7
Social Search

Peter Brusilovsky[1]([✉]) [iD], Barry Smyth[2], and Bracha Shapira[3]

[1] School of Computing and Information, University of Pittsburgh,
Pittsburgh, PA 15260, USA
peterb@pitt.edu
[2] Insight Centre for Data Analytics School of Computer Science,
University College Dublin, Dublin, Ireland
barry.smyth@ucd.ie
[3] Department of software and Information Systems Engineering,
Ben-Gurion University of the Negev, 84105 Beer-Sheva, Israel
bshapira@bgu.ac.il

Abstract. Today, most people find what they are looking for online by using search engines such as Google, Bing, or Baidu. Modern web search engines have evolved from their roots in information retrieval to developing new ways to cope with the unique nature of web search. In this chapter, we review recent research that aims to make search a more social activity by combining readily available social signals with various strategies for using these signals to influence or adapt more conventional search results. The chapter begins by framing the social search landscape in terms of the sources of data available and the ways in which this can be leveraged before, during, and after search. This includes a number of detailed case studies that serve to mark important milestones in the evolution of social search research and practice.

1 Introduction

Search-based information access is the most popular way for people to locate information online, and the text-based query-box and search button have become a ubiquitous user interface component across the gamut of operating systems and apps. Usually when we have a general information need, we start with a query and a search engine, such as Google or Bing, and simple browsing or recommendation engines play a secondary role. Search-based information access has a long research heritage that began with early work on information retrieval during the early 1970s [169]. Somewhat surprisingly, despite at least 40 years of active research and practice, the world of information retrieval and web search has largely focused on a single-user embodiment that still informs modern search engines. By and large, modern search is framed as a single-user activity in which a lone searcher conducts their search in isolation from others, despite the fact that other users may experience similar information needs, or may have satisfied the same or similar needs in the past. Given this, why shouldn't search be viewed as

a more social or collaborative enterprise by harnessing the past work or present needs of others? While a few visionary projects from Memex [39] to Superbook [165] have suggested some ideas about how to leverage the power of community for information finding, it was not until the second part of 1990 that this idea of a more social approach to search—what we will refer to as "social search"— enjoyed a sustained and systematic interest from the research community.

In this chapter, we adopt the major theme of this book by defining social search as any approach to information search that harnesses the information access patterns of other users, whether past or present. One way to achieve this is to leverage the query and selection traces of past searchers to help future searchers to find what they are looking for, either by helping them to formulate better queries or to identify more promising leads by promoting more relevant content. Before the age of the web, the opportunities for collecting sufficient volume of user "search traces" were limited. This changed as the world of the web became the world of search. Today, the volume of queries and selections that feed modern search engines provides a unique resource for social search endeavors. In fact, as search engines continuously look for new ways to improve their performance, such data doesn't just enable social search, but provides an appealing opportunity to radically transform how we search. By 1997, the search engines of the day were already collecting increasing volumes of queries and clicks from their users. Even at this early stage, the Web had outgrown traditional query-based, term-matching, information-retrieval techniques as a viable solution to the challenges of general web search. This led to a explosion of creative ideas, by exploring several kinds of novel search data and user traces to address pressing problems such as query formulation [58,74,85,163] and document ranking [36,53,57,217].

Indeed, the "big idea" for modern search emerged from this early work, as several researchers independently highlighted the power of various forms of search traces – what we might now refer to as social, community, or crowd data – to fundamentally change web search from its information-retrieval origins. Most influential were two novel search engines: Google [36], which leveraged community page linking for better ranking; and DirectHit [202], which leveraged user queries and page exploration behavior for query term suggestion [58] and ranking [57]. Over the next 15 years, this stream of work gradually became an important and active research direction at the crossroads of information retrieval, social computing, and intelligent interfaces. The work produced over these years explored a much larger variety of social traces and suggested many novel ideas for using social information to enhance search.

This chapter attempts to provide a systematic overview of social search as a research stream, with a comprehensive account of the core ideas that underpin modern social search approaches and by highlighting social search in practice through a variety of case studies. It complements two other chapters in this book that offer a deeper analysis of two specific areas of social search: Chap. 8 on network-based social search [92] and Chap. 9 on using tags for social search [149]. The chapter is structured as follows. The next section attempts to frame social search to define key dimensions for its classification. We position social search in a broader context, compare and contrast various definitions of social

search, identify key sources of information for social search, and elucidate the main stages of the information retrieval process where social information can be leveraged. The subsequent sections go on to review the main approaches for using social data in each of the identified stages. Following that, we examine the problem of making social search more personalized with a review of a number of cases studies to demonstrate different ways of addressing the personalization challenge. We conclude with a final review of two case studies that attempt to go beyond the usual borders of social search by connecting social search with other forms of social information access approaches reviewed in this book.

2   Framing Social Search

In this chapter, we attempt to frame social search as a unique approach to web search and information discovery, which sits at the intersection between information retrieval, social computing, and intelligent interfaces. We pay particular attention to the sources of information that drive social search and the social search process. But first, it is useful to review the evolution of search and, in particular, to distinguish between two common viewpoints: using sources of social data to support search versus searching social data sources.

2.1   Defining Social Search

A useful working definition for *social search*, albeit a very broad one, is that it refers to *a group of approaches that use past user behavior to assist current users in finding information that satisfies a specific information need*. Such an information need is typically represented by a query or question and is satisfied by the retrieval of some unit of information, which is typically (but not limited to) a web page. This definition follows the original understanding of the term social search, which emerged from several early papers between 1997 and 2007 during the first decade of research on this topic [8,37,74,76,77,102,140]. This perspective can be also considered as a useful match with terminology used in other well-defined types of social information access, such as social navigation and social bookmarking, although with a different emphasis on how information needs are expressed or satisfied.

It is important to note, however, how the increased popularity of social media and social networking in the decade that followed has encouraged a range of other definitions of social search, many of which differ from one another as well as from its original meaning. Most of these recent definitions framed social search as search within social media content or data. For example, Carmel et al. [41] use the term social search "to describe the search process over 'social' data gathered from Web 2.0 applications, such as social bookmarking systems, wikis, blogs, forums, social network sites (SNSs), and many others". Similarly, Wikipedia defines social search as "a behavior of retrieving and searching on a social searching engine that mainly searches user-generated content such as news, videos, and images related search queries on social media like Facebook, Twitter, Instagram, and Flickr" [203].

On the other hand, Bao et al. [21], who were among the first to use the term "social search" in the context of social media, defined it as "utilizing social annotations for better web search". Amitay et al. [13] also understand social search as "leveraging social information to enhance search results" where by social information authors mean information collected by social systems such social bookmarking and networking services. McDonnell and Shiri [137] define social search as "the use of social media to assist in finding information on the internet". Evans and Chi [67,68] expand this understanding of social search further: "Social search is an umbrella term used to describe search acts that make use of social interactions with others. These interactions may be explicit or implicit, co-located or remote, synchronous or asynchronous."

The first group of these definitions focus on *searching social data sources* and, as such, are somewhat orthogonal to our understanding of social search, since they distinguish social search by the *type of content* to be searched (i.e., social media contents). The second group emphasizes *the use of social data during search*, regardless of what is being searched, and is fully consistent with the meaning of social search used in this paper. At the same time, this definition narrows the scope of the "social data" that can be used to improve search to those data collected by social systems. In contrast, for the purpose of this work, our definition of social search refers to approaches that can improve search using *all* kinds of information traces left by past users/searchers, both inside and outside social media systems. As will be explained below, this chapter considers a variety of past information traces that range from queries and clicks, to votes and tags, to comments and social links, and everything in between. The common denominator is the use of any and all information left, or contributed by, users as a side effect of their natural information-seeking behaviors.

2.2   Sources of Social Information

Following the goal of the book, this chapter attempts to show how search processes can be improved using various kinds of explicit and implicit social information, and specifically traces left by previous users. The introductory chapter of this book [38] offered a classification of such traces and where they might be sourced. Here we focus on a subset of these traces: those that have been found to be useful for users and that have been used to explore search systems. Below, we review the main sources of social information for social search, along with the types of social information that is usually available within these sources. These sources provide one of the two primary dimensions in our classification of social search techniques.

2.2.1   Search Engine Logs

Search engines routinely collect query session logs. A typical query session log includes queries issued by the users and the result pages selected, or "clicked", by the user on the search engine result page (SERP). More sophisticated logging may include additional data, such as dwell time, within-page behavior, and

even post-query browsing. Either way, queries and clicks are perhaps the most important and popular kinds of social information, which can be used alone or in concert to guide social search.

2.2.2 Information Links

User-generated links, created by content authors between information items, such as Web pages, are another important source of community wisdom. Web page authors make extensive use of hyperlinks to create the webs of content that underpin the world wide web. And these links have served as one of the earliest sources of social information used to improve the search process [36,84]. Moreover, beyond traditional "open Web" and its hyperlinks, social search systems now use links created by users of the many and varied Web-based systems the exist today, such as wikis, blogs, and micro-blogs, for example.

2.2.3 Browsing Trails

Surfers of the "open Web" and other Web-based systems leave extensive browsing trails as they travel from content to content. Within a single Web site, clicked links are also routinely collected in web logs and provide yet another source of information-seeking behavior. Beyond these routine browsing traces, traces of browsing behavior can also be accumulated by using various browsing agents [129], proxy-based intermediaries [23] or browser plugins, such as Alexa or HeyStaks [181]. For the interested reader, a comprehensive review of user tracking approaches beyond single-site borders can be found in [83].

2.2.4 Annotation and Comments

User annotations and comments are examples of the "secondary" information content that can accumulate to augment primary content such as Web pages or product descriptions. In a Web context, annotation functionality is provided by various Web annotation systems including the original Annotea project from the WWW Consortium [108]. Indeed, today user-generated comments and reviews are commonplace across most e-commerce sites and have proven to be a powerful source of collective user opinion, helping to guide many users as they consider their choices among an ocean of products and services, from Amazon[1] to Yelp[2].

2.2.5 Curation Systems

We use the term *curation system* to distinguish several types of systems where users collectively collaborate to collect and organize content for future consumption. This includes hierarchically organized directories, such as the Open Directory Project (dmoz.org), guided path systems [81], and page grouping systems [2,177]. Most important among the curation systems, from the perspective of Web search, are social bookmarking systems, which allow their users to openly

---

[1] http://amazon.com.
[2] http://yelp.com.

share various information items (Web pages, photographs, research papers) while providing textual *comments* and *tag-based* annotations.

2.2.6   Blogs and Microblogs

Conventional blogs, as well as micro-blog systems such as Twitter, have emerged as important contemporary sources of primary user-generated content, and are now increasingly used by search systems. Just like in other contexts, an explicit reference to a Web page (i.e, a Web link) or another item (i.e., a movie) in a blog post or a tweet is a signal of its social value, with surrounding text providing a context for mentioning the link.

2.2.7   Social Links

Social links—such as the friendship connections between social network users—carry important information about relationships between users. There are several examples of social search systems that use direct links of various kinds (trust, friendship, etc.), as well as indirect links, such as those formed by the users who have joined the same social group. Social links originate in the social linking systems—originally systems like Friendster and MySpace, and today, platforms like Facebook and LinkedIn—but today the idea of a social link has evolved beyond a simple friendship connection. For example, Twitter allows user to follow and mention other users, while social bookmarking systems usually support "watch" links that help to keep interested users informed about changes and updates.

2.3   The Social Search Process

While conventional search might appear to be an atomic, mostly single-shot process—"submit a query, get results"—it is naturally comprised of several distinct stages. These various all stand to benefit in different ways from different kinds of social information. To better organize our review of social search, this section presents an "anatomy" of the social search process as a sequence of actions to be performed by both the user and the search system.

A sensible and straightforward decomposition of search identifies three obvious stages: "before search", "during search", and "after search". This decomposition was used in the past for the analysis of personalized [139] and social search [68]. It remains useful here because it serves to separate social search enhancements into those that might be provided as *external* supports (before or after search) and therefore potentially provided as third-party enhancements, versus those than are contrived as *internal* supports (during search), and therefore more naturally provided by search engines themselves.

2.3.1   Before Search

This stage includes all actions that must be performed before a query meets the search engine. On the search engine side, it includes resource discovery (known

as *crawling* in Web search context) and indexing. On the user side, it includes query formulation and query expansion. The reason to separate query formulation from query expansion is to emphasize how these processes differ from a user engagement perspective. Query formulation is a user-centered stage during which a search support system can offer valuable assistance. Query expansion is conceived of as a fully automatic process that attempts to improve the query after it is formulated by the user, but before it is received by the search engine.

2.3.2   During Search
This stage includes two related processes performed by the search engine: matching items to the query (and retrieving relevant items) and ranking these items. To distinguish the ranking produced by the search engine from various external or secondary re-ranking approaches, we refer to the former as *primary item ranking*; the latter will usually be referred to as *re-ranking.*

2.3.3   After Search
This stage includes several types of search enhancements that can be implemented after obtaining an original ranked list of search results in order to produce a better search engine results page. Historically, these enhancements were produced by "third party" systems, which took the list of results produced by a search engine and returned it in an enhanced form. However, nowadays, some of these enhancements are done by the search engines themselves, as shown in Table 1. The most popular of these enhancements are various kinds of item *re-ranking and recommendation* that use social or personal information to change or augment the primary ranking; e.g. [50,77,180,183]. Once the order of presentation is determined, a final *item presentation* needs to be generated, including an item summary or *snippet* to provide the user with a synopsis of the item.

Table 1. Stages of (social) search process

	Add-ons	Engines
I Before search		
1. Resource discovery and indexing		✓
2. Query formulation	✓	✓
3. Query expansion	✓	
II During search		✓
4. Matching items to query		✓
5. Primary item ranking		✓
III After search (SERP)		
6. Item re-ranking and recommendation	✓	
7. Item presentation (snippet building)	✓	✓
8. Item augmentation	✓	✓
9. After SERP support	✓	

The presentation of individual items may be also *augmented* to emphasize their social or personal relevance. While this concludes the SERP presentation, some projects have explored opportunities to use social information *after SERP*, i.e., to support the user after a specific SERP item has been accessed, as we discuss later in this chapter.

2.4   The Big Picture View

To summarize, these core stages of the search process are a useful way to organize our review of social search technologies. The following seven sections (3 to 9) review the use of social information to support each step of the search process listed in Table 1. The unique needs of each step help to frame the approaches taken and the types of social information used. Moreover, by grouping social search approaches by stages, we can clarify their commonalities and connections with their underlying social data sources. To better understand this connection, we also offer Table 2 that groups together approaches, which support the same step of the search process while using the same social information source.

Table 2. Stages of social search and their support using various sources of social information

Source of social information	Query formulation	Query expansion	Indexing	Primary ranking	Re-ranking and recommendation	Presentation	Augmentation	After SERP
Search engine logs	[19, 22, 58, 60, 75, 85, 101, 107, 121, 127, 132, 184, 192]	[25, 56, 74, 101, 120, 196]	[14, 61, 80, 82, 87, 114, 151, 159, 170, 171, 209]	[5, 28, 36, 46, 54, 55, 57, 63, 64, 82, 89, 104, 106, 112, 158, 160, 162, 176, 193, 209]	[32–35, 138, 176, 178–180, 183, 188, 200, 201]	[9, 30, 31, 170]	[50, 51, 124, 170, 195]	[9, 48, 50, 51, 72, 76]
Browsing trails		[120]	[215]	[16, 24, 65, 90, 130, 131, 190, 210, 216]			[8, 37, 48, 70, 72, 76]	[48, 72, 76]
Information links		[15, 52, 59, 119]	[44, 53, 78, 87, 117, 151, 153, 198, 215]	[36, 43, 44, 53, 78, 84, 115, 117, 153, 160, 198, 205, 215]				
Annotations and comments			[61, 165]		[145, 146]		[8, 37, 70, 72, 76]	
Curation systems		[4, 128, 207, 213]	[21, 26, 40, 47, 87, 97, 150, 151, 164, 208, 214]	[1, 21, 40, 100, 126, 206, 212, 214]	[103, 150, 208, 211]		[13, 109, 142, 211]	
Blogs and microblogs			[11, 45, 62, 66, 87, 123, 144, 157, 168, 191]	[7, 87, 136]	[7, 157, 168]		[13, 157]	
Social links	[20]			[189]	[77, 95, 176, 178–180]	[9, 30, 31]	[72, 73, 76, 122, 148, 155, 173]	

In the context of this chapter, Table 2 serves as a "big picture" that provides an overview of the whole social search area and highlighting both well-explored areas and "white spots" on the map of social search. In our review and the organization of Table 2 we mostly follow the order of the search process listed Table 1. One exception is the Sect. 5, where we review together tightly interrelated social indexing and matching, which belong to different stages in our classification. This section is placed between query expansion and ranking, where its matching aspect belongs.

3   Query Formulation, Elaboration, and Recommendation

Web search is challenging, not only because of the sheer scale of information that exists online, but also because of the simple lack of information retrieval expertise that most searchers exhibit. For example, it is well documented that most searchers begin their searches with vague or under-specified queries. Early work on web search established that most queries contain only 2–3 terms [186, 187, 204] and rarely offer a clear account of the searcher's real information needs. At the same time, the advanced search features offered by modern search engines (e.g. boolean operators, positional switches, filters etc.) are seldom if ever used; see, for example, [185, 187]. In short, a typical search query is usually woefully incomplete when it comes to the specification of a searcher's real information needs.

In response to this, there has been a significant amount of work undertaken on how to better support searchers by, for example, elaborating vague queries or recommending better ones. As a source of social information for query elaboration and recommendation, most systems use search engine logs. The idea of using such logs for accumulating successful queries was originally suggested by [163], and the first simple approach for using search logs to suggest new terms to narrow a search was patented by DirectHit [58]. The first example of a full-scale query recommendation service based on accumulated queries was described in [85] and was followed by a large body of work on query recommendation [19, 22, 60, 75, 101, 121, 132, 184, 192].

One of the early approaches to query recommendation was the mining of query logs to identify clusters of related queries [17, 19, 156, 197]. For example, in [19] the clustering process was based on the queries, and the terms contained in selected URLs and term-based similarity can be used to rank and recommend a set of queries that are similar to a particular target query. Another popular approach was based on building query networks using various techniques for query association [18, 85, 127, 132, 156, 184]. Yet another approach [22] borrows ideas from recommender systems research to apply collaborative filtering techniques for query recommendations, based on search sessions that contain successful queries.

The approaches reviewed above, and many further variations that exist, rely solely on submitted queries and result selections, and as a result, are less satisfactory as examples of social search, because they lack any real social context other than the set of users who happened to use a particular search engine. The work of Balfe and Smyth [20] serves as a more tangible social counterpoint, in

this regard, by applying query-log analysis and recommendation techniques in an early collaborative search setting. They describe a novel query recommendation technique that suggest a queries that have proven to be successful during past search sessions *within a community of related searchers*. In particular, they propose a novel recommendation ranking metric that prioritizes queries for recommendation that are based on their relevance and coverage characteristics, preferring queries that are not only relevant to the user's likely information need—but that also have the potential to cover a broad set of variations of these needs.

4   Query Expansion

One of the most natural and successful techniques to resolve the query-mismatch problem is to expand the original query with other query terms to better capture the actual user intent or to produce a query that is more likely to retrieve relevant documents. Automatic query expansion (AQE) has been studied since 1960, when it was suggested by Maron and Kuhns [133]. AQE is applied to the original query, as submitted by the user. This query is expanded to include additional terms and is then submitted to the search engine to drive search, and hopefully to retrieve better results. Thus, from the perspective of the user, AQE is a type of query formulation technique that is transparent, in that it is performed in the background in a manner that is typically unseen by the user.

