ISSN:2321-1156

International Journal of Innovative Research in Technology & Science(IJIRTS)

INFORMATION RETRIEVAL IN CONTEXT: A STATE OF THE ART

G.Krishna Raju , P.Padmanabham, A.Govardhan

Abstract : The results provided by the general purpose search engines in response to user queries are often too general and do not satisfy the information needs of the user. This is in part due to the inability of the average user to form an accurate query that reflects his information need. Such queries are often short. If the context in which the user is carrying out his search can be somehow captured, it will help to narrow down the search results thereby enhancing the user experience. Several researchers have strived to improve the user search experience by incorporating the notion of context in information retrieval. Apparently, context can denote different things. This paper is an attempt to summarize the efforts of different researchers and present a state of the art with regards to using context in information retrieval to improve user experience.

Keywords:- context, information retrieval, query intent, personal interest, document quality, search evaluation

1 .Introduction

Context can be viewed broadly in two ways. In the first perspective, context can be defined as the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood and assessed. Although such a point of view is rooted in strategically important disciplines like user behavior, cognition or human interaction, it cannot fully help see how to proceed with what should be observed and computed for implementing context within an IR system. At the other extreme, context can be viewed as the parts that immediately precede and follow a word or passage and clarify its meaning. The definition of context as a text window around a word is easier to implement and is strictly related to the nature of text and is part of common sense.

Search context can be captured from different aspects such as **content, geographical, interaction, and social**. Content variables refer to features observed from text, image, video, audio, link anchors etc.Examples of geographical variables are names added to documents or queries, digital photographs tagged with geographical coordinates and the latitude and longitude of the location associated to a user. Interaction variables refer to the interaction between users and IR systems. Click-through data, data about queries or search sessions and user behavior data (e.g., document retention, display time, eye or mouse movements) are some of the examples. Social variables are observed from social media such as tweets and friendships).

If the items of a context are gathered together, a sort of relation is obtained. Saracevic [153, p. 1918] suggested an understanding of relevance as a relation. According to this understanding, relevance is a relation over information objects and contexts which include information needs, tasks, and other elements. In Saracevic's view, context is an element of relevance ("Relevance has a context") and it is viewed as a complex, dynamic "interaction between a number of external and internal aspects, from a physical situation to cognitive and affective states, to motivations and beliefs, to situations, and back to feedback and resolution." Context is "ambiguous, even amorphous" and at most "context is a plural." In the review of relevance authored by Mizzaro [63], context "includes everything not pertaining to topic and task, but however affecting the way the search takes place and the evaluation of results."

Azzopardi [8] gives a thorough study that starts from theoretical issues, investigates whether and how language models can be an efficient and effective theoretical framework for contextual search, and ends with experiments. Bai et al. [9, 10] are examples of text window-based context with co-occurrence analysis. Bartholomew et al. [11] provide a perspective of the factorial models that are relevant to the notion of computational framework presented in this survey. Bian et al. [13] are worth reading as for the Expectation-Maximization algorithm. Blei et al. [14]'s is the original publication on latent Dirichlet allocations. The notion of geographical variable is discussed by Cai [18]. The remarks made by Chakrabarti et al. [22] on how to build an effective model and avoid bias, overfitting, etc. are useful to a newcomer to machine learning because they explain basic issues in a realistic scenario. Croft and Lafferty [25] survey language models for IR. The study by Efthimiadis [30] describes query expansion.

Inmon [34, 39] introduced the notion of time-variancy and viewed click-through datasets as an instance of data warehouses. Jones and Purves [46] provide a useful reference on the issues of geographical variables. Implicit relevance feedback is explored by Kelly and Belkin [95, 96], Kelly and Fu [97], Kelly et al. [98] and Kelly [48, 49, 50]. The survey by Lalmas and Ruthven [107] provides a precise, recent and exhaustive account of relevance feedback.

Lau et al. [52] address context at difference abstraction levels, from the conceptual, to the logical up to the statistical level. Lau et al. [53] present an interesting application of their theoretical framework and show that the vector space model is still a good baseline for search in context. The notion of geographical variable is also discussed by Reichenbacher [65] and Reichenbacher and De Sabbata [66].

