Semantic Resource Adaptation Based on Generic Ontology Models

Bujar Raufi, Florije Ismaili, Jaumin Ajdari, Mexhid Ferati and Xhemal Zenuni
Faculty of Contemporary Sciences and Technologies, South East European University, Ilindenska 335, Tetovo, Macedonia

Keywords: Semantic Web, Linked Open Data, User Adaptive Software Systems, Adaptive Web-based Systems.

Abstract: In recent years, a substantial shift from the web of documents to the paradigm of web of data is witnessed. This is seen from the proliferation of massive semantic repositories such as Linked Open Data. Recommending and adapting resources over these repositories however, still represents a research challenge in the sense of relevant data aggregation, link maintenance, provenance and inference. In this paper, we introduce a model of adaptation based on user activities for performing concept adaptation on large repositories built upon extended generic ontology model for Adaptive Web-Based Systems. The architecture of the model is consisted of user data extraction, user knowledgebase, ontology retrieval and application and a semantic reasoner. All these modules are envisioned to perform specific tasks in order to deliver efficient and relevant resources to the users based on their browsing preferences. Specific challenges in relation to the proposed model are also discussed.

1 INTRODUCTION

The expansion of the web in the recent years, especially with the proliferation of new paradigms such as semantic web, linked semantic data and the social web phenomena has rendered it a place where information is not simply posted, searched, browsed and read. The web has been transformed into a place where content and user experiences can be adapted in various ways for a single purpose of action, which is meeting user's needs. This adaptation is being conducted through Adaptive Web-Based Systems, which tend to arrange its internal link structure, content, or both, based on user browsing patterns.

User adaptive software systems in general and adaptive web-based systems in particular have been developed in two distinct pillars. One was particularly focused on document space related to the area of Personalized Information Retrieval(PIR) and the other originating from the hypertext space in the field of Adaptive Hypermedia (AH) (Steichen, 2012). PIR extends the classical view of one-size-fits-all paradigm by taking into account historical interactions and finding most relevant documents for a single query. On the other hand, AH approach tends to ease this sort of search by providing the most relevant browsing content and linking it with respect to a rich representation of user characteristics. Such characteristics could be preferences, history or prior knowledge. While PIR is mostly query based, AH is mainly browsing oriented.

An ongoing challenge is the semantic retrieval and adaptation of resources over a huge repositories such as Linked Open Data (LOD) as well as the facilitation of the collaborative knowledge construction by assisting in discovering new things (Bizer, 2009). Likewise, recommending and adapting resources over a huge semantic repositories such as LOD (Linked Data, 2013) still represents an open issue regarding the relevant data aggregation, link maintenance, provenance and inference. Raufi et al. (2011) outlines the importance of adoption of Semantic Web Technologies for content adaptation. The reason for such adoption, which can be furthermore extended to big resource repositories, can be summarized as follows:

1. Semantic web can be used to describe every document or other resources in adaptive web-based system's repository (documents and any other smaller units) according to a given vocabulary.
2. After their semantic description, these resources can become machine processable and conceptually determinable.
3. These aforementioned resources are scaling optimally with no particular increase in processing power. Such examples have been already tested and are up and running in multiple Triple Stores like: AllegroGraph, Stardog, OpenLink Virtuoso,

Raufi B., Ismaili F., Ajdari J., Ferati M. and Zenuni X..
Semantic Resource Adaptation Based on Generic Ontology Models.
DOI: 10.5220/0005096001030108
In *Proceedings of the 9th International Conference on Software Paradigm Trends* (ICSOFT-PT-2014), pages 103-108
ISBN: 978-989-758-037-6

BigOWLIM, Garlik4store, Bigdata(R), YARS2, Jena TDB, Jena SDB, Mulgara, RDF gateway or Jena with PostgreSQL (Large Triple Store, 2014).

4. These semantically describable resources can be further on queried through various endpoints and presented to the user.

