Social approach to context-aware retrieval

Luca Vassena
University of Udine
via delle Scienze, 206
Udine, Italy
vassena@dimi.uniud.it

ABSTRACT

In this paper we present a general purpose solution to Web content perusal by means of mobile devices, named Social Context-Aware Browser. This is a novel approach for the information access based on the users' context, whose aim is to retrieve what the user needs, even if she did not issue any query. Our solution is built upon a social model that exploits the collaborative efforts of the whole community of users to control and manage contextual knowledge, related both to situations and resources. This paper presents a general survey of our solution, describing the idea and presenting an implementation approach.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

Context-aware retrieval, mobile search, social, folksonomy, Web 2.0

1. INTRODUCTION

Context-aware computing is a computational paradigm that has faced a rapid growth in the last few years, especially in the field of mobile devices. A key-role in this new approach is played by the notion of context, that is roughly described as the situation the user is in. This concept encloses important information that could be used to affect the capabilities of mobile devices, adapting them to the user's needs. In particular, contextual data can be used to predict the user needs and to seek and retrieve information, thereby reducing the complexity of the user-device interaction and providing the right information in the right place at the right time. From this point of view, because of the huge amount of contextual information and its heterogeneity and uncertainty, the mobile and context-aware computing environments represent a new challenge for Information Retrieval (IR). The combination of IR and context-aware computing has been named context-aware retrieval [4].

These considerations guided us towards a new approach to Web contents production and fruition, where contextual data are exploited to capture the dynamic nature of the user needs, of the information available, and of the relevance of this information, typical of a mobile user in the real world. This approach is named *Social Context-Aware Browser* and its novelty is threefold. First of all this is a new radical approach that aims at discovering *"the query behind the context"*: to retrieve what the user needs, even if she did not issue any query [7]. Second this is not a domain dependent application, but a new generic way of interaction and information access, able to adapt to every domain. Third, as current models for context-awareness are too limited for very general applications, this approach brings new models built upon the social dynamics at the basis of Web 2.0.

This paper is structured as follows. We first briefly survey related work (Section 2), presenting the Context-Aware Retrieval field and introducing the main ideas behind Web 2.0. We then describe our solution (Section 3), presenting a general survey, the main ideas, and an implementation approach. In Section 4 we present a brief discussion and finally we draw some conclusions and we present future work (Section 5).

2. RELATED WORK

2.1 Context-Aware Retrieval

Context-Aware Retrieval (CAR) is an extension of classical Information Retrieval (IR) that incorporates the contextual information into the retrieval process, with the aim of delivering information to the users that is relevant within their current context [4]. CAR systems are concerned with the acquisition of context, its understanding, and the application of behaviour based on the recognized context [11].

Typical CAR applications present the following characteristics [4]: a mobile user, i.e., a user whose context is changing; interactive or automatic actions, if there is no need to consult the user; time dependency, since the context may change; appropriateness and safety to disturb the user. Although CAR applications can be both interactive and proactive in their communication with the user, we concentrate on the proactive aspects, since they are more relevant to our proposal. Besides, we concentrate on the association between CAR and mobile application, as they can be considered as the prime field for CAR [4].

An example of CAR system is the Ubiquitous Web [5], a solution based on the spontaneous annotation by a community of users of objects, places, and other people with Web accessible content and services. A more general system is represented by the MoBe framework [7]. In this applica-

Appears in the Proceedings of the 1st Italian Information Retrieval Workshop (IIR'10), January 27–28, 2010, Padova, Italy.
http://ims.dei.unipd.it/websites/iir10/index.html

tion, a general inferential framework (based on ontologies and Bayesian networks) combines the information coming from sensors to infer new and more abstract contexts (user activities, needs, etc.), that are used to retrieve and execute the most relevant applications.

2.2 Web 2.0, the social web

With Web 2.0 [9] and social software we represent all web-based services with "an architecture of participation", that is, an architecture featuring a high interaction level among users and allowing users to generate, share, and take care of the content. In the plenty of tools provided by Web 2.0, we are mainly focusing on social bookmarking and folksonomies.

Social bookmarking is a method for organizing, searching, and managing documents of interest among users. In a social bookmarking system, users save links to documents of interest in order to remember or share them with the community. Social bookmarking is strictly related with the concept of folksonomy, that is the practice of annotating and categorizing content in a collaborative way, by means of informal tags. Folksonomies, that is a portmanteau of folk and taxonomy, allow users to easyly and informally descrive documents and content. This represents a powerful combination that has gained popularity as it allows a more natural and simpler management of the knowledge. The use of freely choosen categorizations and the collaborative aspect in fact allow also non-expert users to classify and find information. Folksonomies and social bookmarking for example are used in well-known Web 2.0 systems like Flickr[1], Youtube[2], Del.icio.us[3], etc.

