

# Personalized Entity Search by Sparse and Scrutable User Profiles

Ghazaleh H. Torbati

ghazaleh@mpi-inf.mpg.de  
Max Planck Institute for Informatics  
Saarbrücken, Germany

Andrew Yates

ayates@mpi-inf.mpg.de  
Max Planck Institute for Informatics  
Saarbrücken, Germany

Gerhard Weikum

weikum@mpi-inf.mpg.de  
Max Planck Institute for Informatics  
Saarbrücken, Germany

## ABSTRACT

Prior work on personalizing web search results has focused on considering query-and-click logs to capture users' individual interests. For product search, extensive user histories about purchases and ratings have been exploited. However, for general entity search, such as for books on specific topics or travel destinations with certain features, personalization is largely underexplored. In this paper, we address personalization of book search, as an exemplary case of entity search, by exploiting sparse user profiles obtained through online questionnaires. We devise and compare a variety of re-ranking methods based on language models or neural learning. Our experiments show that even very sparse information about individuals can enhance the effectiveness of the search results.

## CCS CONCEPTS

• Information systems → Personalization.

## KEYWORDS

personalized entity search, sparse user profile, knowledge graph

### ACM Reference Format:

Ghazaleh H. Torbati, Andrew Yates, and Gerhard Weikum. 2020. Personalized Entity Search by Sparse and Scrutable User Profiles. In *2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*, March 14–18, 2020, Vancouver, BC, Canada. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3343413.3378011>

## 1 INTRODUCTION

**Motivation and Problem:** Personalization to improve web search result ranking has been a long-standing theme in information retrieval [34, 36]. With the increasing availability of individual users' online traces and derived traits, personalization is again gaining importance for chatbots, recommender systems, product search, and more. [5] has formulated a vision and research agenda for constructing and leveraging *personal knowledge graphs* (PKG's) in such settings. In this paper, we investigate the role of PKG's for topical entity search, with the challenging case that the only per-user knowledge is a sparse profile obtained from a short questionnaire. In contrast to the "data-hungry" approaches of prior works, we focus on this "minimal PKG" case to strengthen the user's ability

to understand and control her user profile, similar to what major search engines offer for controlling the personalization of ads (e.g., [adssettings.google.com](https://adssettings.google.com)). Note that our case is more ambitious, though: minimal-PKG profiles aim to capture the bare necessities, whereas ads controls often comprise a hundred or more tags for the same user. The fewer traits the profile contains and the more explicit they are (as opposed to learned latent models), the more scrutable and actionable the personalization model becomes from a user perspective.

**State of the Art and its Limitations:** The most important line of exploiting user information for general web search is based on *query-and-click logs* (e.g., [30, 34]). This helps in interpreting user interests and intents for ambiguous queries as well as for identifying salient pages for popular queries, and for suggestions for query auto-completion (e.g., [29]). In all this, cues about the user's location and daytime are a major asset, too (e.g., [8]).

*Recommender systems* have incorporated personalization as well, for ads, products and other contents (e.g., [17, 26, 31]). Here, structured data is leveraged, most notably, purchases or ratings of products, likes of news, YouTube videos, Instagram photos, etc. This field has recently paid attention to *scrutable recommendations* that are comprehensible by end-users and pinpoint the specific data that explains how the recommended item was computed [6, 25, 38]. However, these approaches are at least as "data-hungry" as the search engines, and require extensive user-specific data.

*Entity search* about people, products or events has received great attention and has been incorporated into major search engines (see, e.g., [4, 7] and further references there). This methodology leverages large knowledge graphs to infer the focus of the query and/or return crisp entities as answers. However, except for special cases such as music recommendation [13] and consumer product search [2], there is hardly any work on *personalized* entity search with individual user traits.

**Approach:** This paper explores the direction of personalized entity search, relying solely on a "minimal-PKG" user profile for scrutability. The requirement is that users can fully understand and control the information that drives the personalization (e.g., modify or revoke pieces of a profile), and that this is as sparse as possible while still giving benefits. We consider online *questionnaires* as a source of sparse and scrutable user profiling.

