A Semantic Service-Oriented Platform for Integrated and Personalized Access to Sustainable Projects*

Felipe Costa, Carlos Dantas, Tatiana Sorrentino, Claudia Ribeiro
State University of RN-UERN, Brazil

This paper reports the achieved results of a project that aims at the development of a technological apparatus to support a sustainable development agenda. The resultant platform named ARIDUS concentrates specific services, such as capture and representation of knowledge behind sustainable projects. Initially addressed to the Brazilian Semiarid, ARIDUS can be used in different contexts based on its key features such as service-oriented principles and semantic technologies.

Keywords: SOA, semantic web, sustainability

Introduction

Sustainable development has been considered as an item of high priority in the decision makers’ agenda. This revisited and modern concept of development states that not only environmental issues must be addressed, but also economic and social ones. In other words, considering sustainability, a project must be “environmentally correct, economically feasible and socially fair” (Brundtland, 1987).

Despite the novelty related to sustainability issues, it is clear that computational models are critical to tackle the complex problems involved, as pointed out by the recent report of the National Academy of Sciences (Millett, et al., 2012). Among others, this report lists the heterogeneity of data and decision making under uncertainty as essential elements of computer science research agenda for sustainability.

This paper presents the results of a project that addresses some of the most challenging aspects for provisioning support to sustainable development. The challenges and proposed solutions to provide such functionalities are described in this paper, which is structured as follows.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Motivational scenario presents a scenario that illustrates barriers that could prevent the easy access to relevant information about sustainable projects. ARIDUS architecture overview and its services are presented in the ARIDUS platform. Accessibility and personalization issues are discussed in the case studies. Related

* Acknowledgments: The authors gratefully acknowledge the financial support of CNPq Brazilian Agency (Grant Proc. 475081/2011-0-14/2011).
Felipe Costa, master, Lumen Lab, State University of RN-UERN.
Carlos Dantas, master, Lumen Lab, State University of RN-UERN.
Tatiana Sorrentino, master, Lumen Lab, State University of RN-UERN.
Claudia Ribeiro, doctor, Lumen Lab, State University of RN-UERN.
works present some related work and the conclusion presents some conclusions and perspectives for future work.

**A Motivational Scenario**

The intrinsic complexity concerning the integrated modeling of sustainability issues requires specific strategies. In this work, it was defined the Brazilian semiarid region as a geographic boundary. This area is representative as a case study because of its environmental susceptibility, economic particularities, and social conditions.

Bearing in mind that a trade off among environmental, economic, and social aspects is hard to achieve, decision makers have been hardly challenged to propose feasible and effective projects for this region. Tools to support them have become critical, mainly to help the process of selection sustainable projects, by considering priorities and the available resources. However, this is not a trivial task, since meaningful and complete sustainable projects descriptions are rare. Besides, they are frequently dispersed in different sources, use different data formats, and quite often are saved inside private repositories.

Technical aspects of data source integration have been intensively investigated (Liao et al., 2011). More recently a movement for publishing and connecting data on the web named linked data (Bizer, Heath, Berners-Lee, 2009) has gained momentum. Sustainability and more specifically sustainable development have pushed some boundaries, mainly concerning integrated modeling and assessment (Harris, 2002). However, many other challenging issues have appeared, for instance, mitigating the negative impacts of natural resources on one hand, by maximizing the positive impacts, such as job offering and diversity of economic activities on the other hand. Then it is critical to assess the sustainability level of projects before deciding which of them will deserve public fundraising.

At the Brazilian semiarid, for instance, sustainable agriculture projects based on intercropping and family farming quite often dispute space with wind farm projects that produce clean energy. In this case, political decision must consider not only environmental aspects, but also social and economic ones.

![Figure 1. Motivational scenario.](image)

Many critical specialists questions claim to be answered. Thus, clear information is needed, for instance, to what extent projects planned activities may affect the environmental, economic, and social dimensions.
In order to deepen the implications of problems that have emerged, it is necessary to go beyond interoperability issues. For sake of illustration, two projects should be considered that are suitable for the Brazilian semiarid: intercropping agriculture and recovery of degraded area. Despite of the appeal of their titles, these projects must be meaningfully described in order to have their sustainability level properly assessed. For this reason, sustainable assessment commonly involves different specialists as illustrated by Figure 1.

However, it is important to consider that projects descriptions can potentially encompass subjective perceptions that are highly dependent on specialists’ background. In this context, semantic differences alignment (Euzenat et al., 2011) among perceptions plays a key role. Even if projects were individually considered, the situation would imply dealing with knowledge from different disciplines.