2. Query Intent

Computing statistical distributions of click-through data was studied by Lee et al. [54] who were among the early researchers who postulated a relationship between query intent and click-through data. A relationship was found between query intent and query frequency. In particular, Downey et al. [29] found that users behave according to query frequency or to URL frequency — for example, search session length increases when the query is rare. There is not only a relationship between query intent and search session length; query intent evolves over time, as reported by Kulkarni et al. [105]. The fact that click-through data are not always good predictors especially when queries are rare has been confirmed within contextual advertising by Ashkan et al. [6].
Markov chains are an alternative, yet similar approach to contextual advertising as suggested by Li et al. [55]. The basic idea of Li et al. is that the greater the number of users who clicked an advertisement from a page is, the higher the relevance between the advertisement and the page. A similar yet independently conceived approach based on click-through data and session data is illustrated by Cao et al. [21]. A session is the sequence of queries issued by the same user immediately before the current query; this is a simple instance of context used for detecting the underlying intent.

Attenberg et al. [7] observe the users' trails starting from links displayed in a search engine result page until the (inferred) end of the trail. These detailed data about the users' trails highlight some facts about the relationship between click-through data and query intent.

The effectiveness of click-through data depends on the amount of historical data which are available for estimation and prediction. This is observed by Shen et al. [156] where it is reported that the performance improvement is more substantial for precision at the top 20 documents than for precision at the top 10 documents. An approach to query intent detection using eye-tracking is described by Guo and Agichtein [66].

It is possible to predict query intent by looking at the user's past search behavior according to Teevan et al. [72]. To this end, the authors automatically identified a set of navigational queries from the query logs followed by the same result — this identification is based on click entropy. Teevan et al., however, had to make quite a strong yet acceptable assumption, that is, low click entropy is a good approximation of similar intents.

Query intent detection that is based on search engine result page has been studied within a contextual advertising perspective by Ashkan and Clarke [5]. Query intent detection that is based on classification has been investigated by Broder et al. [17]. Their approach seems promising since the classification accuracy can be maximized by an appropriate quantity of documents given as input. This accuracy rises as the number of documents in a search engine result page increases, and drops when using too few documents due to too little external knowledge, or when using too many results due to extra noise.

Broder et al. [17] have used search engine result pages to obtain additional information for query intent detection. To this end, the authors employ pseudo relevance feedback and assume the top search results to be relevant to the query. As not many results are equally relevant, the given query is dispatched to a general WWW search engine, the top-ranked documents are selected and the WWW pages indicated by these top-ranked documents are retrieved. Then, the document classifier classifies the search results into the same taxonomy into which queries are to be classified. The classifier was trained by human editors who populated the taxonomy nodes with labeled examples.

Ganti et al. [33] use the corpus of advertising bids used in sponsored search. In sponsored search, each advertiser lists the queries against which an ad should be shown — as this is actually a bid, these queries are called bid-phrases. An immediate application of query intent prediction is to suggest queries to the user. The aim is to predict users' tasks based on implicit relevance feedback data (e.g., user behavior). This problem is addressed by Cheng et al. [24] where the authors propose to mine the latent search intent by using their own framework (i.e., SearchTrigger, that is, a query is triggered by the content of the browsed page) to suggest queries to users when they are browsing.

An analysis in workplace is also performed by Campbell et al. [19] .Their approach is centered on a document usage-based similarity matrix which thus defines the contextual relationships between documents. It is worth noting that the idea of context as document network was introduced early by Belkin et al. [12]. Further studies were performed by researchers in automatic hypertext construction who found that the effectiveness provided by automatic document link detection quickly decreases as the user clicks on documents after issuing a query as reported by Melucci [61].

In the current literature, the taxonomy introduced by Broder [26] has become quite well accepted because it allows researchers to simplify the methods for classifying intents. Broder suggests classifying the queries issued to a search engine as informational, navigational, and transactional. According to Dai et al. [26], the classification of queries into navigational and informational is not the only one possible. In electronic commerce, further understanding of commercial intents is crucial.

The idea that there is a significant correlation between a geographical location and an event has been tested and implemented by Abrol an Khan [1] who have proposed a geographical contextual search called TWinner. Geographical names are necessary yet not sufficient to detect geographical intent, thus requiring other query features. An application of this evidence has been reported by Yi et al. [74] and consists of tagging query words using a sort of part-of-speech tagger.

3.Personal Interest

Personalization is not the only term encountered in the literature of contextual search for denoting the adaptation of a contextual search system to the user. Pitkow et al.[64]were among those researchers who distinguished between contextualization and individualization as the two extremes of a wide range of contextual search methods.

