Journal of Universal Computer Science, vol. 16, no. 8 (2010), 1117-1137
submitted: 23/12/09, accepted: 25/2/10, appeared: 28/4/10

Locating and Crawling eGovernment Services A Light-weight Semantic Approach

Luis Álvarez Sabucedo
(Universidade de Vigo, Spain
lsabucedo@det.uvigo.es)

Luis Anido Rifón
(Universidade de Vigo, Spain
lanido@det.uvigo.es)

Abstract: The application of Web 2.0 tools and methodologies in the domain of eGovernment is not yet a fully exploited area due to the immaturity of the software support, and the lack of commitment from Public Administrations. This paper proposes a solution to locate a service which a citizen may be interested in. The solution uses particular features from this environment, such as microformats, metadata and dynamic procedures on the Web. The paper describes a semantic model for the domain, and tools to annotate, publish and crawl services in Public Administrations are discussed in depth. The paper details the entire software platform, and it presents conclusions and a number of proposals for future research efforts.
Key Words: metadata, knowledge management, eGovernment
Category: H.4, M.1, H.3.5

1 Introduction

The use of Web 2.0 technologies in the domain of the eGovernment is not a new feature in the current state of the art. Unfortunately, its application is not as mature as is desirable in this field. Currently, eGovernment projects developing the application of these features make a superficial use of technological resources available within this area. This is due to several factors: lack of maturity in some technologies, the high cost of developing solutions under new paradigms, lack of commitment from Public Administrations (hereafter PAs), etc. Even though some efforts have been launched, it has not yet been fully determined how to take full advantage of these technologies in the domain of eGovernment. Significant applications of this kind of technology relate to increasing content accessibility, fostering the collaboration of citizens in services provided by PA, providing simpler user interfaces, and engaging citizens in related processes. Most successful use cases are linked to politicians who make use of social networks and tools to broadcast videos and spread their ideas. Also, PAs have launched projects based on the use of new interfaces and metadata support to deliver search services. However, these efforts are far from achieving the full potential of this technology.

This paper focuses on solving a particular problem in the domain of eGovernment: locating the service desired by the citizen using a strategy based on annotating and crawling PA contents. Initially this could seem to be a trivial task but, on the contrary, it may even be difficult to locate the correct PA for a certain service. Imagine the case of a citizen looking for support to request a grant from his/her administration to enrol in a university. The user must find out the PAs which are involved among all the possible ones (National Government, Regional Government, Local Government, University, etc.), the documentation required, the correct way to upload the form, etc. It seems to the authors that in the current state such a scenario it is not simple for citizen to undertake all the required operations on their own.

Therefore, this paper is aimed to provide a support for searching for eGovernment services. The objective of the current work is to provide an architecture which enables the citizen to locate a desired service easily and guide his/her access to the service, with additional information in order to correctly fulfil the goal required. To achieve the provision of such a high-level service in a simple and effective manner, some resources from light-weight technologies are used. The first section of the paper provides an overview of the current state of the art in the field (see Section 2).

Upon reviewing the present state of technology in the domain, the authors of the current work propose a model which contributes to the technologies in the field entitled AdministrativeService. This model is reviewed in detail in Section 3. To make this concept functional, the support of currently available semantic technologies (see Section 4) is required to construct the system (see Section 5)

The final software architecture proposed by the authors is based on the collaboration of different agents that collaborate to annotate resources, crawl the web and recover semantically annotated information, as shown in Fig. 1. Therefore, a tool is included to generate semantic annotations on web pages and also to store those on a remote DBMS (see Section 6). Later on, the semantically annotated contents are indexed by means of a crawler (see Section 7). The last step is the provision of a web tool to make the contents already gathered and stored within the system available to the citizen (see Section 8). The system is tested and discussed in Section 9. Finally, in section 10, some final conclusions regarding the initiative are presented.

2 Related works

As already mentioned, eGovernment is currently a domain of current interest with an ever-increasing number of research efforts. As a result, a large number of projects and initiatives have arisen in this domain. A review of such efforts must be presented to demonstrate the current state of the art. The efforts in the domain can be categorized into three groups:

