Semantic based P2P System for local e-Government

Fernando Ortiz-Rodríguez[1], Raul Palma[2], Boris Villazón-Terrazas[2]

[1]Universidad Tamaulipeca
M, Escobedo, 88500 Reynosa. Tamaulipas. México
`fortiz@tamalipeca.net`
[2]Facultad de Informática. Universidad Politécnica de Madrid
Campus Montegancedo, s/n 28860 Boadilla del Monte. Madrid. Spain
`(rpalma, bvillazon)@fi.upm.es`

Abstract. The Electronic Government is an emerging field of applications for the Semantic Web where ontologies are becoming an important research technology. The e-Government faces considerable challenges to achieve interoperability given the semantic differences of interpretation, complexity and width of scope. This paper addresses the importance of providing an infrastructure capable of dealing with issues such as: communications between public administrations across government and retrieval of official and non official documents in a timely, secure and accurate way at the back office. A semantic peer-to-peer approach is proposed to enhance the information management at the e-Government domain; this approach is integrated with a Government Information Retrieval system and it reuses the EGO Model which can be deployed within the e-Government context.

1 Introduction and Motivation

Electronic Government (e-Gov) is an important application field [CS03] for the transformations that governments and public administrations will have to undergo in the next decades. Therefore, to transform the e-Gov into the e-Governance, the e-Gov research needs to be based on a robust theory, on modelling approaches, and on planning. In this scenario, a crucial issue is to manage in different ways the legal knowledge to improve and create semantic systems applications.

The Semantic Web was proposed by Tim Berners-Lee [HBM02] as a new field of research, and according to the World Wide Web Consortium[1] (W3C) the Semantic Web is defined as "an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. It is based on the idea of having data on the Web defined and linked such that it can be used for more effective discovery, automation, integration, and reuse across various applications".

---

[1] http://www.w3.org/2001/sw

The Peer-to-Peer (P2P) research community has recently started to use increasingly ontologies and Semantic Web technologies to enhance the applications and alleviate problems arising in query routing and network topology and in those issues related to the high level of distribution common in a P2P environment such as: the duplication of answers to a single query due to the duplication of information across the network; answers requiring the integration of information that resides at different, independent and uncoordinated peers; and the complexity of search queries due to the lack of a single coherent schema for organizing information sources across the network.

Even though the P2P concept is nothing new, given the decentralized, heterogeneous and dynamic nature of the e-Gov field, a P2P approach will provide an ideal solution to improve the information management processes across the local and autonomous Public Administration on behalf of the citizens and business.

Even though security issues are important for e-Gov domain application area, the scope of this paper only covers basic approaches about this topic. The remaining of this paper is organised as follows. In section 2, the related work is given; in section 3, The EGO Model is shown; in section 4, an Egoster overview is presented. And finally, in section 5 is devoted to conclusions.

2 Related Work

Nowadays the joint efforts put in by different research communities have made possible the birth of the semantic e-Gov. Since e-Gov ontologies are still in their initial state, only a few works carried out in this field are known; thus, in this section we provide a brief state of the art of those works performed in AI, in the law field and in the Semantic Web. The sum up of all these efforts will produce robust ontologies for the e-Gov domain in the near future.

2.1 E-Government within the Semantic Web

Currently, the Semantic Web is a new area of research and applications within the legal system and e-Gov domains [TL02] and is a promise for the Web of the next generation; this new area, which is now used mainly to communicate with people but not with machines, will transform the current web since the capability of communication with machines is one of the main objectives of the Semantic Web. If the Web were equipped with more meaning, every citizen would extract answers in a new, easy and simple way and this action could be carried out by web powered semantics, what would enable citizens and businesses to obtain better information from the government. Web powered semantics could help the e-Gov in two ways: first, by allowing the government to delegate more intelligent tasks to computers and second, by solving daily problems with logic deductions and reasoning. But at present, the web is merely a common framework that allows data to be shared and reused.

Currently e-Gov Semantic applications are still in an experimental phase, but their potential impact on social, economical and political issues is extremely significant.

The main goals of e-Gov are to develop user-friendly and efficient services for the public and the business community, though semantic interoperability is also seen as an important issue to solve within this domain. Some of the works aimed at covering the semantic e-Gov domain are the following: the DIP project[2], the Reimdoc project[3], The IFIP Working Group 8.5[4], the Ontogov project[5], the Egov project[6], HOPS project[7], and the WEBOCRAT project[8].

