Contextual Search: A Computational Framework

Contextual Search: A Computational Framework

Massimo Melucci

Università di Padova
Italy
massimo.melucci@unipd.it

now
the essence of knowledge
Boston – Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is M. Melucci, Contextual Search: A Computational Framework, Foundation and Trends® in Information Retrieval, vol 6, nos 4–5, pp 257–405, 2012

ISBN: 978-1-60198-600-9

Foundations and Trends® in Information Retrieval
Volume 6 Issues 4–5, 2012

Editorial Board

Editorial Scope

Foundations and Trends® in Information Retrieval will publish survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians
Foundations and Trends® in Information Retrieval, 2012, Volume 6, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Foundations and Trends® in
Information Retrieval
Vol. 6, Nos. 4–5 (2012) 257–405

DOI: 10.1561/1500000023

now
the essence of knowledge

Contextual Search: A Computational Framework

Massimo Melucci

Università di Padova, Dipartimento di Ingegneria dell'Informazione, Via G. Gradenigo, 6, Padova, 35131, Italy, massimo.melucci@unipd.it

Abstract

The growing availability of data in electronic form, the expansion of the World Wide Web (WWW) and the accessibility of computational methods for large-scale data processing have allowed researchers in Information Retrieval (IR) to design systems which can effectively and efficiently constrain search within the boundaries given by context, thus transforming classical search into contextual search. Because of the constraints imposed by context, contextual search better focuses on the user's relevance and improves retrieval performance, since the out-of-context aspects of the search carried out by users that are likely linked to irrelevant documents are left apart.

This survey introduces contextual search within a computational framework based on contextual variables, contextual factors and statistical models. The framework adopted in this survey considers the data observable from the real world entities participating in contextual search and classifies them as what we call contextual variables. The contextual variables considered are content, geotemporal, interaction, and social variables. Moreover, we distinguish between

contextual variables and contextual factor: the former is what can be observed, the latter is what cannot be observed, yet this is the factor affecting the user's relevance assessment. Therefore, in this survey, we describe how statistical models can process contextual variables to infer the contextual factors underlying the current search context.

In this survey we provide a background to the subject by: placing it among other surveys on relevance, interaction, context, and behavior; providing the description of the contextual variables used for implementing the statistical models which represent and predict relevance and contextual factors; citing and surveying useful publications to the reader for further examination; providing an overview of the evaluation methodologies and findings relevant to this subject; and briefly describing some implementations of contextual search tools.

Contents

1

Introduction

> Context: from Latin *contextus*, where *con* stands for "together" and *texere* stands for "to weave".
>
> Oxford Dictionary

1.1 Motivation of this Survey

Many researchers with various backgrounds believe that context can enhance the user's experience and improve the system's effectiveness of search. In so doing, they frame Information Retrieval (IR) within the more general notion of contextual search, although from differing viewpoints. The different perspectives at which context has been viewed have led to definitions of context with different potential of implementation.

At one extreme, 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. From this perspective, some publications relevant to contextual search are being written mostly from an information seeking and retrieval point of view. 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 of something written or spoken that immediately precede and follow a word or passage and clarify its meaning.

These two extreme perspectives clearly differ in their potential of implementation. While the definition of context in terms of events and settings cannot obviously be utilized for designing algorithms and data structures, 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. However, this view is quite reductive and considers only one of the ways context occurs.

From one extreme to the other, computational approaches to contextual search followed one another, ranging from sophisticated and computationally expensive approaches to more simple and efficient ones, but each one of them has been useful for writing this survey.

1.2 Definitions and Scope of the Survey

A *variable* is any observable value that is liable to change. Variables can be: qualitative or quantitative; ordinal or not; if ordinal, cardinal or not; if cardinal, integer or not; and so on. When the variables are random, they change according to a probability distribution in such a way that its observation value occurs with a given probability. Such a characterization allows inference to be made on estimation and prediction of potential relationships between variables in such a way that the variation of some independent variables determines the variation of some dependent variables.

A *contextual factor* is any unobservable circumstance or fact of search such as query intent, personal interest and document quality, which affects relevance. We concentrate on three contextual factors: query intent, personal interest, and document quality. (The contextual factors are illustrated in Sections 2, 3, and 4.)

Query intent refers to the objectives of the user who issued the query. In this situation, a query is viewed as a means to accomplish a task such as "dissertation writing," "finding a resource," "bibliography compilation" and a query intent is an objective to be achieved in order

to accomplish the task. Intent is a property of a query and is not necessarily tied to a user in the way a personal interest is.

Personal interest in general refers to the user's state of wanting to know or learn about a thing, a person or an event. An interest is an information need that has the quality of sparking curiosity or holding the user's attention, and may be viewed as a property of an information need that makes the information need crucial to the user.

Document quality refers to the property of a document that is able to be trusted as being up-to-date, authoritative, exhaustive, accurate, reliable, and clear. A high-quality document is considered to be the best of its kind and unlikely to be improved upon.

A *contextual variable* is any set of variables dependent on contextual factors. This survey classifies the contextual variables observed by the real world entities participating in contextual search as: content variables, interaction variables, geographical variables, and social variables. These contextual variables are introduced in Section 1.3.

A *statistical model* is a a set of computable mathematical rules defined over a set of variables or factors for example height and weight are related in a way that they can be plotted as points along a straight line, or the frequency of a term within a document and relevance assessment are related in a way that the higher the frequency, the more likely the document is relevant. In this survey, the rules of a statistical model express computable relationships between contextual variables and contextual factors.

The denotation of context in this survey is thus essentially computational and allows us to introduce a computational framework of contextual search summarized by the following

---

Definition 1.1. *Context* is represented as a set of contextual variables and contextual factors weaved together by statistical models of estimation and prediction.

An objective of this survey is to inform the reader which contextual variables, contextual factors, and statistical models have been utilized in the literature to represent context by means of a

computer system and therefore which of these provide some useful hints about what to use to extend a traditional IR system toward context.

The definitions of contextual variable and contextual factor point out two main aspects: observability and dependence. Observability is a necessary condition of the meaning of context in this survey, that is, context can be operationalized as variables and can therefore be exploited in search only if some variables can be defined and observed; context implementations not referred to as variables are not considered in this survey. A basic example is standard IR: index term occurrence is a contextual variable observed from the document content, in contrast, aboutness is a contextual factor affecting relevance and cannot directly be observed.

Dependence is between contextual factors, relevance, and variables. Research in IR often assumes that a variation of contextual factors reflects upon a variation of the contextual variables. The relationship posited between personal interest and term frequency is an example of the relationship between a contextual variable and a contextual factor; term frequency may increase in a document if this document becomes interesting for a user. Therefore, if some variations of the contextual variables are observed, a variation of the contextual factor is likely to have occurred; for example, if term frequency is higher in a document than in another document, the former is more likely interesting than the latter.

In the computational framework presented in this survey, attention is also paid to discovering the contextual factors that affect relevance assessments. Thus, a variation of relevance assessment is due to a variation of the contextual factors; for example, if query intent is viewed as contextual factor and term frequency is a contextual variable, a variation of frequency may result from a variation of query intent which in turn affects relevance.

Another feature of this survey is the attention paid to statistical models. The statistical models mentioned in this survey provide some advantages in illustrating context. They allow researchers to implement context because these models are suitable for estimating and predicting context starting from the variables observed in objects. Another

advantage is of a computational nature. Most of the statistical models scale up when the size of data available from the user's environment, social network, and personal dimension increases by hundred-fold, thus keeping high levels of effectiveness and efficiency. Everyone prefers computationally efficient approaches to search in context; however, computational efficiency is not a feature of every one of the approaches (e.g., those arising from Artificial Intelligence), despite the computational potential that makes contextual variable implementation more unbridged than in the past.