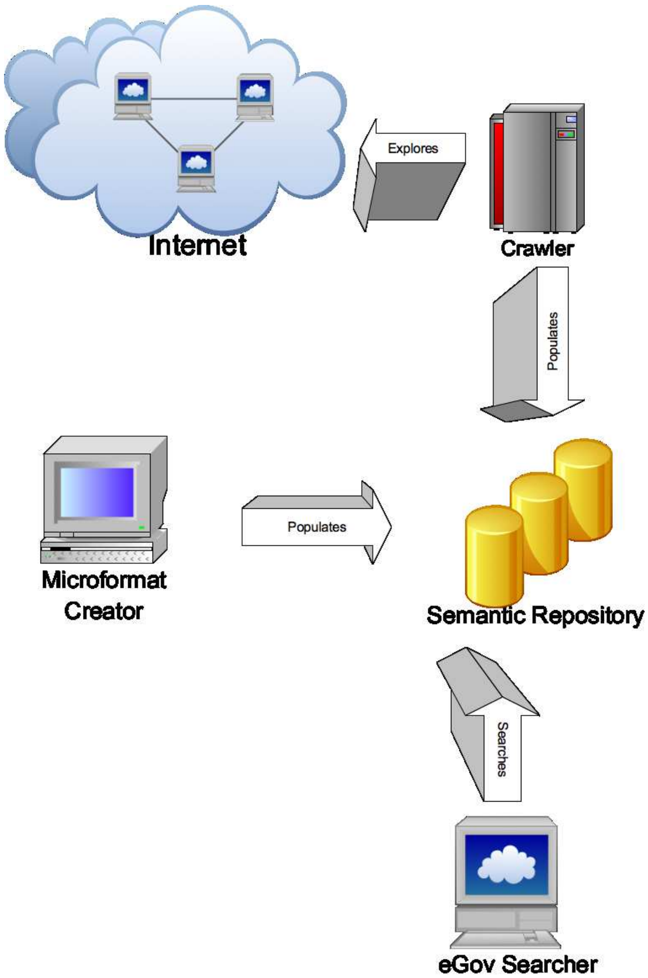

Figure 1: Sketch of the system

– Initiatives from governmental bodies. National governments and international organisms such as the European Union have had an interest in eTechnology since its infancy for interoperability and accessibility issues. Therefore, they have supported the provision of recommendations and interoperability solutions in the domain. As shown later on, the support available for current solutions is limited.

– Standardization bodies and consortia. International standardization bodies have gained an interest in this domain during the last number of years. Most of their work is related to the adoption of their previous specifications/standards to cover the needs of the eGovernment domain.

– Projects developed by research institutions. Resources assigned to this area by many institutions, for instance the European Commission, have resulted in a large amount of projects. As many of them are related to the objectives of this paper, they are reviewed in depth, providing details about how they could benefit from the proposal of the present paper and vice-versa.

2.1 Governmental initiatives

The vast majority of countries have developed their own frameworks to host eGovernment solutions[Guijarro 2007]. Most of them are oriented towards the provision of a framework for interoperability at some level, or supporting tasks related to providing eGovernment solutions. Unfortunately, none of them provide a solid or efficient platform to actually carry out services. The frameworks related to the current proposal are briefly outlined below:

– SAGA (Standards und Architekturen in eGovernment Anwendungen)[KBSt 2005] in Germany is a guideline to follow for the development of solutions in German administrations. This framework does not consider semantic based solutions. It mainly provides us with a set of standards that must be used to deploy solutions in the domain.

– e-GIF (eGovernment Interoperability Framework)[UK GovTalk 2004] in the United Kingdom covers issues related to specifications and policies for any cross-agency collaboration, eService access and content management. It also includes definitions of metadata to mark documents by using RDF[World Wide Web Consortium 2005]. This is one of the official projects making larger use of semantic based solutions. Unfortunately, it only studies light-weight semantic technologies such as RDF or taxonomies. Besides, it does not deal with additional relevant issues within the scope of this project such as Semantic Web Services.

– CCI (Le Cadre Commun d'Intéroperabilité) from ADEA (l'Agence pour le dèveloppement de l'administration électronique)[French Government 2004] in France establishes recommendations for the development of systems related to public services and electronic delivery. Details concerning semantic definitions or services are scarcely considered.

– EIF (European Interoperability Framework)[Enterprise and Unit 2005] for European solutions. establishes a framework for discussion regarding interoperability, but it does not actually endorse any particular catalogue or standard to build eGovernment solutions. It guides European countries in the search for an interoperable framework for pan-European solutions.

- FEAF (Federal Enterprise Architecture Framework)[POPKIN Software 2004] in USA focuses on the description of enterprise models to develop cross-border solutions. It neither considers particular architectures or technologies, nor does it actually deploy systems. However, it is quite useful as a systematic guide for the design of solutions which take into account the integration of all stakeholders.

Most of these projects just describe technological support or methodologies to define operations. As mentioned above, unfortunately a complete set of instructions to develop services or platforms for final solutions is not provided. Further information about this particular topic can be found in [Guijarro 2008].

2.2 Standardization bodies

Several international organisms involved in standardization issues have also devoted efforts to the domain of eGovernment. Some of the most relevant ones involved in horizontal technological development have also launched their own particular groups of interest in the area. The most significant ones are:

- DGRC. The Digital Government Research Center was founded in 1999 by the National Science Foundation[NSF 2005]. Its area of interest is research in ICT applied to eGovernment services. It is involved in the development of several projects, and the information intended for the citizen is provided by means of the newsletter dgOnline[Information Science Institue 2005].
- OMG. The Object Management Group[OMG 2005b], besides its principal projects and initiatives, launched a specific working group for eGovernment, the Government Domain Task Force (GovDTF)[OMG 2005a]. At the time of writing of the current paper, the working group was in their very first steps. One of their biggest points of interest is related to the application of MDA and other OMG specifications to the domain.
- OASIS. The Organisation for the Advancement of Structured Information Standards[OASIS 2005] also has a committee[OASIS 2005] devoted to the study of the applicability of their own technologies to eGovernment. Its work is mainly focused on the articulation and coordination of requirements for XML- and Web services-based standards. This Committee includes several subcommittees (SC): eGov Asia-Pacific SC, eGov Best Practices SC, eGov Core Components SC, eGov ebXML Registry SC, eGov Harmonising Taxonomies, eGov Infrastructure SC, eGov Services SC, and eGov Web Services SC. However, during the course of the preparation of the current paper the research outcomes of this group are rather initial.

- CEN. The European Committee for Standardization[CEN 2005b] launched its own group of interest in the area[CEN 2005a] in February 2005. Additionally, it has undertaken some interesting work by means of some CWAs (CEN Workshop Agreements): CWA 1859 *"Guidance on the Use of Metadata in E-Government"*[CEN 2008c], CWA 1860 *"Dublin Core E-Government application Profiles"*[CEN 2008a] and CWA 13988 *"Guidance information for the use of Dublin Core in Europe"*[CEN 2008b]

- W3C. The eGovernment Interest Group (eGov IG)[W3C 2008] from the W3C is concerned with the exploration of methods to improve access to eGovernment through better use of the Web using open Web standards at any government level. This group has recently delivered several documents of interest for the domain.