Many techniques have been examined for AQE, such as relevance feedback [167] and term distribution analysis [161]. For many years, until relatively recently, the practical benefits of AQE were questioned in web search, as results mainly benefited from improved recall but with a cost to precision [94]; not an ideal combination for most web search tasks.

Recently, and largely due to the huge amount of available data and the low quality of user queries, AQE has been revisited. New AQE techniques have been presented that use new data sources and employ more sophisticated methods for finding new features that can correlate with the query terms. AQE is now regaining popularity as a promising method and has been adopted in various commercial applications, such as intranet and desktop search or domain-specific search engines [42]. However, it is still not commonly employed in major Web search engines, mainly due to their emphasis on precision, but also due to the need for fast response times, which preclude the costly computations that most AQE methods require.

A recent survey [42] classifies AQE techniques into five main groups according to the conceptual paradigm used for finding the expansion features: linguistic methods, corpus-specific statistical approaches, query-specific statistical approaches, search log analysis, and Web data. The linguistic approaches examine global language properties of the query, while the corpus-specific and query-specific techniques analyze the entire database or the query context to identify correlated word features to use for expansion. The search log and web data analysis approaches are more social in nature and are deserving of further discussion.

The use of search logs for AQE was originally suggested in [74]. This paper was one of the first to recognize search improvement based on social data as a

type of "social searching". Search logs typically contain the queries submitted by users, alongwith result URLs that were selected by the searchers. By mining these query-result associations, AQE techniques can, for example, predict the relevance of selected URLs to a given query, as in [175]. Expansion terms can be extracted from the top results, or the clicked results, or from documents that are frequently returned for similar queries [56,74]. The underlying assumption is that users who click on results usually do not make their choice randomly, but rather select results rationally, based on their perceived relevance to their information need. Even if some of the clicks are erroneous or mistaken, query logs remain a valuable source for automatic query expansion, due to the sheer volume of query and click data that mainstream search engines can provide.

Some systems, such as [56], investigate the association between query terms and the documents that were retrieved for the query during a specific search interaction. Their method is based on probabilistic term correlations that are computed by linking query terms to terms in documents that were retrieved in the same session. When a new query is submitted, for every query term, all correlated document terms are selected based on the probability that was pre-computed between them. For every query, the top-ranked terms, by their probability, can be selected as expansion terms. Another example is described in [25]. The authors suggest and compare some variations of associating queries extracted from a freely available large query log with documents of the corpus that is being searched. The association between pairs of documents from the corpus, and a query from the log, is based on close statistical matches between them. For each document in the corpus, query associations are maintained as document surrogates. Then, the terms of the associated queries are used in the search process, either as expansion terms, or as surrogates for document ranking. The authors describe experiments with almost one million prior query associations, which improved the effectiveness of queries by 26%29%, and showed that their method outperformed a conventional query expansion approach using the corpus of past queries.

Other query log analysis methods for AQE look at the association within query terms in the log. For example [196] analyzes term co-occurrence in queries, to discover association patterns for terms; more specifically, term substitution and addition within multi-word queries. The authors look for terms that can syntactically substitute for other terms (e.g. "auto" - "car"; or "yahoo" - "google") and terms that often occur together (e.g. "car" - "insurance"). They frame the term association pattern mining as a probability estimation problem, so that high-probability patterns are used for query expansion and reformulation. Similarly, [101] also look at the association between terms that appear in queries, but rather than analyzing their co-occurrence in a single query, they consider the entire user's query session to infer term associations across sequences of related queries. Query logs can be also combined with other social data for the purpose of AQE. For example, Kramár et al. [120] built communities of like-minded users by mining user browsing behavior and then used query analysis within a user community to produce more user-sensitive query expansion.

Another AQE technique, based on social data, harnesses *Web data sources* – such as web pages, wikis, blogs, and others – for term expansion, especially since query logs are not always available for the developers of AQE approaches, who may be external to the operators of a given search engine. A large variety of external linguistic sources can be used for AQE. One type (not related to social search) is expert-generated data, such as WordNet [143, 213]. However, various kinds of user-generated content, from regular Web pages to Wikipedia, is also a popular choice for AQE. Such data presumably reflects the perception of users about relevant terms and the semantic relationships between their terms.

The anchor text inside Web pages have also been used for AQE and specifically for query refinement (via expansion or substitution of terms) as suggested by Kraft and Zien [119], and as an alternative to query log analysis. Kraft and Zien [119] developed a simple and elegant approach to automatic query refinement using anchor text and median rank aggregation. Briefly, the query is matched with similar anchor texts and additional terms from the matching anchor texts are then used for query expansion. This is based on the notion that anchor text is an strong surrogate for the document to which a corresponding link connects. While this is a fairly sound assumption to make, a number of challenges exist when it comes to using this approach in practice. For example, popular queries may match too many anchor texts, and some anchor texts might be useless if they are automatically generate – rather than hand-coded by the Web page creator – a ranking algorithm is required to assess the utility of any matching anchor texts, to focus on those that are most likely to act as a source of high-quality additional terms. Kraft and Zien [119] used factors such as the weighted number of occurrences of an anchor text, the number of terms in the anchor text, and the number of characters in the anchor text, to rank the anchor texts. Their experiments successfully demonstrated how mining anchor text for query refinement in this way was capable of outperforming similar methods using document collections.

Following on from this work, Dang and Croft [59] constructed an anchor log from the anchor texts in a web test collection. They compared their log-based query reformulation techniques, using both the anchor log and a real query log, and showed that the former produces results that are at least as effective as the latter. Their anchor log was built from the TREC Gov-2 web collection, which contains 25 million web pages crawled from the .gov domain during early 2004. The integration of this algorithm into a search process was based on providing the user with a list of ranked query refinements, from which the user selected one that best suited their needs.

A more recent study by Craswell et al. [52] extended the work of Dang and Croft [59] by fully automating the formulation process. The authors again used a large anchor graph as a linguistic resource for the query rewriting, but introduced new algorithms based on a random-walk approach, which proved to be capable of offering statistically significant improvements in precision.

Wikipedia, DBpedia, FreeBase, and the like, are also considered as effective sources for AQE [4]. Wikipedia is the largest and most popular collaborative effort of volunteer contributors, and is routinely ranked among the ten most

popular websites. It is the leading, and most updated open encyclopedia, and offers wide coverage of a diverse range of topics, events, entities, and more. It is considered to be a reliable data source and has found application in many AQEs. DBpedia [125] is a large-scale multilingual, semantic, knowledge base extracted from Wikipedia info-boxes, and is connected to several other linked data repositories on the Semantic Web. FreeBase was a large collaborative knowledge base consisting of data composed mainly by its community members. The data was collected from various sources, including user-submitted wiki contributions. FreeBase data is now being moved to WikiData. All of these sources have been used to enrich and expand queries with new terms that are inferred as relevant based on the content and/or structure of these other resources. Given a query string, the search engine may identify relevant pages or entities within the source that it uses, such as relevant Wikipedia concepts, and use their titles, categories, or other components to infer relatedness between terms to expand the query. Different approaches rely on different sources of query terms, use different methods to identify relevant information within a source, and rely on different approaches when it comes to selecting and ranking terms for use in AQE.

For example, Arguello et al. [15] used the links that appear in Wikipedia pages, which are relevant to the query string, and the anchor text associated with these links to expand queries. Li et al. [128] used the categories from Wikipedia pages for AQE. Briefly, the starting query is run against a Wikipedia collection; each category is assigned a weight that is proportional to the number of top-ranked articles assigned to it and articles are re-ranked based on the sum of the weights of their corresponding categories.

Elsewhere, Xiong and Callan [207] used FreeBase as their source. They investigated two methods for identifying FreeBase entities that are relevant to a query, and two methods of using these entities for query expansion, by utilizing information derived from Freebase linked entity's description and categories, to select the terms for expansion. Like most of the systems that utilized user generated content for expansion, results are superior to more conventional state of the art AQE methods. According to Xiong and Callan [207], using experiments conducted on the ClueWeb09 dataset with TREC Web Track queries, these methods are almost 30% more effective than state-of-the-art query expansion algorithms. This makes sense: query expansion aims at expanding queries with terms that users would consider relevant in order to enhance the queries; therefore AQE should benefit from harnessing user-generated content for expansion, since it reflects how people view the relationships between terms. A user who assigns a page to a category is signaling that they believe that the topic of the page is related to that category.

5   Social Indexing and Matching

An inherent challenge for all search systems is the surprising degree of variety with which people refer to the same thing. Frequently referred as *the vocabulary problem* [79], it means that people with the same information need are liable to describe it using very different query terms. Sometimes people with different

information needs will formulate similar queries. This makes it exceedingly difficult to index content for future retrieval. This problem has been recognized in the early days of automated search based on manual document indexing. Back then, professional indexers were often blamed for conceptualizing a document incorrectly, or at least in a way that was at odds with searchers of the document. The switch to automatic content-based document indexing, which was expected to resolve this problem, served only to highlight the potential for mismatch between the vocabulary used by content producers (i.e, the authors of the item text) and content consumers (i.e., the search system users who are looking for this item).

As early as 1993, Bollmann-Sdorra and Raghavan [27] argued that the structure of the query space is very different from the structure of the document space, and that it makes sense to consider documents and queries to be elements from different term spaces. Cui et al. citecui2002probabilistic attempted to quantify the differences between the document-space and the query-space by measuring the typical similarities between document vectors and query vectors by using two months of query logs from the Microsoft Encarta search engine. They found very low average similarities between queries and document vectors, with the vast majority of similarities at less than 0.4 (using a standard a cosine similarity metric) and a mean similarity value of only 0.28 across the collection as a whole.

The lack of correspondence between the query space and the document space in traditional information retrieval settings is only amplified in Web search, for two basic reasons. First, as mentioned above, Web search queries tend to be short and vague, often including no more than 2 or 3 query terms [204]. This limits the opportunity for overlap between queries and documents. Second, Web content is unusually sensitive to changing trends and fashions, such that documents that were originally produced for one purpose (and indexed accordingly) might later become more relevant in an entirely different context, which may not be captured by their existing index terms.

Fortunately, while exacerbating the potential for vocabulary mismatches, the Web also offers a potential cure: the social information that started to accumulate rapidly on the Web has emerged as a viable bridge capable of connecting content (producer) and query (consumer) spaces. Indeed, certain unique types of social information are contributed mostly by document consumers who typically refer to a document from the consumer perspective, which may change over time. For example, the query terms used by users seeking a particular item may be useful as a means to augment automatically index terms or extracted item content during query-item matching; other forms of social information in addition to query terms may also be used. These approaches are often referred to as social *document expansion* or *social indexing*. A more flexible and, arguably, safer way to achieve this is *social matching*, by separating content indexing from social information related to a document while taking both kinds of information into account at the query/matching time. Over the last 15 years, an increased variety of social information has been used for social indexing and matching. The remaining part of this section briefly reviews the most popular approaches, organizing them by the type of social information they endeavor to leverage.

5.1   Link Anchors

The first kind of social information explored as a source for enhanced document representation was the anchor text associated with document links, which we have discussed in the previous section on query formulation, elaboration, and recommendation. When a link is created from one document to some target document, the link creator will typically use anchor text that references the target document in a particular way, revealing their own perspectives and biases. The intuition is that this anchor text will likely contain terms that will appear in queries for the target document that are submitted by future searchers who share a similar perspective. These terms may be absent from the original target document and therefore absent from its index; for example the linking document may be created at a much later date and serves to place the target document in an entirely new context that was unknown when it was originally created. These new anchor text terms may provide useful and novel indexing terms for the target document.

The use of anchor text in this way was pioneered by McBryan [135] in the World Wide Web Worm. A few years later, Brin and Page [36] used it as one of the foundations of their Google search engine, along with Web link structure, and Chakrabarti et al. [44] used it for compiling a list of authoritative web resources on a given topic. Following that, several research teams demonstrated the value of anchor text as a source of information in the context of site search and known-item Web search [53,153,198]. This work attracted broad attention to the use of anchor text and helped the approach to emerge as a mainstream information source for document expansion, which is today used by many (if not all) of the major Web search engines.

The study of anchor text also continued as a research stream. For example, Zhou et al. [215] compared two ways to combine anchor text with Web page content for site-level search: a combined representation with a single ranking and a separate representation with a combined ranking. Fujii [78] separately assessed the value of anchor text for answering informational and navigational queries and showed that anchor-based retrieval is especially effective for navigational queries. Koolen and Kamps [117] demonstrated the value of anchor text during ad-hoc search in sufficiently large Web collections. Ritchie et al. [166] expanded this line of work for research literature search by using the text associated with citations as a form of anchor text to expand document representations. Currently, anchor text is considered to be a gold standard for the social expansion of document representations, and it is typically used as a baseline against which to evaluate alternative social expansion approaches [14,87,123,151].

5.2   Query Logs

The second primary source for social document expansion, indexing, and matching is search engine logs. The idea to use queries for social indexing is a natural one, since queries are, by definition, an expression of a user's information needs, albeit a partial one that uses a particular vocabulary. It is

therefore an obvious step to enhance a document representation using query terms [14,114,170,171,209]; or the terms could be used to create an alternative query-based representation [159].

The main problem with this idea is that it is not always straightforward to create a reliable association between documents and queries by using query logs alone. First-generation research on query-document association focused on two natural ideas: associate a query with each of the top-N retrieved documents [163, 170], or with each of the selected documents on a SERP [82,114,151,159,197, 209]. Both approaches were found to be satisfactory for such tasks as document labeling or clustering, but were not able to match the quality of anchor-based document expansion.

A breakthrough in the use of query-based document expansion came once researchers recognized that query logs served to accumulate social feedback across multiple queries and even across multiple search sessions. For instance, users frequently have to go through a sequence of query reformulation, starting with one or more failed queries that provided no satisfactory results, and ending with successful queries that return some satisfactory results. Somewhat surprisingly, it is the failed queries that turned out to provide some of the best terms for document expansion, since these terms indicate the elements of the user's conceptualization of the document that are missing from the document's default representation. Moreover, another insight was that not every selected document truly matches user information needs, since clicks are made on the basis of SERP snippets that rarely give a complete picture of the document in question. Documents accessed at the end of the query session or extensively analyzed documents are more likely to represent the information that the searcher was looking for. The idea of using failed queries for document expansion was originally suggested by Furnas [80] and was based on explicit user feedback. The use of automatic query-based document expansion based on "smart" session mining was suggested by Amitay et al. [14] and independently explored in a few other projects [61,87]. Currently, "smart" mining of query sessions serves as one of the main sources for both social indexing and document ranking. It will be discussed in more detail below.

5.3   Annotations and Tags

While link anchors and queries can be considered as two indirect ways for information consumers to describe their conceptualization of documents, various page annotations and comments offer users a more direct way. The idea to use page annotations for better document indexing was pioneered by the Superbook system [165]. Dmitriev et al. [61] explored the use of annotation for document indexing in a much larger context of intranet search, where document annotations were collected using a special corporate browser, Trevi. This project is notable because it also attempted to incorporate all kinds of available social sources for document indexing, including anchors, query chains, and annotations.

The new stage of work on the use of social annotations for document indexing and search was associated with the rapid rise of social bookmarking systems

[86,91] in the middle of 2000. Social bookmarking systems, such as Del.icio.us[3] and Flickr[4], allowed users to collect and organize various kinds of resources, such as Web pages and images. Most of the systems allowed their users to add free-text annotations and more structured social tags to the bookmarked resources. The use of both kinds of annotations for document indexing and search within these systems was then quite natural. Indeed, in at least some social systems, such as Flickr, it was the only indexable information.

It was also shown how social annotations could be useful in the broader context of Web search. As early as in 2007, two seminal papers [21,150] proposed social search engines that used social annotations, accumulated by Del.icio.us, to build enhanced document representations and offer better retrieval and ranking. A direct comparison of queries, anchors, and annotations as a source of index data demonstrated that tags outperform anchors in several retrieval tasks [151]. This work was expanded in a sequence of follow-up studies that examined the idea using larger datasets [47,97] and experimented with building better tag-based document representations using more sophisticated tag analysis techniques [26,40,87,164,208,214]. Since this book considers tag-based social search as a separate social information access approach, we refer the reader to Chap. 9 for a more extensive discussion of this topic [149].

5.4   Blogs and Microblogs

Among other sources of social information that could be used for social indexing, it is also worth mentioning various kinds of blogs, and specifically microblogs, such as Twitter. Many blog and microblog posts include links to Web pages or mentions of other kinds of indexable objects (such as people, events, or products). Naturally, the content of these posts could be used as a source of information for enriching representations of these objects, just as link anchors can be used for regular Web pages. Lee and Croft [123] explored several kinds of social media in the context of Web search and referred to this kind of data as *social anchors* (similarly, Mishne and Lin [144] referred to Twitter text as *twanchor*).

Over the last five years, this source of information has been extensively explored in several contexts and using several kinds of blogs and microblogs. For example, short posts in Blippr [66] were used to build representations of users and movies, while tweets and Facebook posts were used as a source of terms for Web document representation [11,87,144,157,168,191] and user profiling [3]. In addition to their text content, which is similar in nature to anchor text and annotations, blog and microblog posts offer an important temporal dimension. The natural tendency of blog posts to refer to recent events and actual documents could be used for just-in-time crawling and document expansion with reference to time [45,62], which in turn, is valuable for recency-based ranking.

---

[3] https://del.icio.us.
[4] https://flickr.com.

6 Ranking

When information retrieval systems – from classic search systems to modern web search engines – receive a user query, they have to perform two important tasks: (1) retrieve documents that are relevant to a query (this is also known as query-document matching); and (2) rank these documents based on their relevance to the query. While different models of information retrieval offer a variety of approaches to query-document matching, all these approaches are based on some form of document indexing. As a result, from the perspective of using traces of past users, social matching is equivalent to social indexing, which was reviewed above. Ranking, however, and the subject of this section, uses social traces in significantly different ways.

Ranking search results has been always an important problem in information retrieval. The ability to correctly rank search results has been long advocated as a benefit or modern vector and probabilistic models of information retrieval, in comparison with alternative set-based Boolean models [118]. The goal of ranking is to sort results in descending order of their relevance to the user's true information need. Getting this right is critically important, given that users have a limited practical ability, or willingness, to examine and access many results; if the right result isn't near the top of the ranking, then a user is unlikely to be satisfied by the results of their search, and may even abandon their search session altogether. Eye-tracking studies [88,105] have shown that results located near the top of the ranked list have a much higher probability of being noticed and examined by the user. This probability rapidly decreases as we move down the ranking. Indeed, [113] demonstrated how the judgments of searchers are significantly influenced by the rank position of a result, as users tend to inherently trust in the ability of a search engine to correctly rank results by their relevance; so much so that manipulating lists of results to position relevant results at the end of the ranking failed to attract user attention. In other words, it is not sufficient for a search engine just to have good results somewhere in the returned list (as ensured by matching). It is equally important to position the best results at the top of the ranking if the user is to notice and attend to them.

Traditional ranking approaches in both vector and probabilistic models of information retrieval are based solely on the content of documents (e.g. web pages). Similarly, primary ranking in early Web search engines and search tools was fully based on document content. While this approach works relatively well for smaller-scale search systems, the world of web search quickly outgrew this approach. With the rapid growth of the Web, almost every query produced an increasingly large number of matched Web pages. In this situation, content-based ranking routinely failed to identify the most relevant documents. It was a perfect context to start harnessing other sources of information for ranking, including the social information that naturally accumulates from traces left by past Web searchers. As mentioned in the introduction, the idea of using social traces for ranking was pioneered in parallel by two novel search engines of the day: Google [36], which leveraged page links, and DirectHit [202], which leveraged user page

exploration behavior [57]. This seminal work encouraged a large volume of follow-up research on "social ranking".

