Management of the Citizen's Digital Identity and Access to Multi-version Norm Texts on the Semantic Web[⋆]

Fabio Grandi[1], Federica Mandreoli[2], Maria Rita Scalas[1], and Paolo Tiberio[2]

[1] Alma Mater Studiorum - Università di Bologna,
Dip. di Elettronica, Informatica e Sistemistica, Bologna, Italy
{fgrandi,mrscalas}@deis.unibo.it
[2] Università di Modena e Reggio Emilia,
Dip. di Ingegneria dell'Informazione, Modena, Italy
{mandreoli.federica,paolo.tiberio}@unimo.it

Abstract. This paper describes an ongoing research project involving the implementation of e-Government services on the Semantic Web. In particular, the project is aimed at managing the "digital identity" of citizens on the Internet, enabling them to benefit from "personalized" versions of the online services offered by the Public Administration, which can improve and optimize their involvement in the e-Governance process. The kind of service we will consider is the selective access to norm texts available on Web repositories. The project requires the definition and maintenance of a citizen's ontology, the semantic markup and versioning of the stored norm texts which takes into account the actual applicability to different classes of citizens, the definition and enactment of Web services for the reconstruction of the citizen's digital identity and its classification with respect to the ontology, the design and implementation of a legal document management system for the selective access to personalized norm versions.

1 Introduction

This paper describes the program of an Italian research project entitled "Semantic Web techniques for the management of digital identity and access to norms" currently carried out at the University of Modena and Reggio Emilia in cooperation with researchers from the University of Bologna. The initial idea underlying the definition of the research project was inspired by a definition of *digital identity* taken from a talk by the President of the Italian Data Protection Commission [8]:

> *Transferred into cyberspace, fragmented into a multiplicity of databanks scattered around the world, the personal identity, divided, explodes on the network. Each one of us may be one person, a hundred thousand persons, or no-one. The personality corresponds to the multiple windows which can be opened on the screen.*

In particular, disregarding at the moment all the possible risks and disadvantages, every piece of data concerning a citizen, stored in Public Administration (PA) databases and available online, represents a fragment of the citizen's digital identity which could be useful to *improve and optimize* his/her participation to the e-Governance process. Therefore, the premise of this research lies in the following "facts":

1. A consistent research effort, gaining momentum, in the field of Semantic Web [2], which is a collection of machine-understandable resources that can be shared and used in an automatic way by intelligent computer tools.
2. A wide availability on the Internet of public collections of norm texts (e.g. XML-encoded, also stored in "official" data banks managed by the PA [10]).
3. An online availability of information, managed by the various administrations, that make up fragments of the citizen's "digital identity".
4. An institutional push towards the implementation of E-Government support services, aimed at a higher level of integration and involvement of the citizens, in particular, in the PA activities that concern them directly.

In this scenario, we identify as a missing link the lack of techniques as those described at point (1) for the exploitation of resources such as described at points (2) and (3) to face the application requirements arising from point (4). Such a gap, which needs to be filled, represents the starting point of our project. Therefore, the main effort will be devoted to the development of a technological infrastructure based on the Semantic Web for the citizen's personalized access to PA services. In particular, the service chosen for the project is the access to norm texts, and the personalization of the service consists of the *retrieval and reconstruction of all and only norms that are valid and applicable to the citizen also according to his/her digital identity*.

As far as point (2) is concerned, we want to emphasize how the fast dynamics involved in normative systems actually implies the coexistence of *multiple versions* of the norm texts stored in a database. The proliferation of versions is due to the fact that laws are subject to amendments and modifications in time but also to the fact that some of them have (or acquire) a limited applicability. For instance, an active norm might state a modification to a preexisting passive norm, where the modified norm is applicable to a limited category of citizens only, whereas the rest of the citizens are still subject to the unmodified norm. The adoption of temporal dimensions for the versioning of norm texts stored as XML documents –and the problem of reconstructing consolidated acts with respect to temporal versioning– were considered in previous approaches (e.g. [7, 3]). On the other hand, the problem of partial applicability was not considered before and will be the core of the research described here.