**Example:** Figure 1 shows an excerpt of a questionnaire, obtained by hiring crowd workers at Amazon MTurk. It captures demographic attributes (age, gender, location etc.), personal tastes regarding books, movies and music, and hobbies – all entered as free-form text (as opposed to guiding users through menus which may create bias). The questionnaire has only 10 questions, and users spent 7 minutes, on average, to fill in their answers. For entity search, we consider medium-grained query topics about books: finer-grained

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CHIIR '20, March 14–18, 2020, Vancouver, BC, Canada

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-6892-6/20/03...\$15.00

<https://doi.org/10.1145/3343413.3378011>

Demographics	
What is your age?	<input checked="" type="checkbox"/> 31-50 years old
What is your gender?	<input checked="" type="checkbox"/> Female
Where are you from?	<input checked="" type="checkbox"/> New York, USA
What is your family status?	<input checked="" type="checkbox"/> Married
Preferences	
What are your hobbies?	<input checked="" type="checkbox"/> Reading, Baking and walking
Please name 3 of your favorite books.	<input checked="" type="checkbox"/> Catcher in the Rye Love Poems by Pablo Neruda The Celestine Prophecy
Which kind of books do you like?	<input checked="" type="checkbox"/> Poetry, science fiction, true crime
Please name 3 of your favorite movies/series.	<input checked="" type="checkbox"/> Return of the King The Hobbit Requiem for a Dream
Which genre of movies do you like?	<input checked="" type="checkbox"/> Science Fiction, documentary
Please name your favorite singers/bands/composers.	<input checked="" type="checkbox"/> Blink 182, Portishead

Figure 1: Sparse user profile from questionnaire.

Query: Time travel Domain: Books Results and user's judgement:	
	<b>The Christmas Eve Letter: A Time Travel Novel</b> "Very interesting" - This book sounds emotional and positive, I think I would really enjoy it.
	<b>Time Travel Adventures Of The 1800 Club</b> "Interesting" - Living in New York myself I think this sounds like a very interesting story. I think it could do a lot for people to relax and slow down like people did in the past.
	<b>Divine Hotel: Time Travel Mystery</b> "Not interesting" - I do not find historical fiction book interesting, so I doubt I'd like this one.

Figure 2: Examples of entity search results for the query "time travel" and the user's judgements and justifications.

than merely specifying a genre of interest (e.g., "history" or "science fiction"), but not as specific as aiming for a single entity as an answer. Examples of our queries are: "historical romance", "sci fi noir", and "time travel".

Consider the example query "time travel" on the domain of books, executed for entity pages returned from goodreads.com. The results were personalized for the participants of our MTurk study, including the user whose profile is shown in Figure 1. Figure 2 shows three query results and the user's judgments along with her justification sentences. As shown in Figure 1, she is from "New York", and one of her favorite books is "Love poems from Pablo Neruda". Personalized re-ranking over the top-100 answers correctly inferred that she would find an emotional book "very interesting" and a book with a New-York-based storyline "interesting".

**Contribution:** Our methodology focuses on ranking entities as candidate answers, based on the sparse profiles from questionnaires. We cast this background knowledge into user-specific language models or queries, but treat entities (e.g., favorite books, movies or singers/bands) in the user answers different from free-form text (e.g., talking about hobbies). To this end, we incorporate entity descriptions from general knowledge graphs, and we also consider word embeddings (from word2vec) as an additional input. This way, we derive a suite of *re-ranking methods* for pools of candidate

answers obtained by running a standard search engine on the book community portal goodreads.com, which has both content summaries for millions of books and comments from nearly 100 million users.

The key hypothesis that we test in this study is that even a very small PKG about a user can improve the quality of search answers as assessed by the user herself. To this end, we hired the same MTurk workers for judgements of interestingness who contributed their questionnaire answers. In total, we evaluated 115 user-query pairs from 33 distinct users.

Our experiments compare a variety of re-ranking methods, with different degrees of incorporating sparse user profiles. The results are preliminary, as the study is limited in scale and scope. Nevertheless, our findings indicate that even sparse profiles yield statistically significant benefits over not personalizing at all.

## 2 RELATED WORK

Prior works covered two major dimensions [16]:

1. Creating user models from explicit signals like queries, clicks, likes, social links, etc. [1] or/and rich contents like email histories or desktop data [14, 22].
2. Leveraging this background knowledge for answer ranking, query expansion, and auto-completion suggestion [12, 24, 29].