The first two subsections of this section review the use of social traces for the "primary rankings" produced by retrieval systems and engines, as well as search components of various Web systems with the use of data available to these systems. In this respect, primary ranking approaches differ from various re-ranking and other result-promotion approaches produced by third-party systems, which have no access to search engine data, but which leverage other valuable sources of ranking data. These re-ranking approaches are reviewed separately in the following sections.

6.1   Web Links for Primary Ranking

Web links were the first kind of social information leveraged for ranking purposes. The authors of link-based ranking approaches argue that the availability of links is a feature that distinguishes valuable pages and suggest different ways to promote pages with more links in the ranked list. The first approach, suggested by Carriere and Kazman [43], didn't differentiate incoming and outgoing links and used only local link analysis to re-rank results based on the total number of links (both incoming and outgoing). However, it was the other two link-based ranking approaches—PageRank [36] and HITS [84,115]—that demonstrated the potential power of link-based ranking. PageRank and HITS are frequently referred to together because they were introduced almost at the same time and inspired by similar ideas. Both approaches focused on incoming links and argued that incoming links created by multiple authors signify page value or "authority" [115].

The key difference between PageRank and HITS was the kind of link analysis that was performed. PageRank adopted ideas from citation analysis: papers with a larger number of citations (i.e., incoming links) are more important, but citations by more important papers weight more than citations from less important papers. To combine these two basic ideas, PageRank introduced a recursive calculation of page importance (called PageRank) based on a random walk in the whole Web linking graph [36].

HITS was also based on the results of an analysis of the Web hyperlink structure [44] to identify so-called hubs (sources of outgoing links) and authorities (sources of incoming links) in the web graph. However, the network analysis of HITS is query oriented. It is focused on the top ranked search results and their parent and children pages. Both approaches are explained in more detail in Chap. 8 of this book [92], but suffice it to say that PageRank and HITS inspired a generation of network-based ranking approaches whose reach often extended well beyond web search. Some of these approaches, as they relate to social search, are reviewed below with additional coverage in Chaps. 8 [92] and 9 [149] of this book.

Link anchors, the "other side" of Web links, emerged as a popular source of information for ranking at the same time as link topology [36,44]. For example, the Clever system [44] enhanced the original HITS approach by using an analysis of anchor text to differentially weight the contribution of individual links.

The simplest ranking, using anchor text, can be implemented by re-using the regular ranking approach, but using the anchor text instead of the original content [53,117] or the anchor text merged with the original content [215]. Such approaches can be effective in certain settings. For example, Craswell et al. [53] demonstrated how re-ranking based on link anchor text is twice as effective as ranking based on document content for site finding.

More advanced ranking approaches use separate content-based and anchor-based document representations. The main challenge that remains is to find an appropriate way to combine ranking information produced by these two independent sources. One approach is to blend together the independent rankings; see [198,205,215]. Another approach is to incorporate both sources of information into a single probabilistic ranking model; see [78,153,198].

6.2   Query Logs for Primary Ranking

The idea of using query logs for ranking is based on an assumption that a click on a specific link in a ranked list of results, generated by a specific query, is an indication of the *general importance* of a page behind the link as well as its *relevance* to the query. With this assumption in mind, every click can be considered as a vote in favor of the clicked page, and a query log can be treated as a collection of votes that can be used for ranking. The problem is that this voting data is very noisy, for several different reasons. First, an isolated click on a link is not a reliable vote. As mentioned earlier, link selection is usually based on a small result snippet that might be misleading. Second, as shown by the studies mentioned above, a chance to click on a link depends not only of the true relevance of the document, but also to a large extent on its position of the link in the ranked list; as shown in [113], even highly relevant links tend to be ignored if they are ranked at the bottom of a result list. Third, different users might need different information in response to the same query. Thus, using "click votes" from one group of users could harm the ranking for another group.

Due to the noise in the data, the use of query logs for ranking present a considerable practical challenge, as demonstrated by the rather limited success of Direct Hit [202], the first major search engine to use query session clicks as relevance votes for ranking. Direct Hit did attempt to handle the noise in click-through data by accounting for time spent by a user on the actual page, as well as the position of the clicked link in the ranked list [57]. However, its use of query log data did not prevail and could not compete with the link-based ranking system popularized by Google. More recent research using query log-based ranking focused on addressing the problem of noisy click-through data. For example, clustering can be used (query-centered, document-centered, and user-centered clustering) [17,63,82,176,209] to reduce the noise from isolated "votes". A switch from absolute to the relative treatment of click-through votes [104,106,162], and later, more advanced click models [46,89,158] consider the noise associated with different positions of clicked documents in the ranked lists displayed to users. Using a broader set of features in combination with more advanced machine-learning approaches also led to considerable quality improvements in log-based ranking approaches [5].

The post-Direct Hit research on log-based ranking can be divided into several strands of research. One popular approach explored various graph-based methods to better leverage connections between the queries and pages accumulated in the log. Another popular strand of research harnessed machine-learning techniques by treating search ranking as a type of *learning-to-rank* task. The benefits of this approach include its ability to consider richer feature sets extracted from query logs. Yet another strand of research looked at grouping users into more coherent clusters to generate social ranking within a cluster; namely, using only the click data of like-minded users. This led to good results without the need to use more advanced document representations or ranking approaches.

Graph-based approaches to query log analysis condense the social wisdom represented in the click log in a concentrated graph form that can be explored in several ways. For example, [82, 209] focused on using social data as an additional source of *features* (known as click-though features) to improve document representation for improving matching and ranking. Both groups of authors attempted to address two known problems with click-through indexing: (1) the overall *reliability* of click-based evidence mentioned above; and (2) the *sparsity* of click-through data. While popular queries and pages appear in many query-click pairs, many queries and pages receive too few associations in the logs to be used reliably. To solve both problems, Xue et al. [209] suggested moving from the level of isolated click "votes" to clusters of similar queries and clicked pages. By representing query log data as a bipartite graph, they applied an iterative reinforcement algorithm to compute the similarity between web pages and queries, which fully explore the relationships between web pages and queries. After the similarity between web pages are computed, two similar web pages can "share" queries; for example, queries associated with one page can be assigned to similar pages as extra metadata.

Gao et al. [82] followed the same idea for more reliable query-based indexing based on the expansion of sparse click-through data. Defining this approach as "smoothing", this paper explored two such smoothing techniques: (1) query clustering via Random Walk on click graphs; and (2) a discounting method inspired by the Good-Turing estimator.

Elsewhere, Craswell and Szummer [54] used a random walk approach on a graph-based representation of a click log to produce a probabilistic ranking of documents for a given query. They explored several types of Markov random walk models to select the most effective combination of parameters. Poblete et al. [160] attempted to combine the benefits of link-based and log-based ranking by also applying a random walk model to the integrated graph, which included both static hypertext links between pages and query-page links reconstructed from the log.

The idea of considering log-based ranking as a machine-learning problem was introduced by Joachims [104] as a learning-to-rank problem. While learning-to-rank has been used in the past with explicit feedback data, the intrinsic noise of log data, based on implicit feedback, made it a challenge to use conventional learning-to-rank approaches in this context. To overcome this challenge, Joachims [104] considered more reliable *relative* implicit feedback that could be

obtained by analyzing user clickthrough patterns in the log; for example, a click on the third link in SERP while ignoring the second link indicates that the third document is more relevant to the query than the second, from the user's point of view. In follow-up work, Radlinski and Joachims [162] expanded the original approach by considering multi-query search sessions and using more complex log patterns that allowed for engaging more broadly across sessions with more reliable social wisdom not available within a single query.

Following this pioneering work, learning-to-rank gradually emerged as one of the key approaches to incorporate implicit feedback accumulated in query logs into the ranking of search results. More recent work in this area advanced earlier research by exploring different learning-to-rank approaches and alternative ideas to reduce the impact of log noise. For example, Agichtein et al. [5] re-examined the use of absolute implicit feedback, originally explored in Direct Hit [57]. To deal with the noise, the authors extracted a much larger set of feedback-related features from the log and fused them with a RankNet supervised algorithm to learn a ranking function that best predicts relevance judgments. In particular, they aggregated click frequencies to filter out noisy clicks and used a deviation from expected click probability to more accurately consider the value of a click at different positions in the SERP. Dou et al. [63] followed suit [104] by using the relative feedback approach, but adopted the idea of feedback aggregation and RankNet fusion suggested in [5]. Instead of relying on single feedback cases, the authors aggregated large numbers of user clicks for each query-document pair and extracted pairwise preferences, based on the aggregated click frequencies of documents.

Another important advance in the use of machine-learning techniques for ranking was the introduction of formal *click models*. Click models attempt to model user search behaviors to address rank position bias in a more holistic way than pairwise preferences. Once learned from data, click models can produce a data-informed ranking of web pages for a given query. The idea of learning a model of user search behavior was introduced by Agichtein et al. [6]. Early exploration of several click models was performed in 2008 by Craswell et al. [55] as well as Dupret and Piwowarski [64]. In the following year, several papers [46,89,158] pioneered the use of probabilistic graphical models (i.e., Bayesian networks) to represent and learn click models. In a broad stream of follow-up work, the use of click models and learning-to-rank approaches for log-based ranking have been expanded and refined in many different ways. For example, recently, Katarya et al. [112] extended the approach of [89] using interactive learning. Wang et al. [193] captured non-sequential SERP examination behavior in a click model. Borisov et al. [28] experimented with neural networks as an alternative basis for click models, and Wang et al. [194] explored learning-to-rank approaches in the sparse context of personalized search.

The use of user groups to cluster social wisdom accumulated in query logs into smaller and more coherent communities was pioneered in several projects [10,77,111,179,180]. Some of these projects suggested using existing social groups, such as the implicit communities harnessed by [179,180] based on the origin of search queries. Others attempted to match users and form

groups dynamically [10,111]. For example, Almeida and Almeida [10] suggested a "community-aware" search engine, which used a graph-based representation of a query log to identify multiple communities of interest and associate these communities with documents. The engine then helped users in the search process by matching them to existing communities and providing community-biased ranking by fusing community-based relevance and content relevance. The use of static groups based on demography and other characteristics has also been extensively explored. For example, Teevan et al. [189] demonstrated how the value of "groupization" depends on the type of the group and query category. Work-related groups were found to have no cohesion on social queries, while demographic groups have no cohesion on work-related queries. The author also suggested an alternative approach to group-based ranking, by applying ideas from personalized ranking at the group level. More details on several group-based social search approaches can be found in the Sect. 10.

6.3   Using Browsing Trails and Page Behavior for Primary Ranking

Browsing trails can offer several types of social wisdom that can be used to improve rankings. First, browsing trails provide some evidence of page value and importance through various implicit indicators, such as reading-time, or, with appropriate instrumentation, within-page scrolling, and mousing (for desktops) or "fingering" for mobile browsing [93].

Second, assuming that users have coherent goals when browsing, the sequence of pages followed can help to create associations between pages. This information can be used to enhance network-based algorithms for estimating page importance, such as PageRank. For example, while the fixed links between pages, used by PageRank, already influence navigation pathways, without real user information, PageRank has to assume an equal probability of navigation from one page to all outgoing links. In reality, some pages may be much more important than others, and browsing data could help to capture this information.

The easiest way to collect browsing information is through browsing logs, which are maintained by most Web sites and information systems. However, due to its localized nature, this information can be used only for improving ranking at the site level or for system-level searching. To be useful for Web search, user browsing needs to be captured beyond a single site, which requires some client-side or server-side instrumentation. Client-site instrumentation, such as user agents or browser plug-ins, originally developed for personalized search [83] can be used to collect user browsing data across multiple sites, and usually on a deeper level than site logs. On the other hand, server-side instrumentation, such as the link-following technology used by most modern search engines, can register a user's continuous browsing session starting from an instrumented page or SERP.

Each of these approaches has its own pros and cons. Browser-level data facilitates navigation capture in a much broader context, but because browser-based instrumentation must be installed, tits adoption has been limited. In contrast, link following is now commonplace, but it typically includes only short, truncated browsing segments.

Site-level browsing data, which is the easiest data source to collect, was also the first to be explored for ranking. Xue et al. [210] attempted to improve the PageRank approach by calculating page importance using "implicit links" instead of explicit hypertext links between site pages. The authors argued that implicit links exist between pairs of pages if there are frequent transitions between them (as mined from browsing logs), and that such associations are more valuable in a "small web" context, where explicit links might not reliably reflect main navigation pathways. Their study demonstrated that a PageRank-style algorithm, based on implicit links, provided the best ranking for site-level search, as it outperformed regular PageRank, HITS, and other approaches. This work was followed by a number of like-minded attempts to improve PageRank by using site-level browsing data. For example, Eirinaki and Vazirgiannis [65] suggested an approach to bias the PageRank calculation to rank pages more highly that previous users visited more often. Guo et al. [90] took time spent on a page into consideration and suggested an approach to bias the PageRank calculation so as to give a higher ranking to pages that previous users visited more frequently and for a longer period of time.

The use of post-search browsing trails, collected by the search engines and through the link-following approach, has been investigated by Bilenko and White [24] in the context of their research on recommending search destinations [200]. The nature of post-search trails is different from site-level browsing trails, since each trail originates from a specific query. As a result, the use of this data is more similar to the use of query sessions, rather than the use of general browsing trails. Each page selection and its dwell time indicates a page value for the original query, rather than its general importance. In their work, Bilenko and White [24] explored several approaches to process and use trail data and confirmed that post-search browsing behavior logs provided a strong signal for inferring document relevance for future queries. In particular, their work indicated that using full trails can lead to better results and that using the logarithm of dwell time is the best approach when using this source of information.

More recent work has attempted to go beyond site-level and post-search trails by using server-side and client side instrumentation to collect a broader set of trails that could be used to improve regular Web searching. The most well-known of the these approaches is BrowseRank, which was introduced in [130]. The authors suggest using a navigation graph, augmented with time data in place of a "timeless" link graph, used by PageRank. Instead of the usual approach, which applies PageRank on the top of browsing data, they used a continuous-time Markov process on the user browsing graph as a model for computing the stationary probability distribution of the process as page importance. The authors argued that this approach could leverage this new kind of data better than the original PageRank, a discrete-time Markov process on the web link graph as a model. In a follow-up paper, [131] suggested and evaluated several other approaches for browsing-based estimation of page importance. Using a large "link following" dataset collected by a major search engine, the authors demonstrated that BrowseRank significantly outperforms PageRank and other

simple algorithms. While this type of data, as discussed above, might not be the best match to evaluate BrowseRank, a follow-up work [190] demonstrated that BrowseRank also outperforms PageRank in the context of site-level search. Zhu and Mishne [216] suggested an alternative approach to calculate browsing-based importance using both click order and page dwell time. The authors evaluated ClickRank using a large volume of user browsing logs collected from the Yahoo! toolbar and demonstrated that this method outperforms both PageRank and BrowseRank; see also [16].

6.4   Reranking and Recommending Web Search Results

In this section, we consider how social signals (which might not be available to the search engines and thus cannot be used for primary ranking) might be used to better bring search results to the attention of the end users; for example, understanding the searcher's community may help to influence search results. We will look at how such information can be used to re-rank an original set of results or used to insert new results into an original result list.

To begin with, the I-SPY [178–180] system provides an early example of a form of social ranking or, more correctly, social *re-ranking*, since I-SPY re-ranks a set of results provided by an underlying search engine. We discuss I-SPY in further detail as a later case study, but for now, it is sufficient to say that I-SPY uses click-through data re-ranking based on community interests. Briefly, result pages that have been frequently selected in the past for a query that is the same or similar to the target query are considered to be more relevant than result pages that have been selected less frequently. I-SPY calculates a relevance score based on click-through rates and weighted by query similarity, and uses this score to promote and re-rank results. Moreover, I-SPY leverages click-through data that originates from a community of like-minded searchers by sourcing its search queries from search boxes placed on topical websites. For instance, an I-SPY search box on a motoring web site is likely to attract queries from motoring enthusiasts, and their click patterns will help to differentiate and promote pictures of cars for a query like "jaguar photo", instead of pictures of large cats.

This combination of click-through data and community focus is closely linked to the idea of trust, expertise, and reputation as a ranking signal, which has become increasingly important in recommender systems. Intuitively, not all searchers/users are created equally. Some will have more or less expertise on certain topics and may make better recommendation sources as a result. This idea was first explored in the work on *trust-based recommender-systems* [134,152], in which they explicitly modeled the trustworthiness of users based on how often their past predictions were deemed to be correct. By prioritizing users who are more trustworthy in general or in respect to a particular topic, it was possible to significantly improve recommendation quality. Similar ideas were explored by expertise-based recommender systems that attempted to model user expertise [12]. In parallel, these ideas were adapted for community-based web search in

the work of [32,35] with search results ranked based on the *reputation* of the users who previously selected them for similar queries.

In this sense, we can view reputation as yet another type of social signal that arises from the result selection behavior of searchers, and how this selection behavior helps others in the future. Users who search frequently on a given topic, and whose selections are recommended to, and re-selected by, other searchers can be usefully considered to have a higher reputation on such topics, in comparison to other users who rarely search on these topics, or whose selections are rarely re-selected when recommended. This idea informed the work of [138,183], which proposed an explicit model of search collaboration and considered a variety of graph-based reputation models for use in social search. Briefly, different ways to distribute and aggregate reputation were evaluated as ranking signals, with significant precision benefits accruing to reputation-based ranking compared to alternative ranking approaches. Similar ideas are explored by the ExpertRec system [188], which modeled user expertise by observing search sessions and subsequently promoted search results preferred by experts.

The previous examples are all examples of ranking/re-ranking an original result list using social signals, but we are not limited to the original list of results. Many researchers have looked at ways to make new recommendations by adding novel results into an original result list. For example, White et al. [200,201] describes a technique to exploit the searching and browsing behavior of many users to augment regular search results with popular *destination pages*, which may be located many clicks beyond the original search results. This offers the promise of a significant time-saver for searchers by short-circuiting long navigation trails from SERP results to a final destination. They focused on navigation trails that began with a search query submitted to a popular search engine. Each trail represented a single navigation path (that is, a sequence of clicked hyperlinks) from a search results page to a terminating page, at which point the user went no further. These terminating or *destination* pages are then indexed using their original query terms so that they can be recommended and added to result lists for similar queries in the future. In this way, this approach is related to some ideas presented earlier on the topic of social indexing, in the sense that these destination pages are effectively being socially indexed based on the queries that have led to them.

In fact, authors describe a number of different ways to harness these destination data. For example, in one interface, for a given search query, a set of destinations are presented that have been frequently navigated to after similar queries. In another interface, searchers have recommended destinations that have been navigated to, not for the current query, but for typical follow-on queries (from the same search session). The study found that systems that offer popular destinations lead to more successful and efficient searching compared to query suggestion and unaided Web search but that destinations recommended based on follow-on queries had the potential to be less relevant to searcher needs.

Likewise, the HeyStaks collaborative search system [181,182] also attempts to add new pages to an existing SERP; it too is described in more detail as

a later case study. Briefly, like I-SPY, it organizes searchers and searches into communities, but this time, they are based on user-defined search topics and interests. These interests are captured in so-called *search staks*, which act as repositories of past search histories that encompass a variety of search-related data, including queries, result selections, result ratings, and tags, among others. Searchers continue to search using their favorite search engine, but by using the HeyStaks browser app, their searches are recorded and aggregated in staks that they have created or joined—and the search results that they receive from an underlying search engine are enriched with additional recommendations from relevant staks. For instance, if a user is searching for "ski chalets" on Google, as part of a "ski vacation" stak, then, in addition to Google's results, they may see other results that have been found by other members of the stak, either for similar queries or from different search engines. These recommendations are selected and ranked based on a weighted scoring metric that combines query similarity, selection popularity, and other social signals (rating, tags, and so on).

