

EgoIR: ontology-based information retrieval intended for eGovernment

Fernando Ortiz-Rodríguez, Raul Palma, Boris Villazón-Terrazas

Facultad de Informática. Universidad Politécnica de Madrid
Campus Montegancedo, s/n 28860 Boadilla del Monte. Madrid. Spain
(fortiz, rpalma, bvillazon)@fi.upm.es

Abstract: The eGovernment is a field of applications for the Semantic Web. The eGovernment also is becoming an important research area and faces considerable challenges to achieve interoperability because of the semantic differences of interpretation, complexity and width of scope. It is however an open question how to apply these techniques fruitfully in the eGovernment domain. This paper addresses the importance of providing a semantic application that, within the eGovernment domain, is capable of dealing with the issue of the retrieval of government documentation in a timely and accurate way. In this paper, we present an approach ontology-based for retrieving government information.

1 Introduction

Electronic Government (eGov) is a significant application field [CS03] for the transformations that governments and public administrations will have to undergo in the next decades. Therefore, is necessary to transform the semantic eGov from electronic service to networked societal guidance to achieve the eGovernance (i.e. application of ICT to steer or guide society so as to best public interests and achieve the common good). The eGov 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 applications. The advancing technologies are being able to deal with real world legal knowledge utilized by governments. In concrete, the Semantic Web [HBM02] has the capabilities to face this challenge.

The application of the Semantic Web to the eGov domain is completely new. The Semantic Web is considered to be the infrastructure upon which all intelligent eGov applications will be built in the near future. Even though security issues are important for eGov 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 presented; in section 3, the EGODO Ontology is described; in section 4, the EgoIR information retrieval is shown; in section 5, improved services of EgoIR with a peer to peer are described. And finally, in section 6 is devoted to conclusions.

2 Related Work

There are many systems developed for managing legal information, but only a few deals with legal knowledge. In this section we describe briefly some legal IR systems. In [Wi98], CLIME aims at improving the access and understanding of large collections of legal information through the Internet. CLIME just combines conventional IR with artificial legal reasoning without ontologies. In [DP01], the authors describe the Webocrat system whose goal is to provide new types of communication and service flows from public institutions toward citizens, thus improving the access of citizens to services and information of public administration. This system focuses on security issues. Another work reported in [LTS00] is the EULEGIS, whose main goal is to provide a consistent user interface for legal IR generated in different legal systems and at different legislative levels. This system focuses on user interfaces.

3 EGODO Ontology

Within the Public Administration many processes take place and these must be carried out properly to provide efficient services. For example, the Public Administration works in a decentralized way and the dynamics of this field generates a huge amount of information to be processed, therefore it is necessary to manage this vast amount of information in a transparent and efficient way. Thus, the implementation of eGov ontologies and semantic applications is crucial.

The Documentation Ontology EGODO [Or07] is the core ontology used by the EgoIR. EGODO aims to address the current problems and deficiencies that governments faces nowadays in the management of legal documents. One main objective of this Documentation Ontology is to make all documents, regardless of the format, accessible and secure through any access point across governments; as a secondary objective is to regulate the access of the role actors within eGov and the documentation. EGODO pursue to improve in some way the outcome of the interaction of government and citizens to achieve the e-Governance. EGODO is a part of EGO model. The EGO ontology model [OV06] is being adapted to the legal system of the Spanish and Mexican government. The EGO Ontology Model is one of the first efforts not intended for legal domain but for eGov domain instead.

4 EgoIR

This section describes our approach, EgoIR, which is an ontology-based information retrieval system intended for eGov. This system is the result of integrate the Ontological Workbench WebODE¹, and a text search engine library, Lucene². The main idea is to have a system which can deal with government information.

¹ <http://webode.dia.fi.upm.es/>

4.1 EgoIR Goals

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

4.2 EgoIR Architecture

EgoIR approach solves the information retrieval issues generated by using Semantic Web technologies (e.g., accessibility, accuracy). It is our first approach to link domain ontologies with an information retrieval. The system integration of the EgoIR is built and composed by the Search Client, the Search Server and the Ontology Server modules, which are described in the next subsections. Figure 1 shows the general architecture of the system.

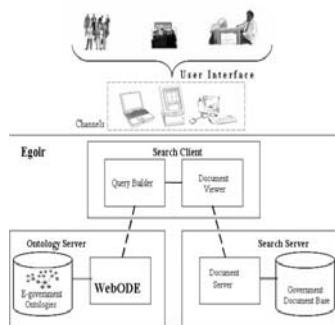


Figure 1. EgoIR system architecture

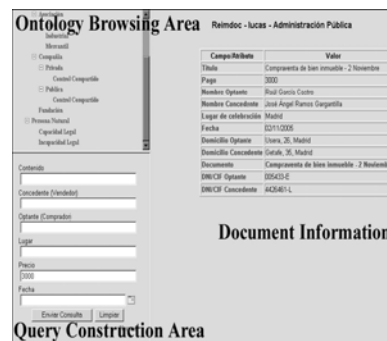


Figure 2. A snapshot of EgoIR user interface

4.2.1. Ontology Server. This module defines how the knowledge is structured in the application domain. It includes the EGODO as the main ontology and several legal ontologies [GOV05] that support this module. Within the ontologies, concept instances are associated with documents. Every time that a new concept instance is added the Ontology Server communicates with the Search Server to index its corresponding document.

