Aspects of Context in Daily Search Activities
Survey about Nowadays Search Habits

Melyara Mezzi and Nadjia Benblidia
LRDSI Laboratory, Computer Science Department, Saad Dahlab University, Blida 1, Blida, Algeria

Keywords: Information Retrieval, Contextual Information Retrieval, Context-awareness, Search Trends.

Abstract: Over almost 70 years of perpetual improvement, Information Retrieval (IR), has had many approaches to satisfy one’s information need. In this paper, we want to put forward and measure the importance of the inclusion of a contextual dimension to enhance the relevance and effectiveness of a search task. Indeed, a search task is no more concerned with a query and a set of documents only, but it is related to a wide range of some extrinsic and intrinsic factors, so called “context”, which became a great challenge these last few years. Besides the emergence and the significance of the use of context in IR, we conducted a survey with 434 internet users to understand their search trends and habits. The results and discussion may contain valuable information for future researchers.

1 INTRODUCTION

Information seeking behaviour is rooted in a need to find information (Han, Wang, M., and Wang, J., 2010). According to Saracevic (2010), information is anything that can change person’s knowledge. Thus, the Information Retrieval (IR) process begins with an anomalous state of knowledge (ASK). Then, many changes in knowledge state are involved. In short, IR is a purposeful process that alters the state of knowledge reacting to an information need or gap. A simple vision of an Information Retrieval System (IRS) was believed to be as follows: (a) The user expresses an Information need by formulating a question (called query), (b) The IRS answers the query and gives back results (texts, images, videos,…etc), (c) The final phase is up to the user who has to evaluate and reformulate her query if the results do not satisfy her request. Today, this vision became somehow obsolete, because the users, their queries, and the desired information were believed to be static. So, the relevance of a document was computed statically between the query and the set of documents ignoring the user, the device, the environment, and the specificities around the search activity which constitute the search context and are as a matter of fact highly variable factors. Besides, with the technology advances, information can nowadays, be accessed everywhere and at anytime which add to the variability and the uniqueness of each search situation. And as no information is context free, the inclusion of a contextual dimension in the classic IR process became a real challenge.

To meet the expectations aforementioned, we have conducted a short survey among 434 Internet users in order to understand and analyze actual user preferences and trends in the area of IR, especially with the huge advances in search devices. We aim to provide researchers with a good starting point in the field of Contextual Information Retrieval (CIR).

The rest of the paper is structured as follows. In section 2, we discuss the significance of context in IR. After that, in section 3 we overview the most important context’s components as an outcome to a study performed on 16 works in the field of CIR. Then, in the 4th section, we present our survey, its analysis, and discussion. Finally, section 5 ends the paper with a discussion and outlooks.

2 CONTEXT SIGNIFICANCE IN INFORMATION RETRIEVAL

According to Agbele et al (2012), context refers to the circumstances in which an event (an IR computing task in our case) takes place. In fact, context is multi-layered; it extends beyond users or systems. It is not self-revealing, nor it is self-evident, but researchers do integrate context, which, they understand intuitively, in IR theory and practice.
In addition, IR task’s context is any information whose change modifies the task’s outcome (Agbele et al., 2012). Thus, an application is believed to be “context-sensitive” or “context-aware” if its structure and behaviour change depending on the context so as to provide relevant information and services for a given user. Research activities on context-aware IR have increased remarkably in recent years and many approaches have been developed to automatically provide users with information and services based on their current situation (Mircesa, Trajkovic, and Ristevska, 2010). But unfortunately, they remain greatly dependent on the field of application (smart-spaces, weather-forecast, tour guides…). In fact, there are no standards.

Context-aware computing was introduced for the first time by Schilit, Adams, and Want (1994) who state: “One challenge of mobile distributed computing is to exploit the changing environment with a new class of applications that are aware of the context in which they are run”. After that, there have been many definitions about the notion of context in IR. One of the most approved definitions is the one given by Dey (2001): “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and application themselves. And by extension, the environment, the application and the user are embedded in”. In short, we can say that Context includes all the intrinsic and extrinsic factors, which are related to a given search task and whose the direct or indirect inclusion in the IR process leads to enhance, whether implicitly or explicitly its effectiveness to convey the right information to the searcher.