2.3 Related projects

A number of projects in the area have aims quite similar to the current research. This section analyzes current efforts and provides a detailed insight on how they can benefit from this work and vice versa. Some of these projects are outlined below:

- The SemanticGov project[N. Loutas and Tarabanis 2008]. This project supported by the 6th Framework Program aims at developing a software infrastructure intended to provide support for PAs. Semantic technologies are expected to play a principal role. The project is based on the use of WSMO to provide interoperability mechanisms. The definition of the preconditions and the post-conditions is quite challenging within this project. It provides a two-level ontology that models high level services in the domain in great detail.

- The Access-eGov project[Stojanovic 2008]. This is an initiative, also in the 6th Framework Program, based on a peer-to-peer and service-oriented architecture that also takes advantage of semantics to improve accessibility and connectivity. It is focused on the reorientation of services to provide support and access to services already provided rather than on the provision of strategies for automatic composition services or semantic-based searching services.

2.4 Discussion

Having reviewed the significant initiatives undertaken by governmental bodies, it is apparent that the intensive work carried out is mainly concerned with providing the basis for future solutions. The set of recommendations and standards proposed mainly deal with interoperability at application level. Only some of them such the eGIF project, provide some metadata to annotate resources. Unfortunately, the provision of a complete model of the domain, by means of ontological models or by any other means, is not accomplished. The main reason for this is due to the lack of maturity of the domain and the great complexity of the tasks involved.

Also, the work performed by international consortia and standardization bodies is quite initial. As long as they are in their initial phases and its scope is mainly limited to the adaptation of their former standards to the domain, their contribution is rather partial. Hopefully, as the domain gets more mature, their work will drive the solutions in the domain as already happens in other related environments such as eLearning or eBusiness.

In relation to the contribution presented, it should be mentioned that it is not simple to take advantage of the work already developed. This situation is due a generalized problem in the domain: the lack of an agreed conceptualization of the domain. As shown in further sections, the definition of the business model to be tackled is not a trivial task and it is driven by a large number of issues that are not common to all projects. Consequently, only minor parts of already developed ontologies and models can be applied in a straightforward manner.

3 Characterization of the domain

The very first step in the provision of the solution is the identification of the environment under consideration. Many different PAs are providing services under different models and based on different concepts. Therefore, it is required to identify some sort of pattern for the description of those services that are currently being provided by PAs.

Presently, it is common to find services modelled under the label identified as LifeEvents in different approaches from both public and academic sources such as [Wimmer et al. 2001, N. Loutas and Tarabanis 2008, Tasmania Government 2009, Slovenia Government 2009, Citizens Information Board 2009]. However, this concept is not always used to refer to precisely the same thing. Also, Most of the time it is used at a very high level of abstraction; thus, it can be found as LifeEvent services such as getting married, losing a wallet, dying, etc. Upon evaluation of the works mentioned, some general lines regarding the so-called LifeEvents can be identified:

- LifeEvents are considered to involve a large number of PAs.
- A number of realizations of the same LifeEvent are possible which depend on further pieces of information.

- Upon its selection, a webform or a set of them are provided to develop the service, in other words, usually no WebService for this implementation is available.

- In most cases, little or none semantic annotations are provided. Consequently, no advanced services regarding discovery or composition are possible.

In the scope of this work, a lower level definition of service is used, that is, the aim is to provide support to manage services from a single PA. At this point the proposal presented in [Álvarez Sabucedo et al. 2008] is brought into focus. This work introduces its own definition of LifeEvent, in a similar manner to solutions already deployed, but it also proposes the use of the element AdministrativeService, a concept used to model services in the context of PAs.

The first element, LifeEvent (hereafter LE), is used to refer to high level services that may involve different PAs, but the latter refers to those services provided by a single PA in a straightforward manner. Actually, AdministrativeServices are those services that take place in just a single office and generate, as output, some documents for the citizen. These last ones can be considered as a second level to deploy LEs, as they act as an interface for services between citizens and LE providers.

These latter ones, the AdministrativeServices (ASs hereafter) are the focus of this work. As the goal intended is to provide support for locating services from PAs in a simple and straightforward manner, they actually support the fulfilment of the aim of the present work. Thus ASs, as they were defined in [Álvarez Sabucedo et al. 2008], are re-used in this work. To describe an AS, it is required to detail the following aspects:

- Title. Brief name for the AS.

- Description. A brief textual description about the service for the citizens.

- Max Life Span. The maximum span of time for the response from the PA before the operation is considered approved/dismissed.

- Public Administration. Information about the PA that it is responsible for the execution of the AS. Therefore, it can be used to decide about the scope of the operation.

- Input documents. The documents the citizen needs to be in possession of to be able to invoke the AS.

- Output documents. These are the documents that will be generated as output in the case that the AS is completed as expected.

- Area. The group of services in which this AS fits.
- Location. The URL where the service is hosted.

At this point, the key importance of documents must be brought into focus. These are the legal contracts linking the PA and the citizen in the fulfilment of an operation or in declaring a status or circumstance. In this framework, they are considered as the input for the invocation of an AS and also its output, that is, the legal proof for a new status or the requirement for a further operation, that is, AS.

4 Semantic Web

The evolution of Web-based platforms shows a clear path from its beginnings to the present day. Firstly, HTML was introduced as a mark-up language to design content and information. Later on, XML was introduced as a technology to represent just data and the use of CSS was applied to describe its representation. The next step in this evolution is the transformation from data (XML-based contents) to information, that is, semantic contents.

