A Two Layered personalized Information Retrieval System

Hassan NADERI, Béatrice Rumpler, and Jean-Marie Pinon

Abstract— **Providing personalized information becomes an important challenge. More and more users require that Web or distributed services can be tailored to their needs. Development of wireless technology increases the use of handheld devices. Therefore, one need is that mobile users can use their handheld devices anywhere to access to digital documents. However, some constraints such as, small screen size; narrow bandwidth of connectivity; short duration of battery, limit the displaying of a long document. This means that, on one hand, user's preferences and goals must be considered to find the most pertinent documents. On the other hand, contextual information like devices capabilities must also be considered to satisfy user's need.**

In this paper we present a two layered system which is able to consider these two aspects.

Index Terms— **context modeling, information retrieval, small screen devices, user modeling.**

I. Introduction

TODAY providing personalized information is an important challenge. More and more users require that the content services can be tailored to the user's needs. Development of wireless technology increases the use of handheld devices. Therefore, one major need is that mobile users be able to use their handheld devices to read some relevant information. However, some constraints such as, small screen size; narrow bandwidth of connectivity; short duration of battery, limit the displaying of a long document [1]. For example at present, wireless's bandwidth restricts downloading time of documents from server. In such a situation we have categorized the information which can be helpful to satisfy the users, in three main groups (fig. 1): User related information, Document related information, and Context related information. By exactly modeling the user's characteristics and the document's properties, we can more efficiently find the most appropriate documents for the users. The context parameters in this matching process could be helpful in order to construct a good representation of information to display on the user's device.

All authors are with the LIRIS, INSA of LYON, Bâtiment Blaise Pascal, 7, Av. Jean Capelle, F69621 Villeurbanne Cedex, France (e-mail: {hassan.nadery, beatrice.rumpler, jean-marie.pinon}@insa-lyon.fr).

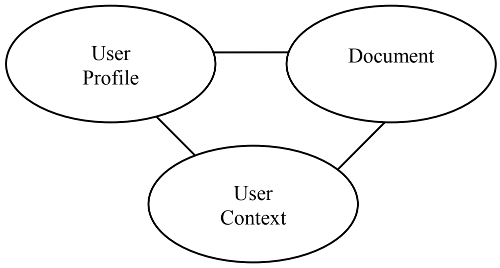

Fig. 1. Information retrieval based on three main information classes.

This means that, on one hand, contextual information likes devices capabilities must be considered to satisfy user's need. By getting the context characteristics we can reduce the complexity and the length of texts and then extract the most important information suited to mobile users. On the other hand, summarization technique plays an important role to provide brief information. In this paper, we integrate the user's context into text summarization process to provide reduced information for mobile users [2]. Our principal objective is to model user's context and integrate it to the summarization process to provide specific information to mobile users according to their context and their personal preferences.

As shown in fig. 2, our system is based on two layers: the document retrieval layer and the information restitution layer.

The document retrieval layer finds the most pertinent documents based on the user's query and the user profile (personalized document retrieval)[3].

The information restitution layer extracts the most appropriate information from the selected documents based on the user's context in order to adapt the document representation. So we deal with two modeling systems: the user modeling and the context modeling. In what follow we will present the user modeling and the context modeling respectively.

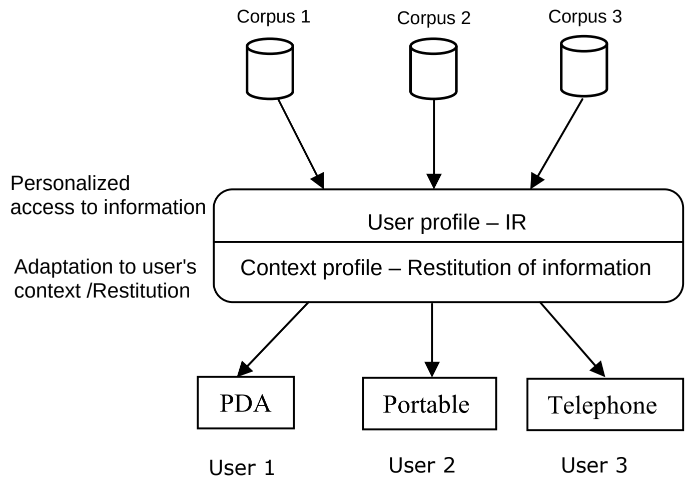

Fig. 2. Two layered information seeking system.