In general, machine-understanding of the information available on the Semantic Web requires a semantic markup of the contents and the availability of automated reasoning tools. In order to let information and its interpretation be shared by several agents (human users or automatic tools), the introduction of common reference *ontologies* becomes necessary [4]. At the state-of-the-art, several proposals have been put forward for the definition and standardization of the formalisms [12], usually with an XML-like syntax, which are needed for the definition of ontologies and the introduction of annotations in the Semantic Web. Such formalisms, typically based on the adoption of De-
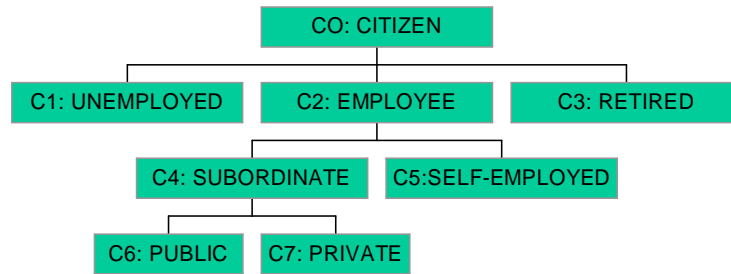
Fig. 1. A simple Civic Ontology

scription Logics as the underlying layer supporting reasoning services [1], also include languages for the formal specification of services which need to be also implemented.

2 The Research Program

The present project implies the creation, maintenance and usage of a legal ontology of the citizen (considered as a passive subject of laws), that we will call **civic ontology** in the following. This ontology corresponds to a classification of citizens based on the distinctions introduced by successive norms that imply some limitation (total or partial) in their applicability. In the following, we will refer to such norms as **founding acts** of the civic ontology. For instance, we can consider as founding acts a small *corpus* of norms ruling the status of citizens with respect to their work position and assume that the definitions introduced there allow us to define the simple ontology in Fig. 1 (simply defined as a class taxonomy induced by the IS-A relationship).

The existence in the ontology of different (sub-)classes of citizens represents, in all respects, a new "versioning dimension" for the modelling and processing of norm texts. Hence, the results of previous work on temporal versioning of XML legal documents [3] will be extended in this research to also introduce references to **civic subclasses** used as a new versioning coordinate. The norm texts stored in an XML repository will thus contain variant parts: single text portions will be annotated as only pertinent to the (maximal extension) subclass to which they are applicable. For example, let us consider a new norm act, say Law No. 1/2004 (e.g. defining tax treatment), whose text portions may separately apply to different classes of citizens: For instance, Article 1, which is made up of four paragraphs, applies to all employees but Paragraph 2 only applies to public servants, whereas Paragraph 3 applies to employees who are also retired persons (i.e. to citizens both having a pension and a new job). Once the ontology has been defined, class information can be added as a semantic markup in the XML encoding of the norm text. For example, our sample Law can be annotated as shown in Fig. 2 (for the sake of simplicity, temporal annotations are not included). Notice that Paragraph 3 of Art. 1 actually applies to the intersection class (`C2 and C3`), since the paragraphs also "inherit" the applicability of their containing article, whereas the whole `C6` is the maximal subclass to which Paragraph 2 can be applied, since it is a subclass of `C2`.

```
<norm type="Law" num="1/2004">
   ...
      <article num="1">
         <applies to="class:C2">
            <paragraph num="1">
                  --- text of Art.1 Par.1 ---
            </paragraph>
            <paragraph num="2">
               <applies to="class:C6">
                  --- text of Art.1 Par.2 ---
               </applies>
            </paragraph>
            <paragraph num="3">
               <applies to="class:C3">
                  --- text of Art.1 Par.3 ---
               </applies>
            </paragraph>
            <paragraph num="4">
                  --- text of Art.1 Par.4 ---
            </paragraph>
         </applies>
      </article>
   ...
</norm>
```

