Enhanced access to eGovernment services: temporal and semantics-aware retrieval of norms*

Fabio Grandi[1], Federica Mandreoli[2], Riccardo Martoglia[2], Enrico Ronchetti[2], Maria Rita Scalas[1], and Paolo Tiberio[2]

[1] Alma Mater Studiorum – Università di Bologna, Italy
{fgrandi,mrscalas}@deis.unibo.it
[2] Università di Modena e Reggio Emilia, Italy
{fmandreoli,rmartoglia,eronchetti,ptiberio}@unimo.it

Abstract. In this paper, we summarize the results of an ongoing research involving the design and implementation of a multi-version repository of norm texts supporting efficient and personalized access in an eGovernment scenario. The research activity is entitled "Semantic web techniques for the management of digital identity and the access to norms". In the context of a complete and modular infrastructure, we defined a multiversion XML data model and developed an XML query processor supporting both temporal and semantic versioning. Semantic versioning is based on the applicability of different norm parts to different classes of citizens and allows users to retrieve personalized norm versions only containing provisions which are applicable to their personal case. The whole infrastructure, which we plan to complete in the near future, will integrate the query answering component with several auxiliary services, including automatic citizen identification and classification and computer-aided update of the repository data.

1 Introduction

Nowadays we are witnessing a strong institutional push towards the implementation of eGovernment support services, aimed at a higher level of integration and involvement of the citizens in the Public Administration (PA) activities that concern them. In this framework takes its place the research activity entitled "Semantic web techniques for the management of digital identity and the access to norms", which we are carrying out as part of the PRIN Italian project "European Citizen in eGovernance: legal-philosophical, legal, computer science and economical aspects" [1]. One of the main objectives of such activity is the development of a system allowing an effective and efficient access to multi-version norm repositories supporting temporal queries and personalization facilities. In fact, the fast dynamics involved in normative systems implies the coexistence of

* An extended version of this paper has been presented at the DEXA EGOV 2005 Conference [4]

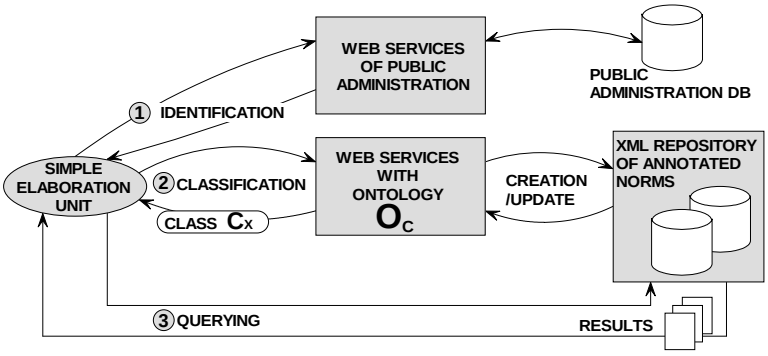

Fig. 1. The Complete Infrastructure

multiple *temporal versions* of the norm texts stored in a repository, since laws are continually subject to amendments and modifications (for instance, it is crucial to reconstruct the consolidated version of a norm as produced by the application of all the modifications it underwent so far). Moreover, a citizen accessing the repository may be interested in finding a *personalized version* of the norm, that is a version only containing articles which are applicable to his/her personal case.

In order to enhance the participation of the citizens to an eGovernance procedure of interest, their automatic and accurate positioning within the reference legal framework is needed. To solve this problem we employ Semantic Web techniques and introduce a *civic ontology*, which corresponds to a classification of citizens based on the distinctions introduced by subsequent norms which imply some limitation (total or partial) in their applicability. In the following, we refer to such norms as *founding acts*. Moreover, we define the citizen's *digital identity* as the total amount of information concerning him/her –necessary for the sake of classification with respect to the ontology– which is available online [9]. Such information must be retrievable in an automatic, secure and reliable way from the PA databases through suitable Web services (*identification services*). For instance, in order to see whether a citizen is married, a simple query concerning his/her marital status can be issued to registry databases. In this way, the classification of the citizen accessing the repository makes it possible to produce the most appropriate version of all and only norms which are applicable to his/her case.

Hence, the resulting complete infrastructure is composed by various components that have to communicate between each other to collect partial and final results (see Fig. 1).

Firstly, in order to obtain a personalized access, a secure authentication is required for a citizen accessing the infrastructure. This is performed through a simple elaboration unit, also acting as user interface, which processes the citizen's requests and manages the results. Then, we can identify the following phases:

– the **identification phase** (step 1 in Fig. 1) consists of calls to identification services to reconstruct the digital identity of the authenticated user on-the-fly. In this phase the system collects pieces of information from all the involved PA web services and composes the identity of the citizen.
– the citizen **classification phase** (step 2 in Fig. 1) in which the classification service uses the collected digital identity to classify the citizen with respect to the civic ontology, by means of an embedded reasoning service. In Fig. 1, the most specific class C_X has been assigned to the citizen;
– finally, in the **querying phase** (step 3 in Fig. 1) the citizen's query is executed on the multi-version XML repository, by accessing and reconstructing the appropriate version of all and only norms which are applicable to the class C_X.