The issues raised by the appropriateness of personalization are addressed by Luxenburger et al. [122] who aim to select the queries that are expected to benefit from the user's history. To this end, the authors introduce different granularity levels of a user profile and propose language models for modeling the user's tasks.

Attenberg et al. [11] pay a great deal of attention on the user activity performed on the sponsored search advertisements displayed by search engines next to conventional search engine result pages. Melucci and White [62] present a formal framework based on vector spaces that captures multiple aspects of user interaction and allows a new mathematical model of implicit relevance feedback to be developed. The model uses display time, document retention, and interaction events to build a multi-faceted user interest profile.

Query expansion is perhaps the most widespread method for extracting evidence about personal interest and in general from context. The paper written by Pitkow et al. [137] was one of the earliest on contextual search and in particular on using query expansion for meeting personal interests. To our knowledge, they were the first to mention the idea of comparing the current query with something else for deciding whether personalization is worth performing.

Similarly, using categories (e.g., those from the ODP) is useful to improve effectiveness according to Ma et al. [123]. When query expansion selects the number of expansion terms depending on the user and on the the user's query, it outperforms both the original ranking and the personalization in the case of a fixednumber of expansion terms as Dang and Croft [27] and Luxenburger et al. [59] report.

A formalization of the combination of click-through data, content and user profiles has been described by Sontag et al. [69]. Basically, probability distributions were extensively used in that paper for modeling every entity playing a role in a contextual search system. Thus, relevance and contextual variables are modeled as random variables, feedback is modeled as probability update through the Bayes rule, decision is supported by divergence measures. When query expansion is insufficient, it might be integrated by the user's search history as proposed by Liu et al. [121] who propose modeling and gathering the user's search history. Jones et al. [47] described a method for query modification that is based on past users' queries, phrase similarity, and query suggestion ranking. Finally, a combination of social variables and geographical variables is described by Kinsella et al. [51]. This is another example of how language models can be exploited for modeling and integrating diverse contextual variables together.

As natural disasters heavily involve people, the user is likely to be more interested in such an event if he is connected with friends or relative involved by the event. Yom-Tov and Diaz [75] investigate how the users' information need is affected by the number of their acquaintances who may be involved by the event.

4. Contextual Search Evaluation

Once a system for contextual search is developed, it is important to evaluate it to see how it performs. Over the years, several standard data collections have been used by researchers from the information retrieval community to evaluate their work. Some such data collections as used by the researchers are presented in this section.

Agosti [2] reports some guidelines on evaluation within DLs which is a natural area where contextual search may be applied. Almeida and Almeida [3] use a company intranet repository. Anast´acio et al. [4] address semi anonymity when geographical variables are exploited. Bai et al. [14, 15] use TREC collections. Bian et al. [13] use LETOR and TREC collections.

Broder et al. [16] use corpora. Campbell et al. [19] use a company intranet repository. Campbell et al. [19] use corpora produced from company intranets. Cao et al. [20] use the ACM KDD Cup data set. Chapelle et al. [23]'s research work report interleaving as an alternative approach to collecting relevance assessments, since the conventional Cranfield-based approach to evaluation is not free of drawbacks although it is the most used and well accepted in non-contextual search.

Dai et al. [49] use "live" WWW. Dang and Croft [50] use TREC collections. Dang and Croft [50]'s work is an example of careful design and detailed implementation allow to collect many useful data about user interaction at no cost and preserving the user's privacy. Diaz [28] uses "live" WWW. Finkelstein et al. [31] use a corpus extracted from a CD-ROM. Freund et al. [32] use corpora produced from company intranets. Guo and Agichtein [34]; Harvey et al. [69] use "live" WWW.

Haveliwala [35] uses a corpus extracted from a WWW site. Hawking and Craswell [36] report on using the WWW track of TREC .gov collection. Hu et al. [75] use logs (around 20 million WWW queries collected from around 650,000 Web users). Hu et al. [76] use a corpus extracted from a WWW site. Ingwersen [37] reports on evaluation from both an information seeking and retrieval and operational point of view. Jansen and Spink [44] use logs (nine major commercial search engine anonymized and well prepared query logs); see also Jansen and Spink [85]; Jansen et al. [41, 62, 43]. Jansen [40]'s paper is a useful side-effect is the public availability of the data set. Joachims [45] uses corpora. Kelly [94] provides a complete account on some approaches to interactive IR evaluation illustrated by Ingwersen. Lau et al. [52, 53] use TREC collections. Li et al. [56, 57] illustrate an interesting approach to automatic training set construction. Liu et al. [58] use a series of small data sets that have been built with user cooperation.