Folksonomies however are criticized because the lack of terminological control could lead to unreliable and inconsistent results [3].

3. SOCIAL CONTEXT-AWARE BROWSER

3.1 Description

The Social Context Aware Browser (sCAB for short) [12] is a general purpose solution to Web content navigation by means of context-aware mobile devices. It allows a "physical browsing": browsing the digital world based on the situations in the real world. The main idea behind sCAB is to empower a generic mobile device with a browser able to automatically and dynamically retrieve and load Web pages, services, and applications according to the user's current context.

The sCAB acquires information related to the user and the surrounding environment, by means of sensors installed on the device or through external servers. This information, combined with the user's personal history and the community behaviour, is exploited to infer the user's current context (and its likelihood). In the subsequent retrieval process, a query is automatically built and sent to an external search engine, in order to find the most suitable Web pages for the sensed context and present them to the user.

As current models for context-awareness are too limited for very general applications like the sCAB, this approach brings new social models for CAR that exploit the collaborative efforts of the community of users. The community, in fact, is encouraged to define the contexts of interest, share, use and discuss them, associate context to content (web pages, applications, etc.), to have a dynamic and more user-tailored context representation and to enhance the process of retrieval based on users' actual situation.

In particular users can freely interact with resources and can define that a resource is useful (or not adapt) to their current context, can associate resources to particular contexts, can explicitly define the context their are in, and finally can browse resources relevant for their current context.

3.2 Model

3.2.1 *Context representation*

We represent the context as a folksonomy. Each tag is banally a keyword or string of text and represents a single contextual value [8]. We divide the contextual tags into two categories:

- Concrete tags: represent the information obtained by a set of sensors. These information can be read from the surrounding environment through physical sensors (e.g., temperature sensor), or can be obtained by other software (e.g., calendar) through logical sensors. Concrete tags that directly refers to sensors values are represented using the *triple tags* notation that are tags that uses a particular syntax (`namespace:predicate=value`) to define extra information.
  For example, `geo:longitude=12.456` is tag for the geographical longitude coordinate whose value is 12.456. Other concrete tags, can be automatically obatined by the sensed values (e.g. `afternoon`, `summer`, ...).

- Abstract tags: represent the high level contextual information that are freely associated by the users to the concrete contexts, in order to detail their context description. Some examples are: `home`, `shopping`, etc.

The difference between the two categories is faded since the contexts cannot be unambiguously assigned to one or the other category. However this partition is helpful in order to distinguish the low level information coming from sensors and the high level contextual information intoduced by users.

The user context is a "cloud" composed by an undefined number of concrete and abstract tags (Figure 1).

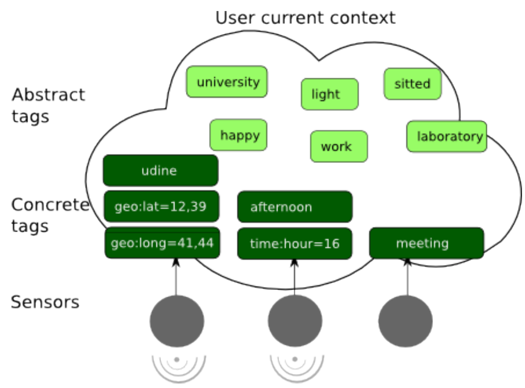

Figure 1: User's current context.

[1]www.flickr.com
[2]www.youtube.com
[3]www.del.icio.us.com

3.2.2 Operations

In the sCAB conceptual model [12] there are six main operations. The first two are performed automatically and continuosly by the system. With the *inference* operation (Figure 2), starting from the concrete tags sensed by sensors, the most relevant abstract tags are retrieved and become part of the user's context representation. Then with the *retrieval* operation (Figure 2), starting from the set of all the tags in the user's current context, the most relevant resources are retrieved. For example, starting from the GPS coordinates, the system enhance the user's context with the abstract tags "`walk out park dog`"; then starting from all the tags, the system retrieves resources relevant to the given context, as Web pages that teaches how to train dogs, etc.

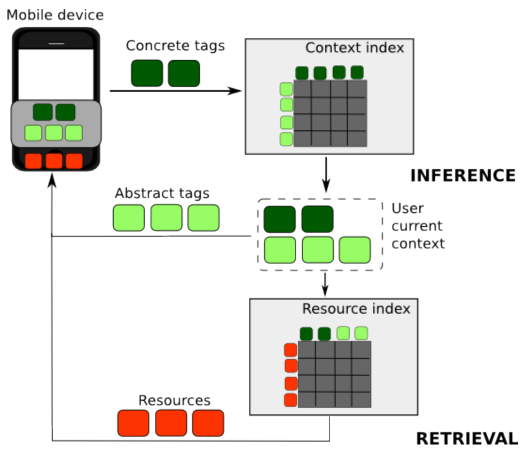

Figure 2: Inference and retrieval operations.