6.5 Ranking and Re-ranking with Social Media Data

Social media data plays a dual role in social search. In its *local* role, socially posted data are used to perform search within the very social system where these data were posted; for example, searching for images or bookmarks in a social sharing system or searching for people in a social network, where comments, tags, and other user-contributed information are vital for finding relevant information. However, more and more frequently, social data are used more *globally* – to improve search ranking beyond the host system. For example, posts in a collaborative Web bookmarking system like Del.icio.us or tweets of Web pages could improve Web search, while posts in a research paper bookmarking systems such as CiteULike[5] or blog posts mentioning research papers could improve searching in academic search systems, such as Semantic Scholar[6]. The use of social data for searching within a social system is naturally a component of primary ranking, since these data are directly available to the social system search component. In contrast, the majority of work for using social data in a more global search context is to perform a social re-ranking of results that are returned by a general search engine. The use of social data could considerably improve both local and global rankings in both primary ranking and re-ranking settings. Moreover, the approaches to improve ranking using social data are quite similar in all these contexts. Here, we review the use of social data for both primary ranking and re-ranking with a focus on the approaches taken, rather than their application contexts.

Using social data could improve ranking in internal and external search systems in three ways. First, as reviewed in Sects. 5.3 and 5.4, a resource (i.e., a paper, a Web page) shared in a social system is usually augmented with comments, tag, or at least some surrounding content (like in tweets and blogs).

---

[5] http://www.citeulike.org.
[6] http://www.semanticscholar.org/.

This content complements the content of the original resource and can be used to improve both matching and ranking in the same way as it can be done by using query texts or anchors.

Second, the very fact that a resource was shared on social media and the scale of this sharing (i.e., the number of tweeting or bookmarking users) could be interpreted as the sign of value, and used for promoting shared content for both internal and external search. Simple approaches based on this idea follow the DirectHit path that treats every sharing event as a vote. For example, in an internal Twitter search, the number of re-tweets is treated as a sign of a tweet's importance, which can be used for tweet ranking [7]. For external ranking, blog posts could be treated as votes for various news articles and leveraged to produce better news article rankings [136]. Similarly, sharing a research paper in social bookmarking system could be treated as a vote that the paper is worth reading [103]. More sophisticated approaches adopt ideas from PageRank to extract importance data from the complex network of connections between users, shared items, and tags. For example, a network of tweets and tweeters could be used to estimate the *authority* of each tweeter and give a higher weight to more authoritative tweets or re-tweets [157, 168].

Finally, resources and posts explicitly shared by individual users provide a much more reliable indication of user interests, as compared with traces left while searching or browsing. A model of user interests extracted from shared social data could be used to further improve ranking by making it more personalized; a considerable share of work on tag-based search was at least partially focused on personalization [1, 40, 126, 150, 208].

Among all kinds of social media data, tags were both the first and the most popular source of information for improving search ranking. The work on tag-based ranking technologies has started with research on improving ranking within social tagging systems. Several advanced approaches for tag-based ranking were reported as early as 2006 [100, 206]. Subsequently, several papers adopted different approaches for using social tagging data to improve the ranking of Web search results [21, 150, 211]. While most early papers explored straightforward approaches for integrating tags into rankings, such as simple "vote counting" to estimate page importance [211] or traditional vector model to measure query-document similarity [150], a increasing share of work focused on more sophisticated uses of tags for ranking.

Network-based ranking approaches could be considered to be the most popular group of such advanced approaches. Most are motivated by Google's PageRank and try to adapt it for social systems. While PageRank and HITS tried to leverage the information encapsulated in a complex network of interconnecting Web pages, the ranking approaches in social tagging systems attempted to leverage information hidden in the even more complex network that is formed by users, tags, and resources. While each Web link establishes a new edge in a network of pages, each tagging event (a user U tags item I with a tah T) creates a new *hyperedge* that connects U, I and T, which could be alternately represented with three edges $U - I$, $U - T$, and $I - T$.

Hotho et al. [100] were the first to suggest an advanced network-based ranking approach, *FolkRank*, to extract information from a tagging network for the internal ranking of social bookmarks in BibSonomy; see Chap. 9 of this book [149]. Bao et al. [21] used similar ideas to leverage the information encapsulated in a social bookmarking system to improve external result ranking. The authors suggested two network-based ranking approaches to optimize Web search. First, they proposed a novel similarity-based ranking algorithm *SocialSimRank* based on the idea that social tags can act as a type of metadata for the similarity calculation between a query and a web page. Second, they described an *importance ranking* approach, *SocialPageRank*, inspired by PageRank, to estimate the popularity of a page based on its tag history. In a similar attempt, Abel et al. [1] suggested a graph-based ranking approach that was inspired by HITS, rather than PageRank.

Other early approaches focused on exploiting tags to build a better semantic representation of the document space (and the user space), which could be leveraged for better ranking. A range of formalisms were used for the included Semantic Web [206], such as topic models [126, 208] and language models [215]. Some radically different approaches were also introduced by researchers from other communities. For example, Zanardi and Capra [212] used ideas from collaborative filtering to combine both social and personal data for tag-based search ranking, di Sciascio et al. suggested a user-driven approach to fuse traditional query-based relevance with tag-based relevance [172]. Over the last 10 years, the research on tag-based ranking expanded into a distinct research direction with papers exploring different ways to leverage graph information, extract semantics, or combine these with other approaches. An extensive review of tag-based ranking for both search in social systems and Web search is provided in Chap. 9 of this book [149].

7   Resource Presentation – Social Summarization

Almost all web search engines follow a similar pattern of result presentation, the so-called *10-blue links* approach. That is to say, results are presented as a simple list of URL links, with each result made up of a title, a URL, and some suitable *snippet* to summarize the result. These snippets play a vital *sense-making* role, as previously discussed. They help searchers to efficiently make sense of a collection of search results, as well as to determine the likely relevance of individual results. Snippets are typically extracts of content from the corresponding web page. They are also typically query-sensitive, in the sense that the selected text is chosen because it is relevant to the current query. For example, the most popular approach is to select text in the web page that contains some or all of the query terms; however, it is not easy to decide which parts of the page and which keywords will be most helpful for the users to recognize a page as a true match to the query. This is one resource presentation task where social data can help.

The idea of using social data to generate page snippets was first suggested in [170]. The authors proposed to use past queries as a component of page snippets

by arguing that they offered the best characterization of a document from an end-user perspective. To generate these kinds of snippets, they used query logs to incrementally build associations between queries and their selected documents. However, these past queries were not truly integrated into a presentation snippet, but instead were added after the snippet summary, and served more as an augmentation rather than a summary. It also provided a one-size-fits-all approach to snippet generation that was not adapted to a specific user or context.

The work of [30,31] resolved both of the shortcomings mentioned above and demonstrated an example of using social signals to guide the generation of "smooth" snippets in a *community-based search context*. They describe an approach to personalizing snippets for the needs of a community of like-minded searchers. Their approach, called *community-focused social summarization*, uses community search behavior – query repetition and result selection regularity – as the basis for generating community-focused snippets. It uses the standard, query-focused snippet generation technique of the underlying search engine, but each time a result is selected, for some query, the corresponding snippet is recorded so that, over time, popular results within a community of searchers come to be associated with a growing set of (possibly overlapping) snippets, based on the queries for which they were selected. Sentence fragments can be scored based on how frequently they recur in snippets, and a social summary can then be generated from the most popular fragments and weighted according to how similar the current query is to the query that resulted in a particular snippet fragment. In this way, highly personalized, query-focused snippets can be generated at a given length specification. In tests, these snippets prove to be superior to those produced by alternative summarization techniques, including those that involve sophisticated natural language processing techniques; see also Alhindi et al. [9] for related ideas on generating group-adapted page snippets.

8 Augmenting Search Results: Annotations and Explanations

In the previous section, we focused on the presentation of search results – generating snippets that present each search result to the user – but that is not the only way to improve the presentation of results for the benefit of searchers. In this section, we consider SERP *augmentation*: the different ways that the SERP can be *decorated* with additional information to assist the searcher. SERP augmentation is, in some aspects, similar to the link augmentation in social navigation reviewed in Chap. 5 of this book [71].

There has been a long history of research into how search results might be presented to users so that they can better understand their relevance. For instance, TileBars [96] introduced a visualization paradigm that offers an explanation of not just the strength of the match between the users query and a given result page, but also the frequency of each term, how each term is distributed in the text, and where the terms overlap within the document. This approach provides the user with additional explanatory information that can help them to come to

a decision as to whether or not each page is relevant to their information needs. In a social search context, explanations can be derived from social data, such as the interactions of other users as they search or explicit social links.

8.1   Query Logs

Earlier, in Sect. 6.2, we discussed the role of query logs for primary ranking, but sequences of queries can also be used for the purposes of result annotation and explanation. As mentioned in Sect. 7, in 2002, Scholer and Williams [170] suggested augmenting document presentation with past queries to result in this document selection. The I-SPY collaborative search engine [178–180], mentioned in Sect. 6.2, and discussed later in Sect. 10.2, is another early example of the use of social annotations, and query annotations in particular. For example, Fig. 1 shows a sample SERP annotated with additional information that is based on how relevant a result is to the searcher's community, the recency of result selections, and the availability of related queries that have caused a result to be selected.

This work was subsequently extended by the *SearchGuide* project [50, 51], which built on-top of I-SPY to provide a even richer augmentation interface [50, 51], and which will be discussed in further detail in Sect. 9. Briefly, for now, SearchGuide (see Fig. 3) provides an enhanced interface to brings similar types of social annotations to bear on regular content pages, in addition to SERPs.

Using query logs to augment search results has also been explored in several other projects. For example, Lehikoinen et al. [124] demonstrated how past users interacted with search results in the context of meta-search in P2P networks. Elsewhere, Wang et al. [195] used query log data to build a topic map of a search space and used this map to augment search results with a set of related topics as the basis for further exploration.

Fig. 1. An example of the augmented SERP used by a version of the I-SPY system that shows a set of results that are augmented with additional icons to reflect how relevant a result is to the searcher's community, the recency of result selections, and the availability of one or more related queries for a result.

8.2   Browsing and Annotation

Research on using navigation and annotation activity of past users to augment search results was performed mostly in parallel with the work on using query sessions, though it was originally motivated by the ideas of social navigation rather than social search. The social navigation research stream in early 2000, as reviewed in Chap. 5 of this book [71], aimed to guide users through an information space by augmenting navigation links with "social wisdom" extracted from the history of past navigation. The main impediment to using this approach in a search context is selecting a meaningful subset of users, such that their browsing data will be helpful in selecting query results. While the use of query sessions for augmentation facilitates focusing on users who issued the same (or similar) queries, and thus are likely to have a high probability of similar information needs, browsing traces can come from all kinds of users. The majority of browsers are likely to have highly different needs from those who issued a specific query. As a result "everyones" browsing behavior will hardly help in a specific search context. However, if a community of like-minded users who share similar goals can be identified, then their browsing behavior may help other community users with their searches.

This basic idea was originally developed in the Knowledge Sea project [37], which leveraged the browsing behavior of users taking the same course in an e-learning context. Knowledge Sea attempted to use social navigation to support several kinds of information access to educational content, in the form of a collection of online textbooks, including browsing, search, and information visualization. While the collection of textbooks can be accessed by students who take different college classes (and thus have different information needs and priorities), Knowledge Sea considered students within the same class as sufficiently like-minded to apply social navigation support. In the context of search, social navigation support was provided by augmenting SERP links with visual cues, which reflected how much each search result had been read and annotated by students *of the same class* (Fig. 2). The browsing-based visual cue was shown as a blue human icon on a blue background. The density of color indicated the cumulative amount of page reading by the user (figure color) and the class (background color). For example, a light icon on a dark background indicated pages frequently explored by other students in class, but so far ignored by the current user. The annotation-based cue used a yellow background color to indicate how many annotations made by students in class each SERP page has, and also indicated how positive were these annotations.

This approach made it easy to recognize pages that members of the class found to be useful, especially if the user had so far paid little attention to them. Several rounds of studies with Knowledge Sea demonstrated that pages with high levels of class browsing and annotation behavior were especially appealing to the students: the presence of annotation considerably affected their navigation and reading choices [37]. Moreover, in the context of search, user result-selection and reading-time data demonstrated how higher levels of browsing behavior, indicated by visual cues, offered a stronger signal of search result relevance, compared to being among the top three results in the ranked list [8].
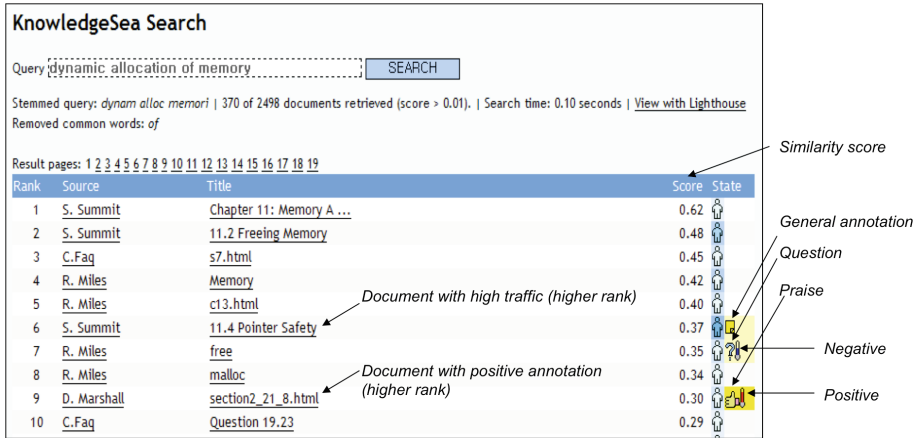
Fig. 2. Augmentation of search results with visual cues to indicate browsing and annotation behavior of students in the same class. Darker blue color indicated pages frequently selected and read by the class, while yellow-orange color indicates pages annotated by users in class (Color figure online)

The idea of group-level augmentation of search results with browsing data was further explored in the ASSIST project [72,76]. Using group-level data collection, ASSIST integrated the ideas of SearchGuide and Knowledge Sea and used both search and browsing data to augment the presentation of search results. This project is reviewed in more details in Sect. 11.1. The use of annotation data to augment search results has been further explored in an exploratory search context, in [70]. In this work, browsing and annotation behavior produced by a group of people working on a set of similar information analysis tasks was used to augment each search result to stress the extent to which the retrieved document has been read and annotated by other users in a group. In addition, the SERP was also augmented with a social "map" icon to show the location of documents with the group footprint among the top 100 retrieved documents.

8.3 Social Media and Social Links

Augmenting search results with data from various social media systems (i.e., bookmarking, tagging, blogging, and microblogging data) may help with SERP augmentation in several different ways. First, since document sharing in social media is a sign of that document's importance, augmenting a SERP document with information about the number of shares (i.e., posted to a bookmarking system, shared in a microblog) could help to guide users to prominent documents. Of course, users who bookmarked or otherwise shared documents may have done so for a variety of reasons, which may have little connection to the goals of their search or the goals of the new searcher. Consequently, to make this idea practical, social data should be carefully collected from a subcommunity of like-minded users; for example, users within the same enterprise, as suggested in [13].

Moreover, additional information provided by users while sharing could also be used to augment shared documents with "social" descriptions that could further help during selection decisions. For example, social bookmarking data could be used to augment search results with related tags [13,109,211], while Twitter data could be used to augment search results with mentioning tweets [157].

Since almost every instance of information sharing in social media is associated with an authenticated and identifiable user, this opens up the opportunity to augment search results with information about the users themselves. This approach was originally used in social bookmarking systems, where it was natural to show who bookmarked a document, regardless of the context in which the document was shown. First attempts to apply "people augmentation" in a broader setting were performed by several IBM researchers in the context of enterprise search, which leveraged information from enterprise social systems such as IBM social bookmarking system Dogear [141]. For example, Millen et al. [142] explicitly injected data retrieved from Dogear (and augmented with some people information) into all search results. Amitay et al. [13] used Dogear and the IBM blog system BlogCentral for a more elaborated augmentation of search results, using a list of people who shared the retrieved documents.

As with other types of social media augmentation, the main challenge with "people" augmentation is ensuring that there is a match to the current searcher's needs: the reasons for one person to share a result might be completely irrelevant to a future search context, and while simple people augmentation can work well in a narrow enterprise search context [13,142,147] its value quickly decreases in broader settings. To address these issue, IBM researchers explored the use of social networks by focusing on the social connections of the searcher [173]. In an enterprise context, social connections are usually professionally oriented and documents shared by connections have a much higher chance to be relevant to future searchers. In addition, the availability of social connections can act as important signals of authenticity and credibility of the shared content. A study of a file sharing system at IBM demonstrated that users are more likely to download a file when the file author is in their social network [173].

By 2011, "people" augmentation for search results had become widespread, reaching major search engines like Google and Bing [148]. It was natural to expect this approach to be beneficial for searchers. However, a sequence of studies of social augmentation of Web search, from 2012–2013, demonstrated the situation to be more complex. An eyetracking study by Muralidharan et al. [148] demonstrated how the social augmentation of search results remained unnoticed in the vast majority of cases. In post-study interviews subjects indicated this kind of annotation to be useful for only a subset of search topics, which they classified as "social" and "subjective". A follow-up study by Fernquist and Chi [73] confirmed that searchers often simply did not notice result annotations, mostly because we have evolved fairly rigid attention patterns when it comes to parsing search result-lists and these patterns tend to focus exclusively on titles and URLs, a form of *intentional blindness*. Moreover, when searchers did notice annotations, they tended to disregard those from strangers or unfamiliar people with uncertain expertise.

In a related study, Pantel et al. [155] performed a utility analysis of social annotations. They produced a taxonomy social relevance to capture and model the different types of features (query features, content features, social connection features) that can influence social relevance in search. Their findings corroborate some of those above. They also established that close social connections and experts on a given search topic provided the most utility, as compared to more distant contacts or those with uncertain expertise. This study also demonstrated how the value of different types of connections (i.e., a work colleague, a personal friend, an expert) is not universal, but depends on the topic of the query. For example, the presence of a friend in augmentation of movie search results increases the value of the result, while the presence of work colleague reduces it. Moreover, [155] described how their approach can be used to predict whether a given social annotation is likely to be relevant to a given query-page pair, which may have applications when it comes to a more selective, and possibly personalized, approach to automatically annotate individual search results.

The idea of selective annotation was further explored by [122], in the context of news reading. The authors consider how different types of annotations affect peoples' news selection behavior, and report on results from experiments looking at social annotations in two different news reading contexts. Although not strictly search-focused the results are relevant because they confirm, unsurprisingly, that the annotations of strangers have no persuasive effects, while the annotation of personal friends do have a positive impact, on article selection and reader engagement. Intriguingly, the results also suggest that annotations do more than simply influence selection: they can make (social) content more interesting by their presence, at least in part, by providing additional context to the annotated content.

9   Beyond the SERP

Finally, in this section, we consider the opportunity to support searchers beyond the SERP, which we have only touched on briefly in what has come before. In a conventional web search setting, once a searcher selects a result, they are redirected to the appropriate URL, where they are effectively left to their own devices. In other words, once they select a result, they leave behind the search engine and any ability for it to further support their search needs, which may or may not be satisfied by the selected page. At the very least, this is a missed opportunity when it comes to helping the searcher to find what they are looking for.

For instance, many search results, depending on the query, will be for high-level landing pages. They may bring the search close to the information they are looking for, but the searcher may have to engage in additional browsing to locate the specific page they need. When planning a vacation, and looking for a hotel, a search engine like Google might bring us to a travel site, or a city-level page, but often not to a specific hotel – even if it does bring the searcher to a hotel page. It is likely that users will want to search further using the hotel

site's own search interface. This begs the question as to whether the primary search engine might be able to further support searchers as they continue to search and browse, with the added benefit that the primary search engine can then learn from these off-SERP interactions. These ideas are related in spirit to some discussed earlier on the topic of augmenting search results, but instead of adding new information to the SERP, it is all about augmenting non-SERP pages, but with search-related information.