2.1 Features of the Information Retrieval Task

As stated in (Saracevic, 2010), the information can be: (a) objects in the world potentially conveying information, (b) what is transferred from people or objects to person’s cognitive systems, or (c) components of internal knowledge in people’s mind. Furthermore, according to Han, Wang, M., Wang, J. (2010), the request for information can either be external or self-initiated. In the same ground, Saracevic (2010) talked about direct (end-user) search and mediation search. Direct searchers are people who seek information by and for themselves, whereas in mediation search, there is an intermediary who acts on the behalf of a person who is actually seeking for information. The mediation can either be informal when it comes to search information for colleagues, family, and friends, or formal when it comes to search for information as a searcher or a teacher. Moreover, two kinds of search are noticeable as reported by (Fujita and Oyama, 2011), (Daoud et al, 2009), and (Mircesa, Trajkovic, and Ristevska, 2010): (a) Navigational (evidential, or pull-based) search, and (b) Thematic (informational, or push-based) search. Navigational search deals with aware users having steady needs. In this case, IR is explicit and the process consists of comparisons with previous knowledge. Whereas, in thematic search, the user inputs the query that explains or describes information related to that the user wishes to collect or research. Hence, IR is implicit and the process consists of seeking for new knowledge whether the needs are known, unknown and poorly defined, or changing. Table 1 shows the advantages and disadvantages of each kind.

<table>
<thead>
<tr>
<th>Table 1: Navigational search VS thematic search.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Navigational search</strong></td>
</tr>
<tr>
<td>Aware users</td>
</tr>
<tr>
<td><strong>Thematic search</strong></td>
</tr>
</tbody>
</table>

To overcome the drawbacks of these search methods, there is a need to contextualize the search task.

3 CONTEXT’S COMPONENTS

Further to researches in the field of CIR, we can observe that each search task is unique and comes under a certain configuration of contextual factors. However, some correlations can be found among a set of search activities of the same user, between two similar users, or between two disjoint users performing a search task in a similar configuration of contextual factors. According to Jilei (2010), context fully describes the searcher, her device, and her surroundings using a wide range of sensed and historic information which forms the backbone for a completely new class of services. There is a real need for categorizing context’s types or components in order to spot the most useful ones according to a given application. Effectively, nowadays, context is more targeted than ever.

As Han, Wang, M., Wang, J. (2010), we agree that task is the driving force that constitutes IR and
real information behaviour. In order to find if there may be other contextual components, we choose sixteen valuable works that made use of context for different purposes. Our goal was to deepen our comprehension of the notion of context according to different use cases and to come out with a categorization of the context factors. What can be noticed is that the use of contextual factors differs from one application to another. Thus, the related works are given just as valuable resources to enrich future researchers with leading theories, models, and results in the area of Contextual IR (CIR).

We find that the IR task is usually interlaced with seven contextual components (table 2), namely: user, queries, device, time, location, environment, and documents. We restricted our focus to those seven contextual factors and to test their coverage, we conducted a short survey about search habits.

