Adaptability Approaches in Digital Libraries

Pavel Boytchev[1], Alexander Grigorov[1], Jeffrey Earp[2],
Krassen Stefanov[1], and Atanas Georgiev[1]

[1] Faculty of Mathematics and Informatics,
St. Kl. Ohridski University of Sofia, Bulgaria
{boytchev, alexander.grigorov, krassen, atanas}@fmi.uni-sofia.bg
[2] Istituto per le Tecnologie Didattiche – C.N.R., Genova, Italy
jeff@itd.cnr.it

Abstract. This paper examines some approaches for endowing digital libraries with adaptability capabilities in order to scaffold and enhance end user experience. The paper provides a general overview of techniques and methods commonly adopted for achieving adaptability. It also discusses how these can be implemented, and to this end illustrates specific examples and guidelines drawn from the practical experience that the authors are currently gaining in the Share.TEC European project, a context in which adaptability is key to managing and responding to considerable diversity in user requirements.

Keywords: adaptability, digital library, user profile, user interface, search portal, education

1 Introduction

Adaptability is the overall capacity of a system to adapt to the user. This adaptation can be controlled explicitly by the user (customization) and implicitly by analyzing user behaviour and interaction with the system (personalization). Adaptability encompasses the collection of statistics about the user, building a user model, analyzing knowledge emerging from this model, determining how the system can adapt better to the user and, finally, the actual adaptation.

According to Barra, Negro and Scarano, adaptability is "the ability to be aware of user's behaviour so that it can take into account the level of knowledge and provide the user with the right kind of documents" [1]. Other authors, like Norvig and Cohn, mention adaptive systems where "much of the searching may go on when the user is not even logged in. The application does more on behalf of the user without constant interaction, and the sophistication comes from a splitting of responsibilities between the program and the user" [2].

In this paper we provide a general overview of the approaches adopted to endow digital libraries with adaptability capabilities, i.e. those features that scaffold the user experience for locating appropriate resources and sharing the burden, as mentioned above. Section 2 of the paper outlines general adaptability architecture first in general and then in more detailed terms. Section 3 describes adaptability features, especially with regard to interface customisation and searching and filtering. An examination of adaptability implementation is provided in Section 4, which looks in particular at user profile /model, search customisation, user statistics, and multilingual support. Finally, some brief conclusions are offered.

2 Adaptability Architecture

In search portals adaptability is closely related to the *user model*. The user model comprises a set of data describing the user, together with the algorithms that generate and use these data. The sum of these constitutes a virtual representation of a real (or potentially real) user and provides the basis for system adaptability.

According to Nurmi and Laine, a user model is "a representation of a user that captures goals/tasks, knowledge/background/experience, interests, traits/cognitive styles (holist, serialist) and context of work" as described in [3]: such a model can be empirical and analytical. So adaptability involves three fundamental, interconnected, mutually-related elements: adaptability system, user model, and user interface.

The relationship between these three elements is represented in Fig. 1. Adaptability is an iterative, continuous process that runs throughout a system's entire lifetime. As users interact with a system, it collects more data about their preferences and goals and fine-tunes itself. The main goal of this continuous process is to converge the virtual model of a user to a model that can provide sufficient data for adequate adaptation. However, when appropriately implemented, this process can also capture sharp shifts in user behaviour and again fine-tune itself considering the new situation. To illustrate this, let's imagine the situation of "Martha", a mathematics teacher who uses a portal for accessing and sharing education-oriented resources of different kinds. For career reasons, Martha may apply for a job as a physics teacher. This will affect her interests and even if she does not change her personal preferences, the system should be able to recognize the shift in her interests and build a new adaptation target.

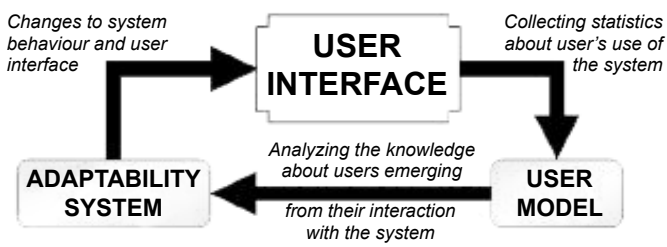

Fig. 1. Adaptability components & their relations.