A computational framework for contextual search like that described in this survey may resemble classical modeling in noncontextual search in which the "best" model is selected for optimizing effectiveness. We think that this approach is not constrictive since in the past there have definitely been statistical models for contextual search that resulted in significant improvements in IR systems. We do not claim that a computational approach is the only approach to explaining contextual search, but we do claim that it is the best approach for making contextual variables useable; relevance feedback is an example of a computational approach to contextual search thoroughly investigated in the past.

Although (or maybe precisely because) investigated and employed for a long time, relevance feedback is still crucial in contextual search, since it mainly relies on content variables and in particular on document content. Indeed, relevance feedback has recently been reevaluated and experimented with huge test collections and very short noisy queries through initiatives such as the relevance feedback track of TREC. However, despite it being relevant, explicit or pseudo-relevance feedback is not addressed in this survey because our focus is on recent developments of contextual search while there are already surveys of relevance feedback and query expansion.

In contrast, implicit relevance feedback is the backbone of the incorporation of behavior in contextual search. The research conducted within implicit relevance feedback has aimed to use the contextual information generated during the interaction between the user and information as implicit evidence of relevance. Hence, a key question is whether implicit relevance feedback can effectively be used in

contextual search systems in comparison with traditional content-based ranking functions or more advanced yet well experimented methods such as anchor text or link analysis algorithms.

As it happens, contextual search is not relevant only to IR, of course, but to other research areas too with which they interact yet they seem rather distant from IR. Examples are Psychology, Mobile Communication, Electronic Commerce, Nomadic Computing, Human Computer Interaction. All these subjects are relevant to this survey although they cannot be looked at thoroughly because the topic of this survey is already vast enough. A few things that are on the side of the context of a document and are not the primary focus are: temporal context (e.g., two e-mail messages sent right after the same event); storage context (e.g., two documents found in the same file system folder); conversational context (e.g., one e-mail message is a reply to another).

1.3 Contextual Variables

This survey considers four types of contextual variables: content, geographical, interaction, and social variables.

Content variables refer to the informative content and relationships of queries and documents. The data are content features observed from text, image, video, audio; link anchors; layout; genre; lexical properties (e.g., part-of-speech tags); user's tags (e.g., image tags or file names); category labels (e.g., Wikipedia category labels); demographic labels (e.g., authorship) and anything used to describe informative contents or to enrich information need representations.

Geographical variables are any variable with the state of existing within or having some relationship with space location. Examples are geographical names added to documents or queries, digital photographs tagged with geographical coordinates, typically the latitude and longitude of the space location perhaps associated to a user.

Interaction variables are observed over time during the interaction between users and IR systems. (Geographical variables are not necessarily referred to a user.) These variables are for example: click-through data; data about queries or search sessions; user

judgments or assessments; user behavior data (e.g., document retention, display time, eye or mouse movements).

Social variables refer to user communities or groups and are observed for example from: "tweets"; social connections (e.g., friendship); hyperlinks (e.g., a link between two WWW pages).

1.3.1 Content Variables

Content is a contextual variable exploited in contextual search to decide whether an additional or special action that is different from time to time should be performed by an IR system when the user is interacting with the informative content managed by the system for meeting his information needs. Content variables can be observed from the documents of a collection, search engine result pages, queries, or from parts of them such as windows, fragments, and passages.

The main *medium* addressed in the literature of this survey is text. It is perhaps the richest source of evidence for predicting context since text is an expression of natural language, that is, the main means used by humans to communicate information and needs. Text can easily be managed because words or terms can be suggested to the user who in turn can understand them by leveraging common cognitive abilities and feed data back into the system: positive words represent what items the user would like to retrieve; negative words indicate what the user does not want; neutral words are not good indicators of her information needs. We are not dealing with multi-lingual text IR; the literature utilized in this survey refers to the English language only.

1.3.2 Geographical Variables

In our view, geographical variables are observed and are instrumental for detecting contextual factors such as query intent and personal interests. Geographical variables differ from other variables due to the intents underlying the queries referring to geographical information. However, they raise issues similar to the issues raised by natural language processing. In particular, when geographical variables are names, the issues are: name or reference detection,

name disambiguation, name clustering, linking or association, name weighting, and document ranking.

Our use of geographical variables complements the view of geography as a relevance factor; for example, Raper [143] defines geographical information needs based on cognitive and geographic criteria and argues that geographic relevance is best defined as a spatio-temporally extended relation between geographic information needs and geographic documents.

1.3.3 Interaction Variables

Whenever users have difficulty in expressing their information needs, contextual variables based on interaction are precious because an IR system can be enabled to automatically deduce a user's interest based on the data generated during the interactions with the system. Indeed, the data observed over time during the interaction of the user with a contextual search system form an interaction history where history also means "finding out." Thus, the value of interaction variables is not only the individual pieces of content, but their organization within a coherent stream of data — it is the observation of these pieces together which makes history valuable; for example, if the user has requested some documents recently, it is likely that the user is in a given context, and the retrieved documents can form the basis of supervised learning for the user's preferences because of recency and not only because of the amount of data.

When the interaction data employed for estimation are very close in time to the user's actions, the estimated models are more closely related to the user than the models that would be estimated with the farthest data in time. The data employed for estimation are very close in time to the user's actions when, for instance, it is of interest to the contextual search system to recognize the correct query sense or intent of the user. On the other hand, implicit relevance feedback information collected over a long period of time is less likely to be very useful for predicting the individual's interests than the immediate search context and feedback information; they may be useful for predicting the interests of a group.

1.3.4 Social Variables

What a user either directly or indirectly learns from or teaches to the communities is a crucial contextual variable because human relationships are at the basis of many conditions. Hence, it is not surprising that quite an appreciable proportion of the literature on contextual search addresses the issues of the social dimensions of the users. Community is meant in a broad sense and is not confined to social networks or similar user organizations; the same algorithms can make the social variables observed from any user community a useable source of evidence for contextual search.

In this survey, we consider some cases of community behaviors where members participate in a collective activity and unwittingly collaborate to build collective knowledge. These kinds of community behaviors differ in the degree to which a member is aware of belonging to a community. For example, the users tagging resources are often aware of their membership to a community (e.g., they log into a system) whereas the users clicking ads are not aware that their clicks are collected and exploited for boosting ad ranking. What the various kinds of community behaviors have in common is that they leverage the (large) size of the communities involved in a way that the large quantity of observed data can be exploited to estimate parameters and discover patterns useful for implementing contextual search.

In IR and related disciplines there have already been research works that to a certain extent investigate how members of a community interact, perhaps indirectly, to building knowledge that is further exploited by the community (the link analysis methods addressed in this section are an example and the earlier bibliometrics is another notable example).

Numerous papers addressing social variables as contextual variables are based on link analysis algorithms since a graph is a natural way to represent a community; nodes are members and edges are relationships between members. However, in this survey, we not only address link analysis but also address other statistical models suited to mining useful information from community contextual variables. To this end, we are drawing the reader's attention to a couple of social variables,

that is, "tweets" and tags, which can efficiently provide useful information about the context on a large scale.

Another research area is known as digital annotation systems. Annotations affixed to digital documents is a little more recent than bibliometrics because the use and production of the digital documents has grown since the 1980s at the earliest; Agosti et al. [5] introduced digital annotation systems. However, since their advent their use is still limited, thus making the exploitation of these data for contextual search through statistical methods difficult. We focus on two types of annotation (ESP games and "tweets") that in contrast to "traditional" annotation affixed to digital documents stimulate the implementation of large scale statistical methods for contextual search.

1.4 Historical Background

In this section we provide a background to contextual search by: placing the subject among other surveys on relevance, interaction, context, and behavior; citing and surveying useful publications to the reader for further examination.

Before the relatively renowned and growing interest in contextual search viewed in the recent literature of IR, context had been on the scene, or perhaps better stated behind the scenes, for many years (perhaps for decades) as the IR literature since the 1970s shows. As the literature is by now quite vast, we can distill only some aspects and issues and cannot be more exhaustive than the publications already available on this topic.