In this paper we introduce an architecture model for semantic resource adaptation over large repositories based on generic ontologies. The generic ontologies in this model are considered those ontologies that can be successfully retrieved and applied from the semantic reasoner in order to perform the adaptation process.

The rest of this paper is organized as follows: section 2 introduces technology background related to semantic web, section 3 describes some related work introduced to the topic, section 4 elaborates the architecture of the proposed model for semantic resource adaptation based on generic ontologies, it describes its mechanisms through a simple use case and opens up some interaction design challenges in relation to resource adaptation and finally, section 5 concludes this paper.

2 BACKGROUND

The main techniques used to represent the semantic web approach in our proposed model is based on Resource Description Framework (RDF), RDF Schema (RDFS) and Web Ontology Language (OWL) which is portrayed as a standard in ontology modeling and representation.

RDF is categorized as a general-purpose triple-based language used for describing information on the Web. RDF statements are defined as *subject-predicate-object triples*. From the perspective of LOD, RDF offers flexibility in the sense of publication, but creates a drawback in the point of view of scalability (Fernandez, 2012). Flexibility of RDF is seen through the aspect of resource exchange (many proprietary formats such as RDF/XML (RDF/XML, 2004), Notation3 (N3) (Berners-Lee, 2011), Turtle (Turtle, 2014) and RDF/JSON (RDF/JSON, 2013)) and resource publication (dereferenceable URIs, RDF data dumps or SPARQL endpoints (SPARQL, 2011). Drawback on the other hand is seen in the process of resource consumption which is constrained by factors such as serialization formats of RDF resources and indexing of the same (Fernandez, 2012).

RDF Schema (RDFS) represents a semantic extension of RDF which provides mechanisms for describing groups of related resources (classes or properties) and the relationships between these resources. RDFS allows extension of the classical triple based RDF in the sense of well defined classes and properties.

Web Ontology Language (OWL) is used to render the available resources to be machine processable, where the meaning of each term and their relationships are explicitly denoted. OWL has more extended mechanisms for expressing meaning and semantics than XML, RDF, and RDFS, thus OWL goes beyond these languages in its ability to represent machine interpretable content on the Web (Cristophides, 2008).

The important aspect related to ontologies is that there are very solid reasoners or inference engines available (HermiT, Fact++, Pellet, RacerPro etc.) that can draw fair conclusions in huge semantic resources at a reasonable time (Dentler, 2011). These vast resources, however, pose a challenge in deciding the most relevant resources for a given task. To address this issue, we introduce a model of adaptation by grasping user activities for performing concept adaptation on large repositories such as linked open data based on extended generic ontology user model for Adaptive Web-Based Systems initially introduced in (Raufi, 2009).

3 RELATED WORK

The undergoing research concerning adaptation of resources in huge semantic repositories such as Linked Open Data is focused around potential identification of complementary feasibilities between Personalized Information Retrieval (PIR) and Adaptive Hypermedia (AH) (Steichen, 2012). The above mentioned possible complementariness gives various ways for hybrid approaches which allow techniques in relation to retrieval process such as: query adaptation, adaptive retrieval and adaptive composition and presentation.

In query adaptation techniques, the focus is given mainly on the process of query disambiguation, personalization and result diversification. All this in relation to medium and small closed corpus semantic repositories and no user model is available. Adaptation technique is focused around semantic reasoning, query expansion or substitution. An example of this group is seen in (Chirita, 2007) where query adaptation is performed through the process of query expansion with new terms depending on the user experience (novice or experienced visitors) or query ambiguity.

Concerning adaptive retrieval techniques, the center of attraction is given mainly to content-concept mapping around small open/closed corpus domain, user model represents an overlay knowledge model based on preferences and adaption algorithms are mainly rule based. One promising semantic adaptation techniques lies in the field of Open Hypermedia Systems (Wiil, 2006). Even though, OHSs do not provide direct adaptation, they can be combined with such systems to provide various personalization strategies. One approach of such model is elaborated in (Fulantelli, 2000).

