Personalization in Text Information Retrieval: A Survey

Jingjing Liu*
School of Library and Information Science, University of South Carolina, Columbia, SC.
E-mail: jliujingjing@gmail.com

Chang Liu*
Department of Information Management, Peking University, Beijing, China. E-mail: imliuc@pku.edu.cn

Nicholas J. Belkin
School of Communication and Information, Rutgers, The State University of New Jersey, New Brunswick, NJ.
E-mail: belkin@rutgers.edu

Personalization of information retrieval (PIR) is aimed at tailoring a search toward individual users and user groups by taking account of additional information about users besides their queries. In the past two decades or so, PIR has received extensive attention in both academia and industry. This article surveys the literature of personalization in text retrieval, following a framework for aspects or factors that can be used for personalization. The framework consists of additional information about users that can be explicitly obtained by asking users for their preferences, or implicitly inferred from users' search behaviors. Users' characteristics and contextual factors such as tasks, time, location, etc., can be helpful for personalization. This article also addresses various issues including when to personalize, the evaluation of PIR, privacy, usability, etc. Based on the extensive review, challenges are discussed and directions for future effort are suggested.

Introduction

The information retrieval (IR) research community has long agreed that major improvement of search performance can only be accomplished by taking account of the users and their contexts, rather than through proposing new retrieval algorithms that have reached a plateau (cf., Keenoy & Levene, 2005; Sparck-Jones, 1995, 2000). As Dumais (2016) noted, the traditional "one-size-fits-all" search strategy that returns the same search results to the same query without considering who submits it or under what circumstances limits search engine performance in providing relevant search results.

*Corresponding author

Received June 24, 2018; revised March 5, 2019; accepted March 21, 2019

 • Published online May 6, 2019 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.24234

An alternative to the traditional search strategy is personalization of IR (PIR), which tailors search toward the individual search and searcher (White, 2016). The idea of considering users and their contexts for improving search effectiveness is not new. For decades, various models in the interactive IR field have incorporated this idea, such as Taylor's (1968) reference interview, Belkin's (1980) ASK (Anomalous State of Knowledge) model, and Dervin's (1992) sense-making theory, to name a few. However, most search engines had offered no or limited personalization features by 2001 (Khopkar, Spink, Giles, Shah, & Debnath, 2003), and only until recent years has personalization been employed in operational system design, for example, search engines return results or suggest queries based on the searcher's location if one is searching for weather, restaurants, and so on. Since then, PIR has become increasingly popular and has attracted a rich amount of effort in both academia and industry.

Personalization systems gather, store, and use information about the users and their situations beyond search queries. PIR is conducted by refining search queries and adapting search results to particular users. Those documents that the current users would desire are ranked toward the top of the result lists. Making the users' search experience as effective and pleasurable as possible is the goal of personalization (Belkin, 2008), and the focus on users and their contexts makes personalized search a compelling area (Pitkow et al., 2002).

Ke, Deng, Ng, and Lee (2005) classified personalization into three categories: content-based, link-based (e.g., PageRank), and function-based (c.f., Ng, Deng, & Lee, 2007). Belkin (2008) pointed out that personalization could be applied in various aspects including search result content, interface presentation, and search mode. Content personalization has been the focus of personalization

TABLE 1. Aspects/factors used for personalization.

General aspects		Specific factors/techniques used
Explicit relevance feedback (ERF)		Ask users for feedback
Implicit relevance feedback (IRF)	Search behaviors	Click-through, document dwelling time, browsing history, etc.
	Search behaviors + contextual factors	Individual differences, knowledge, task, location, time, etc.

research and is accordingly the focus of the current survey of the literature.

This article reviews personalization approaches that use both explicit and implicit sources (Table 1). Using explicit relevance feedback (ERF) means to explicitly ask for users' preference or interest of search terms, results liked, etc. Because of its intrusive nature, ERF is not much favored by users, although there are still attempts using it (Belkin, Cool, Koenemann, Ng, & Park, 1996). On the other hand, implicit relevance feedback (IRF) is used by most existing personalization approaches (Kelly & Teevan, 2003). Implicit sources help the system learn about user preference/interest by inferring or predicting it from observable user behaviors (click, browsing, dwelling time, etc.), and/or contextual factors (user background, tasks, etc.).

Information search is usually conducted to reach a certain goal at a given point in time and space, which altogether form the "context" of the search, and this can be used to enhance queries (Dumais, 2012). Backhausen (2012) noted that in personalization, both behavioral and contextual information are necessary. Although White (2016) differentiated personalization (tailoring search to the individual searcher) from contextualization (tailoring search to the searcher's situation), he also acknowledged that the distinction between them two is often blurred in the literature on mining and modeling search behavior. The current article reviews the literature about both personalization and contextualization in White's (2016) notion, taking into consideration all factors about the users and their contexts.

Although many researchers have noted that contextual factors are important and should be taken into consideration in IR research and system design (e.g., Belkin, 1993; Cool, 2001; Fan, Gordon, & Pathak, 2000; Croft, Cronen-Townsend, & Lavrenko, 2001; Ingwersen & Javerlin, 2005; Kelly, 2006a, 2006b; Dumais, 2007), Dervin (2003) pointed out that context is a term that is most "often used," least "often defined," and "when defined so variously" (p. 112). Context has numerous definitions in various fields, and could be of various types, for example, user context, software context, document context, and network context (Goker, Myrhaung, & Bierig, 2009). Allen (1996) stated that context is not a single thing, but rather is a composite of things comprised of several elements or aspects. Kelly (2006a, 2006b) highlighted the importance of Allen's (1996) matrix metaphor about context and addressed its influence on most discussion of context. Based on this, the current article uses "context" as an umbrella term that includes a variety of factors about both the user (e.g., demographic, knowledge) and their situations (e.g., tasks, time, location; more details in subsequent sections).

This article reviews the literature following the structure outlined in Table 1 after a description of personalization techniques. Personalization approaches are introduced using ERF, and IRF, respectively. Various issues in PIR follow, including when to personalize, evaluation of PIR, privacy, and usability issues. Based on the comprehensive survey, challenges of and future directions for personalization are finally discussed.

Before concluding the Introduction section, the scope of the current article is to be set. This article is a systematic review of the work that has been done in personalization of IR, and it is not intended to present any formal models in the way as was done in Pasi (2010) and Ghorab, Zhou, O'Connor, and Wade (2013). Personalization is a widely attended topic that has been researched in multiple areas, including search engine design, e-commerce (or e-business), human information interaction, and so on. It is beyond the scope of this article to address all these aspects, and indeed, some areas, such as recommender systems used in e-commerce, deserve separate reviews on their own. The focus of the current article is on personalization of text IR that aims at providing the most relevant search results to search queries, taking no account of the following:

- Collaborative filtering-based recommender systems
- Predefined customization based on user preference
- Personalization approaches in other contexts than searching, such as e-learning
- Personalization of search in other media format, such as TV, video, music
- Personalization of search on mobile-specific devices
- Personalization in the collaborative search scenario (the current article only focuses on individual search scenario)

The relevant articles selected in this survey are from multiple sources: databases (including ACM digital library, Library and Information Science Abstracts, PsycINFO, EBSCOhost, and PubMed), books, conferences, seminars, workshops, and so on. that are relevant to search personalization (to the authors' best knowledge). Searches in the above-mentioned databases used keywords "personalization," "personalized information retrieval", "personalized search" in the abstract or title metadata fields. All largely, partially, and marginally relevant search results were saved, and reviewed, with notes taking in spreadsheets about the articles' topics and relevance. Questionable articles were discussed by the authors to determine if they were relevant to the current survey, and which sections they could be included in, on a case-by-case basis.

Personalization Process and Techniques

User Modeling

User modeling is a major element distinguishing personalization and traditional IR systems. User modeling creates user

DOI: 10.1002/asi

profiles that store the users' search interests. Frias-Martinez, Magoulas, Chen, and Macredie (2006) summarized four steps of generating user profiles: (a) data collection that collects information about users, (b) preprocessing that analyzes the collected information to obtain semantic content about the user's interaction with systems, (c) pattern discovery that obtains structural descriptions of what has been learned about user behaviors, and (d) validation and interpretation that analyzes and interprets the structures obtained in the pattern discovery step.

It should be noted that user models could be both long and short term. Although user profiling is usually for long-term modeling of a user's constant and general interests and preferences, the user's instant information need that can be learned for short-term modeling is not always stored in a profile. Instead, such models are generated based on user behaviors in the current search session and used to directly predict document usefulness to the users in their current search, and return the most useful results according to the modeling user interest. This can be seen in many articles to be reviewed in the following sections, especially those predicting document usefulness from users' behaviors and contexts in the current session (see subsections "RLb. Personalization From Behaviors" and "RLd. Personalization From Behaviors and Contexts"). Short-term user information plays the same role as the long-term profile in terms of providing additional information for the system to personalize search results (c.f., Shen, Tan, & Zhai, 2005b), and studies (e.g., Bennett et al., 2012; Eickhoff, Collins-Thompson, Bennett, & Dumais, 2013) have found that combining both could improve personalization performance (more see subsection "RLb. Personalization from behaviors"). This, as expected, indicates that in general, more information about users could lead to a better understanding of users and their interests/intensions, which accordingly lead to better prediction of what they desire in the search.

Personalization Techniques

Two main techniques employed by personalization systems are query expansion and result re-ranking. Query expansion typically gathers additional information about user interest from various sources, represents user interest by some terms, and automatically adds these terms to the initial query for a refined search. The various sources could be different types. Some of them can be obtained from search behaviors (see subsection "RLb. Personalization using behaviors"), for example, preceding queries and search history (Bilenko, White, Richardson, & Murray, 2008; Cai & de Rijke, 2016); previously read documents (Biancalana, Micarelli, & Squarcella, 2008; Bilenko et al., 2008; Shen, Tan, & Zhai, 2005a); visited social bookmark services (Biancalana et al., 2008; Bouadjenek & Bouzeghoub, 2013; Carman, Baillie, & Crestani, 2008); and eye gaze behaviors (Buscher, van Elst, & Dengel, 2009), etc. Some of them can be obtained from other sources, for example, user tasks (Budzik & Hammond, 1999); context around and within queries (Bai, Nie, Bouchard, & Cao, 2007; Kraft, Maghoul, & Chang, 2005); desktop information (Chirita, Firan, & Nejdl, 2006, 2007); and user provided relevant pages (Chen & Sycara, 1998) or their annotations of the results (Jayarathna, Patra, & Shipman, 2013), what they have already known (Kelly, Dollu, & Fu, 2005) or their feedback about the search topics (Belkin et al. (2005), and so on (see PIR using Explicit Relevance Feedback (ERF)).

Result re-ranking techniques reorder search results for users according to document relevance. Some re-ranking systems involve user interaction, for example, the system presents the top n documents to the users for feedback and then refines ranking based on the feedback (e.g., Liu, Yu, & Meng, 2002; Tanudjaja & Mui, 2002). Most re-ranking systems do not require user involvement in re-ranking but have some sorts of pre-settled weighting criteria for re-ranking, giving heavier weight to those documents that match user interests and push them to top ranks (e.g., Agichtein, Brill, Dumais, & Ragno, 2006; Agrawal, Rantzau, & Terzi, 2006; Aktas, Nacar, & Menczer, 2004; Chirita, Nejdl, Paiu, & Kohlschutter, 2005; Chirita, Olmedilla, & Nejdl, 2004; Gauch, Chaffee, & Pretschner, 2003; Liu & Hoeber, 2011; Sieg, Mobasher, & Burke, 2007; Wang et al., 2013; You & Hwang, 2007).

Although query expansion and result re-ranking are two different techniques, they can be incorporated into single systems, for example, Pretschner and Gauch (1999), Pitkow et al. (2002), Shen et al. (2005b), Ferragina and Gulli (2005), and Lv et al. (2006). Incorporating both techniques has been demonstrated to enhance search performance compared to using either technique individually. Despite this result, many approaches reviewed in this article use only one technique, which could possibly be because of multiple reasons, such as testing the effectiveness of the technique that is of interest, or focusing on one technique with the limited computing resources, etc.

PIR Using Explicit Relevance Feedback (ERF)

Starting in the 1970s (e.g., Oddy, 1977; Salton, 1971), there have been relevance feedback approaches to learning user preferences beyond the queries they submit. ERF explicitly asks the user to provide feedback by means of, for example, specifying keywords, selecting and marking documents in terms of their relevance or usefulness, or answering questions about their interests/preferences. Despite that ERF requires users to spend extra effort beyond querying, it has been proved to be quite effective for improving retrieval accuracy (Rocchio, 1971; Salton & Buckley, 1990).

One ERF technique asks users to select favored additional keyword terms out of several candidates extracted from a certain sources. For example, the TaskSieve interface designed by Ahn, Brusilovsky, He, Grady, and Li (2008) provides a task panel on which important terms related to the task are extracted and displayed to users by various font sizes to indicate the different degrees of importance. Another

DOI: 10.1002/asi

ERF technique asks users to mark documents that they find relevant or useful, extracts keyword terms from these documents or users' responses, and then expand queries or re-rank search results. Vechtomova, Karamuftuoglu, and Lam (2003), and Shen and Zhai (2003) both found that compared with the original query search performance, queries expanded by this means showed significant improvement. The Rants interface designed by Gao and Jan (2010) allows users to edit and share search result ranks with each other, and edits of ranks can also be shared among similar queries. In addition, some commercial search engines also had similar approaches to allowing users to rank search results, such as U Rank by Microsoft[1], and SearchWiki by Google (released in November 2008 and discontinued in March 2010).

Some approaches use the information searchers' feedback for query expansion. Jayarathna et al. (2013) took user annotations of search results to help generate personalized results. Kelly et al. (2005) asked users to describe what they already knew and what they wanted to know about the search topics, and elicited terms from their responses for query expansion. Likewise, Belkin et al. (2005) elicited terms from user feedback and found that expanding queries by user provided terms outperformed the initial query. They found, however, that pseudo-relevance feedback (automatically expanding the query using the top n terms in the top search results without user involvement), performed at least as well as ERF. Rode et al. (2005) reached a similar conclusion, that user interaction improves search performance, but not better performance than automatic expansion performance. These findings indicate that although users may be a good source for relevance feedback, its usefulness may be limited with the tradeoff between user effort and performance improvement.