This section is then devoted to providing a summary of and the references to the publications in which contextual search has been thoroughly considered. These publications may provide the reader with complementary information, and give a background to this survey. In particular, this section draws the reader's attention to the papers by Belkin et al. [18]; Ingwersen and Järvelin [78]; Mizzaro [131]; Ruthven [149]; Saracevic [152]; Spink [159].

1.4.1 Relevance

Because users of an IR system assess whether a document is relevant in a context, context has been a crucial aspect of relevance for decades.

Hence, relevance is intrinsically dependent on context. Due to the complexity of context and relevance, the most common IR models are a mere simplification of the reality in which users are called on to assess the relevance of documents to their information needs.

If the items of a context are gathered together, a sort of relation is obtained; actually, a mathematical relation as it is intended by a DBMS. 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 review, 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 [131], context "includes everything not pertaining to topic and task, but however affecting the way the search takes place and the evaluation of results." This definition suggests the view that the user has some context that is not stated in the query but which we could nonetheless model. Mizzaro's paper also cites literature relevant to context introduced as a factor, component or container of the content, user, task, and so on.

1.4.2 Anomalous State of Knowledge

The Anomalous State of Knowledge (ASK) by Belkin et al. is another useful element for understanding contextual search. The first part of the paper reported by Belkin et al. [18] introduces the ASK hypothesis stating "that an information need arises from a recognized anomaly in the user's state of knowledge concerning some topic or situation and that, in general, the user is unable to specify precisely what is needed to resolve that anomaly" [18, p. 62]; the second part reported by Belkin et al. [19] describes an experiment. The information need of the ASK hypothesis stems from a "topic or situation" which might better be named as problematic situation or task. In the words of Belkin et al., "the user, faced with a problem, recognizes that her/his state of

knowledge is inadequate for resolving that problem, and decides that obtaining information about the problem area and its circumstances is an appropriate means toward its resolution."

When the ASK hypothesis is valid, the user is unable to make his information need explicit because what he would be asked to say is precisely what he does not know. A consequence of this impossibility which is relevant to this survey is that, to address the ASK, an IR system should be interactive and iterative, thus calling into play various contextual factors such as query intents, personal interests, and document qualities. Sometimes, the combination of different contextual variables leads to concept networks. Belkin et al. [18, p. 68] defined concept networks as networks of inter-related documents and named them as "formal context." Such a network becomes a description of context and at the same time a source of evidence from which data can be observed to represent context. Networks of concepts have been further elaborated in Agosti et al.[4] within the most naturally interactive system, that is, hypermedia systems.

1.4.3 Interactive Information Retrieval

Ingwersen and Järvelin [78] introduced the Integrated Cognitive Research Framework for IR. The components of this framework are: information objects (e.g., documents); the IT component (e.g., search engines); the interface (e.g., WWW clients); the cognitive actor (e.g., the user); the socio-cultural and organizational context (e.g., the workplace or the community). Between the components, which are depicted in Figure 1.1, there are influence or exchange relationships depicted as unidirectional and bidirectional arrows, respectively, and there are solid or dashed unidirectional arrows corresponding to influence and influence over time, respectively. Within the Integrated Cognitive Research Framework for IR, the definition of context suggested in Ingwersen and Järvelin [78] becomes: "in information seeking and retrieval actors and objects [are] associated with each component of the cognitive information seeking and retrieval framework function as context for their own elementary cognitive structures (intra-object context), as context to one another (inter-object context), and in context of

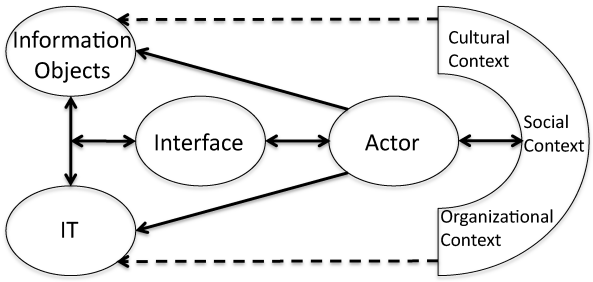

Fig. 1.1 Ingwersen and Järvelin [78]'s Integrated Cognitive Research Framework for IR.

the interaction processes between framework components, which themselves are contextual to each other. In the latter case one may talk about social/organization/cultural as well as systemic contexts. The context of interactive IR processes ranges from algorithmic IR processes in context of interactive IR as well as information seeking processes to information behavior. All information seeking and retrieval components and activities are in context of common social, physical and technological infra-structures as well as their history over time."

As it happens, circumscribing a notion like that of context to something simpler and perhaps simplicistic makes its implementation easier or more understandable than general or perhaps vague definitions. An example is a user interface-oriented notion of context which would help visualize the different components of Ingwersen and Järvelin's framework. Ruthven [148] gives a user interface-oriented notion and states that "Our ideas on context (from both a soft and hard laboratory perspective) often manifest themselves at the interface." Lalmas [106] adopted this definition.

Ruthven [149] some years later provides another definition: context is "a complex set of variables describing our intentions, our personal characteristics, the data and systems available for searching, and our physical, social and organizational environments" or it is also thought as the fact that "personal [context] information can cover any information that we have experienced (such as webpages we have visited), information that we have received (such as email) or information that we have created (such as documents or images). [...] [T]he range of contextual factors that might be important is vast ranging from age, physical

and cognitive ability (which may require altering the presentation of search results as well as the selection of results), learning styles, education level, or mood of searcher. The most common personal context investigated so far is the searchers topical search interests, particularly through applications of information filtering."

To personal context, we may add social context, which is somehow related yet independent of other contexts since it is about "how people use systems and for what purposes. We can mine this information — the context of use — for many purposes including filtering information to obtain better search results" according to Ruthven [149]. From this point of view task is the information problem, for example, finding a holiday destination, writing an essay, giving a lecture, which is the reason why the user expresses his information need through queries, browsing, clicking, etc. Thus, task context covers any information that describes the user's problem and that makes relevance, usefulness or authoritativeness of documents dependent on the task, with all the other variables being equal.

Space–time reality is perhaps the most intuitive and common setting where we experience context. Thus, it is quite straightforward to define physical context as the container of important data for providing situationally relevant information (e.g., GPS coordinates or time). Similarly, environmental context relates to any information about the type of location where the user's search takes place (e.g., whether the user is in a public place, the weather is nice, the roads are congested) according to Ruthven [149].

Contextual search can barely be separated not only from IR and information seeking and retrieval but also from the notion of human information behavior defined by Spink [159] as follows: human information behavior "refers to a wide range of processes which people employ when engaged with information and to related cognitive and social states and effects". In a sense, human information behavior studies are orthogonal to ASK, IR and information seeking and retrieval since they aim at understanding how and why the users interact with information when this information is contained in documents or queries. Spink in particular is interested in the user's behavior during the formulation of the ASK. She defines information seeking and retrieval

as "one sub-process within human information behavior that includes the purposive seeking of information in relation to" an ASK because information seeking and retrieval starts when an ASK has been recognized, continues when relevance assessments have been observed and ends when the ASK has been solved. From this description it is then clear that information seeking and retrieval is as highly dependent on context as human information behavior is. The remarks made by Spink [159] about human information behavior within communities and the personal dimension are relevant to this survey.

1.5 Concluding Remarks and Suggestions

The computational framework underlying contextual factors, contextual variables, and statistical models is the main conceptual contribution of this survey. Other researchers are allowed to place other contextual factors, contextual variables and statistical models in this framework, thus preserving the overall consistency of the illustration of contextual search proposed in this survey. Some results illustrated in the remaining sections may well be placed in more than one contextual factor (e.g., understanding the intent of a given user may be placed in Section 2, in Section 3, or both). However, these decisions are a matter for the researchers implementing this framework. Appendix A briefly illustrates some prototypes of contextual search tools.