In order to supply the desired services, the digital identity is modelled and represented within the system in a form such that it can be translated into the same language used for the ontology (e.g. a Description Logic [3]). In this way, during the classification procedure, the matching between the civic ontology classes and the citizen's digital identity can be reduced to a standard reasoning task (e.g. ontology entailment for the underlying Description Logic [7]).

Furthermore, the civic ontology used in step 2 and 3 requires to be created and constantly maintained: each time a new founding act is enforced, the execution of a **creation/update phase** is needed. Notice that this process (and also the introduction of semantic annotations into the multi-version XML documents) is a delicate task which needs advice by human experts and "official validation" of the outcomes and, thus, it can only partially be automated. However, computer tools and graphic environments (e.g. based on the Protégé platform [8]) could be provided to assist the human experts to perform this task.

For the specification of the identification, classification and creation/update services, we plan to adopt a standard declarative formalism (e.g. based on XML/SOAP [10]). The study of the services and of the mechanisms necessary to their semi-automatic specification will be dealt with in future research work.

2 Personalized access to norms

Up to now, our research has been mainly focused on the querying phase. In particular, we defined an XML data model which combines semantic annotations with temporal versioning and we defined efficient techniques for querying repositories storing legal documents supporting temporal and semantic versioning. We first focused on the temporal aspects and on the effective and efficient management of time-varying norm texts. The proposed data model [6] uses four time dimensions (*validity time*, *efficacy time*, *transaction time*, *publication time*) to correctly represent the evolution of norms in time and their resulting versioning. Then, we enhanced the model to include a semantic versioning mechanism to provide personalized access; the semantic versioning dimension encodes information about the applicability of different parts of a norm text to the relevant classes of the civic ontology defined in the infrastructure [5]. Consider, for in-

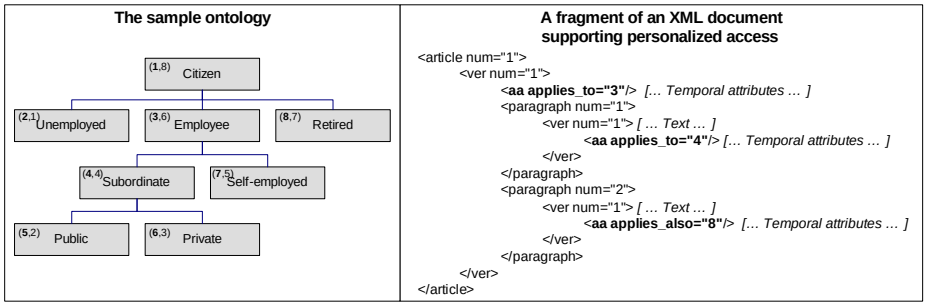

Fig. 2. An example of civic ontology, where each class has a name and is associated to a (pre,post) pair, and a fragment of a XML norm containing applicability annotations.

stance, Fig. 2. The left part of the figure depicts a simple civic ontology built from a small corpus of norms ruling the status of citizens with respect to their work position. The right part of the figure shows a fragment of a multi-version XML norm text supporting personalized access with respect to this ontology. Notice that, at this stage of the project, we manage "tree-like" ontologies defined as class taxonomies induced by the IS-A relationship. This allows us to exploit the pre-order and post-order properties of trees in order to enumerate the nodes and check ancestor-descendant relationships between the classes; such codes are displayed in the upper left corner of the ontology classes in the Figure, in the form: (pre-order,post-order). For instance, the class "Employee" has pre-order "3" which is also its identifier, whereas its post order is "6". The article in the XML fragment on the right of Fig. 2 is composed of two paragraphs and contains applicability annotations (aa). Notice that applicability is inherited by descendant nodes unless locally redefined. Hence, by means of redefinitions we can also introduce, for each part of a document, complex applicability properties including extensions or restrictions with respect to ancestors. For instance, the whole article in the Figure is applicable to civic class "3" ($applies\_to$) and by default to all its descendants. However, its first paragraph is applicable to class "4", which is a restriction, whereas the second one is also applicable to class "8" ($applies\_also$), which is an extension. The reconstruction of pertinent versions of the norm based on its applicability annotations is very important in an eGovernment scenario. The representation of extensions and restrictions gives rise to high expressiveness and flexibility in such a context.

Currently, the queries that can be submitted to our system can contain four types of constraints: temporal, structural, textual and applicability. Such constraints are completely orthogonal and allow users to perform very specific searches in the XML norm repository. Let us focus first on the applicability constraint. Consider again the ontology and norm fragment in Fig. 2 and let John Smith be a "self-employed" citizen (i.e. belonging to class "7") retrieving the norm: hence, the sample article in the Figure will be selected as pertinent, but only the second paragraph will be actually presented as applicable. Furthermore,

the applicability constraint can be combined with the other three ones in order to fully support a multi-dimensional retrieval. For instance, John Smith could be interested in all the norms ...