PIR Using Implicit Relevance Feedback (IRF)

This line of approach requires the system to infer user preference from their behaviors and/or contextual factors. The three sets of elements involved in this approach are:

1) user preference or interest, or document usefulness: this is the core value that a personalization system tries to infer, learn, or predict;
2) search behaviors: this refers to observable user activities in the search process; and
3) context: this sets and conveys the background and environmental information of the users who conduct the search.

Figure 1 illustrates the relationships among these three sets of elements. Dotted arrow lines denote the direction of "influence," in other words, the elements on the left side of the dotted arrow lines influence the elements on the right side (where the arrows point to). Dashed arrow lines denote the relation of "prediction," in other words, the elements on

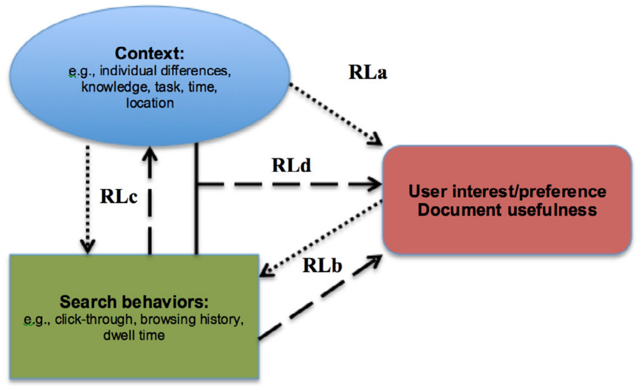

FIG. 1. The multiway relations between document usefulness, user behavior, and contextual factors (dotted arrow line denotes "influence"; dashed arrow line denotes "can predict"; the solid line denotes combination). [Color figure can be viewed at wileyonlinelibrary.com]

the left side of the dashed arrow lines can predict the elements on the right side of (where the arrows point to).

The following describes the relations (RLs) among these three sets of elements.

RLa: Context & Interest. Simply put, RLa indicates the relation that contexts influence users' search interest. Contexts include a variety of factors, including the goal that drives the user to the search, his or her background, knowledge, time, location, and so on, which are all likely to influence the user's search interest, or what document the user may deem useful. Research about influential factors on relevance and usefulness is along this line. For example, Teevan, Dumais, and Horvitz (2010) found that different people's explicit judgments for the same queries differed greatly. Taylor, Cool, Belkin, and Amadio (2007) found that users' relevance judgment was statistically associated with the stage of their search. Because this RL does not directly involve search behaviors, based on which user interest can be predicted, research about this RL is not to be reviewed in more detail.

RLb: Interest & Behaviors. RLb characterizes the connection between search interest and behaviors. Users' search interest could influence search behaviors, and on the other hand, search behaviors could predict user interest. Research on user interest influencing search behaviors is rarely seen (and therefore it is not reviewed), but it seems to be a well-accepted assumption that one's search interest, that is, the goal that one wants to reach in the current search, largely determines the user's search behaviors, for example, what documents she may click in, how long she dwells on the documents. Meanwhile, research on the other direction of the relationship, that is, predicting search interest (mainly represented by query and document relevance or usefulness) from search behaviors was the main approach of early PIR algorithm and system design.

RLc: Context & Behaviors. RLc describes the relation between contexts and behaviors. Contexts could influence users' search behaviors, and on the other hand, search behaviors could predict contexts. Much work has been done to

[1] https://searchengineland.com/u-rank-microsofts-social-search-experimental-site-15018

DOI: 10.1002/asi

explore how contextual factors may influence search behaviors; in recent years, some effort also tried predicting contexts from behaviors.

RLd: (Context + Behavior) & Interest. RLd conveys the idea of combining behaviors with contexts to predict search interest, or in other words, taking contexts into consideration when predicting search interest from behaviors. Although RLs a, b, and c are quite straightforward one- or two- way relations, RLd is three-way which sets up the link between all three elements. This relationship establishes a foundation for a wide range of future research directions in PIR (more discussions later).

The following review starts with RLb, moves on to RLc, and then RLd.

RLb. Personalization Using Search Behaviors

Many PIR approaches predict user interest from search behaviors only, which can be reading time (or dwell time) on search engine result page (SERP) or on documents, clickthrough, document saving behaviors, and so on. Dwell time has been extensively studied as an indicator for inferring document usefulness. The findings have been mixed: some found that reading time had a strong positive correlation with user's rating of the retrieved document's usefulness (e.g., Claypool, Le, Wased, & Brown, 2001; Morita & Shinoda, 1994, 1994), whereas others found that reading time was not significantly related to the user's subsequent relevance judgment (e.g., Kelly & Belkin, 2001). Kelly and Teevan (2003) pointed out that the implications of user behaviors on relevance judgment might be affected by contextual factors such as task, topic, and collection differences (cf., Kelly & Belkin, 2001; see more about this in the subsection on RLd).

Another behavior having received much attention is clickthrough. Workshops on personalization using web search click logs have been organized in conferences such as WSDM (Web Search and Data Mining; cf., Serdyukov, Dupret, & Craswell, 2014), fostering the research and practice of PIR. Joachims (2002) and Radlinski and Joachims (2005) used clickthrough data to infer relevance feedback. Joachims, Granka, Pang, Hembrooke, and Gay (2005) further analyzed the users' decision process using eye-tracking and compared IRF against manual relevance judgments. Although they found that clicks could be biased, which makes the interpretation of clicks as absolute relevance judgments difficult, they also found that that relative preferences derived from clicks were reasonably accurate on average.

In addition, retention (or saving) has been found to be potentially useful as implicit evidence for identifying documents that could be used as seeds for re-weighting (Kelly, Diaz, Belkin, & Allan, 2004). This is quite intuitive. Nevertheless, its usefulness is limited to users explicitly saving documents, which does not always happen.

Many approaches use multiple types of search behaviors from search logs to infer user interest. The behaviors usually are a combination of several out of click-through, browsing, scrolls, mouse movements, querying behaviors, sequential patterns, and so on. Models or algorithms have been built from a set of search behaviors to "recognize" users by topic interest (Wedig & Madani, 2006) or search trail features (White & Drucker, 2007), re-rank search results (Agichtein, Brill, & Dumais, 2006; Agichtein, Brill, Dumais, & Ragno, 2006; Jiang, Leung, & Ng, 2011; Matthijs & Radlinski, 2011; Shen, Chen, Hu, & Yang, 2012; Wei, Sen, Yuan, & Chen, 2009), diversify search results (Radlinski & Dumais, 2006), predict query goal being navigational (searcher has a particular web page in mind) or informational (searcher does not have a particular page in mind) (Lee, Liu, & Cho, 2005), learn user interest (Daoud, Tamine-Lechani, & Boughanem, 2008; Daoud, Tamine-Lechani, Chebaro, & Boughanem, 2009), predict user search interest switch (Guo & Agichtein, 2009), and so on. By demonstrating the effectiveness of IRF in differentiating users or their interests, these studies provided foundations for personalization.

Some studies compare multiple models to reach the best performance. For example, Shen et al. (2012) built 3 personalization models based on click-through data: a matrix factorization click model (MFCM) that relates queries and documents collaboratively, a personalized click model (PCM) that exploits the latent relationships of users, queries and documents, and a hybrid personalized click model (HPCM) that emphasizes query-document interactions while characterizing user variations simultaneously. Evaluation showed that MFCM performs well for capturing latent features of queries and documents, PCM is better for the capability of personalization, and HPCM achieved the greatest improvement by combining the strengths of the other two models. More recently, Chen, Cai, Chen, and de Rijke (2017) developed a personalization algorithm that considers only queries with clicked documents being taken as search context rather than all queries.

Besides search behaviors in one search session (instant and short term), research has also considered search history in previous sessions (long term) (e.g., Albanese, Picariello, Sansone, & Sansone, 2004; Sontag et al., 2012; Ustinovskiy & Serdyukov, 2013; Wedig & Madani, 2006). Regarding the usefulness of short- and long-term behaviors for personalization, Zhu, Callan, and Carbonell (2008) found that incorporating short-term contexts works well for personalized search, and longer history does not provide further improvement. However, Bennett et al. (2012) found that historic behaviors provided substantial benefits at the start of a search session; short-term session behaviors contribute most gains in an extended search session; and the combination of session and historic behaviors outperformed either alone.

Eickhoff et al. (2013) characterized and personalized search for atypical search sessions, that is, instances when users diverge from their search profiles to satisfy information needs outside their regular areas of interest. They found that certain topics such as medical information and technical support were much more likely to arise in atypical sessions, along with query features such as increased term count, more unique terms, and more natural language-type terms. They showed that atypical

DOI: 10.1002/asi

sessions could be successfully identified using features based on short term only, long term only, or a combination of both, and the performance showed an increasing order among the three.

In addition, eye-tracking data was found to be helpful in personalization approaches. Buscher et al. (2009) compared webpage display time and eye-tracking data used as IRF sources for query expansion and result re-ranking. They found that although feedback based on display time was much coarser than feedback from eye tracking, surprisingly, it worked as well as eye-tracking based feedback. Personalization methods using display time and eye tracking data all performed significantly better than a nonbehavior-based baseline and especially improved poor initial rankings of the Web search engine. Buscher, Dengel, Biedert, and van Elst (2012) also demonstrated that eye-gaze data about what was read and how it was read was very valuable to determine whether viewed document parts were relevant to an individual user, and further, that using this information as IRF can greatly improve PIR.

RLc. Relationship Between Contextual Factors and Search Behaviors

RLc is two directional, meaning that contextual factors can influence search behaviors, and can also be predicted from search behaviors. However, most of the related studies belong to the former and therefore this subsection has more weight on this; only a few studies attempted to predict contexts from behaviors; these are reviewed in corresponding sub-topics.

As mentioned before, contexts can refer to a variety of factors. These include both factors about users, for example, demographic information, cognitive style, personality, knowledge, and so on and factors that are nonuser but situation related, for example, task, time, location, and so on. Table 2 lists those factors, which are reviewed in the following subsections.

Demographic information. A rich amount of research can be found in the literature about how demographic characteristics influence information search, which provides evidence for personalizing search based on demographic information. Males and females were found to differ in three aspects: navigation patterns, attitudes and perceptions (Chen & Macredie, 2010); boys were found to be more active than girls while online (Large, Beheshti, & Rahman, 2002); girls had a tendency to be vertical searchers whereas boys had a tendency to be horizontal searchers (Roy & Chi, 2003); males read further in the result list and made fewer regressions when they were reading abstracts (Lorigo et al., 2006); males prefer nonlinear reading patterns more than do females (Liu & Huang, 2008); females often feel disoriented or lost on the Web (Ford, Miller, & Moss, 2001); male students often had positive attitude toward computers and the Web (Schumacher & Morahan-Martin, 2001), and so on. Ford et al. (2001) suggested that Web-based applications should be developed to support adaptation to gender because males and females might need different levels of support.

TABLE 2. Contextual factors.

General aspect	Category	Subcategory (or specifics)
User-related	Demographic information	Gender, age, etc.
	Cognitive style	Field-dependent/independent
	Cognitive ability	Perceptual speed, working memory, dyslexia
	Personality	Big 5 factor
	Knowledge	Domain knowledge, topic knowledge
	Search experience	
Situation-related	Task type	Task product, complexity, difficulty, stage, others
	Location & Time	Location, time
	Information object features	FAQs, list format, genre
	Others	Language, health literacy, social network, etc.

However, research in this area had inconsistent findings. Although the above listed studies identified different search behaviors between males and females, Hupfer and Detlor (2006) showed no significant differences between males and females in search behaviors. Kim, Lehto, and Morrison (2007) reported that females claimed more positive attitudes than males, and Koohang and Durante (2003) found no significant gender users' attitudes toward web-based instruction program.

Some studies suggest that multidimensional demographic attributes (gender, age, marital status, income, education level) should be taken into consideration. Research has shown that different demographic groups differ in their queries and clicked URLs and that search result performance could be improved to some extent by applying demographic information (Weber & Castillo, 2010), or could benefit by using the "groups" they belong to (Teevan, Morris, & Bush, 2009). Mislove, Viswanath, Gummadi, and Druschel (2010) also found that users with common attributes are more likely to be friends and often form dense communities, and they proposed a method of inferring user attributes for detecting communities in social networks.

In addition to these studies exploring the influences of demographic information on search behaviors, some recent effort has been devoted to predicting users' demographic attributes from information behaviors. For example, Otterbacher (2010) inferred gender of movie reviewers by exploiting their writing style, content and metadata. Graepel (2013) infers users' demographic attributes based on users' social and search data on Facebook and achieved good accuracy in prediction. Wang, Guo, Lan, Xu, and Cheng (2016) conducted demographic prediction based on users' purchase history in a retail scenario. Zhong, Yuan, Zhong, Zhang, and Xie (2015) inferred users' demographic attributes from location check-ins on a social media platform in China (weibo.com).

Cognitive style. Cognitive style is a construct used to describe an individual's habitual mode of perceiving, remembering, thinking and problem solving (Riding & Cheema, 1991). It is believed to be stable throughout an individual's

DOI: 10.1002/asi

lifetime, therefore, knowing one's cognitive style could help build long-term user profiles for PIR.

Among the many types of cognitive styles that have been developed, field dependence-independence has been the most frequently studied pair. Field dependent (FD) people typically experience surroundings in a relative global manner, are easily influenced by the environment, and are more passive, that is, willing to accept ideas as presented; conversely, field independent (FI) people usually experience surroundings with an internal perspective, process information with their own structure, and are more active, that is, tend to accept ideas through analysis (Witkin, Moore, Goodenough, & Cox, 1977).

The effects of FD and FI on information search have been examined for two decades or so with rich findings. Compared with FI users, FD users tend to spend longer time in completing search tasks (Palmquist & Kim, 2000), use Boolean search more than Best Matching (Ford et al., 2001), make heavier use of the main menu and previous/next buttons, take more navigational moves, and engage in more duplicated pages (Chen & Ford, 1998), consult the user guide longer in a search interface where they can only follow the restricted links to other pages, and vice versa in the free access version (Dufresne & Turcotte, 1997). FD and FI have also been found to influence users' reactions to the organization of subject categories, presentation of the search results, and screen layout (Chen, Magoulas, & Dimakopoulos, 2005).

Some researchers have attempted to predict cognitive styles from users' behaviors. Frias-Martinez, Chen, and Liu (2007) demonstrated how different classification systems could be used to automatically identify the user's cognitive style based on the user's interactions with a digital library. Nevertheless, more work in this direction is needed.

