowPages.com) accounted for 10 percent or $2.3 billion;

3 Sponsored Search-based Advertising

Abstractly, online advertising can be viewed as a supply and demand market economy. Demand in this context refers to how much (quantity) of ad slot inventory is desired by advertisers. Supply represents how much the market can offer. The quantity supplied refers to the amount of a certain ad inventory publishers are willing to supply when receiving a certain price. In this model, advertisers are demand-side entities and publishers are supply-side entities. In the early days the relationship between advertiser and publisher was direct and driven by large sales teams – it was inefficient and largely informal (by today's online advertising standards at least). This inefficient advertising market place left a lot to be desired by both the advertiser and the publisher. As a result lots of new models and marketplaces such as ad networks, ad exchanges, demand side trading platforms have been developed over the last four or five years (some of these marketplaces will be discussed below).

3.1 Sponsored Search Overview

Sponsored search, though one of the most innovative forms of advertising on the Internet, turns out to be one of most easily understood forms of online adverting. Here ads are embedded within the search engine results page (SERP), generally on the North and East ports of the SERP (corresponding to the top and right hand side of the SERP). Sponsored search is dominated by the CPC model. It is a direct relationship between the search engine and the advertiser or the advertiser agency. From a user's perspective, a search engine serves one key need: that of trying to satisfy a searcher's information need which is expressed via the searcher's query. This is accomplished by providing a ranked list of organic web pages and documents of various types (maps, videos, etc), sometimes known as hits, (where each page is presented in terms of a title, snippet and corresponding URL) and a ranked list of sponsored results presented in the north and east sections of the SERP. Within the online advertiser world, the search engine plays the role of the publisher and the auctioneer. As auctioneer, the search engine gets to select which ads can participate in the auction, the ranking of these ads and the cost of a click (should the display of an ad result in a click). Viewed this way, the SERP can be simply broken into the organic set of results and the sponsored set of results generated by the organic search engine and the sponsored search engine respectively.

3.2 Organic Search Engine

An organic web search engine is designed to search for information on the World Wide Web servers. A modern day organic search engine can be decomposed into the following core modules: crawler, webpage repository, document index, query processor, document ranker, logging, analytics, SERP generator and administrators console. Figure 1 presents these components and their relationship to each other. For completeness, a brief summary of these components is provided as many of them will be used in other online advertising ecosystems. A search engine operates in the following order: 1. Web crawling; 2. Indexing; 3. Ranking documents for user search queries; 4 SERP generation.

Crawler. One of the core components of Web search engines is the ability to retrieve and store the billions of pages available across the World Wide Web. The crawler sometimes, also known as a spider, is primarily responsible for retrieving these pages, and storing them locally. A Web crawler is a computer program (or agent) that browses the World Wide Web in an automated manner. Web crawlers generally store a copy of all the visited pages for later processing by the featurization process and by the indexing process (that encodes the downloaded pages in an efficient data structure to provide fast retrieval). In general, a crawler is provided with an initial list of URLs to visit – called the seeds. As the crawler visits these URLs, it identifies all the hyperlinks in visited pages and adds them to the list of URLs to visit – called the crawl frontier. URLs from the frontier are recursively visited according to a set of crawling policies. Exclusions can be made by the use of robots.txt. Policies include how deep to crawl, limits on how many pages are crawled from a site, and the order in which a URLs are visited.

Inverted Index and Scoring Process. The crawled web pages need to be stored in a manner that is computationally efficient (SERPs need to be generated in response to queries within 100s of milliseconds) and with minimum storage requirements (say 10-20% of the original raw page size). To enable this, web pages are stored in an inverted index data structure; this structure exploits the sparse nature of web documents. Within this framework, it is common to represent Web pages in terms of the words, or phrases that occur in the titles, headings, body content, or other special HTML fields called meta-tags that occur in the raw HTML page. These words or phrases can be more generally referred to as features or characteristics of the webpage. An inverted index is an index data structure storing a key-value mapping from features, such as words or numbers, to their locations in a document or a set of documents. An everyday example of an inverted index is the index at the back any text book, where the index words are the keys, and the associated values are a list of pages where the corresponding word(s) occur. In web search the purpose of an inverted index is to allow fast full text searches, at a cost of increased processing when a document is added to the database. Within the context of web search and, more broadly, digital advertising, one can generalize the use of inverted indexes beyond that of word-based keys. The key type in the inverted index mapping can be extended to any feature, f, that can be used to characterize a web page p (e.g., category of webpage, a word that occurs on the webpage, a postal code corresponding to the geographic area that had the highest density of visitors to this webpage in the past 3 days). The value associated with each key is generally referred to as the postings list. In this case, it could be a linked list of posting records. A posting could be simply a document identifier denoting that this feature occurred or is associated with this document and a corresponding payload. The posting payload could be a number corresponding to the number of times a word occurred in a document, (sometimes known as term frequency (TF) of a term in the document) and a list of numbers corresponding to the position of each word occurrence within a document. A posting is formally represented as tuple $<ID, W>$, where ID is the page identifier (generally ranges from 1 to the number of unique pages N), and W, which is assumed for the purposes of ease of presentation to be limited to a number, known as a weight; e.g., this weight could be the term frequency of f within p. For efficiency of memory usage and query scoring time, each posting list is sorted in increasing order of *page ID*. As a result, there are numerous ways to encode postings lists to reduce memory requirements [5]. Generally, since billions of documents are being indexed, the documents are divided into non-overlapping subsets where each subset is indexed separately on its own server; at query time responses from each server need to be merged.

Features. The featurization step in Figure 1 is concerned with representing each page in terms of features that summarize this page and distinguish it from other web pages. These features can be purely based on the textual content of the page (such as words and phrases). Values for these features can be simply frequency counts or normalized frequencies (frequency of word / total number of words). In addition, features can be based on the web graph of hyperlinked web pages such as PageRank [6] where the feature value corresponds to a probability that a web browser (user) who follows links randomly (with teleportation capabilities) for a long time will be at a particular page if polled randomly. This probability can provide a query-independent measure of the popularity (and quality) of the page. Another family of features could be behavioral/historical features, such as the number of times a page has been clicked on from a SERP. Many other features families are commonly used. In summary, an inverted index is used extensively within large web-scale systems such as organic web search, ad search, and in other sparse data problems such as user behavior problems.

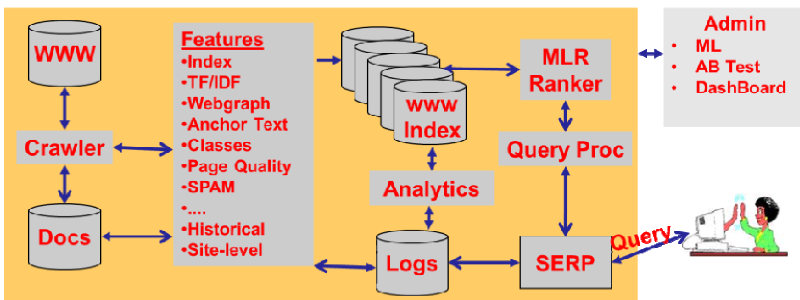

Fig 1: Core components of a web search engine.

Ranking Documents for user search queries. Web search engine index sizes are in the order of tens of billions of web pages of pages. For example, the website, http://www.worldwidewebsize.com/, estimates Google's index to be around 30 billion pages. Typically, search engines need to process tens of thousands of queries per second (QPS). This is generally a distributed, multistep process consisting of the following steps: query parsing, query featurization, scoring pages in the index, sorting them based on score, and generating a results page. Query parsing is concerned with splitting the query into words (tokenizing) and assigning a role to each word or sequence of words such as determining if a word is a proper noun, a determiner, a preposition, or a part of a phrase. This can be accomplished using part of speech taggers, lexicon lookup approaches, or Markov models (HMMs or CRFs [7]). Query featurization includes the features output by query parsing and possibly many other features. For example, the zip code from where the user typed the query, or the day of week, etc. At its simplest, this feature vector could just be a list of query words, as is the case in traditional information retrieval, or could be an augmented feature vector as briefly described above. Subsequently this query feature vector is taken and a join is performed over the pre-sorted feature posting lists in the index. One of the common approaches to performing this multi-way join is to use a document-at-a-time (DAAT) evaluation. Most of the DAAT algorithms in the literature are merge-joins with improved efficiency [8]. To evaluate a query using a DAAT algorithm, a cursor *Cf* is created for each feature f in the query vector, and is used to access *f's* posting list. During the join, the cursors are moved in a coordinated fashion forward over the posting lists. While traversing the posting list in this fashion one accumulates documents that have all the query features. One can use these features to assign a relevance score (sometimes interpreted as a similarity) to each of the accumulated documents. This similarity score between page *p* and query vector *q*, can be as simple as summing contributions over all common features of the query vector and the document vector, and over all features that are local to the query and webpage as follows (this is commonly known as a dot product):

$$\text{Score(q, p)} = \sum_{f \in q \cap p} \mathrm{w_q} * w_p + \sum_{f \in qo} 1 * w_q + \sum_{f \in po} 1 * w_p \quad \textbf{(1)}$$

Where the weights w_p and w_q measure the importance of the feature *f* on the query side and on the page side respectively, and *qo* and *po* denote features on the query side only and page side only.

