CHALLENGES IN ADAPTIVE WEB INFORMATION SYSTEMS: DO NOT FORGET THE LINK!

GEERT-JAN HOUBEN

Technische Universiteit Eindhoven
PO Box 513, NL-5600 MB Eindhoven, The Netherlands
E-mail: g.j.houben@tue.nl

In this invited talk we will consider a number of issues that are interesting research topics for the area of Web Engineering and WIS (Web Information System) Design Methodologies. Since this is a workshop we will raise a few questions and identify challenges for the research area, and hope to provide food for the discussions and panels in this workshop. In order to indicate some interesting research topics in relation to methods and design approaches for WIS, we take the perspective of adaptation and personalization. Not only is the aspect of adaptation a challenging one in Web Engineering where the (business) interest in personalization is much higher than the general understanding of the adaptation engineering process. Moreover, the field of Adaptive Hypermedia has considered adaptation for some time and some important lessons can be learned from that research. Topics that are discussed are device adaptation, user modeling, data integration, interaction and dynamics, and quality. While modern trends in Web Engineering, e.g. Semantic Web, consider content engineering in the new environment, we do argue that for the effective use of the Web paradigm it is crucial not to underestimate the value of the hypermedia navigation, and therefore not to forget the hyperlink in the engineering process.

1 WIS and Adaptation

When we consider the World Wide Web (WWW) and its development, we can clearly see how this has evolved from being a HTML-based platform that combines content with presentation of information. We also see the hypertext or hypermedia-based nature, with link and navigation as core concepts. Through languages like HTML the Web offers a uniform way of accessing the content. In terms of engineering the day-to-day practice shows that handcrafting is the standard when it comes to producing information on the WWW.

Out of the research field of hypertext/hypermedia, that preceded the WWW research for decades, an interesting subfield emerged that concentrates on the aspect of adaptation in hypermedia. While hypermedia typically breaks down into pages and links, adaptive hypermedia identifies content adaptation (adaptive presentation) and link adaptation (adaptive navigation support). See [1] for a lot of material from adaptive hypermedia research. Out of all characteristics of adaptive hypermedia, we can single out the typical handcrafting nature of constructing adaptive hypermedia. In response to this, the problem of authoring has received a lot of attention from adaptive hypermedia researchers.

Out of this research many Adaptive Hypermedia Systems (AHS) have evolved. They are tools and engines for adaptive content delivery. Within

limitations they adapt the access to the content. A reference model for AHS is AHAM [2] that distinguishes a Domain Model (DM), that describes the subject domain at the conceptual level, a User Model (UM), that describes the user's knowledge, preferences, background, experience, etc., and an Adaptation Model (AM), that describes how the actual adaptation should be done, depending on the domain model and the user model. While some that see the possibilities of AHS ask themselves whether adaptation is a kind of magic, the construction of adaptive applications simply uses the knowledge behind the material available. Combining knowledge about the content and knowledge about the users and exploiting straightforward techniques of content mark-up, user modeling and adaptation specification, the main challenge is to structure all this knowledge properly: with a proper knowledge structuring the actual adaptation is straightforward.

Let us now take a look at the concepts of WIS and WIS engineering. For now we can suffice with the definition of a WIS as an information system that uses Web technology in its front-end or back-end. Typically we see that a WIS contains information that gets instantiated dynamically, and therefore strongly separates content and presentation: in fact, in specifying the WIS functionality much of the attention is paid to controlling this access and presentation for the users.

WIS engineering is considered to be different from the traditional authoring we know from the WWW. One important reason for that is that the so-called deep Web with databased content is quite different from the so-called surface Web with manually crafted content. It implies dealing with data that is dynamic, considered at schema-level, and used in high volumes. Applying these WIS in business contexts implies the need to reuse or copy many of the proven principles from database-based Information System (IS) engineering: in a sense the engineering of a WIS is first and foremost of course the engineering of an IS. Most of the research in WIS Engineering Methodologies, exemplified by the previous IWWOST workshops, aims to study how to exploit the navigation and browsing paradigm in the content of the information system.

In order to set the stage for discussing this research on methodologies we sometimes refer to Hera [3] and its design process and models. Hera distinguishes a Conceptual Model, that expresses the relevant domain. Other models in this first Semantic Layer like the Integration Model are out of scope here. In the second Application Layer Hera distinguishes the Application Model that describes the navigation structure of the Web (hypermedia) presentations, and offers as basic modeling primitives the concepts of slices (content containers) and slice relationships (navigation between slices). In that layer also the adaptation is specified.