2.2 Semantic-based Peer- to-Peer systems

The SWAP [Eh03] project (Semantic Web and Peer-to-Peer) combines the P2P paradigm with Semantic Web technologies. In this sense, the SWAP environment which is based on the JXTA[9] platform aims at providing a general view of the knowledge each peer has. It facilitates the access to different information sources and enables the user to take advantage of other peer's knowledge. There exist many domain specific applications that are instances of the SWAP environment, like Bibster[10], for bibliographic metadata or Oyster[11], for ontology metadata [PH05]. Another semantic-based approach is pSearch [CZM02], a decentralized non-flooding P2P information retrieval system. pSearch distributes document indices through the P2P network and this distribution is based on document semantics generated by Latent Semantic Indexing (LSI).

Edutella[12] is another P2P system based on the JXTA platform, which offers similar base functionality as the SWAP system. Edutella focuses on the exchange of learning material.

In the e-Gov application domain, there exist some initial efforts, such as The X-Leges System[13], which is a P2P system for legislative document exchange.

3 EGO Model

We use a reference model to focus on and build a common understanding of the problem stated; Figure 1 shows the different actors within the e-Gov.

---

[2] http://dip.semanticweb.org

[3] http://reimdoc.atosorigin.es

[4] http://falcon.ifs.uni-linz.ac.at/research/ifip85.html#aim

[5] http://www.ontogov.com/

[6] http://www.egov-project.org

[7] http://www.bcn.es/hops/

[8] http://www.webocrat.org/

[9] http://www.jxta.org/

[10] http://bibster.semanticweb.org/

[11] http://oyster.ontoware.org/

[12] http://edutella.jxta.org/

[13] http://www.cnipa.gov.it/site/_files/Mecella_FURORE_7_4_2005_.ppt
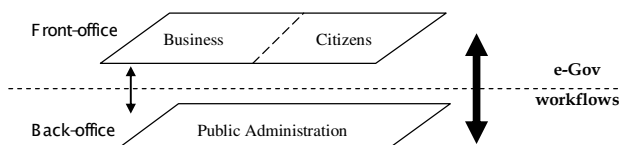
Fig 1. The e-Government Reference Model

At the Back-office, the main actor is the Public Administration; it has many processes inside which should work properly to provide efficient services. The dynamics of the Public Administration provides a huge amount of information to be processed and these data should be managed in a transparent and efficient way.

Within the Public Administration many processes take place and these must be carried out properly to provide efficient services; since the Public Administration functions in a decentralized way and the dynamics of this field generates a huge amount of information to be processed, it is necessary to manage this vast amount of information in a transparent and efficient way. Therefore, the implementation of e-Gov ontologies and applications is crucial.

The main objective of designing this initial model in the e-Gov domain is aims to represent the legal issues behind the governments. This model should work as a framework to deploy semantic e-Gov systems, under the law and regulations approach.

The EGO Model [Or06, OV06] reuses parts of the first two layers of LRI-Core model [Br04] and is being adapted to the legal system of the Spanish government. The EGO Model is one of the first efforts not intended for legal domain but for e-Gov domain instead, which is a domain that needs to consider the law, regulations, citizen services, administrative processes, best-practices, and also the different languages spoken within the nation.

4 Egoster

In this section we provide our approach to solve the document retrieval across local public administration and to improve the services in the back office. The reason we proposed a P2P approach instead of centralized solution relies on the fact that both share decentralized model and local autonomy. Furthermore, given the growing nature of the government and the back office services availability expectations, the P2P solution will provide the scalability and fault tolerance needed. This semantic-based P2P system was developed to apply advanced technology and create a new way of accelerating the IT capabilities. This will help the government to work better at the back-office on behalf of the citizens and business.

The transformation of isolated Public Administration into Peers must improve the communication between them and the semantics, which are added to help reduce the interoperability problem.

The benefits carried out on the implementation of a semantic P2P, has many different impacts, on the social stage: improving of the service to citizen across the different Public Administrations, a considerable reduction of budget in back-office

services, easy retrieving and interchange of legal documents at the back-office; at the technological stage we can say that the e-Gov is at the present time a testbed for semantic applications, and it is important to prove a P2P system on a real and complex environment.

4.1 Egoster goals

Egoster aims to exchange and retrieve the government documentation from e-Gov back office to allow final users (citizen and business) to get what they need in any time and any place where a peer exists.
In particular, Egoster goals are:
- To aid Public Administration to find and share legal documents within the back office.
- To provide interoperability between e-Govs offices.
- To improve citizens-government interaction.
- To offer a scalable technology that could be adapted to future requirements of the Public Administration.