For instance, the project team of a sustainable agriculture project should have deep knowledge about plants and their specific characteristics, as well as climate conditions and the requirements to apply the techniques. They also need to be aware of economic and social indicators, such as return investment rate, human rights, and safety. In fact, many sustainability indicators have been used to assess the level of sustainability of projects (Singh et al., 2009).

The aforementioned aspects do not exhaust research topics, but they were enough to motivate this proposal, which consists in a service-oriented platform boosted by semantic technologies, whose technical details are discussed at the next section.

**Aridus Platform**

**Architecture Overview**

ARIDUS platform has been conceived under openness and flexibility principles, by means of a four-layer service-oriented architecture, as depicted by Figure 2. Heterogeneous data sources, such as relational databases, public repositories, metadata, and knowledge bases, are virtually grouped and treated at the Storage Layer.

The Semantic Integration Layer presents services that use ontology to find out similarities between different terms used as input. The main purpose of this layer is to amplify the search space by leveraging integration issues to more abstract levels.

At the Extraction and Annotation Layer, services deal with metadata that process documents before use them in semantic queries by using SPARQL\(^1\). Particular attention is paid to textual documents, such as thesis and projects related to the Brazilian semiarid. External services like GATE\(^2\) are planned to text analysis and automatic knowledge extraction. For now, manual annotation service allows the inclusion of links to access such documents.

The Presentation Layer is the most abstract layer. It is dedicated to present the answers for questions, such as those depicted by Figure 1. Services at this layer have two major concerns: content recommender and accessibility.

\(^1\) http://www.w3.org/TR/rdf-sparql-query/
\(^2\) http://gate.ac.uk/
ARIDUS Services

ARIDUS services have been developed as complementary facilities to extend applications functionalities, by improving context-aware information gathering. In this sense, there was no particular concern about ARIDUS data presentation. This section presents technical details of ARIDUS services by taking Figure 2 as a guide.

1. **DataSetManager**: keeps essential information to access data sources described by VoID (Vocabulary of the Interlinked Datasets), such as URI and foaf:homepage, dcterms:title;

2. **AlignmentManager**: deals with RDF alignments imported to ARIDUS by using external services like Aser³, a third party service for ontology alignments;

3. **UserProfileManager**: keeps user profile information, including preferences and constraints;

4. **DocBaseManager**: deals with textual documents;

5. **DocBaseWrapper**: Endpoint SPARQL to access documents at ARIDUS DocBase;

6. **QueryRewriter**: rewrites SPARQL queries by using imported alignments for some terms used as input;

7. **QueryProcess**: orchestrates sequences of SPARQL tasks for data sources integration;

8. **QueryExecutor**: executes SPARQL composites of endpoints having as a basis the FedX framework⁴;

9. **PersonProfile**: deals with user profile ontology;

10. **ContentRecommender**: seeks content according to user profile interests;

11. **ContentAdapter**: makes dynamic changes in web pages structure in order to adapt its content to the user profile;

12. **ContentViewer**: deals with the visualization of content after the execution of ContentRecommender and ContentAdapter services.

---

³ http://aserv.inrialpes.fr/
⁴ http://www.fluidops.com/fedx/
ARIDUS services can be executed individually or as a composite to accomplish more complex tasks. This is the case of the query executor, which includes alignment, documents selection, and SPARQL rewritten services.

It is worth noting that the nature of the challenges presented by the motivational scenario (see Figure 1) is not technological only. Actually, the modeling aspects related to the expertise extraction from textual documents have been proved to be more demanding than the application of novel technologies. The next sections discuss some of these aspects, namely personalization and accessibility.

Conceptual Model for Personalization

ARIDUS services use a semantic user profile to improve the level of personal content recommendation. This user profile named PerSoN includes SIOC\(^5\) that provides the description of online communities and FOAF\(^6\) concepts for user profiles description as depicted by Figure 3.

PerSoN also includes concepts related to accessibility defined by @dapt ontology that describes visual

---

\(^5\) http://sioc-project.org/ontology

\(^6\) http://www.foaf-project.org/
impairments concepts as illustrated by Figure 4. @dapt concepts includes the degree of blindness the user has. It also encompasses specific information related to the context in which the user is located, and uses it to adapt dynamically the content delivery.

In a pragmatic way, PerSoN ontology represents an enriched user profile, by including not only disabilities but also social networks information, as depicted by Figure 5. Despite its simplistic appearance, this integrated model allows achieving personalized content.