Earlier, we referred to the SearchGuide system [50,51], and how it went one step further than SERP augmentation by also supporting searchers as they navigated beyond the SERP. SearchGuide uses a browser plug-in to augment regular web pages with search information that is relevant to the current session (see Fig. 3). For example, SearchGuides navigation bar provides a visualization of a pages "computational wear"; see also [98]. This navigation bar is calibrated to the length of the page and visualizes the distribution of query terms within the page. Each icon acts as a hyperlink to a query term occurrence within the page,
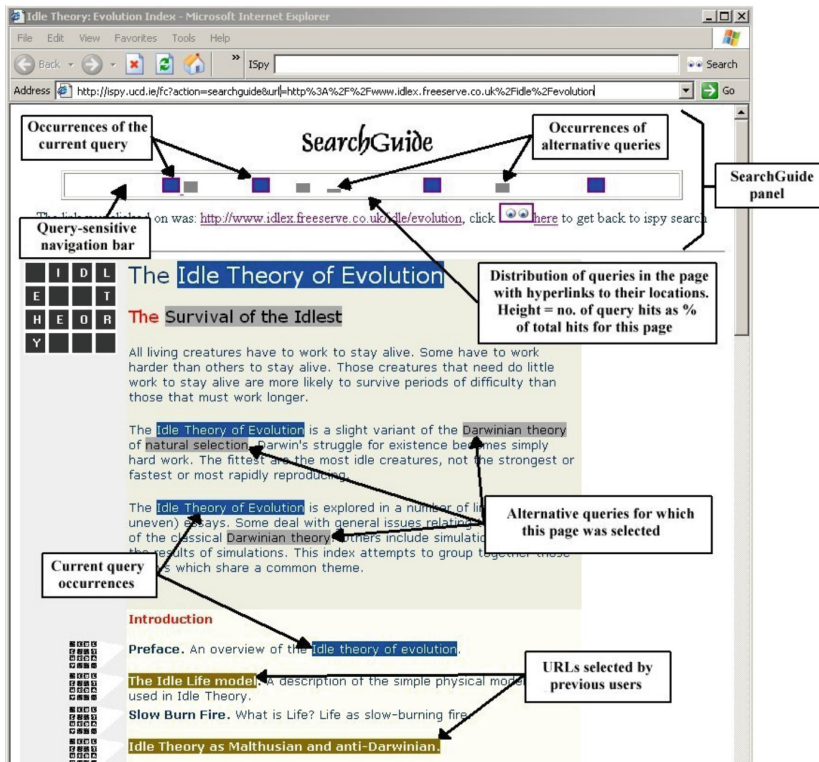
Fig. 3. An example of SearchGuide in operation, showing: the query-sensitive navigation bar; annotated page content, to emphasize occurrences of the current query; related queries within the main page content; and highlighted links that were selected by previous users. (Color figure online)

so the user can jump rapidly to regions of the page which appear to interesting, but without having to read or scan through the other (possibly irrelevant) content.

Page content is also augmented: *virtual signposts* (see also [199]) bring the user's attention to key areas of the page that relate to their query or to related queries. Finally, SearchGuide provides additional navigation support by highlighting hyperlinks within the page that have been frequently selected by users for *similar queries*, thereby serving to identify common navigation trails that past searchers have followed. The ideas in SearchGuide were further extended as part of the ASSIST project, which considered an even broader set of social feedback. This project is reviewed in more detail in Sect. 11.1.

10   Personalizing Social Search

One of the challenges faced by social search is where to draw the line between social data integration and the resulting adaptation of the search experience. Social search traces can come from a wide variety of people with many different types of information needs, but blindly using the search traces from all users may not help a particular user who has specialized needs.

This differentiates social search from other social recommendation techniques, where users are matched by a profile of their interests (see, for example, Chap. 10 of this book [116]), rather than a time-sensitive slice of current needs. In this chapter, we offer a deeper review of several case studies of practical systems, which have been developed to support more personalized social search experiences, chosen to convey the evolution of personalized search. The interested reader is encouraged to follow the citations provided for more detailed information on each case study.

10.1   Antworld

The AntWorld system [29, 110, 111, 174] was one of the first ad-hoc search engines to implement the sharing of community knowledge in order to improve the accuracy and speed of finding information on the Web. AntWorld supports users in resolving information "quests" rather than simple queries, as it attempts to understand the context of the user's information need. Following the world of ants as a computational metaphor, the system implements an asynchronous collaboration mode, where information trails from user quests are "deposited" for other community members to follow, just as ants leave pheromone trails to food sources. The AntWorld system accommodates the posting and sharing of communal knowledge as community members share their gained knowledge with the communal repository by providing feedback on how well specific search results answer their particular information needs.

For each user "quest" (Fig. 4), formulated as keywords (short description), and a longer natural language text (long description), the system computes and stores a summarized quest profile. In addition to the text of the quest, the profile contains the pages that the user browsed after receiving the system's response to

the quest, as well as their judgment about the relevance of each of these pages. This additional information reflects, to some degree, the contextual information about the user's need and their relevant level of knowledge about the domain of the quest. The quest profile is analyzed and stored, and is used to guide users who search for similar quests in the future.

During a user's interaction with the system, their quest profile is dynamically built and matched against stored quests. The system presents the user with a list of pages that other users judged to be useful for similar quests. The system also puts an ant icon next to pages that were found to be useful for similar quests. As the user provides more feedback, the system's confidence in the quest profile increases, and it is able to identify similar quests with higher accuracy. Quest profiles are represented as vectors, using the vector space model with several variations of the *TF-IDF* scheme. The terms included in the short and long description of the quest are assigned higher weights than the text of the documents that the users judged as relevant. The similarity between the user quest and the quests stored in the system is computed using cosine similarity. For further technical details, see [29].
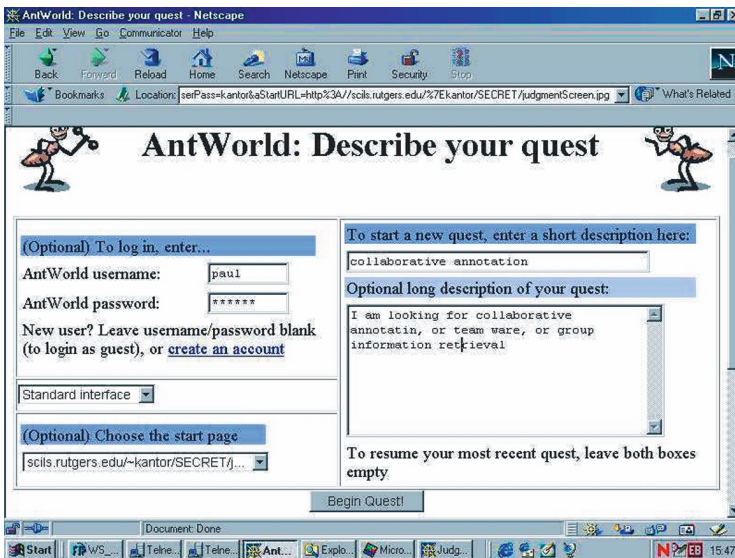

Fig. 4. AntWorld: quest description interface

A user study performed to evaluate AntWorld's potential effectiveness, and especially the extent to which users would make the effort and contribute knowledge to the community was presented in [174]. The experiment was conducted with students who used the AntWorld system to find specific information for their course assignments. In one session, users were not extrinsically motivated, while another session included an extrinsic incentive for providing evaluations

(pizza coupons for the most contributing user). The results show that the extrinsically motivated group exhibited a more significant contributing behavior than the less active group without the extrinsic motivations. A clear conclusion was drawn about the need for some type of extrinsic motivation, in order to encourage users to provide feedback, since the productivity benefits of the system did not prove motivation enough. It might well be that if a similar study were conducted today, when social networks are popular and sharing knowledge and feedback has become habitual to users, results would be different.

Following AntWorld, the SERF system implemented a similar idea in a library search context [107]. In a manner similar to AntWorld's "quests", SERF encouraged users to submit extended and informative queries and collected feedback from users as to whether search results met their information needs. The system used the feedback to provide recommendations to later users with similar needs. Over time, the SERF system learned from the users about which documents were valuable for which information needs. One difference between AntWorld and SERF is that AntWorld builds a dynamic quest profile that is adapted during the user's interaction in relation to an information need, and updates the list of similar quests and related documents as the search goes on. SERF provides only one list of similar queries and their relevant documents after the user submits their query. Initial user studies to evaluate SERF concluded that recommendations based on prior users with similar queries could increase the efficiency, and potentially the effectiveness, of library website search at Oregon State University, where an experiment was conducted. In respect to user's willingness to provide feedback, the results followed the findings of AntWorld, with relatively low participation observed.

10.2 I-SPY

In this section and the next, we summarize a pair of related collaborative/social search approaches – I-Spy and HeyStaks – which have both been mentioned earlier in this chapter. They have been chosen as early and influential examples of collaborative search and social search with the aim of making traditional web search more personalised with respects to the needs and interests of groups, or communities, of like-minded searchers.

I-Spy is an early example of a *collaborative* web search engine [178–180]. It was developed as a meta-search engine which drew its results from an number of underlying search engines: queries to I-SPY were dispatched to a variety of underlying (third-party) search engines, such as Google and Bing, and their result lists were normalised and aggregated to provide I-SPY with an initial set of results. These combined results were then ranked and returned to the search using a variety of social signals; see Fig. 5.

I-SPY used an implicit model of a search community, by using the source of search queries as a proxy for topically related searches and searchers. Thus, for example, by hosting an I-SPY searchbox on a wildlife site one would expect queries for "Jaguar photos" to result in clicks for pages with photographs of the wildcat, rather than the motor car, or Apple operating system. In contrast,
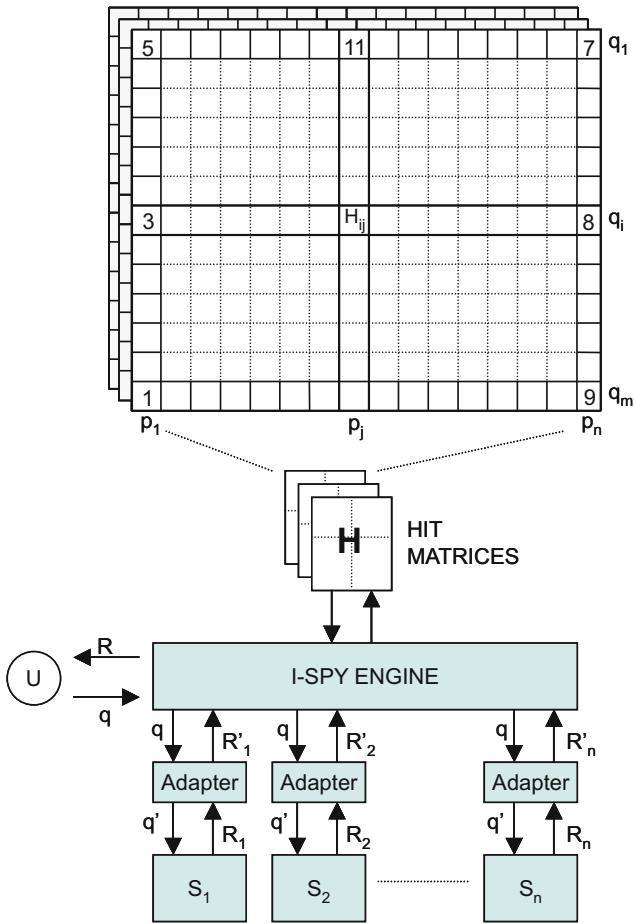
Fig. 5. The I-Spy system architecture showing the I-SPY collaborative search engine, search engine adaptors, and core hit-matrix as the basis for result relevance.

an I-SPY search box hosted on an antique automobile site might also attract "jaguar photos" queries but their selection histories would, presumably, link to car related pictures.

I-SPY records the past queries of users from a given community and their corresponding result selections. These data are stored in a data structure called a *hit matrix* (see Fig. 5). Then, for a new target query, q_t, the relevance of a page p_j is calculated as the proportion of selections for p_j given q_t. This simple relevance metric was extended in [178–180] to accommodate page selections for queries that were *similar* (based on term-overlap) to q_t. Thus, the relevance of a page p_j depends on a weighted-sum of its selections for similar queries. If the page was selected for many similar queries, then it received a higher relevance score than if it was selected for fewer, less similar queries. In this way I-SPY leveraged

Fig. 6. An example I-Spy result-list showing a set of results retrieved for a given query. These results originate from underling search engines such as Google and Bing but are re-ranked based on social search data, such as past selections for similar queries.

a combination of conventional web search approaches (used of its underlying search engines) to identify a set of candidate results while harnessing community selection behavior in order to rank results.

An example of I-SPY in operation is presented in Fig. 6, which shows the result-list returned for the rather ambiguous query "cbr". In this example, the query originated from an I-SPY search box hosted on a AI research site, and therefore "cbr" referred to *case-based reasoning*, a form of machine learning. As a result, the results returned are all examples of case-based reasoning or AI related results, as reflected by the past selections of other searchers in this community.

10.3 HeyStaks

More recently, HeyStaks [181,182] built on many of the ideas developed as part of I-SPY, but provided a more flexible social search experience and was motivated by a number of problems with the original I-SPY approach:

1. I-SPY used an *implicit* form of community, based on the origin of the search queries, as discussed above. At the time, it was common to host mainstream search-engine query-fields on third-party sites, but in the end this did not gain traction—especially when browsers implemented more dynamic "navigation fields" which allowed users to freely enter queries or web addresses – and it became obvious that most people interacted with search engines via their favorite search engine interface.

2. Many early users expressed a desire to be able to create their own search communities, based on different interests that they might have (e.g. work-related, travel-related, personal, etc.), but without the need to host search boxes on sites that were beyond their control.
3. I-SPY expected users to transition to an entirely new search interface when most searchers just wanted to "search as normal."
4. I-SPY used a limited set of social signals (essentially just result selections) as the basis for its judgments, and, as the social web evolved, it became clear that users engaged in many other types of search-related activities such as the tagging or sharing of results, as previously discussed in this chapter.

HeyStaks was developed with these shortcomings in mind. It was implemented using a browser plugin and toolbar, which carried a two-fold advantage. First, the toolbar was always available to the user through their browser, which allowed users to interact with HeyStaks at any time, rather than only during search. They could create or join staks (see below), tag pages, share results, and perform other actions. Second, it made it possible to seamlessly integrate HeyStaks with the user's preferred search engine. This enabled HeyStaks to capture queries and page selections and allowed it to directly augment the search interface of a search engine, such as Google or Bing; see Fig. 7.
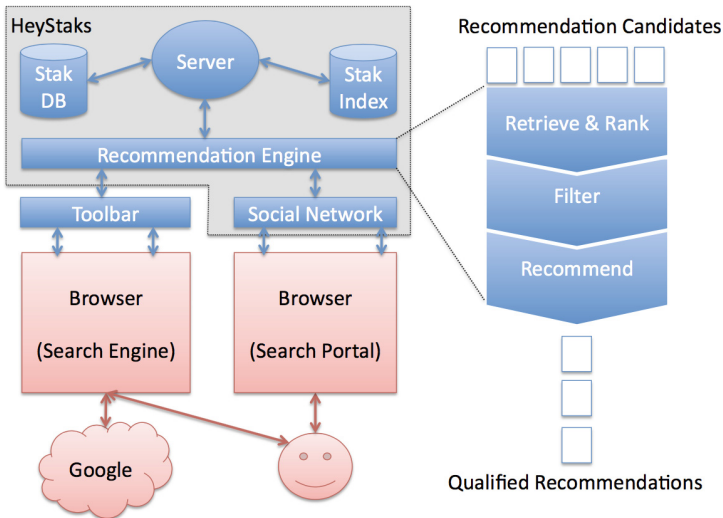

Fig. 7. HeyStaks is implemented as a back-end social search and recommendation system with a user-facing browser toolbar, thereby providing a seamless integration between HeyStaks and an existing search engine, such as Google.

The social and collaborative focus of HeyStaks was based on the ability of users to create *search staks* as types of folders for their search experiences. For example, a searcher might create a stak called "Canada Trip" as a repository for search information generated as they researched an upcoming trip to Canada;
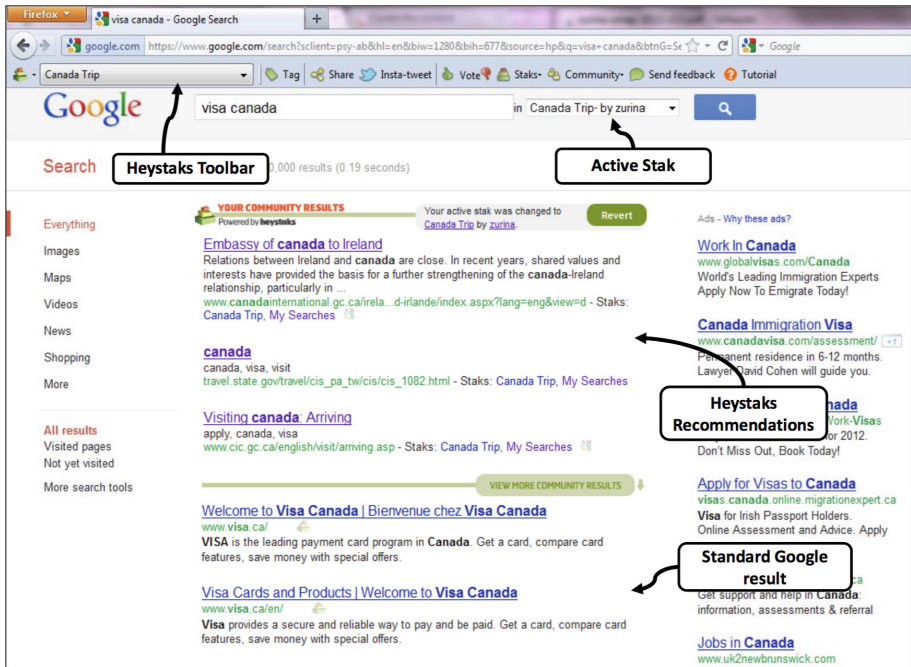
Fig. 8. An example of HeyStaks in action with Google. The HeyStaks toolbar is shown as a browser toolbar and provides the user with access to various features such as stak creation and sharing, and feedback options, such as voting and tagging. The screenshot shows a Google SERP that has been augmented with HeyStaks various augmentations, including a set of top-3 community promotions at the head of the Google result-list and annotations on regular Google results to indicate other community-relevant results.

creating a stak is a simple matter of completing a short pop-up form by using an option on the HeyStaks toolbar.

Next, as stak members search, their queries and selections are associated with a particular stak. Staks can be "shared" with others so that their searches will also be added to the staks they choose to join. Like I-Spy, HeyStaks tracks queries and result selections, but it also records other forms of search-related actions. For example, users can explicitly tag or share pages and can provide explicit relevance feedback in the form of positive and negative votes. These signals are integrated with queries and result selections in order to determine the relevance of a page for a new query, with a greater weighting given to explicit indicators of relevance, such as tagging and sharing results, while less weighting is given to implicit signals, such as a result selection. In this way, HeyStaks implements a version of a hit-matrix with each stak and uses this at search time to generate and rank recommendations.

An example, is shown in Fig. 8 which shows the results returned for the query "canada visa" based on our searcher's "Canada Trip" stak. The screenshot shows the regular results returned by Google as part of the normal Google SERP, but

in addition, there are a number of results promoted to the top of the SERP by HeyStaks. These recommendations are results that stak members have previously found to be relevant for similar queries and help the searcher to discover results that friends or colleagues have found interesting, results that may otherwise be buried deep within Google's default list of results. Google's regular results can also be marked as community-relevant, and the screenshot shows how the bottom two results shown are tagged with the HeyStaks icon to indicate this.

