MEDI-ADAPT

A Distributed Architecture for Personalized Access to Heterogeneous Semi-structured Data

Rim Zghal Rebaï, Amel Corinne Zayani, Ikram Ammous and Abdel Majid Ben Hamadou
Institute Multimedia, Information Systems and Advanced Computing Laboratory,
Higher Institute of Computer and Multimedia, PO Box 242, Sakiet Ezzit, 3021 Sfax, Tunisia

Keywords: AHA!, Navigation Adaptation, Content Adaptation, DARPA I3, Mediation, Semi-structured Data, XML, RDF.

Abstract: The increased volume of data in digital form has led to a wide variety of syntactic and semantic data. Thus, the user may encounter several problems related to the heterogeneity, distribution and volume of returned information. As a matter of fact, systems are needed to solve part or all of these problems. In this paper we propose and illustrate a distributed architecture that enables personalized access to collections of heterogeneous and distributed semi-structured documents (XML, RDF, SMIL). This architecture is based on an extension on the reference architecture of mediation systems DARPA I3 by adding an adaptation layer, based on the reference architecture for adaptive hypermedia systems AHA!, that takes into account the user's request, his context (profile, device, network, etc..) and data sources context (minimum bandwidth, necessary characteristics of the device to display data, etc.).

1 INTRODUCTION

Actually, data sources have become increasingly heterogeneous and distributed all over the world. As a result, the data volume grows and the user can not access to relevant information not only to his needs but also to his context. Thus, mediators are provided to resolve the problem of access to these sources regardless of their natures, semantics and locations. But these mediators, like IRO-DB (Gardarin, 1995), XMedia (Dang-Ngoc, 2008), (Kerzazi, 2008), etc., are unable to adapt the data and provide always the same information to users despite differences in their contexts (profile, device, network, etc.). This clearly shows the need for systems able to deal with both mediation and adaptation to provide a unified and transparent access to distributed heterogeneous data sources taking into account the user s' needs and context and the data sources' context.

With the lack of this type of systems, we propose an architecture of a distributed system mainly establishing components for solving problems related to remote and transparent access to document collections, adaptation of data and context management. The main objective of this architecture is to offer the user an adaptation of the content and navigation while accessing to collections of semi-structured documents that are distributed, syntactically and semantically heterogeneous and not designed to be adapted.

The paper is organized as follows: In section 2, we present a state of the art of some works dealing with mediation and adaptation. In Section 3, we introduce our proposal for a distributed architecture treating jointly mediation and adaptation. We focus in section 4 on the main layers of this architecture and their components. We conclude and give directions for future work in Section 5.

2 RELATED WORKS

The mediation systems allow the integration of heterogeneous and distributed data sources. All the proposed systems in the literature share a common reference architecture DARPA I3 (Wiederhold, 1992). This architecture has three layers: client, mediation and sources. From this reference architecture, different generations of systems have been developed based on the choice of pivot model and language. The first mediation systems were based on the relational model such as Multibase

Zghal Rebaï R., Corinne Zayani A., Ammous I. and Majid Ben Hamadou A..
MEDI-ADAPT - A Distributed Architecture for Personalized Access to Heterogeneous Semi-structured Data.
DOI: 10.5220/0003932402590263
In *Proceedings of the 8th International Conference on Web Information Systems and Technologies* (WEBIST-2012), pages 259-263
ISBN: 978-989-8565-08-2

(Landers, 1982), Mermaid (Templeton, 1987), etc. In the 90s, a second generation of object-oriented systems appeared such as IRO-DB (Gardarin, 1995), GARLIC (Carey, 1995), DISCO (Tomasic, 1998). Then, with the achieved success by XML (W3C, 1998), which is accepted as a standard for semi-structured data, a generation based on this language has emerged, such as XMedia (Dang-Ngoc, 2008). However, with the advent of the Semantic Web (Tim, 1999), XML has shown some limitations since it can provide only descriptive metadata, unlike RDF (Manola, 2004), which can provide both descriptive and semantic metadata. That's why, another RDF-based generation of mediation systems emerged. The advantage of such systems is that they take into account the semantics of data as the proposed systems in (Vdovjak, 2001), (Kerzazi, 2008).

