PHYSICAL DOCUMENT ADAPTATION TO USER'S CONTEXT AND USER'S PROFILE

Hassan Naderi and Béatrice Rumpler
INSA LYON, Bâtiment Blaise Pascal, 7, Av. Jean Capelle F69621 Villeurbanne Cedex, France

Keywords: Mobile device, Document adaptation, Physical Document Adaptation (PDA), Context, User's delivery preferences.

Abstract: Modern technology promises mobile users Internet connectivity anytime, anywhere, and using any device. However, given the constrained capabilities of mobile devices, the limited bandwidth of wireless networks and the varying personal preferences, effective information access requires the development of new computational patterns. The variety of mobile devices available today makes device-specific authoring of web content an expensive approach. The problem is further compounded by the heterogeneous nature of the supporting devices and the users' behaviour. This research investigates the challenges posed by these problems, and proposes a context-aware adaptation framework to bridge the gap between the existing Internet content and today's heterogeneous computing environments.

1 INTRODUCTION

The rapid advances in wireless communication technologies with their integration into mobile devices such as mobile phones, PDA, and portable computers provide a technical infrastructure that enables a broad range of mobile services and applications. These devices are designed to help people access and to manage information. While they provide greater access to online information, they also suffer from significant limitations such as small screen-sizes, slow connection, and limited input capabilities. A typical cell phone screen is 200 times smaller than a standard PC monitor. This imposes limits on one's ability to locate and display the right information quickly at the right time, and therefore the information overload problem itself becomes more serious. Because these access devices are becoming more and more popular, there is a need to make the next generation of information retrieval systems capable of automatically adapt the found document to the *user's situation* (cf. § 4, Figure. 1).

In section 2, we study the actual approaches to adapt a document to a given situation. In section 3 we present our proposition to adapt a given document to the user's situation. In section 4 we define the concept of user's situation. In section 5 we show how the user's situation influences the process of Physical Document Adaptation (PDA). In order to simplify and clarify the process of PDA we have categorized the documents' components in two main classes: structural components and logical components (section 5). In section 6 we propose a formula to estimate the optimal size of the adapted document which is based on the user's situation.

2 RELATED WORKS

Since now, there have been great efforts to provide a unified web which is accessible from many types of devices. Device independent access to information has been identified as a fundamental principle of the W3C Device Independent Working Group (Gimson 2003). Several approaches for providing adaptive web content delivery over heterogeneous and dynamic environments have been investigated. Often in these approaches, *markup languages* play a major role to provide an adapted content for different contexts. Some of these approaches try to extend existing languages and content models with new elements and attributes (Lewiz 2002, Mandyam 2002). The other approaches try to define new languages which are completely device independent (Grassel 2002). Their objective is to build a new

Naderi H. and Rumpler B. (2007).
PHYSICAL DOCUMENT ADAPTATION TO USER'S CONTEXT AND USER'S PROFILE.
In *Proceedings of the Ninth International Conference on Enterprise Information Systems - HCI*, pages 92-97
DOI: 10.5220/0002384300920097

markup language that can be used to create the new generation of documents which can be displayed on various devices in different contexts. These approaches use a new tag set to highlight the most pertinent parts of a document according to different contexts. These approaches have some limitations: 1- They require some efforts from the author and they are time-consuming. 2- They produce different versions of a document for different contexts. Therefore finding an appropriate version of a multi-version document to a given context is a difficult task because there is not an ideal method to match the user's context to the author's defined document's context. 3- The author has to produce the different versions of a document for each context. As there are a lot of contexts, it's unlikely to produce a specific version of a document for each context. 4- It may be feasible for the next generation of documents but as we have mentioned there is a serious need to adapt the actual documents (Kurz 2004).

