

Context-aware retrieval going social

Luca Vassena
University of Udine - Dep. of Mathematics and Computer Science
Via delle Scienze 206,
33100 Udine, Italy
luca.vassena@dimi.uniud.it

Abstract

In this paper we present the Social Context-Aware Browser, a general purpose solution to Web content perusal by means of mobile devices. This is not just a new kind of application, but it is a novel approach for the information access based on the users' context. With the aim of overtaking the limits in current approaches to context-awareness, our solution 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 some scenarios, presenting the model to information access, open problems and future challenges.

Keywords: context-aware retrieval, mobile search, social, folksonomy, evaluation

1. INTRODUCTION

The widespread diffusion of mobile devices and, with them, of real-world mobile users, have moved the static world of classical and Web IR towards an always changing context-based world. Dynamism and evolving situations have become the central elements of the environment where information retrieval is asked to operate. The dynamic nature of the user needs, of the information available, and of the relevance of this information, call for new approaches to application development, user-device interaction, and information seeking. So, the notion of *context* (roughly described as the situation the user is in), and the information it conveys, are gaining increasing importance for the development of new IR systems. When combined with context-awareness, IR has been named Context-Aware Retrieval (CAR) [4].

We can imagine a user seeking information on the Web. In a traditional situation, she has to manually interact with search engines, making explicit her information need into a query and filtering out the not relevant retrieved resources. If this can be acceptable in the everyday use of a desktop system, it becomes a serious issue when the same task has to be carried out in a mobile environment. These considerations guided us towards a new approach to Web contents production and use named *Social Context-Aware Browser*. The novelty of the proposed approach 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 [11]. 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 for CAR that exploit the collaborative efforts of the community of users.

In this paper we first briefly survey CAR system (Section 2). In Section 3 we present the main idea of our approach, introducing some scenarios, and underlying how current models for contextual knowledge management are not suitable for our solution. We then present our new model for the information access based on context (Section 4): we present the conceptual model and the open questions, focusing on possible methods to evaluate the model. At the end, in Section 5 we draw some conclusions and present future work.

2. 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 [7]. CAR systems are concerned with the acquisition of context,

its understanding, and the application of behaviour based on the recognized context [15]. Thus the CAR model includes, among the classical IR model elements, the user's context, that is both used in the query formulation process and associated with the documents that are candidates for retrieval.

Typical CAR applications present the following characteristics [7]: 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 [7].

A first exploitation of user's contextual information for the retrieval of documents is represented by the idea of "virtual Post-It" [3, 5]. More advanced examples are CoolTown [8], AmbieSense [13], Physical Mobile Interaction [2], where mobile devices exploit contexts to extract information and services associated with physical entities. An extension of the previous approaches is represented by the Ubiquitous Web [9], that is 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 [11]. In this application, 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.

3. SOCIAL CONTEXT-AWARE BROWSER

3.1. Description

The Social Context Aware Browser (sCAB for short) 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.

Considering the sCAB usage we can find four main scenarios of interaction, where values from sensors and resources candidates for retrieval can be enhanced with contextual information. In the first case, directly on the basis of the information provided by sensors, a resource is retrieved. For example, inside a museum, the sCAB perceives the wireless network named *Museum X* and retrieves the Web page whose content is the presentation of that museum. With a subsequent step, we annotate resources with contextual information, in order to retrieve them in the right situation. For example a Web page describing an historical fact can be enhanced with location information (GPS coordinates), to be automatically retrieved only when users are in that place. In the same way we can enhance the knowledge related to contexts, making an abstraction on the sensors' values. For example information about activities can be related to a particular combination of sensors' values, in order to retrieve resources based on the activity the user is doing: when the user is taking the dog out for a walk, Web pages that speak about dogs training are retrieved. In the last and more general step, we can imagine a user in a museum: when she is near an artwork, a detailed description is presented on her device. In a crowded situation, on the contrary, a detailed description is not useful as users can have difficulties in seeing the paintings, thus a navigable high resolution picture can be more interesting. In this case, both resources and sensors' values are enhanced with contextual information.

In a such general and large scale approach, contextual knowledge is continuously associated (added, removed, and modified) with resources and low level sensors information. This entails the creation and management of a huge amount of knowledge related to contexts. Thus, the main question is to understand who is the provider of that knowledge and how it has to be defined.

