Automatic Hypermedia Generation for Ad-hoc Queries on Semi-structured Data

Geert-Jan Houben, Paul De Bra*

Department of Computing Science
Eindhoven University of Technology
PO Box 513, 5600 MB Eindhoven, the Netherlands
Tel: + 31 40 2472733
E-mail: {houben, debra}@win.tue.nl
(* currently on leave at the University of Antwerp, Belgium)

ABSTRACT
This paper describes research on the automatic generation of hypermedia or Web-based presentations for semi-structured data resulting from ad-hoc queries. We identify how different aspects of adaptation, such as personalization and customization, influence the generation process. We address important aspects of the software that facilitates the generation process.

KEYWORDS: Hypermedia generation, semi-structured data, ad-hoc queries, personalization, customization, adaptation.

INTRODUCTION
The generation of hypermedia presentations plays an important role in information management on the Web. In applications for Web modeling and querying, information extraction and integration, or Web site construction and restructuring [4], the generation of hypermedia aspects for given data is essential. At the same time Web-based information systems have become a popular vehicle for disseminating information. Using the concepts known from the Web users can access information more effectively than in most legacy systems based on strongly structured data, e.g. relational databases. Especially in applications where the data have a semi-structured nature, a Web-based or hypermedia approach benefits the users: they can exploit structures that are (implicitly) available, while browsing the data looking for the desired information.

Our target applications are collections or libraries of digital (semi-structured) data. Concrete examples concern real-estate sales, mail-order catalogs, auctions, (used) car sales or electronic (TV) program guides (EPG). The applications allow users to ask for data (by posing an ad-hoc query) and react by automatically producing a hypermedia presentation for these data. The main idea is that the data (possibly retrieved from different heterogeneous sources) have an internal structure that the application can use to provide the end-user with a "richer" hypermedia or Web-based presentation. In the HERA research project we investigate how the information system can derive a hypermedia presentation that presents the data to the user in a way that is adapted to the user's situation. The prime goal is that the derived presentation "adds value" based on knowledge about the data, the user, and the application itself. HERA also develops a software architecture that facilitates this generation process, using XML as the main data format.

QUERY SPECIFICATION
The data that need to be presented are typically the result of a query asked to the underlying information system. Perhaps the end-user is not aware of that, since this might be hidden through an elegant browsing interface. For the system, however, it is a query and the returned data represent the query result. For the generation process the system tries to extract knowledge from the query specification in order to produce a suitable presentation [5].

We use a query language that allows for the specification of the kinds of queries that typically occur in our applications. We concentrate on ad-hoc queries that can be seen as minor deviations from standard queries, assuming that for a set of standard queries predefined presentations are available [5]. The user asks to hide certain attributes or to show extra ones. The user also enters object selection criteria. We could use a language like XML-QL for this purpose. However, due to the lack of a standard query language for XML, we currently use XSLT in the HERA software to express the different transformation of the data (in XML).

This research is related to work on view specifications for XML data [1]. The main difference is that in HERA the query transformations are influenced by a number of practical issues, e.g. general or domain-dependent heuristics.

QUERY RESULT TRANSFORMATION
The generation process starts from XML data that represent the query result, and eventually transforms them into a

specification that is given to the Presentation Manager, the software responsible for the actual presentation in the user's browser. The transformation is based on different aspects of "intelligence" used in the system:

- Heuristics-based rules incorporate general design knowledge on presentation generation. The current software includes rules on access and navigation principles for large collections of data, e.g. on the use of an index or a guided tour. See [5] for details.
- Personalization represents personal preferences as specified by the user. Being able to influence the generation process enables the user to style the generation presentation to his or her needs. Such personalization includes addition or removal of attributes and also preferences regarding layout.
- Customization acknowledges the use of different platforms. The EPG for instance will present data about TV programs differently on a PC, a TV screen, a palmtop, or a remote control unit.
- The application contains also internal knowledge, specified by the application's author, to express the specific goals that the application wants to reach. E.g. the application may disallow the removal of attributes that represent advertisements.

This intelligence is specified as separate transformations on the XML data. The current HERA software uses XSLT, but the final architecture will provide friendly user interfaces for the specification of this intelligence.

ADAPTATION

This use of intelligence in the generation is closely related to the concept of adaptation in hypermedia. Using results from our research on adaptive hypermedia [3], we develop a model for this generation intelligence that distinguishes the different aspects of the adaptation by identifying a Domain Model, a User Model, and an Adaptation Model. The goal is to obtain a system that operates in such a way that the end-user better understands the adaptation that is performed.

ADDED HYPERMEDIA VALUE

To specify the internal structure of the data we use concepts from (Extended) RMM (see [6] for the basics). In an Application Diagram slices specify how the data of an element are structured and partitioned in view of presentation and navigation in a browser. These slice structures include both intra-element relationships and inter-element relationships.

The different kinds of intelligence, discussed in the previous section, specify transformations to the data. While some of the transformations aim at adapting to the situation, a core idea is that the transformations try to "add value" to the given data. This is based on a principle of "maximal access relationships" (like in DHymE [2]). It means that we exploit the structure in the data, make it explicit in terms of relationships and thus provide the user with more ways to navigate from one data element to another. New relationships are added to provide navigational access. New data elements are constructed, for example to represent higher-level objects that offer easy access to large collections of elements. The exact transformations being applied are a combination of the system's intelligence and the user's query specification (that can influence the process [5]).

The construction of these new elements and relationships implies that the system knows how to present them. The intelligence incorporates constraints for this construction, e.g. for the layout and inclusion of "obligatory" attributes.

HERA aims at finding practical solutions for the creation of new, volatile relationships for the given data. The basic principle is that the user is offered as much freedom as possible when navigating through the data. It implies that the HERA system exploits the structures "hidden" in the data to provide navigational relationships. In particular, we consider three types of relationships, or "dimensions":

- Hypermedia and navigational mechanisms.
- Spatial relationships.
- Temporal relationships.

Note that adaptation to the user's platform can imply a change between these hypermedia dimensions.

CONCLUSION

The generation of hypermedia or Web-based presentations has become an important aspect of the retrieval of data from an information system. The target applications of the HERA research described here typically contain large collections of data with a semi-structured nature. The HERA software exploits the structures that are (implicitly) available, adds relationships to the data (by means of different kinds of query transformations), and makes these relationships the basis for navigation. Thus, the user is provided with a presentation that is adapted to the user's situation and that offers the user easier browsing of the data.

REFERENCES

1. Abiteboul, S. On Views and XML, SIGMOD Record, 28, 4 (1999), 30-38.
2. Bieber, M. Hypertext and Web Engineering, Proc. ACM Hypertext'98 (1998), ACM Press, 277-278.
3. De Bra, P., Houben, G.J. and Wu, H. AHAM: A Dexter-based Reference Model for Adaptive Hypermedia, Proc. ACM Hypertext'99 (1999), ACM Press, 147-156.
4. Florescu, D., Levy, A. and Mendelzon, A. Database Techniques for the World-Wide Web: A Survey. SIGMOD Record, 27, 3 (1998), 59-74.
5. Houben, G.J. and De Bra, P. World Wide Web Presentations for Volatile Hypermedia Database Output, Proc. WebNet97 (1997), AACE, 229-234.
6. Isakowitz, T., Stohr, E. and Balasubramanian, P. RMM: A Methodology for Structured Hypermedia Design, Comm. ACM, 38, 8 (1995), 34-44.