Fig. 2. A fragment of XML norm text containing semantic annotations

The operation of reconstruction of a consolidated act (e.g. as defined in [3]) will then be extended to return all and only contents which are pertinent, not only to the temporal coordinates, but also to the civic subclass of interest. In addition to such a reconstruction operation, we will also consider other advanced query modalities, like norm retrieval by keywords and similarity search, for which physical organizations and optimization algorithms will also be studied. To this end, extensions to the system architecture and the implementation solutions presented in [3] will be considered. Notice that the new versioning dimension is orthogonal to the temporal dimensions but has interesting interactions with them, as explained in the example which follows. Let us consider, as a new founding act, a pension reform which states the following: "Paragraph 3 of Art. 1 of Law No. 1/2004 is abrogated. However, it continues to be applicable until 2010 to persons retired since 2000." As a consequence, the ontology must be updated as shown in Fig. 3 with the addition of the subclass C8 containing citizens retired since 2000. Hence, after the enactment of the reform, Paragraph 3 of the sample Law No. 1/2004 does not belong any longer to Art. 1 of the Law version currently in force, since it has been abrogated. However, the previous version has still efficacy from now to the end of 2010 for the subclass C8. Therefore, the temporal data model proposed in [3] for multiversion norm texts, including maintenance operations, must be carefully extended to correctly deal also with the new versioning dimension. Such an extension is currently under development at this stage of the project.
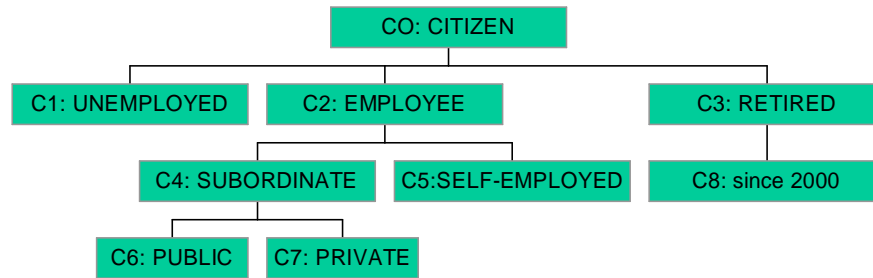
Fig. 3. The Civic Ontology after the pension reform

The civic ontology, for the sake of availability in a Semantic Web scenario, will be specified using a standard formalism (e.g. based on RDF, RDF-S, OWL [12]). In particular, we intend to adopt an OWL sublanguage which can be translated onto a decidable Description Logic for which efficient reasoning algorithms exist. For example, [6] showed how the basic reasoning task of *ontology entailment* in the OWL Lite sublanguage can be translated into *knowledge base satisfiability* in the $\mathcal{SHIF}(\mathbf{D})$ Description Logic, for which an effective and highly-optimized implemented reasoner exists (RACER, [5]). Other useful reasoning tasks, including *instantiation* with respect to an ontology (which can be used to test applicability for a single citizen identity) and *subsumption* with respect to an ontology (which can be used to test applicability for a complete subclass of citizens) can be reduced to *ontology entailment* problems. The specifications for the ontology definition and the design of a case-study ontology are also currently under development (in cooperation with the philosophical-juridical Research Unit).

In this context, the citizen's digital identity will be defined as the total amount of information concerning him/her, which is necessary for the sake of classification with respect to the ontology. All such information must be retrievable in an automatic way from the PA databases. To this purpose, facilities for querying PA databases must be provided and implemented through standardized access services. In order to supply the desired services, the digital identity will be modelled and represented within the system in a form such that it can be translated into an expression of the same Description Logic used for the ontology. In this way, during the consolidated act reconstruction procedure, the matching between the annotation of a text portion and the citizen's digital identity to ascertain applicability can be reduced to a standard reasoning task for the underlying Description Logic. The reconstruction and the classification operations will be encapsulated into suitable Web services.

Each time a new founding act is enforced, the civic ontology needs to be updated and its consistency re-checked. Actually, the ontology update process cannot be fully automated, since it is a delicate task which needs advice by human experts and "official validation" of the outcomes. However, computer tools and graphic environments (e.g. based on the Protégé platform [11]) could be provided to assist the human experts to perform this task. Moreover, the introduction of a new founding act must also trigger the specification of a new Web service aimed at retrieving from the network the infor-

mation necessary to verify the position of a citizen with respect to the distinguishing features newly introduced by the founding act. For example, if the new law states some benefit for former public servants retired since 2001, the ontology must be enriched with a new subclass corresponding to such a description, the norm will be annotated with a reference to the new subclass and, at the same time, a new Web service must be specified in order to verify whether a citizen belongs to the new subclass by querying the database of the public body paying out pensions (i.e. INPDAP, for Italian public servants). The specification of such services could be completely automated or, more likely, will be effected through a semi-automated process involving a human expert by means of an "intelligent" interactive editor, which is used for the recording of the new laws in legal databases. Once formally specified (and "officially" validated), such services (**identification services**[3]) will anyway allow a completely automated verification, by effectively and efficiently supplying the fragment of the citizen's digital identity which can be used for the desired high-level services.