Ma et al. [60] use "live" WWW. Sanderson [67] surveys the most general issues of Cranfield style-based evaluation. Shen et al. [68] use TREC collections. Spink [70] discusses the potential of user behavior and interaction variables. Teevan et al. [71] use "live" WWW. Yue and Joachims [76] use TREC collections. van Rijsbergen [73] and the publications cited in Section 1.4 are worth reading from an evaluation point of view.

5. Conclusions

Modeling and implementing context is not sufficient means for improving IR effectiveness, since additional research areas such as Economics , Cognition should be explored to help the IR researchers to better understand contextual search . If Research is done on the combining of different stat statistical methods it would be of more effective nature , the method at hand is applied to only text if it is applied for of non-textual sources of evidence and content is a great opportunity for the researcher in contextual search . That there is so much competition in "context" software indicates just how important all the giants of technology think it will be. Google, of course, has an early head start. (It also already has a close relationship with Everything.Me Contextual-based software still uses search engines to locate and deliver information, but users no longer have to interact with a search engine directly.

References:

[1] S. Abrol and L. Khan, "Twinner: Understanding news queries with geocontent using Twitter," in Proceedings of the 6th Workshop on Geographic Information Retrieval (GIR '10), New York, NY, USA, pp. 10:1–10:8, 2010.

[2] M. Agosti, "Digital libraries," in Advanced Topics in Information Retrieval, (M. Melucci and R. Baeza-Yates, eds.), Information Retrieval. Springer, 2011.

[3] R. B. Almeida and V. A. F. Almeida, "A community-aware search engine," in Proceedings of the International Conference on World Wide Web (WWW '04), New York, NY, USA, pp. 413–421, 2004.

[4] I. Anast´acio, B. Martins, and P. Calado, "Using the geographic scopes of Web documents for contextual advertising," in Proceedings of the Workshop on Geographic Information Retrieval (GIR '10), New York, NY, USA, pp. 18:1–18:8, 2010.

[5] A. Ashkan and C. L. Clarke, "Characterizing commercial intent," in Proceedings of the ACM Conference on Information and Knowledge Management (CIKM '09), New York, NY, USA, pp. 67–76, 2009.

[6] A. Ashkan, C. L. Clarke, E. Agichtein, and Q. Guo, "Classifying and characterizing query intent," in Proceedings of the European Conference *on IR Research on* Advances in Information Retrieval (ECIR '09), Berlin, Heidelberg, pp. 578–586, 2009.

[7] J. Attenberg, S. Pandey, and T. Suel, "Modeling and predicting user behavior in sponsored search," in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '09), New York, NY, USA, pp. 1067–1076, 2009.

[8] L. Azzopardi, "Incorporating context within the language modeling approach for ad hoc information retrieval," PhD thesis, University of Paisley, 2005.

[9] J. Bai, J.-Y. Nie, G. Cao, and H. Bouchard, "Using query contexts in information retrieval," in Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '07), New York, NY, USA, pp. 15–22, 2007

[10] J. Bai, D. Song, P. B. andJ. Y. Nie, and G. Cao, "Query expansion using term relationships in language models for information retrieval," in Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '05), New York, NY, USA, pp. 688–695, 2005.

[11] D. Bartholomew, F. Steele, and I. Moustaki, Analysis of Multivariate Social Science Data. Statistics in the social and behavioral sciences series. CRC Press, 2008.

[12] N. J. Belkin, R. Oddy, and H. M. Brooks, "ASK for information retrieval. Part 1: Background and theory," Journal of Documentation, vol. 38, pp. 61–71, 1982.

[13] J. Bian, T.-Y. Liu, T. Qin, and H. Zha, "Ranking with query-dependent loss for Web search," in Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '10), New York, NY, USA, pp. 141–150, 2010.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," Journal of Machine Learning and Research, vol. 3, pp. 993–1022, 2003.

[15] A. Broder, "A taxonomy of Web search," SIGIR Forum, vol. 36, pp. 3–10, September 2002.