Generally, when two users submit the same request (by query or link), mediators provide the same answers despite the difference of contexts. This means that the mediators do not deal with adaptation and do not take into account the context of the user when accessing to sources.

In the literature, several models and architectures dealing with adaptation are proposed. The first reference model is "Dexter" (Halasz, 1994). It is a local-oriented model proposed for hypertext applications. An extension of this model is proposed in (De Bra, 1999) called AHAM (Adaptive Hypermedia Application Model). This latter was extended in the local context to take into account semi-structured XML documents (Zayani, 2008).

The emergence of client / server architecture has led to the development of AHA! (Adaptive Hypermedia Architecture) (De Bra, 2003). It is a web-oriented architecture based on AHAM and considered as a reference architecture for adaptive hypermedia systems. Among the architectures based on AHA! we cite CA-WIS (Soukkarieh, 2010) which is proposed to take into account web services.

Another type of mediator-oriented architecture is proposed to allow the resolution of adaptation and mediation problems together. The only one is proposed in (Kostadinov, 2008). It offers a personalized access to many distributed relational data sources and uses XML as a pivot language in the mediator level. It adapts the content taking into account the user's profile and the available sources' quality. But in this work: (i) all of data are relational; (ii) the syntactic heterogeneity of these data is supposedly resolved; (iii) the semantic heterogeneity is neglected; and (iv) the navigation adaptation is not treated.

We notice that all the already mentioned studies adapt homogenous and known in advance data, except (Zayani, 2008) which adapts unknown local semi-structured data. Moreover, the navigation adaptation is not treated by the majority of these studies despite its importance, especially in the case of a large amount of data like in distributed environments. In these environments, the user can be easily disoriented and cannot get the required information, thus the benefits of navigation and content adaptation are more visible and efficient. That's why we propose a mediator-oriented architecture able to deal with; (i) the navigation and content adaptation, (ii) the access to semantic and syntactic heterogeneous and distributed semi-structured data (XML, RDF, SMIL) by using RDF as a pivot language in the mediator; and (iii) takes into account the variety in the contexts of users and sources.

3 THE MEDI-ADAPT ARCHITECTURE

We propose a distributed architecture that ensures the adaptation of content and navigation while accessing to heterogeneous sources not designed to be adapted.

It is considered to be innovative because, it executes multiple functions; it (i) deals with the navigation and content adaptation in a distributed environment; (ii) uses a mediator that provides transparent access to collections of semi-structured documents; (iii) treats syntactic and semantic heterogeneity of sources; (iv) takes into account both the user's context and the data sources' context.

This architecture is essentially based on a combination of DARPA I3 (reference architecture for mediation systems) and AHA! (reference architecture for hypermedia adaptive systems). We extended the DARPA architecture by adding an adaptation-providing layer. This layer is essentially based on the AHA!.

The architecture that we propose is consequently composed of four layers: client, adaptation, mediation and sources. As it is illustrated in Figure 1.

The client layer allows the user-system interaction. Thus, the user makes his access (by query or link) using various devices (PC, PDA, cell phone, etc.).

The adaptation layer receives the user's demand from the client layer and adapts the content and navigation according to the user's needs and context as well as the sources' context.

The mediation layer provides a unified and transparent access to different sources, regardless their syntax, semantics and location.

The sources layer integrates several collections of semi-structured heterogeneous and distributed documents.