On the first dimension, [34] pioneered the analysis of user interests and activities reflected in query, click and mail histories, and possibly even other online contents written or read by a user. [28] focused on short-term context, like browser sessions, to infer the user's interest and personalize interactive search. Numerous follow-up works addressed the analysis and usage of query-and-click logs and browsing sessions. To learn from this kind of expressive but highly noisy data, [1] introduced predictive models with learning-to-rank features, whereas [15] and [33] explored the use of similarity signals from taxonomies and ontologies. [37] learned models of user interests for proactive "zero-query" search. [8] studied the important role of user location.

On the second dimension, prior works explored personalization for ranking as well as query expansion and query suggestions. For *personalized ranking*, [32] developed methods for incorporating user-specific priors into language models. The interplay of a user's long-term behavior and short-term context for personalized ranking has been investigated in [8, 10]. [35] and [9] addressed the issue of selective personalization: when to incorporate user profiles.

Another line of research addresses *query expansion* for personalization. [11, 39] utilize folksonomy data, like user-provided tags in social bookmarking communities, as a source for expanding a user's queries. [22] personalizes email search via word embeddings learned from email histories. [14] proposes methods for harnessing a user's desktop files (incl. email). The viability of all these methods relies on the availability of large collections of user data.

The same assumption holds for prior work on *query auto-completion* [12, 29], perhaps the most successful line of personalization in major search engines. The underlying user data ranges from long-term query-and-click histories to browser histories to email contents, in addition to location and daytime as short-term context.

For *entity search*, to the best of our knowledge, prior work on personalization is scarce. CLEF had a series of competitions on book recommendations [21], but this relied on posts, tags, reviews

and ratings by many users in the LibraryThing community and the Amazon shop. The closest to our work is [3] on personalized product search. It is based on learning embeddings for users and items in the same semantic space, by leveraging user-written reviews on item pages. However, such rich data about individual users is not easily available for general entity search.

### 3 METHODOLOGY

This section discusses how we cast user input from questionnaires into user profiles, and how these are incorporated into different kinds of rankers: language models, BM25, and neural methods.

#### 3.1 User Profiles

In contrast to prior works based on extensive logs of user queries, clicks and other activities, we focus on sparse and concise models of user-specific interests and tastes. To this end, we designed a small questionnaire and hired crowd workers at MTurk to fill in their answers. As shown in Figure 1, these profiles cover basic demographics and personal attributes like hobbies, favorite books and book genres, favorite movies and movie genres, and favorite singers or music bands. The advantage of this “minimal-PKG” approach is that such a profile is easily comprehensible by the respective user, and easy to control as the user may want to change it when her interests evolve or she becomes concerned about privacy. This kind of scrutability and controllability is impossible with a huge log and latent models derived from massive user data.

As most fields in the questionnaire are free-form text, we treat the profiles as text documents, like a statistical language model, or a bag-of-words model, or a term-sequence model, depending on which ranking method is adopted.

**Incorporating Entity Descriptions:** The user profiles include named entities like favorite books, movies and musicians. To take advantage of the sparse data to its fullest potential, we employ named entity disambiguation to link the entity mentions to their respective entries in the Yago knowledge base and in Wikipedia. Our implementation uses the AIDA tool [19]; for experiments we manually checked (and corrected a few of) the computed links to eliminate a potential source of erroneous drift. For the user models (language model or other), we did not include the entity names themselves to avoid overfitting to specific entities. Instead, we incorporated the descriptions of these entities from the knowledge base and Wikipedia. For the test case of books, we selected the first paragraph of each book’s Wikipedia article. Typically, this gives summary information about the book’s story.

#### 3.2 Ranking Methods

Given a pool  $M$  of non-personalized results for a query  $q$  and a user  $u$ , we want to re-rank the results such that the ranking reflects the user’s individual interests. All ranking methods treat query answers as text documents about specific entities. In our experiments, these are pages from goodreads.com, each featuring a single book – typically in the form of a content summary and user comments.

**3.2.1 Statistical Language Models.** This method adopts a query-likelihood model where the score of document  $d \in M$  for query  $q$  is proportional to the Kullback-Leiber divergence between the language models of  $q$  and  $d$ . As we want to capture the interestingness of a document for a user, rather than general relevance for the

query, we incorporate the user profile as another language model. We use a mixture model over the estimated language models  $\theta_q, \theta_d, \theta_u$  for the query, document and user, respectively, and a general background model  $\theta_C$  (based on ClueWeb09) for smoothing.