[16] A. Broder, M. Ciaramita, M. Fontoura, E. Gabrilovich, V. Josifovski, D. Metzler, V. Murdock, and V. Plachouras, "To swing or not to swing: learning when (not) to advertise," in Proceedings of the ACM Conference on Information and Knowledge Management (CIKM '08), New York, NY, USA, pp. 1003–1012, 2008.

[17] A. Broder, M. Fontoura, E. Gabrilovich, A. Joshi, V. Josifovski, and T. Zhang, "Robust classification of rare queries using Web knowledge," in Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '07), New York, NY, USA, pp. 231–238, 2007.

[18] G. Cai, "Relevance ranking in geographical information retrieval," SIGSPATIAL Special, vol. 3, pp. 33–36, July 2011.

[19] D. R. Campbell, S. J. Culley, C. A. McMahon, and F. Sellini, "An approach for the capture of context-dependent document relationships extracted from bayesian analysis of users' interactions with information," Information Retrieval, vol. 10, pp. 115–141, 2007.

[20] H. Cao, D. H. Hu, D. Shen, D. Jiang, J.-T. Sun, E. Chen, and Q. Yang, "Context-aware query classification," in Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR'09), New York, NY, USA, pp. 3–10, 2009.

[21] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, "Context-aware query suggestion by mining click-through and session data," in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '08), New York, NY, USA, pp. 875–883, 2008.

[22] D. Chakrabarti, D. Agarwal, and V. Josifovski, "Contextual advertising by combining relevance with click feedback," in Proceedings of the International Conference on World Wide Web (WWW '08), New York, NY, USA, pp. 417–426, 2008.

[23] O. Chapelle, T. Joachims, F. Radlinski, and Y. Yue, "Large-scale validation and analysis of interleaved search evaluation," ACM Transactions on Information Systems, vol. 30, pp. 6:1–6:41, March 2012

[24] Z. Cheng, B. Gao, and T.-Y. Liu, "Actively predicting diverse search intent from user browsing behaviors," in Proceedings of the International Conference on World Wide Web (WWW '10), New York, NY, USA, pp. 221–230, 2010

[25] W. Croft and J. Lafferty, eds., Language Modeling for Information Retrieval,volume 13 of Kluwer International Series on Information Retrieval. Kluwer Academic Publishers, 2002.

[26] H. K. Dai, L. Zhao, Z. Nie, J.-R. Wen, L. Wang, and Y. Li, "Detecting online commercial intention (OCI)," in Proceedings of the International Conference on World Wide Web (WWW '06), New York, NY, USA, pp. 829–837, 2006

[27] V. Dang and B. W. Croft, "Query reformulation using anchor text," in Proceedings of the ACM International Conference on Web Search *and Data Mining (WSDM '10)*, New York, NY, USA, pp. 41–50, 2010.

[28] F. Diaz, "Integration of news content into Web results," in Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '09), New York, NY, USA, pp. 182–191, 2009.

[29] D. Downey, S. Dumais, D. Liebling, and E. Horvitz, "Understanding the relationship between searchers' queries and information goals," in Proceedings of the ACM Conference on Information and Knowledge Management (CIKM '08), New York, NY, USA, pp. 449–458, 2008.

[30] E. Efthimiadis, "Query expansion," in Annual Review of Information Science and Technology (ARIST), vol. 31, chap. 4, (M. Williams, ed.), Medford, NJ, pp. 121–185, 1996.

[31] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G.Wolfman, and E. Ruppin, "Placing search in context: the concept revisited," ACM Transactions on Information Systems, vol. 20, pp. 116–131, 2002.

[32] L. Freund, E. G. Toms, and C. L. Clarke, "Modeling task-genre relationships for IR in the workplace," in Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '05), New York, NY, USA, pp. 441–448, 2005. [33] V. Ganti, A. C. K¨onig, and X. Li, "Precomputing search features for fast and accurate query classification," in Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '10), New York, NY, USA, pp. 61–70, 2010

[34] Q. Guo and E. Agichtein, “Towards predicting Web searcher gaze position from mouse movements,” in Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA, pp. 3601–3606, 2010.

[35] T. H. Haveliwala, “Topic-sensitive PageRank,” in Proceedings of the International Conference on World Wide Web (WWW ’02), New York, NY, USA, pp. 517–526, 2002

[36] D. Hawking and N. Craswell, “The very large collection and Web tracks,”in TREC: Experiment and Evaluation in Information Retrieval, chap. 9,(E. Voorhees and D. Harman, eds.), MIT Press, 2005.