In adaptive resource composition and presentation, a more distinctive grouping between information retrieval and adaptive hypermedia techniques is evident. This distinction is seen from the perspective of adaptive behavior, user involvement and modeling, adaptation algorithms and their scalability. In this sense, techniques such as: faceted ranking search, adaptive navigation, user intent elicitation, rule based and semantic reasoning for adaptive composition from one side and adaptive summarization, link annotation, cluster selection, preference summarization for adaptive presentation from the other are available (Steichen, 2012).

It is worth mentioning that partial solutions of hybrid systems that tend to fuse the approaches of PIR and AH are available. One such solution is given in (Shekhar, 2013) where the application is consisted of three modules: web extraction module which contains the user browsing history, the semantic knowledge base module which contains the ontology in accordance with the user's browsing history and finally the reasoner module which parses through the ontology and generates a result based on a search query. Another methodology is seen in (Bielikova, 2006) and (Bielikova, 2008) which tends to model the content of adaptive web application based on ontology. The ontology is modeled from a domain model and a domain dependent part of a user model. While the former work focuses more on content modeling using an ontology, the latter tends to design the ontological representation of a model which is suitable for further reuse.

Main drawback of the above elaborated approaches lies mainly in their closed corpus nature, considering that, to the best of our knowledge, they have not been designed and tested on large repositories such as the Web of Data. In the section that follows, we elaborate a model designed specifically to address the complexity and vastness of large semantic repositories with a particular emphasis on Linked Open Data.

4 PROPOSED ARCHITECTURE

In order to address the issues risen above, a hybrid architecture that focuses on the issues of both personalized information retrieval in semantic level and resource adaptation is proposed. The architecture is envisioned of four main modules with specific tasks. The semantic resource adaptation architecture is illustrated in Figure 1 and is consisted of the following elements:

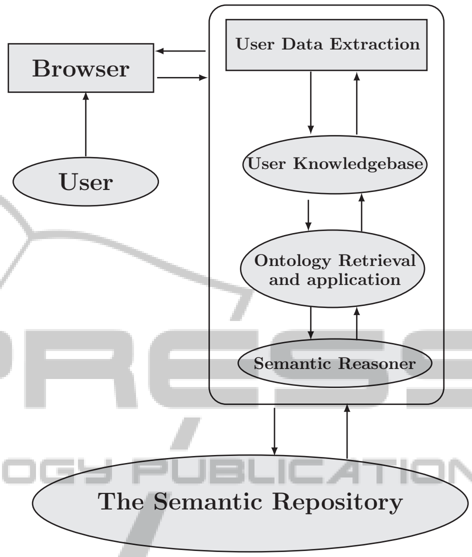

Figure 1: The architecture of a semantic resource adaptation.

User data extraction module will extract the user browsing activities and insert them to the user knowledge base (ontology) as object instances or data properties. User browsing activities are kept in various formats (logs, database tables, etc) on the browser side and requires additional intermediary tasks for such extractions and data mappings.

User knowledge base represent an extended version of the ontology initially introduced in (Raufi, 2009) that depicts user activities on the semantic level. These activities usually comprise user sessions, visited content and generated user views as part of the adaptation. Browsing activities from the data extraction phase given earlier, are also stored on the ontology in the form of object instances or data properties.

Ontology retrieval and application is a phase of ontology aggregation and fetching from various repositories based on user browsing behavior. For instance, say a user is interested in a particular topic, and in order to adapt to that concrete user activity, a specific ontology from publicly available repositories (Cupboard, Knoodl, Schemapedia, SchemaWeb, TONES, etc.) should be acquirable.

Finally, the *reasoner* module makes it possible the connection between the previous modules to work together in the background. The reasoner draws conclusions on instances and data properties based on the classes, class hierarchies and various restriction properties defined explicitly in the ontologies. The important part of the architecture lies in the user model defined as the user knowledge base and the ontology retrieval as depicted in Figure 1 above.