A number of evaluation studies have been reported in the literature to describe the utility of HeyStaks in practical search settings; see [181–183]. Key to the HeyStaks proposition is that searchers need a better way to organize and share their search experiences, as opposed to the largely ad-hoc and manual mechanisms (email, word of mouth, face-to-face collaboration) that are currently the norm. HeyStaks provides these features, but do users actually take the time to create staks? Do they share these staks or join those created by others? Briefly, studies show that users do engage in a reasonable degree of stak creation and sharing activity; for example, on average, beta users created just over 3.2 new staks and joined a further 1.4. Perhaps this is not surprising: most users are likely to create a few staks and share them with a small network of colleagues or friends, at least initially.

Moreover, 85% of users engaged in search collaborations. The majority consumed results that were produced by at least one other user, and on average, these users consumed results from 7.5 other users. In contrast, 50% of users produced knowledge that was consumed by at least one other user, and each of these producers created search knowledge that was consumed by more than 12 other users on average. While users often re-selected promotions that stemmed from their own past search histories in a stak, 33% of the time, they selected results that had been contributed by other stak members. Thus, there is evidence that many users were helping other users and many users were helped by other users.

10.4   Social Search Engine – Search with Social Links

SSE [176] is a social search engine that uses both the collaborative analysis of search logs (similar to the AntWorld collaborative Quest idea) and the data obtained from the user's social network to personalize search results. Unlike other collaborative/social engines (e.g., I-SPY and HeyStaks), the user is not required to explicitly form search communities for various search topics for which collaboration is desired, but rather the system searches for relevant social ties. SSE looks for queries that were submitted by the user's friends, based on their social network, that are relevant to the user's current need. SSE merges results obtained from the collaborative analysis and the social network analysis, with results obtained by implementing standard search engines, to produce personalized and more accurate results for users. SSE integrates existing social network data (users friendships) and network metrics to rank documents. It considers the opinions of close friends about similar topics to the user's query topics as the more important metric for estimating document relevance. In addition, SSE integrates the socially sourced results with standard search results to better

balance precision and recall. This balance is achieved by including results based on the opinions of friends and, at the same time, including results that were not previously identified by the these friends.

SSE consists of a standard, underlying search engine (based on Lucene.Net 2.4.0) that realizes standard keyword-based retrieval and which is expanded with modules to implement the two algorithms, based on query logs and social links. SSE also includes a merging algorithm that integrates the search results from all algorithms into a unified ranked list. The first personalization algorithm is a collaborative algorithm that looks for documents that received positive feedback from users for similar queries. The similarity between queries is based on the query terms, as well as on the documents that were returned for the query in a similar process to the AntWorld algorithm. The second personalization algorithm uses users' social links and follows the intuition that a document that was considered relevant by a close friend (from the user's social network) is more relevant than a document suggested by a more distant friend. The system, therefore, maintains a Friendship Value $FV(U_i, F_j)$ for each member of the social network and their friends in the network. The $FV(U_i, F_j)$ - friend value between user (U_i and another member on the network, F_j, is the centrality of the friend in the network normalized by the geodesic distance between them (i.e., the shortest path between (U_i and F_j). SSE builds on the idea of sharing knowledge between users of a community in order to enhance a user's search results. However, SSE is unique in integrating knowledge from two sources, both from the set of users' friends and from any other users that had the same need in the past and rated relevant documents for that need. SSE uses these sources, along with the user's personal profile, to personalize the search results; i.e., the system re-ranks results obtained from a standard search engine according to both of the above-mentioned sources.

Figure 9 presents the SSE's main processes. As soon as the user submits a query through the GUI, it is sent to a traditional SE whose results are returned to the ranker module. The query is also submitted to the Social Filter, which consults both the Collaborative knowledge base (KB) and the Social KB. The Social Filter returns a set of ranked documents from the set of documents that were previously seen by other users. As a last step, the ranker merges the two lists of documents (from the traditional SE and the Social Filter) and returns the merged list to the SSE GUI that displays the final set of ranked documents to the user. The user may evaluate any of the retrieved documents via the SSE GUI to enrich the system with additional feedback. For further results, the reader is referred to [176].

SSE demonstrated the benefit of personalizing search results by making use of collaborative knowledge and data from social networks. While sharing knowledge between users with similar needs improves search results, the integration of social information contributes further improvements. The SSE algorithm used the social network metric to indicate the strength of friendship between a current user and the user who is the source of the document to be returned as an indicator of the document's relevance for the current user. Specifically, a combination
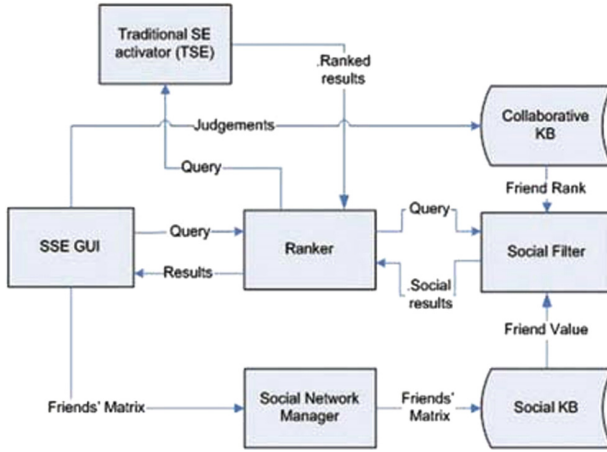
Fig. 9. The SSE architecture and outline recommendation model.

of the centrality of members in the network and the geodesic distance between the two users was used. As opposed to other community-based engines, SSE does not require users to join communities or to look for relevant searches that were performed by other users, but rather computes and identifies user communities, on the fly, that are based on network metrics. Hence, SSE could be used for short-term information needs; that is, ad hoc queries which constitute the most common usage of SEs. Although somewhat limited in scale, user studies were conducted to evaluate SSE (60 students in a lab and 7 search topics) that did not allow for a broad generalization of results, the trend was clearly towards better result accuracy when social information was used. Thus, given the popularity of social networks today, the integration of friendship data from several social networks has become very important, and recent publications [7] follow a similar approach, including a recent patent by Google [95].

11   Expanding the Borders of Social Search

As the chapter shows, social search technologies have demonstrated their ability to support various steps of the search process and leverage a range of social traces for this purpose. Yet, the majority of existing social search projects are very narrow in their coverage. Quite typically, a single social search system supports just one aspect of the search process and uses one kind of social data. We believe that overcoming this limitation is another important challenge of social search as a research field. It means developing approaches and systems that can use multiple kinds of data, support a wider set of search steps, and even support other kinds of information access. In the final section of this chapter, we showcase two further projects that go beyond the usual borders of social search systems to connect social search technologies with other types of social information access. The ASSIST platform, which was discussed earlier, demonstrates

how social search can be integrated with social navigation, another social information access technology reviewed in Chap. 5 of this book [71]. ASSIST shows how these technologies can collaborate by using community browsing data to improve search, but also by using social data collected during search to improve navigation. We also consider the Aardvark [99] system, which crosses the boundary between traditional social searching and social Q&A technology, reviewed elsewhere in Chap. 3 of this book [154].

11.1   ASSIST – From Social Search to Social Navigation

The ASSIST platform [48,72,76] is a general-purpose approach to incorporating social visual cues into existing information access systems. ASSIST was designed to integrate elements of social search and social navigation into a single platform to assist users in both searching and browsing by using both active and passive social guidance. The motivation for ASSIST included I-SPY [179,180] and Knowledge Sea [69]. While both systems used link annotation with social visual cues, I-SPY used search log data exclusively and focused on supporting user search, while Knowledge Sea used browsing data exclusively and focused on supporting navigation. The first attempts to use I-SPY to support browsing [30], and Knowledge Sea to support search [8] demonstrated how these approaches could be integrated, and led to the joint work on ASSIST.

ASSIST collects search and browsing data on a group level, and uses it in an information-exploration context (during search and browsing). The first implementation of this idea, called ASSIST-ACM, was developed for exploring research papers in the ACM Digital Library [72,76]. It combined the hit matrix of I-SPY (Fig. 5) with a similar browsing-based hit matrix to offer I-SPY visual cues (Fig. 10) to the user as they used the ACM Digital Library.

To explain the work of the ASSIST platform in more detail, we will refer to a more advanced version of the system, ASSIST-YouTube, which was designed
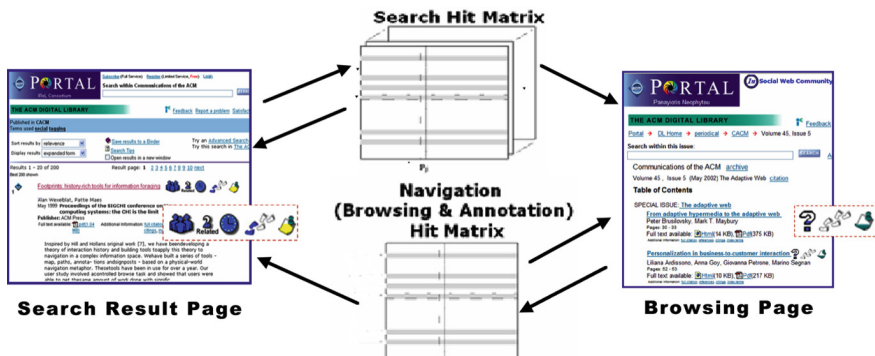

Fig. 10. ASSIST-ACM systems collects past user search and browsing traces and used them to help future users in both search and browsing.

to provide enhanced social supports, but for YouTube users [48,49]. ASSIST-YouTube offers a number of enhancements to the standard YouTube interface to improve search and browsing capabilities. In terms of active recommendations, ASSIST re-ranks lists of videos, offered by YouTube, to reflect accumulated community preferences.

In a search context, ASSIST re-ranks YouTube's search results in response to a user query Q and according to their relevance to Q. ASSIST leverages similar search history data as that used by I-SPY to assign relevance scores to videos, based on past search interactions. The relevance of video item I to query Q is estimated by the number of times I has been selected in response to Q as a fraction of the total number of selections across all items for Q. As in I-SPY, ASSIST also identifies videos that have been selected for similar queries (using a simple term-overlap similarity metric) and weights their relevance to their associated query Q_i by the similarity of Q_i to Q. These promotion candidates are ranked according to their weighted relevance score and are placed at the top of the result list for query Q.

In a browsing context, ASSIST re-ranks YouTube's *related videos*, which are displayed alongside the video that is currently being watched. This list is a valuable source of complementary content for engaging in browsing activities, and thus the position of videos within this list is important. ASSIST re-ranks the related video list, according to the items' contextual browsing popularity.

To provide passive recommendations, ASSIST augments content hyperlinks with visual social cues throughout the interface, highlighting areas of interest and suggesting paths through the wider information space. The presence of these cues signals previous encounters by community members with the content of the link. For example, when a user mouses over a cue icon, they are presented with the items search and browsing history by community members (Fig. 2). The search history presented conveys to users how the associated content has been chosen by a community member in relation to a query, as well as the strength of the item-query relationship (i.e. the relevance score). The mouse-over also includes a list of all queries that have led to the selection of the video in the past (see Fig. 2). By selecting these queries, the user can begin a new search, which essentially allows them to query YouTube for "more videos like this" with comparatively little effort. The query list is ordered by the strength of the item-query relationship. The mouse-overs are also used to provide the user with Amazon-style users who watched this video subsequently watched these recommendations. As mentioned earlier, in Sect. 2.1, if previous users engaged in browsing behavior after viewing a particular video (i.e., they selected a related video), this fact is recorded in the browse-hit matrices. By recommending videos that were subsequently watched in the mouseover provided alongside a hyperlink, the user may choose to skip watching the top-level video and go straight to one of the recommended videos.

A live-user evaluation (21 participants over a 14-week period of ASSIST-YouTube usage in a leisure-oriented context) uncovered three major types of user behavior in YouTube: traditional goal-directed search; direct browsing, following an externally recommended link; and casual browsing by watching interesting,

but not search-specified videos. While the ASSIST-YouTube social recommendation engine was designed to assist only the first type of activity, the nature of its browsing support component also makes it useful for the social support of casual browsing. However, the social support of casual browsing turned out to be more challenging than the social support of goal-directed browsing. While the search goals of the users of a specific community have some reasonable overlap, their casual browsing is driven by their general interests, rather than focused goals. In particular, the evaluation highlighted how the effects and types of social enhancements should be engineered to match the user task with the target content repository. The original social support in ASSIST was engineered for a goal-driven search task, which needs to be done with minimal effort, versus an entertaining exploratory task with fewer time constraints.

11.2   Aardvark – From Social Search to Social Q&A

Aardvark [99] was yet another take on social search. Although it was subsequently acquired by Google (and subsequently shut down), it remains as a useful case-study in an alternative vision of social search. Aardvark is not a conventional web search engine, like Google or Bing: it is not designed to find web pages in response to user queries. Instead, it is closer to a Q&A service where users provide 'queries' in the form of questions, and the 'results' are answers returned by other users. We view this as a form of search, and include Aardvark here because of the central role that social information plays in the sourcing and ranking of its results (answers).

Briefly, the central idea behind Aardvark was to harness the knowledge of individuals to directly answer questions and queries posed by users. Its key contribution was its role as a social search platform, which attempted to capture and index the expertise and interests and social networks of users in order to match these users with incoming queries. Users ask a question using a variety of channels; for example, by instant message, email, web input, text message, or even voice. Aardvark routes the question to people in the user's extended social network who are likely to be able, and available, to answer it. Accordingly, there were four key features that distinguished Aardvark from more conventional search engines:

1. *Social Indexing.* Conventional search engines focus on indexing documents, but Aardvark focuses on indexing people and, in particular, on modeling user-topic and user-user relationships, rather than query-document relationships.
2. *Question Classification.* On receiving a new question (query) Aardvark must classify the topic of the question so as to relate it to users with the right type of expertise and interests.
3. *Question Routing.* Aardvark must route a new question to the right subset of users. It does this by using a variety of information, including an aspect model that captures the topical content of the question (and candidate answerers), social network information in order to connect the questioner with people they may know, and availability information based on historical responsiveness/availability data.

4. *Answer Ranking.* Once answers come to be generated, they must be ranked so that, ideally, the most helpful answer appears at the top of the ranking. Aardvark performed this ranking based on a number of factors, including topic expertise and availability, for example.

Compared to a traditional web search engine, where the challenge lies in finding the right document to satisfy a user's information need, the challenge in a social search engine like Aardvark lies in finding the right person to satisfy a user's information need. Moreover, while traditional search engines emphasize the authority of pages when it comes to ranking, with Aardvark, trust, which is based on intimacy and reputation, is critical. Moreover, we can expect that the type of questions that a user might submit to a search engine like Aardvark to be different from those entered into a search engine like Google. The latter are optimized for information retrieval and information discovery, whereas we might expect the former to be better suited for uncovering insightful user opinions on topics that are more challenging to convey as a simple search query.

The early evidence from Aardvark was promising. For example, [99] described the results of a preliminary trial to evaluate Aardvark against Google. This trial found, for example, that Aardvark was capable of successfully answering 71.5% of questions submitted with a mean answer rating of 3.93, as compared to 70.5% of questions answered by Google with a mean rating of 3.07. Admittedly, this trial was somewhat biased towards Aardvark in the sense that the 200 or so test questions originated with Aardvark and thus were not necessarily indicative of a typical Google search query. Nevertheless, it shows that for at least these type of questions Aardvark's human-powered social search has the potential to reliably deliver high-quality answers.

As a final, case study in this chapter, Aardvark serves as a useful indication that many opportunities remain for delivering more social search experiences beyond traditional search engines. This is likely to be especially important as the world of the web continues to move towards a mobile-first future.

12   Conclusions

Today, search is among the most popular ways that people access information and the search box has become an ever-present user interface component across all operating systems and most applications where information access and discovery is relevant. This is no more obvious than in the world of Web search, where search engines like Google, Bing, and Baidu continue to serve the information needs of millions of searchers, every hour of every day.

In the quest to consistently provide the best search experience to users, the world of Web search, in particular, has been evolving, and in this chapter we consider how the increasingly social world of the web has come to bear on web search. Specifically, this chapter attempts to review how web search has become (and will continue to become) more and more social, as new social signals come to be leveraged to make web search more relevant and personal for end users and communities of like-minded users.

To do this, we framed our treatment of social search along two key dimensions. First, we considered the sources and types of social information available, from links and logs, to tags and trails, annotations, and comments. Second, we considered three key stages of the search process and the opportunities to influence searching *before*, *during*, and *after* search. Accordingly, the main contribution of this chapter included a comprehensive review of how these sources of social information can be used throughout the traditional steps of the search process, including enhancing query formation, content indexing, result ranking, and final result presentation. Throughout this chapter, we provided concrete examples from the literature of the many and varied systems that have implemented different approaches to social search, and have concluded with a number of detailed case studies to highlight a number of seminal systems that served as important milestones in the evolution of social search to address the important challenges of this field.