4.2 Egoster Architecture

The Egoster system is implemented as an instance of the Swapster System architecture. Swapster provides the link between peer-to-peer technology and Semantic Web technology. It gives the semantics that are necessary for effective and precise knowledge sharing that is lacking in the majority of current peer-to-peer applications that allow only searching by keywords. Swapster also allows users to have their individual views on knowledge treating every user as a peer which may ask queries to the network of peers, and let them share knowledge effectively in a decentralized framework with low administration overhead.
Furthermore, in order to achieve an efficient and innovative infrastructure for the public administration offices, Egoster exploits the benefits obtained by the integration of Semantic Web technologies in a Peer-to-Peer network and combine them with Information Retrieval capabilities to provide its main functionalities (formulating queries, routing queries and processing queries).
With respect to security we are implementing the following security levels:
- Basic electronic signatures that can be used for the minimum level of transactions where peers simply want to ensure that the people at the other end can verify their identities.
- Qualified certificates, in which electronic signature is only one part of the authentication and authorization information stored on the certificate.

4.2.1 Egoster Functionalities
The government document collection [GOV05] is located at each public administration office and is stored locally in the file system of each peer (i.e. birthday certificate, real-state transaction documents, etc.). Each peer also implements a search

engine for the Information Retrieval capabilities. The search engine processes the legal document collection to create internally access structures that allow fast document location. These access structures, called indices, are stored locally in the file system. The search engine connects to the Legal Ontologies in order to obtain synonyms of the query concepts. In Egoster system the search engine is based on Reimdoc [GOV06] IR, which uses the Jakarta Lucene[14] Search library. The legal ontologies are used to provide a kind of thesauri to the search engine. Legal ontologies are used to improve the search process; these ontologies were built with the methodology METHONTOLOGY [GFC03] and the ontological engineering workbench WebODE [GFC03].

Formulating queries: The user can search for legal documents using simple keyword searches, or using more advanced, semantic searches. Here, queries are formulated in terms of two ontologies: (1) the legal document ontology which describes the properties of the legal documents, (2) a geographic hierarchy of local administration offices (i.e. the Spanish local administration offices), which describes all the local administration offices classified by the city and geographic region. This means queries can refer to fields like name, nationality, etc. (using the legal document ontology) or queries may refer to geographic location (using the geographic hierarchy). After we conclude the evaluation of the first version of Egoster, we expect to have a bigger amount of documents of different types, so we foresee the inclusion of the type of the document as another search criterion.

Routing queries: The user may query a single specific peer (e.g. their own computer, because it is the local administration office where the document resides, or users may want to query another peer in particular because this peer is known in advance as the location where the information resides), or a specific set of peers (e.g. all the members under the same manager), or the entire network of peers (e.g. when the user has no idea where to search), in which case queries are routed automatically in the network. In the latter case, queries are routed through the network depending on the expertise of the peers, describing which geographic space of the geographic hierarchy a peer is knowledgeable about. In order to achieve this expertise based routing, a matching function determines how closely the semantic content of a query matches the expertise of a peer [Ha04]. In the future, as we mention above, the type of the documents (i.e. birth certificate) that each peer has is going to be considered for the definition of the expertise of the peer as well.

Processing queries: In order to process a query, whether the query comes from the user interface or from other peers, the peer tries to answer the query or distribute it further according to the content of the query. The decision about which peers a query should be sent is based on the knowledge about the expertise of other peers. If the peer can answer the query, it uses its local search engine (i.e. the Information Retrieval system).

---

[14] http://lucene.apache.org/

4.3 Application Scenario

In this section we provide snapshots of the first prototype of the Egoster system. We have used Egoster to aid Public Administration to find and share government documents within the back office. The system will be installed in every Public Administrator office. In order to illustrate Egoster potential, we have adopted, as a case study, some extracts of a project aiming at seeking e-Gov strategies within Spain. The impact of this system has had an important impact on the back office.