The steps involved in the creation of the ontology comprises of the following aspects:

- Definition of concept classes and hierarchies
- Definition of concept properties
- Definition of concept relationships and
- Definition of concept instances and data properties

Main concept classes in the user knowledge base ontology are as follows: *User, Session, Preferences, UserView, Content, Page and AtomicUnit*. Each of these classes are hierarchically defined in a parent-child relationship. To summarize shortly, a user can have a particular session, preference and a view which in turn is consisted of a resource content such as pages or more fine grained elements we coined as atomic units (Raufi, 2009). All of the above entities are related at semantic levels through concept relationships such as: *hasSession, hasPreference, hasUserView* etc.

The complete ontology and their relationships is depicted as in Figure 2.

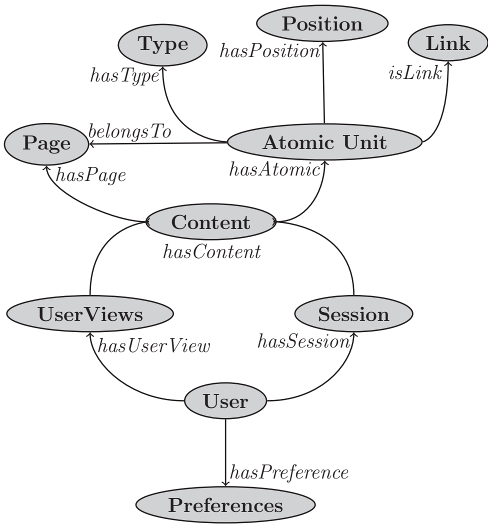

Figure 2: The User Ontology.

During ontology retrieval and application phase, the *preferences* class from user ontology given above serves as an input to ontology retrieval. After the ontology is retrieved, a proper class hierarchy, domain attributes as well as datatype identification is performed through ontology parsing. Finally, the retrieved elements are returned to Content class in user knowledge base which will serve as inputs for *UserView* generation.

The detailed view of ontology retrieval, parsing and application is illustrated in Figure 3. The dashed lines illustrate the process of resource return to Content class of user knowledge base ontology.

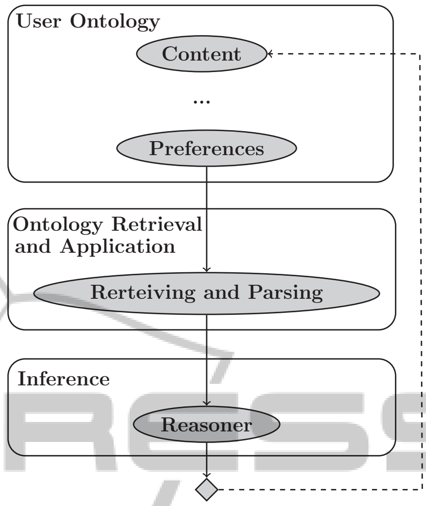

Figure 3: Detailed view of ontology retrieval and application.

4.1 Semantic Resource Adaptation: A User Scenario

John is visiting Berlin for the first time. He is using his mobile device to explore interesting sites around the city. A friend has recommended him to visit Reichstag's building, which he types in his semantically enabled mobile web browser. The browser shows important details about Reichstag's building along with a link 'Directions', which shows John how to get to the building (the browser does this based on John's smartphone geolocation). John decides to click on the link and follow the directions. Leveraging geoname (GeoNames, 2014) information, the browser suggests John to also visit Jakob Kaizer Haus (the browser now suggests other important sites following John's trail to get to Reichstag's building). John initially did not know anything about this building, but since it is on his way, he is glad to explore it.

The typical course of events for this particular user scenario seen from systems perspective involves the following steps:

Step 1: User visits a particular semantically capable web page.

Step 2: User wants to know more for a particular resource.

Step 3: The application checks for other relevant resources based on retrieved ontologies and derived inference.

Step 4: These resources, in the form of highlighted content or other visually aided methods is presented to the user.