References

1. Abel, F., Baldoni, M., Baroglio, C., Henze, N., Krause, D., Patti, V.: Context-based ranking in folksonomies. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, pp. 209–218. ACM (2009)
2. Abel, F., Frank, M., Henze, N., Krause, D., Plappert, D., Siehndel, P.: GroupMe! - where semantic web meets web 2.0. In: Aberer, K., et al. (eds.) ASWC/ISWC -2007. LNCS, vol. 4825, pp. 871–878. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_63
3. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Analyzing user modeling on Twitter for personalized news recommendations. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) UMAP 2011. LNCS, vol. 6787, pp. 1–12. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22362-4_1
4. Aggarwal, N., Buitelaar, P.: Query expansion using Wikipedia and DBpedia. In: Forner, P., Karlgren, J., Womser-Hacker, C. (eds.) CLEF (Online Working Notes/Labs/Workshop) (2012)
5. Agichtein, E., Brill, E., Dumais, S.: Improving web search ranking by incorporating user behavior information. In: The 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 19–26. ACM Press (2006)
6. Agichtein, E., Brill, E., Dumais, S., Ragno, R.: Learning user interaction models for predicting web search result preferences. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3–10. ACM (2006)
7. Agrawal, R., Golshan, B., Papalexakis, E.: Whither social networks for web search? In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1661–1670. ACM (2015)
8. Ahn, J.W., Farzan, R., Brusilovsky, P.: Social search in the context of social navigation. J. Korean Soc. Inf. Manag. **23**(2), 147–165 (2006)
9. Alhindi, A., Kruschwitz, U., Fox, C., Albakour, D.: Profile-based summarisation for web site navigation. ACM Trans. Inf. Syst. **33**(1), 4 (2015)
10. Almeida, R.B., Almeida, V.A.F.: A community-aware search engine. In: The Thirteenth International World Wide Web Conference, WWW 2004, pp. 413–421. ACM Press (2004)

11. Alonso, O., Bannur, S., Khandelwal, K., Kalyanaraman, S.: The world conversation: web page metadata generation from social sources. In: Proceedings of the 24th International Conference on World Wide Web, pp. 385–395. ACM (2015)

12. Amatriain, X., Lathia, N., Pujol, J.M., Kwak, H., Oliver, N.: The wisdom of the few: a collaborative filtering approach based on expert opinions from the web. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 532–539. ACM (2009)

13. Amitay, E., Carmel, D., Har'El, N., Koifman, S.O., Soffer, A., Yogev, S., Golbandi, N.: Social search and discovery using a unified approach. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, pp. 199–208. ACM (2009)

14. Amitay, E., Darlow, A., Konopnicki, D., Weiss, U.: Queries as anchors: selection by association. In: Proceedings of the 16th ACM Conference on Hypertext and Hypermedia, pp. 193–201 (2005)

15. Arguello, J., Elsas, J.L., Callan, J., Carbonell, J.: Document representation and query expansion models for blog recommendation. In: 2nd International Conference on Weblogs and Social Media, ICWSM 2008, pp. 10–18 (2008)

16. Ashyralyyev, S., Cambazoglu, B.B., Aykanat, C.: Incorporating the surfing behavior of web users into PageRank. In: 22nd ACM International Conference on Information and Knowledge Management, pp. 2351–2356. ACM (2013)

17. Baeza-Yates, R., Hurtado, C., Mendoza, M.: Improving search engines by query clustering. J. Am. Soc. Inf. Sci. Technol. **58**(12), 1793–1804 (2007)

18. Baeza-Yates, R., Tiberi, A.: Extracting semantic relations from query logs. In: KDD 2007: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 76–85. ACM (2007)

19. Baeza-Yates, R.A., Hurtado, C.A., Mendoza, M.: Query recommendation using query logs in search engines. In: Current Trends in Database Technology - EDBT 2004 Workshops, Heraklion, Crete, Greece, 14–18 March 2004, Revised Selected Papers, pp. 588–596 (2004)

20. Balfe, E., Smyth, B.: Improving web search through collaborative query recommendation. In: Proceedings of the 16th European Conference on Artificial Intelligence, pp. 268–272 (2004)

21. Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., Su, Z.: Optimizing web search using social annotations. In: Proceedings of the 16th International Conference on World Wide Web, Banff, Alberta, Canada, 8–12 May 2007, pp. 501–510 (2007)

22. Baraglia, R., Cacheda, F., Carneiro, V., Fernández, D., Formoso, V., Perego, R., Silvestri, F.: Search shortcuts: a new approach to the recommendation of queries. In: Proceedings of the 2009 ACM Conference on Recommender Systems, pp. 77–84 (2009)

23. Barrett, R., Maglio, P.P.: Intermediaries: an approach to manipulating information streams. IBM Syst. J. **38**(4), 629–641 (1999)

24. Bilenko, M., White, R.W.: Mining the search trails of surfing crowds: identifying relevant websites from user activity. In: The 17th International Conference on World Wide Web, pp. 51–60. ACM (2008)

25. Billerbeck, B., Scholer, F., Williams, H.E., Zobel, J.: Query expansion using associated queries. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management, pp. 2–9. ACM, New York (2003)

26. Bischoff, K., Firan, C.S., Nejdl, W., Paiu, R.: Can all tags be used for search? In: Shanahan, J.G., Amer-Yahia, S., Zhang, Y., Kolcz, A., Chowdhury, A., Kelly, D. (eds.) The 17th ACM Conference on Conference on Information and Knowledge Management, pp. 203–212. ACM Press (2008)

27. Bollmann-Sdorra, P., Raghavan, V.V.: On the delusiveness of adopting a common space for modeling IR objects: are queries documents? J. Am. Soc. Inf. Sci. **44**(10), 579–587 (1993)

28. Borisov, A., Markov, I., de Rijke, M., Serdyukov, P.: A neural click model for web search. In: Proceedings of the 25th International Conference on World Wide Web, pp. 531–541. International World Wide Web Conferences Steering Committee (2016)

29. Boros, E., Kantor, P.B., Neu, D.J.: Pheromonic representation of user quests by digital structures. In: Proceedings of the 62nd Annual Meeting of the American Society for Information Science, pp. 633–642 (1999)

30. Boydell, O., Smyth, B.: From social bookmarking to social summarization: an experiment in community-based summary generation. In: Proceedings of the 2007 International Conference on Intelligent User Interfaces, 28–31 January 2007, Honolulu, Hawaii, USA, pp. 42–51 (2007)

31. Boydell, O., Smyth, B.: Social summarization in collaborative web search. Inf. Process. Manag. **46**(6), 782–798 (2010)

32. Briggs, P., Smyth, B.: On the role of trust in collaborative web search. Artif. Intell. Rev. **25**(1–2), 97–117 (2006)

33. Briggs, P., Smyth, B.: Harnessing trust in social search. In: Advances in Information Retrieval, 29th European Conference on IR Research, Rome, Italy, 2–5 April 2007, Proceedings, pp. 525–532 (2007)

34. Briggs, P., Smyth, B.: Trusted search communities. In: Proceedings of the 2007 International Conference on Intelligent User Interfaces, 28–31 January 2007, Honolulu, Hawaii, USA, pp. 337–340 (2007)

35. Briggs, P., Smyth, B.: Provenance, trust, and sharing in peer-to-peer case-based web search. In: Althoff, K.-D., Bergmann, R., Minor, M., Hanft, A. (eds.) ECCBR 2008. LNCS (LNAI), vol. 5239, pp. 89–103. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85502-6_6

36. Brin, S., Page, L.: The anatomy of a large-scale hypertextual (web) search engine. In: Ashman, H., Thistewaite, P. (eds.) Seventh International World Wide Web Conference, vol. 30, pp. 107–117. Elsevier Science B.V, Amsterdam (1998)

37. Brusilovsky, P., Farzan, R., Ahn, J.W.: Comprehensive personalized information access in an educational digital library. In: The 5th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 9–18. ACM Press (2005)

38. Brusilovsky, P., He, D.: Introduction to social information access. In: Brusilovsky, P., He, D. (eds.) Social Information Access. LNCS, vol. 10100, pp. 1–18. Springer, Cham (2018)

39. Bush, V.: As we may think. The Atlantic, July 1945

40. Cai, Y., Li, Q.: Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In: The 19th ACM Conference on Information and Knowledge Management (CIKM 2010), pp. 969–978. ACM (2010)

41. Carmel, D., Zwerdling, N., Guy, I., Koifman, S.O., Har'el, N., Ronen, I., Uziel, E., Yogev, S., Chernov, S.: Personalized social search based on the user's social network. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, pp. 1227–1236. ACM (2009)

42. Carpineto, C., Romano, G.: A survey of automatic query expansion in information retrieval. ACM Comput. Surv. **44**(1), 1:1–1:50 (2012)

43. Carriere, S.J., Kazman, R.: WebQuery: searching and visualizing the web through connectivity. In: Sixth International World Wide Web Conference, pp. 1257–1267. Elsevier (1997)

44. Chakrabarti, S., Dom, B., Raghavan, P., Rajagopalan, S., Gibson, D., Kleinberg, J.: Automatic resource compilation by analyzing hyperlink structure and associated text. In: Proceedings of the Seventh International Conference on World Wide Web 7, vol. 30, pp. 65–74. Elsevier Science Publishers B.V. (1998)
45. Chang, Y., Dong, A., Kolari, P., Zhang, R., Inagaki, Y., Diaz, F., Zha, H., Liu, Y.: Improving recency ranking using Twitter data. ACM Trans. Intell. Syst. Technol. 4(1), Article No. 4 (2013)
46. Chapelle, O., Zhang, Y.: A dynamic Bayesian network click model for web search ranking. In: Proceedings of the 18th International Conference on World Wide Web, pp. 1–10. ACM (2009)
47. Choochaiwattana, W.: Using social annotation to improve web search. Ph.D. thesis (2008). http://d-scholarship.pitt.edu/7832/
48. Coyle, M., Freyne, J., Brusilovsky, P., Smyth, B.: Social information access for the rest of us: an exploration of social YouTube. In: Nejdl, W., Kay, J., Pu, P., Herder, E. (eds.) AH 2008. LNCS, vol. 5149, pp. 93–102. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70987-9_12
49. Coyle, M., Freyne, J., Farzan, R., Smyth, B., Brusilovsky, P.: Reducing click distance through social adaptive interfacing. In: ReColl 2008, International Workshop on Recommendation and Collaboration at 2008 International Conference on Intelligent User Interfaces (2008)
50. Coyle, M., Smyth, B.: SearchGuide: beyond the results page. In: De Bra, P.M.E., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 296–299. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-27780-4_36
51. Coyle, M., Smyth, B.: Supporting intelligent web search. ACM Trans. Internet Techn. **7**(4), 20 (2007)
52. Craswell, N., Billerbeck, B., Fetterly, D., Najork, M.: Robust query rewriting using anchor data. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, pp. 335–344. ACM, New York (2013)
53. Craswell, N., Hawking, D., Robertson, S.: Effective site finding using link anchor information. In: 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 250–257. ACM Press (2001)
54. Craswell, N., Szummer, M.: Random walks on the click graph. In: 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 239–246 (2007)
55. Craswell, N., Zoeter, O., Taylor, M., Ramsey, B.: An experimental comparison of click position-bias models. In: Proceedings of the 2008 International Conference on Web Search and Data Mining, pp. 87–94. ACM (2008)
56. Cui, H., Wen, J.R., Nie, J.Y., Ma, W.Y.: Probabilistic query expansion using query logs. In: Proceedings of the 11th International Conference on World Wide Web, pp. 325–332. ACM (2002)
57. Culliss, G.: Method for organizing information, 20 June 2000. http://www.google.com/patents/US6078916, US Patent 6,078,916
58. Culliss, G.: Method for organizing information, 11 January 11 2000. http://www.google.com/patents/US6014665
59. Dang, V., Croft, B.W.: Query reformulation using anchor text. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 41–50. ACM, New York (2010)
60. Dennis, S., Bruza, P., McArthur, R.: Web searching: a process-oriented experimental study of three interactive search paradigms. J. Am. Soc. Inf. Sci. Technol. **52**(2), 120–133 (2002)

61. Dmitriev, P., Eiron, N., Fontoura, M., Shekita, E.: Using annotations in enterprise search. In: 15th International Conference on World Wide Web, pp. 811–817. ACM (2006)
62. Dong, A., Zhang, R., Kolari, P., Bai, J., Diaz, F., Chang, Y., Zheng, Z., Zha, H.: Time is of the essence: improving recency ranking using Twitter data. In: Proceedings of the 19th International Conference on World Wide Web, pp. 331–340. ACM (2010)
63. Dou, Z., Song, R., Yuan, X., Wen, J.: Are click-through data adequate for learning web search rankings? In: Proceedings of the 17th ACM Conference on Information and Knowledge Management, pp. 73–82. ACM (2008)
64. Dupret, G., Piwowarski, B.: A user browsing model to predict search engine click data from past observations. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 331–338. ACM (2008)
65. Eirinaki, M., Vazirgiannis, M.: Usage-based PageRank for web personalization. In: The Fifth IEEE International Conference on Data Mining, pp. 130–137 (2005)
66. Esparza, S.G., O'Mahony, M.P., Smyth, B.: On the real-time web as a source of recommendation knowledge. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 305–308. ACM (2010)
67. Evans, B.M., Chi, E.H.: Towards a model of understanding social search. In: 2008 ACM Conference on Computer Supported Cooperative Work, pp. 485–494. ACM (2008)
68. Evans, B.M., Chi, E.H.: An elaborated model of social search. Inf. Process. Manag. **46**(6), 656–678 (2010)
69. Farzan, R., Brusilovsky, P.: Social navigation support through annotation-based group modeling. In: Ardissono, L., Brna, P., Mitrovic, A. (eds.) UM 2005. LNCS (LNAI), vol. 3538, pp. 463–472. Springer, Heidelberg (2005). https://doi.org/10.1007/11527886_64
70. Farzan, R., Brusilovsky, P.: Social navigation support for information seeking: if you build it, will they come? In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 66–77. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02247-0_9
71. Farzan, R., Brusilovsky, P.: Social navigation. In: Brusilovsky, P., He, D. (eds.) Social Information Access. LNCS, vol. 10100, pp. 142–180. Springer, Cham (2018)
72. Farzan, R., Coyle, M., Freyne, J., Brusilovsky, P., Smyth, B.: ASSIST: adaptive social support for information space traversal. In: 18th Conference on Hypertext and Hypermedia, pp. 199–208. ACM Press (2007)
73. Fernquist, J., Chi, E.H.: Perception and understanding of social annotations in web search. In: 22nd International World Wide Web Conference, WWW 2013, Rio de Janeiro, Brazil, pp. 403–412, 13–17 May 2013
74. Fitzpatrick, L., Dent, M.: Automatic feedback using past queries: social searching? In: Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 306–313 (1997)
75. Fonseca, B., Golgher, P., De Moura, E., Pôssas, B., Ziviani, N.: Discovering search engine related queries using association rules. J. Web Eng. **2**(4), 215–227 (2003)
76. Freyne, J., Farzan, R., Brusilovsky, P., Smyth, B., Coyle, M.: Collecting community wisdom: integrating social search and social navigation. In: International Conference on Intelligent User Interfaces, pp. 52–61. ACM Press (2007)
77. Freyne, J., Smyth, B.: An experiment in social search. In: De Bra, P.M.E., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 95–103. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-27780-4_13

78. Fujii, A.: Modeling anchor text and classifying queries to enhance web document retrieval. In: The 17th International Conference on World Wide Web, pp. 337–346. ACM (2008)

79. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. Commun. ACM **30**(11), 964–971 (1987)

80. Furnas, G.W.: Experience with an adaptive indexing scheme. SIGCHI Bull. **16**(4), 131–135 (1985)

81. Furuta, R., Shipman III, F.M., Marshall, C.C., Brenner, D., Hsieh, H.W.: Hypertext paths and the world-wide web: experience with Walden's paths. In: Bernstein, M., Carr, L., Østerbye, K. (eds.) Eight ACM International Hypertext Conference (Hypertext 1997), pp. 167–176. ACM (1997)

82. Gao, J., Yuan, W., Li, X., Deng, K., Nie, J.: Smoothing clickthrough data for web search ranking. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 355–362. ACM (2009)

83. Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User profiles for personalized information access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 54–89. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_2

84. Gibson, D., Kleinberg, J., Raghavan, P.: Inferring web communities from link topology. In: Ninth ACM International Hypertext Conference, pp. 50–57. ACM Press (1998)

85. Glance, N.: Community search assistant. In: Proceedings of the 6th International Conference on Intelligent User Interfaces, pp. 91–96. ACM (2001)

86. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. J. Inf. Sci. **32**(2), 198–208 (2006)

87. Graus, D., Tsagkias, M., Weerkamp, W., Meij, E., de Rijke, M.: Dynamic collective entity representations for entity ranking. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, pp. 595–604. ACM (2016)

88. Guan, Z., Cutrell, E.: What are you looking for?: an eye-tracking study of information usage in web search. In: ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 407–416. ACM Press (2007)

89. Guo, F., Liu, C., Kannan, A., Minka, T., Taylor, M., Wang, Y., Faloutsos, C.: Click chain model in web search. In: Proceedings of the 18th International Conference on World Wide Web, pp. 11–20. ACM (2009)

90. Guo, Y.Z., Ramamohanarao, K., Park, L.A.F.: Personalized PageRank for web page prediction based on access time-length and frequency. In: Lin, T.Y., Haas, L., Kacprzyk, J., Motwani, R., Broder, A., Ho, H. (eds.) The 2007 International Conference on Web Intelligence, pp. 687–690. IEEE (2007)

91. Hammond, T., Hannay, T., Lund, B., Scott, J.: Social bookmarking tools (i): a general review. D-Lib Mag. **11**(4) (2005)

92. Han, S., He, D.: Network-based social search. In: Brusilovsky, P., He, D. (eds.) Social Information Access. LNCS, vol. 10100, pp. 277–309. Springer, Cham (2018)

93. Han, S., He, D., Yue, Z., Brusilovsky, P.: Supporting cross-device web search with social navigation-based mobile touch interactions. In: Ricci, F., Bontcheva, K., Conlan, O., Lawless, S. (eds.) UMAP 2015. LNCS, vol. 9146, pp. 143–155. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20267-9_12

94. Harman, D.: Information retrieval. In: Relevance Feedback and Other Query Modification Techniques, pp. 241–263. Prentice-Hall, Inc. (1992)

95. Harrington, T., Shenoy, R., Najork, M., Panigrahy, R.: Social network recommended content and recommending members for personalized search results, 3 February 2015. https://www.google.com/patents/US8949232, US Patent 8,949,232

96. Hearst, M.A.: TileBars: visualization of term distribution information in full text information access. In: Human Factors in Computing Systems, CHI 1995 Conference Proceedings, Denver, Colorado, USA, 7–11 May 1995, pp. 59–66 (1995)

97. Heymann, P., Koutrika, G., Garcia-Molina, H.: Can social bookmarking improve web search? In: The International Conference on Web Search and Web Data Mining (WSDM 2008), pp. 195–206 (2008)

98. Hill, W.C., Hollan, J.D., Wroblewski, D., McCandless, T.: Edit wear and read wear. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3–9. ACM, New York (1992)

99. Horowitz, D., Kamvar, S.: The anatomy of a large-scale social search engine. In: The 19th International Conference on World Wide Web, pp. 431–440. ACM (2010)

100. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: search and ranking. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 411–426. Springer, Heidelberg (2006). https://doi.org/10.1007/11762256_31

101. Huang, C.K., Chien, L.F., Oyang, Y.J.: Relevant term suggestion in interactive web search based on contextual information in query session logs. J. Am. Soc. Inf. Sci. Technol. **54**(7), 638–649 (2003)

102. Huberman, B.A., Adamic, L.A.: Novelty and social search in the world wide web. CoRR cs.MA/9809025 (1998). http://arxiv.org/abs/cs.MA/9809025

103. Jiang, J., Yue, Z., Han, S., He, D.: Finding readings for scientists from social websites. In: 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1075–1076 (2012)

104. Joachims, T.: Optimizing search engines using clickthrough data. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2002), pp. 133–142. ACM (2002)

105. Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G.: Accurately interpreting clickthrough data as implicit feedback. In: 28th Annual International ACM SIGIR Conference, pp. 154–161. ACM Press (2005)

106. Joachims, T., Radlinski, F.: Search engines that learn from implicit feedback. Computer **40**(8), 34–40 (2007)

107. Jung, S., Harris, K., Webster, J., Herlocker, J.L.: Serf: integrating human recommendations with search. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, pp. 571–580. ACM (2004)

108. Kahan, J., Koivunen, M.R., Prud'Hommeaux, E., Swick, R.R.: Annotea: an open RDF infrastructure for shared web annotations. Comput. Netw. **39**(5), 589–608 (2002)

109. Kammerer, Y., Nairn, R., Pirolli, P., Chi, E.H.: Signpost from the masses: learning effects in an exploratory social tag search browser. In: 27th International Conference on Human Factors in Computing Systems, pp. 625–634 (2009)

110. Kantor, P., Boros, E., Melamed, B., Neu, D., Menkov, V., Shi, Q., Kim, M.H.: Ant world (demonstration abstract). In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, p. 323. ACM (1999)

111. Kantor, P.B., Boros, E., Melamed, B., Meñkov, V., Shapira, B., Neu, D.J.: Capturing human intelligence in the net. Commun. ACM **43**(8), 112–115 (2000)

112. Katariya, S., Kveton, B., Szepesvári, C., Wen, Z.: DCM bandits: learning to rank with multiple clicks. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning, vol. 48, pp. 1215–1224. JMLR.org (2016)

113. Keane, M., O'Brien, M., Smyth, B.: Are people biased in their use of search engines? Commun. ACM **51**(2), 49–52 (2008)

114. Kemp, C., Ramamohanarao, K.: Long-term learning for web search engines. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) PKDD 2002. LNCS, vol. 2431, pp. 263–274. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45681-3_22

115. Kleinberg, J.: Authoritative sources in a hyperlinked environment. J. ACM **46**(5), 604–632 (1999)

116. Kluver, D., Ekstrand, M., Konstan, J.: Rating-based collaborative filtering: algorithms and evaluation. In: Brusilovsky, P., He, D. (eds.) Social Information Access. LNCS, vol. 10100, pp. 344–390. Springer, Cham (2018)

117. Koolen, M., Kamps, J.: The importance of anchor text for ad hoc search revisited. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 122–129. ACM (2010)