Such re-adaptation cannot occur immediately; the system would need time to monitor and collect data in order to change its behaviour.

This general model can be considered in greater detail by considering specific processes and data. The two main adaptability-related activities that users usually perform in digital libraries are *setting preferences* and *searching* for digital content. When a user sets preferences (like language or topic of interest) this is commonly stored in that user's model. The model can contain three clusters of data:

- *explicit preferences* defined by the user and used for customization;
- *implicit preferences* that the system defines automatically by analyzing data about user behaviour and his/her preferences;
- *user behaviour* data generated from the user statistics.

Searching and filtering in a repository produces two clusters of data: (1) the list of *search results* (the user can *browse*, *display*, *comment* or *annotate* them); and (2) the raw *user statistics*, containing unprocessed data like searched keywords, found results, displayed records, written comments and annotations, ratings, etc.

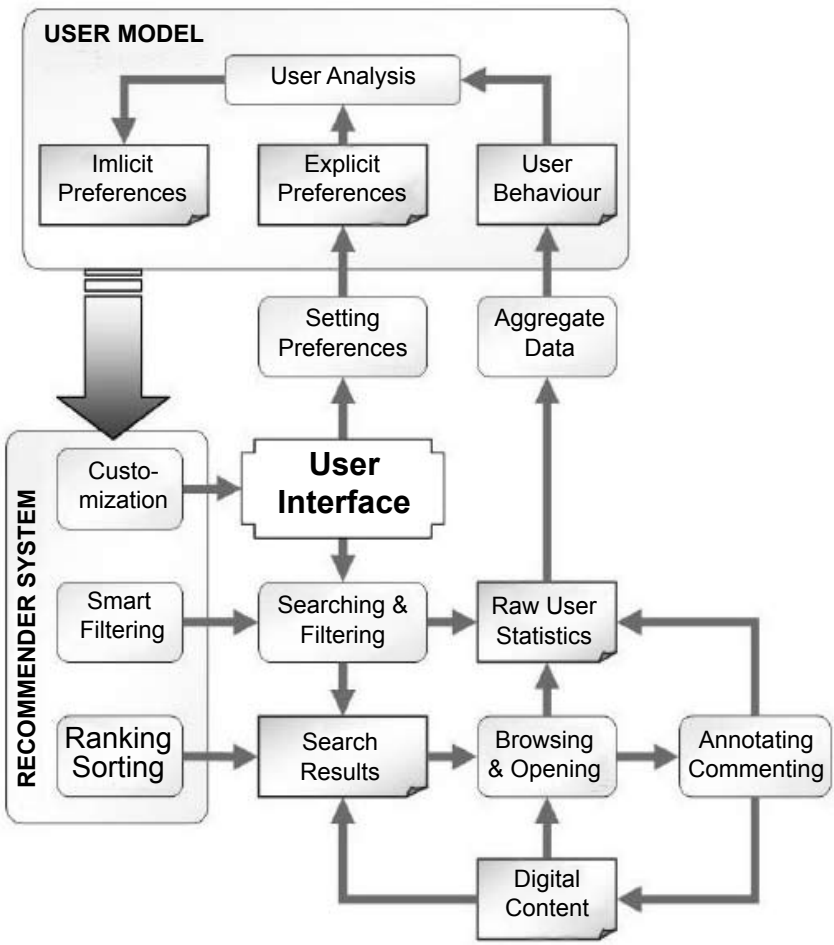

Fig. 2. A suggested adaptability architecture.

For these data to be used, they must be *aggregated and processed* into a representation form that describes *user behaviour*. As explained earlier, analysis of these data allows generation of the user's implicit preferences.