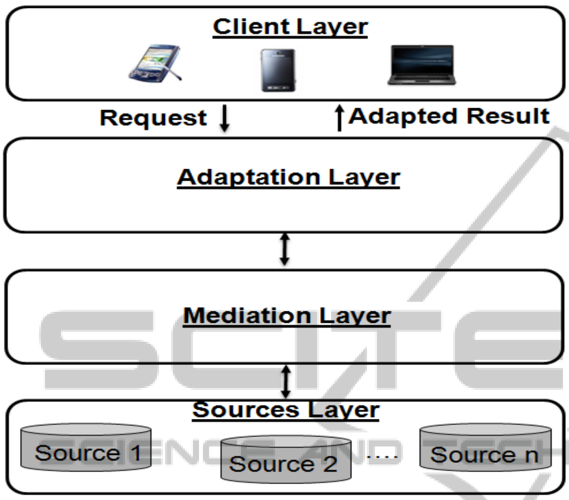

Figure 1: The MEDI-ADAPT architecture.

4 THE DESCRIPTION OF THE MAIN PROPOSED LAYERS

The main layers of the proposed architecture are the adaptation layer and the mediation layer. These two layers collaborate together to provide the content adaptation (Peter Brusilovsky,1996), the personalized access to the sources layer and the navigation adaptation (Peter Brusilovsky,1996) taking into account different contexts.

We detail below their components and their performance.

4.1 The Adaptation Layer

In the adaptation layer there are: a User Context Manager (UCM), a Query Interpreter (QI), a Content Adaptation Engine (CAE), an Updating Profile engine (UPE), a Navigation Adaptation Engine (NAE), a Response Generator (RG) and a Navigation Manager (NM).

This layer is illustrated in Figure 2.

Each component of this layer provides a well defined role:

- The UCM manages all components of the user's context: profile and environment (network and device). It is composed of a Profile Manager (PM), a User Model (UM) and an Environment Manager (EM). The UM stores the user's profile as proposed in (AHA!, (Zayani,2008), CA-WIS). We suggest that the user's profile is divided into two parts: one contains data concerned with content adaptation and another one contains data concerned with navigation adaptation;
- The CAE and the UPE, respectively, adapts the content and updates the user's profile. They are similar to the downstream adaptation mechanism and the updating profile mechanism proposed in (Zayani, 2008);
- The NAE adapts the navigation. It is based on an extension of the upstream adaptation engine proposed in (Zayani, 2008) by tacking into account the navigation adaptation;
- The RG performs the transformations generated by the NAE on the results data before being displayed to the user as it is the case with FAWIS (De Virgilio, 2007) and CA-WIS;
- The QI receives the user's request, analyses and transforms it into a SPARQL (Prud'hommeaux, 2008) (the W3C Recommendation for an RDF query language) query that is understandable to all the system components.

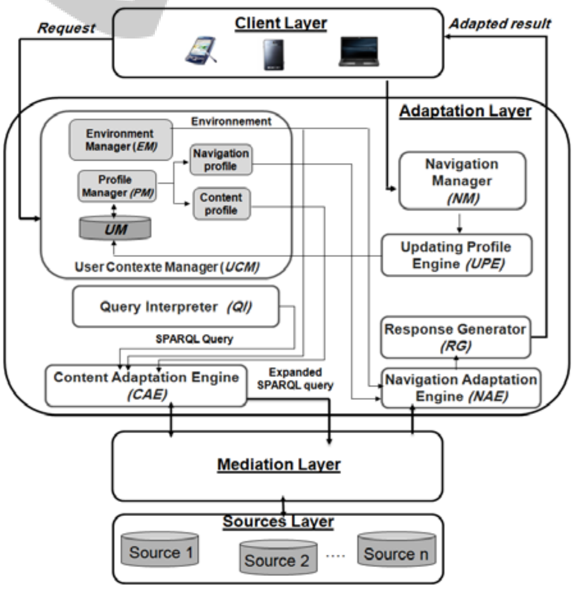

Figure 2: The adaptation layer's components.

When the user launches a request to the system, the UCM treats four actions simultaneously: by the PM, it (i) recognizes the user (ii) looks for his profile from the UM (iii) analyses it in order to differentiate between its two parts (content and navigation) and by the EM, it detects the user's environment (the used device and the network's characteristics). The QI receives the user's request, analyses and transforms it into a SPARQL query. Then, the CAE expands this query with the extracted data from the content part of the profile and sends it to the mediator. The latter brings the documents results from sources and sends them to the NAE. This engine receives from the UCM the navigation part of the profile and the user's environment characteristics to make the navigation adaptation and sends the result to the RG which, in its turn, generates the final result and sends it to the user.