118. Korfhage, R.R.: Information Storage and Retrieval. Wiley Computer Publishing, Hoboken (1997)

119. Kraft, R., Zien, J.: Mining anchor text for query refinement. In: Proceedings of the 13th International Conference on World Wide Web, pp. 666–674. ACM, New York (2004)

120. Kramár, T., Barla, M., Bieliková, M.: Disambiguating search by leveraging a social context based on the stream of user's activity. In: De Bra, P., Kobsa, A., Chin, D. (eds.) UMAP 2010. LNCS, vol. 6075, pp. 387–392. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13470-8_37

121. Kruschwitz, U., Lungley, D., Albakour, M.D., Song, D.: Deriving query suggestions for site search. J. Am. Soc. Inf. Sci. Technol. **64**(10), 1975–1994 (2013)

122. Kulkarni, C., Chi, E.: All the news that's fit to read: a study of social annotations for news reading. In: 2013 ACM SIGCHI Conference on Human Factors in Computing Systems, CHI 2013, Paris, France, 27 April–2 May 2013, pp. 2407–2416 (2013)

123. Lee, C., Croft, B.: Incorporating social anchors for ad hoc retrieval. In: Proceedings of the 10th Conference on Open Research Areas in Information Retrieval, pp. 181–188 (2013)

124. Lehikoinen, J., Salminen, I., Aaltonen, A., Huuskonen, P., Kaario, J.: Metasearches in peer-to-peer networks. Pers. Ubiquitous Comput. **10**(6), 357–367 (2006)

125. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., et al.: DBpedia-a large-scale, multilingual knowledge base extracted from Wikipedia. Seman. Web **6**(2), 167–195 (2015)

126. Lerman, K., Plangprasopchok, A., Wong, C.: Personalizing image search results on flickr. In: AAAI07 workshop on Intelligent Information Personalization (2007). http://arxiv.org/abs/0704.1676

127. Li, L., Otsuka, S., Kitsuregawa, M.: Finding related search engine queries by web community based query enrichment. World Wide Web **13**(1), 121–142 (2010)

128. Li, Y., Luk, W.P.R., Ho, K.S.E., Chung, F.L.K.: Improving weak ad-hoc queries using Wikipedia as external corpus. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 797–798. ACM (2007)

129. Lieberman, H.: Letizia: an agent that assists web browsing. In: The Fourteenth International Joint Conference on Artificial Intelligence, pp. 924–929 (1995)

130. Liu, Y., Gao, B., Liu, T.Y., Zhang, Y., Ma, Z., He, S., Li, H.: BrowseRank: letting web users vote for page importance. In: 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 451–458. ACM (2008)

131. Liu, Y., Liu, T.Y., Gao, B., Ma, Z., Li, H.: A framework to compute page importance based on user behaviors. Inf. Retr. **13**(1), 22–45 (2010)

132. Ma, H., Yang, H., King, I., Lyu, M.R.: Learning latent semantic relations from clickthrough data for query suggestion. In: Shanahan, J.G., Amer-Yahia, S., Zhang, Y., Kolcz, A., Chowdhury, A., Kelly, D. (eds.) The 17th ACM Conference on Conference on Information and Knowledge Management: CIKM 2008, pp. 709–718. ACM Press (2008)

133. Maron, M.E., Kuhns, J.L.: On relevance, probabilistic indexing and information retrieval. J. ACM **7**(3), 216–244 (1960)

134. Massa, P., Avesani, P.: Trust-aware recommender systems. In: 2007 ACM Conference on Recommender Systems, pp. 17–24. ACM (2007)

135. McBryan, O.A.: GENVL and WWWW: tools for taming the web. In: The 1st International World Wide Web Conference, pp. 79–90 (1994)

136. McCreadie, R., Macdonald, C., Ounis, I.: News article ranking: leveraging the wisdom of bloggers. In: RIAO 2010 - 9th RIAO Conference (2010)

137. McDonnell, M., Shiri, A.: Social search: a taxonomy of, and a user-centred approach to, social web search. Program **45**(1), 6–28 (2011)

138. McNally, K., O'Mahony, M.P., Smyth, B.: A comparative study of collaboration-based reputation models for social recommender systems. User Model. User-Adapt. Interact. **24**(3), 219–260 (2014)

139. Micarelli, A., Gasparetti, F., Sciarrone, F., Gauch, S.: Personalized search on the world wide web. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 195–230. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_6

140. Millen, D., Whittaker, S., Yang, M., Feinberg, J.: Supporting social search with social bookmarking. In: The HCIC 2007 Winter Workshop (2007). http://www.hcic.org/hcic2007/papers.phtml

141. Millen, D.R., Feinberg, J., Kerr, B.: Dogear: social bookmarking in the enterprise. In: SIGCHI Conference on Human Factors in Computing Systems, CHI 2006, pp. 111–120 (2006)

142. Millen, D.R., Yang, M., Whittaker, S., Feinberg, J.: Social bookmarking and exploratory search. In: Bannon, L.J., Wagner, I., Gutwin, C., Harper, R.H.R., Schmidt, K. (eds.) ECSCW 2007. Springer, London (2007). https://doi.org/10.1007/978-1-84800-031-5_2

143. Miller, G.A.: WordNet: a lexical database for English. Commun. ACM **38**(11), 39–41 (1995)

144. Mishne, G., Lin, J.: Twanchor text: a preliminary study of the value of tweets as anchor text. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1159–1160. ACM (2012)

145. Muhammad, K., Lawlor, A., Rafter, R., Smyth, B.: Great explanations: opinionated explanations for recommendations. In: Hüllermeier, E., Minor, M. (eds.) ICCBR 2015. LNCS (LNAI), vol. 9343, pp. 244–258. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24586-7_17

146. Muhammad, K.I., Lawlor, A., Smyth, B.: A live-user study of opinionated explanations for recommender systems. In: Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI 2016, Sonoma, CA, USA, 07–10 March 2016, pp. 256–260 (2016)

147. Muller, M., Millen, D., Feinberg, J.: Information curators in an enterprise file sharing system. In: 11th European Conference on Computer Supported Cooperative Work, ECSCW 2009 (2009)

148. Muralidharan, A., Gyongyi, Z., Chi, E.: Social annotations in web search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1085–1094. ACM (2012)

149. Navarro Bullock, B., Hotho, A., Stumme, G.: Accessing information with tags: search and ranking. In: Brusilovsky, P., He, D. (eds.) Social Information Access. LNCS, vol. 10100, pp. 310–343. Springer, Cham (2018)

150. Noll, M.G., Meinel, C.: Web search personalization via social bookmarking and tagging. In: Aberer, K., et al. (eds.) ASWC/ISWC -2007. LNCS, vol. 4825, pp. 367–380. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_27

151. Noll, M., Meinel, C.: The metadata triumvirate: social annotations, anchor texts and search queries. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2008, vol. 1, pp. 640–647. IEEE (2008)

152. O'Donovan, J., Smyth, B.: Trust in recommender systems. In: IUI 2005: Proceedings of the 10th International Conference on Intelligent User Interfaces, pp. 167–174. ACM Press (2005)

153. Ogilvie, P., Callan, J.: Combining document representations for known-item search. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 143–150. ACM (2003)

154. Oh, S.: Social Q&A. In: Brusilovsky, P., He, D. (eds.) Social Information Access. LNCS, vol. 10100, pp. 75–107. Springer, Cham (2018)

155. Pantel, P., Gamon, M., Alonso, O., Haas, K.: Social annotations: utility and prediction modeling. In: The 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, OR, USA, 12–16 August 2012, pp. 285–294 (2012)

156. Parikh, N., Sundaresan, N.: Inferring semantic query relations from collective user behavior. In: Shanahan, J.G., Amer-Yahia, S., Zhang, Y., Kolcz, A., Chowdhury, A., Kelly, D. (eds.) The 17th ACM Conference on Conference on Information and Knowledge Management: CIKM 2008, pp. 349–358. ACM Press (2008)

157. Phelan, O., Mccarthy, K., Smyth, B.: Yokie - a curated, real-time search and discovery system using Twitter. In: RSWEB 2011: 3rd Workshop on Recommender Systems and the Social Web at RecSys 2011 (2011)

158. Piwowarski, B., Dupret, G., Jones, R.: Mining user web search activity with layered Bayesian networks or how to capture a click in its context. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, pp. 162–171. ACM (2009)

159. Poblete, B., Baeza-Yates, R.: Query-sets: using implicit feedback and query patterns to organize web documents. In: The 17th International Conference on World Wide Web, pp. 41–50. ACM (2008)

160. Poblete, B., Castillo, C., Gionis, A.: Dr. Searcher and Mr. Browser: a unified hyperlink-click graph. In: Shanahan, J.G., Amer-Yahia, S., Zhang, Y., Kolcz, A., Chowdhury, A., Kelly, D. (eds.) The 17th ACM Conference on Conference on Information and Knowledge Management: CIKM 2008, pp. 1123–1132. ACM Press (2008)

161. Porter, M.F.: Implementing a probabilistic information retrieval system. Inf. Technol.: Res. Dev. **1**(2), 131–156 (1982)

162. Radlinski, F., Joachims, T.: Query chains: learning to rank from implicit feedback. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 239–248. ACM (2005)

163. Raghavan, V.V., Sever, H.: On the reuse of past optimal queries. In: 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 344–350. ACM (1995)

164. Ramage, D., Heymann, P., Manning, C., Garcia-Molina, H.: Clustering the tagged web. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, pp. 54–63 (2009)

165. Remde, J.R., Gomez, L.M., Landauer, T.K.: SuperBook: an automatic tool for information exploration—hypertext? In: The ACM Conference on Hypertext, Hypertext 1987, pp. 175–188 (1987)

166. Ritchie, A., Robertson, S., Teufel, S.: Comparing citation contexts for information retrieval. In: Shanahan, J.G., Amer-Yahia, S., Zhang, Y., Kolcz, A., Chowdhury, A., Kelly, D. (eds.) The 17th ACM Conference on Information and Knowledge Management: CIKM 2008, pp. 213–222. ACM Press (2008)

167. Rocchio, J.J.: Relevance feedback in information retrieval (Chap. 14). In: Salton, G. (ed.) The SMART Retrieval System: Experiments in Automatic Document Processing. Prentice-Hall Series in Automatic Computation, pp. 313–323. Prentice-Hall, Englewood Cliffs (1971)

168. Rowlands, T., Hawking, D., Sankaranarayana, R.: New-web search with microblog annotations. In: Proceedings of the 19th International Conference on World Wide Web, pp. 1293–1296. ACM (2010)

169. Salton, G.: The SMART Retrieval System: Experiments in Automatic Document Processing. Prentice-Hall, Englewood Cliffs (1971)

170. Scholer, F., Williams, H.E.: Query association for effective retrieval. In: ACM 11th Conference on Information and Knowledge Management, CIKM 2002, pp. 324–331 (2002)

171. Scholer, F., Williams, H., Turpin, A.: Query association surrogates for web search. J. Am. Soc. Inf. Sci. **55**(7), 637–650 (2004)

172. di Sciascio, C., Brusilovsky, P., Veas, E.: A study on user-controllable social exploratory search. In: 23rd International Conference on Intelligent User Interfaces. ACM (2018)

173. Shami, N.S., Muller, M., Millen, D.: Social search and metadata in predicting file discovery. In: Fifth International AAAI Conference on Weblogs and Social Media, pp. 337–344. AAAI Publications (2011)

174. Shapira, B., Kantor, P.B., Melamed, B.: The effect of extrinsic motivation on user behavior in a collaborative information finding system. J. Am. Soc. Inf. Sci. Technol. **52**(11), 879–887 (2001)

175. Shapira, B., Taieb-Maimon, M., Moskowitz, A.: Study of the usefulness of known and new implicit indicators and their optimal combination for accurate inference of users interests. In: Proceedings of the 2006 ACM Symposium on Applied Computing, pp. 1118–1119. ACM (2006)
176. Shapira, B., Zabar, B.: Personalized search: Integrating collaboration and social networks. J. Am. Soc. Inf. Sci. Technol. **62**(1), 146–160 (2011)
177. Shipman, F.M., Marshall, C.C.: Spatial hypertext: an alternative to navigational and semantic links. ACM Comput. Surv. **31**(4es), Article No. 14 (1999)
178. Smyth, B., Balfe, E., Boydell, O., Bradley, K., Briggs, P., Coyle, M., Freyne, J.: A live-user evaluation of collaborative web search. In: IJCAI-2005, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, 30 July–5 August 2005, pp. 1419–1424 (2005)
179. Smyth, B., Balfe, E., Briggs, P., Coyle, M., Freyne, J.: Collaborative web search. In: IJCAI-2003, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 9–15 August 2003, pp. 1417–1419 (2003)
180. Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M., Boydell, O.: Exploiting query repetition and regularity in an adaptive community-based web search engine. User Model. User-Adapt. Interact. **14**(5), 383–423 (2004)
181. Smyth, B., Briggs, P., Coyle, M., O'Mahony, M.: Google shared. A case-study in social search. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 283–294. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02247-0_27
182. Smyth, B., Coyle, M., Briggs, P.: Heystaks: a real-world deployment of social search. In: Sixth ACM Conference on Recommender Systems, Dublin, Ireland, 9–13 September 2012, pp. 289–292 (2012)
183. Smyth, B., Coyle, M., Briggs, P., McNally, K., O'Mahony, M.P.: Collaboration, reputation and recommender systems in social web search. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 569–608. Springer, Boston (2015). https://doi.org/10.1007/978-1-4899-7637-6_17
184. Song, Y., Zhou, D., He, L.: Query suggestion by constructing term-transition graphs. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, pp. 353–362. ACM (2012)
185. Spink, A.: Web search: emerging patterns. Libr. Trends **52**(2), 299–306 (2003)
186. Spink, A., Bateman, J., Jansen, B.J.: Users' searching behavior on the excite web search engine. In: Proceedings of WebNet 1998 - World Conference on the WWW and Internet, Orlando, Florida, USA, 7–12 November 1998
187. Spink, A., Jansen, B.J., Wolfram, D., Saracevic, T.: From e-sex to e-commerce: web search changes. IEEE Comput. **35**(3), 107–109 (2002)
188. Sun, J., Zhong, N., Yu, X.: Collaborative web search utilizing experts' experiences. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 120–127 (2010)
189. Teevan, J., Morris, M., Bush, S.: Discovering and using groups to improve personalized search. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, pp. 15–24. ACM (2009)
190. Trevisiol, M., Chiarandini, L., Aiello, L.M., Jaimes, A.: Image ranking based on user browsing behavior. In: 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 445–454. ACM (2012)
191. Uherčík, T., Šimko, M., Bieliková, M.: Utilizing microblogs for web page relevant term acquisition. In: van Emde Boas, P., Groen, F., Italiano, G., Nawrocki, J., Sack, H. (eds.) SOFSEM 2013: Theory and Practice of Computer Science. LNCS, vol. 7741, pp. 457–468. Springer, Heidelberg (2013)

192. Vahabi, H., Ackerman, M., Loker, D., Baeza-Yates, R., Ortiz, A.: Orthogonal query recommendation. In: Proceedings of the 7th ACM Conference on Recommender Systems, pp. 33–40. ACM (2013)

193. Wang, C., Liu, Y., Wang, M., Zhou, K., Nie, J.Y., Ma, S.: Incorporating non-sequential behavior into click models. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 283–292. ACM (2015)

194. Wang, X., Bendersky, M., Metzler, D., Najork, M.: Learning to rank with selection bias in personal search. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 115–124. ACM (2016)

195. Wang, X., Tan, B., Shakery, A., Zhai, C.: Beyond hyperlinks: organizing information footprints in search logs to support effective browsing. In: ACM International Conference on Information and Knowledge Management 2009 (CIKM 2009), pp. 1237–1246 (2009)

196. Wang, X., Zhai, C.: Mining term association patterns from search logs for effective query reformulation. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management, pp. 479–488. ACM (2008)

197. Wen, J.R., Nie, J.Y., Zhang, H.J.: Query clustering using user logs. ACM Trans. Inf. Syst. **20**(1), 59–81 (2002)

198. Westerveld, T., Kraaij, W., Hiemstra, D.: Retrieving web pages using content, links, URLs and anchors. In: Tenth Text REtrieval Conference, TREC 2001, pp. 663–672 (2002)

199. Wexelblat, A., Maes, P.: Footprints: history-rich tools for information foraging. In: Proceeding of the CHI 1999 Conference on Human Factors in Computing Systems: The CHI is the Limit, Pittsburgh, PA, USA, 15–20 May 1999, pp. 270–277 (1999)

200. White, R., Bilenko, M., Cucerzan, S.: Leveraging popular destinations to enhance web search interaction. ACM Trans. Web **2**(3), 1–30 (2008)

201. White, R.W., Bilenko, M., Cucerzan, S.: Studying the use of popular destinations to enhance web search interaction. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, 23–27 July 2007, pp. 159–166 (2007)

202. Wikipedia: Direct Hit Technologies, September 2015. https://en.wikipedia.org/wiki/Direct_Hit_Technologies

203. Wikipedia: Social search, November 2016. https://en.wikipedia.org/wiki/Social_search

204. Wolfram, D., Spink, A., Jansen, B.J., Saracevic, T.: Vox populi: the public searching of the web. JASIST **52**(12), 1073–1074 (2001)

205. Wu, M., Hawking, D., Turpin, A., Scholer, F.: Using anchor text for homepage and topic distillation search tasks. J. Assoc. Inf. Sci. Technol. **63**(6), 1235–1255 (2012)

206. Wu, X., Zhang, L., Yu, Y.: Exploring social annotations for the semantic web. In: 15th International Conference on World Wide Web, pp. 417–426. ACM Press (2006)

207. Xiong, C., Callan, J.: Query expansion with freebase. In: Proceedings of the 2015 International Conference on the Theory of Information Retrieval, pp. 111–120. ACM (2015)

208. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 155–162. ACM (2008)

209. Xue, G., Zeng, H., Chen, Z., Yu, Y., Ma, W., Xi, W., Fan, W.: Optimizing web search using web click-through data. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, pp. 118–126. ACM (2004)
210. Xue, G.R., Zeng, H.J., Chen, Z., Ma, W.Y., Zhang, H.J., Lu, C.J.: Implicit link analysis for small web search. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informtaion Retrieval, pp. 56–63. ACM (2003)
211. Yanbe, Y., Jatowt, A., Nakamura, S., Tanaka, K.: Can social bookmarking enhance search in the web? In: JCDL 2007: 2007 Conference on Digital Libraries, pp. 107–116. ACM Press (2007)
212. Zanardi, V., Capra, L.: Social ranking: uncovering relevant content using tag-based recommender systems. In: Proceedings of the 2008 ACM Conference on Recommender Systems, pp. 51–58. ACM (2008)
213. Zhang, J., Deng, B., Li, X.: Concept based query expansion using WordNet. In: Proceedings of the 2009 International e-Conference on Advanced Science and Technology, pp. 52–55. IEEE (2009)
214. Zhou, D., Bian, J., Zheng, S., Zha, H., Giles, C.L.: Exploring social annotations for information retrieval. In: The 17th International Conference on World Wide Web, pp. 715–724. ACM (2008)
215. Zhou, J., Ding, C., Androutsos, D.: Improving web site search using web server logs. In: CASCON 2006: Proceedings of the 2006 Conference of the Center for Advanced Studies on Collaborative Research. ACM (2006)
216. Zhu, G., Mishne, G.: ClickRank: learning session-context models to enrich web search ranking. ACM Trans. Web **6**(1), 1:1–1:22 (2012)
217. Zhu, J., Hong, J., Hughes, J.: PageRate: counting web users' votes. In: Proceedings of the Twelfth ACM Conference on Hypertext and Hypermedia, pp. 131–132. ACM (2001)