**Incorporating Word Embeddings:** Optionally, we integrate word embeddings; our implementation uses pre-computed word2vec vectors. The language model is augmented by a translation model, largely following [23]. Using the cosine between word-embedding vectors as term-term similarity  $sim$ . In the following ranking equation,  $div$  is the Kullback–Leibler divergence,  $p(w|u) = \frac{sim(w,u)}{\sum_{u' \in V_d} sim(u',u)}$  are the probabilities for the word-word relatedness model,  $V_q$  and  $V_u$  denote the query and user vocabularies,  $p(w|\theta)$  are the estimated probabilities of word  $w$  in each language model  $\theta$ ,  $\mu$  is the Dirichlet smoothing parameter, and  $\lambda$  determines the relative weight of the query and user models.

$$\begin{aligned} rank(q, d) &\propto \lambda div(\theta_q \| \theta_{d_\mu}) + (1 - \lambda) div(\theta_u \| \theta_{d_\mu}) = \\ &\lambda \sum_{w \in V_q} p(w|\theta_q) \log \frac{p(w|\theta_q)}{\sum_{u \in V_d} p(w|u)p(u|\theta_d) + \mu p(w|\theta_C)} + \\ &(1 - \lambda) \sum_{w \in V_u} p(w|\theta_u) \log \frac{p(w|\theta_u)}{\sum_{u \in V_d} p(w|u)p(u|\theta_d) + \mu p(w|\theta_C)} \end{aligned} \quad (1)$$

**3.2.2 BM25 Ranking.** To personalize the BM25 scoring, we use query expansion, treating the user profile as a bag-of-words. As we use it only for re-ranking a pool of candidate results, the query is fixed and the user profile becomes the actual query. When incorporating entity descriptions, they are included in this query.

**3.2.3 Neural Ranking.** We adopted the state-of-the-art methods DRMM [18] and PACCR [20] as representatives for neural ranking. Analogously to BM25, we focus on re-ranking and thus treat the user profile as the query, as a term sequence where each term is represented as an embedding vector using word2vec.

## 4 EXPERIMENTS

### Research Questions

Our experiments target the following research questions.

- RQ1: To what extent can sparse user profiles improve rankings towards individual interests? We compare non-personalized and personalized versions of all ranking methods of Section 3.2.
- RQ2: Do entity descriptions improve the ranking? We investigate ranking variants with and without entity models (except for neural methods which cannot handle long text as query input).
- RQ3: Do word embeddings improve the ranking by semantic similarities between terms? We examine this by running the language-model-based ranker with and without embeddings.

### Data

We conducted an Amazon MTurk study in which we recruited 33 people for a two-stage task. In the first stage, users create sparse profiles via questionnaires as explained in Section 3.1. In the second stage, for search-result evaluation, we asked the same workers for their judgements of *interestingness* on results of a small set of self-selected queries (which they deemed of personal interest). Queries were medium-grained gathered by crowdsourcing. Examples from a total set of 50 queries are: african books, greek mythology, historical fiction, memoirs and autobiography, novels made into movies,

Methods	nDCG@20			nDCG@5			Precision@1		
	Query	+ User Profile	+ Profile + Entities	Query	+ User Profile	+ Profile + Entities	Query	+ User Profile	+ Profile + Entities
LM	0.783	<b>0.797</b>	0.772	0.557	<b>0.576</b>	0.525	0.765	<b>0.817</b>	0.687
LM+WV	<b>0.782</b>	0.781	0.776	<b>0.558</b>	0.548	0.541	<b>0.748</b>	0.722	0.704
BM25	0.781	<b>0.799</b>	0.776	0.555	<b>0.579</b>	0.536	0.774	<b>0.809</b>	0.696
DRMM	0.752	<b>0.782</b>	-	0.493	<b>0.544</b>	-	0.644	<b>0.730</b>	-
PACRR	0.766	<b>0.792</b>	-	0.53	<b>0.572</b>	-	0.67	<b>0.809</b>	-
Commercial SE	0.757	-	-	0.505	-	-	0.652	-	-

**Table 1: Results for different rankers when using the queries, user profiles, and user profiles enhanced by entity descriptions.**

scandinavian suspense, sword and sorcery, time travel, true crime, and victorian society.