The last set of elements in the suggested adaptability architecture (Fig. 2) make up the *recommender system*. These elements use data from the user model, namely the explicit and implicit user preferences, and decide how to accommodate system functionality and behaviour for better matching with these preferences. In this architecture, the recommender system is composed of several modules that control different aspects of adaptability. The *customization* module is responsible for adapting the user interface. This includes features such as language preferences and layout/colour schemes. The module for *smart*

filtering provides additional filters depending on user interests (either explicitly stated or implicitly deduced). This module also provides initial values for filtering criteria. The *ranking* module affects the sorting order of search results, promoting those records which are likely to best represent the user's intentions.

It should be noted that the architecture represented in Fig. 2 regards system adaptability towards *individual* users, i.e. it does not take into consideration collaborative behaviour, community interests or special interest groups.

3 Adaptability Features

The adaptive behaviour of a digital library system is mainly (although not exclusively) aimed at identifying those digital resources that best suit users' needs, without asking them to enumerate the requirements in detail. Whenever possible, users are spared the trouble of explicitly expressing these needs as query parameters; the values are inferred from the user model (see Fig. 2).

This section enumerates and describes various adaptability features that can usually be found in digital library applications; the assumed point of view is that of the end user. It should be noted that the listed features are not intended as those that necessarily ought to be implemented in any digital library interface. Some of the adaptations may be inappropriate or inapplicable when factors like the computational power, storage limitation, algorithmic complexity and data availability within a given system are taken into account.

The user interface for digital libraries typically uses adaptability approaches for implementing the following features: (1) *adaptation of the user interface* based on explicit user preferences given in the user profile; (2) *adaptation of search results presentation*, based either on explicit user preferences or on implicit preferences derived from user behaviour and available statistical data; and (3) *specific recommendations* about resources and social interactions, based either on explicit user preferences or on implicit preferences derived from user behaviour and available statistical data. When considering adaptability features, we must recognise the tension that exists between a user's more persistent characteristics and his/her momentary interests, which may vary in accordance with context shifts, either temporarily or permanently. Accordingly, system suggestions ought to be presented in the form of *default values* (e.g. for query fields) that the user can always override.

3.1 Overview of interface customization

User customization of the digital library interface is a kind of adaptability which is explicitly controlled by the user and is stored in the user profile. One major component of the user interface is interface language. Many systems propose different ways of changing the interface language, either through a manually set "Language" field or automatically via the user's profile or default browser setting.

Manual interface language change is volatile: it is not saved when the user logs out from the system. At the next login, the interface language will be determined by the user's personal profile setting, which by contrast is persistent and overrides default browser language settings. This last means proves useful when the user is unregistered or when the registered user has not defined her/his interface language in the profile.

Other adaptability opportunities that digital libraries typically provide fall within the category of so-called *skin adaptability*, i.e. the chance to change personal page characteristics like text font, style, size and text and background colour. Adaptability regarding the user view of the information that a given system provides can be handled by explicit preferences. A typical example is where the user has options for personalising the individual home page, e.g. displaying and re-arranging different sections of particular personal interest.

3.2 Searching and filtering

Clearly, the essential function of digital libraries is to help users easily locate resources that fit their needs by providing suitable browsing, searching (and filtering) functions. Put simply, many systems handle search as follows: a user query is received and analyzed, the query is executed in the repository, and a result set is generated. In this scenario, the onus is largely on the user to frame the query in such a way as to (hopefully) generate a satisfactory result list; the system provides very little "background" support, though "foreground" support like help info and FAQs may be on hand.

In Section 4 a digital library initiative called Share.TEC is introduced in which adaptability functions are presently being implemented to meet specific user needs (for more details see below). In this particular case, query system behaviour can be described as follows:

1. A query is received and analyzed, new user model information is extracted, and the query is executed in the repository, yielding a result set.
2. If more results are needed, the query filters are relaxed. The results returned from the transformed query are ranked lower.
3. The user model is updated to take user model information into account.