2 WIS Engineering and Adaptation

Adaptation (and personalization) is a design aspect that recently gained much more attention, but when we take a close look at the different WIS design methodologies we can argue that the adaptation aspect is still lightly supported in the methodologies. In general, it is still hard to specify adaptation, there are not many tools to support the designer, and in practice it turns out to be hard to combine adaptation with other relevant aspects of WIS design, such as performance to name just one. One of the main applications of adaptation is personalized Web access. All users of the WWW know the numerous MyWhatever portals that give a personalized access to information, but only a few people know how to construct such a personalized access effectively. Another good and interesting example of adaptation is device adaptation or device dependency in presenting information. Many methodologies, including Hera, allow to specify rules that define conditions for presenting information. Such rules can express that a certain piece of information, e.g. a picture, is only shown when the user's platform has the capability to support that information format and is thus able to show that information. These rules can sometimes be context-independent, as are the typical device dependency rules, e.g. show pictures only in case of image-capability, while fewer methods allow for context-dependent conditions, i.e. conditions that depend on the actual state of the current browsing process. This latter kind of dependency is often used in fields like e-learning where the state of the browsing process is interpreted as an indication of the state of the learner, and therefore the information that is provided is made dependent on that learner's state: show a student only that information that the teacher thinks is relevant given the current knowledge state of the student.

When we limit ourselves here to the device adaptation, already a number of issues have been raised, for example in the fields of adaptive hypermedia and e-learning.

- While the technology allows us to carry many smart Internet-aware devices, e.g. camera, MP3-player, phone, GPS etc., it is not clear how to characterize device capabilities more precisely, both in terms of how they present information as in the aspect of what data they generate. A proper question in relation to the diversity of devices and device families is also how to specify device capabilities: per device or per device feature? Related to the issue of evolution, we should also think on how to support that WIS will work on future devices with presently unknown features.
- What about real-time issues in connection with device dependency, when applying adaptation technologies in dynamic and runtime settings? Do the adaptation techniques allow for an effective application in large-scale high-volume high-speed settings?
- A fundamental question for the developers of WIS design methodologies

is where device adaptation modeling has its place in the WIS design process? Most of the current research proposals focus on a small number of device aspects but appear to neglect that this device dependency is only one of the issues in the global application design process.

3 User Modeling

Closely connected to the issue of user adaptation is the process of user modeling as the basis for the adaptation. We first define two kinds of adaptation. Adaptability is the kind of adaptation in which based on a (static) user profile, the application delivers information in accordance with the conditions from the profile. Typical examples are device dependency where the information presentation is geared towards the user device, or the way in which the application adds advertisements for the user. In this class, the user profile is often obtained through explicit registration by the user or by considering user groups or stereotypes to identify the users. The second kind of adaptation that can be distinguished is adaptivity. Then the application delivers information based on a (dynamic) user model. Typical examples can be seen in the areas of e-learning and in sales Web sites. With this adaptivity the user model is often constructed and maintained persistently by observing the users browsing behavior.

User modeling is necessary to adapt to knowledge, goal, interest, background, or navigation history. In fact, good user modeling is essential to getting the adaptation right. There are a number of interesting research problems when it comes to user modeling, often studied in a multidisciplinary context like in e-learning:

- The main question in adaptation is what is actually stored in the user model? Each application and each application area has of course its own requirements, but also for the sake of proper and effective adaptation the right insight in the user's state needs to be available and the user model should supply that information.
- Often nice adaptive applications fail because of the Cold Start phenomenon: how is the system provided with enough user model information to be able to start functioning properly right away (and not giving the users an annoying start-up phase which actually might make the users dislike the system and leave)?
- Since usually applications do not exist stand-alone, it is also essential to know how to share the user model between different WIS. This implies questions concerning the universality and standardization of UM data, but also concerning standard access protocols for UM data.
- A really important research problem is related to the privacy and legal issues with respect to storing and exchanging user models. This might

include also issues connected with the security involved in the user model storage and exchange.

- Applications that allow the users to interact themselves with the user model data stored for them, require a careful design of the support for that. It is challenging to decide on how much freedom the user gets to change user model values and how the system interprets that user influence.

4 Data Integration

Data integration refers to the process of retrieving the data that can subsequently be presented. This data is typically gathered from multiple diverse sources. The diversity is related to aspects like the format, quality, and access protocol for the data.