In reality, scoring each document as described above is impractical (due to millisecond time constraints on query turnaround time). As a result, various approaches have been proposed to address this constraint resulting, typically, in a multi-phase retrieval process. This multi-phase process can consist of two or three phases, where each phase can be functionally similar to the scoring process described above with the later phases consisting of more features (with some of these features being dependent on the distribution of the results set to-date). It is common to use a two-phase process using an operator such as the WAND operator in the first phase to improve first phase efficiency [9].

SERP Generation. The organic SERP consists of a list of snippets of text extracted from each of the resultant web pages. This snippet serves the same role as an ad creative and needs to be carefully crafted (automatically).

Evaluation. Relevance is core to evaluation in information retrieval. Here, relevance typically denotes how well a retrieved document or set of documents meets the information need of the user. *Relevance* is often viewed as multifaceted. One core facet of relevance relates to *topical* relevance or *aboutness* [10], i.e., to what extent the topic of a result matches the topic of the query or information need. Another facet of relevance is based on user perception, sometimes referred to as *user* relevance; it encompasses *topical* relevance and possibly other concerns of the user such as timeliness, authority or novelty of the result. Depending on the query yet another facet of relevance comes into play that of geographical aboutness, .i.e., to what extent the location of a result, say in the case of a business listing, matches the location of the query or information need. Performance of systems is generally captured by relevance-centered metrics such as discounted cumulative gain [11] or precision at specific ranks (say up rank 5) [12]. For example, DCG is defined here up until rank *P* as follows:

$$DCG@P = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2 i} \quad \textbf{(2)}$$

The constants rel_i are the relevance scores generally provided by human assessors. Human assessors typically judge documents with respect a query on an ordinal scale, with labels such as “Perfect”, “Excellent”, “Good”, “Fair”, and “Bad”. These are then mapped to a numeric scale. Though relevance is generally provided by a team of professional assessors, it is sometimes obtained via crowd sourcing services such as the Amazon Mechanical Turk [13].

Machine Learnt Ranking Functions. Recent work on applying machine learning approaches to optimizing ranking functions has led to huge gains over hand-tuned expert systems. The field is commonly known as learning to rank or machine-learned ranking (MLR) and is a type of supervised or semi-supervised machine learning problem in which the goal is to automatically construct a ranking function from training data. Training data generally consists of query, webpage, and associated relevance triples <*q, p, relevance*>, with some partial order specified between web pages associated with the same query. This score or label may be provided manually by human assessors (or editors), who judge the relevance for each selected <*query, page*> pair. It is not feasible to check relevance of all query document pairs, so typically a technique called top-k pooling is used — only top few documents, retrieved by some existing ranking models are checked [14]. Alternatively, training data labels may be derived automatically by analyzing click-through logs (i.e. search results which received clicks from users) [15]. Training data is used by a learning algorithm to produce a ranking model, which can then be used to compute a relevance score for a document with respect to a given query. MLR problems can be reformulated as optimization problems with respect to evaluation metrics (such as DCG). This has lead to the development of a plethora of very powerful pair-wise and list-wise learning algorithms [16].

3.3 Sponsored Search Engine

Sponsored search, unlike organic search, is concerned with selecting and presenting text ads in response to a searcher’s query with the express goal of satisfying the searcher’s information need while in parallel delivering the advertiser’s message; these text ads are embedded along with organic search results on the SERP. This can be viewed as “*not seeing advertising as entertainment or an art form, but as a medium of information…*“ [17].

Sponsored search system strongly parallels the functionality of organic web search but differs in subtle ways that make the automatic targeting of ads much more challenging. One could frame the problem of targeting ads as a web search (IR) problem. Here the organic user query, *q*, and user, *u*, become the query vector (as is the case in web search) but the index differs to that in web search; it is created from features extracted from the ad creative, the keywords associated with the ad (for targeting), and sometimes the text of the corresponding landing pages. An ad search engine operates in a similar way to an organic search engine; it consists of the following core steps: 1. landing page crawling; 2. indexing; 3. ranking ads for user search queries; 4 presentation of ads within SERP. In paid search, the crawler crawls the landing page associated with an ad creative; this is generally a much simpler and focused crawling process than for web search. Indexing uses, in addition to the features used in web search, a set of features specialized for advertising. An ad can be viewed as consisting of terms (unigram or n-gram) taken from the ad creative, or terms generated from the keywords that the advertiser associated with the ad, or terms extracted from the landing page. In addition, these n-gram terms can be further qualified by the zone in which they occur. For example, an n-gram term, such as, "ipod case", extracted from the creative title might be featurized as "ipod case:adTitle". These and other features will be discussed in detail below. Ranking of ads can be based on a hand-coded scoring function or based on a machine learning optimization strategy like those described in the MLR section above.

ECPM-based Ranking of Ads. For a search engine (publisher) to show an advertiser's ad, the advertiser (or their representative, e.g., search engine marketer (SEM); SEMs are advertising agencies that manage the search ad campaigns of large companies) must (1) create a text ad creative; (2) select the key search words or phrases that they would like their ad to appear alongside (specifying exact match or a broader syntactic and semantic match); (3) set other targeting constraints such as geo-constraints (e.g., show my ad in the San Francisco area only); and (4) set budgeting and delivery constraints. For the latter point, more concretely advertisers must set the following: their budget or total advertising spend for this ad campaign or overall; the start and end times for their ad campaigns and their maximum cost-per-click (CPC), which is the maximum amount an advertiser is willing to spend to have one searcher click on the ad and be transferred to the advertiser's website. All of this information is known only by the advertiser and the publisher.

One of the biggest challenges for the search engine is to establish how much a click is worth such that the search engine makes money at a fair market value and that all advertisers are getting a good ROI. This type of decision-making falls squarely in the field of economics where recent work in auction theory combined with game-theoretic analyses of auctions can be leveraged to create marketplace mechanisms, i.e., auctions (rules for bidding, and pricing), that enable both the search engine and advertisers to operate collectively to maximize the social welfare of the marketplace. In other words, it is necessary to create an auction mechanism such that the advertiser pays up to one's maximum value of the impression/click without remorse (for paying too much or missing opportunities) and by doing so the search engine earns optimal revenue. Initial auction models for sponsored search operated as follows: When a searcher inputs a search query, the search engine conducts a sealed auction across all ads with keywords matching the query (corresponding to a partial match, or as an exact match of the keywords associated with an ad). This auction both (1) induces a ranking of ads and (2) establishes a price per click on the winning ad(s). This induced ranking and pricing scheme can take many forms depending on the type of auction used. For example, it could be based exclusively on the bid price associated with the triggered keyword(s). The pricing for this model could be based upon second pricing where the advertiser is charged the bid associated with the ad in second place. For multi-slot SERPs, ads are selected based upon bid price and charged for clicks based upon the bid price of the ad in the next rank down in the ranked list; this is known as a *generalized second price auction*. This was the ranking strategy adapted by Goto.com/Overture when they introduced the first CPC-based cost model for online advertising back in 1997 [2]. This approach can be easily gamed by unscrupulous advertisers leading to many issues ranging from spamming-like behaviors to revenue reduction (due to active gaming-like behaviors that lead to reduced CPCs). For example, an advertiser could associate one's ads with a broad spectrum of keywords (possibly keywords that have nothing to do with the ads being targeted) and bid at high levels to insure impressions. Within this ranking model these ads would be shown to searchers thus leading to spam-like behavior (i.e., showing ads that have nothing to do with the information need or intent of the user's search). Ultimately, this type of activity would lead to reduced revenue (since there would be fewer clicks) and annoyed searchers. To overcome these limitations with bid-based ranking while maintaining keyword-based targeting Google developed a yield-based ranking model that optimized both revenue (via higher CPCs) and ad relevance. This ranking is based on a product of the advertiser bid price for that keyword and an ad quality score corresponding to the expected user behavior in terms of clicks on that ad. This product score is known as the expected CPM (ECPM) and is defined as follows:

$$\mathrm{ECPM}(a) = \mathrm{P(click(a)} \mid \mathrm{p, a, u}) * \mathrm{CPCBid_a} * 1000 \quad \textbf{(3)}$$

