Semantic Application Integration by Learning Ontologies

Bojan Cestnik, Stojan Košti

Abstract: *Integrating computer applications is as a difficult and intricate task. The integration can be performed on various levels, starting from simpler data exchange level, to more complex and knowledge intensive levels. The paper presents the role of ontologies as metadata in integrating computer applications on higher semantic levels. We present our experience in automating information exchange between several actors, where application interoperability is one of the most important critical success factors that contributes to the overall impact of the ICT support. We illustrate the approach on the state project called National Housing Savings Schema that implements G2B cooperation between the Housing Fund and several business banks.*

Key words: *Application integration, Ontologies, Knowledge representation, Interoperability.*

INTRODUCTION

When integrating applications, ontologies often play a substantial role in logical understanding of the problem domain [6]. Ontologies have for long been used as semantic models for structured presentation of concepts and their relations that facilitate the exchange of knowledge between humans and machines. Accepting ontologies as such helps solving the information integration problem by turning different process and data models into views that are understandable to both end-users and application integration designers. Proper understanding of application semantics is, therefore, a common prerequisite for successful application integration.

In science, Ontologies have been used for decades to systematize scientific information in the form of a common vocabulary. Ontologies as such can be regarded as supportive means that facilitate reasoning and information analysis for a given problem domain. Therefore, they are mostly used for formally representing domain knowledge. They typically contain objects, concepts, properties and relations between objects. Moreover, ontologies can serve as useful generalizations that make the problem domain more comprehensible to the concerned stakeholders.

In this paper we address the following question: can we use existing semi-automatic ontology construction tools to speed up the traditionally manual process of application integration. Traditional way of manual construction is time consuming and typically requires several phases to complete. As a case study for the application integration using ontology as metadata we took the National Housing Savings Schema (NHSS) domain, which has some typical characteristics that justify the implementation of e-government integration principles [1]. First, NHSS is a state project governed by the legal Act. Second, there are several actors involved, including the Ministry, the Housing Fund, several business banks, and citizens. And third, NHSS offers an incentive financial instrument to help citizens acquire suitable housing facilities. Therefore, the citizens that are saving their money in NHSS are entitled to gain access to exact and up-to-date information about the status of their savings, while the other three cooperating actors have to provide services for accessing such information. Although providing such services to citizens seem relatively simple at first sight, various roles of the involved actors are causing increasing complexity that calls for an efficient implementation of integration and interoperability principles [1].

The main purpose of this article is to provide information technology viewpoint on application integration based on constructed ontologies. We primarily focus on integration and interoperability issues related to exchanging information between the government institutions and business banks included in NHSS project with a goal to speed up and add value to the business process. First, we describe the context and the underlying business process. Then, we elaborate on some of the most crucial phases of application integration with other actors in the process, discussing the issues related to

the information technology. Next, we present a few topics concerning the actual implementation of supporting the underlying business process. Finally, we conclude by pointing out the most important findings in the paper.

RELATED WORK

Traditionally, ontology construction is a manual task that consists of determining interesting domain concepts and establishing a hierarchy of such concepts. The process uses a special sort of description language, in which common domain knowledge is represented. In the last decade several computer programs have been developed that support and speed–up such manual ontology construction, for example Protégé [4]. To demonstrate the program's ability we took the legal act governing the NHSS domain as a case to construct the domain ontology. For the task we employed an expert from the field and provide her with the support of using Protégé shown in figure 1. The concepts from the ontology typically form the core structure for exchanging information in application integration with other actors. This step is crucial in achieving conceptual understanding necessary for successful cooperation between the responsible developer teams.

