A Semantical Reasoning Framework for eGovernment of the French Social Welfare System

Comte Frédéric and **Leclère Michel**
Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier,
LIRMM 161 rue Ada,
34392 Montpellier cedex 5 - France
{frcomte,leclere}@lirmm.fr

eGovernment and the French social welfare system

The french social welfare system is managed by national institutes. They are in charge of managing contributions from active people. These contributions are subsequently used to provide settlements to retired people. Nowadays there are around 200 institutes in charge of 27 million working people and 13 million retired people.

The IRCEM group is one of the major institutes of the French national welfare system which manages family household service professions: 10% of French citizens (around 6 million) are directly concerned by IRCEM activities. The IRCEM group has an information and communication department that runs the group's Information System (IS). The priorities of this department are: service quality (particulary eServices) and efficient data sharing with its partners. In this context, we are investigating semantical technologies with the aim of improving the quality of provided services.

For 2 years the number of eGovernment services in France is increasing thanks to a governmental agency called ADAE (now DGME). Through the ADELE[1] project, the French government promotes IS development in order to modernize French administration functioning, to help administrations identify supply and demand in the area of eGovernment, and to provide technical guidelines to enhance administration IS interoperability. Technical guidelines of the ADELE project do not take into consideration semantical web technologies such as RDF (G. Klyne 2004) or OWL (Dean & Schreiber 2004). The IRCEM group and LIRMM laboratory are working together to come up with semantical solutions adapted to IRCEM's needs.

Our first work was to analyse the eGovernment problems of the IRCEM group. This led us to study the context of development of French electronic administration projects. Based on these observations, we present the framework that we are currently developing to take into account semantical needs of the IRCEM group in this context.

[1]cf. http://www.adae.gov.fr/adele/

eGovernment problems and semantical approach

An eGovernment process generally involves collaborative management of resources, generally a very large amount of data, and different IS able to manage data and assist users with powerful functionalities. In the following, we highlight three main problems encountered in eGovernement systems and how a semantical approach helps to solve them.

Poor document management (Klischewski 2003b). eGovernment IS are generally based on database management systems (DBMS). However, information retrieval processes cannot only involve requesting a DBMS because eGovernment processes have to deal with different kinds of electronic resources which are often not structured (sometimes in a paper-based form). Moreover, the lack of visibility concerning what information is available or not on different media, cause the administration to lose knowledge and increase forms on different media.

To solve this kind of problem, we work with IRCEM's experts to construct an *ontology of social welfare services for family household service professions*. This will make it possible to index electronic resources and represent the content of paper-based resources in a semantic annotation base. Hence, a query on this base will enable the operator to find relevant information from different media, and a way to access it.

Lack of interoperability (Wimmer & Traunmüller 2002). Different problems hamper the interoperability between eGovernment IS: a first is syntactical incompatibility due to heterogeneity in technologies used by different ISs, but technical guidelines provided by the ADELE project could lead to solutions to this problem in the near future; a second is semantical incompatibility due to the lack of semantic annotation; Moreover even when annotations exist, utilisation of different metadata (due to the fact that ADELE guidelines do not recommend any semantic technologies) is a barrier to the information retrieval process (Klischewski 2003a).

As RDF and OWL are, respectively, the resource annotation language and the ontology representation language recommended by W3C and they are widespread on the web, we will use them to annote IRCEM resources: eServices, forms, documentation, etc. We will also use them to represent the ontology. Utilisation of these standards provides syntactical interoperability but also semantical interoperability thanks to the formal semantics of these languages defined in (Hayes 2004) and (Patel-Schneider, Hayes, & Horrocks 2004).

Absence of intelligent mechanisms. Passive management of eGovernment data is not sufficient if we consider the social range of an eGovernment process. If a user submits a request about law texts, answers from the system have to be relevant and complete: the process has to guarantee that answers are correct and that no other relevant answer has been forgotten (for important requests a forgotten resource can lead to a very bad legal situation). Besides, an eGovernment IS contains explicit but also implicit knowledge. Implicit knowledge must not be lost but could also be utilized: a naive example to illustrate this problem is an IS which can provide the date of birth of a person but cannot answer a request about the age.