"Semantics", as an IT research field, was born in the early 2000's. In May 2001, Sir Tim Berners-Lee published the foundational article presenting the semantic web to the world [Berners-Lee et al. May-2001]. According to this article, "the Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users."

The gist of this idea is to make machines capable of understanding the information within the web. This feature will allow them to perform more complex interactions without the need for human support. According to the previous article, the semantic web is: "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Therefore, semantics can be considered as the next step in the evolution mentioned above. In the current state of art, it is not enough to provide only data. Current software agents would benefit from any additional features available only in the case that contents are actually "understandable" for them. This is the point where semantics enters into the scenario. Semantics is compelled to provide information, not just data. In other words, semantics introduces meaning into the data in order to allow computers to deal with this information in a more interoperable manner.

The tools that drive this technology are ontologies. This word taken from philosophy and linguistics is used in the domain of ICTs to speak about "a formal, explicit specification of a shared conceptualization of a domain of interest"

[Gruber 1993]. Thus, using this tool, a support is provided to present abstract information about a certain domain in a concrete way, by means of a machine-understandable data format.

To express an ontology in a formal manner, several languages were proposed in the last years [Gómez-Pérez et al. 2003] and are currently available to researchers. However, OWL (Ontology Web Language) [W3C 2004], the W3C Official Recommendation, can be considered as the preferred support in the semantic research community. It is intended to provide a fully functional way to express ontologies. To make different levels of complexity possible, OWL provides different sublanguages with increasing expressivity: OWL Lite, OWL DL and OWL Full. By using OWL, we are addressing a W3C recommendation that can be considered as solid and interoperable support for the provision of this solution.

Unfortunately, all the power within this tool, OWL, may be not needed and, on the contrary, it may introduce too much complexity to develop simple-to-use solutions. This must be avoided, particularly in those cases where the semantic features of the system are quite close to the human user. To deal with these cases, there are other options available such as the use of microformats[Khare 2008]. According to its creators [Khare and ¸Celik 2006]: "Designed for humans first and machines second, microformats are a set of simple, open data formats built upon existing and widely adopted standards."

So, microformats can be considered as a light-weight semantic tool to represent information. The price to pay for this simplicity to use and possibilities for quick adoption relates to the loss of power to express complex rules, relations among concepts and inference possibilities. Despite all those problems, microformats could be considered as a successful technology in the scope of the Semantic Web, as it is one of the most used technologies. This is due, partially at least, to the possibility for microformats to be expressed in terms of XHTML [W3C 2009c] code so they can be used immediately on a web site directly in sight of the user. As shown in the present paper, this technology has been proven to be the balanced choice to address the current problem.

5 Describing knowledge

In order to describe the knowledge in the system, the use of an ontological support is the most convenient tool. In the current case, former works in the area, in particular, the ontological support presented in [Álvarez Sabucedo et al. 2008] will be taken as a starting point. This contribution proposes OWL support to describe the domain of eGovernment (an excerpt is shown on Figure 2).

This semantic information involves a large effort in the modelling of the business model implicit in the problem and its aim is to take into account all relevant features in the system. In particular:

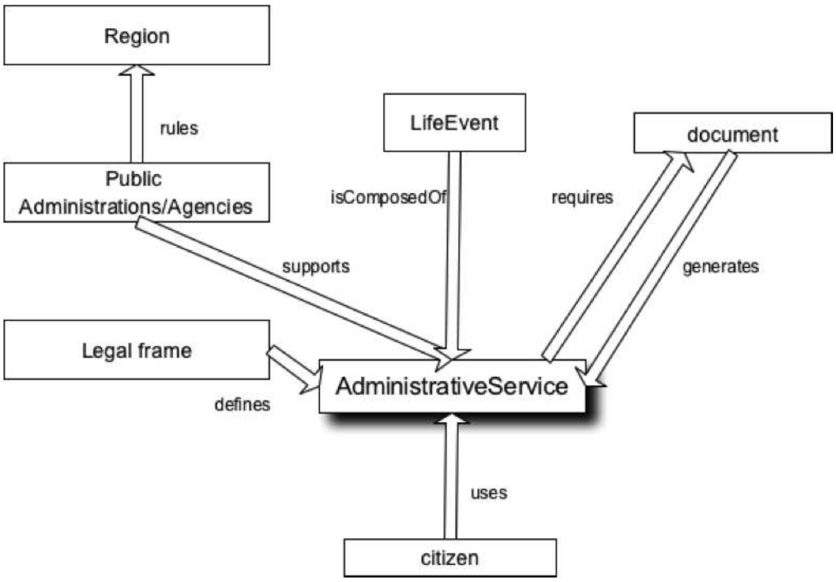

Figure 2: Simplified model of the ontology to define the business model

- Modelling the artefacts LE and AS.
- Support for annotating the legal framework for ASs.
- Management of legal documents from citizens.
- Relations of locate which PA is in charge for each AS.
- Connections between required and generated documents in each AS.

However, in the present approach this level of semantic information could exceed actual requirements of the system. Besides, it could be difficult for normal users or even civil servants to generate the full semantic descriptions. So, in the context of this solution it was decided to take advantage of a microformat specification [Khare and ¸Celik 2006]. To keep a balance between the complexity of the system and its simplicity of use, some restrictions were put on the semantic model to make it fit the proposal of microformats. Therefore, the solution is just concerned with modelling ASs, and for its definition, a number of relevant fields were identified taking the ontological model in Fig. 2 as a basis:

- name: name of the service itself.
- description: brief description of the service.

- PA: public entity in charge for the service.
- Input Document: name of the documents required to invoke the service.
- Output Document: name of the documents generated as output for the service.
- MaxSpan: maximum delay in the execution of the service, expressed in days.
- URL: address of the web page where the service is accessible.
- Areas: Areas of interest for the AS.