The other four operations are strictly related to the user interaction: the main two are *definition* and *annotation* (Figure 3). The *definition* is used to manage the contextual information and it is performed when a user directly define her context, or when she provides contextual tags during the annotation of a resource. In particular, this operations manages the associations between concrete and abstract tags, and the strength of their relationships. The *annotation* on the contrary is used to manage the association between contextual tags and resources and it is performed when the users link resources to particular contexts. We can imagine a user at a park with her dog: she wants to associate to her context a particular Web page teaching dog training. For this reason she bookmarks that resource with the contextual tags "`out dog park sunny train`". Doing so, first the added abstract tags are related to the sensed concrete tags and for all the users with a similar concrete tag cloud, these abstract tags (or part of them) can become part of the their context representation. Second, that particular Web page is enhanced with all the tags, and it will be automatically proposed to users every time they will be in a similar context.

As the users are the main actors in the process of context definition and resource annotation, problems related to the quality of context and resources are likely to appear. To cope with this problem we propose the adoption of a social evaluation/reputation mechanism. We exploit the ideas presented in [6]: every element in the model (users, contexts, resources) has a score that increases or decreases based on the community behavior. The score of each user is used to

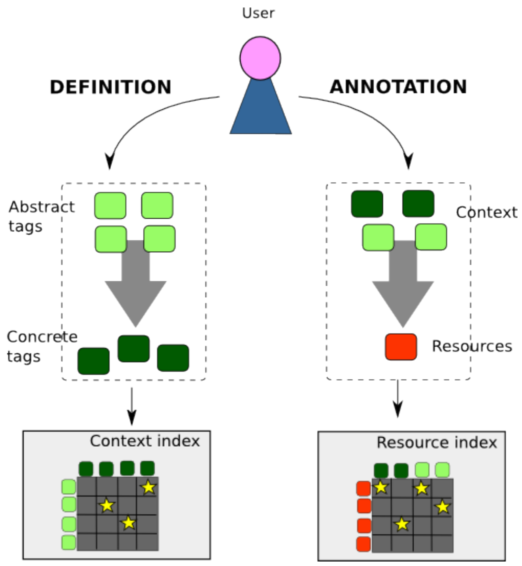

Figure 3: Definition and annotation operations.

weight the operations she performs, while the scores of contextual tags and resources define their quality and relevance. If a resource annotated with contextual information is never used in that context, the related score decreases and more relevant resources will stand out.

3.3 Implementation approach

Concrete and abstract tags, and resources are the main elements in our implementation model. Concrete tags, as output of sensors, are exploited to retrieve the most relevant abstract tags, and in the same way all the tags are exploited to retrieve the most relevant resources.

In the following sections we show an implementation proposal and how the different operations in the model have effect on the system, from a low level point of view.

3.3.1 Indexes

We exploit two indexes. In the first one, called *contexts index*, abstract tags are indexed over concrete tags, while in the second one, called *resources index*, resources are indexed over the set of all tags (both concrete and abstract). The proposed approach is community based, thus the indexes and the inferential system are managed by remote servers and not stored on the mobile device. Since the approach is similar for both the indexes, we are going to show just the first one.

The contexts index is a matrix that describes the frequency of abstract tags over the concrete ones. Each column corresponds to a concrete tag, and each row corresponds to an abstract tag. Each entry in the matrix has three values (Figure 4):

- U_{ij}: represents the user that has associated the abstract tag i to the concrete tag j first;
- S_{ij}: a score that defines how relevant the abstract tag i is for the concrete tag j. This value is in the interval $[0, 1]$;
- σ_{ij}: steadiness value that defines how steady is the association between the abstract tag i and the concrete

tag j.

	c_1	c_2	$\dots$
a_1			
a_2		$(U_{22}, S_{22}, \sigma_{22})$	
$\vdots$			

Figure 4: *Contexts index* example

Intuitively, since not all the abstract tags can be related to all concrete tags, the proposed index will be a very sparse matrix. At the same time, because of the very high number of both concrete and abstract tags, the index can assume very huge dimensions. However a lot of research is being performed on indexes designing and analysis, also in the CAR field [2]. The related discussion is out of the scope this work.