Figure 2 shows the initial user interface of the system. On the left side we can observe the query area. In this query area the user can do a search according to specific fields of the document concept structure. Besides, the user can choose the target location of the document. On the right side we can see the results area. In this area, in the upper side we can see the list of documents that match the user query. In the bottom area we see the instance attribute values of the document. An additional window appears and shows the government document that has been retrieved. Once the user clicks on the local path of a given document, it opens and shows the government document content.
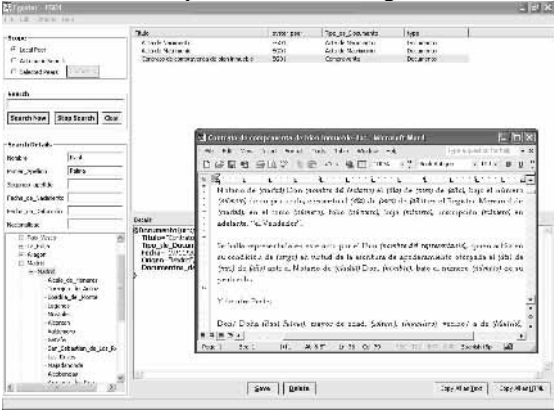

Fig. 2. A snapshot of Egoster retrieving a document

5   Conclusions

In this paper, we have shown how a semantic P2P system can be used to improve the communications between governments and citizens. The governments are formed of decentralized administrations, which are complex and dynamics, this brings us an issue that is solved with P2P approach. We also have shown how the inner capabilities of the Egoster system fit perfectly at the e-Gov domain.

The Swapster architecture on which Egoster is based has been successfully used in different applications (i.e. Bibster, Oyster, etc.). This provides confidence to use this technology in the e-Gov domain.

The e-Gov must take into account that being online does not rely only on a low cost services or a socio-political image for government, however governments need to explore and pursue new opportunities and leave behind old models and envision semantic models that empower the citizen.

In the future, we will work towards improving the actual Egoster application with the ongoing deployment and evaluation on real world environment, additionally we plan to enhance the system so as to have mobile e-Gov, to improve security functionalities using the European Electronic Signature Standardization and to provide a summary of the retrieved documents.

Bibliography

[Br04]   Breuker, J.: Constructing a Legal Core Ontology: LRI-Core.University of Amsterdam, Leibniz Center for Law, Amsterdam, the Netherlands (2004)

[Ha04]   P. Haase, R. Siebes, and F. van Harmelen. Peer selection in peer-to-peer networks with semantic topologies. In Proceedings of the International Conference on Semantics in a Networked World (ICNSW'04), Paris, June 2004.

[CZM02]   Chunqiang, T., Zhichen, X., Mahalin, M.: pSearch: Information Retrieval in Structured Overlays. ACM HotNets-I (2002)

[CS03]   Curtain, G., Sommer, M.: The World of E-Government. The Haworth Press. (2003)

[Eh03]   Ehrig, M., Haase, P., Siebes, R., Staab, S., Stuckenschmidt, H., Studer, R., Tempich, C.: Invited Contribution. The SWAP Data and Metadata Model for Semantics-based Peer-to-Peer Systems. In: MATES. Volume 2831 of Lecture Notes in Computer Science., Springer (2003) 144–155

[TL02]   Traunmüller, R., Lenk, K.: Electronic Government: First International Conference, EGOV 2002, Aix-en-Provence, France, September 2-5,. Proceedings (Lecture Notes in Computer Science) (2002)

[GFC03]   Gómez-Pérez, A.; Fernández-López, M.; Corcho, O. Ontological Engineering. Springer Verlag. 2003

[GOV05] Gómez-Pérez, A., Ortiz-Rodríguez, F., Villazón-Terrazas, B.: Legal Ontologies for the Spanish e-Government. In Proceedings of the 11th Conference of the Spanish Association for Artificial Intelligence. Santiago de Compostela, Spain (2005) To appear in CAEPIA'05 post-proceedings volume. Springer Verlag 2006

[GOV06] Gómez-Pérez, A., Ortiz-Rodríguez, F., Villazón-Terrazas, B.: Ontology-Based Legal Information Retrieval to Improve the Information Access in e-Government. In Proceedings of the WWW2006, Edinburgh, Scotland (2006) and In Demo and Poster Proceedings of the 3rd European Semantic Web Conference. Budva, Serbia y Montenegro (2006)

[HBM02] Hendler, J., Berners-Lee, T., and Miller, E.: Integrating Applications on the Semantic Web, (2002) http://www.w3.org/2002/07/swint.html

[Or06]   Ortiz-Rodríguez, F.: Mexican E-Government Ontologies: an adaptation. In Proceedings of the Fourth Latin American and Caribbean Conference. Mayaguez, Puerto Rico (2006)

[OV06]   Ortiz-Rodríguez, F., Villazón-Terrazas, B.: EGO Ontology Model: law and regulation approach for E-Government. In WS Proceedings of the 3rd European Semantic Web Conference. Budva, Serbia y Montenegro (2006)

[PH05]   Palma R. Haase Peter. Oyster - Sharing and Re-using Ontologies in a Peer-to-Peer Community. International Semantic Web Conference 2005: 1059-1062