3.2. Social approach

In current approaches to context-awareness, that manage several dimensions of context, the knowledge is usually provided by a small group of experts (application developers or specific domain experts). This is due to the difficulties in representing contexts. In fact, in order to fully capture the concept of context, these approaches are based on categorizations and ontologies, and implicates the strict definition of the contextual information. Moreover contexts are defined a priori, and there is no way to dynamically extend the contextual values adopted or to enhance their representation at run-time: the operations of modeling contexts and using context-aware applications are rigidly separated. This is the reason why current approaches show a trade off between the generality of applications and the depth of context representations. Applications that fully manage several contextual dimensions are confined to limited fields (e.g. Smart Homes), while general applications work only on a narrow notion of context (e.g. in location-based applications the context is represented just by location and time).

The high generality and the deep context representation we aim at with the sCAB, require both a dynamic nature and a huge amount of information to be categorized and modeled (to represent both contexts and contexts-resources associations). For these reasons, current approaches are not suitable for the sCAB.

Starting from these considerations, we propose a novel model for CAR, that aims at overtaking the just defined limits, exploiting the social dynamics underlying the Web 2.0. In fact we believe that only the collaborative effort of a community can provide the right tool for a comprehensive definition, management and use of context, in an open architecture as the sCAB. In particular, we do not want a priori contexts definitions made by experts, and we do not want people to be just passive users. Rather, through collaborative annotation, the community of users is encouraged to define the contexts of interest, share, use and discuss them, associate context to content (web page, applications, etc.), to have a dynamic and more user-tailored context representation and to enhance the process of retrieval based on users' actual situation.

4. PROPOSED APPROACH

4.1. Conceptual model

Users, contexts and resources are the main elements of our model (Fig. 1). A *user*, in the real world, is engaged in some activities, she has some needs, and she perceives her surrounding environment through her senses and the sensors on her mobile device. *Contexts* are virtual representations of the users current situations. *Resources* represent every kind of content that could be useful for the user to accomplish her needs.

Instead of using rigid categorizations built upon ontologies and terminologies, we represent the context by means of a folksonomy. This representation allows an easy and informal modelling of context, giving the opportunity also to non-expert users to classify and find context-related information. Thus each context is represented by a tag cloud and the tags (little squares in Fig. 1) are socially *defined* by the users themselves. The tag-context association is done both explicitly, when the tags are added directly by the user, and implicitly, when the tags are derived from the interaction of the community with resources.

Six are the main operations in the model [12].

Contexts definition: users can explicitly use tags to represent contextual information. For example, the user can enhance the values provided by sensors on the mobile device ("concrete" values) with her own tags, as `'out dog walk leash park play ball'`. Doing so these "abstract" tags are stored in a remote repository and they are linked with the concrete ones. For all the users with the same or a similar concrete tag cloud, the abstract tags (or part of them) can be retrieved and become part of the representation of their context.

Contexts inference: the values provided by sensors are combined with the contextual tags defined by the whole community, and tags that best describes the user's situation are retrieved. For example, starting from the GPS coordinates, the current user's context could be enhanced with the tags `'walk sunny park'`.

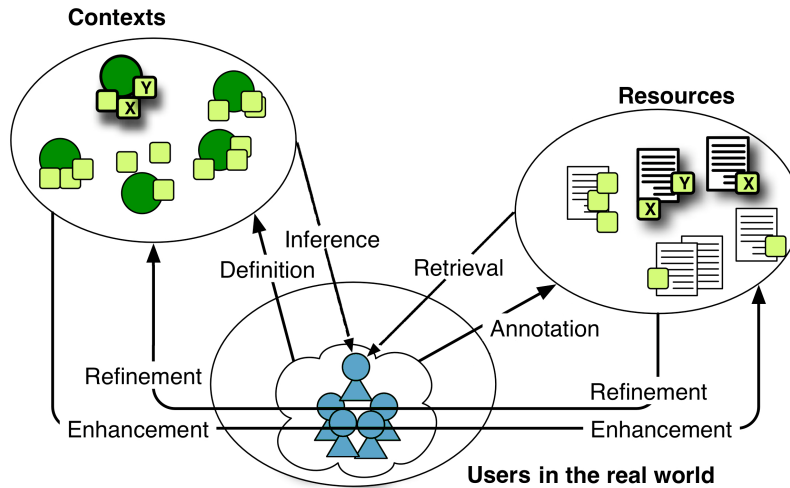


FIGURE 1: A conceptual model for social CAR

Resources annotation: users can explicitly annotate resources with contextual information to allow the community of users to automatically retrieve them when they are in the suggested context. For example a user can associate a classical music web-radio with the context ``out dog walk``, in order to listen to music when she is out with her dog.