We consider that inferential mechanisms are essential for eGovernment services and should be characterized formally in order to guarantee the *operational quality of services* to users. To provide these guarantees, we use an Artificial Intelligence Knowledge Representation (AI KR) formalism called the $\mathcal{SG}$ *family* (Baget & Mugnier 2001), which is a subformalism of Conceptual Graphs (CG) (Sowa 1976), logically founded (reasonings being sound and complete w.r.t. First Order Logic (FOL) semantics).

Semantical Reasoning Framework

Due to the complexity and distributivity of an eGovernment process, it cannot be considered that all information is known. For this reason, the closed world assumption of classical databases would be too restrictive for an eGovernment service; we have thus chosen to work under the open world assumption. The architecture of our semantical framework is based on a resource-metadata-ontology paradigm: each IRCEM resource (and those from their collaborators) will be identified by a unique name and characterised by its metadata. Metadata vocabulary comes from the ontology. This ontology is a kind of heavyweight ontology allowing complex representations like type definitions or axioms used in an inferential mechanism or in a validation mechanism.

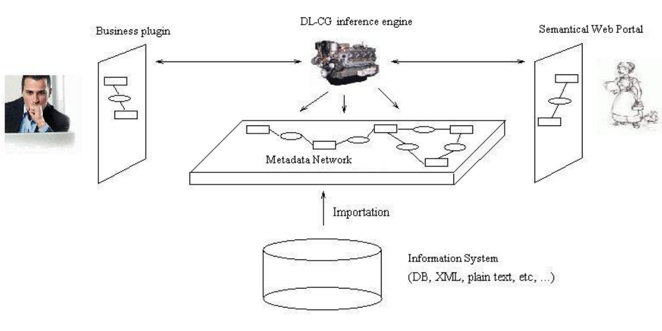

Figure 1: Semantical reasoning framework

The architecture of the semantical reasoning framework is mainly composed of (see Figure 1):

- a *semantical web portal (SWP)* which allows end-users to acces eServices. In (Comte 2004), we describe SWP as a common web interface that is able to: (1) display semantic networks representing metadata, (2) provide a graphical way to specify requests, and (3) provide direct acces to resources (electronic). SWP will allow IRCEM subscribers to retrieve specific eServices (e.g online subscription), specialized law texts, official forms, etc;
- a *business plugin* providing more extended functionalities than SWP for IRCEM employees. This plugin will be integrated in existing IRCEM IS;
- a set of *protocols and processes* able to
  1. import RDF and OWL data into our AI KR formalism; for reasonings purposes this importation has to preserve semantics (Baget 2005),
  2. link different components of the architecture,
  3. share data with different external partners;
- A *Logic kernel* composed of the domain ontology, the metadata network and the inference engine.

Inference engine

An inference engine is always dedicated to a particular AI KR formalism. There are three main criteria to characterize a KR formalism: *epistemological* criterion, which represents the expressivity of the formalism, *computationnal efficiency*, and *convenience* criterion (also called *human window* or *conciseness*), which represents the facility to describe something in a KR formalism.

On one hand, Description Logic (DL) (Baader & Nutt 2003), very widespread in the semantic web community, meets the first two criteria but not the convenience criterion. From a computational viewpoint, these DL systems are well designed for terminological inferences, like classification or realization.

On the other hand, the $\mathcal{SG}$ *family* meets the convenience criterion because knowledge is represented graphically and reasonings are graphically performed. This formalism is well adapted to assertional knowledge (the basic formalism, i.e. *simple graphs*, has an expressivity equivalent to RDF/S formalism (Baget 2005)). In particular, it allows to represent unidentified cyclic knowledge structures which are not expressable in DL. See for instance the following example in OWL abstract syntax:

Class(houseHoldProfessional complete intersectionOf(Person restriction(workingPlace value _:x)) restriction(workFor intersectionOf(Person someValuesFrom(restriction(live value (_:x))))))

At the terminological level, the $\mathcal{SG}$ *family* has representation features that differ from these of DL. For example, DLs facilitate representation of relation cardinalities, negation or disjunction, while $\mathcal{SG}$ *family* permits to represent general axioms of ontologies that can be used as rules or constraints. Based on graph theory, the $\mathcal{SG}$ *family* benefits from a substantial amount of algorithmic results in this domain.

To be in line with previous criteria, we will develop an *heterogeneous DL-CG system*. A DL system will be used for terminological inferences, while a CG system will be used for assertional reasonings. Logic kernel offer two kinds of heterogeneous service (other inferences can be reduced to these two) for logical data management: *information retrieval* and *annotation validation*.