The reader should note that Input/Output Documents are very important in this approach as they are the pre and post conditions for the execution of the service. In order to invoke the service, it is required to be in possession of all necessary documents and, conversely, as proof of the actual fulfilment of the operation, the service provider will generate all required documents.

This way, an owl instance is transformed into a simpler RDF representation:

Of course, generic Cascading Style Sheets (CSS) [W3C 2009a] are also provided. This CSS ensures that the contents are properly displayed on a web page in the case that the corresponding web master decides to use it.

Under these circumstances, information can be translated from an OWL individual to microformat code and the other way round. This would be quite useful for the current purposes in order to store and recover the information in the most convenient manner for each case.

6 Annotating web pages

It must be kept in mind that the effectiveness of this approach depends on the amount of users which are utilizing the system. Asking PAs to update their web contents may result in a low rate of successful implantation of the solution. To tackle this issue, a simple to use software tool is provided, whose aim is to perform two different tasks: creating the microformat content and storing this information in a remote database.

The software tool will provide a simple user interface (see Figure 3). The user can fill in a form about a certain service with all the data required for its full characterization. Upon the completion of this information, two different options are available. The first option is to generate the microformat code itself (see button "Generate Code" in Fig. 3). In this case, if the user is the civil servant in charge for the actual web page, the code generated can be introduced in the web page using a XHTML representation of this semantic information. If so, the information created will remain available to any user that may access the web page in the future.

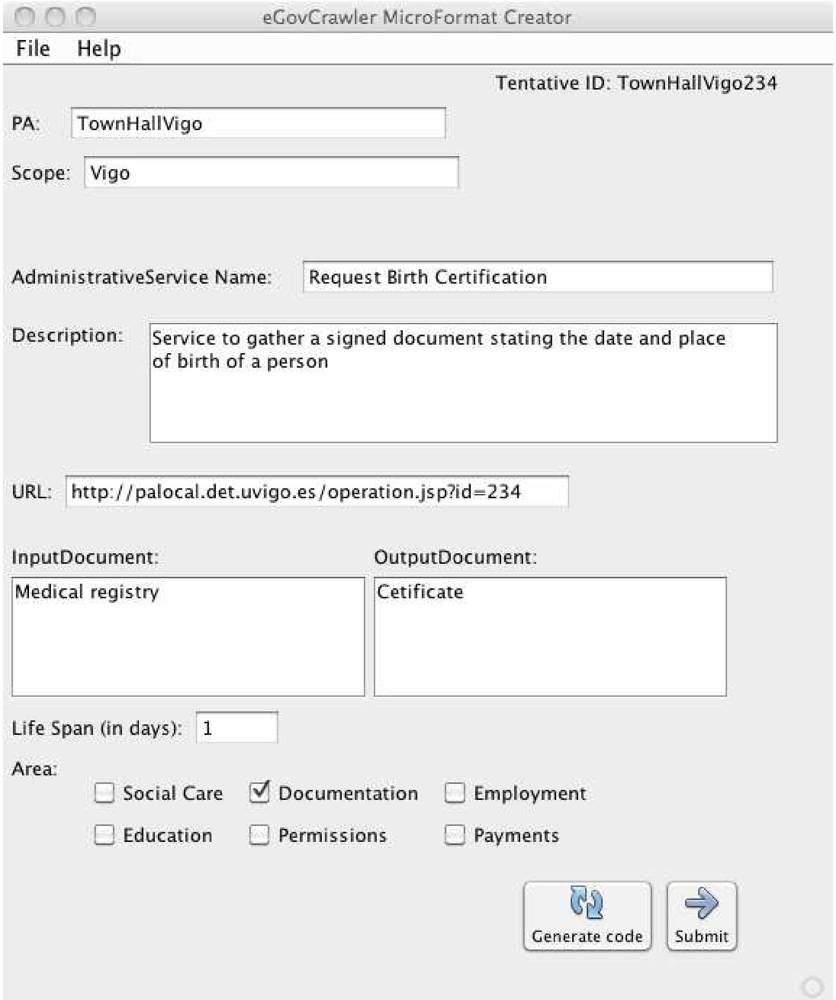

Figure 3: Annotating services

Given that usually the person generating these contents may not be in charge of updating the web, this platform offers an alternative approach to take full advantage of mining tools and semantic data recovery. Once the information is introduced, the user can submit the information (see button "Submit" on Fig. 3). This button will invoke a Web Service that will introduce the information on a remote server where all these pieces of data are stored. Thus, this information will not be lost and, as shown in further sections, users can still benefit from this effort.

7 Crawling the web

The system proposed also has the ability to explore the World Wide Web in search for new services. Those services, once located, are analyzed and all available information is inferred and stored in the system. Actually, as previously mentioned, the information within the system is managed by a database populated with semantic annotations. Besides the method based on annotations made directly by users (see Section 6), the system will acquire new contents by means of its own exploration of the web for contents annotated by the microformats introduced.

A crawler performs this task. Developed in Java, this software agent is implemented using a number of software libraries to download HTTP content, scan HTML contents, and manage information using libraries such as Jaxen[Codehaus 2009] and Jena[Hewlett-Packard 2005].

To perform this task, the system is fed with an initial set of known URLs corresponding to certain web pages with contents relevant for the application.

Once the target web is located, and downloaded, the crawler performs two different operations:

- Explores the web in search for the information introduced in the web page itself using the microformat annotations.
- Scans the HTML code of the page in search for further links that may contain new pages to supply the process with data again.