II. USER MODELING

User modeling is a process including the acquisition and formalization of user knowledge and preferences, but also the construction, the exploitation and the evolution of the user model. Without a user model a system will perform exactly in the same way with all users, because there is no basis to behave in a different manner. But users are different: they have different background, different knowledge about a subject, different preferences, goals and interests. To individualize, personalize or customize actions, a user model is needed to perform the selection of individualized responses. Today user modeling plays an important role in Information Retrieval systems: “personalized IR is a direction of research on the crossword of IR techniques and user modeling.” [4].

In our research we have defined a typology for the user profile based on six dimensions:

Personal data. The "personal data" dimension includes the user's identity information (name, first name, date and birthplace, etc). This factual information is independent of the retrieval system. It is also relatively stable in time: some are perennial (name, first name, sex, date of birth) while others evolve according to a rather weak dynamics (matrimonial situation, age, spoken languages, handicaps...). They constitute long term user characteristics.

Domain of interest. The "domain of interest" dimension expresses the general characteristics of information the user wishes to obtain. These characteristics are generally described using suitable languages of representation. In some cases, the domain of interest is described by the user's selected documents. Finally, the “domain of interest” dimension covers the data closed to the user interest.

Quality. The quality dimension gathers a part of the constraints the user can express. It includes all requirements relating to the data and the sources providing the data, such as the reliability of the sources, the freshness of the data, and the author reputation about a document according to the context and the application.

Preferences of delivery. This dimension covers the flexible constraints for the restitution of information to the user. The existence of preferences of delivery induced that the user or the information system is conscious of a particular situation about the context (particularly the device). For example in some cases, the restitution of information can be blocked if an adequate adaptation of the information is not possible. The particular situation is sometimes restrictive (for example, reduced dimensions of a mobile screen) or advantageous (for example, an important band-width). Constraints expressed in this dimension can have a strong impact on the process of data selection.

Security/Safety. It includes all information about the privileges or the secret concerning the data, the user or the treatment. It can be involved at several levels of the information treatment: during the request process, during the treatment of this request (for example for an access control to the resource of system), and during the information delivery process.

History of execution. Historical dimension is distinguished from other dimensions in the sense that it evolves the user profile. Information of this dimension describes the interactions between the user and the system. This dimension can be used to manage the evolution of the user domain of interest.

III. CONTEXT MODELING

To facilitate the programming of context-aware applications we need to categorize the context fields [5]. The categorization helps application designers to select which context field will be useful in their applications. Nowadays some works about context classification are done. [6] suggests four types of context: computing device, user characteristics, physical environment and temporal. The categories provided by [7] are location, environment, identity and time. In our proposition, we divide contextual information into three categories (fig. 3):

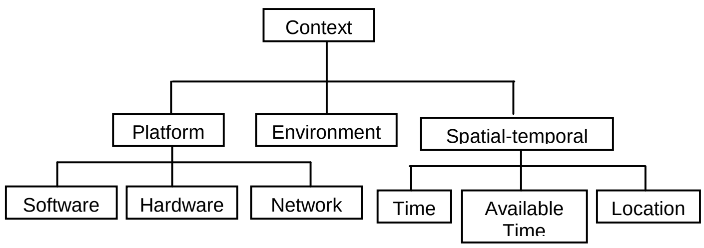

Fig. 3. Context profile

Platform context: divided in three sub-categories:

- Device capabilities: they are basic hardware feature such as screen size, memory capacity, etc.
- Network properties: most of these contexts frequently change so they must be continuously known by the system to

use the recent value such as bandwidth, protocol, etc.