We conclude by giving some bibliographic references relevant to the computational framework introduced in this survey and to the general notion of contextual search. Alpaydin [7] describes support vector machines. Alpaydin [7] is a reference on machine learning. Azzopardi [12] 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. [14, 15] are examples of text window-based context papers with co-occurrence analysis, an interesting modeling of contextual factors based on language models and an analysis of domain knowledge and language model combination. Bartholomew et al. [17] provide a perspective of the factorial models that are relevant to the notion of computational

framework used in this survey. Bian et al. [23] are worth reading as for the Expectation-Maximization algorithm. Blei et al. [24]'s is the original publication on latent Dirichlet allocations. The notion of geographical variable is discussed by, for example, Cai [33]. The remarks made by Chakrabarti et al. [41] 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 [47] survey language models for IR. The study by Efthimiadis [57] describes query expansion whereas the paper by Carpineto and Romano [38] is an up-to-date survey of this topic. Feller [60]; Levinson et al. [113]; Rabiner and Juang [140] are some reference publications on Bayes' rule, Markov chains and hidden Markov models. Halmos [68] explains Singular Value Decomposition and in general vector spaces. The paper written by Hu et al. [75] is easy to read and has a computational flavor. As for interaction variables, the reader may want to spend some time reading Inmon [79, 80] who introduced the notion of time-variancy, since click-through datasets may be viewed as an instance of data warehouses. The special journal publication edited by Jones and Purves [90] is a useful reference on the issues of geographical variables. The papers on implicit relevance feedback by Kelly and Belkin [95, 96]; Kelly and Fu [97]; Kelly et al. [98]; Kelly [92, 93, 94] are definitely worth reading. The survey by Lalmas and Ruthven [107] provides a precise, recent and exhaustive account of relevance feedback. Lau et al. [108] address context at difference abstraction levels, from the conceptual, to the logical up to the statistical level. Lau et al. [109] present an interesting application of their theoretical framework and show that the vector space model is still a good baseline for search in context. Metzler and Croft [130] illustrate conditional random fields. Ponte and Croft [138] introduce language models for IR. The notion of geographical variable is also discussed by Reichenbacher [144]; Reichenbacher and De Sabbata [145]. The paper written by Shannon [155] is the reference for entropy. The papers on exploratory search by White et al. [172]; White and Kelly [173]; White et al. [170, 171]; White and Roth [174]; White [169] are also useful reading.

References

[1] S. Abrol and L. Khan, "Twinner: Understanding news queries with geo-content using Twitter," in *Proceedings of the 6th Workshop on Geographic Information Retrieval (GIR '10)*, New York, NY, USA, pp. 10:1–10:8, 2010. ISBN 978-1-60558-826-1. doi: 10.1145/1722080.1722093. URL http://doi.acm.org/10.1145/1722080.1722093.

[2] E. Agichtein, E. Brill, and S. Dumais, "Improving Web search ranking by incorporating user behavior information," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '06)*, New York, NY, USA, pp. 19–26, 2006. ISBN 1-59593-369-7. doi: http://doi.acm.org/10.1145/1148170.1148177. URL http://doi.acm.org/10.1145/1148170.1148177.

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

[4] M. Agosti, R. Colotti, and G. Gradenigo, "A two-level hypertext retrieval model for legal data," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '91)*, New York, NY, USA, pp. 316–325, 1991. ISBN 0-89791-448-1. doi: http://doi.acm.org/10.1145/122860.122892.

[5] M. Agosti, N. Ferro, E. Panizzi, and R. Trinchese, "Annotation as a support to user interaction for content enhancement in digital libraries," in *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '06)*, New York, NY, USA, pp. 151–154, 2006. ISBN 1-59593-353-0. doi: http://doi.acm.org/10.1145/1133265.1133296. URL http://doi.acm.org/10.1145/1133265.1133296.

[6] 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. ISBN 1-58113-844-X. doi: http://doi.acm.org/10.1145/988672.988728.

[7] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2004.

[8] I. Anastácio, 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. ISBN 978-1-60558-826-1. doi: 10.1145/1722080.1722103. URL http://doi.acm.org/10.1145/1722080.1722103.

[9] 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. ISBN 978-1-60558-512-3. URL http://doi.acm.org/10.1145/1645953.1645965.

[10] 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. ISBN 978-3-642-00957-0. URL http://dx.doi.org/10.1007/978-3-642-00958-7_53.

[11] 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. ISBN 978-1-60558-495-9. doi: http://doi.acm.org/10.1145/1557019.1557135.

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

[13] L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak, "Spatial variation in search engine queries," in *Proceedings of the International Conference on World Wide Web (WWW '08)*, New York, NY, USA, pp. 357–366, 2008. ISBN 978-1-60558-085-2. doi: 10.1145/1367497.1367546. URL http://doi.acm.org/10.1145/1367497.1367546.

[14] 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. ISBN 978-1-59593-597-7. doi: http://doi.acm.org/10.1145/1277741.1277747.

[15] 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. ISBN 1-59593-140-6. doi: http://doi.acm.org/10.1145/1099554.1099725.

[16] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999. doi: 10.1126/science.286.5439.509. URL http://www.sciencemag.org/content/286/5439/509.abstract.

[17] D. Bartholomew, F. Steele, and I. Moustaki, *Analysis of Multivariate Social Science Data*. Statistics in the social and behavioral sciences series. CRC Press, 2008. ISBN 9781584889601.

[18] 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. ISSN 0022-0418.

[19] N. J. Belkin, R. Oddy, and H. M. Brooks, "ASK for information retrieval. Part 2: Results of a design study," *Journal of Documentation*, vol. 38, pp. 145–164, 1982. ISSN 0022-0418.

[20] M. Bendersky, W. B. Croft, and Y. Diao, "Quality-biased ranking of Web documents," in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '11)*, New York, NY, USA, pp. 95–104, 2011. ISBN 978-1-4503-0493-1. doi: 10.1145/1935826.1935849. URL http://doi.acm.org/10.1145/1935826.1935849.

[21] P. N. Bennett, F. Radlinski, R. W. White, and E. Yilmaz, "Inferring and using location metadata to personalize Web search," in *Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR '11)*, New York, NY, USA, pp. 135–144, 2011. ISBN 978-1-4503-0757-4. doi: 10.1145/2009916.2009938. URL http://doi.acm.org/10.1145/2009916.2009938.

[22] K. Bharat, "Searchpad: Explicit capture of search context to support Web search," *Computer Networks*, vol. 33, pp. 493–501, June 2000. ISSN 1389-1286. doi: http://dx.doi.org/10.1016/S1389-1286(00)00047-5. URL http://dx.doi.org/10.1016/S1389-1286(00)00047-5.

[23] 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. ISBN 978-1-60558-889-6. doi: http://doi.acm.org/10.1145/1718487.1718506.

[24] 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. ISSN 1532-4435. doi: http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

[25] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks ISDN Systems*, vol. 30, pp. 107–117, April 1998. ISSN 0169-7552. doi: http://dx.doi.org/10.1016/S0169-7552(98)00110-X. URL http://dx.doi.org/10.1016/S0169-7552(98)00110-X.

[26] A. Broder, "A taxonomy of Web search," *SIGIR Forum*, vol. 36, pp. 3–10, September 2002. ISSN 0163-5840. doi: http://doi.acm.org/10.1145/792550.792552. URL http://doi.acm.org/10.1145/792550.792552.

[27] 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. ISBN 978-1-59593-991-3. doi: http://doi.acm.org/10.1145/1458082.1458216.

[28] 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.

ISBN 978-1-59593-597-7. doi: http://doi.acm.org/10.1145/1277741.1277783. URL http://doi.acm.org/10.1145/1277741.1277783.

[29] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the Web," in *Proceedings of the International Conference on World Wide Web (WWW '02)*, Amsterdam, The Netherlands, May 2002. URL http://www9.org/.

[30] A. Z. Broder, P. Ciccolo, M. Fontoura, E. Gabrilovich, V. Josifovski, and L. Riedel, "Search advertising using Web relevance feedback," in *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM '08)*, New York, NY, USA, pp. 1013–1022, 2008. ISBN 978-1-59593-991-3. doi: 10.1145/1458082.1458217. URL http://doi.acm.org/10.1145/1458082.1458217.