The process to locate new suitable web pages to increase the pool of pages known by the system is quite simple. The system just scans the HTML code downloaded in search for tags a and $frame$. Those HTML entities include the attribute *src* that points to web pages of potential interest, in the same web site or in external ones. All of them are used for future iterations of the crawler.

However, the most relevant operation for the current purposes is the second one: obtaining the information from the web pages under analysis. This operation is undertaken on the HTML content already downloaded for each target. The crawler must look for the microformat information included there. This operation is performed using Xpath[W3C 2009d] expressions designed to extract the information compliant with the solution presented and Xerces[Apache Software Foundation 2009]. Therefore, it was only possible to perform evaluations on ad-hoc contents. However, the results obtained by the prototype turned out to be successful in extracting the required information. As occurs with the tool mentioned in Section 6, as the information is *learned* by the system, a Web Service[W3C 2009b] is invoked to store it in the server. This Web Service will accept the fields identified in Section 5 as input parameters. They will be used to create a new instance of the ontology with that information and inserted into the DBMS. This is done by means of the Jena library in a quite straightforward manner.

8 Gathering services

The final goal of the present work involves the provision of a simple-to-use mechanism to store and recollect information on the web regarding services provided to citizens by PAs. In order to make that possible, a semantic support (see Section 5) was introduced which was used to annotate web pages (see Section 6) that would be indexed later on by a crawler (see Section 7). The next and ultimate step is gathering that information from a common location where it was stored by all the software agents involved in this solution. Therefore, citizens, the final users of the proposed solution, can take full advantage of the system.

In this approach, the searches are conducted by means of a web interface (see Figure 4) where the citizen introduces information about the desired AS, a normal service from his/her perspective. With this information, the semantic engine will present all the data matching the expressed conditions as a response.

This web interface supports also some features regarding the personalization of the profile of the citizen. Thus it is possible to take into account information such as the region of the operations, the groups of interest for services, documents he/she may be in possession of, etc.

Once the web form is filled in, the system will generate the SPARQL[W3C 2006] query containing both the information from the web form and the information from the profile of the citizen. The queries generated are similar to the following example:

```
SELECT ?ASid WHERE
 {
   ?ASid rdf:type as:AS .
   ?ASid as:isSupportedBy :PA_X .
   ?ASid as:requires :Doc_X .
   ?ASid as:generates :Doc_Y
 }
```

So this query will return those ASs that are assigned to the PA PA_X, that require Doc_X to be invoked and that generate Doc_Y as output. Queries introduced on the system are executed on the server that manages the DBMS using the support provided by Jena. The result of the execution of this query is presented to the user and direct navigation to the service provided along with a detailed explanation of the AS.

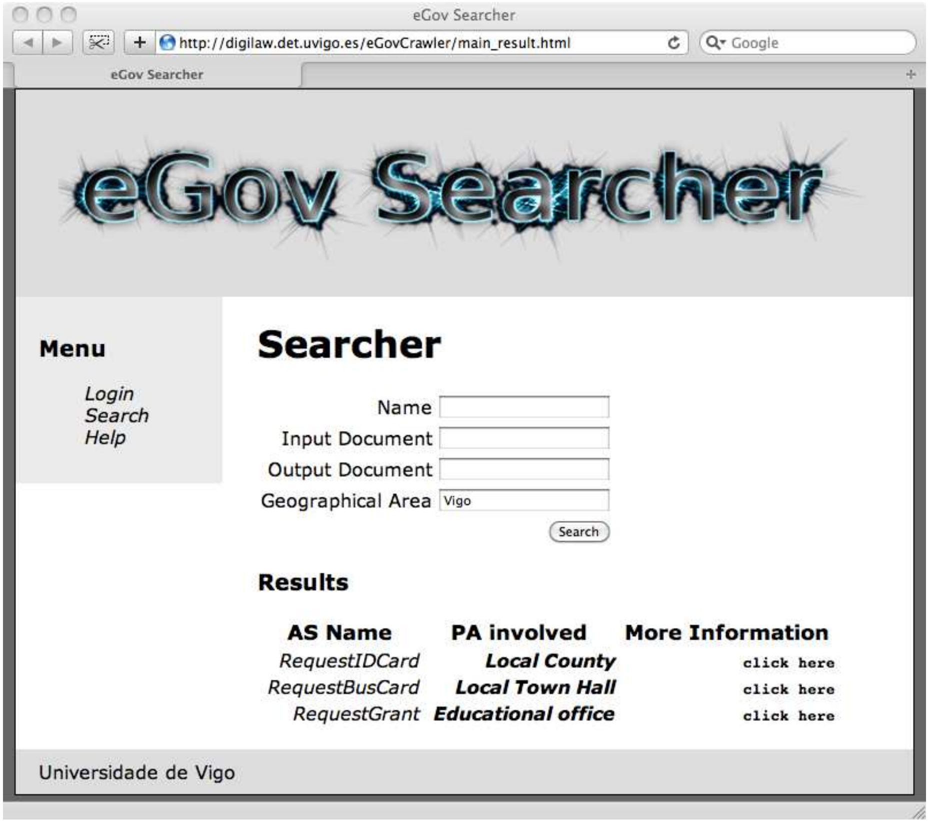

Figure 4: Web Interface for searching ASs

9 Discussion

Once the system was fully developed, a testing phase was required. Thus, authors decided to apply the developed system to a set of services offered on current web pages from different Public Administrations. A group of 20 people were selected to make the annotations to a set of 80 different services. Upon the finalization of the tests, some conclusions were clear for the working team:

- Eventhough the system was designed to be simple to use, some training was required to make it possible for unskilled users to make use of the platform.