Alternate courses of events might also involve:

Step 5: User browsing patterns are highly idiosyncratic so that a proper preference and consequently, a relevant resource cannot be identified.

Step 6: In case of step 5, no adaptive resource is delivered.

The above elaborated use case also implies several user interaction challenges that also need to be addressed. These challenges can be viewed from the perspective of resource entities, user focus preservation and user friendliness.

4.2 User Interaction Challenges

The key benefit of large semantic repositories such as linked open data seen from user's perspective mentioned above, is the arrangement of integrated data access from heterogeneous data sources which initially might not be accessible to users (Bizer, 2009). In this direction the following several challenges could be identified:

Entity-centered User Interface: In a typical web browsing scenario users navigate from page to page, which represents the page-centered navigation. This type of navigation, however, is not suitable for LOD as it does not leverages the full potential of these large semantic repositories. To offer a heightened user experience a new paradigm of entity-centered navigation should be adopted. Such, navigation will allow a fine-grained interaction between entities that will generate a more specific entity to entity access compared to page to page. In order to offer the best user experience, the entity-centered navigation should allow bidirectional browsing between underlying entities. For instance, referring to the above mentioned scenario the user moves from *Raichstag* subject entity to *Jakob Kaizer Haus* object entity connected through a particular predicate like e.g. *onTheWaySites*. Despite that such scenario should be unambiguous from the machine processing point of view, it must be natural and cognitively plausible from user experience perspective.

Maintaining User's Focus: When navigating large semantic repositories, users could lose their focus from the amount of options offered. Maintaining user's focus based on the selected preferences is a challenging task. For instance, while John, on his way to Reichstag's building he should be offered with other relevant sites on his way to destination.

User Friendly Information Representation: The LOD functions on the principle of displaying data organized in triplets: object-predicate-subject. Typically, such information is shown as it is without guiding the user through the process of information discovery. While such navigation is effective, it undermines the serendipity of information, which results in low user experience. For instance, when John searches for Reichstag he is offered with directions link, which depicts how the interface should anticipate user's possible future interests.

Some aspects of the issues presented are partially addressed in (Catasta, 2009) through Sigma search engine. However the approach gives an insight of the limits in capabilities and functionalities of the system when huge repositories are the case.

5 CONCLUSIONS AND FUTURE WORK

In this paper, an architecture of semantic resource adaptation on massive repositories is introduced. The system tends to adapt its content based on user preferences, retrieved external ontologies and reasoning mechanism. The process of adaptation is done entirely through reasoners and reasoning techniques taken from semantic web technologies. The fetched resources are returned to user ontology as content and proper user views.

The future work would involve:

- Development on specific parts of the architecture modules and implementing the system as a whole. Some part of the architecture elaborated above are implemented such as the user knowledgebase as given in Figure. 2 and some initial results regarding reasoning on that knowledgebase.
- Exploring various ways of reasoning mechanisms and reasoners against our architecture. Comparison analysis between reasoners such as Pellet, Fact++ and Hermit on one hand and SWRL rule language on the other hand would be interesting from the sense of retrieved resources relevant to the user.
- Evaluation of the aforementioned proposed architecture against these large linked repositories from the perspective of its efficiency, resource recall and precision. One particular interesting evaluation would be from the perspective of scalability having in mind that we are dealing with huge amounts of data.
- Investigating best interaction modality to provide the retrieved resources to users. This will involve

experimenting with various adaptive interfaces to offer users rich user experience.

If these future task are to be addressed properly, we expect that the proposed architecture will substantially facilitate the access to linked open data as well as realize the full potential of the web not only as document but also as a data information space.

REFERENCES

Bizer, C. Heath, T., Berners-Lee, T. (2009). Linked data-the story so far. In *International Journal on Semantic Web and Information Systems (IJSWIS).* 1-22. 5/3

Berners-Lee, T., Conolly. D. (2011). Notation3 (N3): A readable RDF syntax. *W3C Recommendation* [Online]. http://www.w3.org/TeamSubmission/n3/ Accessed: April, 2014.