Where *p, a,* and *u* denote the context of the impression in terms of the target page (where the ad is shown; the can be simply the organic search query or the organic part of the resulting SERP), the ad, and the user respectively; and $CPCBid_a$ corresponds to the bid price that the advertiser has agreed to pay for a click on this ad in the case of a first price auction. This $P(click(a)|p, a, u)$ component is generally referred to as a quality score and can be viewed as a proxy for the click-through rate (CTR) of an ad shown in the context of a target page, user and ad. The *1000* multiplier in the above ECPM equation corresponds to 1,000 impressions (presentations of the ad to users). The ECPM-based scores induce a ranking and thereby determine the placement of ads on the search engine page with the top-ranked ad being placed in the top slot, and second ranking in the second slot, and so on. Pricing for this ranking model can take many forms. For now let's assume a second price auction mechanism whereby the advertiser is charged as follows for a click. The cost per click corresponds to the minimum it takes for an ad to maintain its position in the ECPM-based rank. For example, let's assume two ads ad_1 and ad_2 with corresponding bid prices Bid_1 and Bid_2 and quality scores Q_1 and Q_2 respectively. Let's assume that the rank scores for ad_1 is greater than ad_2. This is more succinctly stated as follows:

$$Bid_1 \times Q_1 > Bid_2 \times Q_2 \quad \textbf{(4)}$$

For ad_1 to maintain its current rank (using some basic algebra on the above equation) then Bid_1 needs to be at least:

$$Bid_1 \geq \frac{Bid_2 \times Q_2}{Q_1} \quad \textbf{(5)}$$

For a click on ad ad_1 the search charges the advertiser this minimum (plus an arbitrary small amount, say, one penny) to maintain its current ranking. This can be more naturally written in terms of CPC notation as follows:

$$CPC_1 = \frac{Bid_2 \times Q_2}{Q_1} \quad \textbf{(6)}$$

Where the advertiser pays a cost per click of CPC_1 for a click on ad_1. This is commonly known as the generalized second price auction (GSP) [2] and is used by most major search engines despite provable imperfections that will be discussed in Section 5. In its simplest form, if all click-through rates are equal, an advertiser's payment per click is the bid of the next-highest bidder. Since the context is implicit below for a given expression the notation *Pr(click(a)|p, a, u)* is simplified to *Pr(Click)*.

Sponsored Search versus Organic Search. Ranking for organic search can be reduced to generating the best rank ordering for a set of webpages with respect to a *query* (based on, say, *topical* relevance). Sponsored search differs in that it needs to balance the need of selecting and presenting text ads that satisfy the searcher's information need while in parallel delivering the advertiser's message. This joint objective is optimized by ECPM-based ranking (also known as revenue-based ranking). Estimating *Pr(Click)* accurately is key to creating an optimal ranking and pricing. For ads that have been shown many times in the context of an organic *SERP*, *p*, (here the user, u is marginalized), the estimate for *Pr(Click)* is the binomial maximum likelihood estimate; this is then simply the number of observed successes divided by the number of trials, i.e. clicks/impressions, or CTR. However this data requirement is generally very difficult to meet. Like most web distributions, the distribution of impressions and clicks follows a power law curve where the short head corresponds to ads with a high frequency of impressions and clicks on keywords and a long tail corresponds to rare events where an ad gets a small number of impressions and clicks on different keywords. In addition, since click rates for ads are generally very low (a recent study reported a CTR of 2.6% for ads on SERPs [18]) these estimates tend to have high variance. For example, if the true CTR for an Ad is 2.6% then it must be shown 1,000 times before one can be 95% confident that this estimate is within 1% (about 40% relative error) of the estimated CTR, i.e., in the range [1.6%, 3.6%]. This interval is narrowed down to [2.3, 2.9] after 10,000 impressions. To converge on confident estimates of CTR can represent a significant investment by the publisher (in this case the search engine). For example, it could cost $40 to converge on the 95% interval of 2.6% ±1% if the CPC were $1.60 or $400 for the tighter interval of 2.6% ±0.3%. Errors in these estimates can be detrimental to the search engine as it might lead to a loss in revenue (under charge due to higher estimates of Pr(click)) or missed opportunities for advertisers (who are priced out by a low estimate for CTR). The CTR rates on an overall SERP basis is 25% [19] corresponding about 3% CTR on any ad slot; most ads, however, will exhibit a CTR of less than 1% resulting in hugely rare events. This challenge is further compounded by a non-stationary marketplace: Sponsored search is a non-stationary market place consisting of hundreds of millions of ads, with a significant portion of new ads and advertisers entering (and leaving) the marketplace on a daily basis. Additionally, the constraints associated with current ads (such as keywords, geo, demographics) may also change. As a result there is a significant portion of ads for which there is limited or no information on Pr(Click) estimates.

Generally one is dealing with hundreds of millions of ads (order of 10^8) and billions of SERPs (10^9) and over a billion users(10^9). The cross product of these three large dimensional variables yields a massive event space (10^{26}) most of which is unexplored (cannot possibly observe user interactions for most <query, ad> pairs as there are not enough SERPs to do this and it would come at a very heavy cost to the publisher (the search engine) and user. This type of phenomenon is commonly known as the curse of dimensionality [20] and is well lamented in the field of machine learning. More concretely, with an estimated trillion (10^{12}) queries issued worldwide in 2009 [19] by about 1.3 billion people issuing about 700 queries per year, one can only hope to see 10^{22} events or about one hundredth of one percent of the possible events. Finally scalability is a challenge, paralleled only by organic search. In 2009, an estimated trillion (10^{12}) queries were issued worldwide by about 1.3 billion people corresponding to about 700 queries per user per year or through-put of 34,000 queries per second (QPS) globally. This is an extremely large number of SERPs. As a result of these key difference between organic search and sponsored search, many research opportunities have opened up that push the limits of areas such as machine learning, information retrieval, statistics, and large scale computing. Some challenges and current work in these areas are reviewed next.