- The semantic support provided within the proposal to express the required level of knowledge was sufficient for the purposes of the solution. Using the set of fields identified, users feel comfortable locating and describing services.

– A particular language had to be set. As the different PAs selected for testing were located in different geographical areas with different languages, the need to set a certain language for tagging and for searching soon became evident.

– The requirement for a moderator to supervise the suitability of the tags introduced can be overlooked. Although some users used the system incorrectly, as most users used it properly, the contribution of the firsts was not relevant.

– As the number of new instances pile up, it can be experienced a large delay in operations can be evidenced such as memory loading the entire ontology. Tests with more than 200 instances loading in memory using the Jena library on an average server[1] takes too much time for an interactive use.

– Users provided abundant feedback about the tool developed. As is common in software projects, and even more on proof-of-concepts, these remarks can or should be taken into account for subsequent versions of the prototype in order to provide additional features regarding usability and functionality.

The experimental results, even in a such limited environment, turned out be quite promising. However, it is still needed to consider the real scenario. In order to tackle this goal, support for average users is required and it is necessary to guarantee a minimum amount of data in the system to achieve the critical mass required for the final success of the system.

The testing of the crawler was conducted in a fully automatic manner and, due to the constrains imposed on the system (a particular set of microformats on a medium size pool of possible web to explore), the results were quite satisfactory in terms of amount of data gathered.

10 Conclusion

The real ethos of eGovernment is to provide services centered on the citizen that improve the quality of services using the support of ICTs. The proposal presented in this paper is focused on the task of locating the service the citizen is actually looking for. Even though this may seem a trivial task at first glance, actually it is the opposite. Citizens may get overwhelmed by the large amount of different PAs involved in a particular area or by the broad number of options available to perform the same service.

By means of ICTs, this platform offers a simple-to-use service that makes it possible for a citizen to locate the service more suitable for his/her circumstances. To undertake this task, a semantic based approach was chosen. During

[1] Tests were conducted on a Intel Core Quad Q9450 at 2.66Ghz and 4 GB of RAM running Ubuntu and MySQL as the database support.

the development of the final solution, different tools were provided with distinct goals, that when combined constitute a global solution. Firstly, a tool to annotate the contents was developed (see Section 6). These annotations are gathered by a different tool (see Section 7). Finally, all these pieces of information are presented to the user by means of a Web-based tool (see Section 8). These tools were designed to work all together but in fact they can be used under different approaches to fit into other projects already developed.

During the development of this solution, semantics proved to be the cornerstone that maintains the coherence of the whole system, as it contributes the knowledge that makes the information within the platform understandable. Even though the use of heavy-weight semantic technologies was considered as a suitable option in this context, a light-weight semantic approach based on an already developed OWL ontological support turned out to be the most accurate option, as demonstrated by the final results.

The model of the domain proposed is based on the use of the artefact identified as AdministrativeService. This concept is used to model those services provided by Public Administrations directly to citizens, it is based on previous works [Álvarez Sabucedo et al. 2008] and it is quite similar to other already existing solutions from different PAs [Wimmer et al. 2001, N. Loutas and Tarabanis 2008, Tasmania Government 2009, Slovenia Government 2009, Citizens Information Board 2009].

With this approach, services available for citizens are placed in a common pool. This pool provides highly useful information and can be considered as one of the outcomes of the proposal. It must be kept in mind that in this digital repository, services from different PAs are stored using the same language, due to the semantic definition of the business model. This provides new opportunities to Public Administrations to deliver advanced interoperability tools for services which were previously unrelated. Also, the repository of available services is supposed to be updated and filled with proper contents due the philosophy of the crawler that keeps on scanning the Web in search for new content all the time. Anyhow, as the architecture was developed as a proof-of-concept, further deployments must bear in mind some additional considerations regarding legal issues in each country.

This approach is based on the collaboration of users and service providers, public administrations or agencies, willing to take part in the system. This is currently a feature of utmost importance for the proposal to succeed. The commitment of all parts implied in this environment is required to come up with an actual solution that may overcome the current state of the art. Therefore, to make that possible, an open definition of AdministrativeServices and the tools for its application were released, creating the microformat-based annotation and storing the contents as they are defined in the context of this work. Achieving the critical mass for this project will be the key to its final success.

Acknowledgment

This work has been funded by the Ministerio de Educación y Ciencia through the project "Servicios adaptativos para e-learning basados en estándares" (TIN2007-68125-C02-02) and by the Xunta de Galicia, Consellería de Innovación e Industria "SEGREL: Semántica para un eGov Reutilizable en Entornos Locais" (08SIN006322PR).

References

[Álvarez Sabucedo et al. 2008] Álvarez Sabucedo, L., Anido Rifón, L., Míguez Pérez, R., Santos Gago, J.: "Providing standard-oriented data models and interfaces to egovernment services. a semantic-driven approach"; Computer Standards & Interfaces; doi:10.1016/j.csi.2008.09.042 (2008), 14.

[Apache Software Foundation 2009] Apache Software Foundation: "The Apache Xerces Project, Web site http://xerces.apache.org/"; (2009); web available at http://xerces.apache.org/.

[Berners-Lee et al. May-2001] Berners-Lee, T., Hendler, J., Lassila, O.: "A new form of web content that is meaningful to computers will unleash a revolution of new possibilities"; Scientific American; (May-2001).

[CEN 2005a] CEN: "CEN Group on e-Government"; Web available (2005a); http://www.cenorm.be/cenorm/businessdomains/businessdomains/isss/activity/e-government.asp.

[CEN 2005b] CEN: "CEN home page"; Web available (2005b); http://www.cenorm.be/cenorm/index.htm.