The advanced query function provides users with a mask to specify parameter values: these are used by the query engine to select those records that exactly match the required values. Parameters in the advanced query mask can take initial values from: (1) the preferences stored in the user profile; (2) the history of user interaction; and (3) information associated to the similarity ring the user belongs to (grouping of fellow-users with similar characteristics). The advanced query mask can be presented to the user in two modes: *complete*, where all the result elements are shown, possibly with preset values that the user can change; and *simplified*, where elements that have been assigned a default value are hidden from the user. A typical example of the latter is language preferences, which are likely to remain constant in time, although the user can always switch to the complete view to alter these preset values.

The approach adopted in Share.TEC is for search result lists to be customizable, i.e. the user determines exactly what fields of the records are to be displayed along with the resource title, e.g. content provider, author's name, resource type. Another customization possibility is to reorder the generated search result from the default relevance setting to a different parameter like date, location, etc.

Another option under consideration is for a personalized basic search form that groups the user's most commonly used search fields. Initially, this could contain a preset list of basic filters.

An adaptability feature already implemented in Share.TEC is *Similar documents*. This option prompts the system to find similar documents to the one considered by comparing only their metadata records. To allow flexibility, a *Level of adaptability* option is to be proposed. This customization allows the user to specify what level of adaptability he/she requires for the returned search results. Levels might be:

- *No adaptability* – results are returned as they are in machine-dependent order. This option provides results in the fastest possible way;
- *Basic adaptability* – results are returned in accordance with metrics based exclusively on the matching of query and result metadata. This option provides relatively fast results ordered by relevance.
- *Full adaptability* – results are returned and ordered according user preferences, interests and history. This option provides the most accurate ranking at the potential cost of performance.

Level of adaptability may well be provided in Share.TEC as an advanced option. However for novice users, "full adaptability" could be set as a default level of adaptability if the performance cost is affordable.

4 Adaptability implementation

The principles and features of adaptability in digital libraries as described in the above sections find suitable application in the EC co-funded Share.TEC[1] project. Share.TEC is developing a federated digital library system designed to provide culturally-aware access to resources related to the field of Teacher Education (TE) across Europe. This field is made up of people with very different backgrounds, ideas and assumptions, and with very different requirements; the community of system users is expected to reflect a variety of languages and cultures. Central to Share.TEC's mission is that the system being developed should have capabilities to support diversity, and a key aspect of that support lies in adaptability.

In this section we describe in more detail the features being integrated into the Share.TEC system in order to provide as much scope for adaptability as possible.

The *user profile* is used to provide flexibility and adaptability to the system. The user model is composed of three top-level components:

- *the quasi-static profile*, which includes personal information about the user that is unlikely to change over a period of a year;
- *history of interaction with the system*;
- *counters* that allow implementation of special heuristics.

When first registering, new users lack the history required to fuel practical adaptability features and so they acquire an initial "novice" status. In this phase they inherit the characteristics of the closest fellow-users and the adaptability functions are driven accordingly. This status lasts only until the history size reaches a given threshold.

[1] Share.TEC - SHAring Digital REsources in the Teaching Education Community, eContentplus programme (ECP 2007 EDU 427015); http://www.sharetecproject.eu/.

The user model (profile) consists of some more or less static features (affiliated institution, teaching target level, language, etc.) and dynamic features like history of interaction with the system (ratings, queries, annotations, resources visited, etc.) and some counters. Most of the data changes dynamically so for efficiency reasons the user profile is not stored in the digital library itself but in the internal user interface database. However, the user profile is closely integrated with an important component of Share.TEC's semantic layer, namely the Teacher Education Ontology (TEO) [4], which is stored in the digital library. User features such as Professional Area, Experience Area, Teacher Practice Context that refer to TEO branches assume values related to the corresponding TEO nodes. One of the results from this is multilingual support as the values associated to TEO nodes are expressed in several languages. When users edit their individual profile, they select from a list of values extracted from TEO in their own (or preferred) language.

The dynamic part of the user profile is used mainly by the recommender system to highlight relevant resources and user recommendations. These profile features are used to compute the distance between users based on the metrics defined. For details on the implementation of the recommender system, see [5].