The judgements were graded "*not interesting*" (0), "*interesting*" (1), "*very interesting*" (2) or "*don't know*" (*discarded*). To ensure faithful judgements, we required a justification sentence for each judgement; examples are shown in Figure 2. Our pool of results contained 50 queries and 100 results to be judged per query. As it was not feasible to evaluate all 100 results for each user and query, we randomly selected 20 answers per query for per-user assessment. We did not choose the top 20 results from a baseline ranking (or pool rankings) in order to reduce bias towards (global) popularity as a ranking criterion. We obtained judgements for 115 user-query pairs by 33 users with 47 distinct queries – 2163 judged items in total.

### Experimental Setup

To investigate the utility of sparse user profiles for our setting, we evaluated a range of retrieval methods both with and without the user profiles. The methods' performance was evaluated using normalized Discounted Cumulative Gain (nDCG@5 and nDCG@20) and Precision@1 (P@1) applied to condensed lists, with all unjudged results filtered out. This follows [27], which argues that this approach to handling partial judgements is preferable to other metrics. To compute precision, a result was considered good when deemed "very interesting" or "interesting" by the respective judge.

Due to the limited size of our dataset (despite considerable spending on MTurk), we minimized the tuning of hyper-parameters and picked reasonable defaults where possible. With the neural models, we used ten-fold cross-validation (with each of the ten folds containing unique queries): eight folds for training and the remaining two for validation and testing. These models were trained with a softmax loss over pairs of documents. Due to the substantial increase in query length, we do not consider using entity descriptions with neural models. We used the following hyper-parameters:

- $\lambda$  determines the relative influence of query and user models. To test extreme cases, we either set  $\lambda = 0$  or  $\lambda = 1$ .
- $\mu$ , the Dirichlet prior parameter for smoothing, was set to the average document length for each query.
- N-gram order in the language model approach is set to 1.
- When incorporating embedding similarity into the language model, we discarded terms with a similarity below  $T = 0.5$ .
- $\theta_C$  is the background model, for which we use ClueWeb09.
- We set  $k_1 = 1.5$  and  $b = 0.75$  with BM25, after determining that a grid search on our folds could not improve results.
- DRMM used an IDF gate with LCH-normalized histograms.
- PACRR used unigrams through trigrams, 32 filters, a  $k$ -max of 2, and a combination layer of size 32.

## 5 RESULTS

The results are shown in Table 1 (LM denotes the language model method, LM+WV additionally uses word2vec embeddings). As a reference point, we include the quality of the original ranking obtained from a commercial search engine.

Regarding RQ1, we observe that incorporating sparse user profiles consistently improves results across methods and across metrics, with the exception of LM+WV. We performed a paired t-test over all samples of non-personalized vs. personalized rankings; the p-values for the hypothesis that personalization is beneficial were below 0.01 for all metrics.

Regarding RQ2 and RQ3, neither word embeddings nor entity descriptions helped to improve rankings, though. Their performance even dropped below the non-personalized baselines. We believe that the word2vec embeddings are too broad and diluted the focus of our queries. This would call for user-specific embeddings, but it is an open issue on how to obtain these. The negative impact of entity descriptions is due to the Analyzing incorrectly high-ranked results indicated that cues from the out-of-scope entity descriptions can be misleading due to the ambiguity of words. This calls for extending our model to consider other metrics such as entity specificity (i.e. selectively using entities and descriptions).

**Ablation Study:** We performed an ablation study with further restriction of user profiles, to investigate how minimal the PKG could be while still being beneficial. We evaluated two variants: 1) omitting the book-related fields from questionnaires (but keeping fields on movies and music), and 2) keeping solely the demographic attributes and the hobbies field. The first restriction led a notable loss in nDCG and precision, but still gave decent quality, whereas the variant with minimal profiles degraded. For example, when using the BM25 ranker with 80% nDCG@20 and 81% precision@1 with full profiles, leaving out the book fields gave 79% nDCG@20 and 77% precision@1. Capturing the user's interests and tastes is crucial, but does not have to be domain-specific, like books.

## 6 CONCLUSION

Given the limited scope and scale of our study, the results and findings are preliminary. Nevertheless, they indicate that even sparse user profiles have potential to improve ranking quality through personalization. Our work in progress involves comparisons to richer profiles obtained by gathering chat-like user-to-user conversations. The overriding goal is to understand upper bounds of personalized quality and gain insight on how well these can be approximated with different extents of user profiling.