We believe that the user's situation can be used to roughly determine the characteristics of the adapted document. In this paper, we propose a mechanism to automatically adapt a given document based on the user's situation. For example, if the user's situation indicates that the user has a mobile device unable to display the pictures, then the system must eliminate the document's pictures

3 DOCUMENT ADAPTATION

Indeed, our adaptation system is based on two types of adaptations: Physical Document Adaptation (PDA) and Semantic Document Adaptation (SDA). By separating the PDA from the SDA, it will be easier to build a document suited to the user's situation. In the two following paragraphs we will briefly explain each of these adaptation types and then the PDA will be elaborated in this paper.

SDA answers to this question: which textual parts of a document must be selected to be displayed on the user's device to semantically satisfy the user's request? The SDA approaches are similar to the document summarization methods. In this step, the wasted information will be eliminated from the original document, based on the user's situation and his query in order to enrich the document's content.

PDA is the process of selection, modification or elimination of the physical components of the document (cf. §5) based on the user's situation. Recognizing the size of the adapted document (the total size of the selected components to create the adapted document) plays an important role in a successful adaptation. In other words, a competent adaptation system must be able to estimate the size of the adapted document according to the given situation.

The size of an adapted document is a *dynamic* value related to the user's situation. More the user's situation is restricted more the size of adapted document must be reduced. An adaptation system which is not able to estimate the size of an adapted document may generate a too long or too short document. For example if the system sends a lot of information to a mobile user who is time restricted, he can't efficiently find what he wants. Inversely, he will not be satisfied when the system sends him insufficient information.

In what follow we explain our methodology to create the *textual part* of adapted document. The steps are:

1- Extracting and sorting the pertinent phrases based on the user's requirements.

2- Estimating the size of the textual part of the adapted document based on the user's situation.

3- Calculating the number of phrases (N) which must be displayed on the user's device.

4- Selecting the N most pertinent phrases form the sorted list of phases.

5- Displaying N selected phrases to user *by respecting their order in the original document*.

Knowing the number of phrases to display on the user's device is essential to respect the order of the phrases in the original document. We propose a formula in section 6 to estimate the optimal size of the adapted document.

So far it has been shown that the *user situation* has an important role in document adaptation. In what follows we describe our definition of *user's situation*, which is essential in the PDA process as well as for estimating the optimal size of an adapted document.

4 USER'S SITUATION

In most of the proposed systems, the user's preferences and the context's characteristics (such as user's device) are grouped in the same category. In such systems, if a user uses two different devices, the system will associate him two different profiles. This problem is because the context includes the user's characteristics.

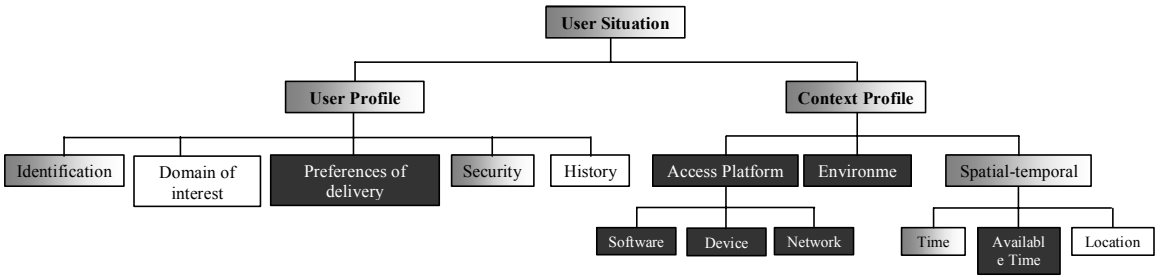

Figure 1: User's situation components.

In order to solve this problem in our system, we propose to separate the *user profile* and the *context profile*. Thus in our system both of them characterize the user's situation (Figure 1). The parameters in Figure 1 are represented by three different colours to illustrate their different roles in the semantic or physical adaptation processes. The white elements are related to the SDA, the black elements are related to the PDA, and the components that can be used for SDA and PDA (i.e. user's personal data) are shown by the grey-white elements. Thus according to Figure 1, SDA is roughly based on the user profile and PDA is based on both the context profile and the user profile (Figure 2).