3.4 Research Challenges and Opportunities

Characterizing ad selection and pricing as an online learning problem. Most of the major search engines (Bing, Google, Yahoo, Yandex) rank ads and price clicks using ECPM or yield-based ranking. Inaccurate estimates of the Pr(Click) can have adverse effects on user satisfaction and on advertiser ROI, which, ultimately, effects the revenue of the search engine. Consequently search engines are very motivated to accurately model Pr(click). To simplify the problem of ad selection and pricing based upon ECPM, one can assume that all ads have the same bid without losing generality. Framed like this the problem reduces to accurate estimation of Pr(Click) for a huge variety of ad contexts (user, publisher, and ad combinations). As was presented previously this problem suffers from the curse of dimensionality with little or no data about most possible contexts where an ad could be served. To overcome this problem various commercial entities have bootstrapped ad selection (i.e., accurately estimate Pr(Click)) in different ways leveraging both domain expertise and more statistically principled approaches. These schemes typically combine an exploit strategy that focuses on the short term and immediate revenue gains based upon accurate estimates of Pr(Click), generally for a small subset of possible events, with an explore strategy, where a small portion of query traffic is devoted to exploration that tends to be more speculative in nature. The goal of exploration is to discover other ad contexts where an ad can deliver high ROI for the advertiser and high revenue for the search engine. In addition, recent studies such as Broder et al. [21] and Shanahan et al. [22] have also started to look at lifetime value of users as opposed to the one-shot nature of the explore/exploit approaches discussed here; this is not discussed further here. Typically, one can frame all explore and exploit approaches as a form of online learning or sequence learning. Online learning is a model of induction that learns one instance at a time [23]. The goal in online learning is to predict values for instances, i.e., here, the Pr(Click) on ad in a given context. The key defining characteristic of online learning is that soon after a prediction is made, a more accurate estimate is discovered. This information can then be used to refine the prediction system using learning. The goal of learning is to make predictions that are close to the true values. More formally, an online algorithm proceeds as a sequence of trials. Each trial can be decomposed into the following steps. First the algorithm generates an estimate of Pr(Click), say, in terms of its mean and variance values. Then algorithm receives feedback via a click or lack of click on an ad in a specific context. The third stage is the most crucial as the learning algorithm can use this user feedback to update its hypothesis for future trials. The ultimate goal of the learning algorithm in the context of ad serving could be to maximize short-term revenue while balancing the need to learn about unknown ad behaviors. Because online learning algorithms continually receive feedback, the algorithms are able to adapt and learn in difficult situations. Online learning algorithms have been studied extensively leading to many theoretical findings on learnability that help guide exploration and exploitation dilemma, even, in an ad serving setting. To aid with understanding it useful to cast the problem of ad serving as a multi-arm bandit problem (a slot machine with multiple arms) where each arm corresponds to an ad with an unknown Pr(Click) and where each machine corresponds to a *<user, target page>* context. Let's assume one machine (corresponding to one ad context). When an arm is pulled it generates a revenue equivalent to the bid with a probability Pr(Click), The goal of online learning is to generate a finite series of arm pulls (ad servings) that maximizes the total expected revenue (this is often referred to reward in the online learning literature). This arm selection rule is often referred to as a policy. A bandit policy or allocation rule is an adaptive sampling process that provides a mechanism to select an arm at any given time instant based on all previous pulls and their outcomes. These series of arm pulls (ad servings) could be based upon a random policy (selecting an arm at random) or based on an exploit policy where an arm is selected based upon it expected revenue (based on its most current estimates). Thus, an arm with a worse empirical mean but high variance might be preferred to an arm with a better mean but low variance (exploration). After the sampling is continued for a while the online learning algorithm should learn enough to sample the arm that will provide the highest payoff (exploitation). A good sampling scheme should reach this point as efficiently as possible. A popular metric to measure the performance of a policy is called *regret*, which is the difference between the expected reward obtained by playing the best arm and the expected reward given by the policy under consideration [24]; it corresponds to the number of mistakes made (i.e. a mistake is when a suboptimal arm/ad is selected). A large body of online learning literature has considered the problem of constructing policies that achieve tight upper bounds on regret as a function of the time horizon (which is modeled as the total number of arm pulls) for all possible values of the expected revenues. This work has culminated in the following key finding: assume that each arm is assigned a priority function which is a sum of the current empirical payoff plus a factor that depends on the estimated variability (as opposed to a function of the expected revenue and its variance) [25]. Sampling the arm with the highest priority at any point in time, one explores arms with little information and exploits arms which are known to be good based on accumulated empirical evidence. With increasing sampling size, the sampling variability reduces and one ends up converging to the optimal arm. Consequently one cannot construct a variance adjustment factor to make the regret (or number of mistakes) better than log(N), thereby providing a benchmark for evaluating policies. Though online learning provides a very nice framework to analyze ad selection behavior it has many practical limitations. For example, the current framework assumes just one winner, whereas most ad contexts have the ability to show multiple ads. In addition, as a result of the curse of dimensionality, it is not possible to explore and then exploit all possibilities. To address this shortcoming Pandey et al. developed an approximation using a hierarchical model with a multi-stage sample strategy combined with a Bayesian model to do online learning at different levels of resolution [26]. Shrinkage and other back-off strategies can also be applied. Agarwal et al. [27] report an interesting hierarchical Bayesian model that address some of the idiosyncrasies of online advertising such as short lifetimes, delayed feedback and non-stationary rewards with compelling results. They also provide a good review of recent literature in this area. Yet another alternative approximation is based upon building regression models to predict click behavior as function of ad, user and target page characteristics. This can be accomplished using commodity learning frameworks such as logistic regression, though care has to be taken in training set construction as will be highlighted below to avoid biases that can provide misleading probability estimates. This latter approach and variations thereof are commonly used in practice.

Predicting Pr(Click). Reported studies on modeling Pr(click) fall into a number of categories: collaborative filtering approaches; predictive approaches; and evaluation approaches. All approaches leverage user click behavior (that can be harvested from log files) to model the Pr(Click) estimates; this can be viewed as a form of implicit relevance modeling. In addition, hybrid models have been developed where predicted CTRs are combined with empirical data counts using, say, Beta updating. This is in contrast to modeling ranking functions for organic search, where the objective is to model topical (and sometimes user) relevance and use such models to rank documents with respect to user information needs as expressed through queries; in this context, it is common to obtain explicit relevance judgments from human labeler(s) though some recent studies [28] have shown that learnt ranking models from implicit judgments give performances that are within 4% (in terms of normalized DCG) of models built using explicit relevance judgments. In addition, to this stricter predictive (scoring) requirement, relevance needs to be modeled beyond users' topical needs and the commercial intent and commercial opportunity also needs to be considered. Previous studies [29] have shown that advertising relevance involves complex inference processes that go beyond surface-form semantic models commonly used in traditional information retrieval and organic web search (typically unigram modeling with limited semantics such as the role of the n-gram in the query or sentence, e.g., a city or person's name). It can involve a complex interplay between the message contained in the ad creative (words, images, video), the target audience, the pitch level of the ad (to produce increases in brand awareness or to achieve other marketing objectives such as product sales), and user modeling. For example, if a consumer has already purchased a product they may not need further brand-based advertising. Understanding and modeling some of these interplays is a very fruitful area of research and product development that cuts across all aspects of digital advertising. It is an area largely unexplored with few theoretical models and where guidance is often achieved via empirical studies.

An example of such a study is reported in Richardson et al. [30], where they tackled the problem of estimating Pr(Click) for sponsored search ads from impression and click event log data using a logistic regression model:

$$\Pr(Click) = \frac{1}{1+Exp(-\sum_{i=1}^{n} w_i x_i)} \tag{7}$$

The problem of predicting Pr(Click) was modeled using training data of the form <*Ad, Term, CTR*> where *Ad* is the ad being potentially embedded in the SERP, *Term* is the search term (some or all of the query terms) input by the searcher, and *CTR* is binomial maximum likelihood estimate of the *CTR*. Each ad *Ad* contains the following information: bid terms; creative title and body; display URL; landing page and the number of times the ad was presented and clicked in the context of the keyword *Term*. Note: advertisers may specify whether an ad is displayed under the rules of exact match, or broad match. In the exact match case, the searcher query must exactly match the bid terms. In the broad match case, the bid terms can be related more loosely, such as being a subset of the query words. Richardson et al. used both exact match and broad match examples in their study. The dataset consisted of 10,000 advertisers, and one million example triples <*Ad, Term, CTR*>. The following is a list of some of the key feature families used in this study and represent common practice:

- Historical data: a smoothed CTR estimate for <*Ad, Term*> based on other ads with this keyword *Term*
- Appearance: features that describe the appearance of the ad creative such as the number of words in title or body; the presence of capitalization; the presence of punctuation; word length statistics
- Presence of certain call to action words such as "*buy*", "*join*", etc.
- Reputation based on URL: number of page-views and in-links; does the URL have a "*.com*", Etc.
- Landing page quality: is it WC3 compliant? Etc.
- Text Relevance: does the keyword match with ad title? Ad body? What is the fraction of match?

The learnt logistic regression models yields 29% improvement of mean squared error (MSE) over the baseline model (of always predicting the average click through rate of the training dataset). Though the paper focused on predicting the click-though rate for new ads, a similar model could be used for ads displayed repeatedly. Alternatively, data permitting, a Pr(Click) could be estimated empirically via maximum likelihood or using a hybrid of both approaches.

An alternative approach to predicting Pr(Click) for new *<Ad, Term>* pairs is based upon hierarchical clustering using keyword-advertiser matrix where the matrix entries correspond to the empirical CTR for ad *<Ad, Term>* pair [31]. This can be seen as a type of collaborative filtering. Their approach to estimating the CTR for low-volume or novel terms builds on the hypothesis that the more closely terms are related, the closer their CTRs. Following clustering, each node in the hierarchy is characterized by the average CTR for that node. Subsequently, predictions for the CTR of <Ad, Term> were based upon a variety of strategies: backoff strategies (analogous to the back-off method used for smoothing n-gram probabilities in language models [32]) that backed-off or smoothed the CTR estimates across parent, and grand parent nodes in the derived hierarchy; based on historical data (where available) such as average CTR of the entire <Ad, Term> dataset and the CTR of the <Ad, Term> in a previous time period. Regelson and Fain [31] show that the clustering-derived back-off model yields the most accurate predictions (of the examined models). Despite this, this approach is susceptible to huge variances for a keyword (e.g., the CTR for the keyword "surgery" has a huge variance across all the ads with which it is associated; the max CTR for<Ad, "surgery"> is 5 times the average [31]), which can be compounded even more when borrowing strength (smoothed) from clusters of keywords.