Contextual retrieval: based on users' current context and on the contextual information associated to resources, the most relevant resources are retrieved.

Refinement: information about contexts are refined based on the interaction between users and resources. For example, if a lot of users work with a Web application in the same context ``work``, probably this resource is related to that context and it is automatically annotated with it.

Enhancement: information related to resources are refined based on the interaction of users within their contexts. For example if a user uses resources annotated with the context ``work``, probably she is working, and the representation of her context can be enhanced with this information.

Although 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.

4.2. Issues and open questions

Several are the issues related to the proposed approach: they go from implementation/system/architecture issues, to telecommunications issues, from annotation/Web 2.0 issues, to interface/HCI, to evaluation issues. The proposed work do not address all these, but focuses on the annotation, Web 2.0 and evaluation issues. The main goal is to improve the effectiveness of the Context-Aware Browser, supporting and enabling the effective and efficient access to relevant information. Thus we want to understand if the exploitation of the crowdsourcing is viable for the management of the concept of context in a general approach as the sCAB.

In particular several critical questions can be found in the presented approach. Starting from a low level, the questions are related for example to context representation: it is better to exploit a simple folksonomy or to introduce a level of complexity (e.g. logic operator over tags)? Tagging a tag can introduce useful information? Tags should be associated with probability values to describe the uncertainty related to contexts?

Other interrogative points are related to the association between tags and the elements in the model. New tags are continuously added to situations and resources, leading to a huge amount of tags for each of them. Thus the question is how to understand, given a resource or situation, which are the most relevant contextual tag. All, just the last ones used, or have we to consider the number of times a tag has been related to a resource? Or rather we can imagine a more complex approach based on social evaluation, where every association (e.g. tag–resource) has a score that increases or decreases based on the community behaviour and on how users are representative for the community. Could this approach improve the relevance computation or does it introduce a useless layer of complication?

Once we have identified relevant contextual tags, which is the best strategy to combine the user’s context tags with the ones associated to resources, to retrieve the most relevant resources for the given context? How can the system extrapolate contextual knowledge from the simple amount of tags provided by users? Can current machine learning and artificial intelligence techniques cope with this problem? As the users are the main actors in the process of contexts definition, how can we ensure the quality of the information provided? Moreover how can we easily manage the problem of communities and subcommunities?

Finally there are some general questions. This vision presents an extension of the idea of Web, where resources and documents are indexed not only based on their content, but also based on the context which they are relevant to. Could this model be helpful in understanding the relationships between context and content, and how do they reflect one to each other? Which is the connection between context and information needs? Is the context alone enough to understand what the user needs and how she needs it (textual information, audio, image)?

4.3. System evaluation

Although the conceptual model is clear, several alternative strategies to the problems presented exist, so it is important to compare their effectiveness. Moreover the understanding of the good and weak points of the proposed model is the first step toward its realization.

4.3.1. Evaluation approaches

Evaluation is an issue of paramount importance in IR. Depending on resources and aims, different evaluation approaches can be adopted; benchmark-based and user-centred are the main ones. The benchmark evaluation (e.g. TREC initiative, <http://trec.nist.gov/>) follows the Cranfield model, giving emphasis on controlled laboratory tests, without user interaction. Benchmarks are system centred and they directly focus on the evaluation of different implementation strategies. Within the CAR field, benchmarks have been for example exploited to have a rough evaluation during the first stage of development [10, 11].

While benchmarks concentrate on details, user evaluation approach works on the IR system from a much broader point of view. The main purpose is to study usability and interaction, to understand how the system satisfy the user’s needs, and, in general, how well the user, the retrieval mechanism, and the database interact extracting information under real-life operational conditions [1]. Within the CAR field, this kind of evaluation has been for example exploited in [6, 16].

In the last years hybrid evaluation models have been studied to combine the advantages of both the previously presented approaches. These ones are mainly addressed to improve the evaluation in the interactive information retrieval field (IIR). The overall purpose is to facilitate the evaluation of IIR systems as realistically as possible, taking into account the dynamic natures of information needs and relevance as well as reflects the interactive information searching and retrieval processes (as in a user-based study), though still in a relatively controlled evaluation environment [1].

4.3.2. Suggested methodology

With this considerations in mind, we propose a multistage approach, where implementation and evaluation processes will proceed hand in hand. Different stages of work will require different evaluation solutions, and we believe that early stage benchmark evaluations, followed by user studies, is an effective methodology to

be applied to systems like the sCAB. We will move from pure laboratory studies, to simulated environments with users involvement, to a real world operational environment.