Bielikova, M., Moravcik, M. (2008). Modeling the Reusable Content of Adaptive Web-Based Applications Using an Ontology. In *Advances in Semantic Media Adaptation and Personalization Studies in Computational Intelligence.* Vol. 93. pp. 307-327

Bielikova, M., Moravcik, M., (2006). Modeling the Content of Adaptive Web-Based System Using an Ontology. In First International Workshop on Semantic Media Adaptation and Personalization, SMAP '06. vol., no., pp.115-120

Catasta, M., Cyganiak, R., Tummarello, G. (2009). Towards ECSSE: live Web of Data search and integration. In *Proceedings of the Semantic Search 2009 Workshop at WWW2009.*

Chirita, P., Firan, C. S., and Nejdl, W. (2007). Personalized query expansion for the web. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* SIGIR '07, pp. 7-14

Cristophides, V., Collard, M., Gutierez, C. (2008). Semantic Web, Ontologies and Databases. In *VLDB Workshop, SWDB-ODBIS.* Lecture Notes In Computer Science (LNCS). 5002

Davis, I., Steiner, Th., Le Hors, A.J. (2013) RDF 1.1 JSON Alternate Serialization (RDF/JSON). *W3C Recommendation. 2013*

Dentler, C., Cornet, R., Teije, T.A, Keizer, N. (2011). Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile. In *Journal of Semantic Web.* (1) 1-5. IOS Press.

Fernandez, H. D. (2012). Binary RDF for scalable publishing, exchanging and consumption in the web of data. In *Proceedings of the 21st international conference companion on World Wide Web (WWW12).* 133-138.

Fulantelli, G., Rizzo, R., Arrigo, M., Corrao, R. (2000). An Adaptive Open Hypermedia System on the Web. In *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH '00).* Brusilovsky, P., Stock, O. and Strapparava. C. (Eds.). Springer-Verlag, London, UK, UK, 305-310.

GeoNames geographical database. (2014). The GeoNames geographical database. [Online] http://www.geonames.org/ Accessed May, 2014

Raufi, B., Luma, A., Zenuni, Xh., Ismaili, F. (2011). Semantic Adaptation Approach for Adaptive Web-Based Systems. In *Proceedings of the 2nd International Science and Technology Conference (ISTEC'2011).* 705-708. Istanbul, Turkey

Raufi, B. Ismaili, F. Zenuni, X. (2009). Modelling a complete ontology for adaptive web based systems using a top-down five layer framework. In *Proceedings of the 31st International Conference on Information Technology Interfaces (ITI 2009).* pp. 511,518, 22-25 June 2009. Dubrovnik/Croatia

Shekhar, M., Saravanaguru RA. K. (2013). Semantic Web Search based on Ontology Modeling using Protege Reasoner. arXiv preprint arXiv:1305.5827.

Steichen, B. Ashman, H. Wade, V. (2012). A comparative survey of Personalised Information Retrieval and Adaptive Hypermedia techniques. In *Journal of Information Processing and Management.* 48, 4, pp. 698-724.

SPARQL 1.1. Query Language. (2011). *W3C Recommendation. 2011*

RDF/XML Syntax Specification (Revised). (2004). *W3C Recommendation. 2004*

Turtle - Terse RDF Triple Language. (2014). *W3C Recommendation.* [Online] http://www.w3.org/TR/2014/REC-turtle-20140225/ Accessed April, 2014

Linked Data. (2013). [Online] http://www.w3.org/standards/semanticweb/data. Accessed April, 2014

Large Triple Store. (2014). [Online] http://www.w3.org/wiki/LargeTripleStores Accessed April, 2014

Wiil, U. (2006). Open Hypermedia: Systems, Interoperability and Standards. In *Journal of Digital Information.* 1(2). Retrieved from http://journals.tdl.org/jodi/index.php/jodi/article/view/9/9