By considering the motivational scenario (see Figure 1), it means that ARIDUS users can obtain relevant content about sustainable projects directly from social networks. It also means that such content can be presented in a way that improves the user experience, since their preferences and constraints are taken into account.

Figure 5. FOAF, SIOC and @dapt alignment.

Semantic Personalized Services

In this work, semantic rules were used as a filter for getting recommended content automatically. Such functionality is played by two services previously described at ARIDUS services section. The first one, UserProfileManager deals with user profile. The second one, ContentRecommender (service 10) registers the users’ preferences and constraints.

ARIDUS ContentRecommender uses two different strategies.

One is based on the automatic record of the user’s accessed content. The other strategy is based on the users’ interests.

ARIDUS services use semantic similarity techniques to calculate the distance between the user’s interests and the content obtained, whose algorithm is explained by Table 1.

Table 1

Semantic Recommendation Algorithm

<table>
<thead>
<tr>
<th>Input: content, user organizing the algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output: recommended link</td>
</tr>
<tr>
<td>Procedure:</td>
</tr>
<tr>
<td>(1) Initialize variables</td>
</tr>
<tr>
<td>(2) Establish Connection with repositories</td>
</tr>
<tr>
<td>(3) Accept query parameters</td>
</tr>
<tr>
<td>(4) Access user profile</td>
</tr>
<tr>
<td>(5) Structure the SPARQL query</td>
</tr>
<tr>
<td>(6) Execute SPARQL query</td>
</tr>
<tr>
<td>(7) Get results</td>
</tr>
<tr>
<td>(8) While (results not = {})</td>
</tr>
<tr>
<td>(9) Calculate semantic similarity</td>
</tr>
<tr>
<td>(10) If match</td>
</tr>
<tr>
<td>(11) Show recommended link</td>
</tr>
<tr>
<td>(12) endIf</td>
</tr>
<tr>
<td>(13) endWhile</td>
</tr>
<tr>
<td>(14) End</td>
</tr>
</tbody>
</table>
Some parameters used by SPARQL queries can be dynamically obtained in a configuration file called \textit{extract.properties}, exemplified by Table 2.

\begin{table}[h]
\centering
\begin{tabular}{|l|}
\hline
PREFIX=dc:<http://purl.org/dc/elements/1.1/>  
SELECT=doc,title  
DISTINCT=true  
ORDERBY=?doc  
LIMIT=5  
\hline
\end{tabular}
\caption{Configuration File for SPARQL Query}
\end{table}

It is important to notice that ARIDUS services use internal and external resources. To illustrate what ARIDUS features represent, a situation is considered where a project leader tries to find out similar sustainable projects.

By using ARIDUS services, all searches that the leader makes are registered, in order to help in future recommendation tasks. The terms that are presented in titles, subjects, and keywords are used as input to calculate semantic similarity distances. In social networks the terms are extracted from posts.

The main benefit of this approach is to expand the search space, by improving the possibility to meet the users’ expectations of getting relevant information.

\textbf{Case Studies}

This section presents two case studies that show the dynamic features of ARIDUS services. The first one emphasizes three important features of ARIDUS: semantic integration, personalization, and accessibility. The second one focuses on social networks by considering posts as an important source of information. ARIDUS functionalities were conceived to integrate, in an open and virtual manner, data sources about sustainable projects and good practices.

\textbf{Semantic Integration, Personalization, and Accessibility}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure8.png}
\caption{Website adapted and linked content recommendation.}
\end{figure}
In this case study, many documents about sustainable practices were integrated by means of a metadata repository namely ARIDUSDB. Besides, an external database namely infoteca-e aggregates thousand of projects about sustainable techniques. It is the service that collects and provides access to information about technologies developed by EMBRAPA, the Brazilian Agricultural Research Corporation. It contains many collections comprised by booklets, books, radio, and television programs in accessible language.

The semantic integration of these resources is automatically activated. Words from titles and abstract of a given project are used as input to calculate semantic similarity distances. This strategy allows discovering similar and related projects.

Figure 8 illustrates an example of search mediated by ARIDUS platform. The first part of the image shows the original infoteca-e page that allows searching projects in the context of agriculture domain. As a result, it is possible to see information about the authors, date, and title.

In this case study, the original page was modified to add ARIDUS services in a transparent way. During the experiment the user looked for projects about corn plantation. A specific project is showed in the upper part of the screen. At this time the ARIDUS services capture and register the actions made, in background, by saving key terms about the subject searched. Such terms were then used to extend the results.