Ciaramita et al. [33] explore a number of machine learning approaches for ranking ads (as opposed to predicting CTR) where ranking models were generated and evaluated using click logs. They use click behavioral data from a large scale commercial Web search engine (Yahoo!). They formulate the problem of ranking a set of ads given a query as a learning task and investigate three learning methods of increasing complexity based on the perceptron algorithm: a binary linear classifier, a linear ranking model and a multilayer perceptron, or artificial neural net. To generate labeled data from the click logs, they adapt Joachims' blocking strategy [34]. This implicit labeling approach makes a conservative assumption that a click can only serve as an indication that an ad is more relevant than the ads ranked higher but not clicked, but not as an absolute indication of the ad relevance. As such the clicks on the top ranked ad do not carry any information, because the top ranked ad cannot be ranked any higher. Clicks at rank one are dropped from training and evaluation. For each clicked ad, they create a block which consists of the clicked ad and the non-clicked ads that ranked higher (lower in the SERP), yielding a total of 123,798 blocks in their study. In each block, they assign a score of *"+1"* to the clicked ad and *"-1"* to the ads that were ranked higher but were not clicked. Their results show that it is possible to learn and evaluate directly on click data encoding pair-wise preferences following simple and conservative assumptions. They considered the following feature families: word overlap between the query and ad creative; cosine similarity of the query and various aspects of the ad (such as ad title, ad description, and bid terms); and correlation features such as pointwise mutual information (PMI) between terms of the query q and the bidded terms of an ad measured over queries in the Yahoo search query logs. They report that metrics, such as precision at 1 and mean reciprocal rank, increase with the complexity of the model with the online multilayer perceptron learning performing the best for the reported experiments.

Bias Correction. Though machine learning approaches like those reported in [30] provide very encouraging results, they suffer from a number of biases such as training data composition and highly correlated features. For example, in the case of Naïve Bayes where features are assumed to be conditional independent, the presence of correlated features will lead to over prediction. In other cases, datasets may be down-sampled to improve discrimination in the learnt model. Consequently, the probabilities predicted by supervised learning algorithms can be under or over estimated. Previous research [35], [36] generates de-biased (corrected) probability estimates from supervised learning algorithms (such as support vector machines, Naïve Bayes, etc.) using a variety of approaches such as Platt scaling (a sigmoidal-based maximum likelihood approach) [37], isotonic regression[38], and smoothing probabilistic estimates within decision tree leaf nodes. Calibration significantly improves the performance of most of these learning algorithms on a variety of learning problems.

Keyword Suggestion. When an advertiser creates an ad campaign one of the most important steps is setting up the keyword targeting constraints. i.e., on which queries should the ad been shown to the searcher. Many automatic approaches to keyword suggestion that have been developed and deployed have generally borrowed and extended ideas from IR and NLP. Various starting points can be adapted: start with some seed keywords and suggest; start with landing page (and URL) and suggest; start with the ad creative, landing page and keywords and suggest. Starting with seed keywords, Wu et al. [39] compared a pseudo-relevance feedback system with an active learning approach. As a starting point for each keyword phrase they generated a corresponding characteristic or SERP-400 document. This document contained retrieved snippets from the top 400 search-hits; stop-words were removed and words were stemmed. The top 400 terms, weighted by TFIDF, in each seed's characteristic document are selected as candidate suggestion terms (inverse document frequency, IDF, is based upon a corpus of characteristic documents). Three approaches were primarily compared by Wu et al. [39]: a baseline based on the TFIDF ranking of suggested keyword unigrams; an approach based on pseudo-relevance feedback where the top five candidate terms (based upon TFIDF) and the bottom five terms are used to expand the original query using an approach such as Rocchio [40] for query expansion; and an active learning approach. The active learning approach used the following feature set for each <seed phrase, candidate unigram> pair as predictor variables: (1) Candidate-TF and Candidate-TFIDF corresponding to the TF and TFIDF of the candidate unigram in the seed phrase's SERP-400 document and the overall SERP-400 corpus respectively; (2) Seed-TF: the frequency of the seed in the search snippets document of the candidate unigram (SERP-400 of the candidate unigram); (3) SERP similarity: the similarity of the seed-phrase's SERP and the candidate's SERP. (4) URL overlap: the number of common URLs between the seed-phrase's SERP and the candidate's SERP. The learning algorithm used by Wu et al. is based upon Transductive Experimental Design (TED) [41], which tends to select candidates (for labeling and training) that are hard-to-predict and representative for unlabeled candidates with a core learning algorithm based upon regularized linear regression with Gaussian kernels. Overall, the active learning approach outperforms all reported approaches on an evaluation set of 100 seed phrases (based on average precision); these seed keyword phrase were based on popular category names. Though Wu et al. focus on unigram suggestion their approach could be extended to consider multiword phrases (using, for example, a language model approach discussed next).

Ravi et al. tackle the problem of keyword suggestion starting from a landing page perspective [42]. They explore a number of approaches: a cosine similarity-based model; support vector machines; and hybrid approach based on a language model combined with a translation model. Reviewing the latter approach only, to find good quality phrases, Ravi et al. used a bigram language model (i.e., using bigrams with back-off to unigrams) that was constructed from 76MM queries taken from a web search engine query log. Scoring bid-phrases was accomplished using a translation model over a <*bid phrase, landing page*> parallel corpus based on the *IBM Model 1* translation model. A sample of 10,000 pages was used for testing purposes. In this study for evaluation purposes, *ROUGE*, a commonly used translation and summarization metric (based upon measuring the unordered overlap between terms in the suggested bid phrase and terms in all gold bid phrases), and a normalized edit distance were used. Overall, the combination of bigram language model and translation outperformed all other systems. Other approaches to keyword suggestion not covered in here include query expansion to provide enhanced ad matching; Broder et al. report such a study using a taxonomy of 6,000 categories to enhance ad selection with compelling results. [43] This will be indirectly covered later in the behavioral targeting section (Section 4.3).

Metrics. Evaluation is an important part of online advertising. While traditional information retrieval measures such as precision, recall, mean reciprocal rank, fbeta etc., [10] and standard machine learning measures (such as mean squared error, MSE [7]) can be used, studies have highlighted weaknesses in these measures for ranking tasks. This has lead to the development of highly customized metrics and evaluation methods such as expected reciprocal rank (ERR) [28] ; and the replay-match method [44]. Lots more domain dependant metrics are called for to provide better model discriminating power while also minimizing experimental costs during live AB-testing similar in spirit of those approaches proposed by Cartlette and others for information retrieval metrics - see [45] for a review of key work in this area.

Borrowing from Direct Marketing. Online advertising targeting problems have many similarities with direct marketing systems and can borrow a lot from almost 70 years of experiences and ideas. Examples of such studies include [46] and [47].

Forecasting. Forecasting is a key component of online advertising. This component can be used by an advertiser to predict how many impressions and potential clicks an ad (ad creative, keywords, bids, and budget) will generate. This forecasting feedback can then be used to refine a campaign's constraints. Wang et al. present a very innovative way to do ad impression forecasting using an information retrieval framework for contextual advertising [48].