[31] S. Brody and P. Kantor, "Automatic assessment of coverage quality in intelligence reports," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers* — vol. 2*, (HLT '11)*, Stroudsburg, PA, USA, pp. 491–495, 2011. ISBN 978-1-932432-88-6. URL http://dl.acm.org/citation.cfm?id=2002736.2002834.

[32] C. Buckley and E. Voorhees, "Retrieval evaluation with incomplete information," in *Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR '04)*, pp. 25–32, 2004.

[33] G. Cai, "Relevance ranking in geographical information retrieval," *SIGSPATIAL Special*, vol. 3, pp. 33–36, July 2011. ISSN 1946-7729. doi: 10.1145/2047296.2047304. URL http://doi.acm.org/10.1145/2047296.2047304.

[34] 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. ISSN 1386-4564. doi: http://dx.doi.org/10.1007/s10791-006-9016-2.

[35] 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. ISBN 978-1-60558-483-6. doi: http://doi.acm.org/10.1145/1571941.1571945.

[36] 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. ISBN 978-1-60558-193-4. doi: http://doi.acm.org/10.1145/1401890.1401995.

[37] N. Cardoso and M. J. Silva, "Query expansion through geographical feature types," in *Proceedings of the 4th ACM workshop on Geographical information retrieval (GIR '07)*, New York, NY, USA, pp. 55–60, 2007. ISBN 978-1-59593-828-2. doi: 10.1145/1316948.1316963. URL http://doi.acm.org/10.1145/1316948.1316963.

[38] C. Carpineto and G. Romano, "A survey of automatic query expansion in information retrieval," *ACM Computing Surveys*, vol. 44, pp. 1:1–1:50, January 2012. ISSN 0360-0300. doi: 10.1145/2071389.2071390. URL http://doi.acm.org/10.1145/2071389.2071390.

[39] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proceedings of the International Conference on World Wide Web (WWW '11)*, New York, NY, USA, pp. 675–684, 2011. ISBN 978-1-4503-0632-4. doi: 10.1145/1963405.1963500. URL http://doi.acm.org/10.1145/1963405.1963500.

[40] M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging topic detection on Twitter based on temporal and social terms evaluation," in *Proceedings of the International Workshop on Multimedia Data Mining (MDMKDD '10)*, New York, NY, USA, pp. 4:1–4:10, 2010. ISBN 978-1-4503-0220-3. doi: 10.1145/1814245.1814249. URL http://doi.acm.org/10.1145/1814245.1814249.

[41] 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. ISBN 978-1-60558-085-2. doi: http://doi.acm.org/10.1145/1367497.1367554.

[42] 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. ISSN 1046-8188. doi: 10.1145/2094072.2094078. URL http://doi.acm.org/10.1145/2094072.2094078.

[43] Z. Cheng, J. Caverlee, K. Y. Kamath, and K. Lee, "Toward traffic-driven location-based Web search," in *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '11)*, New York, NY, USA, pp. 805–814, 2011. ISBN 978-1-4503-0717-8. doi: 10.1145/2063576.2063693. URL http://doi.acm.org/10.1145/2063576.2063693.

[44] 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. ISBN 978-1-60558-799-8. doi: http://doi.acm.org/10.1145/1772690.1772714.

[45] K. Collins-Thompson, P. N. Bennett, R. W. White, S. de la Chica, and D. Sontag, "Personalizing Web search results by reading level," in *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '11)*, New York, NY, USA, pp. 403–412, 2011. ISBN 978-1-4503-0717-8. doi: 10.1145/2063576.2063639. URL http://doi.acm.org/10.1145/2063576.2063639.

[46] D. Crandall and N. Snavely, "Modeling people and places with internet photo collections," *Queue*, vol. 10, pp. 30:30–30:44, May 2012. ISSN 1542-7730. doi: 10.1145/2208917.2212756. URL http://doi.acm.org/10.1145/2208917.2212756.

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

[48] W. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*. Addison Wesley, 2009.

[49] 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.

ISBN 1-59593-323-9. doi: http://doi.acm.org/10.1145/1135777.1135902. URL http://doi.acm.org/10.1145/1135777.1135902.

[50] 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. ISBN 978-1-60558-889-6. doi: http://doi.acm.org/10.1145/1718487.1718493. URL http://doi.acm.org/10.1145/1718487.1718493.

[51] G. De Francisci Morales, A. Gionis, and C. Lucchese, "From chatter to headlines: harnessing the real-time Web for personalized news recommendation," in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '12)*, New York, NY, USA, pp. 153–162, 2012. ISBN 978-1-4503-0747-5. doi: 10.1145/2124295.2124315. URL http://doi.acm.org/10.1145/2124295.2124315.

[52] E. Di Buccio, "Design, implementation and evaluation of a methodology for utilizing sources of evidence in relevance feedback," PhD thesis, University of Padua, 2011.

[53] 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. ISBN 978-1-60558-390-7. doi: http://doi.acm.org/10.1145/1498759.1498825.

[54] B. Doerr, M. Fouz, and T. Friedrich, "Social networks spread rumors in sublogarithmic time," in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC '11)*, New York, NY, USA, pp. 21–30, 2011. ISBN 978-1-4503-0691-1. doi: 10.1145/1993636.1993640. URL http://doi.acm.org/10.1145/1993636.1993640.

[55] 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. ISBN 978-1-59593-991-3. doi: 10.1145/1458082.1458143. URL http://doi.acm.org/10.1145/1458082.1458143.

[56] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins, "Stuff I've seen: A system for personal information retrieval and re-use," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '03)*, New York, NY, USA, pp. 72–79, 2003. ISBN 1-58113-646-3. doi: http://doi.acm.org/10.1145/860435.860451. URL http://doi.acm.org/10.1145/860435.860451.

[57] 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.

[58] P. Erdős and A. Rényi, "On random graphs," *Publicationes Mathematicae*, vol. 6, 1959.

[59] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences (A Mathematikai Kutató Intézet Közleményei)*, vol. 5, pp. 17–61, 1960.

[60] W. Feller, *An Introduction to Probability Theory and its Applications,* volume 1. Wiley, 3rd ed., 1968.

[61] 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. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/503104.503110.

[62] 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. ISBN 1-59593-034-5. doi: http://doi.acm.org/10.1145/1076034.1076110.

[63] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, "Analysis of geographic queries in a search engine log," in *Proceedings of the International Workshop on Location and the Web (LOCWEB '08)*, New York, NY, USA, pp. 49–56, 2008. ISBN 978-1-60558-160-6. doi: 10.1145/1367798.1367806. URL http://doi.acm.org/10.1145/1367798.1367806.

[64] V. Ganti, A. C. König, 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. ISBN 978-1-60558-889-6. doi: http://doi.acm.org/10.1145/1718487.1718496. URL http://doi.acm.org/10.1145/1718487.1718496.

[65] E. Garfield, "Citation analysis as a tool in journal evaluation," *Science*, vol. 178, pp. 471–479, 1972.

[66] 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. ISBN 978-1-60558-930-5. doi: http://doi.acm.org/10.1145/1753846.1754025. URL http://doi.acm.org/10.1145/1753846.1754025.

[67] K. Gyllstrom, C. Soules, and A. Veitch, "Activity put in context: Identifying implicit task context within the user's document interaction," in *Proceedings of the International Symposium on Information Interaction in Context (IIiX '08)*, New York, NY, USA, pp. 51–56, 2008. ISBN 978-1-60558-310-5. doi: http://doi.acm.org/10.1145/1414694.1414707.

[68] P. Halmos, *Finite-dimensional Vector Spaces, Undergraduate Texts in Mathematics*. New York, USA: Springer, 1987.

[69] M. Harvey, I. Ruthven, and M. J. Carman, "Improving social bookmark search using personalised latent variable language models," in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '11)*, New York, NY, USA, pp. 485–494, 2011. ISBN 978-1-4503-0493-1. doi: http://doi.acm.org/10.1145/1935826.1935898. URL http://doi.acm.org/10.1145/1935826.1935898.