Cognitive ability. Various types of cognitive abilities have been studies regarding how they influence online information search. Perceptual speed (PS) characterizes one's speed in comparing and scanning to find visual objects such as figures or symbols (Ekstrom, French, Harman, & Dermen, 1976). It has been found that high PS users located relevant materials more effectively, that is, showing higher precision and recall (Allen, 1992); and more efficiently, that is, using less time (Al-Maskari & Sanderson, 2011). High PS users also reported lower levels of workload (Brennan, Kelly, & Arguello, 2014). Working memory (WM) involves holding information in mind and working with information that is no longer perceptually present (Diamond, 2013). High WM searchers have been found to complete search tasks faster (Gwizdka, 2010), and performed more actions (Gwizdka, 2017). MacFarlane et al. (2010) examined search differences between dyslexic (DS) and nondyslexic (ND) searchers and found that DS searchers had fewer searches, and examined fewer documents in total, but examined more documents per iteration.

Given that many cognitive abilities tend to be stable, predicting cognitive abilities from search behaviors seems to be useful. With the detected differences between searchers with different levels of cognitive ability, the above-mentioned studies build foundations for predicting cognitive abilities from search behaviors. However, it should be noted that predicting might not be straightforwardly easy because of the interplay of the multiple aspects of cognitive ability. For example, both high PS and high WM users were found to complete tasks faster than their counterparts; given the observation of short task completion time, how could it be determined that the searcher is high PS, or high WM, or both? This issue requires more research effort.

Personality. Heinström (2003, 2005, 2006) examined the relationships between Big 5 factors (Neuroticism, Extraversion, Openness to experience, Agreeableness, and Conscientiousness) and search behaviors using questionnaires. The research found that information behaviors could be connected to all the tested personality dimensions. Schmidt and Wolff (2016) revisited the above relationship and found selective correlations of slight and intermediate strength between personality dimensions and search behaviors. There are relatively few studies characterizing the relationship between personality and search behaviors, and it seems that more effort is needed to infer personality from search behaviors.

Knowledge. Knowledge in IR research can be subject domain knowledge or task topic knowledge.[2] The former is one's knowledge of a general subject domain of the search task; the latter is one's knowledge of the specific search task topic. As can be seen from their definitions, these two types of knowledge are measured by different methods and have led to different research findings. The following first introduces studies about subject domain knowledge, and then studies about task topic knowledge.

Among the studies examining the effects of domain knowledge in IR research, some examined how it affects search performance measured by precision and recall (e.g., Allen, 1991; Marchionini, 1989). These studies hardly detected relationships between domain knowledge and search performance, making it difficult to draw conclusions for personalization. On the other hand, some studies looked at user's search behaviors and detected relationships between domain knowledge and search behaviors. Higher domain knowledge was found to be associated with less use of a thesaurus for term suggestion (Hsieh-Yee, 1993), more effective term selection when using a thesaurus (Sihvonen & Vakkari, 2004), less use of synonyms and combinations (Hsieh-Yee, 1993), wider and more specific vocabulary in search terms (Vakkari, Pennanen, & Serola, 2003), more efficient selection of concepts to include in the search (Wildemuth, Freund, & Toms, 2014), fewer errors in the reformulation of search tactics (Wildemuth, 2004), more queries per search task (Zhang, Anghelescu, & Yuan, 2005), longer queries (White, Dumais, & Teevan, 2009; Zhang et al., 2005), less time spent on webpages (Duggan & Payne, 2008), and more keyword-based than tag-

[2] Another type of knowledge, search knowledge, is treated in this article as search experience that is discussed below.

based queries in a social tagging system (Kang, Fu, & Kannampallil, 2010). But there are also inconsistent findings, for example, Duggan and Payne (2008) found that higher domain knowledge was associated with shorter queries in the specific domain of football. Despite such inconsistencies, all these studies implied the possible benefits of providing different systems or system features to users with different levels of domain knowledge, so that the system (features) best support different search tactics of query formulation and reformulation.

Regarding users' knowledge of search task topics, previous studies have found that users with higher topic knowledge issued longer and more complex queries (Hembrooke, Granka, Gay, & Liddy, 2005), used elaborations as a reformulation strategy more often than simple stemming and backtracking modifications (Hembrooke et al., 2005), used more search expressions (Allen, 1991), had shorter document reading time (Kelly & Cool, 2002) and document display time (Kelly, 2006b), and had a higher ratio of saved documents to total viewed documents (Kelly & Cool, 2002). All these results indicate the possibility of inferring topic familiarity from search behaviors. Kumaran, Jones, and Madani (2005) built a classifier to effectively differentiate documents using various document features (e.g., stop-word, line-length) to match searchers' different levels of topic knowledge. This method can be effective in biasing result ranking for topic knowledge and can help in inferring the knowledge of users who read different categories of documents. Meanwhile, further efforts are needed to tell which specific documents may be predicted as useful based on topic knowledge and reading time, and/or the user's saving, viewing, and other behaviors. Some researchers further attempted to predict knowledge from search behaviors such as first dwell time on SERPs, number of saved documents, number of unique queries, etc., and obtained satisfactory results (e.g., Liu, Liu, Cole, Belkin, & Zhang, 2012).

Search experience. Studies have found that compared with search novices, search experts with more search experience tended to have better search performance (Ahmed, McKnight, & Oppenheim, 2004), higher proficiency in locating websites (Lazander, Biemans, & Wopereis, 2000), and modify the system's default search field and search results display format (Li, Zhang, Liu, & Zhang, 2006). Meanwhile, it should be noted that some research indicated no significant relationship between user's previous search experience and the levels of relevance of search results. For example, Ford et al. (2001) did not detect any relationship between levels of relevance and user's experience variables, which fails to "confirm a number of previous findings in relation to the role of experience" in search (p. 1063). Similar to the above factors, this area also calls for effort on predicting search experience from behaviors.

Tasks. In IR, task is a significant element and has been continuously attracting research attention over time. Essentially, users search for information for the purpose of reaching their goals, or more specifically, solving their tasks at hand. an extensive amount of effort has been spent on examining the effects of different tasks or task types on information searchers' behaviors and performance, which provides significant evidence for personalization based on tasks.

A commonly seen basis of this stream of research is to classify tasks into different types along some task feature(s). These include, for example, closed versus open-ended tasks (Marchionini, 1989); specific vs. general tasks (Qiu, 1993); factual, descriptive, instrumental, and exploratory tasks (Kim, 2006); fact-finding vs. information gathering (Kellar, Watters, & Shepherd, 2007; Toms et al., 2007); learning about a topic, making a decision, finding out how to, finding facts, and finding a solution (Freund, 2008), and so on. With various standards and definitions of task classification, it is difficult to compare findings across studies, so it necessary to have standard classification schemes. Li and Belkin (2008) developed a quite comprehensive task classification scheme which includes several dimensions to classify tasks: task product, objective complexity, subjective complexity, difficulty, and urgency to name a few. The following reviews studies along some of these dimensions.

Task product. Along the dimension of task product, Li and Belkin (2008) classifies tasks into several categories: intellectual (a task producing new ideas or findings), decision or solution (a task making a decision or solving a problem), factual (a task locating facts, data, etc.), and mixed (of two or more of the above). Studies have found that task product significantly affected the number of IR systems consulted and result pages viewed (Li & Belkin, 2010; Liu et al. 2010), time to accomplish task (Liu et al., 2010), query length, search success (Li & Belkin, 2010), and eye movement (Cole et al., 2010). Using similar task type categories, Kellar et al. (2007) found that compared with fact-finding tasks, information-gathering tasks required participants to view more pages, use Web browser functions more heavily, and spend more time completing the task.

For this dimension of task type, there has been effort on predicting task types from search behaviors. Using TREC 2014 session track tasks, Mitsui, Liu, and Shah (2018) tried predicting whether task type was intellectual or factual. They found that even though significant behavioral features improve prediction over baselines, the improvement was minimal. They suggested that in some cases, considering personal patterns might help with effective prediction.

Task complexity. Li and Belkin (2008) defines the degree of objective task complexity according to the number of activities in a work task or the number of information source types in a search task. It has been found to affect many aspects of interactive information searching behavior, for example: the number of visited documents, the number of search sources, the number of queries, time used to complete tasks (Li & Belkin, 2010; Liu et al., 2010), and eye movement (Cole et al., 2010). White, Ruthven, and Jose (2005) defined task complexity in a similar way, according to the number of potential information sources and type of information required to complete a task. Their study found that users preferred IRF for more complex tasks, but they preferred ERF for less complex tasks. This study implies that to avoid

DOI: 10.1002/asi

task bias, task complexity should be considered when designing systems involving IRF or ERF. Because different types of RF are appropriate for tasks with different levels of complexity, it might be beneficial to use both types of RF simultaneously in a system which can automatically detect task complexity and switch between the two modes of RF.

Other researchers have looked at task complexity using different definitions. Byström and Järvelin (1995) defined task complexity from the worker's point of view based on "*a priori* determinability of, or uncertainty about, task outcomes, process, and information requirements" (p. 194). Vakkari (1999) followed this definition in his proposed model relating task complexity and information actions. Though their definition is different from Li and Belkin (2008), this line of research obtained similar findings; that more complex tasks require more user interactions, and, for example, that task complexity is related to information type and information source selection.

Task difficulty. One task type dimension that is closely related to but different from complexity is difficulty. As with complexity, difficulty has been determined both "objectively," for instance, according to precision of search results (e.g., Zhang, Liu, Cole, & Belkin, 2015) and subjectively, although task difficulty much more usually emphasizes the task doer's subjective perception of how difficult the search is (Campbell, 1988; Li & Belkin, 2008; Wildemuth, Freund, & Toms, 2014; Kelly, 2015). Similar to task complexity, task difficulty is also an important aspect having been found to show significant effects on information search, including: number of queries (Kim, 2006; Liu et al., 2010; Kelly, Arguello, Edwards, & Wu, 2015), query length and diversity (Kelly et al., 2015), number of SERP clicks (Kelly et al., 2015), number of visited documents (Kim, 2006; Gwizdka, 2008; Liu et al., 2010; Kelly et al., 2015), number of relevant/saved documents (Gwizdka, 2008; Kim, 2006), task completion time (Kim, 2006; Gwizdka, 2008; Liu et al., 2010; Kelly et al., 2015), document dwell time (Liu et al., 2010), and so on.

Some studies attempted to predict search task difficulty from search behaviors, including: search effort, navigational speed (Gwizdka & Spence, 2006), the number of SERPs, the number of visited documents, and the number of relevant documents (Gwizdka, 2008; Liu, Gwizdka, Liu, & Belkin, 2010; Liu et al., 2012; Arguello, 2014). It was also found that task difficulty could be well predicted in the early stage of the search process (Liu et al., 2012).

Stage of task. Although not included in the Li and Belkin (2008) scheme, stage of task is an important contextual factor in IR and has been investigated for decades regarding the information seeker's affective, emotional, and physical action changes during the information seeking process. Kelly's (1963) construct theory, Taylor's (1968) four levels of information need along the different stages of search, and Kuhlthau's (1991) ISP (Information Seeking Process) model are all classic theses about task stage. Task stage has been found to affect query term selection (Vakkari, 2000; Vakkari, 2001; Vakkari & Hakala, 2000), relevance judgment criteria (Taylor et al., 2007; Vakkari, 2000; Vakkari, 2001; Vakkari & Hakala, 2000), the use of IRF and ERF by users in search (White et al., 2005), document dwell time (Liu & Belkin, 2015), and utility of search interface features (Huurdeman, Wilson, & Kamps, 2016).

Task types along other dimensions. There are other task classification dimensions studied in the literature. Task structure, that is, the relationship between the sub-tasks in complex tasks, was found to affect search behaviors (Toms et al., 2007) and play roles in inferring document usefulness from first dwell time (Liu & Belkin, 2010; Liu & Belkin, 2015). Task level, that is, whether a task required judging a document as a whole, or only a part or parts of a document, for usefulness, was found to affect task completion time, number of visited documents, number of search sources, number of queries (Liu et al., 2010), and eye movement (Cole et al., 2010). Some recent efforts contribute to the prediction of task types (Liu, Cole, Baik & Belkin, 2012), which is a further step toward personalizing search based on task types.

Location and time. Besides search engines such as Google that personalize search based on the searcher's location, the literature has seen more effort in this direction. Welch and Cho (2008) presented a method to automatically identify queries that would benefit from localization. They asked real users to identify localizable queries, then determined a set of relevant features of these queries, and then used conventional machine learning techniques to classify these queries. Jones, Hassan, and Diaz (2008) found that the introduction of geographic features slightly improved search performance for queries containing geographical information. O'Brien, Luo, Abou-Assaleh, Gao, and Li (2009) proposed a model for personalizing search results in a local search engine by adding a geographical dimension. Bennet, Radlinski, White, and Yilmaz (2011) personalized search using both the searchers' physical location and a more general notion of locations of interest for Web pages which were computed using implicit user behavioral data that characterize the most location-centric pages. Shokouhi (2013) found the user's location and long-term search history were the most effective for personalizing auto-completion rankers. Belkin et al. (2003, 2004) found that although geographical information can potentially improve search performance, giving it too much weight (compared to topical information) decreased performance.

Time information has also been found useful for personalizing search. Mei and Church (2008) found that query volumes and search difficulty (click entropy) varied by time of day and day of week. They suggested that variables such as IP addresses, time, and geographical information could be viewed as convenient surrogates for more sensitive market segmentation variables.

Information object features. Kelly, Murdock, Yuan, Croft, and Belkin (2002) and Murdock, Kelly, Croft, Belkin, and Yuan (2007) found that relevant documents' types (with different document features) varied according to search task types

DOI: 10.1002/asi

(procedure oriented or fact oriented). Specifically, documents with FAQs and lists tended to be judged relevant in procedure-oriented tasks rather than fact-oriented tasks. This indicates that the types of search tasks might predict which types of document are relevant. AbdulJaleel et al. (2003) found that document genre helped improve search performance in some cases but not in general. Given these restrictions, for the examined factors to be effectively used in personalization system design, more research is needed to clarify when and how they should be considered and incorporated in personalization.

Other contextual factors. Lopes (2013) studied several contextual features that can be used in supporting query formulation for personalized search: language, health literacy, and topic familiarity. Evaluation results showed that a personalization system biasing query suggestions using English proficiency and health literacy outperformed a system providing non-personalized query suggestions, which outperformed a system without query suggestion.