[37] P. Ingwersen, “The user in interactive information retrieval evaluation,” in Advanced Topics in Information Retrieval, (M. Melucci and R. Baeza-Yates, eds.), Springer, 2011.

[38] W. H. Inmon, Building the Data Warehouse. Wiley, 1996.

[39] W. H. Inmon, “The data warehouse and data mining,” Communications on ACM, vol. 39, pp. 49–50, November 1996.

[40] B. J. Jansen, Instructions for Obtaining Search Engine Transaction Logs. 2011. URL http://faculty.ist.psu.edu/jjansen/academic/transaction logs.html.

[41] B. J. Jansen, D. L. Booth, and A. Spink, “Determining the user intent of Web search engine queries,” in Proceedings of the International Conference on World Wide Web (WWW ’07), New York, NY, USA, pp. 1149–1150, 2007.

[42] B. J. Jansen, D. L. Booth, and A. Spink, “Determining the informational, navigational, and transactional intent of Web queries,” Information Processing and Management, vol. 44, pp. 1251–1266, May 2008

[43] B. J. Jansen, D. L. Booth, and A. Spink, “Patterns of query reformulation during Web searching,” Journal of the American Society for Information Science and Technology, vol. 60, pp. 1358–1371, July 2009.

[44] B. J. Jansen and A. Spink, “How are we searching the world wide web?:A comparison of nine search engine transaction logs,” Information Processing and Management, vol. 42, pp. 248–263, January 2006.

[45] T. Joachims, “Optimizing search engines using clickthrough data,” in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD ’02), New York, NY, USA, pp. 133–142, 2002.

[46] C. B. Jones and R. S. Purves, “Geographical information retrieval,” International Journal of Geographical Information Science, vol. 22, pp. 219–228, 2008.

[47] R. Jones, B. Rey, O. Madani, and W. Greiner, “Generating query substitutions,” in Proceedings *of the International Conference on World Wide Web(WWW ’06)*, New York, NY, USA, pp. 387–396, 2006.

[48] D. Kelly, “Measuring online information seeking context. Part 1: Background and method,” Journal of the American Society in Information Science and Technology, vol. 57, no. 13, pp. 1729–1739, 2006.

[49] D. Kelly, “Measuring online information seeking context. Part 2: Findings and discussion,” Journal of the American Society in Information Science and Technology, vol. 57, no. 13, pp. 1862–1874, 2006.

[50] D. Kelly, “Methods for evaluating interactive information retrieval systems with users,” Foundations and Trends in Information Retrieval, vol. 3, pp. 1–224, 2009.

[51] S. Kinsella, V. Murdock, and N. O’Hare, ““I’m eating a sandwich in glasgow”:modeling locations with tweets,” in Proceedings of the International Workshop on Search and Mining User-generated Contents (SMUC ’11), New York, NY, USA, pp. 61–68, 2011

[52] R. Y. Lau, P. D. Bruza, and D. Song, “Belief revision for adaptive information retrieval,” in Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR ’04), New York, NY, USA, pp. 130–137, 2004.

[53] R. Y. K. Lau, P. D. Bruza, and D. Song, “Towards a belief-revision-based adaptive and context-sensitive information retrieval system,” ACM Transactions on Information and Systems, vol. 26, no. 2, pp. 1–38, 2008

[54] U. Lee, Z. Liu, and J. Cho, “Automatic identification of user goals in Web search,” in Proceedings of the International Conference on World Wide Web (WWW ’05), New York, NY, USA, pp. 391–400, 2005.

[55] T. Li, N. Liu, J. Yan, G. Wang, F. Bai, and Z. Chen, “A Markov chain model for integrating behavioral targeting into contextual advertising,” in Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD ’09), New York, NY, USA, pp. 1–9, 2009.

[56] X. Li, Y.-Y.Wang, and A. Acero, “Learning query intent from regularized click graphs,” in Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR ’08), New York, NY, USA, pp. 339–346, 2008

[57] X. Li, Y.-Y. Wang, D. Shen, and A. Acero, “Learning with click graph for query intent classification,” ACM Transactions on Information and Systems, vol. 28, pp. 12:1–12:20, July 2010.

[58] F. Liu, C. Yu, and W. Meng, “Personalized Web search by mapping user queries to categories,” in Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM ’02), New York, NY, USA, pp. 558–565, 2002.