Benchmark evaluation will be our first step, and will be exploited to evaluate detailed implementation solutions, like, for example, different algorithms to assess the relevance of tags for situations and resources. Benchmarks do not substitute the user testing evaluation. Rather, several early stage benchmark experiments could provide more solid basis for the subsequent user testing, that can thus be more focused. For example, knowing which is the best strategy allows us to give to users just one prototype, instead of different prototypes, one for each strategy.

Once the system has attained the desired levels of accuracy and effectiveness, we can apply an IIR evaluation methodology, involving users in a controlled environments, following the ideas presented [1, 14, 16]. In particular this step will consist in an iterative process based on the design-evaluation cycle described in [14]: starting from some hypothesis, we will build/refine a prototype, that will be evaluated, and the results will be the basis for the next cycle hypothesis.

Finally a broader user-centred evaluation will help us to understand if the sCAB is effective in the real world. This last stage does not involve laboratory experiments anymore, but only studies in an operational environment. In this stage, we must move beyond performance and usability and consider utility or impact measures [16]. That is, how do the proposed system change the work that users are doing?

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. Introducing the main ideas, we have shown how current approaches to contextual knowledge management are unsuitable for our solution. Thus the sCAB is not merely an application, but it is a novel paradigm for the information access based on context, where the community of users is called to manage the contextual knowledge through collaboration and participation. We presented some scenarios and the conceptual model, suggesting a possible way to evaluate it.

The project is an initial stage. As future work we aim at answer all the open questions, advancing in the same time with the evaluation of precise aspects of the model and with its implementation.

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] G. Broll, S. Siorpaes, E. Rukzio, M. Paolucci, J. Hamard, M. Wagner, and A. Schmidt. Supporting mobile service usage through physical mobile interaction. *IEEE Intl. Conf. on Pervasive Computing and Communications*, pages 262–271, 2007.
- [3] P. J. Brown, J. D. Bovey, and C. Xian. Context-aware applications: from the laboratory to the marketplace. *IEEE Personal Communications*, 4(5):58–64, 1997.
- [4] P. J. Brown and G. J. F. Jones. Context-aware retrieval: Exploring a new environment for information retrieval and information filtering. *Personal and Ubiquitous Computing*, 5(4):253–263, 2001.
- [5] J. Burrell, G. K. Gay, K. Kubo, and N. Farina. Context-aware computing: A test case. In *Proc. of the 4th Intl. Conf. on Ubiquitous Computing (UbiComp '02)*, pages 1–15, London, UK, 2002. Springer-Verlag.
- [6] A. Göker and H. Myrhaug. Evaluation of a mobile information system in context. *Information Processing & Management*, 44(1):39–65, 2008.
- [7] 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.
- [8] T. Kindberg, J. Barton, J. Morga, G. Becker, I. Bedner, D. Caswell, P. Debaty, G. Gopal, M. Frid, V. Krishnan, H. Morri, C. Pering, J. Schettino, B. Serra, and M. Spasojevic. People, places, things: Web presence for the real world. In *3rd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2000)*, volume 7, 2002.

- [9] D. Lopez de Ipiña, J. I. Vazquez, and J. Abaitua. A context-aware mobile mash-up platform for ubiquitous web. In *Proc. of 3rd IET Intl. Conf. on Intelligent Environments*, pages 116–123, 2007.
- [10] S. Menegon, Davide Mizzaro, E. Nazzi, and L. Vassena. Benchmark evaluation of context-aware web search. In *Workshop on Contextual Information Access, Seeking and Retrieval Evaluation (CIRSE) in conjunction with the 31th European Conference on Information Retrieval*, 2009.
- [11] 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 (IiX '08)*, pages 65–71, 2008.
- [12] 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.
- [13] H. Myrhaug, N. Whitehead, A. Göker, T. E. Faegri, and T. C. Lech. Ambiesense—a system and reference architecture for personalised context-sensitive information services for mobile users. In *Proc. of the Second European Symposium on Ambient Intelligence (EUSAI '04)*, 2004.
- [14] D. Petrelli. On the role of user-centred evaluation in the advancement of interactive information retrieval. *Inf. Process. Manage.*, 44(1):22–38, 2008.
- [15] A. Schmidt. *Ubiquitous Computing - Computing in Context*. PhD thesis, Lancaster University, 2003.
- [16] J. Scholtz. Metrics for evaluating human information interaction systems. *Interacting with Computers*, 18:507–527, 2006.