Some studies use semantic contexts to personalize search. Leung, Fung, and Lee (2011) proposed a method that constructs a network of concepts related to different semantic interpretations of the query. Concepts were extracted from the search results of each query. Leung, Lee, Ng, and Fung (2012) built concept-based user profiles from query history, search results, and clickthrough data. Steichen, O'connor, and Wade (2011) also proposed providing personalization according to semantics besides social and open-web resources. Tiwari, Husain, Srivastava, and Agrawal (2011) used the content semantics and the structural properties of a website to improve the effectiveness of web personalization.

Some studies made use of social networks as sources for personalization. Kashyap, Amini, and Hristidis (2012) used three types of feedback from social network groups with decreasing strength: the document preference of the user herself, that of the users in her social group, and that of other users in the network. Wang and Jin (2010) presented a personalization system that retrieves information from multiple social systems (e.g., blogs, social bookmarks and mutual tags) and creates an interest profile for each user by integrating different streams of information. They also found that integrating information from multiple social systems led to better personalized results than relying on the information from a single social system. Teevan et al. (2009) explored the similarity of query selection, desktop information, and explicit relevance judgments across people grouped in different ways. They found that the explicitly formed groups were similar in many aspects when considering queries related to their group's theme but were less cohesive for off-theme queries. Their findings also demonstrated that groupization (as compared to personalization) improves on personalization for several group types, particularly for explicit groups and group-related queries.

RLd. Personalization Using Behaviors and Contextual Information

As mentioned previously, previous studies looking at the relationship between document usefulness and dwell time (e.g., Kelly & Belkin, 2001; Morita & Shinoda, 1994) have generated seemingly conflicting findings. The studies had different settings. Morita and Shinoda (1994) attempted to design a filtering system based on monitoring user behaviors in a usenet system, and they found a strong tendency for users to spend a greater length of time reading those articles rated as interesting than those rated as not interesting. In a different setting which was interactive in nature, Kelly and Belkin (2001) asked users to perform search tasks, and they found that the length of time that a user spent viewing a document was not significantly related to the user's subsequent relevance judgment. They suggested that contextual factors such tasks, document collection, and searching environment may affect the relationship between document relevance and reading time. Kelly and Belkin (2002) further hypothesized personalized behavioral models and proposed an example of how users' level of topic familiarity may affect users' reading time for relevant versus nonrelevant documents.

Kelly and Belkin (2004) found that dwell time differed significantly according to specific tasks and specific users, which provided evidence that inferring the usefulness of a document from dwell time should be tailored toward individual tasks and/or users. White and Kelly (2006) found that tailoring display time threshold based on task information improved IRF algorithm performance over a general dwell time-only threshold, and that dwell time was able to successfully predict document usefulness when the task information is considered. Liu and Belkin (2015) further found that in a parallel-structured multisession task, users' knowledge of task topics could help interpret first dwell time (the time duration from first viewing a document to first exiting the document) as a reliable indicator of document usefulness. This demonstrated the possibility of personalizing search using dwell time for users with different knowledge levels. Liu, Liu, and Yan (2018) examined the relationships between time constraints, search behaviors, and document usefulness judgment. They found that time constraints and usefulness had interaction effects on first dwell time, indicating that knowing time constraints helps predict document usefulness from dwell time.

Some effort has been spent on designing personalization models or systems using search behaviors and contextual factors. Teevan, Dumais, and Horvitz (2005b) explored rich models of user interests, built from both behaviors (e.g., issued queries and visited webpages) and contexts (e.g., documents and emails the user has read and created). Ahmed, Teo, Vishwanathan, and Smola (2012) used click history and contextual factors including article entities and concepts, and recency to personalize news stories to the current users. Liu, Belkin, and Cole (2012) modeled search behaviors in different search tasks to infer document usefulness and then used this information to modify queries. They found that specific prediction models for task types outperformed a general prediction model. White et al. (2013) explored a task-based method to personalize users' current search results. They mined historic search-engine

DOI: 10.1002/asi

logs to find other users performing similar tasks to the current user, and then leveraged their on-task behavior to identify Web pages to promote in the current ranking. This study demonstrates the value of considering search tasks in addition to just search queries during personalization. Agichtein, White, Dumais, and Bennett (2012) developed prediction models of future task continuation by considering features about tasks, engagement effort and focus, user profile, and repeating status.

Issues in Personalization

When to Personalize

Research has indicated that personalization could have various effects in various situations, making it an important issue to consider when to personalize. Teevan, Dumais, and Liebling (2008) found a lot of variation across queries in the benefits that can be achieved through personalization. Teevan, Dumais, and Horvitz (2005a) found that personalization has variant effectiveness on various query types, and that personalization is not necessary on less ambiguous queries. Teevan, Liebling, and Geetha (2011) also found that different people often use the same queries to navigate to different resources. Dou, Song, and Wen (2007) found that personalized search had significant improvement over common web search on some queries, but it had little effect on other queries, for example, queries with small click entropy. It even harmed search accuracy under some situations. Patel, Tendulkar, and Chakraborti (2013) found that for single word expansions, 69.10% queries benefited from user specific expansion, whereas 30.90% did not.

Teevan, Dumais, and Horvitz (2007) quantified the potential value of personalizing search results based on the investigation of the diverse goals, or task types, that people have when they issue the same query, and the ability of search engines to address such diversity. They took the gap in normalized DCG (discounted cumulative gain) between the individual and group personalized ranking as an indication of the potential gain that can be achieved by personalizing rankings. Predictive models were built to identify queries that could benefit from personalization. Luxenburger, Elbassuoni, and Weikum (2008) proposed a model that could detect when the user's search and browse history was not appropriate for aiding the user in satisfying her current information quest. Li, Yang, Qi, Li, and Zhao (2010) found that queries issued by more than six and less than 20 users had the biggest potential for personalization in AOL, and queries issued by more than six and less than 4000 users had the biggest potential for personalization in Sogou (a Chinese search engine). They attributed the difference between the two systems to the many recommended queries on Sogou which were informational and were clicked by many users. Chen, Yang, Li, Zhao, and Qi (2010) investigated two kinds of strategies for predicting query's potential for personalization: classification and regression and found that the classification model performed better for predicting query potential for personalization.

Diversity

A related issue to personalization is diversity. This is a significant issue, especially when considering that personalization, while trying to return the most useful results toward the particular searcher, could also potentially limit the search results to only certain specific aspects of users' queries or reduce the chances to be exposed to new ideas (Carbonell & Goldstein, 1998). The concept of "filter bubbles" (Pariser, 2011) addresses the possibility of intellectual isolation and the neglect of those important but less clicked information sources brought by filtering and personalization.

The relationship between personalization and diversity seems to be conflicting, but they could be combined. Personalization does not focus on specific topics, but more on the task and the context. Personalization aims to maximize the satisfaction for each of the individual users and generates models to provide search results that are like individual users' search interests. Diversity is often seen as a contradictory approach to personalization because one goal of diversity can be said to be to satisfy as many users as possible with a single result list. Radlinski and Dumais (2006) presented three methods for diversifying search results for a given query using past query reformulations. They suggested that the top-k results of a query might not be diverse enough to contain representative documents corresponding to different interpretations of the query, thus might not be useful for personalized re-ranking of results. They explored methods of generating related queries to yield a more diverse set of search results.

But another goal of diversity is to ensure that the searcher is not burdened with repetitive information, and to ensure that all aspects of a topic are presented. These were primary motivations of Carbonell & Goldstein (1998)'s MMR, and this suggests a possibility of complementarity of some types of diversity and personalization. More recent research has explored methods to complement diversity with personalization to enhance each other. Vallet and Castells (2012) introduced a method that combined personalization and diversification, and the experiments showed that this method improved both the accuracy and the diversity values. They suggested that diversification has a role in personalization when the system lacks information about searchers' preferences or when users' preferences are diverse themselves. Diversity and personalization can be combined in different ways, which needs further research.

Evaluation

Evaluation is crucial for understanding whether any proposed personalization technique is valuable, and it is challenging to accomplish, given that evaluation criteria and measurements for this purpose are usually not simple or straightforward but multifaceted and complex. There has not been an agreed overall framework for PIR in the current literature, and different PIR approaches have used various means for evaluation, including batch experiments, online experiments using A/B tests (also named bucket

DOI: 10.1002/asi

testing, which is to split the users into two groups and assign them with two versions of the same system variable), and laboratory-based user studies. Carterette (2018) describes the limitations and advantages of the various methods: batch experimentation makes strong assumptions about users and their needs in context, but are replicable; user studies have small N, high cost, and high variance, but yields deep information; and online experimentation using A/B tests requires a large user base and has some ethical problems, but gives robust results.

There are only a few studies that have attempted online evaluation of personalization by conducting controlled experiments, with real-user participants working in the personalized versus non-personalized systems. Ma, Pant, and Sheng (2007) compared users' performance as well as their perceptions on using these two types of systems. Most personalization systems have been evaluated in a non-interactive and simulated way. Usually, the evaluation study collected user search logs and usefulness judgments and compared the performance of the user's original queries with that of the expanded queries using their proposed personalization techniques/algorithms. Some studies, for example, Matthijs and Radlinski (2011), evaluated their proposed personalization system using both types of methods: (a) an offline evaluation using document judgments obtained from a small number of users for 72 queries to assess potential approaches; (b) an online evaluation with 41 participants using the system for two months to issue thousands of queries as part of their day to day web search activities.

The traditional criteria used for evaluation mainly include search effectiveness and time efficiency. Search effectiveness includes: (a) results ranking effectiveness (e.g., DCG, MRR), for example, Qiu and Cho (2006), Sun, Zeng, Liu, Lu, and Chen (2005), Speretta and Gauch (2005), Teevan et al. (2005a), and Dou et al. (2007); (b) results precision (e.g., MAP), for example, Chirita et al. (2006), Shen, Tan, and Zhai (2005), Shen and Zhai (2003), Lv et al. (2006), Wang et al. (2013), and White et al. (2013); and (c) accuracy of prediction, for example, Qiu and Cho (2006), Sun et al. (2005), and Liu, Yu, and Meng (2004). Studies using time efficiency as the evaluation criterion include Ma et al. (2007) and Lv et al. (2006). More evaluation measures have been suggested, including perceived relevance (Saracevic, 1996), learning (Eickhoff, Gwizdka, Hauff, & He, 2017; Hansen & Rieh, 2016), affect (Lopatovska & Arapakis, 2011), and using neuro-physiologic data as a source of metrics (Gwizdka, 2018).

Carman et al. (2008) evaluated personalization system performance using a social bookmarking website, del.icio.us, as a substitute for query logs and click-through data that are not always readily available because of issues such as privacy. The evaluation study showed that social tag data can be used to approximate user queries to search engines and that there is enough information in users' bookmark history to successfully personalize search results.

From a different perspective, evaluation of system functions using bucket testing may be disrupted by personalization. That is, the testing can lead to incorrect inference when the system's personalization function is on, because the contents provided to the test and control buckets of users are different. Das and Ranganath (2013) used the Yahoo personalized and nonpersonalized modules to test the impact of personalization on system evaluation using bucket testing. They developed a method that considered user interest in addition to click-through. This approach tended to provide a more accurate interpretation of the bucket testing results in presence of personalization. The method was demonstrated to be effective by their experiment using Yahoo.

Hannak et al. (2013) addressed the issue of measuring web search personalization by asking what features about the users (or in the user profiles) can lead to system providing personalization and how effective personalization is. Following the idea of comparing search results for the same query in different conditions, the authors created various Google accounts with varying information on gender, location, time, search history, operating systems, and so on. The study found that on average, 11.7% of search results showed differences because of personalization, with higher probabilities for results toward the bottom. The highest effect of personalization on search was related to political issues, news, and local businesses. The study also found that although the level of personalization is significant, there were very few user properties that led to personalization: only being logged in to Google and the location (IP address) of the user's machine result in measurable personalization; all other attributes did not result in levels of personalization beyond the baseline noise level.

Recently, Jones, Belkin, Lawless, and Pasi (2018) organized a workshop on evaluation of PIR, aimed at seeking to establish such an agreed overall framework for both user-centered studies and laboratory-based algorithmic research for PIR, which can be applied to not only single queries, but also multiple queries in a session or multiple sessions. Continuous effort is needed to build an agreed framework for PIR evaluation.

Privacy

The privacy issue cannot be avoided in PIR because personalization relies on collecting user information, whether it is the users' explicit release of search interest or the systems' collecting of implicit search behaviors. The privacy issue has become more and more significant and salient, especially after it was brought to the public attention by the cases of leading IT companies such as AOL's search data leak (Hafner, 2006), Facebook's lukewarm commitment to privacy (Anderson, 2018), and Google Plus's shutting down after the discovery of a security bug (Wakabayashi, 2018).

The issue of privacy in PIR has attracted a fair amount of research attention. Studies have found that although participants see the usefulness of personalization for online advertising, it could negatively affect user experience, or brings worries and confusion (e.g., O'Donnell & Cramer, 2015; Rohrer & Boyd, 2004). Matic, Pielot, and Oliver (2017) found that more than half of the displayed advertisements elicited positive reactions, for example, surprise or enthusiasm, which provides evidence that personalization and perceived relevance of online ads may boost its acceptance by customers, with improved control and transparency on how the data is being used. Golbeck (2017) noted that although personalization is powerful in helping users finding information, its benefits needs to be balanced with users' concerns about privacy.

Some privacy protection approaches have been made on the server side and others on the client side. Xie, Zeng, and Ma (2002) conducted personalization on an edge server, which moves the representation tier closer to the user by caching the static output of a Web server to reduce the user's perceived latency. In their prototype system named Avatar, two parts of user models were built: static profile and dynamic profile, which learned through the user's pervious browsing pattern and access history.

De Grande and Zorzo (2006) designed a system that provides privacy to the user in both implicit and explicit information gathering for the purpose of personalization. The MASKS (Managing Anonymity while Sharing Knowledge to Servers) system, placed between the users and the Internet, introduces anonymity to user browsing without blocking information to the sites, by removing communication control information that identifies a user. Rykowski and Cellary (2004) proposed to apply user-defined software agents to build Virtual Web Services, being combinations of several virtual and real services, accessed by the user as it would be a single service. Virtual services were created for and by users, allowing rich personalization of service functionality, data format and presentation, adjusting to different hardware and software environments. Majumder and Shrivastava (2013) attempted to capture Online Search Platform (OSP)'s personalization for a user in a new data structure called the personalization vector, a weighted vector over a set of topics, and presented efficient algorithms to learn it. The approach treated OSPs as black-boxes and extracted the personalization vector by mining only the output, specifically, the personalized content and the content without any user information, and the differences in these types of content. This approach enables users to access their profile information while protecting their privacy. Sendhilkumar and Geetha (2008) designed the User Conceptual Index (UCI), a client-side data collection instrument, by tracking the user interactions from the browser. The system automatically constructed page ontologies and compared user profiles with their search queries using the ontologies to perform personalized search on the client side. Ahmad, Rahman, and Wang (2016) proposed a topic-based privacy protection solution on the client side, in which each user query is submitted with k additional cover queries that act as a proxy to disguise users" intent from a search engine.