As PDA is affected by both user and context profiles, sometimes some conflicts may appear between the user's delivery preferences and the user's context. For example, a user may want to see a picture (preference) but his screen (context) is not able to display it. So a decision engine has been designed in our system to solve such conflicts between the user's context and the user's delivery preferences (Lum 2002, Naderi 2006a).

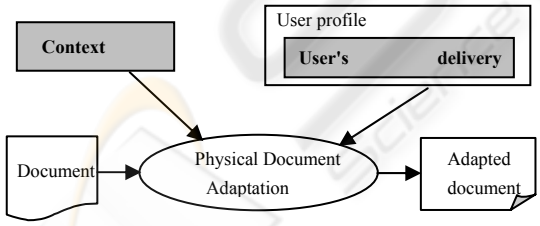

Figure 2: PDA based on context and user's preferences.

As mentioned, adapting a document is based on the user's situation. The more the user's situation is restricted the more the original document must be adapted. So characterizing the user's situation plays an important role for adaptation systems. In what follows we will explain the main parts of the user's situation: the context profile and the user profile.

4.1 Context Profile

The context's components include the user's device, the network, the user's available time, etc. (Figure 1). These components impose restrictions on the process of PDA. For example a mobile device may be unable to display a picture because of the network's speed or because of the device's capabilities. As it has been shown in Figure 1, the only element from the context profile which is related to SDA is ***Location***. It means that this element is a SDA related component and it has no influence on PDA process.

In our proposition, we separate the contextual information into three main categories:

Access platform: divided in three sub-categories:
- Device: represents the basic hardware features such as screen size, memory capacity, etc.
- Network: such as the network's price, bandwidth, reliability, etc.
- Software: such as available browser, virtual machine, etc.

Spatial-temporal: includes three parts:
- Available time: user's availability time.
- Time: represents the current time and date.
- Location: user's absolute or relative address.

Environment: the conditions which can influence the user's behaviour (noise...).

In the section 5, we will explain how a document can be physically adapted, based on the black elements (Figure 1).

4.2 User Profile

The user profile is another component of the user's situation, which plays an important role in the process of the SDA. The main role of user profile is to help the adaptation system to semantically adapt a given document to the user's situation. Without a user model, a system works exactly in the same way with all users. But users are different: they have different background, different knowledge about a subject, different preferences, goals and interests.

The user's interactions will be registered in the user profile in order to recognize his preferences and domain of interest. As it has been shown in Figure 1 "user's delivery preferences" is the only black element from the user profile which affects the process of PDA. The user's preferences will be recorded in this component to more properly display a given document on his device. For example, a user may indicate in his profile that he doesn't want to see any pictures on his device, or he just wants to see the title and a small summary of the document.

We have defined a typology of the user profile based on five dimensions:

- ***Identification:*** includes the user's identity information such as name, job, etc.

- ***Domain of interest***: expresses the characteristics of information that the user wishes to obtain. It affects the process of SDA.

- ***Preferences of delivery:*** It covers the flexible constraints for restitution of information to the user. PDA depends on this part of the user profile.

- ***Security:*** includes all information about the security and privacy such as login/password.

- ***History:*** manages the interactions between the user and the system. It can be used to find pertinent information to users (Naderi 2006b).

5 PHYSICAL DOCUMENT ADAPTATION

In the previous section we have introduced two profiles which define the user's situation. Both of them have an important role in the PDA process. In this section we study the different possibilities for the document adaptations to the user's situation.

As we have mentioned before, PDA means selecting, adapting and/or eliminating document's components (or fragments) to satisfy user's and context's restrictions. Defining the document fragmentation is an important part of PDA. The fragmentation is used to adapt the document to different contexts. Once document fragmentation is defined, the system can easily use the different components according to the user's situation (context + user's preferences).

To simplify the PDA process, we have defined two types of components: *structural components* and *logical components*. According to these document's components, we have defined two types of PDA's respectively: *structural document adaptation* and *logical document adaptation*. In what follows we describe these two types of PDA.