3.3.2 Users' score

In our approach two values are associated to each user and they define the goodness of the user in working with contextual information:

- S_{U_c}: a score that defines how good the user is in associating concrete tags to abstract tags;
- S_{U_r}: a score that defines how good the user is in associating resources to contexts;

As previously, we are concentrating only on the management of values related to concrete and abstract tags, since the approach is exactly the same working at the higher level of tags and resources.

Every time a new relation between abstract and concrete tags is created with a *definition* ("filling a hole" in the index), the user who performed the operation is associated to that relation. Then on the basis of how the community interacts with those contextual information, the user's score will be update. It is calculated as follows: for each association among tags ij performed by the user U, S_{U_c} corresponds to the mean of the products $\frac{\sigma_{ij}}{\sigma_{max}} \times S_{ij}$, where σ_{max} is the max steadiness value in the index.

New associations have a low steadiness value, thus their score, as their have not steadied yet, will have low influence on the user's score. Good associations will have high score and steadiness values, and they will reflect on high users' score. In the same way, low users' scores are due to bad associations between contextual tags. Since $S_{ij} \in [0, 1]$, also $S_{U_c} \in [0, 1]$.

In this approach, for simplicity, only new associations between tags are considered for the computation of the users' score. An extension could consider all the existing associations. In this way a user is "good" because she defines good new associations and because she exploits existing good association.

3.3.3 Values update

The proposed indexes are not static, but the values related to the association between concrete and abstract tags and resources are continuosly updated, based on the interaction of users with resources in context.

With every *definition* operation the values in the *contexts index* are updated according to the following system (for the values in the *resources index* with the *annotation* operation the approach is similar) :

- $\sigma_{ij}(t_{i+1}) = \sigma_{ij}(t_i) + S_{U_c}(t_i) \times \beta$
- $v = \dfrac{\sigma_{ij}(t_i) \times S_{ij}(t_i) \pm S_{U_c}(t_i) \times \beta}{\sigma_{ij}(t_{i+1})}$
- $S_{ij}(t_{i+1}) = \begin{cases} v & if\ v > 0 \\ 0 & otherwise \end{cases}$

where t_i represents a discrete time instant and t_{i+1} the subsequent time instant.

While the score is a value in the interval $[0, 1]$, the steadiness is an always increasing value. The higher the steadiness of an association is, the more stable the association is, and then the lesser effect each update operation will have. The user's score is exploited for the update of the values in the index. It can both increase an association, or decrease it (e.g. a user removes a tags from his context). The higher the user's score is, the more effective the update operation will be. This means that good users have more influence on the system than bad users. Finally, β is a parameter greater than 0 and it is used to weight the user score: operation performed explicitly by users (inclusion or removal of abstract tags) have more effect than implicit update performed automatically based on the interaction of the community with the resources.

3.3.4 Inference and retrieval

The *inference* and *retrieval* operations works respectively on the first and second index, but they are similar, thus in the following we are explaining just the inference one.

The approach is the following:

1. starting from the concrete tags in input, we consider only the set of abstract tags that have been associated at least with one of the concrete tags;
2. for each abstract tag we compute a *rank* value, to define an order of relevance for the abstract tags;
3. in order to limit the number of retrievd tags, we retrieve the abstract tags whose *rank* value is higher than the mean of all *rank* values.

The *rank* value is computed following an adapted version of the tf.idf weighting scheme. In particular for each considered abstract tag a_i we have:

- $A = \sum_{c_j} \sigma_{ij} \times S_{ij}$, for each sensed concrete tag c_j
- $B = \dfrac{|C|}{|\{c :\ a_i \in c\}|}$, where $|C|$ is the total number of sensed concrete tags, and $|\{c :\ a_i \in c\}|$ is the number of concrete tags to which the abstract tag a_i has been associated;
- $rank\ value = A\alpha \times B\beta$, where α, β are parameters exploited to weight the different values.

Some considerations can be drawn. First, more are the concrete tags in the current context to which an abstract tag is associated, the higher will be its rank value. Second, abstract tags with high score and steadiness will have an higher rank value. Third, abstract tags related to particular sets of concrete tags will have an higher rank value than very general ones that are associated to an high number of concrete tags (high frequency).

In addition, starting from this basic approach, we can enhance the rank value computation exploiting other information. For example a reasonable idea is to weight the tags based on their age in the user's context representation, giving more importance to the newest tag. In this we enhance the importance of new contexts.

4. DISCUSSION

Although the conceptual ideas are clear, the implementation approach we propose is in an initial stage of definition. We suggested a possible solution, but several are the ways to refine it and several are the algorithms to be exploited. For this reason the evaluation hold an important role in our work: since different alternative solution exist, it is important to evaluate them and compare their effectiveness.