Shen, Tan, and Zhai (2007) conducted a systematic examination of the privacy preservation issue in personalization. They analyzed four levels of privacy protection: pseudo identity, group identity, no identity, and no personal information, as well as three types of personalization software architecture: server-side, client-side, and client–server cooperative personalization. Client-side personalization was found to have advantages over the server-side approach in privacy preservation, as the first three levels could be easily achieved technology wise. With some challenges being solved, the fourth level, that is, no personal information, could also be realized in personalization system design. Kobsa, Knijnenburg, and Livshits (2014) conducted a study to evaluate users' perception of privacy for client-side personalization (CSP) services. They used four personalization providers, in three of which user data were sent to the systems and in another data remained on users' smartphone. The study obtained encouraging results that CSP is likely to raise perceived protection of privacy. The study also found that privacy concerns influenced the disclosure of demographic data only, whereas satisfaction has an effect merely on context data. Increasing the perceived protection for CSP yields noticeable improvements in user satisfaction and disclosure. This is promising because accordingly, more disclosure typically leads to better personalization.

Usability

One criticism of adaptive/personalization systems is their potential for violation of the usability principles of direct manipulation systems, that is, controllability, predictability, transparency, and unobtrusiveness. Bakalov et al. (2013) proposed an approach to controlling adaptive behavior in recommender systems by allowing users to get an overview of personalization effects, view the user profile that is used for personalization, and adjust the profile and personalization effects to their needs and preferences. A user study evaluating a biomedical literature system portal with seven participants showed that users favored the controllable personalization on the usefulness, usability, user satisfaction, transparency, and trustworthiness of personalized systems. Although Bakalov et al.'s (2013) approach was originally for recommender systems (that are not the focus of this review), the usability issue brought up in their article applies to personalization approaches discussed in the current article, and the user evaluation from their study has implications for personalization system design. On the other hand, this sheds lights on the currently reviewed area that more research is needed to study the usability issue, to evaluate the personalization interface, and to better understand searchers' preferences regarding the interface aspect in personalization systems.

The Gap Between "Characterizing the Influence" and "Predicting"

Some relations (RLs) in Figure 1 are two-directional. For example, the relation between contexts and behaviors (RLc) represents two research activities: (a) characterizing how contextual factors influence search behaviors and (b) predicting contextual factors from behaviors. For a long time, the Human

DOI: 10.1002/asi

Information Behavior (HIB) research community has conducted an extensive amount of research on the former aspect, that is, how contextual factors (e.g., tasks, knowledge, gender, educational background, etc.) influence information behaviors. However, much less has been done on the latter aspect, for example, predicting the searchers' task type, gender, or educational background from search behaviors, except a few studies on predicting searchers' knowledge levels from behaviors (e.g., Liu, Liu, Cole, Belkin, & Zhang, 2012; Arguello, 2014).

It should be noted that although these two research activities present two relations between the same two entities (that is, contexts and search behaviors), these two relations are not necessarily symmetrical. In other words, that one specific contextual factor significantly influences search behaviors does not necessarily mean that this contextual factor can be accurately predicted as a significant factor from search behaviors.

An example to illustrate this issue is Mitsui et al. (2018). The researchers explored the gap between (a) characterizing the relationship between search intention types and search behaviors, and (b) predicting search intention types from search behaviors. Their findings showed that although intention types could have significant effects on search behaviors, predicting intention types from search behaviors did not receive significant results. This gap between "charactering the influence of contexts on behaviors" and "predicting contexts from behaviors" indicates that in personalization, the research activity of "predicting" needs to be conducted for detecting the significant behavioral factors that can predict contexts; significant behavioral factors that can predicting contexts cannot be inferred from what behavioral factors are significant influenced by contexts.

Discussion and Conclusions

Following the frameworks illustrated in Table 1 and Figure 1, we have reviewed existing research and practice in the area of PIR. We also discussed issues in PIR including when to personalize, diversity, evaluation, privacy, usability, and the asymmetry of influence and prediction of personalization factors. Some discussion of these has been made during reviewing, and in this final section, we summarize these, to give an overall picture regarding what has been done, and what needs to be done.

Among the four relations depicted in Figure 1, RLb, especially the prediction of user interest from search behaviors, has gained much attention and effort. RLc has drawn much attention regarding how contexts influence search behaviors, but not the other way around, that is, predicting contexts from behaviors. Attention on RLd is growing, but there is still much left to do. In the rest of this section, some major points about the limitations of current PIR research as well as the future directions are discussed.

We mentioned the gap between "relationship" and "prediction." The rich body of literature has provided substantial evidence on how various factors affect information search behaviors and performance. The literature also demonstrated how personalization could be performed based on users' behaviors and contextual factors. However, existing research has spent more effort along the former line, that is, research that identified the effects of various contextual factors on user behaviors or performance, than on the latter, that is, those that apply personalization based on prediction. It seems to us that more research should be conducted to build on or make use of the findings of the former line, to develop better methods of prediction that tailor search for the specific users and their specific contexts.

This includes approaches to the continuous and extended consideration of various contextual factors. For example, user task has been found to play significant roles in searching, and so it would be beneficial to research how a system can predict task types (e.g., along various task facets in Li and Belkin (2008) scheme) from observed users' behaviors, and how the system can provide better search results based on the prediction of task types. TREC 2016 initiated a new track of Task Track,[3] with the goals of evaluating system's understanding of user tasks and evaluating the relevance of retrieved documents with respect to underlying tasks in queries. Once the IR system obtains the task information that users are searching for, the system would decide whether personalization for this type of tasks is needed, and if needed, what type(s) of personalization would be the most effective and efficient for the current tasks. Meanwhile, it should also be noted that given the gap between the relationship and the prediction, it is challenging to explore prediction algorithms that can produce high prediction performance.

Other than tasks, approaches to studying contextual factors in personalization can spread to other aspects that have received less than enough attention, such as the various types of user characteristics. Having been examined quite extensively regarding its role on search behaviors, knowledge can be better used in predicting document usefulness for the current users. So are the various cognitive characteristics. Others, such as language, reading levels, and so on, need more research attention both in their relationships with search behaviors and in predicting document usefulness from search behaviors.

In addition, although commercial search engines have been personalizing search according to time and location, the granularity can be more detailed, in-depth, and broad. Location can be extended to the sociological sense, such as being in office, at home, or other places. Likewise, time can also be used on a sociological sense, such as to classify time into one's work time, vacation time, or special event time, and so on. Of course, when considering from these aspects, the privacy issue needs to be considered because of the fact that some people may not like to be monitored about their status.

Regarding search behaviors, commercial search engines are effectively using some types of previous search

[3] http://www.cs.ucl.ac.uk/tasks-track-2016/

DOI: 10.1002/asi

behaviors, both individual and group, successfully. Although previous research has covered a wider, and to some extent, deeper scope, one big issue is that many behaviors previously studied cannot be used for real-time prediction in personalization. Specifically, a great number of the behavioral or performance variables examined in the previous studies were on a whole task-session level, such as time spent to complete the whole task, total number of queries, total number of pages viewed and saved, and effectiveness (recall, precision) or efficacy (number of saved documents out of all viewed) of the search. All of these variables cannot be obtained until the end of a session. Although these results can in general be used to predict task type a posteriori, it is not easy to make use of these findings into adaptive search. Lower level behavioral variables that can be captured and used in real-time are needed, for example, document dwell time, number of pages per query, and so on.

Still, more research can be conducted in personalizing search using a combination of multiple components illustrated in Figure 1. This can be the wide range of search behaviors, the various contextual factors, and the combination of search behaviors and contextual factors in different formats. The above all indicate multiple aspects of gaps open for future research.

The reviewed research in PIR has devoted most of its efforts to descriptive and predictive analysis so far. The descriptive analysis demonstrates how various contextual factors and individual characteristics could influence users' search behaviors and preferences. The predictive analysis focuses on generating predictive models of users' search preferences, task types and other contextual factors, users' individual characteristics, and so on. Future research should step further and extend the research to explore what types of assistance the systems can provide based on the prediction about users' preference, contextual factors, and so on, and how to provide the needed types of assistance. It would be ideal if the system can have predesigned various types of assistance and activate the appropriate ones based on the predicted need.

Besides the framework that we followed in reviewing the literature, we also reviewed various related issues regarding personalization, including when to personalize, evaluation, privacy, and usability issues. All of these areas are important to PIR and can be further enriched toward the increasing maturity of PIR. The ideal status of PIR would be that the system becomes the very "personal assistant" to the user, providing the desired information that she needs in her specific situation, with the appropriate amount of control that she wanted, whereas it is also the system that she trusts.

To conclude this review, although there have been many and various approaches to PIR, it is still in its developing phase toward the goal of providing desired information that meets the specific searcher's needs, in the person's specific context. We believe a way toward this goal is continuing effort spent on personalization in both research and practice in IR and related fields.

References

AbdulJaleel, N., Corrada-Emmanuel, A., Li, Q., Liu, X., Wade, C., & Allan, J. (2003). UMass at TREC 2003: HARD and QA. In Proceedings of TREC 2003.

Agichtein, E., Brill, E., & Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. In Proceedings of 29th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '06) (pp. 19–26).

Agichtein, E., Brill, E., Dumais, S., & Ragno, R. (2006). Learning user interaction models for predicting web search results preferences. In Proceedings of 29th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '06), pp. (3–10).

Agichtein, E., White, R. W., Dumais, S. T., & Bennett, P. N. (2012). Search, interrupted: Understanding and predicting search task continuation. In Proceedings of 35th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '12), pp. 315–324.

Agrawal, R., Rantzau, R., & Terzi, E. (2006). Context-sensitive ranking. In Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data (pp. 383–394).

Ahmad, W.U., Rahman, M.M., & Wang, H. (2016). Topic model based privacy protection in personalized Web search. Proceedings of 39th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '16), (pp. 1025–1028).

Ahmed, A., Teo, C.H., Vishwanathan, S.V.N., & Smola, A. (2012). Fair and balanced: Learning to present news stories. In Proceedings of the 5th International Conference on Web Search and Data Mining (WSDM '12), 333–342.

Ahmed, S.M.Z., McKnight, C., & Oppenheim, C. (2004). A study of users' performance and satisfaction with the web of science IR interface. Journal of Information Science, 30(5), 459–468.

Ahn, J., Brusilovsky, P., He, D., Grady, J., & Li, Q. (2008). Personalized web exploration with task models. In Proceedings of the 17th International Conference on World Wide Web (WWW '08) (pp. 1–10).

Aktas, M., Nacar, M., & Menczer, F. (2004). Personalizing PageRank based on domain profiles. In Proceedings of WebKDD 2004: KDD Workshop on Web Mining and Web Usage Analysis, 83-90.

Albanese, M., Picariello, A., Sansone, C., & Sansone, L. (2004). Web personalization based on static information and dynamic user behavior. In Proceedings of the 6th Annual ACM International Workshop on Web Information and Data Management (WIDM 04), 80-87.

Allen, B. (1991). Topic knowledge and online catalog search formulation. The Library Quarterly, 61(2), 188–213.

Allen, B. (1992). Cognitive differences in end user searching of a CD-ROM index. In Proceedings of 15th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '92) (pp. 298–309). New York, NY: ACM.

Allen, B. (1996). Information needs: A person-in-situation approach. In P. Vakkari, R. Savolainen, & B. Dervin (Eds.), Information Seeking in Context: Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different contexts (pp. 111–122). London: Taylor Graham.

Al-Maskari, A., & Sanderson, M. (2011). The effect of user characteristics on search effectiveness in information retrieval. Information Processing & Management, 47(5), 719–729.

Anderson, M. (2018). Facebook privacy scandal explained. CTV News. Retrieved October 11, 2018, from https://www.ctvnews.ca/sci-tech/facebook-privacy-scandal-explained-1.3874533.

Arguello, J. (2014). Predicting search task difficulty. In Proceedings of the 36th European Conference in Information Retrieval (ECIR'14) (pp. 1–12).

Backhausen, D. T. J. (2012). Adaptive IR for exploratory search support. In Proceedings of the 35th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 12) (p. 992).

Bai, J., Nie, J.-Y., Bouchard, H., & Cao, G. (2007). Using query contexts in information retrieval. In Proceedings of the 30th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '07) (pp. 15–22). Ámsterdam, The Netherlands.

DOI: 10.1002/asi

Bakalov, F., Meurs, M.-J., König-Ries, B., Sateli, B., Witte, R., Butler, G., & Tsang, A. (2013). An approach to controlling user models and personalization effects in recommender systems. In Proceedings of the 18th International Conference on Intelligent User Interfaces (IUI '13) (pp. 49–56).

Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, 5, 133–143.

Belkin, N.J. (1993). Interaction with texts: Information retrieval as information-seeking behavior. In: Information retrieval '93. Von der Modellierung zur Anwendung (pp. 55–66). Konstanz: Universitaetsverlag Konstanz.

Belkin, N.J. (2008). Some(what) grand challenges for information retrieval. Keynote lecture presented at the European conference on information retrieval, Glasgow, Scotland, 31 march 2008. ACM SIGIR Forum, 42(1), 47–54.

Belkin, N.J., Chaleva, I., Cole, M., Li, Y.-L., Liu, Y.-H., Muresan, G., Smith, C.L., Sun, Y., Yuan, X.-J., & Zhang, X.-M. (2004). Rutgers' HARD track experiences at TREC 2004. In Proceedings of TREC 2004.

Belkin, N.J., Cole, M., Gwizdka, J., Li, Y.-L., Liu, J.-J., Muresan, G., Roussinov, D., Smith, C.A., Taylor, A., & Yuan, X.-J. (2005). Rutgers Information interaction lab at TREC 2005: Trying HARD. In Proceedings of TREC 2005. Gaithersburg, MD: NIST.

Belkin, N.J., Cool, C., Koenemann, J., Ng, K.B., & Park, S. (1996). Using relevance feedback and ranking in interactive searching. In Proceedings of the Fourth Text Retrieval Conference (pp. 181–210). Gaithersburg, MD: NIST.