- Software components: such as navigator browser, virtual machine, etc. They must be considered to respect software platform of devices. For example, if a mobile device has no PDF reader, the document must be transformed to another format the device can understand, like WML for example.

Spatial-temporal context: which contains the information about user's current location and time, it includes two parts:

- Spatial: user's absolute address such as coordinates (x,y) and relative address such as street name.
- Temporal: user's available time, time zone etc. For instance, if a user is very busy, he only wants some key words as answer.

Environment context: In ubiquitous computing, a user is not a single entity. He must react with his environment. This environment consists of physical information and conditions which can influence the user's behavior. For example, users don't accept to read a long document in a noisy environment.

IV. THE SYSTEM'S ARCHITECTURE

In this section we will describe the architecture of our two layered system. As we have mentioned in the previous sections (and it has been shown in the fig. 4), our system is based on two separated components: Content adaptation and Context adaptation.

The Content adaptation component maintain the user profile (which we have defined in the previous sections) in order to find the more pertinent document to the context component. This component has been shown by a dashed rectangular on the right of fig. 4. When a user connects to the system for the first time, he must spend some time to supply some personal information to fill out some parts of his profile such as: personal data, quality, preference of delivery and security/safety. Subsequently whenever the user connects to system, the content adaptation component records the user's interactions such as the user's queries and the user's feedback in order to update the two other parts of the user profile such as: the domain of interest and the history of execution. When a user sends a query to the system, this component find the most appropriate information's resources (corpus of documents) based on the quality's section in his profile. Content Decision Engine helps this component to find the most pertinent documents by providing a list of concept–weight pairs form the user profile. The output of this section is a pertinent document which will be supply to the context adaptation section.

The purpose of context adaptation component is to construct a good representation of information (document) based on the user's context such as the network parameters, user's device, user's availability time and etc. When a user connects to the system, the system automatically records his context's characteristics (such as network's speed, screen size,

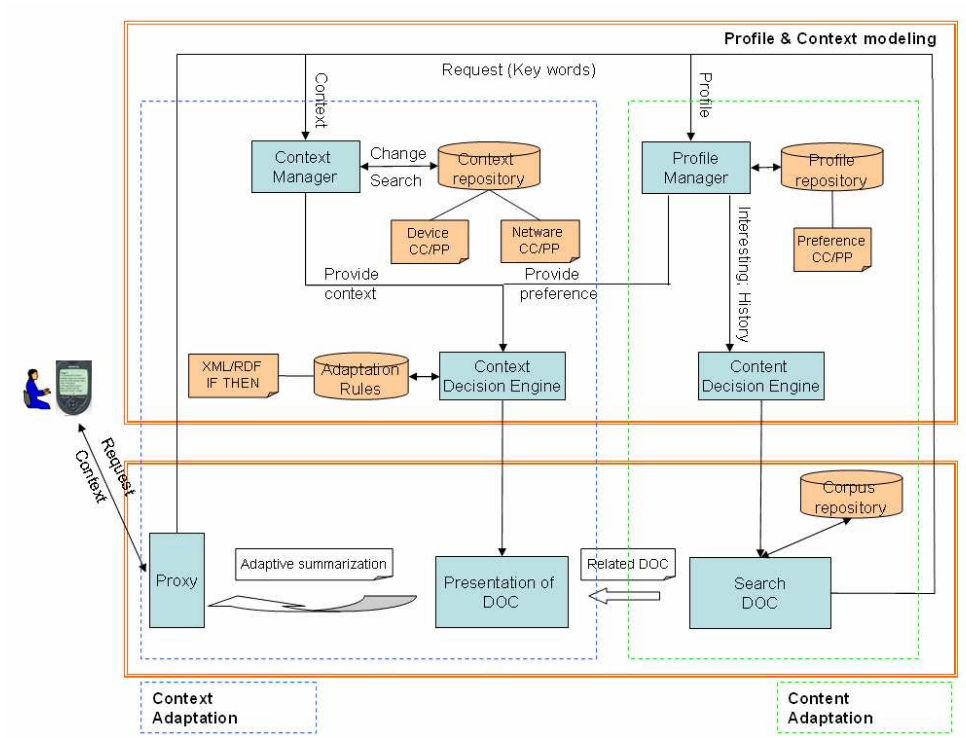

Fig. 4. The two layered personalized information retrieval system.

memory) in the platform context part of the context model in the format CC/PP [8,9].