Throughout each session, the NM detects the user-system interactions (visited links, accessed documents, duration of each document consultation, session's length, etc.) and sends them to the PUE.

4.2 The Mediation Layer

The mediation layer is mainly based on the basic components of XMedia (Tuyet et al., 2008): a Metadata Manager (MM), a Query Decomposer (QD), a Query Executor (QE), a Result Reconstructor (RC), a Metadata Base (MB) and a Memory Cache (MC). Except the Query Interpreter (QI) which is moved, as it is mentioned above, to the adaptation layer.

This layer is illustrated in Figure 3.

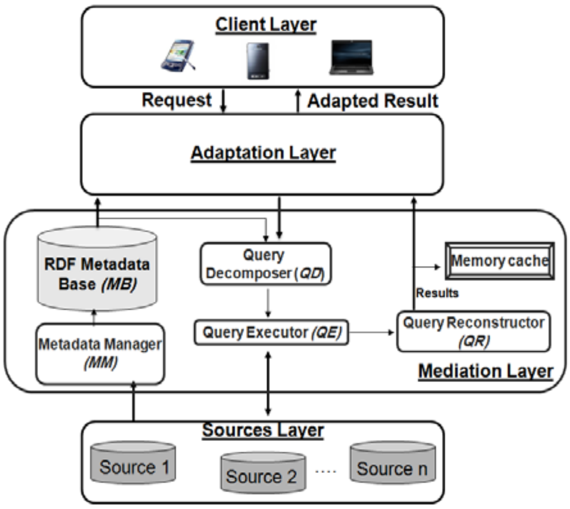

Figure 3: The mediation layer's components.

The components of this layer work as follows:

- The Metadata Manager (MM) provides descriptive metadata for each documents' collection, interprets and converts them to the pivot language RDF.
- The Query Decomposer (QD) decomposes the SPARQL query into atomic queries according to each single collection. It identifies the appropriate metadata sources, locates the collections and creates a query execution plan.
- The Query Executor (QE) executes the atomic queries near the sources, receives the results and sends them to the MC and the RC.
- The Result Reconstructor (RC) gathers all the returned results and sends them to the NAE.
- The Metadata Base (MB) stores the metadata related to documents' collections.
- The Memory Cache (MC) stores all the submitted data to the adaptation layer during a session.

When the CAE (component of the adaptation layer) sends a query to the mediation layer, the QD receives and decomposes it into atomic queries according to each single collection. It identifies the appropriate metadata sources, locates the collections and creates a query execution plan. The QE receives all the atomic queries and executes them near the sources, then it receives the results from there and sends them to the MC and the RC. The latter gathers all the returned results and sends it to the NAE which performs the navigation adaptation task as it is mentioned above. Finally, the RG sends the final result to the user.

Permanently, the MM connects with different sources to provide descriptive metadata for each documents' collection, converts all of them to RDF to be stored in MB and identifies the global schema.

5 CONCLUSIONS

In this paper, we presented a distributed architecture that solves both mediation and adaptation problems together. It is essentially based on a combination of DARPA I3 (reference architecture for mediation systems) and AHA! (Reference architecture for hypermedia adaptive systems).

On the one hand, this architecture offers a unified and transparent access to distributed collections of semi-structured documents through the mediation layer. On the other hand, it offers a navigation and content adaptation taking into account the user and sources contexts through the adaptation layer. The syntactic and semantic heterogeneity of data is solved by the use of RDF as a pivot model at the mediation layer.

There are many perspectives that we are willing to realize. First, we plan to propose and evaluate a navigation adaptation method on which the Navigation Adaption Engine (NAE) will be based (already in the evaluation step). Second, we aim to propose a content adaptation method on which the Content Adaptation Engine (CAE) will be based. Third, we will suggest methods that treat the data heterogeneities (syntactic and semantic) and distribution. Also, we intend to propose, at the Profile Manager (PM), a method that reduces the user profile to the most relevant content. Finally, we are going to implement a prototype based on the proposed architecture.