4 Contextual and Display Advertising

Embedding text ads within publisher's online media such as a web pages, portals and online/mobile/offline applications is commonly referred to as *contextual advertising*. This is also known as *content match*. Here the page where the ad is embedded is also known as the target page. Targeting of ads is largely based on the content of the target page, geographical constraints and demographic constraints. In terms of systems architecture, contextual advertising and sponsored search overlap to a large degree. Contextual advertising is generally provided to advertisers via ad networks whose key function is the aggregation of ad space *supply* from publishers and matching it with advertiser *demand.*

4.1 Research challenges and opportunities in contextual advertising

Contextual advertising presents many of the same challenges as that of sponsored search. For example, it parallels sponsored search with regard to estimating quality and ranking. However, contextual advertising is a lot more challenging as the user information needs and interests are not as crisply defined. For example the intent, think commercial intent, of the user reading a target page is less clear than when a user types an information need into a search engine. A user may be less open to ads, whereas in sponsored search a searcher is in an information gathering mode and ads may well satisfy the information need (and not be ignored). As a result users are more prone to click on sponsored search ads. On average, CTRs for contextual ads is ten times less than that of sponsored search ads. With an order of magnitude lesser training data, campaigns suffer from bootstrap issues. In short, targeting ads in contextual advertising is much more challenging and goes beyond the traditional IR notion of topical relevance and geographical relevance; these are much more aligned with demand and preference studies [49]. In an attempt to characterize these user preferences and intent, a number of studies have focused on characterizing human propensity to buy online, examining the effects of factors such as gender, age, and trust of online vendors (see [50] and [51] for more details). In addition, various studies have examined the language and format of advertising (explicit call to action via words such as buy now and subliminal messages). This notion of user relevance and interest closely aligns with an online marketer's perspective, where the goal is identify a well-defined target market or target audience. This is one of the first and key elements to a marketing strategy. Target markets are groups of people separated by distinguishable and noticeable aspects. Target markets can be separated into:

- Geographic segmentations (user's location)
- Demographic/socio-economic segmentation (gender, age, income occupation, education, sexual orientation, household size, and stage in the family life cycle)
- Psychographic segmentation (similar attitudes, values, and lifestyles)
- Behavioral segmentation (occasions, degree of loyalty)
- Product-related segmentation (relationship to a product)

Several studies have tackled this notion of audience creation from an information need and interest perspective within contextual advertising. For example, Ribeiro-Neto et al. [52] reported a pure information retrieval approach based on topical relevance to ad ranking (targeting) in the context of a Brazilian online newspaper. Ads and landing pages were indexed using word tokens (unigrams) that occur in the text-based ad creative and corresponding landing page. Here the target page can be viewed as a query, albeit a potentially long query, that characterizes an information need for ads that are topically related to its content. In this spirit, Ribeiro-Neto et al. explore traditional information retrieval techniques based on a TF.IDF cosine similarity while also addressing the vocabulary mismatch between the language used in the target page and in the ad (i.e., the words in the ad creative, the associated keyword portfolio, and the corresponding landing page) using approaches such as query (i.e., target page) expansion. The reported study (though limited to one hundred queries) showed that query expansion combined with indexing of the ad keywords and terms extracted from the landing page outperformed other information retrieval approaches.

Targeting is a two-way street, where the advertiser specifies constraints that characterize target online media or online audience and where the publisher provides characterizations of the media and audience available through the publisher media resources. One of the key challenges here is to abstract the target page to a level that connects with advertiser keyword and category constraints. Approaches to this can be organized as follows: keyword extraction; target page classification; and translation of target pages.

Targeting ads using keywords dominates sponsored search advertising and with no modification (on the advertiser's side) these ads can also be targeted at webpage level once the page is represented as a collection of keywords. Yih et al. [53] refer to this problem as harvesting keywords from web pages. They train a logistic regression model system using a set of positive and negative examples of the form <*keyword n-gram, target page, label*>, where keyword n-grams of up to length five were considered in the context of a target page to determine if a work or multiword phrase represents the topical focus of the page. Candidate keyword examples (positive and negative) were manually extracted from a sample of webpages. Two types of models were trained: one where a keyword n-gram is classified as a positive or negative phrase; and a more fine-grained model where each token was classified as B (beginning of a keyphrase, when the following word is also part of the keyphrase), I (inside a keyphrase, but not the first or last word), L (last word of a keyphrase), U (unique word of a keyword of length 1), and finally O (outside any keyword or keyphrase). The latter classifier requires a post processing step commonly known as the phrase-level inference step in information extraction (IE), where probabilities are assigned at the sequence level. This is commonly accomplished using Viterbi-like algorithms to find the most probable word label sequence assignment for each sentence [7]. This work is closely related to [54] where keyword extraction was accomplished using a combination of part of speech tagging, phrase chunking and lexicon-based approaches and [55] where a machine learning-based approach was used for keyword extraction in emails [55]. On the advertising front, suggesting keywords for advertisers to associate (and bid) with their ads is also a fundamental component of the contextual targeting system. This can be viewed as a keyword suggestion tool whereby keywords phrases related to already selected keywords (and bids) and to the keywords extracted from the ad creative and landing page can be used as a basis to find and suggest other keywords. Carrasco et al. [56] tackle this problem by the clustering of a keyword-advertiser bi-partite graph (where each cell value in this connectivity matrix, corresponding to an edge in this graph, represents the keyword bid by an advertiser). They validated the generated clusters by measuring intra cluster similarity based upon semantic similarity (cosine similarity); this was accomplished using words associated with each bid phrase by human editors. High similarity measures were achieved using a top-down clustering approach.

Broder et al. [57] explore how to characterize a target page based upon a taxonomy of 6,000 topics in addition to the usual keyword characterization. Each node in this taxonomy is associated with a list of queries. In their study both target page and ad are assigned to topics using a nearest neighbor approach where the similarity measure is based on a TF_IDF scoring between all query terms at a topic node and keywords associated with the target page or ad. They show that combining this topic-based (they refer to this as semantic) similarity and a cosine similarity measure between the keywords extracted from the target page and the ad leads to a 25% improvement for mid-range recalls of the combined model over the pure cosine (syntactic) model. The topic similarity score is a weighted product between every combination of page class and ad class category where the weight is the inverse distance of the pair of nodes in the taxonomy.

Predicting a quality score for an ad in the context of a target page presents similar challenges to scoring an ad for a SERP. Abstractly, similar approaches can be taken, however, there are differences that arise since the target page for contextual advertising is a web page (as opposed to a search engine results page consisting of 10 blue links and corresponding snippets) and the content of this page may not always be available for targeting at ad call time; since the ad selection engine has only a couple hundred of milliseconds available to select an ad thus preventing fetching and doing a text analysis of the target page content (for keyword extraction purposes). That being said, a number of studies have focused on using machine learning approaches to ranking ads for contextual advertising. For example, Lacerda et al. [58] proposed to use a genetic programming algorithm to select a ranking function which maximizes the average precision on the training data. They show that the ranking functions selected in this way are 61% more accurate than the baseline proposed (a cosine TF_IDF based ranker) in Ribeiro-Neto et al [52]. Murdock et al. explore a different ranking algorithm using support vector machines (SVM) with some novel input features that focus on the vocabulary mismatch between target pages and ads [59]. Specifically, they use a ranking SVM approach which optimizes a function directly related to Kendall Tau (as opposed to margin-based objective functions that most traditional SVM learning algorithms follow). Kendall Tau directly uses the level of disagreement between an ideal query ranking (provided in this study by human editors) and a ranking induced by the ranking SVM [34] and serves as a lower bound on the average precision. In this study three categories of features are explored: traditional features based on cosine similarity and term overlap between the ad and the target page; features based on statistical machine translation; and features based on machine translation evaluation. Overall, this study showed that using both sets of translation-based features produce statistically significant improvements in precision at rank one compared to their baseline, a summed cosine similarity between the target page and each of the ad fields (title, keyword, summary, and landing page). Ciaramita et al. [33] extend this work by looking at other novel features that capture semantic associations between the vocabularies of the ad and the Web page. The examined features look at distributional co-occurrence patterns between lexical items in the ad and target page.

Serving ads for target pages in an advertising network has similar round trip requirements as in sponsored search. When a target page requests ads for the first time limited processing of the target page can be carried out due to latency and bandwidth constraints. Anagnostopoulos et al. [60] compare a number of alternative lightweight processing strategies of the target page such as page summarization that can be executed on the client-side. The resultant approaches yield targeting performances comparable with full-page processing.

4.2 Display Advertising

While contextual advertising is sometimes included within the larger realm of display advertising, the term is often reserved for graphical (as opposed to text ads). It includes all forms of advertising excluding text-based ads (i.e., excludes sponsored search, contextual and classifieds) and plays on a range of different online media. This type of advertisement ranges from traditional banner ads in a variety of size and shapes, to video ads, and can be positioned within a webpage or app or can be incorporated into pop-ups. In 2009 it made up 35% of online revenue in the US [4]. While in the past this form of advertising was primarily attractive to brand advertisers (due to the rich communication bandwidth afforded by images and video), today's display ads are often used for direct response as well. Recently, advertisers have begun to measure both brand and direct marketing effects of their display campaigns - this is called brand response. In a study conducted by the Atlas Institute [61], users exposed to both search and display convert at an even higher rate – 22 percent better than search alone and 400 percent better than display only.