Belkin, N.J., Kelly, D., Lee, H.-J., Li, Y.-L., Muresan, G., Tang, M.-C., Yuan, X.-J., & Zhang, X.-M. (2003). Rutgers' HARD and web interactive track experiences at TREC 2003. In Proceedings of TREC 2003.

Bennet, P.N., Radlinski, F., White, R. W., & Yilmaz, E. (2011). Inferring and using location metadata to personalize web search. In Proceedings of 34th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '11) (pp. 135–144).

Bennett, P., White, R. W., Chu, W., Dumais, S.T., Bailey, P., Borisyuk, F., & Cui, X. (2012). Modeling the impact of short- and long-term behavior on search personalization. In Proceedings of 35th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '12) (pp. 185–194).

Bi, B., Shokouhi, M., Kosinski, M., & Graepel, T. (2013). Inferring the demographics of search users:social data meets search queries. In Proceedings of the 22nd International Conference on World Wide Web (WWW '13) (pp. 131–140).

Biancalana, C., Micarelli, A., & Squarcella, C. (2008). Nereau: A social approach to query expansion. In Proceedings of the 10th ACM Workshop on Web Information and Data Management (pp. 95–102).

Bilenko, M., White, R. W., Richardson, M., & Murray, G.C. (2008). Talking the talk vs. Walking the walk: Salience of information needs in querying vs. browsing. Proceedings of 31th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '08) (pp. 705–706).

Bouadjenek, M.R., Hacid, H., Bouzeghoub, M. (2013). LAICOS: An open source platform for personalized social web search. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13) (pp. 1446–1449).

Brennan, K., Kelly, D., and Arguello, J. (2014). The effect of cognitive abilities on information search for tasks of varying levels of complexity. In Proceedings of the Symposium for Information Interaction in Context (IIiX '14) (pp. 165–174). New York, NY: ACM.

Budzik, J., & Hammond, J.K. (1999). Watson: Anticipating and contextualizing information needs. In Proceedings of the 62nd Annual Meeting of the American Society for Information Science.

Buscher, G., Dengel, A., Biedert, R., & van Elst, L. (2012). Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. ACM Transactions on Interactive Intelligent Systems (TiiS), 1 (2), 9.

Buscher, G., van Elst, L., & Dengel, A. (2009). Segment-level display time as implicit feedback: A comparison to eye tracking. In Proceedings of 32th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '09) (pp. 67–74).

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. Information Processing and Management, 31(2), 191–213.

Cai, F. & de Rijke, M. (2016). Selectively personalizing query auto-completion. In Proceedings of 39th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '16) (pp. 993–996).

Campbell, D.J. (1988). Task complexity: A review and analysis. Academy of Management Review, 13(1), 40–52.

Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 335–336).

Carman, M. J., Baillie, M., & Crestani, F. (2008). Tag data and personalized information retrieval. In Proceedings of the 2008 ACM Workshop on Search in Social Media (SSM 08) (pp. 27–34).

Carterette, B. (2018). Offline evaluation of personalization using logged data. In Proceedings of CHIIR 2018 Workshop on Evaluation of Personalization in Information Retrieval. Retrieved from https://wepir.adaptcentre.ie/papers/WEPIR_2018_Keynote_abstract.pdf.

Chen, C., Yang, M., Li, S., Zhao, T., & Qi, H. (2010). Predicting query potential for personalization, classification or regression? In Proceedings of 33rd Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '10) (pp. 725–726).

Chen, L., & Sycara, K. (1998). Webmate: A personal agent for browsing and searching. In Proceedings of the 2nd International Conference on Autonomous Agents and Multiagent Systems (pp. 132–139).

Chen, S.Y., & Macredie, R. (2010). Web-based interaction: A review of three important human factors. International Journal of Information Management, 30(5), 379–387.

Chen, S.Y., & Ford, N.J. (1998). Modeling user navigation behaviours in a hypermedia-based learning system: An individual differences approach. Knowledge Organization, 25(3), 67–78.

Chen, S.Y., Magoulas, G.D., & Dimakopoulos, D. (2005). A flexible interface design for web directories to accommodate different cognitive styles. Journal of the American Society for Information Science and Technology, 56(1), 70–83.

Chen, W., Cai, F., Chen, H., & de Rijke, M. (2017). Personalized query suggestion diversification. In Proceedings of the 40th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '17) (pp. 817–820).

Chirita, P.-A., Firan, C.S., & Nejdl, W. (2006). Summarizing local context to personalize global web search. In Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM 2006) (pp. 287–296).

Chirita, P.-A., Firan, C.S., & Nejdl, W. (2007). Personalized query expansion for the web. In Proceedings of the 30th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '07) (pp. 7–14). Ámsterdam, The Netherlands.

Chirita, P.A., Nejdl, W., Paiu, R., & Kohlschutter, C. (2005). Using ODP metadata to personalize search. In Proceedings of 28th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '05), August 2005, Salvador, Brazil.

Chirita, P.-A., Olmedilla, D., & Nejdl, W. (2004). PROS: A personalized ranking platform for web search. In Proceedings of the International Conference on Adaptive Hypermedia and Web-Based Systems, 2004.

Claypool, M., Le, P., Wased, M., & Brown, D. (2001). Implicit interest indicators. In Proceedings of the 6th International Conference on Intelligent User Interfaces (IUI 2001) (pp. 33–40).

Cole, M., Gwizdka, J., Bierig, R., Belkin, N.J., Liu, J., Liu, C., Zhang, J., Zhang X. (2010, August). Linking search tasks with low-level eye movement patterns. Presented at Proceedings of the 28th European Conference on Cognitive Ergonomics (ECCE 2010), Delft, The Nethelands.

Cool, C. (2001). The concept of situation in information science. In M. E. Williams (Ed.), Annual review of information science & technology (Vol. 35, pp. 5–42). Medford, NJ: Information Today, Inc.

Croft, W.B., Cronen-Townsend, S., & Lavrenko, V. (2001). Relevance feedback and personalization: A language model perspective. In DELOS

Workshop: Personalization and Recommender Systems in Digital Libraries (pp. 49–54).
Daoud, M., Tamine-Lechani, L., & Boughanem, M. (2008). Learning user interests for a session-based personalized search. In Proceedings of the Symposium for Information Interaction in Context (IIiX '08) (pp. 57–64).
Daoud, M., Tamine-Lechani, L., Chebaro, B., & Boughanem, M. (2009). A session based personalized search using an ontological user profile. In Proceedings of the 2009 ACM Symposium on Applied Computing (SAC '09) (pp. 1732–1736).
Das, A. & Ranganath, H. (2013). When web personalization misleads bucket testing. In Proceedings of the 1st Workshop on User Engagement Optimization (UEO '13) (pp. 17–20).
De Grande, R.E. & Zorzo, S.D. (2006). Privacy protection without impairing personalization by using the extended system MASKS and the extended contextualized P3P privacy policies. In Proceedings of the 12th Brazilian Symposium on Multimedia and the Web (WebMedia'06) (pp. 89–98).
Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In J.D. Glazier & R.R. Powell (Eds.), Qualitative research in information management (pp. 61–84). Englewood Cliffs, CO: Libraries Unlimited.
Dervin, B. (2003). Given a context by any other name: Methodological tools for taming the unruly beast. In B. Dervin, L. Foreman-Wernet, & E. Lauterbach (Eds.), Sense-making methodology reader: Selected writings of Brenda Dervin (pp. 111–132). Gresskill, NJ: Hampton Press.
Diamond, A. (2013). Executive functions. Annual Review of Psychology, 64, 135–168.
Dou, Z., Song, R., & Wen, J.-R. (2007). A large-scale evaluation and analysis of personalized search strategies. In Proceedings of the 16th International Conference on World Wide Web (WWW 2007), (pp. 581–590).
Dufresne, A., & Turcotte, S. (1997). Cognitive style and its implications for navigation strategies. In B. Du Boulay & R. Mizoguchi (Eds.), Artificial intelligence in education knowledge and media learning system (pp. 287–293). Kobe, Japan: Amsterdam IOS Press.
Duggan, G. B. & Payne, S. J. (2008). Knowledge in the head and on the web: Using topic expertise to aid search. In Proceedings of the International Conference on Human Factors in Computing Systems (CHI 08) (pp. 39–48).
Dumais, S.T. (2007). Information retrieval in context. In Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07) (p. 2).
Dumais, S.T. (2012). Putting context into search and search into context. In Proceedings of 35th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '12) (p. 1021).
Dumais, S.T. (2016, October). Personalized search: Potential and pitfalls. In Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM'16), Indianapolis, Indiana.
Eickhoff, C., Collins-Thompson, K., Bennett, P.N., & Dumais, S. (2013). Personalizing atypical web search sessions. In Proceedings of the 6th International Conference on Web Search and Data Mining (WSDM '13) (pp. 285–294).
Eickhoff, C., Gwizdka, J., Hauff, C., & He, J. (2017). Introduction to the special issue on search as learning. Information Retrieval Journal, 20 (5), 399–402.
Ekstrom, R.B., French, J.W., Harman, H.H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. Princeton, NJ: ETS.
Fan, W., Gordon, M. D., & Pathak, P. (2000). Personalization of search engine services for effective retrieval and knowledge management. In ICIS '00 Proceedings of the twenty first international conference on Information systems (pp. 20–34).
Ferragina, P., & Gulli, A. (2005). A personalized search engine based on web-snippet hierarchical clustering. In Proceedings of the 14th International Conference on World Wide Web (WWW 2005) (pp. 801–810).
Ford, N., Miller, D., & Moss, N. (2001). The role of individual differences in Internet searching: an empirical study. Journal of the American Society for Information Science and Technology, 52(12), 1049–1066.
Freund (2008). Exploring task-document relations in support of information retrieval in the workplace (Unpublished Dissertation). University of Toronto.
Frias-Martinez, E., Chen, S.Y., & Liu, X. (2007). Automatic cognitive style identification of digital library users for personalization. Journal of the American Society for Information Science and Technology, 58(2), 237–251.
Frias-Martinez, E., Magoulas, G., Chen, S., & Macredie, R. (2006). Automated user modeling for personalized digital libraries. International Journal of Information Management, 26(3), 234–248.
Gao, B. J. & Jan, J. (2010). Rants: A framework for rank editing and sharing in web search. In Proceedings of the 19th International Conference on World Wide Web (WWW 2010) (pp. 1245–1248).
Gauch, S., Chaffee, J., & Pretschner, A. (2003). Ontology-based personalized search and browsing. Web Intelligence and Agent Systems: An international Journal, 1(3–4), 219–234.
Ghorab, M.R., Zhou, D., O'Connor, A., & Wade, V. (2013). Personalised information retrieval: survey and classification. User Modeling and User-adapted Interaction, 23(4), 381–443. Retrieved from https://link.springer.com/content/pdf/10.1007%2Fs11257-012-9124-1.pdf
Goker, A., Myrhaung, H., & Bierig, R. (2009). Context and information retrieval. In A. Goker & J. Davies (Eds.), Information retrieval: Searching in the 21s century (pp. 131–157). Chichester, U.K.: Wiley.
Golbeck, J. (2017). I'll be watching you: Policing the line between personalization and privacy. In Proceedings of User Modeling, Adaptation, and Personalization (UMAP '17) (p. 2).
Guo, Q. & Agichtein, E (2009). Beyond session segmentation: Predicting changes in search intent with client-side user interactions. In Proceedings of 33rd Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 09) (pp. 636–637).
Gwizdka, J. (2008). Revisiting search task difficulty: Behavioral and individual difference measures. In Proceedings of the American Society for Information Science & Technology, Columbus, OH.
Gwizdka, J. (2010, August). Using stroop task to assess cognitive load. Presented at the Proceedings of the 28th Annual European Conference on Cognitive Ergonomics (ECCE), Delfts, The Nerthelands.
Gwizdka, J. (2017). I can and so I search more: Effects of memory span in search behavior. In Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (CHIIR '17) (pp. 341–344). New York, NY: ACM.
Gwizdka, J. (2018). Neuro-physiological data as a source of evaluation metrics for personalized IR. In Proceedings of CHIIR 2018 Workshop on Evaluation of Personalization in Information Retrieval. Retrieved from https://wepir.adaptcentre.ie/papers/WEPIR_2018_paper_1.pdf.
Gwizdka, J., & Spence, I. (2006). What can searching behavior tell us about the difficulty of information tasks? A study of Web navigation. In Proceedings of the Annual Conference of The American Society for Information Science & Technology 2006.
Hafner, K. (2006). Researchers yearn to use AOL logs, but they hesitate. New York Times, Retrieved October 11, 2018, from https://www.nytimes.com/2006/08/23/technology/23search.html.
Hannak, A., Krishnamurthy, B., Sapie˙zy´nski, P., Lazer, D., Wilson, C., Kakhki, A. M., & Mislove, A. (2013). Measuring Personalization of Web Search. In Proceedings of the 22nd International Conference on World Wide Web (WWW '13) (pp. 527–536).
Hansen, P., & Rieh, S.Y. (2016). Editorial: Recent advances on searching as learning: An introduction to the special issue. Journal of Information Science, 42(1), 3–6.
Heinström, J. (2003). Five personality dimensions and their influence on information behavior. Information Research, 9(1). http://informationr.net/ir/9-1/paper165.html
Heinström, J. (2005). Fast surfing, broad scanning and deep diving - the influence of personality and study approach on students' information-seeking behavior. Journal of Documentation, 61(2), 228–247.
Heinström, J. (2006). Broad exploration or precise specificity: Two basic information seeking patterns among students. Journal of the American Society for Information Science and Technology, 57(11), 1440–1450. https://doi.org/10.1002/asi.20432
Hembrooke, H.A., Granka, L.A., Gay, G.K., & Liddy, E.D. (2005). The effects of expertise and feedback on search term selection and subsequent learning. Journal of the American Society for Information Science and Technology, 56(8), 861–871.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science, 44(3), 161–174.

Hupfer, M.E., & Detlor, B. (2006). Gender and web information seeking: A self concept orientation model. Journal of the American Society for Information Science and Technology, 57(8), 1105–1115.

Huurdeman, H.C., Wilson, M.L., & Kamps, J. (2016). Active and passive utility of search interface features in different information seeking task stages. In Proceedings of ACM Conference on Human Information Interaction and Retrieval (CHIIR 2016) (pp. 3–12).

Ingwersen, P., & Javerlin, K. (2005). Information retrieval in context: IRix. ACM SIGIR Forum, 39, 31–39.