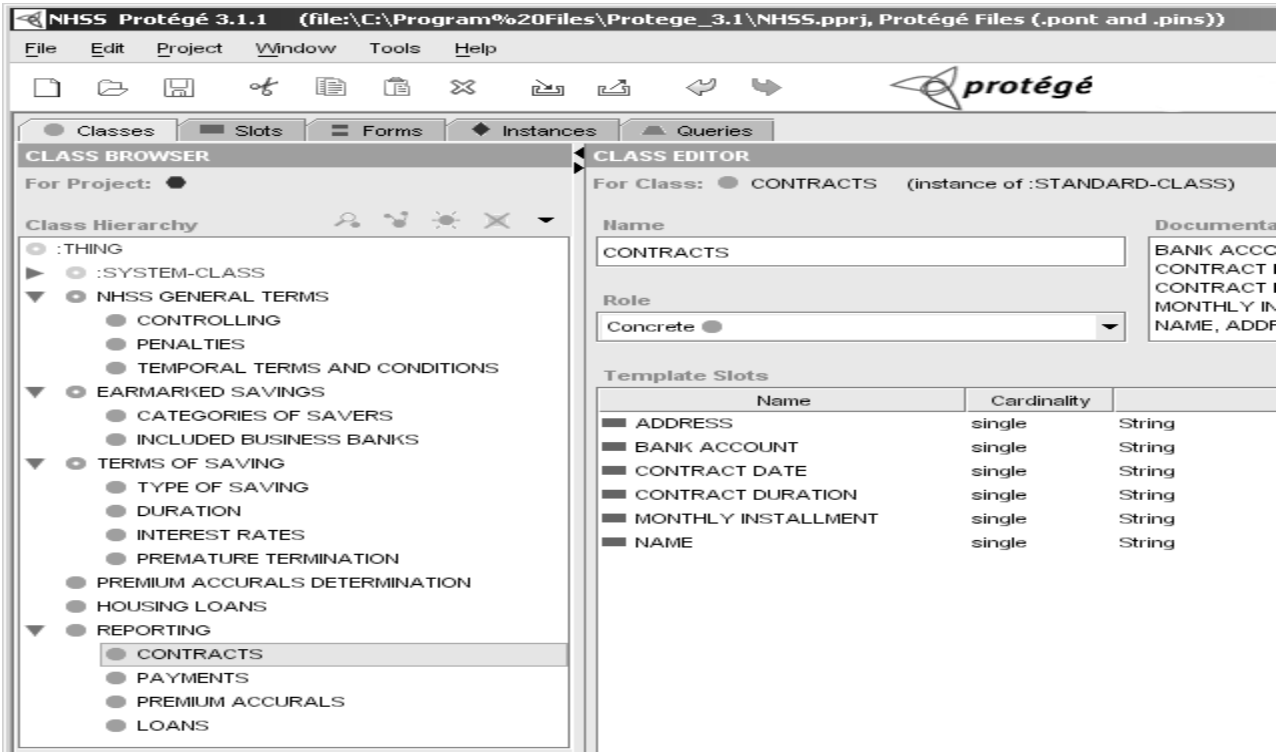

Fig.1: Manually constructed ontology by the legal expert in Protégé.

Ontology construction has taken part in many important on-going e-government projects that deal with knowledge management and ontologies. One example is EU project named QUALEG (Quality of Service and Legitimacy in e-Government) [7]. Its aim is to enable local governments of France, Poland and Germany to manage their policies more transparent and enable adaptability of the proposed solutions. The project showed many advantages of using technology based on ontology, such as preventing redundancy of data representation, enabling adaptability and supporting the realization of the importance ascribed by the local language to topics through the use of multiple synonyms.

The similar EU funded project that deals with this context is the LEX-IS project. Its main objective is to improve the legislation process in the National Parliaments through enhancing public participation with the use of technology-based tools like ontologies.

Developing ontologies helped all of the involved parties to easily locate and interpret important information [7].