[70] A. Hassan, R. Jones, and F. Diaz, "A case study of using geographic cues to predict query news intent," in *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '09)*, New York, NY, USA, pp. 33–41, 2009. ISBN 978-1-60558-649-6. doi: 10.1145/1653771.1653780. URL http://doi.acm.org/10.1145/1653771.1653780.

[71] 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. ISBN 1-58113-449-5. doi: http://doi.acm.org/10.1145/511446.511513.

[72] 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.

[73] L. Hong, A. Ahmed, S. Gurumurthy, A. J. Smola, and K. Tsioutsiouliklis, "Discovering geographical topics in the Twitter stream," in *Proceedings of the International Conference on World Wide Web (WWW '12)*, New York, NY, USA, pp. 769–778, 2012. ISBN 978-1-4503-1229-5. doi: 10.1145/2187836.2187940. URL http://doi.acm.org/10.1145/2187836.2187940.

[74] D. Horowitz and S. D. Kamvar, "The anatomy of a large-scale social search engine," in *Proceedings of the International Conference on World Wide Web (WWW '10)*, New York, NY, USA, pp. 431–440, 2010. ISBN 978-1-60558-799-8. doi: 10.1145/1772690.1772735. URL http://doi.acm.org/10.1145/1772690.1772735.

[75] D. H. Hu, Q. Yang, and Y. Li, "An algorithm for analyzing personalized online commercial intention," in *Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD '08)*, New York, NY, USA, pp. 27–36, 2008. ISBN 978-1-60558-277-1. doi: http://doi.acm.org/10.1145/1517472.1517476.

[76] J. Hu, G. Wang, F. Lochovsky, J.-T. Sun, and Z. Chen, "Understanding user's query intent with Wikipedia," in *Proceedings of the International Conference on World Wide Web (WWW '09)*, New York, NY, USA, pp. 471–480, 2009. ISBN 978-1-60558-487-4. doi: http://doi.acm.org/10.1145/1526709.1526773. URL http://doi.acm.org/10.1145/1526709.1526773.

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

[78] P. Ingwersen and K. Järvelin, *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, 2005.

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

[80] W. H. Inmon, "The data warehouse and data mining," *Communications on ACM*, vol. 39, pp. 49–50, November 1996. ISSN 0001-0782. doi: http://doi.acm.org/10.1145/240455.240470. URL http://doi.acm.org/10.1145/240455.240470.

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

[82] 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. ISBN 978-1-59593-654-7. doi: http://doi.acm.org/10.1145/1242572.1242739. URL http://doi.acm.org/10.1145/1242572.1242739.

[83] 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. ISSN 0306-4573.

doi: 10.1016/j.ipm.2007.07.015. URL http://portal.acm.org/citation.cfm?id=1351187.1351372.

[84] 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. ISSN 1532-2882. doi: 10.1002/asi.v60:7. URL http://portal.acm.org/citation.cfm?id=1568763.1568772.

[85] 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. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2004.10.007. URL http://dx.doi.org/10.1016/j.ipm.2004.10.007.

[86] 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. ISBN 1-58113-567-X. doi: http://doi.acm.org/10.1145/775047.775067.

[87] T. Joachims, *The Support Vector Machine Software Package*. http://svmlight.joachims.org, 2012.

[88] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '05)*, New York, NY, USA, pp. 154–161, 2005. ISBN 1-59593-034-5. doi: http://doi.acm.org/10.1145/1076034.1076063. URL http://doi.acm.org/10.1145/1076034.1076063.

[89] T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay, "Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search," *ACM Transactions on Information and Systems*, vol. 25, April 2007. ISSN 1046-8188. doi: 10.1145/1229179.1229181. URL http://doi.acm.org/10.1145/1229179.1229181.

[90] C. B. Jones and R. S. Purves, "Geographical information retrieval," *International Journal of Geographical Information Science*, vol. 22, pp. 219–228, 2008. doi: 10.1080/13658810701626343. URL http://www.tandfonline.com/doi/abs/10.1080/13658810701626343.

[91] 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. ISBN 1-59593-323-9. doi: 10.1145/1135777.1135835. URL http://doi.acm.org/10.1145/1135777.1135835.

[92] 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. doi: http://dx.doi.org/10.1002/asi.v57:13.

[93] 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. doi: http://dx.doi.org/10.1002/asi.v57:14.

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

[95] D. Kelly and N. J. Belkin, "Reading time, scrolling and interaction: Exploring implicit sources of user preferences for relevance feedback during interactive information retrieval scrolling and interaction: exploring implicit sources of user preferences for relevance feedback," in *Proceedings of the Annual International ACM Conference on Research and Development in Information retrieval (SIGIR '01)*, New York, NY, USA, pp. 408–409, 2001. ISBN 1-58113-331-6. doi: http://doi.acm.org/10.1145/383952.384045. URL http://doi.acm.org/10.1145/383952.384045.

[96] D. Kelly and N. J. Belkin, "Display time as implicit feedback: Understanding task effects," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '04)*, New York, NY, USA, pp. 377–384, 2004. ISBN 1-58113-881-4. doi: http://doi.acm.org/10.1145/1008992.1009057.

[97] D. Kelly and X. Fu, "Elicitation of term relevance feedback: an investigation of term source and context," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '06)*, New York, NY, USA, pp. 453–460, 2006. ISBN 1-59593-369-7. doi: http://doi.acm.org/10.1145/1148170.1148249.

[98] L. Kelly, Y. Chen, M. Fuller, and G. J. F. Jones, "A study of remembered context for information access from personal digital archives," in *Proceedings of the International Symposium on Information Interaction in Context (IIiX '08)*, New York, NY, USA, pp. 44–50, 2008. ISBN 978-1-60558-310-5. doi: http://doi.acm.org/10.1145/1414694.1414706.

[99] J. Y. Kim, K. Collins-Thompson, P. N. Bennett, and S. T. Dumais, "Characterizing Web content, user interests, and search behavior by reading level and topic," in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '12)*, New York, NY, USA, pp. 213–222, 2012. ISBN 978-1-4503-0747-5. doi: 10.1145/2124295.2124323. URL http://doi.acm.org/10.1145/2124295.2124323.

[100] 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. ISBN 978-1-4503-0949-3. doi: 10.1145/2065023.2065039. URL http://doi.acm.org/10.1145/2065023.2065039.

[101] J. Kleinberg, "Authorative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, pp. 604–632, 1999.

[102] A. Kotov, P. N. Bennett, R. W. White, S. T. Dumais, and J. Teevan, "Modeling and analysis of cross-session search tasks," in *Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR '11)*, New York, NY, USA, pp. 5–14, 2011. ISBN 978-1-4503-0757-4. doi: 10.1145/2009916.2009922. URL http://doi.acm.org/10.1145/2009916.2009922.

[103] R. Kraft, C. C. Chang, F. Maghoul, and R. Kumar, "Searching with context," in *Proceedings of the International Conference on World Wide Web (WWW '06)*, New York, NY, USA, pp. 477–486, 2006. ISBN 1-59593-323-9. doi: http://doi.acm.org/10.1145/1135777.1135847.

[104] R. Kraft, F. Maghoul, and C. C. Chang, "Y!Q: contextual search at the point of inspiration," in *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '05)*, New York, NY, USA, pp. 816–823, 2005. ISBN 1-59593-140-6. doi: http://doi.acm.org/10.1145/1099554.1099746.

[105] A. Kulkarni, J. Teevan, K. M. Svore, and S. T. Dumais, "Understanding temporal query dynamics," in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '11)*, New York, NY, USA, pp. 167–176, 2011. ISBN 978-1-4503-0493-1. doi: 10.1145/1935826.1935862. URL http://doi.acm.org/10.1145/1935826.1935862.