### 4 SURVEY ABOUT NOWADAYS SEARCH HABITS

In order to understand the trends and users’ intents

---

**Table 2: Important contextual factors in an Information Retrieval task.**

<table>
<thead>
<tr>
<th>Components</th>
<th>Example</th>
<th>Sources</th>
<th>Related works</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search task</td>
<td>Personal calendars can be used to discover user’s current task</td>
<td>Forms, events in the calendars, query logs, feedback</td>
<td>(Abowd et al., 1999), (Dey, 2001), (Belkin et al., 1999), (Ingwersen and Järvelin, 2005), (Kamvar and Baluja, 2006), (Poveda et al., 2010), (Saracevic, 2010).</td>
</tr>
<tr>
<td>User</td>
<td>Sana usually browses technology news when waiting the subway in working days morning.</td>
<td>Profiling, user mining, forms and feedbacks, search logs, personal data and content, contact list, social network.</td>
<td>(Abowd et al., 1999), (Dey, 2001), (Belkin et al., 1999), (Bertrand, Egyed-Zsigmond, Calabretto, 2012), (Bouidghaghen, 2009), (Daoud et al., 2009), (Ingwersen and Järvelin, 2005), (Kessler, 2007), (Kostadinov et al., 2004), (Poveda et al., 2010), (Ryan, Pascoe, and Morse, 1999), (Ryu et al., 2010), (Saracevic, 2010), (Tamine and Bahsoun, 2006).</td>
</tr>
<tr>
<td>Queries</td>
<td>Composite Capabilities/ Preference Profile (CC/PP) proposes an infrastructure to describe device capabilities and user preferences. Used for content presentation</td>
<td></td>
<td>(Abowd et al., 1999), (Bertrand, Egyed-Zsigmond, Calabretto, 2012), (Ingwersen and Järvelin, 2005), (Kessler, 2007), (Poveda et al., 2010), (Ryan, Pascoe, and Morse, 1999), (Ryu et al., 2010).</td>
</tr>
<tr>
<td>Device</td>
<td>The doctor uses her tablet in a hospital to search about the suitable diagnosis.</td>
<td>System clock, calendars,</td>
<td>(Belkin et al., 1999), (Bouidghaghen, 2009), (Daoud et al., 2009), (Ryan et al., 2010).</td>
</tr>
<tr>
<td>Time</td>
<td>According to a time where a user search for a restaurant we can deduce the type of food he’s searching for “break-fast”, “lunch”, …etc</td>
<td>We can use infrared, Bluetooth and WIFI signal strength to determine indoor locations and GPS for outdoor locations.</td>
<td>(Abowd et al., 1999), (Bertrand, Egyed-Zsigmond, Calabretto, 2012), (Bouidghaghen, 2009), (Brown and Jones, 2001), (Kessler, 2007), (Poveda et al., 2010), (Ryan, Pascoe, and Morse, 1999), (Ryu et al., 2010).</td>
</tr>
<tr>
<td>Location</td>
<td>City guides, weather forecasting, products and services marketing</td>
<td>Environment sensors, device perserviveness (Bluetooth, accelerometers…)</td>
<td>(Abowd et al., 1999), (Bertrand, Egyed-Zsigmond, Calabretto, 2012), (Bouidghaghen, 2009), (Brown and Jones, 2001), (Kessler, 2007), (Poveda et al., 2010), (Ryan, Pascoe, and Morse, 1999), (Ryu et al., 2010).</td>
</tr>
<tr>
<td>Environment</td>
<td>Find all the participants for a meeting saved as an event in the calendar.</td>
<td>Web, intranet, or personal texts, images, videos…etc</td>
<td>(Belkin et al., 1999), (Ingwersen and Järvelin, 2005), (Saracevic, 2010).</td>
</tr>
</tbody>
</table>
in IR and come out with significant patterns for our upcoming research in CIR, we conducted a short survey among 434 anonymous online users (mostly Facebook and Linkedin users).

4.1 Sample Data

Based on the afore-mentioned influential context factors that can be found in the literature, ten leading questions have been formulated and formatted. Then, we broadcasted the Google Form link through some social network groups and also provided a printed version to students (about 12% of the participants). The participants were from 27 nationalities which contributed to enrich the study, but unfortunately, since the study was carried out online many socio-demographic categories have been excluded. Furthermore, despite the fact that many similar surveys have been conducted already, our main focus was to understand the habits and preferences behind actual daily search tasks knowing that several technological advances occurred this last decade. We took special care to formulate the study in the simplest possible form in order to provide researchers in the field of CIR with a clear view about contemporary search preoccupations.