LEARNING ONTOLOGY FOR SEMANTIC APPLICATION INTEGRATION

The quantity of knowledge stored in the form of scientific publications, articles and books grows exponentially in the recent years. Consequently, the interest in using text-mining technologies evidently rises, primarily because effectively following the progress in large quantity of published knowledge is almost impossible even in a relatively narrow field of interest. Written text typically contains large quantity of information that is usually encoded in a way hard to comprehend for automatic approaches. In spite of that fact, information technology spawned, besides the area of text mining, several important application areas like information retrieval, computational linguistics, text categorization, ontology learning and hypothesis formation.

With the emergence of new text mining technologies ontologies can be constructed semi-automatically by processing available text documents [e.g. 2]. Since manual ontology construction is a complex and intricate process that requires both skill and diligence, the need for developing more active and helpful computerized support is evident. In the last decade several approaches for facilitating semi-automatic construction of ontologies have been developed and successfully used in practice, making the process of ontology construction more effective and viable. One example of a tool for interactive construction of ontologies from text documents is OntoGen [3], which has already been proven successful in several real-world applications. A user can form concepts, edit them thematically and assign documents to the formed concepts. By implementing modern machine learning techniques OntoGen can help users in all crucial phases of ontology construction, suggesting concepts and their names and automatically assigning documents to the proposed concepts [3].

An example of ontology constructed with OntoGen on the legal act covering the NHSS is presented in Figure 2. The constructed top-level ontology gives an insight into the structure of the studied domain; it is, therefore, particularly useful in the process of obtaining initial acquaintance with the domain. Note that it provides some sort of "birds-eye view" on the domain under study. It is worth mentioning that the constructed ontology was meaningful and correct to the legal expert from the NHSS field; besides, the top-level structure is very similar to the manually constructed ontology in Protégé from figure 1.

Besides the mentioned activities the Housing Fund managed to improve also the transparency of the NHSS business process. There are three different actors involved in this process: the Fund, business banks and citizens. The Fund's role is central, since it carries out the process and it is responsible for direct communication with the other two actors. One of the banks is engaged in all eight public tenders and concludes almost half of all saving contracts. This bank's group is one of the largest state international finance groups with 1.500 correspondence connections in 140 states. Requirements for saving in the NHSS before year 2006 were very strict and savers had to pay their monthly fee strictly within certain month. In case they forgot payment for one month it had to me carried in next one such as current one. In case saver had forgotten his payment more than once in whole period of saving, he lost all benefits such as premiums and ability to get long-term loan. To avoid this unpleasant situation, the Fund improved communication with savers in a way saver can look at situation on his saving contract in every moment. This is possible with link, which is put on the Fund's web page and updated monthly. Only thing the saver need is the username and password. Last year the Fund and the bank implemented a step further by employing Web Service technology to enable interoperability of their information systems. This enables savers to see updated situation on their saving contracts.