5.1 Structural Document Adaptation

The structural components include the objects such as: texts, figures, tables, diagrams, formulas, etc. Different devices can display the different structural components with different properties. For example a device can display simple tables but it can't display complex pictures. By having the device's characteristics (in the context profile) such as the screen size, screen resolution, number of colours etc; the system can recognize the device's ability to display a particular component. The aim of the system, in the structural document adaptation, is to select or modify the structural component. If the adaptation is not possible the system will eliminate the component before displaying the document.

The information concerning the ability of a device to display a particular component can be automatically extracted from the device's characteristics or it can be manually registered in the device's profile by the user. By referring to the device's profile, the system determines whether a component can be represented, or must be modified or eliminated from a given document.

Beside the context's role, to select, modify or eliminate the structural components; the user himself can specify his preferences to display the different structural component. The user, by specifying his preferences in his profile, enables the system to select the appropriate components to construct an appropriate adapted document. Therefore the system will be able to automatically select, modify or eliminate the structural components from the original document by referring to the device's profile (a part of context's profile) and the user's delivery preferences (a part of user's profile).

5.2 Logical Document Adaptation

In this step, the structural document adaptation subsystem sends a partially and structurally adapted document to the logical document adaptation subsystem. The logical components are the textual parts of a document such as title, authors, abstract, keywords, sections, subsections, conclusion, references, document's summary, etc. The context profile has no impact on the selection of the logical components. Only the user can specify his preferences about the representation of the logical components. For example a user may only be interested in reading the title and some keywords of a document. Another user may be interested to have an automatic summary of the document.

Consequently the system can define the appropriate document modifications by referring to the user profile (structural and logical adaptation) and the context profile (structural adaptation). Therefore roughly all of the possible document's modifications (structural, logical and semantic) can be found in the user profile and in the context profile. Thus by having enough information about the situation, a document can be *automatically* adapted to the user situation. The size and the content of the generated document are the only important elements which can't be directly determined from the two mentioned profiles. In what follows, we propose a formula to estimate the optimal size of an adapted document, based on the user situation.

6 SUMMARY'S SIZE ESTIMATION

There have been some attempts to calculate the optimal size of the document's summary. Sweeny (Sweeney 2003) has mentioned some static percentage to generate an adapted document. For example he suggests that a summary of 3% is the most optimal value for the mobile phones (without considering the size of the document!). But as previously mentioned, the size of an adapted document is not static and it must be dynamically calculated, based on the user's situation. Although the user can't explicitly specify the size of the needed information for the different situations, however it can be calculated from the user's preferences and the other contextual parameters. In this section we try to list all of the parameters which can influence the adapted document's size. Subsequently, we propose a formula to show the relationship between these parameters and the size of the adapted document.

- Screen size N(ch): This is one of the most important parameters for calculating the adapted document's size. The more the screen size is small, the more the size of the delivered information must decrease. The unit we choose to measure the screen size is the *number of characters which can be displayed on the device's screen*. So it depends on some of the user's preferences such as the font's size registered in the user profile.

- User's availability time T_U(s): The more the user is time restricted the more the size of the adapted document must decrease. The value of T_U can be explicitly supply by user, extracted from his agenda, or it can implicitly be estimated based on the user's daily behaviours.

- Battery's availability time T_B(s): Contrary to the desktop devices, the energy of a mobile device is limited, so the system must consider the remaining time of battery's charge.

- Network network_effect: As mobile networks are slower and more unreliable than wired networks, the efficient time of connecting a mobile device to network may be less than a desktop PC one.

- Speed of reading V(ch/s): the size of adapted document depends on the speed of reading of user. This speed can be calculated explicitly by testing the user, or implicitly by observing the user's behaviour.

- Input facility α: This factor represents the required time to get a new page. When the screen size is small the user needs more time to interact with system to display the new page and the system also needs some time to display the required page.