One of the major differentiators of display advertising from a ranking and ad selection perspective is that images and video lack the kind of machine friendly features that text provides. Humans interpret and understand images and videos in ways that are not expressible by TFIDF type measures. To complicate things, display ads have one or two orders of magnitude smaller CTRs than search. Since many display ads are not easily clickable, advertisers also measure view-through conversions which are extremely rare events; with action rates typically in the range of 10e-5, it makes learning and statistical aggregation extremely difficult. Ad selection and optimization in display advertising therefore revolves around making accurate predictions with efficient explore/exploit strategies. Due to the paucity of traditional machine learning features, collaborative filtering is an interesting alternative which is being explored [62]. A recent trend in display advertising is to target audiences rather than contexts. Users are segmented by data-driven modeling techniques into audiences that may be interested in the advertiser's products and therefore more likely to respond to the message. Display advertising is a large and growing field – however the scope of this chapter is simply to give an overview of the display landscape rather than an in-depth treatment.

4.3 Research challenges and opportunities in Display Advertising

Behavioral Targeting and Retargeting. One of the key technologies behind display advertising is behavioral targeting (BT) which is targeting of ads based on a user's browsing behavior. This is commonly used by e-commerce websites and has been more recently adapted en masse by companies providing sponsored search [63] (e.g., expand user query by terms in previous queries), contextual advertising, and display advertising networks. In search, behavioral targeting can be accomplished using query expansion schemes such as Rocchio [64] to expand the original context with terms from recently browsed pages (or transactions) and then weighting these terms with contributions from the base context combined with weight from positive examples (e.g., clicked pages in user browse history), and weight from negative examples (skipped pages in browse history). See [65] for a good example of a post-hoc study of IR approaches applied to BT and [66] where Chen et al, built a linear Poisson regression model from fine-grained user behavioral data to predict click-through rate (CTR) from user history. Online learning techniques can also be use to model another aspect of BT by generating "Look-alike" models from users who exhibit positive behavior (e.g., purchase an advertiser's product) and from users who don't purchase. Exploration techniques can be used to generate interesting candidate users that could provide feedback to accelerate Look-alike learning. An alternative way to accomplish behavioral targeting is to perform offline categorization of a user based on one's browsing behavior and then constrain the ads results set based upon this categorization. This could be modeled using simple bag-of-word techniques or machine learning. Retargeting is a very high ROI model of advertising where users that have previously visited the advertiser's website but have not purchased, or may be likely to purchase again are targeted by the advertiser.

5 Economic Models of Online Advertising

One of core concepts in online advertising is the digital marketplace where publishers present their ad slots for sale to advertisers who wish to purchase these slots for the purpose of showing ads. In online advertising this price is established via a sealed auction. The goal is to create an auction that encourages bidders to bid their true value (known as an incentive compatible auction). Such an auction mechanism helps advertisers avoid buyer's remorse and enables publishers to get paid a fair market value. One such auction mechanism is the second price auction where in the case of a single item auction the winning advertiser (corresponding to the advertiser with the highest bid) pays the bid associated with the second ranking advertiser. Second price auctions were introduced by Nobel Laureate William Vickrey, who was one of the first people to use game theory to develop and study auction mechanisms. When multiple items are being auctioned in the same auction, the more general form of the second price auction can be used. This is commonly known as a generalized second price auction. Generalized second price auctions are commonly used in the online advertising world and have been demonstrated as an effective means of allocating ads to publishers slots at companies such as Google, Yahoo and Turn [2]. Generalized second price auctions are not incentive compatible and as a result a new type of auction was developed that addresses this weakness. This is known as the Vickrey, Clarke, Groves (VCG) auction. Despite their lack of truth-telling, GSPs are the de facto standard for online auctions. This is primarily due to the ease of understanding of the auction mechanism (for both ranking and pricing) versus VCG. Both auction mechanism design (sometimes known as inverse game theory) and bidder design (analyzing effective bidding strategies) are very active areas of research within economics and online advertising with many conferences and journals devoted to the subject.

6 New Trends and Issues

This section is a highly condensed overview of new trends in online. Mobile advertising (advertising on cell-phones, be it SMS-based, application-based or browser-based) is one of the fastest growing segments in digital advertising and comes with its own challenges of performance and relevance measurements (e.g., clicks are uncommon in mobile). IP-based TV is a very new area that will transform a once broadcast advertising medium into a more personalized marketing experience. Real time bidded ad exchanges are fast-growing marketplaces where publishers bring their inventory to sell to advertisers who wish to advertise on that inventory. Individual ad impressions are auctioned off in real-time by publishers to buyers on the exchange. To take advantage of such real-time exchanges, demand side platforms (DSP), a new type of trading desk, empowers the advertiser or ad agency to make complex data driven decisions to evaluate, optimize, and purchase media and audiences across different media sources and exchanges via intuitive user interfaces. Challenges here include performing real-time bid optimization at an unprecedented scale of bid requests, each of which much be evaluated (in 2010 requests for bid is estimated to peak at 200,000 QPS across all exchanges for the US alone - compared to 34,000 QPS for search). Data exchanges are a relatively new entity in online advertising where third party suppliers and users of consumer intent and behavioral data congregate to sell and buy this data (in order to enhance targeting). The challenges here include mining user intent from hugely rare events sequences. Dynamic creatives refer to creatives constructed on the fly, typically based on the audience or context. Imagine an ad that shows a different image for a female user versus a male user – where the data about the gender of the user is either bought from a third party on a data exchange or algorithmically inferred based on the user's browsing patterns. Another major current topic of research and development concerns multiple touch points in the advertising chain (search, multiple display vendors, etc) and how an advertiser can attribute conversions to individual ad impressions and thus to individual intermediaries in the advertising chain (this is also known as credit assignment). Social advertising leverages historically "offline" dynamics such as peer-pressure, recommendations, and other forms of social influence to target ads based an individual's social network or affinity network. This type of advertising, being very new, presents many algorithmic and computational challenges and opportunities to leverage recent work in social network analysis such as information diffusion.

Privacy and fraud (an estimated $2 billion problem in the US alone), though not discussed here, will continue to be important areas with big needs for technology solutions.

7 Conclusions

From an information science perspective the field of online advertising is very active in terms of research and development. This is fueled by an annual revenue stream of $65 billion that continues to grow at a rate of 10% or more. In addition, as more of the traditional broadcast media sources (such as TV) move online, and the use of smart phones and handheld computers become more pervasive, the need for better ways to optimize the consumers advertising experience through personalization will become even greater. As was highlighted in this chapter information science will continue to be one of the cornerstones in making this happen.

References

1. Internet Advertising Bureau - About the IAB, http://www.iab.net/about_the_iab.
2. Edelman, B., Ostrovsky, M., Schwarz, M., Fudenberg, T.D., Kaplow, L., Lee, R., Milgrom, P., Niederle, M., Pakes, A.: Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords. American Economic Review. 97, (2005).
3. AdEx – the definitive guide to the size and scale of European online advertising, 2009 Report, http://www.iabeurope.eu, (2010).
4. IAB Internet Advertising Revenue Report 2009. Pricewaterhouse Coopers LLP (2009).
5. Witten, I.H., Moffat, A., Bell, T.C.: Managing Gigabytes: Compressing and Indexing Documents and Images, Second Edition. Morgan Kaufmann (1999).
6. Brin, S., Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks. 30, 107-117 (1998).
7. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE. pp. 257–286 (1989).
8. Sort-merge join - Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Sort-merge_join.
9. Broder, A.Z., Carmel, D., Herscovici, M., Soffer, A., Zien, J.: Efficient query evaluation using a two-level retrieval process. Proceedings of the twelfth international conference on Information and knowledge management. pp. 426–434ACM, New York, NY, USA (2003).
10. Croft, B., Metzler, D., Strohman, T.: Search Engines: Information Retrieval in Practice. Addison-Wesley Publishing Company, USA (2009).
11. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Trans. Inf. Syst. 20, 422–446 (2002).
12. Manning, C., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge, MA (2008).
13. Amazon Mechanical Turk - Welcome, https://www.mturk.com/mturk/welcome.
14. Manning, C.D., Raghavan, P., Schtze, H.: Introduction to Information Retrieval.