Jayarathna, S., Patra, A., & Shipman, F. (2013). Mining user interest from search tasks and annotations. In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM '13) (pp. 1849–1851).

Jiang, D., Leung, K.W.-T., & Ng, W. (2011). Context-aware search personalization with concept preference. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11) (pp. 563–572).

Joachims, T. (2002). Optimizing search engines using clickthrough data. In Proceedings of the ACM Conference on Knowledge Discovery and Datamining (SIGKDD), 2002 (pp. 133–142).

Joachims, T., Granka, L., Pang, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In Proceedings of 28th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 2005) (pp. 154–161).

Jones, G.J.F., Belkin, N.J., Lawless, S., & Pasi, G. (2018). WEPIR 2018: Workshop on evaluation of personalisation in information retrieval. In Proceedings of ACM Conference on Human Information Interaction and Retrieval (CHIIR '18) (pp. 386–388).

Jones, R., Hassan, A., and Diaz, F. (2008). Geographic features in Web search retrieval. In Proceedings of ACM GIR Workshop on Geographic Information Retrieval (pp. 57–58).

Kang, R., Fu, W-T. & Kannampallil, T. G. (2010). Exploiting knowledge-in-the-head and knowledge-in-the-social-web: Effects of domain expertise on exploratory search in individual and social search environments. Presented at the Proceedings of the ACM Conference on Computer-Human Interaction (CHI 2010), Atlanta, GA.

Kashyap, A., Amini, R., & Hristidis, V. (2012). SonetRank: Leveraging social networks to personalize search. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM '12) (pp. 2045–2049).

Ke, Y., Deng, L., Ng, W., & Lee, D.L. (2005). Web dynamics and their ramifications for the development of web search engines. Computer Network, 50, 1430–1447.

Keenoy, K., & Levene, M. (2005). Personalization of web search. Lecture Notes in Computer Science, 3169, 201–228.

Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing web-based information-seeking tasks. Journal of the American Society for Information Science and Technology, 58(7), 999–1018.

Kelly, D. (2006a). Measuring online information-seeking context, part 1. Background and method. Journal of the American Society for Information Science and Technology, 57(13), 1729–1739.

Kelly, D. (2006b). Measuring online information-seeking context, part 2. Findings and discussion. Journal of the American Society for Information Science and Technology, 57(14), 1862–1874.

Kelly, D. (2015). When effort exceeds expectations: A theory of search task difficulty. In ECIR Supporting Complex Search Task Workshop '15.

Kelly, D., Arguello, J., Edwards, A., & Wu, W. (2015). Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In Proceedings of the 2015 International Conference on The Theory of Information Retrieval (ICTIR '15) (pp. 101–110).

Kelly, D., & Belkin, N. (2002). A user modeling system for personalized interaction and tailored retrieval in interactive IR. In Proceedings of the Annual Meeting for the American Society for Information Science & Technology (ASIST '02) (pp. 316–325).

Kelly, D., & Belkin, N.J. (2001). Reading time, scrolling and interaction: Exploring implicit sources of user preference for relevance feedback. In D. H. Kraft, W.B. Croft, D.J. Harper, & J. Zobel (Eds.), Proceedings of the 24th International Conference on Research and Development in Information Retrieval (ACM SIGIR '01) (pp. 408–409). New York, NY: ACM.

Kelly, D., & Belkin, N.J. (2004). Display time as implicit feedback: Understanding task effects. In Proceedings of the 27th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '04), Sheffield, UK (pp. 377–384).

Kelly, D., & Cool, C. (2002). The effects of topic familiarity on information search behavior. In Proceedings of the 2nd ACM-IEEE Joint Conference on Digital Libraries (JCDL '02) (pp. 74–75).

Kelly, D., Diaz, F., Belkin, N.J., & Allan, J. (2004). A user-centered approach to evaluating topic models. In Proceedings of the 26th European Conference on Information Retrieval (ECIR '04) (pp. 27–41).

Kelly, D., Dollu, V.D., & Fu, X. (2005). The loquacious user: A document-independent source of terms for query expansion. In Proceedings of 28th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '05) (pp. 457–464).

Kelly, D., Murdock, V., Yuan, X., Croft, W.B., & Belkin, N.J. (2002). Features of documents relevant to task- and fact- oriented questions. In Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM '02) (pp. 645–647).

Kelly, D., & Teevan, J. (2003). Implicit feedback for inferring user preference: A bibliography. SIGIR Forum, 37(2), 18–28.

Kelly, G.A. (1963). A theory of personality: The psychology of personal constructs. New York: Norton.

Khopkar, Y., Spink, A., Giles, C.L., Shah, Pl, & Debnath, S. (2003). Search engine personalization: An exploratory study. First Monday, 8(7). Retrieved from http://www.firstmonday.org/issues/issue8_7/khopkar/.

Kim, D.Y., Lehto, X.Y., & Morrison, A.M. (2007). Gender differences in online travel information search: Implications for marketing communications on the internet. Tourism Management, 28(2), 423–433.

Kim, J. (2006). Task difficulty as a predictor and indicator of web searching interaction. In Proceedings of the International Conference on Human Factors in Computing Systems (CHI '06) (pp. 959–964).

Kobsa, A., Knijnenburg, B.P., & Livshits, B. (2014). Let's do it at my place instead? Attitudinal and behavioral study of privacy in client-side personalization. In Proceedings of the International Conference on Human Factors in Computing Systems (CHI '14) (pp/ 81–90).

Koohang, A., & Durante, A. (2003). Learners' perceptions toward the webbased distance learning activities/assignments portion of an undergraduate hybrid instruction model. Journal of Information Technology Education, 2, 105–113.

Kraft, R., Maghoul, F., & Chang C.C. (2005). Y!Q: Contextual search at the point of inspiration. In Proceedings of the 14th International Conference on Information and Knowledge Management (CIKM '05) (pp. 816–823).

Kuhlthau, C.C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42, 361–371.

Kumaran, G, Jones, R., & Madani, O. (2005). Biasing web search results for topic familiarity. In Proceedings of the 14th International Conference on Information and Knowledge Management (CIKM '05) (pp. 271–271).

Large, A., Beheshti, J., & Rahman, T. (2002). Gender differences in collaborative web searching behavior: An elementary school study. Information Processing & Management, 38(3), 427–443.

Lazander, A.W., Biemans, H.J.A., & Wopereis, I.G.J.H. (2000). Differences between novice and experienced users in searching information on the world wide web. Journal of the American Society for Information Science, 51(6), 576–581.

Lee, U., Liu, Z., & Cho, J. (2005). Automatic identification of user goals in web search. Proceedings of the 14th International Conference on World Wide Web (WWW '05), 391–400.

Leung, K. W.-T., Fung, H. Y., & Lee, D. L (2011). Constructing concept relation network and its application to personalized web search. In Proceedings of the 14th International Conference on Extending Database Technology (EDBT '11) (pp. 413–424).

DOI: 10.1002/asi

Leung, K.W.-T., Lee, D.L., Ng, W., & Fung, H.Y. (2012). A framework for personalizing web search with concept-based user profile. ACM Transaction on Internet Technology, 11(4), 17.

Li, D., Yang, M., Qi, H., Li, S., & Zhao, T. (2010). Reexamination on Potential for Personalization in Web Search. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters (pp. 701–709). Association for Computational Linguistics.

Li, Y., & Belkin, N.J. (2008). A faceted approach to conceptualizing tasks in information seeking. Information Processing & Management, 44, 1822–1837.

Li, Y., & Belkin, N.J. (2010). An exploration of the relationships between work task and interactive information search behavior. Journal of the American Society for Information Science and Technology, 61(9), 1771–1789.

Li, Y., Zhang, X., Liu, J., & Zhang, Y. (2006). Trained vs. untrained searchers' interaction with search features in digital libraries: A case study. In Proceedings of the Annual Meeting for the American Society for Information Science & Technology (ASIST 2006).

Liu, C., Belkin, N. J. & Cole, M. (2012). Personalization of search results using interaction behaviors in search sessions. In Proceedings of 35th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '12) (pp. 205–214).

Liu, C., Cole, M., Baik, E. & Belkin, N.J. (2012) Rutgers at the TREC 2012 Session Track. In: Proceedings of TREC 2012, the 21st Text Retrieval Conference. Gaithersburg, MD: NIST. Retrieved on 14 December 2018 from http://trec.nist.gov/pubs/trec21/papers/ruiiltrec.session.final.pdf

Liu, C., Liu, J., & Yan, Z. (2018). Personalizing information retrieval using search behaviors and time constraints. In Proceedings of ACM Conference on Human Information Interaction and Retrieval (CHIIR '18) (pp. 261–264).

Liu, F., Yu, C., & Meng, W. (2002). Personalized web search by mapping user queries to categories. In Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM '02), pp. (558–565).

Liu, F., Yu, C., & Meng, W. (2004). Personalized web search for improving retrieval effectiveness. IEEE Transactions on Knowledge and Data Engineering, 16(1), 28–40.

Liu, H. & Hoeber, O. (2011). A Luhn-inspired vector re-weighting approach for improving personalized web search. In 2011 IEEE/WIC/-ACM International Conferences on Web Intelligence and Intelligent Agent Technology (pp. 301–305).

Liu, J. & Belkin, N. J. (2010). Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. In Proceedings of 33rd Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '10) (pp. 26–33).

Liu, J., & Belkin, N.J. (2015). Personalizing information retrieval for multi-session tasks: Examining the roles of task stage, task type, and topic knowledge on the interpretation of dwell time as an indicator of document usefulness. Journal of the Association for Information Science and Technology, 66(1), 58–81.

Liu, J., Cole, M., Liu, C., Belkin, N.J., Zhang, J., Bierig, R., … Zhang, X. (2010). Search behaviors in different task types. In C. Lagoze, L. Giles, & Y.-F. Li (Eds.), Proceedings of the 10th ACM/IEEE-CS Joint Conference on Digital Libraries (ACM JCDL '10) (pp. 69–78). ACM: New York, NY.

Liu, J., Gwizdka, J., Liu, C. & Belkin, N.J. (2010) Predicting task difficulty for different task types. Presented at ASIS&T '10: Proceedings of the 73rd ASIS&T Annual Meeting. Retrieved from http://onlinelibrary.wiley.com.proxy.libraries.rutgers.edu/doi/10.1002/meet.14504701214/pdf

Liu, J., Liu, C., Cole, M., Belkin, N.J., & Zhang, X. (2012). Examining and predicting search task difficulty. In Proceedings of the 21st International Conference on Information and Knowledge Management (CIKM '12) (pp. 1313–1322). New York, NY: ACM.

Liu, Z., & Huang, X. (2008). Gender differences in the online reading environment. Journal of Documentation, 64(4), 616–626.

Lopatovska, I., & Arapakis, I. (2011). Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction. Information Processing & Management, 47(4), 575–592.

Lopes, C.T. (2013). Context-based health information retrieval. ACM SIGIR Forum, 47(2), 141–142.

Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., & Gay, G. (2006). The influence of task and gender on search and evaluation behavior using Google. Information Processing & Management, 42(4), 1123–1131.

Luxenburger, J., Elbassuoni, S., & Weikum, G. (2008). Matching task profiles and user needs in personalized web search. In Proceedings of the 1st International Conference on Web Search and Data Mining (WSDM '08) (pp. 689–698).

Lv, Y., Sun. L., Zhang, J., Nie, J.-Y., Chen, W., & Zhang, W. (2006). An interactive implicit feedback approach to personalized search. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL (pp. 585–592).

Ma, Z., Pant, G., & Sheng, O.R.L. (2007). Interest-based personalized search. ACM Transactions on Information Systems, 25(1), 1–38.

MacFarlane, A., Al-Wabil, A., Marshall, C.R., Albrair, A., Jones, S.A., & Zaphiris, P. (2010). The effect of dyslexia on information retrieval: A pilot study. Journal of Documentation, 66(3), 307–326.

Majumder, A. & Shrivastava, N. (2013). Know your personalization: learning topic level personalization in online services. In Proceedings of the 22nd International Conference on World Wide Web (WWW '13) (pp. 873–882).

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. Journal of the American Society for Information Science, 40(1), 54–66.

Matic, A., Pielot, M., & Oliver, N. (2017). "OMG! How did it know that?" Reactions to highly-personalized ads. In Proceedings of User Modeling, Adaptation, and Personalization (UMAP '17) (pp. 41–46).

Matthijs, N. & Radlinski, F. (2011). Personalizing Web Search using Long Term Browsing History. Proceedings of the 4th International Conference on Web Search and Data Mining (WSDM '11) (pp. 25–34).

Mei, Q. & Church, K. (2008). Entropy of search logs: How hard is search? With personalization? With backoff? In Proceedings of the 1st international Conference on Web Search and Data Mining (WSDM '08) (pp. 45–540).

Mislove, A., Viswanath, B., Gummadi, K. P., & Druschel, P. (2010). You are who you know: inferring user profiles in online social networks. In Proceedings of the 3rd International Conference on Web Search and Data Mining (WSDM 2010) (pp. 251–260). New York, NY: ACM.

Mitsui, M., Liu, J., & Shah, C. (2018, March). The paradox of personalization: Does task prediction require individualized models? Presented at CHIIR '18: 2018 Conference on Human Information Interaction Retrieval, New Brunswick, NJ.

Morita, M., & Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. In W.B. Croft & C.J. van Rijsbergen (Eds.), Proceedings of the 17th International Conference on Research and Development in Information Retrieval (ACM SIGIR '94) (pp. 272–281). New York, NY: ACM.

Murdock, V., Kelly, D., Croft, W.B., Belkin, N.J., & Yuan, X. (2007). Identifying and improving retrieval for procedural questions. Information Processing and Management, 43, 181–203.

Ng, W., Deng, L., & Lee, D.L. (2007). Mining user preference using spy voting for search engine personalization. ACM Transactions on Internet Technology, 7(4), 19.

O'Brien, P., Luo, X., Abou-Assaleh, T., Gao, W., & Li, S. (2009). Personalization of content ranking in the context of local search. In 2009 IEEE/-WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (Vol. 1, pp. 532–539). IEEE Computer Society.

Oddy, R.N. (1977). Information retrieval through man-machine dialogue. Journal of Documentation, 33(1), 1–14.

O'Donnell, K., & Cramer, H. (2015). People's perceptions of personalized ads. In Proceedings of the 24th International World Wide Web Conference (WWW '15) (pp. 1293–1298).

Otterbacher, J. (2010). Inferring gender of movie reviewers: exploiting writing style, content and metadata. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM 2010), Toronto, Ontario, Canada.