- The main problem concerning data integration is how to model and specify how the content for the WIS is retrieved from multiple distributed heterogeneous sources. This includes how to search for relevant sources in the collection of available sources or on the WWW, and the subsequent retrieval and possible alignment. The alignment should provide the mapping of the data from the source format to the format of the WIS. In case of WIS this is more than just a syntactic transformation, but can include also other aspects on a semantic level, even source discovery. Typical for a WIS application is that, compared to traditional distributed database applications, the types involved are much more different and mixed media types. Moreover, there is typically a much larger difference in schema granularity. The main consequence of this problem is the fundamental question how this integration aspect is handled in the entire WIS design process.

- The data in a WIS is also influenced by time. Evolution in the sources and their data, just as the decay of the information are factors that need to be considered in the integration specification. Compared to traditional database integration the volatility of data is higher and the changes, both in the data and in the sources, are more frequent.

- A research topic which deserves more attention than it gets is the question how to deal with the vast amounts of legacy content that are around. Most applications that are currently developed use data that is already around in more traditionally implemented information systems, and therefore it is beneficial to obtain effective and efficient mechanisms to include legacy content. Experience shows that for the successful adoption of applications also the start-up needs to be considered thoroughly: a technically sound application without useful content to start with often leads to a failure.

- When looking at the specific aspects of integration with respect to adaptation, there are a number of research problems that require attention. When we take the adaptation for device dependency, we have to know how to deal with connection customization? Given the fact that within an application there are considerations to be made regarding the available resources (e.g. due to bandwidth), and the user demands and preferences, then it is a challenge for the application and adaptation designer to ensure the required quality by an appropriate source selection for retrieving the content.

An important observation when one considers the way integration is often treated in current WIS methodology research, is that one tries to exploit all the results from distributed information systems and databases. While this gives nice results, it also leads to the danger that one applies the integration only at the content level, in Hera terms in the Conceptual Model. However, as WIS applications base their success on the hypermedia nature of the information presentation, it is essential that the integration also touches the navigation structures involved, in Hera terms the Application Model.

5 Interaction and Dynamics

Looking at how Web applications develop, we observe that from the one-way communication that was so typical for the first generations of Web applications, we now deal with two-way communication where the user does not just interact by simply following (hyper)links, but also plays a part in the information exchange by adding data, e.g. by filling in data in forms provided by the application.

- The main research question is finding out what the role of interaction should be during the entire WIS modeling and design process. Part of this is the problem how user dialog design can be combined with the design based on hyperlinks typical for earlier generations of applications. It is important to realize that the challenge is to add input and update facilities to the presentation-and-publication-based front-end that goes with a WIS.

- With this interaction, the challenge for adaptation design is how the adaptation can be based on data that is not only determined beforehand (and fixed), but also data that is added dynamically through the user interaction.

- A related problem is how to deal with the multimodal issues involved in modern WIS. For example, voice-interaction and audio play a role next to the traditional (visual) hyperlink mechanism. This is more than just different platforms for presentation: is also means different ways to

capture the interaction by the user and the user's environment, and the issue of where the initiative is: with the user (environment), with the system, or mixed?

This combination of adaptivity and user-interaction brings to the traditional hypermedia navigation view the additional aspect of dynamics, in the data and its navigation access structure. In approaches like Hera this dynamics is supported by having next to the static modeling primitives for the navigation structure of the information presentation, primitives to model input forms, data manipulation operations, and presentation visibility or appearance conditions.

Considering dynamics in the WIS applications, brings lots of researchers to have a look at the Semantic Web idea and the technologies that evolve from it. A main reason is that it also brings a natural demand for considering both data and process aspects together in the design. Particularly the process aspects can benefit from the advantages of using the concept of Web Services. A service-oriented architecture is considered to be an effective basis for supporting dynamic aspects in WIS design. Another good reason for a Semantic Web-inspired approach is the inherent support for metadata and its role in WIS design. Although one can argue that the Semantic Web, as vision to target non-humans on the Web, is still rather abstract, and it is unclear whether it will ever exist fully and if so when, there are a number of technologies and languages that give a more central role for metadata and can therefore help WIS design a lot, e.g. RDF, OWL.