[106] M. Lalmas, "Aggregated search," in *Advanced Topics in Information Retrieval*, (M. Melucci and R. Baeza-Yates, eds.), Springer, 2011.

[107] M. Lalmas and I. Ruthven, "A survey on the use of relevance feedback for information access systems," *Knowledge Engineering Review*, vol. 18, no. 1, 2003.

[108] 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. ISBN 1-58113-881-4. doi: http://doi.acm.org/10.1145/1008992.1009017.

[109] 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. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/1344411.1344414.

[110] 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. ISBN 1-59593-046-9. doi: http://doi.acm.org/10.1145/1060745.1060804. URL http://doi.acm.org/10.1145/1060745.1060804.

[111] R. Lempel and S. Moran, "SALSA: The stochastic approach for link-structure analysis," *ACM Transactions on Information and Systems*, vol. 19, pp. 131–160, 2001.

[112] J. Leveling, "Exploring term selection for geographic blind feedback," in *Proceedings of the ACM Workshop on Geographical Information Retrieval (GIR '07)*, New York, NY, USA, pp. 43–48, 2007. ISBN 978-1-59593-828-2. doi: 10.1145/1316948.1316961. URL http://doi.acm.org/10.1145/1316948.1316961.

[113] L. Levinson, L. Rabiner, and M. Sondhi, "An introduction to the application and theory of probabilistic functions of Markov process to automatic speech recognition," *The Bell System Technical Journal*, vol. 62-II, pp. 1035–1075, 1983.

[114] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the International Conference on World Wide Web (WWW '10)*, New York, NY, USA, pp. 661–670, 2010. ISBN 978-1-60558-799-8. doi: http://doi.acm.org/10.1145/1772690.1772758.

[115] 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. ISBN 978-1-60558-671-7. doi: http://doi.acm.org/10.1145/1592748.1592750.

[116] W. Li, X. Wang, R. Zhang, Y. Cui, J. Mao, and R. Jin, "Exploitation and exploration in a performance based contextual advertising system," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '10)*, New York, NY, USA, pp. 27–36, 2010. ISBN 978-1-4503-0055-1. doi: http://doi.acm.org/10.1145/1835804.1835811.

[117] 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. ISBN 978-1-60558-164-4. doi: http://doi.acm.org/10.1145/1390334.1390393. URL http://doi.acm.org/10.1145/1390334.1390393.

[118] 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. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/1777432.1777435. URL http://doi.acm.org/10.1145/1777432.1777435.

[119] M. D. Lieberman and H. Samet, "Multifaceted toponym recognition for streaming news," in *Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR '11)*, New York, NY, USA, pp. 843–852, 2011. ISBN 978-1-4503-0757-4. doi: 10.1145/2009916.2010029. URL http://doi.acm.org/10.1145/2009916.2010029.

[120] J. Lin, R. Snow, and W. Morgan, "Smoothing techniques for adaptive online language models: Topic tracking in tweet streams," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '11)*, New York, NY, USA, pp. 422–429, 2011. ISBN 978-1-4503-0813-7. doi: http://doi.acm.org/10.1145/2020408.2020476. URL http://doi.acm.org/10.1145/2020408.2020476.

[121] 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. ISBN 1-58113-492-4. doi: http://doi.acm.org/10.1145/584792.584884.

[122] J. Luxenburger, S. Elbassuoni, and G. Weikum, "Matching task profiles and user needs in personalized Web search," in *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM '08)*, New York, NY, USA, pp. 689–698, 2008. ISBN 978-1-59593-991-3. doi: http://doi.acm.org/10.1145/1458082.1458175.

[123] Z. Ma, G. Pant, and O. R. L. Sheng, "Interest-based personalized search," *ACM Transactions on Information and Systems*, vol. 25, p. 5, 2007. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/1198296.1198301.

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

[125] M. Melucci, "A basis for information retrieval in context," *ACM Transactions on Information and Systems*, vol. 26, pp. 1–41, 2008. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/1361684.1361687.

[126] M. Melucci, *Search Engines and Rank Correlation,* volume 4 of *Library and Information Science,* chapter 8. pp. 203–224. Emerald, 2012.

[127] M. Melucci and L. Pretto, "PageRank: When order changes," in *Proceedings of the European Conference on IR Research on Advances in Information Retrieval (ECIR '07)*, Rome, Italy, pp. 581–588, 2007.

[128] M. Melucci and J. Rehder, "Using semantic annotations for automatic hypertext link generation in scientific texts," in *Proceedings of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*, (N. Ashish and C. Goble, eds.), Sanibel Island, Florida, USA: Sun SITE Central Europe (CEUR), October 2003. http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-83/.

[129] 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. ISBN 978-1-59593-803-9. doi: http://doi.acm.org/10.1145/1321440.1321480.

[130] D. Metzler and W. B. Croft, "A Markov random field model for term dependencies," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '05)*, New York, NY, USA, pp. 472–479, 2005. ISBN 1-59593-034-5. doi: http://doi.acm.org/10.1145/1076034.1076115.

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

[132] National Institute of Standards and Technology (NIST), *Text Retrieval Conference WWW site.* http://trec.nist.gov.

[133] K. B. Ng, P. Kantor, T. Strzalkowski, N. Wacholder, R. Tang, B. Bai, R. Rittman, P. Song, and Y. Sun, "Automated judgment of document qualities," *Journal of American Society Information Science and Technology*, vol. 57, pp. 1155–1164, July 2006. ISSN 1532-2882. doi: 10.1002/asi.v57:9. URL http://dx.doi.org/10.1002/asi.v57:9.

[134] Oxford Dictionary, *The New Oxford American Dictionary.* 2005.

[135] E. Palmer, *Graphical Evolution: An Introduction to the Theory of Random Graphs.* Wiley-Interscience, 1985.

[136] E. Peserico and L. Pretto, "HITS can converge slowly, but not too slowly, in score and rank," in *Proceedings of the Annual International Conference on Computing and Combinatorics (COCOON '09)*, Berlin, Heidelberg, pp. 348–357, 2009. ISBN 978-3-642-02881-6. URL http://dx.doi.org/10.1007/978-3-642-02882-3_35.

[137] J. Pitkow, H. Schütze, 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. ISSN 0001-0782. doi: http://doi.acm.org/10.1145/567498.567526. URL http://doi.acm.org/10.1145/567498.567526.

[138] J. Ponte and W. Croft, "A language modeling approach to information retrieval," in *Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR '98)*, Melbourne, Australia, pp. 275–281, August 1998.

[139] R. S. Purves, P. Clough, C. B. Jones, A. Arampatzis, B. Bucher, D. Finch, G. Fu, H. Joho, A. K. Syed, S. Vaid, and B. Yang, "The design and implementation of SPIRIT: A spatially aware search engine for information retrieval on the internet," *International Journal of Geographical Information Science*, vol. 21, pp. 717–745, 2007. doi: 10.1080/13658810601169840. URL http://www.tandfonline.com/doi/abs/10.1080/13658810601169840.

[140] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[141] F. Radlinski and T. Joachims, "Query chains: learning to rank from implicit feedback," in *Proceedings of the ACM International Conference on Knowledge Discovery in Data Mining (KDD '05)*, New York, NY, USA, pp. 239–248, 2005. ISBN 1-59593-135-X. doi: 10.1145/1081870.1081899. URL http://doi.acm.org/10.1145/1081870.1081899.

[142] F. Radlinski, R. Kleinberg, and T. Joachims, "Learning diverse rankings with multi-armed bandits," in *Proceedings of the International Conference on Machine Learning (ICML '08)*, New York, NY, USA, pp. 784–791, 2008. ISBN 978-1-60558-205-4. doi: 10.1145/1390156.1390255. URL http://doi.acm.org/10.1145/1390156.1390255.

[143] J. Raper, "Geographic relevance," *Journal of Documentation*, vol. 63, no. 6, pp. 836–852, 2007.

[144] 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. ISBN 978-1-60558-457-7. doi: 10.1145/1507136.1507146. URL http://doi.acm.org/10.1145/1507136.1507146.