The *search component* of the Share.TEC portal also uses the user profile and TEO intensively. It is based on the Solr search engine [6] and uses Lucene style queries. The search component performs semantic query expansion that consists of the following steps:

- *Initialization phase* - the query is analyzed and user profile data is extracted;
- *Expansion with user preferences* - the query is expanded in a way so that resources matching the user profile are ranked higher;
- *Ontology based expansion* - the query is expanded to return results that can also be associated to the related classes in TEO;
- *Multilingual expansion* - the query is expanded to include as synonyms the names of the TEO nodes translated in the available languages. In this way resources in different languages can be found;
- *Recommender based expansion* - the resource recommender is used to get a list of the most-viewed resources according to user profile. Then the query is extended to boost the rank of these resources.

To achieve a level of adaptability, the system relies on the *user statistics*: explicit user preferences and profile traits; cultural context features (when available); inferredprofile and stereotype. Whenever these three sources conflict, the priority is given to explicit information from the user.

The user profile consists of two logical components. The first, considered here as "static", is qualitative in nature and contains mostly explicit (i.e. not inferred) data such as name, professional profile, language, country, etc, maybe including a specification of interests by means of keywords.

The second component is dynamic and quantitative, and consists of an array of numerical indicators associated to specific user interests and features. Each indicator is incremented whenever the user performs an operation that reveals an interest. For example, the Share.TEC system may have counters for each of the top-level knowledge area classes: when individual users query the repository specifying a mathematics subject, or they annotate a resource, which is

linked to a maths subject, the counter associated to mathematics is incremented. The system could probably have a subset of counters devoted to the knowledge area, one subset for the possible values of context, one for the didactic strategy, etc. To avoid an excessively high number of counters, only the most relevant ones should be chosen for the implementation. Each user has a set of counters and after a period of system usage their values provide a sketch of that individual user's interests.

This is entirely based on the language selection that the user makes in the portal interface. The system currently supports six languages: Bulgarian, Dutch, English, Italian, Spanish and Swedish. The system also grants *multilingual support* through a multilingual metadata model, which guarantees that all concepts from the TEO ontology and the common metadata model are translated in advance to the predefined set of languages [7].

5 Conclusion

In this paper we have presented an overview of adaptability approaches in digital libraries and have given recommendations on how these approaches can be implemented in practice. We have also provided a detailed example of such an implementation in the Share.TEC system, where adaptability is central to supporting the expected diversity in users and their requirements, a key aspect of the project's ambitions. The system itself is still under development and the current architecture for adaptivity is a cornerstone for supporting diversity at a cross-cultural level.

References

1. Barra, M., Negro, A. & Scarano, V. When the Teacher learns: a Model for Symmetric Adaptivity. Proceedings of Second Workshop on Adaptive Systems and User Modeling on the World Wide Web, Banff, Canada. (1999)
2. Norvig P. & Cohn D., "Adaptive Software", Harlequin Incorporated, http://norvig.com/adapaper-pcai.html
3. Nurmi P. & Laine T., "Introduction to user modeling", University of Helsinki, http://www.cs.helsinki.fi/u/ptnurmi/teaching/UM/Seminar_040907%20-%20Introduction.pdf
4. Alvino, S., Bocconi, S., Boytchev, P., Earp, J., Sarti, L., "Sharing digital resources in Teacher Education: an ontology-based approach", In Proceedings of First International Conference on Software, Services & Semantic Technologies, Eds.: D. Dicheva, R. Nikolov and E. Stefanova, Sofia, Bulgaria, pp. 52-59 (2009)
5. Tosato, P. "Metrics for Resource-User Matching", Published as deliverable D4.2 of Share.TEC project (2009), http://www.share-tec.eu/content/1/c6/04/41/02/17D42Metrics forresourceusermatching.pdf
6. Apache Solr home page, available at: http://lucene.apache.org/solr/
7. Alvino S, Bocconi S., Earp J, Sarti L., "Ontology and Metadata Models" Published as Deliverable 2.3 Share.TEC project (2009). http://www.share-tec.eu/content/1/c6/04/41/02/DEL2_3Finalreleaseversion.pdf