As a result not only projects about corn plantation were discovered, showed as links at the bottom of the page, but also ARIDUS recovered other related projects, such as pineapple (the second link), which can be produced in regions traditionally used for corn cultivation. This result was achieved as a consequence of semantic similarity techniques.

Concerning technical details it is important to state that the recommended content was initially obtained from infoteca-e that is available in the Dublin Core standard. To carry out this experiment many projects were imported and saved into the ARIDUSDB repository. Other documents could be added directly by researchers by using ARIDUS services, as an effort to improve the quantity and quality of projects provided.

Figure 8 also shows the results of another critical functionality of ARIDUS services related to the automatic adaptation for accessibility purpose. This functionality is carried out by ARIDUS ContentAdapter service, which makes changes according to international and Brazilian standards for accessibility.

Social Networks

The second case study involves social networks. It is aimed at describing how ARIDUS functionality can be extended in order to include social networks as a source of recommended content. In this sense, semantic integration of heterogeneous bases, which was subject of the first case study reported, has to consider explicitly other types of knowledge.

In fact, the knowledge that emerges from social networks is pretty much diverse in type and form. In this virtual environment many media formats can be gathered to express one single idea. This is precisely what motivated this case study. More specifically, it is the capacity of getting recommendation from posts, by extending the search space.

The goal was to discover relevant posts by using the relationships among people. To tackle this problem, ARIDUS contentRecommender uses SIOC and FOAF concepts as shown by Figure 4. Twitter was selected as a representative network and one of the most popular social networks. Semantic technologies allow establishing semantic links among Twitter posts in order to discover information about people’s interests and needs (Tao et al., 2012). Table 3 shows partial results of semantic inference over Twitter users’ profiles returning users who
share similar interests.

Table 3

<table>
<thead>
<tr>
<th>User’s Who Share Similar Interests</th>
</tr>
</thead>
<tbody>
<tr>
<td>NICK NAME</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>Jorgaf</td>
</tr>
<tr>
<td>Dasilva_uk</td>
</tr>
<tr>
<td>AIKA_informiert</td>
</tr>
</tbody>
</table>

To start receiving recommended posts, ARIDUS services must be running and integrated to Twitter. For sake of simplicity, they were manually integrated, but as future work it is planned to offer ARIDUS services as a plugin that can group the services.

**Related Work**

General aspects of data integration have been widely investigated during the last decades (Liao, et al., 2011; Haslhofer, 2010), and so it has been discussed the application of semantic technologies in this context (Bergamaschi, Castano, Vincini, 2009). Commercial tools have also been developed to provide bridges among data sources. They commonly use semantic web technologies, such as Asio Scout. ARIDUS project adopts semantic web services as a means to integrate disparate data sources. Preliminary case study developed in ARIDUS project, described at some particular difficulties were detected (Pereira, Dantas, Ribeiro, 2012), such as capturing and representing of expertise behind the agriculture project used, made explicit limitations on applying general frameworks, as reported by literature.

Semantic frameworks for agriculture domain have been developed, for instance IAIF (Izza, 2009), whose main objective is to allow the extraction of knowledge from various fields related to agriculture. Another project, reported at presents similarities to IAIF framework (Shoaib, Basharat, 2010), but it also includes transformation rules conceived by expert systems to improve knowledge representation by ontology. Personalization and accessibility are key aspects to ARIDUS services mechanisms. Despite these topics have been already addressed, real life context as discussed still deserve additional more investigation.

**Conclusion and Future Work**

This paper presented a service-oriented architecture named ARIDUS that addresses semantic integration of heterogeneous data sources about sustainable projects. This domain has an interdisciplinary perspective that combines environmental, social, and economic information. The complexity involved has demanded specific functionalities to support the decision making process in this context.

By using a motivational scenario as a starting point, the key features of the proposed platform were presented, such as the use of semantic web services. This technology was used to integrate heterogeneous data sources. This scenario was motivated by a real life necessity to integrate documents about sustainable solutions suitable to the Brazilian semiarid. In fact, the main goal of ARIDUS project is to build a virtual knowledge base to help those who need to define policies for the sustainable development of this rich but fragile biome.

The use of semantic technologies allowed representing knowledge into ontology to support many ARIDUS services, such as modeling user’s profile. It also allows integration among social networks avoiding

7 http://www.bbn.com/technology/knowledge/asio_scout
duplication of information. Modeling aspects have been considered the most challenging demands at ARIDUS project.

The management of predefined priorities is another key feature for supporting decision making in this context. Such functionality must be improved by automatic mechanism boosted by reasoning inference, such as those used by alignment services.

References