The survey motivated users for details surrounding their daily search activities, presented as yes/ no, single response, and multiple choices questions including sections for suggestions. First, users were invited to provide background information about their gender, age, activity, and whether they own a Smartphone or not. Table 3 summarizes the participant types.

More precisely, the study concerned six leading issues: (a) favourite source of information, (b) favourite devices used for search activities, (c) favourite search categories, (d) number of keywords usually used, (e) most important contextual factors, and (f) collaboration in the search activity.

4.2 Results and Discussion

In the following, we will try to analyze the results. But before, it is important to mention that for a deep understanding of them, we performed a cross tabulation analysis, which shows -mostly- very harmonious results regardless to the different types of demographic categories.

4.2.1 Favourite Source of Information

About 62.2% of the searchers concede preferring ‘Famous search engines’ to perform their search activities whereas, 19.7% choose ‘Social networks’ and 10.9% ‘Forums’. Besides, a minority of searchers 5.1% and 2.1% admit using, respectively, ‘Mobile applications’ and ‘other sources of information’ like dedicated web portals, less known and more targeted search engines, digital libraries, internal society or university databases, library catalogues, faceted search engines, personal content (documents, emails, and bookmarked web pages), and finally, computational knowledge engines such as wolfram. Unfortunately, these results (Figure 1) gave rise to our apprehension about the preference of researchers towards famous search engines, which, are agreed to provide powerful search results for trivial queries. But, they lose out personalization and customization of the results according to internet surfers’ needs and purpose. And consequently, they miss effectiveness if the needs are unknown, dealing with thematic search activities for example.

4.2.2 Favourite Devices Used for Search Activities

Results show (Figure 2) that 39% of searchers use mostly ‘Laptops’ while searching, whereas 23.1% still prefer ‘Desktops’, 21% ‘Smartphone’, 15% ‘Tablets’, and only 1.9% of searchers use their ‘Mobile phones’ in daily search tasks. This means that despite the spread of mobile technologies,
people still make use of desktops when it comes to perform their daily search activities. Moreover, we can notice that among all the mobile devices, laptops are the favourite, and it is quite understandable because of their ease of use in terms of interaction fluency, query typing, and clarity of results presentation.

Figure 2: Search devices statistics.

4.2.3 Favourite Search Categories

Concerning favourite search categories, unlike the study of Kamvar and Baluja (2006), we found that ‘Technology’ outclasses the other categories with 19.69%, nearly followed by ‘News and events’ with 16.41%, and ‘Science’ 14.86%. The remaining proposed categories obtained the scores showed in table 4, beginning with the highest.

Table 4: Results of search categories.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>19.69%</td>
</tr>
<tr>
<td>News &amp; events</td>
<td>16.41%</td>
</tr>
<tr>
<td>Science</td>
<td>14.86%</td>
</tr>
<tr>
<td>Entertainment</td>
<td>8.75%</td>
</tr>
<tr>
<td>Health &amp; food</td>
<td>8.57%</td>
</tr>
<tr>
<td>Travels</td>
<td>6.65%</td>
</tr>
<tr>
<td>Society &amp; communication</td>
<td>6.11%</td>
</tr>
<tr>
<td>Local services</td>
<td>5.56%</td>
</tr>
<tr>
<td>Games &amp; hobbies</td>
<td>4.10%</td>
</tr>
<tr>
<td>Industry</td>
<td>3.01%</td>
</tr>
<tr>
<td>Others</td>
<td>1.46%</td>
</tr>
</tbody>
</table>

Despite the differences between the mentioned search categories, we wanted to find some patterns concerning the types of needs behind the queries. The survey results show, analogically to the study of Broder (2002), that the respondents were most willing to perform informational (thematic) search than navigational one.