Some of the research questions and problems related to the Semantic Web influence:

- How are the data and process issues combined, and what is the role of orchestration (and dynamics) in the (static) navigation-based content specification?
- How to deal with the many different languages and approaches to specify services (interoperability)?
- If we combine navigation and services, this adds problems to the adaptation specification, since adaptation can be hidden or encapsulated inside services.
- Having (annotation) metadata does not automatically imply interoperability: how can metadata get shared and exchanged? We see ontologies being used now, for example for exchanging user model data, but it turns out to be still problematic to do this effectively.
- Where is the metadata coming from? From the main suppliers of PC software? Is annotation something for the masses? If we assume that metadata need to be created, then it is also essential that there are mechanisms for XML/RDF/OWL export from backend data.

6 Quality

As most of the adaptation specification is based on metadata, many systems use rules in one way or another to express properties of the content, the user, the information provider (business), and other factors that need to be taken into account. The specification of the information in these rules and constraints in Semantic Web languages like RDF and OWL can also help with model verification ensuring the quality of the design: checking that the model specifications are correct and consistent can help the WIS designers (and the adaptation engineers) a lot. This holds specially for the runtime and dynamic properties which of course are hard to foresee in detail for the designers.

- One of the first difficulties in adaptation engineering is how to combine the different rules and how to weigh the different factors. For example, in a commercial or educational application it is clear that the preferences of the client or student need to be taken into account, but up to a certain limit as decided by the shop or teacher. Related to this, is the problem that adaptation usually combines many different aspects, and the specification via rules can lead to a fragmented handling of the adaptation. Think of the way device dependency is specified independently from the pedagogical aspect in an e-learning application.

- From the previous it is clear that particularly for adaptation engineering tools for authoring are needed. With these tools the quality of the design can be verified. This means that consistency can be checked, consistency between requirements and design, but also consistency between the models that are used within the model-driven design. Note that many researchers now concentrate on usability evaluation, which definitely should not be forgotten, but there is another vast area of quality control to be tackled in adaptation.

7 Adaptive Web Information Systems

When we consider WIS in which adaptation plays a significant role and that therefore deserve to be called Adaptive Web Information Systems (AWIS), it is wise to observe that next to the navigation-related issues there are many other design and engineering aspects that are part of information system engineering in general: think of implementation issues, like the choice for servlets, CGI, or different server/client-side technologies, like mobility, caching, replication, and reuse of services (components). It is a major challenge for research in the field of AWIS design methodologies to integrate and combine all different design aspects, some of which we have discussed here, into one coherent and consistent approach. A good example is the issue of performance which of course cannot be jeopardized for the sake of the ultimate personalization adaptation. For the same reason it could be useful if there would be a generally

accepted definition of UML profiles for designing WIS and user-adaptation, just as a collection of design patterns for WIS and AWIS models.

We conclude that when we evolve from WIS to AWIS, and put the adaptation on top of the available navigation and browsing paradigm, we should note that the combination of hypermedia adaptation (specially adaptivity) and dynamics lead to an interesting relationship between content and navigation. At the same time we see how we go from presentation of data to interaction with content, and the ideas from the Semantic Web initiative and its languages like RDF and OWL support that development.

When we want to make the move from art to engineering in WIS, it is essential to realize that with the Web the role of content in information systems has changed. We keep on making information systems in the more or less traditional way, but we also want to exploit the new possibilities, e.g. metadata, interaction, adaptation, which gives a new context for content engineering. The first step in Web engineering is therefore to *get the content right!*

Metadata helps bridging the gap, and a Semantic Web perspective helps dealing with metadata. It does so for example by supporting the sharing and interoperability of models of the content (CM in Hera) via ontologies. However, we see hardly any attempts to create a similar support for the navigational design models (AM in Hera).

It is therefore crucial to also move from art to link engineering, and to make the second step the one that allows to *get the (hypermedia) navigation right!* Navigation is an important factor in conveying the semantics of the information, and an effective mechanism to involve the user in the communication process. Navigation adaptation is crucial to make the right content available, and an effective mechanism to ease the publishing of content. Therefore, we conclude with the statement:

Do not forget the link!

References

1. Adaptive Hypertext & Hypermedia Homepage, http://wwwis.win.tue.nl/ah
2. De Bra, P., Houben, G.J., Wu, H.: AHAM: A Dexter-based Reference Model for Adaptive Hypermedia. In Proc. The 10th ACM Conference on Hypertext and Hypermedia, ACM Press (1999)
3. Vdovjak, R. Frasincar, F., Houben, G.J., Barna, P.: Engineering Semantic Web Information Systems in Hera, Journal of Web Engineering, Vol.2, No. 1&2 (2003)