Cambridge University Press, New York, NY, USA (2008).
15. Joachims, T., Radlinski, F.: Search Engines that Learn from Implicit Feedback. Computer. 40, 34–40 (2007).
16. Cao, Z., Liu, T.: Learning to rank: From pairwise approach to listwise approach. In Proceedings of the 24th International Conference on Machine Learning. pp. 129–136 (2007).
17. Ogilvy, D.: Ogilvy on advertising / David Ogilvy. Orbis Publishing, London : (1983).
18. Burns, E.: SEMs Sees Optimization PPC, http://www.clickz.com/showPage.html?page=3550881.
19. Khan, I.T.: Nothing but Net. JP Morgan (2010).
20. Bellman, R.: Dynamic Programming. Dover Publications (2003).
21. Broder, A.Z., Ciaramita, M., Fontoura, M., Gabrilovich, E., Josifovski, V., Metzler, D., Murdock, V., Plachouras, V.: To swing or not to swing: learning when (not) to advertise. ACM Conference on Information and Knowledge Management. pp. 1003-1012 (2008).
22. Shanahan, J., Van den Poel, D.: Determining optimal advertisement frequency capping policy via Markov Decision Processes to maximize click through rates. NIPS 2010 Workshop on Machine Learning for Online Advertising. (2010).
23. Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning, and Games. Cambridge University Press, New York, NY, USA (2006).
24. Lai, T., Robbins, H.: Asymptotically efficient adaptive allocation rules. Advances in Applied Mathematics. pp. 4–22 (1985).
25. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time Analysis of the Multiarmed Bandit Problem. Mach. Learn. 47, 235–256 (2002).
26. Pandey, S., Agarwal, D., Chakrabarti, D., Josifovski, V.: Bandits for taxonomies: A modelbased approach. In In Proc. of the SIAM International Conference on Data Mining. SDM (2007).
27. Agarwal, D., Chen, B., Elango, P.: Explore/Exploit Schemes for Web Content Optimization. ICDM. pp. 1-10 (2009).
28. Chapelle, O., Metlzer, D., Zhang, Y., Grinspan, P.: Expected reciprocal rank for graded relevance. Proceeding of the 18th ACM conference on Information and knowledge management. pp. 621–630ACM, New York, NY, USA (2009).
29. Vestergaard, T., Schroder, K.: The language of advertising / Torben Vestergaard and Kim Schroder. B. Blackwell, Oxford ; New York : (1985).
30. Richardson, M., Dominowska, E., Ragno, R.: Predicting clicks. Presented at the the 16th international conference, Banff, Alberta, Canada (2007).
31. Regelson, M., Fain, D.C.: Predicting click-through rate using keyword clusters. Electronic Commerce (EC). ACM (2006).
32. Katz, S.M.: Estimation of probabilities from sparse data for the language model component of a speech recognizer. IEEE Transactions on Acoustics, Speech and Signal Processing. pp. 400–401 (1987).
33. Ciaramita, M., Murdock, V., Plachouras, V.: Online learning from click data for sponsored search. Proceeding of the 17th international conference on World Wide Web. pp. 227–236ACM, New York, NY, USA (2008).
34. Joachims, T.: Optimizing search engines using clickthrough data. KDD. pp. 133-142 (2002).
35. Niculescu-mizil, A., Caruana, R.: Predicting Good Probabilities with Supervised Learning. In Proc. Int. Conf. on Machine Learning (ICML. pp. 625–632 (2005).
36. Zadrozny, B., Elkan, C.: Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. In Proceedings of the Eighteenth International Conference on Machine Learning. pp. 609–616Morgan Kaufmann (2001).
37. Platt, J.C.: Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. ADVANCES IN LARGE MARGIN CLASSIFIERS.

pp. 61–74MIT Press (1999).
38. Härdle, W.: Robertson, T., WrighT, F.T. and R.L. Dykstra: Order restricted statistical inference. Statistical Papers. 30, 316-316 (1989).
39. Wu, H., Qiu, G., He, X., Shi, Y., Qu, M., Shen, J., Bu, J., Chen, C.: Advertising keyword generation using active learning. In WWW'09. pp. 1095–1096 (2009).
40. Rocchio, J.: Relevance Feedback in Information Retrieval. The SMART Retrieval System. pp. 313–323 (1971).
41. Yu, K., Bi, J., Tresp, V.: Active learning via transductive experimental design. Proceedings of the 23rd international conference on Machine learning. pp. 1081–1088ACM, New York, NY, USA (2006).
42. Ravi, S., Broder, A., Gabrilovich, E., Josifovski, V., P, E., Pang, B.: Automatic Generation of Bid Phrases for Online Advertising.
43. Broder, A.Z., Fontoura, M., Gabrilovich, E., Joshi, A., Josifovski, V., Zhang, T.: Robust classification of rare queries using web knowledge. Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 231–238ACM, New York, NY, USA (2007).
44. Li, L., Chu, W., Langford, J.: An Unbiased, Data-Driven, Offline Evaluation Method of Contextual Bandit Algorithms. CoRR. abs/1003.5956, (2010).
45. Carterette, B., Kanoulas, E., Yilmaz, E.: Low cost evaluation in information retrieval. SIGIR. p. 903 (2010).
46. Hill, S., Provost, F., Volinsky, C.: Network-based marketing: Identifying likely adopters via consumer networks. Statistical Science. 22, 2006 (2006).
47. Larose, D.T.: Discovering Knowledge in Data: An Introduction to Data Mining. Wiley-Interscience (2004).
48. Wang, X., Broder, A., Fontoura, M., Josifovski, V.: A search-based method for forecasting ad impression in contextual advertising. Presented at the , Madrid, Spain (2009).
49. Menzel, H.: Information Needs and Uses in Science and Technology. Annual Review of Information Science and Technology. 1, 41-69 (1966).
50. Levin, A., Levin, I., Weller, J.: A Multi-Attribute Analysis of Preferences for Online and Offline Shopping: Differences Across Products, Consumers, and Shopping Stages. Journal of Electronic Commerce Research. 6, 281-290 (2005).
51. Zhou, L., Dai, L., Zhang, D.: Online shopping acceptance model – a critical survey of consumer factors in online shopping. Journal of Electronic Commerce Research. 8, 41–61.
52. Ribeiro-Neto, B., Cristo, M., Golgher, P.B., Silva de Moura, E.: Impedance coupling in content-targeted advertising. Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 496–503ACM, New York, NY, USA (2005).
53. Yih, W.: Finding advertising keywords on web pages. In Proceedings of WWW. pp. 213–222ACM Press. ISBN (2006).
54. Grefenstette, G., Shanahan, J.G.: Document Souls: Joining Personalities to Documents to produce proactive documents engaged in contextualized, independent search, Context-Based Information Retrieval. (2005).
55. Goodman, J., Carvalho, V.R.: Implicit Queries for Email. CEAS. (2005).
56. Carrasco, J., Fain, D., Lang, K., Zhukov, L.: Clustering of bipartite advertiser-keyword graph. Proc. International Conference on Data Mining (ICDM'03). , Melbourne, Florida (2003).
57. Broder, A.Z., Fontoura, M., Josifovski, V., Riedel, L.: A semantic approach to contextual advertising. SIGIR. pp. 559-566 (2007).
58. Lacerda, A., Cristo, M., Gonçalves, M.A., Fan, W., Ziviani, N., Ribeiro-Neto, B.A.: Learning to advertise. SIGIR. pp. 549-556 (2006).
59. Murdock, V., Ciaramita, M., Plachouras, V.: A noisy-channel approach to contextual

advertising. Proceedings of the 1st international workshop on Data mining and audience intelligence for advertising. pp. 21–27ACM, New York, NY, USA (2007).
60. Anagnostopoulos, A., Broder, A.Z., Gabrilovich, E., Josifovski, V., Riedel, L.: Just-in-time contextual advertising. CIKM. pp. 331-340 (2007).
61. The Combined Impact of Search and Online Display Advertising - – Why Advertisers Should Measure Across Channels. Atlas Institute.
62. Agarwal, D., Chen, B.: fLDA: matrix factorization through latent dirichlet allocation. Proceedings of the third ACM international conference on Web search and data mining. pp. 91–100ACM, New York, NY, USA (2010).
63. Google to Offer Ads Based on Interests - NYTimes.com, http://www.nytimes.com/2009/03/11/technology/internet/11google.html?_r=1.
64. Rocchio, J.J.: Relevance feedback in information retrieval. The SMART Retrieval System -- Experiments in Automatic Document Processing. pp. 313–323Prentice Hall, Englewood Cliffs, NJ (1971).
65. Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., Chen, Z.: How much can behavioral targeting help online advertising? Proceedings of the 18th international conference on World wide web. pp. 261–270ACM, New York, NY, USA (2009).
66. Chen, Y., Pavlov, D., Canny, J.F.: Large-scale behavioral targeting. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 209–218ACM, New York, NY, USA (2009).