CC/PP stands for Composite Capabilities/Preferences Profile, and is a system for expressing device capabilities and user preferences. With CC/PP, a user with a specific preference or disability-related need can clarify that even though their browser handles millions of colors, they personally can only distinguish certain colors. Or, perhaps the user navigates exclusively with a keyboard or stylus.

The aim of context manager is to monitor the user's context and update the context repository. Context decision engine [10] is a rule based system which calculates the size of the output based on the user's restitution preferences and the user's context. Document Presentation takes the pertinent document from the Content Adaptation Component and produces the final version of the document in order to represent on the user's device, based on the supplied information from Context Decision Engine. Document Presentation parses the document and extracts its feathers and then reduces the document's size by getting the appropriate parameters form the Context Decision Engine.

The Context Adaptation component has been represented in a vertical rectangular on the left of the fig. 4.

V. Conclusion

In this paper we have described the core of our two layered system in order to represent the most pertinent information to user, based on his preferences and his context. We have proposed a two layered architecture system which separates the content adaptation component form the context adaptation component in an IR system. In our system we categorized the user information in six main classes of information: personal data, domain of interest, quality of resource, preference of restitution, safety/security, and history of execution. These components can be statically or dynamically filled out in different phases of the system's execution. In our system we have divided the context characteristics in three classes: platform, spatial-temporal and environment. We didn't deal with the environment characteristics in order to simplicity. This system is enabled to recognize the preference and the context of the user and adapt the information to the user's context.

VI. Acknowledgment

This research is partially supported by the French Ministry of Research and New Technologies under the ACI program devoted to Data Masses (ACI-MD), project #MD-33.

VII. References

[1] G. Chen and D. Kotz, A Survey of Context-Aware Mobile Computing Research, Department of Computer Science, Dartmouth College, 2000.

[2] T. Lemlouma, N. Layaïda, Adapted Content Delivery for Different Contexts, SAINT 2003 Conference, Orlando, Florida, USA, January 27-31, 2003. IEEE Computer Society Publication.

[3] Naderi H. and Rumpler B., PERCIRS: a PERsonalized Collaborative Information Retrieval System, to be appeared in INFORSID, Tunisie, March 2006.

[4] Croft W.B., Approaches to intelligent information retrieval, Information Processing and Management, 23 (4): 249-254, 1987.

[5] Anind K. Dey and Gregory D. Abowd Towards a Better Understanding of Context and Context-Awareness, Proc. 1st Int'l Symp. Handheld and Ubiquitous Computing (HUC 99), Lecture Notes in Computer Science, no. 1707, Springer-Verlag, Heidelberg, Germany, 1999.

[6] G. Chen and D. Kotz, A Survey of Context-Aware Mobile Computing Research, Department of Computer Science, Dartmouth College, 2000.

[7] Ryan, N., Pascoe, J., Morse, D. Enhanced Reality Fieldwork: the Context-Aware Archaeological Assistant. Gaffney,V., van Leusen, M., Exxon, S. (eds.) Computer Applications in Archaeology 1997.

[8] Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies, W3C Working Draft, March 15, 2000. http://www.w3.org/TR/2001/WD-CCPP-struct-vocab

[9] Resource Description Framework (RDF) Model and Syntax Specification1.0, W3C Candidate Recommendation, Mar 27, 2000. http://www.w3.org/TR/rdf-schema/

[10] W.Y. Lum and F.C.M. Lau, A Context-Aware Decision Engine for Content Adaptation, IEEE Pervasive Computing, Vol. 1, No. 3, July-September 2002, 41–49.