Even if the knowledge related to the whole community is exploited to infer and refine the current context of single users, the proposed model differentiates the personal from the community level, giving more importance to the first one. For example if a user annotates a situation as "`play`", she is considered to be in "`play`" context, even if most people annotate the same situation as "`work`". On the contrary, if a user is for the first time in a situation (e.g. location never visited), her context is refined just with the information from the community. Considering the previous example, as most people annotate the situation with "`work`", the user is considered to be in "`work`" context.

In the last case, the assumption performed by the system in order to provide the user with relevant resources could be wrong. However this is not a problem. Since we are working with people, it will be hardly possible to provide results that totally satisfy each user, due the intrinsic difference of views and needs in a community. Rather our solution aims at and averagely good behavior.

Talking about the indexes, we have seen how the related information are changed dynamically based on community interaction. However this is not the only possible approach. We can imagine complementary approaches that can support the community statistical one. For example, we could use some geographic gazetteer for associating geonames to geographic coordinates provided from the concrete tags, so as to reinforce the rank of associated abstract tags that contain the same geographic names or names of close localities. The geonames could be useful also for retrieving more relevant resources, those containing the geonames ore close geonames.

5. CONCLUSIONS

In this paper we have presented the Social Context-Aware Browser, a general purpose solution to Web content perusal by means of mobile devices. The sCAB is a novel approach for the information access based on context, where the community of users is called to manage the contextual knowledge, both related to situations and resources, through collaboration and participation. In particular we presented a general survey, the main ideas, and an implementation approach.

As future work we aim at implementing a prototype of the proposed system, and, in particular, we suggest a multistage approach, where implementation and evaluation processes will proceed hand in hand. As first step we want to exploit benchmarks to evaluate detailed implementation solutions, like, for example, different algorithms to assess the relevance of tags for situations and resources. After that, we plan to apply an IIR evaluation methodology, involving users in a controlled environments, following the ideas presented [1, 10]. Finally a broader user-centred evaluation will help us to understand if the sCAB is effective in the real world.

Acknowledgements

The authors acknowledge the financial support of the Italian Ministry of Education, University and Research (MIUR) within the FIRB project number RBIN04M8S8, and the region Friuli Venezia Giulia. This research has been partially supported by MoBe Ltd. (www.mobe.it), an academic spin-off company specializing in software for mobile devices.

6. REFERENCES

[1] P. Borlund. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3):8–3, 2003.

[2] A. Göker, S. Watt, H. I. Myrhaug, N. Whitehead, M. Yakici, R. Bierig, S. K. Nuti, and H. Cumming. An ambient, personalised, and context-sensitive information system for mobile users. In *EUSAI '04: Proceedings of the 2nd European Union symposium on Ambient intelligence*, pages 19–24. ACM, 2004.

[3] S. A. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Arxiv preprint cs.DL/0508082*, 2005.

[4] G. J. F. Jones and P. J. Brown. Context-aware retrieval for ubiquitous computing environments. In *Mobile HCI Workshop on Mobile and Ubiquitous Information Access*, volume 2954, pages 227–243. Springer LNCS, 2004.

[5] D. Lopez de Ipiña, J. I. Vazquez, and J. Abaitua. A context-aware mobile mash-up plaftorm for ubiquitous web. In *Proc. of 3rd IET Intl. Conf. on Intelligent Environments*, pages 116–123, 2007.

[6] S. Mizzaro. Quality control in scholarly publishing: A new proposal. *J. of the Am. Soc. for Information Science and Technology*, 54(11):989–1005, 2003.

[7] S. Mizzaro, E. Nazzi, and L. Vassena. Retrieval of context-aware applications on mobile devices: how to evaluate? In *Proc. of Information Interaction in Context (IIiX '08)*, pages 65–71, 2008.

[8] S. Mizzaro, E. Nazzi, and L. Vassena. Collaborative annotation for context-aware retrieval. In *ESAIR '09: Proceedings of the WSDM '09 Workshop on Exploiting Semantic Annotations in Information Retrieval*, pages 42–45. ACM, 2009.

[9] T. O'Reilly. What is web 2.0, design patterns and business models for the next generation of software, 2005.

[10] D. Petrelli. On the role of user-centred evaluation in the advancement of interactive information retrieval. *Inf. Process. Manage.*, 44(1):22–38, 2008.

[11] A. Schmidt. *Ubiquitous Computing - Computing in Context.* PhD thesis, Lancaster University, 2003.

[12] L. Vassena. Context-aware retrieval going social. In *3rd Symposium on Future Directions in Information Access (FDIA).*, 2009.