- which contain paragraphs (*structural constraint*) dealing with health care (*textual constraint*), ...
- which were valid and in effect between 2002 and 2004 (*temporal constraint*), ...
- which are applicable to his personal case (*applicability constraint*).

Such a query can be issued to our system using the standard XQuery syntax. To this purpose, we defined two functions which can be used in the XQuery WHERE clause to specify temporal and applicability constraints [4], whereas structural constraints are implicit in the XPath expressions used in the query.

Notice that the temporal constraints can involve all the four available time dimensions (publication, validity, efficacy and transaction), allowing high flexibility in satisfying the information needs of users in the eGovernment scenario. In particular, by means of validity and efficacy time constraints, a user is able to extract consolidated current versions from the multi-version repository, or to access past versions of particular norm texts, all consistently reconstructed by the system on the basis of the user's requirements and personalized on the basis of his/her identity.

In order to test both the efficacy of the approach and the efficiency of the proposed technique, we built a prototype system implementing the data model. The system exploits "XML-native" architecture, as it is composed of a Multi-version XML Query Processor designed on purpose, which is able to manage the XML data repository and to support all the temporal, structural, textual and applicability query facilities in a single component.

The designed Processor stores the XML norms not as entire documents but by converting them into a collection of ad-hoc temporal tuples, representing each of its multi-version parts (i.e. paragraphs, articles, and so on); during the query processing these data structures are exploited to efficiently perform structural join algorithms [2] we specifically devised and tuned for the temporal/semantic multi-version context. Textual constraints are handled by means of an inverted index. The benefits of our approach are manifold:

- by querying ad-hoc and temporally-enhanced structures (which have a finer granularity than the entire documents managed by standard XML engines), the system is able to access and retrieve only the strictly necessary data;
- only the parts which are required and which satisfy the temporal constraints are used for the reconstruction of the retrieved documents;
- there is no need to retrieve whole XML documents and build space-consuming structures such as DOM trees.

The system is thus able to provide very fast computation efficiency and good scalability qualities in several query contexts, as we experimentally proved through several experiments [4].

3 Conclusions and future work

Our current research work is devoted to the extensions of the current framework and to the development of the complete technological infrastructure enabling our approach to be self-contained and usable in a large Web-based eGovernment scenario. The adoption of a tree-like civic ontology –that is based on a taxonomy induced by the IS-A relationship– is sufficient to satisfy basic application requirements as far as applicability constraints and personalization services are concerned. However, more advanced application requirements include a more sophisticated ontology definition. As a matter of fact, we are currently working in extending the framework in order to fully support generic graph-like ontologies, containing, for instance, multiple interconnected IS-A taxonomies and equivalence relations between the classes. To this aim, the XML documents annotation scheme and their storage organization need to be enhanced. Further, we are also considering the use of new ad-hoc indexes on the semantic attributes to enhance the efficiency of the query processing algorithms even more for highly-selective applicability queries. On the other hand, further work will also include the assessment of our developed systems in a concrete working environment, with real users and in the presence of a large repository of real legal documents. A civic ontology based on a corpus of real norms (concerning infancy schools) is currently under development.

References

1. Semantic web techniques for the management of digital identity and the access to norms. PRIN Project. http://www.cirsfid.unibo.it/eGov03/.
2. S. Al-Khalifa, H.V. Jagadish, J. M. Patel, Y. Wu, N. Koudas, and D. Srivastava. Structural joins: A primitive for efficient xml query pattern matching. In *Proc. of 18th Int'l Conf. on Data Engineering (ICDE 2002)*, pages 141–154, San Jose, CA, 2002.
3. F. Baader, I. Horrocks, and U. Sattler. Description logics for the semantic web. *Künstliche Intelligenz*, 16(4):57–59, 2002.
4. F. Grandi, F. Mandreoli, R. Martoglia, E. Ronchetti, M. R. Scalas, and P. Tiberio. Personalized access to multi-version norm texts in an egovernment scenario. In *Proc. of DEXA EGOV 2005*, Copenhagen, Denmark, 2005.
5. F. Grandi, F. Mandreoli, R. Martoglia, and M. R. Scalas. Efficient management of multi-version xml documents for e-government applications. In *Proc. of WEBIST 2005*, Miami, FL, 2005.
6. F. Grandi, F. Mandreoli, and P. Tiberio. Temporal modelling and management of normative documents in xml format. *Data & Knowledge Engineering*, 54(3), 2005.
7. Ian Horrocks and Peter F. Patel-Schneider. Reducing owl entailment to description logic satisfiability. In *Proc. of ISWC 2003*, pages 17–29, Sanibel Island, FL, 2003. LNCS No. 2870.
8. Owl plugin for protégé. http://protege.stanford.edu/plugins/owl/, 2004.
9. S. Rodotà. Introduction to the one world, one privacy session. In *Proc. of 23rd Data Protection Commissioners Conf.*, Paris, France, 2001.
10. Web services activity. W3C Consortium, http://www.w3.org/2000/xp/Group/, 2004.