## REFERENCES

- [1] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno. 2006. Learning user interaction models for predicting web search result preferences. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 3–10.
- [2] Qingyao Ai, Yongfeng Zhang, Keping Bi, Xu Chen, and W Bruce Croft. 2017. Learning a hierarchical embedding model for personalized product search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 645–654.
- [3] Qingyao Ai, Yongfeng Zhang, Keping Bi, Xu Chen, and W. Bruce Croft. 2017. Learning a Hierarchical Embedding Model for Personalized Product Search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 645–654. <https://doi.org/10.1145/3077136.3080813>
- [4] Krisztian Balog. 2018. *Entity-Oriented Search*. The Information Retrieval Series, Vol. 39. Springer.
- [5] Krisztian Balog and Tom Kenter. 2019. Personal Knowledge Graphs: A Research Agenda. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2019, Santa Clara, CA, USA, October 2-5, 2019*. 217–220.
- [6] Krisztian Balog, Filip Radlinski, and Shushan Arakelyan. 2019. Transparent, Scrutable and Explainable User Models for Personalized Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*. 265–274.
- [7] Hannah Bast, Björn Buchhold, and Elmar Haussmann. 2016. Semantic Search on Text and Knowledge Bases. *Foundations and Trends in Information Retrieval* 10, 2-3 (2016), 119–271.
- [8] Paul N. Bennett, Filip Radlinski, Ryan W. White, and Emine Yilmaz. 2011. Inferring and using location metadata to personalize web search. In *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011*. 135–144.
- [9] Paul N. Bennett, Milad Shokouhi, and Rich Caruana. 2015. Implicit Preference Labels for Learning Highly Selective Personalized Rankers. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval, ICTIR 2015, Northampton, Massachusetts, USA, September 27-30, 2015*. 291–300.
- [10] Paul N. Bennett, Ryan W. White, Wei Chu, Susan T. Dumais, Peter Bailey, Fedor Borisuyk, and Xiaoyuan Cui. 2012. Modeling the impact of short- and long-term behavior on search personalization. In *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012*. 185–194.
- [11] Claudio Biancalana, Fabio Gasparrini, Alessandro Micarelli, and Giuseppe Sansonetti. 2013. Social semantic query expansion. *ACM TIST* 4, 4 (2013), 60:1–60:43.
- [12] Fei Cai and Maarten de Rijke. 2016. Selectively Personalizing Query Auto-Completion. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*. 993–996.
- [13] Ben Carterette. 2019. *Music Recommendation at Spotify (Tutorial Slides)*. <http://sigir.org/afirm2019/slides/16.%20Friday%20-%20Music%20Recommendation%20at%20Spotify%20-%20Ben%20Carterette.pdf>
- [14] Paul-Alexandru Chirita, Claudiu S. Firan, and Wolfgang Nejdl. 2007. Personalized query expansion for the web. In *SIGIR 2007: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, July 23-27, 2007*. 7–14.
- [15] Paul-Alexandru Chirita, Wolfgang Nejdl, Raluca Paiu, and Christian Kohlschütter. 2005. Using ODP metadata to personalize search. In *SIGIR 2005: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, August 15-19, 2005*. 178–185.
- [16] M. Rami Ghorab, Dong Zhou, Alexander O'Connor, and Vincent Wade. 2013. Personalised Information Retrieval: survey and classification. *User Modeling and User-Adapted Interaction* 23, 4 (01 Sep 2013), 381–443. <https://doi.org/10.1007/s11257-012-9124-1>
- [17] Carlos A. Gomez-Urbe and Neil Hunt. 2016. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Trans. Management Inf. Syst.* 6, 4 (2016), 13:1–13:19.
- [18] Jiafeng Guo, Yixing Fan, Qingyao Ai, and W. Bruce Croft. 2016. A Deep Relevance Matching Model for Ad-hoc Retrieval. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*. 55–64.
- [19] Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenauf, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust Disambiguation of Named Entities in Text. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL*. 782–792.
- [20] Kai Hui, Andrew Yates, Klaus Berberich, and Gerard de Melo. 2018. Co-PACRR: A Context-Aware Neural IR Model for Ad-hoc Retrieval. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018*. 279–287.