REFERENCES

Brusilovsky, P., 1996. Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction, 6(2-3):87–129*.

Carey, M., Haas, L., Scwartz, P., 1995. *Towards Heterogeneous Multimedia Information Systems: The Garlic Approach*, in RIDE-DOM, pages 124-131, 19.

Dang-Ngoc, T., Gardarin, G., 2008. Conception et Evaluation de XQuery dans une architecture de médiation “Tout-XML”. *CoRR abs/0806.4920.*

De Bra, P., Geert-Jan, H., Hongjing, W., 1999. AHAM: A Dexter-based Reference Model for Adaptive Hypermedia. *Hypertext 99 Darmstadt Germany*.

De Bra, P., Aerts, A., Berden, B., De Lange, B., Rousseau, B., Santic, T., Smits, D., Stash, N., 2003. AHA! The Adaptive Hypermedia Architecture. *HT’03.*

De Virgilio, R., Torlone, R., Houben, G., 2007. Rule-based Adaptation of Web Information Systems. *World Wide Web.*

Gardarin, G., Fankhauser, P., Finance, B., Klas, W., Ramfos, A., Gannouni, S., Pastre, D., Legoff, R., 1995. IRO-DB: A Distributed System Federating Object and Relational Databases. *Object-oriented Multibase Systems, Prentice-Hall.*

Halasz, F., Schwartz, M., 1994. The dexter hypertext reference model. *ACM, 37(2):30–39.*

Kerzazi, A., Chniber, O., Navas-Delgado, I., F.Aldana-Montes, J., 2008. A Semantic Mediation Architecture for RDF Data Integration. *SWAP’08.*

Kostadinov, D., Bouzeghoub, M., Lopes, S., 2008. Accès personnalisé à des sources de données multiples. *Ingénierie des Systèmes d'Information 13(4): 59-82.*

Landers, T., Rosenberg, R. L., 1982. An overview of MULTIBASE. *In distributed databases, H.J Shneider editor, North-holland*.

Manola. Frank., Milles, E., 2004. RDF Primer. W3C Recommendation, *http://www.w3.org/RDF/.*

Prud'hommeaux, E., Seaborne, A. (eds), 2008. SPARQL Query Language for RDF. W3C Recommendation, *http://www.w3.org/TR/rdf-sparql-query/.*

Soukkariah, B., 2010. Technique de l’Internet et ses langages: Vers un Système d’Information Web restituant des Services Web sensibles au contexte Doctoral thesis, University of Toulouse 3 – Paul Sabatier, Toulouse, France.

Templeton, M., Brill, D., Dao, S. K., Lund, E., Ward, P., Chen, A. L., Macgregor, R., 1987. Mermaid: a Frontend to Distributed Heterogeneous Databases. *In proceedings of the IEEE conference on Data Engineering.*

Tim, B., Mark, F., 1999. *Weaving the web, chapter Machines and the Web*, pages 177 – 198. Harper, San Francisco.

Tomasic, A., Raschid, L., Valduriez, P., 1992. Scaling Access to Heterogeneous Data Sources with DISCO, *IEEE Data and knowledge Engeneering, 10(5):8, No. 3, 3849.*

Vdovjak, R., Houben, G., 2001.RDF Based Architecture for Semantic Integration of Heterogeneous Information Sources, *In Workshop on Information Integration on the Web, pp. 51-57*.

Wiederhold, G., 1992. Mediators in the architecture of future information systems. *IEEE Computer Magazine, Vol. 25, No. 3, 3849.*

W3C, 1998. EXtensible Markup Language (XML) 1.0. Technical report, World Wide Web Consortium (W3C), Technical report, February.

Zayani, A., 2008. Contribution à la définition et à la mise en œuvre de mécanismes d’adaptation de documents semi-structurés. Doctoral thesis, University of Toulouse 3 – Paul Sabatier, Toulouse, France.