Engineering Semantic Web Information Systems in Hera

Richard Vdovjak, Flavius Frasincar, Geert-Jan Houben, and Peter Barna

Eindhoven University of Technology
P.O. Box 513, NL-5600 MB Eindhoven, the Netherlands
{*richardv, flaviusf, houben, pbarna*}@*win.tue.nl*

ABSTRACT

Web Information Systems (WIS) use the Web paradigm and Web technologies to retrieve information from sources connected to the Web, and to present the information in a web hypermedia presentation to the user. Hera, a model-driven WIS design methodology specifies the process of integrating web data and transforming it to a hypermedia presentation to be browsed by an end-user. The process is divided into the following principal phases: integration and data retrieval, and presentation generation.

Categories and Subject Descriptors

H.1 [**Information Systems**]: Models and principles; H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia—*Architectures, Navigation*; D.2.2 [**Software Engineering**]: Design tools and techniques

General Terms

Algorithms, Design, Management

Keywords

WIS, SWIS, RDF(S), Semantic Web, XSLT

1. INTRODUCTION

The diverse audience of the World Wide Web with different platforms (e.g. PC, PDA, WAP phone) causes that the one-fits-all approach typical for the design of traditional information systems is not suitable anymore for Web Information System (WIS) [5] design. Adaptation of the presented content based on a concrete user and platform is a vital feature of a WIS. Moreover, the dynamic nature of Web data asks for the automated generation of hypermedia presentations as WIS output. The main issue in WIS design is therefore the specification of the *hypermedia generation* process including the specification of which sources to use and how to map their data (*integration*) into the system.

There are several engineering frameworks for WIS design, e.g. XAHM [2], WebML [3], UWE [7], and Araneus [9]. This paper presents our model-driven approach called Hera and introduces the transformation software that builds the heart of the hypermedia presentation generation process.

2. HERA METHODOLOGY

A WIS in the Hera perspective generates a hypermedia presentation in response to a user query. The data is retrieved from the data repository composed of heterogeneous data sources distributed over the Web.

WWW2003, May 20–24, 2003, Budapest, Hungary.
ACM xxx.

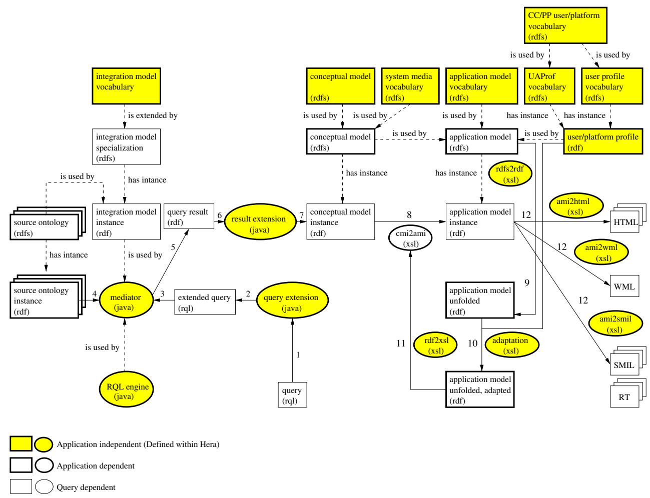

Figure 1: Integration and data retrieval, and presentation generation

During the generation of the output we distinguish the following phases: *integration and data retrieval*, and *presentation generation*. Figure 1 shows the details of the phases described in the following sections.

We chose to represent all information in RDF(S) [1, 8] because of RDFS flexibility (supporting schema refinement) and extensibility (allowing the definition of new concepts and properties). Due to the lack of full-fledged RDF-aware transformation processors and thanks to the XML serialization of RDF, we use an XSLT processor (Saxon 7.0) for our model transformations. For the purposes of data retrieval we use an advanced RDF(S) query language: RQL [6].

3. INTEGRATION AND DATA RETRIEVAL

In this phase several autonomous sources are connected to a conceptual model by creating channels through which the data will populate the concepts from a conceptual model on request.

The *conceptual model* (CM) provides a uniform semantic view over multiple data sources and describes a problem domain. CM is composed of concepts and concept properties that together define the domain ontology.

The *integration model* (IM) addresses the problem of relating concepts from the source ontologies to those from the CM. This problem can also be seen as the problem of merging or aligning ontologies. The efforts to automate the solution to this problem usually do not offer good results. In our approach we currently rely on a domain expert to articulate CM concepts in the semantic language of sources. By instantiating the integration ontology the designer specifies the links between the CM and the sources.

The *integration model ontology* (IMO) is a meta-ontology describing integration primitives that are used both for ranking the sources within a cluster and for specifying links between them and the CM. The main concepts in the IMO are *Decoration* and *Articulation*.

Decorations serve as a means to label "appropriateness" of different sources (and their concepts) grouped within one semantically close cluster, while *articulations* describe actual links between the CM and the source ontologies and clarify also the notion of the concept's uniqueness which is necessary to perform joins from several sources.

While the *integration* phase (instantiating the IM) is performed only once, prior to the generation of the presentation, the *data retrieval* phase is performed for every query. In this phase the query is extended and split into several sub-queries which are then routed to the appropriate sources. Subsequently, the results are gathered and transformed into a CM instance.