- *Information retrieval:* Let Q be a query graph, an information retrieval service consists of computing some answer graphs A matching Q, where each $a \in A$ is obtained by deduction from both the annotation base and the ontology. This problem states that information retrieval involves the return of annotations justifying the answers;
- *Validation:* Is an annotation consistent with a consistent ontology? This step, which is supervised by the annotator (who is the data resource designer or a knowledge engineer depending on the context), ensures that annotations are correctly using ontological vocabulary. The annotation is first satured with rules issued from ontological axioms considered as implicit (thus used in an inferential way), and then constraints issued from axioms restricting the validity of the knowledge are checked.

Use of the $\mathcal{SG}$ *family* in a logical kernel allows us to import some results about deduction, information retrieval and consistency checking services. $\mathcal{SG}$ deduction with type definitions was shown to be sound and complete w.r.t. FOL semantics in (Leclère 1997). Consistency checking of a knowledge base in this formalism, with finite expansion rule set, was shown to be decidable in (Baget & Mugnier 2002). These results will allow us to guarantee the *operational quality of service* for the semantical reasoning framework.

A Logical kernel is currently implemented as an heterogenous *DL-CG* system. The CG system is based on CoGiTaNT (Salvat & Genest 1998). We plan to associate it with a DL system such as RACER (Haarslev & Möller 2001) or FaCT (Horrocks 1998). SWP and business plugin will be implemented with classical technologies and extended with a version of CoGui, which is a graphical interface for CoGiTaNT.

Further works

User's could legitimately have high expectations concerning eGovernment process quality because these processes deal with personal data. The ADELE project and IRCEM eGovernment problems have led us to investigate, and define, a kind of guarantee for eGovernment services that we call *operational quality of service*. This guarantee will contribute to an ADELE imperative, to: "set up a pact of trust with French people".

In further works, we will have to investigate potential erroneous annotations: a *reparation service* will be useful to repair these annotations in a semi-automatic way (collaboration with a supervisor). Finally, through user tests, we also plan to compare the convenience of keyword/formulary requests and semantic network requests.

References

Baader, F., and Nutt, W. 2003. *The Description Logic Handbook*. Cambridge press. chapter 2, Basic Description Logics.

Baget, J.-F., and Mugnier, M.-L. 2001. The $\mathcal{SG}$ Family: Extensions of Simple Conceptual Graphs. In *Proc. IJCAI'01*, volume 1, 205–210.

Baget, J.-F., and Mugnier, M.-L. 2002. The complexity of rules and constraints. *JAIR* 16:425–465. http://www.jair.org/abstracts/baget02a.html.

Baget, J. F. 2005. Rdf entailment as a graph homomorphism. In *ISWC*, volume 3729 of *LNCS*, 82–96.

Comte, F. 2004. Un portail sémantique pour la e-administration. Groupe des Ecoles de Télécommunications. GET.

Dean, M., and Schreiber, G. 2004. Owl web ontology reference. Technical report, W3C.

G. Klyne, J. C. 2004. Resource description framework (rdf): Concepts and abstract syntax. Recommandation, W3C.

Haarslev, V., and Möller, R. 2001. RACER system description. *Lecture Notes in Computer Science* 2083.

Hayes, P. 2004. Rdf semantics. Recommandation, W3C.

Horrocks, I. 1998. Using an expressive description logic: FaCT or fiction? 636–647.

Klischewski, R. 2003a. Semantic web for e-government. In *EGOV 2003*, 288–295. Springer Verlag Berlin.

Klischewski, R. 2003b. Towards an ontology for e-document management in public administration. In *CACM45*, 48–54.

Leclère, M. 1997. Reasoning with type definitions. In *Proc. ICCS'97*, volume 1257 of *LNAI*, 401–415. Springer.

Patel-Schneider, P.; Hayes, P.; and Horrocks, I. 2004. Owl web ontology language semantics and abstract syntax. Recommandation, W3C.

Salvat, E., and Genest, D. 1998. A plateform allowing typed nested graph: how cogito became cogitant. In *springer verlag LNAI*, volume 1453, 154–161.

Sowa, J. F. 1976. Conceptual Graphs. *IBM Journal of Research and Development*.

Wimmer, M., and Traunmüller, R. 2002. Towards an integrated platform for online one-stop government. In *ERCIM news*, volume 48, 14–15.