4.2.4 Number of Keywords per Query

The 44.7% of respondents admitted using from one to three keywords, 44.5% from four to six, and 10.8% more than six keywords. Undeniably, the less keyword, the searcher uses, the harder it is for the IRS, to please their need of information. For instance, we have noticed that 45.51% of Smartphone users utilize from one to three keywords, whereas 43.26% use from four to six. Contrariwise, this trend was reversed for respondents without Smartphones with, respectively, 42.62% and 47.54% as shown in figure 3.

Figure 3: Keywords statistics for Smartphone and non-Smartphone users.

These results resemble barely to the study conducted by Kamvar and Baluja (2006), who reported that mobile users’ queries are shorter and therefore more ambiguous. Indeed, we remark that the two sample results (i.e. smartphone and non-smartphone users) are nearly similar and this is due to the technological advance concerning smartphones that are nowadays as powerful as some laptops. Nevertheless, the results obtained in this section about keywords, indicates the need to rely on the context factors surrounding the search activity.

4.2.5 Most Important Contextual Factors

We noticed that the most important contextual factors (Figure 4) are: ‘Accuracy’ with 38.29%, then ‘Time’ (freshness of the information) with 23.9%, followed by ‘Results and content personalization’ with 12.13%, ‘Personal preferences’ with 11.3%, ‘Location’ with 11.18%, and ‘Social network preferences’ with 2.85%. Finally, 0.36% of respondents chose the option ‘Other’, and gave some suggestions. We retain trustworthiness and genuineness of the information sources, the results’ ranking and referencing, and website speed. This question was somehow the core of our study, since our main focus was about the importance of extrinsic and intrinsic contextual factors in any search activity.

The two most interesting outcomes are:
- The accuracy and freshness of the information are more important than their relation to the notion of location. This differs from the perspective of Ryu et al (2010), who
classified contextual factors that prompt information needs as follows—beginning with the most influential: location, time, conversation, and activity, and also Kamvar and Baluja (2006) who classified them as follows: activity, location, time, and conversation. Instead, this confirms the trend concerning the interaction at a very large scale (allowed by social networks mostly), where, everyone is a world citizen without known boundaries, nor territorial limitations of knowledge.

– The limit between ‘Personal preferences’ and ‘Social network preferences’ is small. That is to say people do take into account the view of their (physical and virtual) social network proportionally to their own ‘Personal preferences’. According to Evans and Chi (2008), external environment (i.e. people) may be valuable information resources for one’s information search process. In their paper, Evans and Chi (2008) state that recently, searchers have observed direct user cooperation during web-based information seeking. Active collaboration may occur under some circumstances, where users interact together remotely, asynchronously, and even involuntarily and implicitly. They are indeed, influenced by their friends, collaborators, as well as by their social network. This is why the opinion of this latter is as important as their own yet most people do prefer performing their research alone as found in the question concerning the collaboration in research.

4.2.6 Collaboration in Search Activity

84.33% of respondents concede that they rather perform a search activity alone. Whereas, 11.06% prefer being surrounded by real (physical) friends, and 4.15% choose to rely on their social network circles. Moreover, 0.46% of respondents gave suggestions that support overall that most searches are performed independently, but at times can be conducted collaboratively. This does depend on the need. These results (Figure 5) support that effectively, the IR task can either be external or self-initiated.

![Figure 5: Search activity statistics.](image)

5 CONCLUSIONS

Throughout years and with the advance of technology, search task became more flexible, allowing a wider range of choices between different sources of information, devices, and search categories. Moreover, the perspective of an eventual collaboration became possible, regardless of the location of the different searchers. In this paper, the significance of the inclusion of a contextual dimension was discussed. Moreover, we tried to inquire about actual search trends taking into account the technological advances. Thus, we introduced our short survey with its detailed results and analysis, which we expect will provide future researchers with valuable information. We retain the inclination of users towards: (a) social network preferences proportionally to their own personal preferences, also (b) users concern about accuracy and time, and finally (c) shorter and thus more ambiguous queries. Consequently, our upcoming work will consist on the formalization and testing of a CIR model centred on the IR task.

REFERENCES