Palmquist, R.A., & Kim, K.S. (2000). Cognitive style and on-line database search experience as predictors of web search performance.

DOI: 10.1002/asi

Journal of the American Society for Information Science, 51(6), 558–566.
Pariser, E. (2011). The filter bubble: What the internet is hiding from you. New York: Penguin Press.
Pasi, G. (2010). Issues in personalizing information retrieval. IEEE Intelligent Informatics Bulletin, 11(1), 3–7.
Patel, B., Tendulkar, A. V., & Chakraborti, S. (2013). Towards selective user specific query expansion. In Proceedings of the 1st Workshop on User Engagement Optimization (UEO '13) (pp. 21–24).
Pitkow, J., Schütze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., … Breuel, T. (2002). Personalized search: A contextual computing approach may prove a breakthrough in personalized search efficiency. Communications of the ACM, 45(9), 50–55.
Pretschner, A., & Gauch, S. (1999). Ontology based personalized search. In Proceedings of ICTAI '99 (pp. 391–398).
Qiu, F., & Cho, J. (2006). Automatic identification of user interest for personalized search. In Proceedings of the 15th International Conference on World Wide Web (WWW 2006) (pp. 727–736).
Qiu, L. (1993). Analytical searching vs. browsing in hypertext information retrieval systems. Canadian Journal of Information and Library Science, 18(4), 1–13.
Radlinski, F., & Dumais, S. (2006). Improving personalized web search using result diversification. In Proceedings of 29th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '06) (pp. 691–692).
Radlinski, F., & Joachims, T. (2005). Query chains: Learning to rank from implicit feedback. In Proceedings of the 11th Conference on Knowledge Discovery and Data Mining (KDD), 239–248.
Riding, R.J., & Cheema, I. (1991). Cognitive styles—an overview and integration. Educational Psychology, 11(3/4), 193–215.
Rocchio, J.J. (1971). Relevance feedback in information retrieval. In The SMART Retrieval System: Experiments in Automatic Document Processing (pp. 313–323). Prentice-Hall Inc.
Rode, H., Ramirez, G., Westerveld, T., Hiemstra, D., & de Vries, A.P. (2005). The lowlands' TREC experiments 2005. In Proceedings of TREC 2005.
Rohrer, C., & Boyd, J. (2004). The rise of intrusive online advertising and the response of user experience research at Yahoo! In Proceedings of CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04) (pp. 1085–1086). ACM.
Roy, M., & Chi, M.T.H. (2003). Gender differences in patterns of searching the web. Journal of Educational Computing Research, 29(3), 335–348.
Rykowski, J., & Cellary, W. (2004). Virtual web services: Application of software agents to personalization of web services. In Proceedings of the 6th international conference on Electronic commerce (ICEC '04), 25–27 October, 2004, Delft, The Nerthelands (pp. 409–418).
Salton, G. (Ed.). (1971). The SMART retrieval system: Experiments in automatic document processing. Englewood Cliffs, NJ: Prentice-Hall.
Salton, G., & Buckley, C. (1990). Improving retrieval performance by retrieval feedback. Journal of the American Society for Information Science, 41(4), 288–297.
Saracevic, T. (1996). Relevance reconsidered 1996. In Proceedings of the Second Conference on Conceptions of Library and Information Science (CoLIS 2) (pp. 201–218).
Schmidt, T. & Wolff, C. (2016). Personality and information behavior in Web search. In Proceedings of the Annual Meeting for the American Society for Information Science & Technology (ASIST 2016).
Schumacher, P., & Morahan-Martin, J. (2001). Gender, Internet and computer attitudes and experiences. Computers and Human Behavior, 17, 95–110.
Sendhilkumar, S. & Geetha, T.V. (2008). Personalized ontology for Web search personalization. In COMPUTE '08 Proceedings of the 1st Bangalore Annual Compute Conference 2008.
Serdyukov, P., Dupret, G., & Craswell, N. (2014). Log-based personalization: the 4th web search click data (WSCD) workshop. In Proceedings of the 7th ACM international conference on Web search and data mining. WSDM'14 (pp. 685–686).
Shen, S., Chen, W., Hu, B., & Yang Q. (2012). Personalized click model through collaborative filtering. In Proceedings of the 5th International Conference on Web Search and Data Mining (WSDM '12), 323–332.
Shen, X., Tan, B., & Zhai, C. (2005a). Context-sensitive information retrieval using implicit feedback. In Proceedings of 28th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '05) (pp. 43–50).
Shen, X., Tan, B., & Zhai, C. (2005b). Implicit user modeling for personalized search. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM '05) (pp. 824–835).
Shen, X., Tan, B., & Zhai, C. (2007). Privacy protection in personalized search. SIGIR Forum, 41(1), 4–17.
Shen, X., & Zhai, C. (2003). Exploiting query history for document ranking in interactive information retrieval. In Proceedings of 26th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '03) (pp. 377–378).
Shokouhi, M. (2013). Learning to personalize query auto-completion. Proceedings of 36th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 13) (pp. 103–112).
Sieg, A., Mobasher, B., & Burke, R. (2007). Web search personalization with ontological user profiles. In Sixteenth ACM Conference on Conference on Information & Knowledge Management (pp. 525–534). ACM.
Sihvonen, A., & Vakkari, P. (2004). Subject knowledge improves interactive query expansion assisted by a thesaurus. Journal of Documentation, 60(6), 673–690.
Sontag, D., Collins-Thompson, K., Bennett, P.N., White, R.W., Dumais, S., & Billerbeck, B. (2012). Probabilistic models for personalizing web search. In Proceedings of the 5th International Conference on Web Search and Data Mining (WSDM 12) (pp. 433–442).
Sparck-Jones, K. (1995). Reflections on TREC. Information Processing and Management, 31(3), 291–314.
Sparck-Jones, K. (2000). Further reflections on TREC. Information Processing and Management, 36(1), 37–85.
Speretta, M., & Gauch, S. (2005). Personalized search based on user search histories. In Proceedings of the 2005 IEEE / WIC / ACM International Conference on Web Intelligence (WI 2005) (pp. 622–628). IEEE Computer Society.
Steichen, B., O'connor, A., & Wade, V. (2011). Personalisation in the wild – providing personalisation across semantic, social and open-web resources. In Proceedings of the 22nd ACM conference on Hypertext and hypermedia (HT '11) (pp. 73–82).
Sun, J.-T., Zeng, H.-J., Liu, H., Lu, Y., & Chen, Z. (2005). CubeSVD: A novel approach to personalized web search. In Proceedings of the 14th International Conference on World Wide Web (WWW 2005) (pp. 382–390). New York, NY: ACM.
Tanudjaja, F., & Mui, L. (2002). Persona: A contextualized and personalized web search. In Proceedings of 35th Annual Hawaii International Conference on System Sciences (HICSS '02) (pp. 67–75). New York, NY: IEEE.
Taylor, A.R., Cool, C., Belkin, N.J., & Amadio, W.J. (2007). Relationships between categories of relevance criteria and stage in task completion. Information Processing & Management, 43, 1071–1084.
Taylor, R.S. (1968). Question negotiation and information seeking in libraries. College and Research Libraries, 29(3), 178–194.
Teevan, J., Dumais, S.T., & Horvitz, E. (2005a). Beyond the commons: Investigating the value of personalizing web search. In Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA).
Teevan, J., Dumais, S.T., & Horvitz, E. (2005b). Personalizing search via automated analysis of interests and activities. In Proceedings of 28th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '05) (pp. 449–456).
Teevan, J., Dumais, S.T., & Horvitz, E. (2007). Characterizing the value of personalizing search. In Proceedings of 30th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '07) (pp. 757–758).
Teevan, J., Dumais, S.T., & Horvitz, E. (2010). Potential for personalization. ACM Transactions on Computer-Human Interaction, 17(1), 4.
Teevan, J., Dumais, S.T., & Liebling, D.J. (2008). To personalize or not to personalize: Modeling queries with variation in user intent. In

Proceedings of 31st Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR ’08) (pp. 163–170).
Teevan, J., Liebling, D.J., & Geetha, G.R. (2011). Understanding and predicting personal navigation. In Proceedings of the 4th International Conference on Web Search and Data Mining (WSDM 11) (pp. 85–94).
Teevan, J., Morris, M.R., & Bush, S. (2009). Discovering and using groups to improve personalized search. In Proceedings of the 2nd International Conference on Web Search and Data Mining (WSDM 09) (pp. 15–24).
Tiwari, R.G, Husain, M., Srivastava, V., & Agrawal, A. (2011). Web personalization by assimilating usage data and semantics expressed in ontology terms. In International Conference and Workshop on Emerging Trends in Technology (ICWET 2011) (pp. 516–521).
Toms, E., MacKenzie, T., Jordan, C., O’Brien, H., Freund, L., Toze, S., Dawe, E., & MacNutt, A. (2007). How task affects information search. In N. Fuhr, N. Lalmas, & A. Trotman (Eds.), Workshop Pre-Proceedings in Initiative for the Evaluation of XML Retrieval (INEX) 2007 (pp. 337–341).
Ustinovskiy, U. & Serdyukov, P. (2013). Personalization of Web-search using short-term browsing context. In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM 13) (pp. 1979–1988).
Vakkari, P. (1999). Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. Information Processing and Management, 35, 819–837.
Vakkari, P. (2000). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. In RIAO ’00 Content-Based Multimedia Information Access, Volume 1 (pp. 894–907).
Vakkari, P. (2001). A theory of the task-based information retrieval. Journal of Documentation, 57(1), 44–60.
Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. Journal of Documentation, 56(5), 540–562.
Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal research. Information Processing & Management, 39(3), 445–463.
Vallet, D. & Castells, P. (2012). Personalized diversification of search results. In Proceedings of 35th Annual ACM International Conference on Research and Development in information retrieval (SIGIR ‘12) (pp. 841–850).
Vechtomova, O., Karamuftuoglu, M., & Lam, E. (2003). Interactive search refinement techniques for HARD tasks. TREC HARD track 2003. In: Proceedings of the 2003 Text REtrieval Conference (TREC 2003). Retrieved on 14 December 2018 from http://trec.nist.gov/pubs/trec12/papers/uwaterloo-olga.hard.ps
Wakabayashi, D. (2018). Google Plus will be shut down after user information was exposed. New York Times. Retrieved October 11, 2018, from https://www.nytimes.com/2018/10/08/technology/google-plus-security-disclosure.html.
Wang, H., He, X., Chang, M.-W., Song, Y., White, R.W., Chu, W. (2013). Personalized ranking model adaptation for Web search. In Proceedings of 36th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 13) (pp. 323–332).
Wang, P., Guo, J., Lan, Y., Xu, J., & Cheng, X. (2016). Your cart tells you: Inferring demographic attributes from purchase data. Proceedings of the Ninth ACM international Conference on Web Search and Data Mining (WSDM ‘16) (pp. 173–182).
Wang, Q. & Jin, H. (2010). Exploring online social activities for adaptive search personalization. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM 10) (pp. 999–1008).
Weber, I., & Castillo, C. (2010). The demographics of web search. In Proceedings of 33rd Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 2010) (pp. 523–530).
Wedig, S., & Madani, O. (2006). A large-scale analysis of query logs for assessing personalization opportunities. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006) (pp. 742–747).
Wei, C., Sen, W., Yuan, Z., & Chen, L. (2009). Algorithm of mining sequential patterns for web personalization services. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 40 (2), 57–66.
Welch, M.J. & Cho, J. (2008). Automatically identifying localizable queries. In Proceedings of 31st Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 08) (pp. 507–514).
White, R., & Kelly, D. (2006). A study of the effects of personalization and task information on implicit feedback performance. In Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM ‘06), (pp. 297–306).
White, R.W. (2016). Interaction with search systems. New York, NY: Cambridge University Press.
White, R.W., Chu, W., Hassan, A., He, X., Song, Y., & Wang, H. (2013). Enhancing personalized search by mining and modeling task behavior. In Proceedings of the 22nd International Conference on World Wide Web (WWW 13) (pp. 1411–1420). New York, NY: ACM.
White, R.W., Drucker, S.M. (2007). Investigating behavioral variability in web search. In Proceedings of the 16th International Conference on World Wide Web (WWW 2007) (pp. 21–30).
White, R.W., Dumais, S., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. In R. Baeza-Yates, P. Boldi, B. Ribeiro-Neto, & B.B. Cambazoglu (Eds.), Proceedings of the 2nd International Conference on Web Search and Data Mining (ACM WSDM ‘09) (pp. 132–141). New York, NY: ACM.
White, R.W., Ruthven, I., & Jose, J.M. (2005). A study of factors affecting the utility of implicit relevance feedback. In Proceedings of 28th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 2005) (pp. 35–42).
Wildemuth, B.W. (2004). The effects of domain knowledge on search tactic formulation. Journal of the American Society for Information Science & Technology, 55(3), 246–258.
Wildemuth, B.W., Freund, L., & Toms, E.G. (2014). Untangling search task complexity and dfficulty in the context of interactive information retrieval studies. Journal of Documentation, 70(6), 1118–1140.
Witkin, H.A., Moore, C.A., Goodenough, D.R., & Cox, P.W. (1977). Field-dependent and field-independent cognitive styles and their educational implications. Review of Educational Research, 47(1), 1–64.
Xie, X., Zeng, H.-J., & Ma, W.-Y. (2002). Enabling personalization services on the edge. In Proceedings of the Tenth ACM International Conference on Multimedia (pp. 263–266).
You, G. & Hwang, S. (2007). Personalized ranking: A contextual ranking approach. In Proceedings of the 2007 ACM Symposium on Applied Computing (SAC ‘07) (pp. 506–510). New York: ACM.
Zhang, X.-M., Anghelescu, H.G.B., & Yuan, X. (2005). Domain knowledge, search behavior, and search effectiveness of engineering and science students. Information Research, 10(2), 21.
Zhang, X.-M., Liu, J., Cole, M., & Belkin, N.J. (2015). Predicting users’ domain knowledge in information retrieval using multiple regression analysis of search behaviors. Journal of the Association for Information Science and Technology, 66(5), 980–1000.
Zhong, Y., Yuan, N. J., Zhong, W., Zhang, F., & Xie, X. (2015). You are where you go: Inferring demographic attributes from location check-ins. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (pp. 295–304). New York, NY: ACM.
Zhu, Y., Callan, J., & Carbonell, J. (2008). The impact of history length on personalized search. In Proceedings of 31st Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR 08) (pp. 715–716). New York, NY: ACM.

DOI: 10.1002/asi