[145] T. Reichenbacher and S. De Sabbata, "Geographic relevance: different notions of geographies and relevancies," *SIGSPATIAL Special*, vol. 3, pp. 67–70, July 2011. ISSN 1946-7729. doi: 10.1145/2047296.2047310. URL http://doi.acm.org/10.1145/2047296.2047310.

[146] S. Robertson, M. Vojnovic, and I. Weber, "Rethinking the ESP game," in *Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*, New York, NY, USA, pp. 3937–3942, 2009. ISBN 978-1-60558-247-4. doi: http://doi.acm.org/10.1145/1520340.1520597. URL http://doi.acm.org/10.1145/1520340.1520597.

[147] D. E. Rose and D. Levinson, "Understanding user goals in Web search," in *Proceedings of the International Conference on World Wide Web (WWW '04)*, New York, NY, USA, pp. 13–19, 2004. ISBN 1-58113-844-X. doi: http://doi.acm.org/10.1145/988672.988675. URL http://doi.acm.org/10.1145/988672.988675.

[148] I. Ruthven, "The context of the interface," in *Proceedings of the International Symposium on Information Interaction in Context (IIiX '08)*, New York, NY, USA, pp. 3–5, 2008. ISBN 978-1-60558-310-5. doi: http://doi.acm.org/10.1145/1414694.1414697.

[149] I. Ruthven, "Information retrieval in context," in *Advanced Topics in Information Retrieval,* chapter 8, (M. Melucci and R. Baeza-Yates, eds.), Springer, 2011.

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

[151] M. Sanderson and Y. Han, "Search words and geography," in *Proceedings of the ACM Workshop on Geographical Information Retrieval (GIR '07)*, New York, NY, USA, pp. 13–14, 2007. ISBN 978-1-59593-828-2. doi: 10.1145/1316948.1316952. URL http://doi.acm.org/10.1145/1316948.1316952.

[152] T. Saracevic, "Relevance: A review of and a framework for thinking on the notion in information science," *Journal of the American Society for Information Science*, vol. 26, pp. 321–343, 1975.

[153] T. Saracevic, "Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 1915–1933, 2007.

[154] J. G. Shanahan, "Digital advertising: An information scientist's perspective," in *Advanced Topics in Information Retrieval*, (M. Melucci and R. Baeza-Yates, eds.), Berlin, Germany, pp. 209–237, 2011.

[155] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[156] 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. ISBN 1-59593-140-6. doi: http://doi.acm.org/10.1145/1099554.1099747.

[157] G. Smith, P. Baudisch, G. Roverson, M. Czerwinski, B. Meyers, D. Robbins, and D. Andrews, "Groupbar: The taskbar evolved," in *Proceedings of OZCHI 2003: New Directions in Interaction, Information Environments, Media and Technology*, pp. 34–43, 2003.

[158] 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. ISBN 978-1-4503-0747-5. doi: 10.1145/2124295.2124348. URL http://doi.acm.org/10.1145/2124295.2124348.

[159] A. Spink, *Information Behavior: An Evolutionary Instinct.* Springer, 1st ed., 2010. ISBN 3642114962, 9783642114960.

[160] 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. ISBN 1-59593-034-5. doi: http://doi.acm.org/10.1145/1076034.1076111.

[161] J. Teevan, S. T. Dumais, and E. Horvitz, "Potential for personalization," *ACM Transactions on Computer-Human Interactions*, vol. 17, pp. 1–31, 2010. ISSN 1073-0516. doi: http://doi.acm.org/10.1145/1721831.1721835.

[162] J. Teevan, S. T. Dumais, and D. J. Liebling, "To personalize or not to personalize: Modeling queries with variation in user intent," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '08)*, New York, NY, USA, pp. 163–170, 2008. ISBN 978-1-60558-164-4. doi: http://doi.acm.org/10.1145/1390334.1390364.

[163] 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. ISBN 978-1-4503-0493-1. doi: http://doi.acm.org/10.1145/1935826.1935848. URL http://doi.acm.org/10.1145/1935826.1935848.

[164] J. Teevan, M. R. Morris, and S. Bush, "Discovering and using groups to improve personalized search," in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM '09)*, New York, NY, USA, pp. 15–24, 2009. ISBN 978-1-60558-390-7. doi: 10.1145/1498759.1498786. URL http://doi.acm.org/10.1145/1498759.1498786.

[165] K. Tuite, D.-Y. H. N. Snavely, N. Tabing, and Z. Popovic, "Photocity: Training experts at large-scale image acquisition through a competitive game," in *Proceedings of the Annual Conference on Human Factors in Computing Systems (CHI '11)*, New York, NY, USA, pp. 1383–1392, 2011. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979146. URL http://doi.acm.org/10.1145/1978942.1979146.

[166] D. Tunkelang, *Faceted Search*. Morgan and Claypool, 2009.

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

[168] L. von Ahn and L. Dabbish, "Labeling images with a computer game," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '04)*, New York, NY, USA, pp. 319–326, 2004. ISBN 1-58113-702-8. doi: http://doi.acm.org/10.1145/985692.985733. URL http://doi.acm.org/10.1145/985692.985733.

[169] R. W. White, "Using searcher simulations to redesign a polyrepresentative implicit feedback interface," *Information Processing Management*, vol. 42, pp. 1185–1202, 2006. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2006.02.005.

[170] R. W. White, P. Bailey, and L. Chen, "Predicting user interests from contextual information," in *Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR '09)*, New York, NY, USA, pp. 363–370, 2009. ISBN 978-1-60558-483-6. doi: http://doi.acm.org/10.1145/1571941.1572005.

[171] R. W. White, M. Bilenko, and S. Cucerzan, "Studying the use of popular destinations to enhance Web search interaction," in *Proceedings of the Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '07)*, New York, NY, USA, pp. 159–166, 2007. ISBN 978-1-59593-597-7. doi: http://doi.acm.org/10.1145/1277741.1277771. URL http://doi.acm.org/10.1145/1277741.1277771.

[172] R. W. White, J. M. Jose, and I. Ruthven, "A task-oriented study on the influencing effects of query-biased summarisation in Web searching," *Information Processings Management*, vol. 39, pp. 707–733, 2003. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/S0306-4573(02)00033-X.

[173] R. W. White and D. Kelly, "A study on the effects of personalization and task information on implicit feedback performance," in *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '06)*, New York, NY, USA, pp. 297–306, 2006. ISBN 1-59593-433-2. doi: http://doi.acm.org/10.1145/1183614.1183659.

[174] R. W. White and R. A. Roth, *Exploratory Search: Beyond the Query-response Paradigm*. Morgan and Claypool, 2009.

[175] Wikipedia, *Definition of Dendrogram*. http://en.wikipedia.org/wiki/Dendrogram, October 2011.

[176] 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. ISBN 978-1-60558-487-4. doi: 10.1145/1526709.1526774. URL http://doi.acm.org/10.1145/1526709.1526774.

[177] 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. ISBN 978-1-4503-0757-4. doi: 10.1145/2009916.2009970. URL http://doi.acm.org/10.1145/2009916.2009970.

[178] 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. ISBN 978-1-60558-205-4. doi: http://doi.acm.org/10.1145/1390156.1390310.

[179] 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. ISBN 978-1-60558-516-1. doi: http://doi.acm.org/10.1145/1553374.1553527.

[180] 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. ISBN 1-59593-140-6. doi: 10.1145/1099554.1099652. URL http://doi.acm.org/10.1145/1099554.1099652.

[181] 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. ISBN 978-1-60558-495-9. doi: http://doi.acm.org/10.1145/1557019.1557131.

[182] 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. ISSN 1559-1131. doi: 10.1145/2109205.2109206. URL http://doi.acm.org/10.1145/2109205.2109206.

[183] 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. ISBN 978-1-60558-160-6. doi: 10.1145/1367798.1367811. URL http://doi.acm.org/10.1145/1367798.1367811.