The above situational parameters limit the efficient users time. By considering these parameters the appropriate size of the adapted document (S) can be calculated as follow:

$$S = V \times T' = V \times (T - t) = V \times [T - (\frac{S}{N} - 1) \times \alpha] \Rightarrow$$

$$S = \frac{V \times (T - \alpha)}{1 + \frac{V}{N} \times \alpha} = \frac{N \times V \times (T - \alpha)}{N + V\alpha} \quad (1)$$

In this formula ***t*** is the elapsed time because of contextual restrictions such as the small screen size. It's equal to the number of displayed page (*S/N-1*) multiplied by the required time to change one page (*α*). According to the above formula, the more the screen size (*N*) decreases, the more the number of displayed pages (*S/N*) increases and therefore the more the elapsed time increases. ***T*** is the efficient user's availability time that can be calculated form the following formula:

$$T = \min(T_U, T_B) \times network\_effect$$
$$where \;\; 0 \le network\_effect \le 1 \quad (2)$$

The above formula permits to calculate the size of a textual summary which simply can be used to calculate the number of phrases which must be displayed on the user's device. When the user is just interested in the textual contents (or the original document doesn't include the other physical components such as picture) the above formula can effectively work. But sometimes the users may be interested in the other physical components. In this case, the user requires some time ***M*** to read the

physical components (i.e. pictures, tables, graphs etc). So the formula (1) will be replaced by:

$$S = V \times T' = V \times (T - M - t) = \frac{NV \times (T - M - \alpha)}{N + V \times \alpha} \qquad (3)$$

The calculation of the required time to read the non textual physical components is out of the scope of this paper and requires more efforts. The user can explicitly specify the average value of M in his profile.

Therefore by the above formulas our system will be able to estimate the optimal size of the information which must be displayed on the user's device.

7 CONCLUSION

We have proposed a methodology to create documents adapted to the user's situation. The user's situation in this paper includes two main profiles: context profile and user profile. Physically Document Adaptation (PDA) and Semantically Document Adaptation have been defined and then we have explained how the situational parameters can be used to physically adapt a given document to the user's requirement. We have divided the PDA process into two sub-processes: structural document adaptation and a logical document adaptation. We have explained that knowing the size of an adapted document plays an important role in an adaptation system. So we have proposed a new formula which is able to estimate the size of the adapted document, based on the user's situation.

ACKNOWLEDGEMENTS

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.

REFERENCES

Gimson, R. et al., 2003, Device Independence Principles, W3C Working Group, http://www.w3.org/TR/di-princ

Grassel, G., Lauff, M., and Spriestersbach, A., 2002, Definition and Prototyping of Renderer-Independent ML. Nokia, SAP and IBM Germany. In *W3C Workshop on Device Independent Authoring Techniques,* SAP University, St. Leon-Rot, Germany.

Kurz, B. Popescu, I. and Gallacher, S., 2004, FACADE - a Framework for Context-aware Content Adaptation and Delivery, In *Second Annual Conference on Communication Networks and Services Research,* Canada.

Lewis, R., 2002, A Device Independent Method for Web Site Authoring. In *W3C Workshop on Device Independent Authoring Techniques*, SAP University, St. Leon-Rot, Germany.

Lum, W.Y. & Lau, C.M., 2002, A Context-Aware Decision Engine for Content Adaptation, In *Pervasive Computing*, IEEE Press, Vol. 1, No. 3.

Mandyam, S., Vedati, K., Kuo, C. and Wang, W., 2002, User Interface Adaptations: Indispensable for Single Authoring. In *W3C Workshop on Device Independent Authoring Techniques*, SAP University, St. Leon-Rot, Germany.

Naderi H., Rumpler B., Pinon J.M., 2006a, A Two Layered personalized Information Retrieval System, In *MAPS'06 : International Workshop on Multimodal and Pervasive Services ICPS'06.* IEEE Press, France.

Naderi H., Rumpler B., Pinon J.M., 2006b, An Efficient Collaborative Information Retrieval System by Incorporating the User Profile, In *AMR 2006*, to appear in Springer Lecture Notes in Computer Science.

Sweeney, S., Crestani, F., 2003, Supporting Searching on Small Screen Devices Using Summarization. In *Proceedings of Mobile HCI 2003 International Workshop*. Lecture Notes in Computer Science, volume 2954.