For the specification of reconstruction, classification and identification services, we also intend to adopt a declarative formalism which is as standard as possible (e.g. based on XML/SOAP [13], like WSDL, DAML-S, BPEL4WS). The study of services and of the mechanisms necessary to their semi-automatic specification will be carried out in the second year of the research.

3 Conclusions

The inspiring objective of this research is the study of information systems and services aimed at facilitating the involvement of the citizen in the e-Governance process. In order to enable a right participation of the citizens to an administrative procedure of interest, their correct and accurate positioning within the reference legal framework is needed.

As a reference application scenario we consider citizens accessing a Web portal (like [9]) to find out useful information concerning their status, including the text of norms relevant for specific circumstances concerning their lives (e.g. unemployment, divorce, minorities' membership etc.). To this purpose, we want to take into account the citizen's digital identity and, with the help of a computer system, exploit its online availability to access the suitable norm versions which are applicable to the citizen's case.

The existence of a technological infrastructure which would automatically enable this kind of service, demonstrating its feasibility and usefulness, would constitute a good starting point for the possible removal of all the hindrances (starting from legal and bureaucratic ones) which prevent such services from being implemented in the e-Governance practice. At least, just to put a foot in the door. This is our commitment for

---

[3] In order to avoid unauthorized accesses to protected information, we assume citizens can be uniquely and trustworthy recognized on the Web (e.g. through an electronic ID card or digital signature) so that they can be always granted the privileges to read all and only their data from the various PA information servers. In this way, the required portions of their digital identity can effectively be reconstructed on demand via the activation of the appropriate identification services.

the research project described here, which will be carried out during this and in the next year.

References

1. F. Baader, I. Horrocks, U. Sattler, "Description Logics for the Semantic Web", *Künstliche Intelligenz* 16(4):57-59, 2002.
2. T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", *Scientific American* 284(5):34-43, 2001.
3. F. Grandi, F. Mandreoli, P. Tiberio, M. Bergonzini, "A Temporal Data Model and Management System for Normative Texts in XML Format", Proc. *Intl' Workshop on Web Information and Data Management (WIDM 2003)* (in conj. with CIKM 2003), New Orleans (November 2003), LA, ACM Press, New York, pp. 29–38, 2003.
4. Nicola Guarino (Ed.), *Formal Ontology in Information Systems*, IOS Press, Amsterdam, 1998.
5. V. Haarslev, R. Möller, "RACER System Description", Proc. *Intl' Joint Conf. on Automated Reasoning (IJCAR 2001)*, Siena (June 2001), Italy, LNCS 2083, Springer, 701-705.
6. Ian Horrocks, Peter F. Patel-Schneider, "Reducing OWL Entailment to Description Logic Satisfiability", Proc. *Intl' Semantic Web Conf. (ISWC 2003)*, Sanibel Island (October 2003), FL, LNCS 2870, Springer, pp. 17–29, 2003.
7. M. Palmirani, R. Brighi, "Norma-System: A Legal Document System for Managing Consolidated Acts", Proc. *Intl' Conf. on Database and Expert Systems Applications - DEXA 2002*, Aix-en-Provence (September 2002), France, 310-320.
8. S. Rodotà, Introduction to the "One World, One Privacy" Session, 23rd International Conference of Data Protection Commissioners, Paris (Sept. 2001), France, `http://www.paris-conference-2001.org/eng/contribution/rodota_contrib.pdf`
9. The Italian e-Government Portal, `http://www.italia.gov.it/`
10. The "Norme In Rete" (Norms on Net) Home Page, `http://www.normeinrete.it/`
11. The OWL Plugin for Protégé Home Page, Stanford University, `http://protege.stanford.edu/plugins/owl/`
12. The Web Ontology Group Home Page, W3C Consortium, `http://www.w3.org/2001/sw/WebOnt/`
13. The Web Services Activity Home Page, W3C Consortium, `http://www.w3.org/2000/xp/Group/`