- [21] Marijn Koolen, Toine Bogers, Maria Gäde, Mark Hall, Iris Hendrickx, Hugo Huurdeman, Jaap Kamps, Mette Skov, Suzan Verberne, and David Walsh. 2016. Overview of the CLEF 2016 Social Book Search Lab. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, Norbert Fuhr, Paulo Quaresma, Teresa Gonçalves, Birger Larsen, Krisztian Balog, Craig Macdonald, Linda Cappellato, and Nicola Ferro (Eds.). Springer International Publishing, Cham, 351–370.
- [22] Saar Kuzi, David Carmel, Alex Libov, and Ariel Raviv. 2017. Query expansion for email search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 849–852.
- [23] Saar Kuzi, Anna Shok, and Oren Kurland. 2016. Query Expansion Using Word Embeddings. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*. 1929–1932.
- [24] Nicolaas Matthijs and Filip Radlinski. 2011. Personalizing web search using long term browsing history. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011*. 25–34.
- [25] Changhua Pei, Yi Zhang, Yongfeng Zhang, Fei Sun, Xiao Lin, Hanxiao Sun, Jian Wu, Peng Jiang, Junfeng Ge, Wenwu Ou, and Dan Pei. 2019. Personalized re-ranking for recommendation. In *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019*. 3–11.
- [26] Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). 2015. *Recommender Systems Handbook*. Springer.
- [27] Tetsuya Sakai. 2007. Alternatives to bpre. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 71–78.
- [28] Xuehua Shen, Bin Tan, and ChengXiang Zhai. 2005. Implicit user modeling for personalized search. In *Proceedings of the 2005 ACM CIKM International Conference on Information and Knowledge Management, Bremen, Germany, October 31 - November 5, 2005*. 824–831.
- [29] Milad Shokouhi. 2013. Learning to personalize query auto-completion. In *The 36th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '13, Dublin, Ireland - July 28 - August 01, 2013*. 103–112.
- [30] Fabrizio Silvestri. 2010. Mining Query Logs: Turning Search Usage Data into Knowledge. *Foundations and Trends in Information Retrieval* 4, 1-2 (2010), 1–174.
- [31] Brent Smith and Greg Linden. 2017. Two Decades of Recommender Systems at Amazon.com. *IEEE Internet Computing* 21, 3 (2017), 12–18.
- [32] David Sontag, Kevyn Collins-Thompson, Paul N. Bennett, Ryan W. White, Susan T. Dumais, and Bodo Billerbeck. 2012. Probabilistic models for personalizing web search. In *Proceedings of the Fifth International Conference on Web Search and Web Data Mining, WSDM 2012, Seattle, WA, USA, February 8-12, 2012*. 433–442.
- [33] Sofia Stamou and Alexandros Ntoulas. 2009. Search personalization through query and page topical analysis. *User Modeling and User-Adapted Interaction* 19, 1 (01 Feb 2009), 5–33. <https://doi.org/10.1007/s11257-008-9056-y>
- [34] Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2005. Personalizing search via automated analysis of interests and activities (SIGIR Test-of-Time Award 2017). In *SIGIR 2005: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, August 15-19, 2005*. 449–456.
- [35] Jaime Teevan, Susan T. Dumais, and Daniel J. Liebling. 2008. To personalize or not to personalize: modeling queries with variation in user intent. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008, Singapore, July 20-24, 2008*. 163–170.
- [36] Ji-Rong Wen, Zhicheng Dou, and Ruihua Song. 2018. Personalized Web Search. In *Encyclopedia of Database Systems, Second Edition*.
- [37] Liu Yang, Qi Guo, Yang Song, Sha Meng, Milad Shokouhi, Kieran McDonald, and W. Bruce Croft. 2016. Modeling User Interests for Zero-Query Ranking. In *Advances in Information Retrieval - 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20-23, 2016. Proceedings*. 171–184.
- [38] Yongfeng Zhang and Xu Chen. 2018. Explainable Recommendation: A Survey and New Perspectives. CoRR abs/1804.11192 (2018). <https://arxiv.org/abs/1804.11192>
- [39] Dong Zhou, Xuan Wu, Wenyu Zhao, Séamus Lawless, and Jianxun Liu. 2017. Query expansion with enriched user profiles for personalized search utilizing folksonomy data. *IEEE Transactions on Knowledge and Data Engineering* 29, 7 (2017), 1536–1548.