4.2.2, Search Server. This module incorporates two sub-modules: a Query Builder and a Document Viewer. The Query Builder connects to the Ontology Server, in order to access the ontologies, browse them and obtain concepts to build the query by using a graphical interface. It sends the query to the Search Server. The Document Viewer connects to the Search Server, in order to retrieve the government documents satisfying the query, and to the Ontology Server to browse and display the documents.

² <http://jakarta.apache.org/lucene>

4.2.3. Search Client. The Search Server module is based on Lucene and processes the Government Document Base to create internally access structures. These structures (called indices) allow fast document location and are stored locally in the file system of the operating system. The Government Document Base consists of electronic documents that are stored in the file system.

Figure 2 shows the EgoIR user interface. On the left side we can see the ontology browsing area and the query construction area, and on the right side the document information. Currently the document's annotation process is manually done. When a concept instance is created, using WebODE interface, the values from its instance attributes as well as the electronic document are indexed using Lucene.

4.3 EgoIR Functionalities

EgoIR as an eGov ontology-based information retrieval system that provides the following main functionalities: Query Construction, Document retrieval and Type of documents.

4.3.1. Query Construction. In this option the user, through a guided interface, navigates among the ontologies, exploring concepts and relationships; and with the selected concepts, the user builds the query. Once the user has finished adding concepts to the query, the user sends the query to the document server.

4.3.2. Document Retrieval. In this option the system retrieves the government documents that match the user query constructed. The documents coming from the document Server are retrieved and showed in a list form. If the user wants to choose a given document he only has to click on the document. Then, the system launches a window with the selected document.

5 Government Document Exchange

The EgoIR is reused by Egoster [OPV06] peer to peer system. Egoster is a java-based system that exploits semantic web techniques in order to provide an efficient way to exchange information among public administration offices. This system is located at each public administration office and is stored locally in the file system of each peer. The Egoster system is implemented as an instance of the Swapster System architecture. Swapster [Eh03] provides the link between peer-to-peer technology and Semantic Web technology.

6 Conclusions

In this paper we have presented our approach to manage and maintain a set of government documents within public administration, which aims to retrieve Government documents in a timely and accurate way at the Back-office.

The Back-office powered by EgoIR will be able to provide document retrieval to citizens and businesses with an im-proved service which consists in retrieving, sharing and exchanging the exact information which they are looking for, and only that specific information. This is an approach of an entirely new wave of eGov knowledge systems. At this time we can mention the utility of ontologies within an information retrieval in two senses. On one hand, as a social impact, ontologies provide the means to aid users with government terms, avoiding the mistakes at the query construction; and on the other hand, mostly technical, being a key to the develop the Semantic Web and improving interoperability on the government applications.

Our future work includes many challenges. In particular we will improve the performance of the EgoIR and we will focus on further enhancement of the ontology-based retrieval mechanism by means of natural language processing, the automatic annotation, and the security issues and to improve the retrieving capabilities with the added peer to peer technology.

References

- [CS03] Curtain, G., Sommer, M., Vis-Sommer, V. (editors): *The World of eGovernment*. The Haworth Press 2003
- [DP01] Dridi, F., Pernul, G.: *The Webocracy project: Overview and Security Aspects*. In Schnurr et al., *Professionelles Wissensmanagement: Erfahrungen und Visionen*. Shaker Verlag, Aachen, 2001.
- [Eh03] Ehrig, M., et.al.: *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
- [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 eGovernment*. In *CAEPIA'05 post-proceedings revised papers volume*. Springer Verlag 2005
- [HBM02] Hendler, J., Berners-Lee, T., and Miller, E. *Integrating Applications on the Semantic Web*, 2002 <http://www.w3.org/2002/07/swint.html>
- [LTS00] Lytikinen, V., Tiitinen, P., Salminen, A.: *Challenges for European Legal Information Retrieval*. Proceedings of the IFIP 8.5 Working Conference on Advances in Electronic Government. Zaragoza, Spain, 2000.
- [Or07] Ortiz-Rodríguez, F.: *EGODO and Applications: sharing, retrieving and exchanging legal documentation across eGovernment*. In *Proceedings of SW4Law at ICAIL 2007, the International Conference on AI and Law*. Stanford, Palo Alto, CA, 2007
- [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
- [OPV06] Ortiz-Rodríguez, F., Palma, R. Villazón-Terrazas, B.: *Semantic based P2P System for local eGovernment*. In *Proceedings of Informatik 2006. GI-Edition-Lecture Notes in Informatics (LNI) Dresden*. Germany 2006
- [Wi98] Winkels, R.G.F.: *CLIME: Legal Information Serving Put to the Test*. Pre-proceedings of the Second French-American Conference on AI and LAW. Available online at: <http://www.Iri.jur.uva.nl/~winkels/papers/AIL98-paper.html>