[59] J. Luxenburger, S. Elbassuoni, and G. Weikum, "Matching task profiles and user needs in personalizedWeb search," in Proceedings of the ACM Conference on Information and Knowledge Management (CIKM '08), New York, NY, USA, pp. 689–698, 2008.

[60] Z. Ma, G. Pant, and O. R. L. Sheng, "Interest-based personalized search," ACM Transactions on Information and Systems, vol. 25, p. 5, 2007.

[61] M. Melucci, "An evaluation of automatically constructed hypertexts for information retrieval," Journal of Information Retrieval, vol. 1, pp. 57–80, 1999.

[62] M. Melucci and R. W. White, "Utilizing a geometry of context for enhanced implicit feedback," in Proceedings of the ACM Conference on Conference on Information and Knowledge Management (CIKM '07), New York, NY, USA,pp. 273–282, 2007.

[63] S. Mizzaro, "Relevance: The whole history," Journal of the American Society for Information Science, vol. 48, pp. 810–832, 1997.

[64] J. Pitkow, H. Sch¨utze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel, "Personalized search," Communications of ACM, vol. 45, pp. 50–55, September 2002.

[65] T. Reichenbacher, "Geographic relevance in mobile services," in Proceedings of the International Workshop on Location and the Web (LOCWEB '09), New York, NY, USA, pp. 10:1–10:4, 2009.

[66] T. Reichenbacher and S. De Sabbata, "Geographic relevance: different notions of geographies and relevancies," SIGSPATIAL Special, vol. 3, pp. 67–70, July 2011.

[67] M. Sanderson, "Test collection based evaluation of information retrieval systems," Foundations and Trends in Information Retrieval, vol. 4, 2010.

[68] X. Shen, B. Tan, and C. Zhai, "Implicit user modeling for personalized search,"in Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '05), New York, NY, USA, pp. 824–831, 2005.

[69] D. Sontag, K. Collins-Thompson, P. N. Bennett, R. W. White, S. Dumais,and B. Billerbeck, "Probabilistic models for personalizing Web search," in Proceedings of the ACM International *Conference on Web Search and Data Mining (WSDM '12)*, New York, NY, USA, pp. 433–442, 2012.

[70] A. Spink, Information Behavior: An Evolutionary Instinct. Springer, 1st ed., 2010.

[71] J. Teevan, S. Dumais, and E. Horvitz, "Personalizing search via automated analysis of interests and activities," in Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR '05), New York, NY, USA: ACM Press, pp. 449–456, 2005.

[72] J. Teevan, D. J. Liebling, and G. Ravichandran Geetha, "Understanding and predicting personal navigation," in Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '11), New York, NY, USA, pp. 85–94, 2011.

[73] C. van Rijsbergen, Information Retrieval, chapter 6, pp. 144–183. London:Butterworths, second ed., 1979.

[74] X. Yi, H. Raghavan, and C. Leggetter, "Discovering users' specific geo intention in Web search," in Proceedings of the International Conference on World Wide Web (WWW '09), New York, NY, USA, pp. 481–490, 2009.

[75] E. Yom-Tov and F. Diaz, "Out of sight, not out of mind: On the effect of social and physical detachment on information need," in Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR '11), New York, NY, USA, pp. 385–394, 2011.

[76] Y. Yue and T. Joachims, "Predicting diverse subsets using structural svms," in Proceedings of the International Conference on Machine Learning (ICML'08), New York, NY, USA, pp. 1224–1231, 2008.

[77] Y. Yue and T. Joachims, "Interactively optimizing information retrieval systems as a dueling bandits problem," in Proceedings of the Annual International Conference on Machine Learning (ICML '09), New York, NY, USA, pp. 1201–1208, 2009.

[78] Y. Zhou and W. B. Croft, "Document quality models for Web ad-hoc retrieval," in Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '05), New York, NY, USA, pp. 331–332, 2005.

[79] G. Zhu and G. Mishne, "Mining rich session context to improve Web search,"in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '09), New York, NY, USA, pp. 1037–1046, 2009.

.

[80] G. Zhu and G. Mishne, "Clickrank: Learning session-sontext models to enrich Web search ranking," ACM Transactions on Web, vol. 6, pp. 1:1–1:22, March 2012.

[81] Z. Zhuang, C. Brunk, and C. L. Giles, "Modeling and visualizing geo-sensitive queries based on user clicks," in Proceedings of the International Workshop on Location and the Web (LOCWEB '08), New York, NY, USA, pp. 73–76, 2008.