[CEN 2008a] CEN: "Dublin Core eGovernment Application Profiles"; Web available (2008a); ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14860-00-2003-Nov.pdf.

[CEN 2008b] CEN: "Guidance information for the use of dublin core in europe"; Web available (2008b); ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa13988-00-2003-Apr.pdf.

[CEN 2008c] CEN: "Guidance on the use of metadata in egovernment"; Web available (2008c); ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14859-00-2003-Nov.pdf.

[Citizens Information Board 2009] Citizens Information Board: "Citizen Information - Irish Government"; Web available (2009); http://www.citizensinformation.ie/index/LIFE.

[Codehaus 2009] Codehaus: "jaxen: universal XPath engine"; Web available (2009); http://jaxen.codehaus.org/.

[Enterprise and Unit 2005] Enterprise, D., Unit, I. I.: "European interoperability framework for pan-european egovernment services"; Web available (2005); http://europa.eu.int/idabc/en/document/2319/5644.

[French Government 2004] French Government: "ADEA"; Web available (2004); http://www.adae.gouv.fr/adele/.

[Gruber 1993] Gruber, T.: "A translation approach to portable ontology specifications"; Knowledge Acquisition; (1993), 199–220.

[Guijarro 2007] Guijarro, L.: "Interoperability frameworks and enterprise architectures in e-government initiatives in europe and the united states"; Government Information Quarterly; (2007), Elsevier.

[Guijarro 2008] Guijarro, L.: "Semantic interoperability in egovernment initiatives"; Comput. Stand. Interfaces, doi:10.1016/j.csi.2007.11.011; (2008), Elsevier.

[Gómez-Pérez et al. 2003] Gómez-Pérez, A., Fernández-López, M., Corcho, O.: Ontological Engineering; Springer, 2003.

[Hewlett-Packard 2005] Hewlett-Packard: "Jena"; Web available (2005); http://www.hpl.hp.com/semweb/.

[Information Science Institue 2005] Information Science Institue: "dgOnline"; Web available (2005); http://www.dgrc.org/outreach/dgonline_latest.jsp.

[KBSt 2005] KBSt: "Saga"; Web available (2005); http://www.kbst.bund.de/-,182/SAGA.htm.

[Khare 2008] Khare, R.: "The next (small) thing on the semantic web?"; IEEE Internet Computing; (2008), 8.

[Khare and ¸Celik 2006] Khare, R., ¸Celik, T.: "Microformats: a pragmatic path to the semantic web"; WWW 2006; 1-59593-323-9/06/0005; 865–866; W3C; ACM, 2006.

[N. Loutas and Tarabanis 2008] N. Loutas, V. P., Tarabanis, K.: "Providing public services to citizens at the national and pan-european level using semantic web technologies: The semanticgov project"; Procededings of the Sixth Eastern European eGov Days, Czech Republic; (2008).

[NSF 2005] NSF: "National science foundation"; Web available (2005); http://www.nsf.gov/about/.

[OASIS 2005] OASIS: "Oasis"; Web available (2005); http://www.oasis-open.org/.

[OASIS 2005] OASIS: "OASIS e-Government TC"; Web available (2005); http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=egov.

[OMG 2005a] OMG: "E-gov dsig"; Web available (2005a); http://egov.omg.org/.

[OMG 2005b] OMG: "Object management group"; Web available (2005b); http://www.omg.org/.

[POPKIN Software 2004] POPKIN Software: "FEAF"; Web available (2004); http://government.popkin.com/frameworks/feaf.htm.

[Seaborne 2007] Seaborne, A.: "RDQL - A Query Language for RDF"; Web available (2007); http://www.w3.org/Submission/RDQL/.

[Slovenia Government 2009] Slovenia Government: "State Portal of the Republic of Slovenia"; Web available (2009); http://e-uprava.gov.si/e-uprava/en/portal.euprava.

[Stojanovic 2008] Stojanovic, L.: "FIT Project "; (2008); www.accessegov.org.

[Tasmania Government 2009] Tasmania Government: "Service Tasmania Online"; Web available (2009); http://www.service.tas.gov.au/LifeEvents/.

[UK GovTalk 2004] UK GovTalk: "e-GIF"; Web available (2004); http://www.govtalk.gov.uk/.

[W3C 2004] W3C: "Web ontology language"; Web available (2004); http://www.w3.org/2004/OWL/.

[W3C 2006] W3C: "SPARQL Query Language for RDF"; Web available (2006); http://www.w3.org/TR/rdf-sparql-query/.

[W3C 2008] W3C: "egovernment interest group"; Web available (2008); http://www.w3.org/2007/eGov/IG/wiki/Main_Page.

[W3C 2009a] W3C: "Cascading style sheets"; (2009a); web available at http://www.w3.org/Style/CSS/.

[W3C 2009b] W3C: "Web services activity"; Web available (2009b); http://www.w3c.org/2002/ws/.

[W3C 2009c] W3C: "XHTML 1.0 The Extensible HyperText Markup Language (Second Edition)"; Web available (2009c); http://www.w3.org/TR/xhtml1/.

[W3C 2009d] W3C: "Xml path language"; Web available (2009d); http://www.w3.org/TR/xpath.

[Wimmer et al. 2001] Wimmer, M., , Wimmer, M., Krenner, J.: "An integrated online one-stop government platform: The egov project"; Project, 9th Interdisciplinary

Information Management Talks, Proceedings, Schriftenreihe Informatik, Universitätsverlag Trauner; 329–337; 2001.

[World Wide Web Consortium 2005] World Wide Web Consortium: “Rdf”; Web available (2005); http://www.w3c.org/RDF/.