4. PRESENTATION GENERATION

In the presentation generation phase the retrieved data is transformed into a hypermedia presentation suitable for the user platform and for the user preferences. The presentation generation is composed from three steps: the application model generation, the application model instance generation, and the presentation data generation.

The *application model* (AM) describes the navigational aspects of the hypermedia presentation. The AM is composed of slices and slice relationships (aggregation and navigation) that together define the navigation ontology. A slice is a meaningful presentation unit of some media items originating from different CM concepts.

In the *application model generation* the AM is converted to an AM template (an "empty" AM instance). In the *application model instance* generation the AM is instantiated with the retrieved data. The adaptation (association of appearance conditions to slice references based on the User/Platform Profile) is also realized in this step. In the presentation data generation the application model instance is translated to browser interpretable code. Figure 2 depicts how three different translations appear in three different browsers: HTML browser, WML browser, and SMIL browser.

5. CONCLUSIONS

Hera is a model-driven methodology which uses different models for different aspects involved in the design of web information systems. This paper briefly describes Hera's integration model and the adaptable application model in order to support an automated process of generating adaptable hypermedia presentations from different sources. As a web ontology language is still in its infancy [4] we chose to represent Hera models in RDF(S). In order to represent the Hera different models we provided appropriate RDF(S) extensions. The RDF/XML model serialization enabled the use of XSLT stylesheets as transformation specifications between the different model instances. This approach proved to be satisfactory as one does not need to use the RDF(S) inference rules in the transformation specification. As future work we plan to use (depending on their existence): a mature web ontology language for representing the Hera models and a web ontology-aware (or at least an RDF(S)-aware) transformation language to be used for the specification of the Hera transformations.

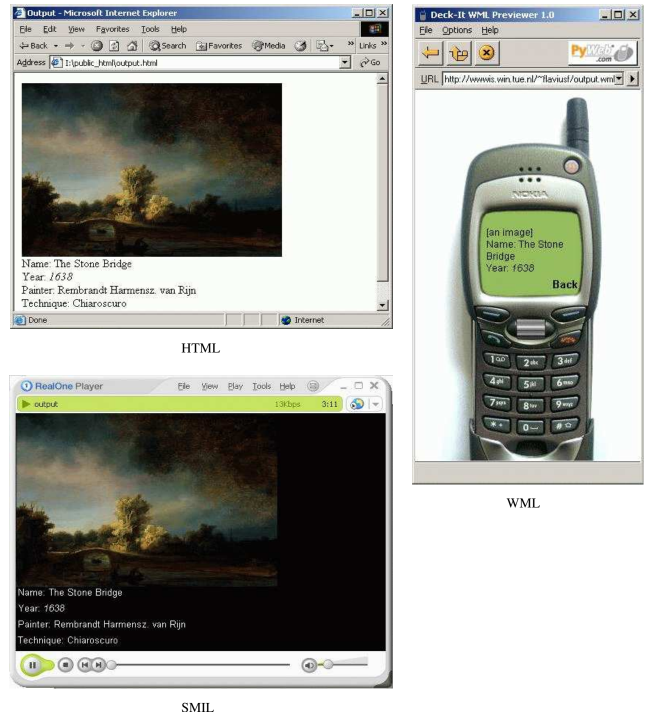

Figure 2: Hypermedia presentation in different browsers

6. REFERENCES

[1] D. Brickley and R. Guha. Rdf vocabulary description language 1.0: Rdf schema. W3C Working Draft 30 April 2002.

[2] M. Cannataro, A. Cuzzocrea, C. Mastroianni, R. Ortale, and A. Pugliese. Modeling adaptive hypermedia with an object-oriented approach and xml. In *Second International Workshop on Web Dynamics*, 2002.

[3] S. Ceri, P. Fraternali, and M. Matera. Conceptual modeling of data-intensive web applications. *IEEE Internet Computing*, 6(4):20–30, 2002.

[4] M. Dean, D. Connoly, F. van Harmelen, J. Hendler, J. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. Owl web ontology language 1.0 reference. W3C Working Draft 29 July 2002.

[5] T. Isakowitz, M. Bieber, and F. Vitali. Web information systems. *Communications of the ACM*, 41(1):78–80, July 1998.

[6] G. Karvounarakis, V. Christophides, D. Dimitris Plexousakis, and S. Alexaki. Querying rdf descriptions for community web portals. In *17iemes Journees Bases de Donnees Avancees*, pages 133–144, 2001.

[7] N. Koch, A. Kraus, and R. Hennicker. The authoring process of the uml-based web engineering approach. In *First International Workshop on Web-Oriented Software Technology*, 2001.

[8] O. Lassila and R. R. Swick. Resource description framework (rdf) model and syntax specification. W3C Recommendation 22 February 1999.

[9] G. Mecca, P. Atzeni, A. Masci, P. Merialdo, and G. Sindoni. The araneus web-base management system. In *ACM SIGMOD International Conference on Management of Data*, pages 544–546. ACM, 1998.