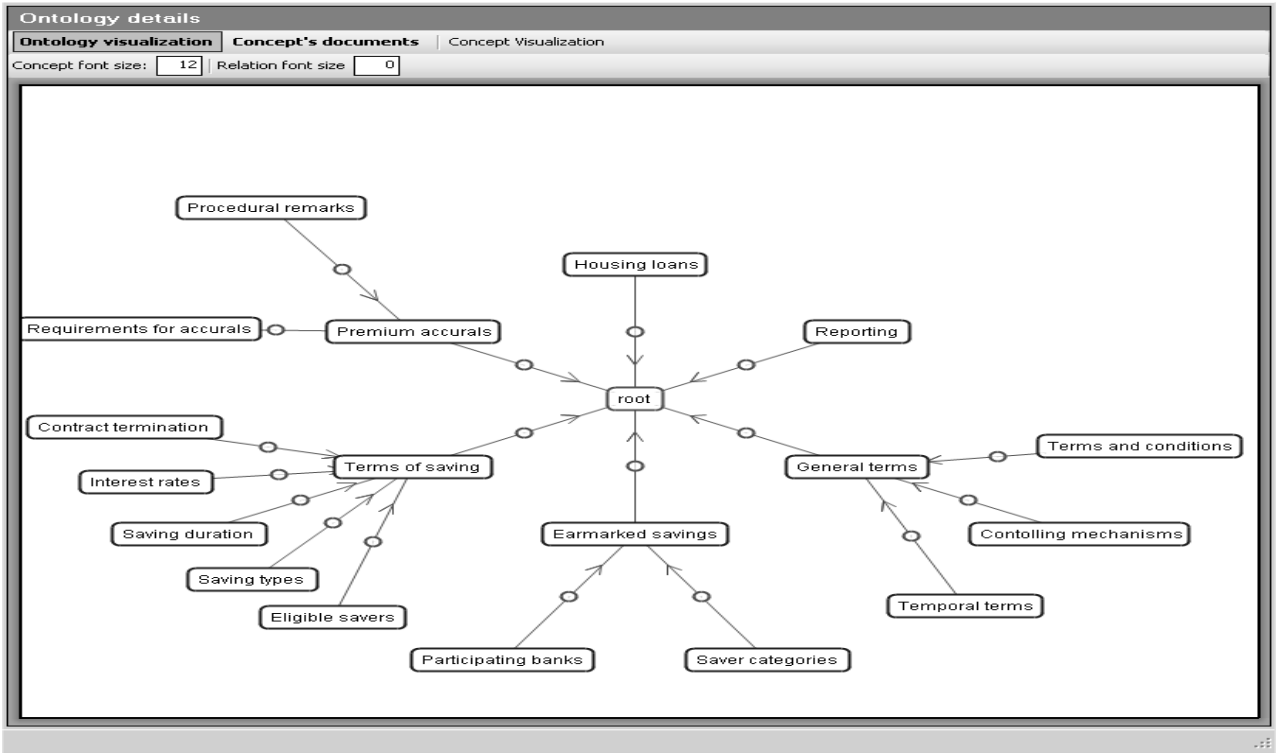

Fig. 2: Semi-automatically constructed ontology with *OntoGen* from the underlying legal act.

By definition, interoperability is the ability of two or more systems, or components to exchange information and to use the information that has been exchanged. With respect to software, the term interoperability is used to describe the capability of different programs to exchange data via a common set of business procedures, and to read and write the same file formats and use the same protocols. In IT we can distinguish between several types of interoperability that can be ranked hierarchically from application interoperability, through semantic and enterprise interoperability to environment interoperability. Most of the up-to-date e-Government interoperable solutions seem to function solely on the application level; moreover, even on this level they function in a rather fragmented manner. However, we believe that one of the most important challenges in the next decade will be shifting the interoperability paradigm to the level of semantic and enterprise interoperability, which will, among other things, in the long run rely heavily on already developed semantic web, metadata and knowledge management technologies.

The Fund and the banks carry on the majority of interdependent work to support the NHSS business process implementation. At first sight banks could deal with the task alone since it mostly falls in the category of their regular business. However, after each completed year the savers are entitled to receive a state premium accrual if they had paid their instalments regularly and according to the established plan. Since the Fund is responsible for validating savers compliance to the plan, it needs to acquire data about the contracts and paid instalments. To actively keep an eye on the things, the Fund issued a protocol that governs the necessary data exchange between the banks and the Fund. The quantity of the exchanged data is relatively high; however, the frequency is quite moderate since the exchange has to take place only once or twice every month. The exchange protocol is robust enough, since it relies on the ontology build for NHSS domain. The ontology is considered as a basis for completely automated interoperability solution using message queuing technology.

Since the Fund decided to maintain its own database for monitoring saving accounts, it was a natural move to make it available to interested banks and citizens

through a web application. In such way the banks are able to use the Fund's database for advising citizens and to proactively handle potential problems like missing instalments etc. The web application takes into account the sensitivity of the data so it allow secure internet access HTTPS for both, banks and citizens. Here, the citizens can observe the status of their saving account, which gives them additional sense of control over the saving maters. The introduction of the Fund's G2C service had also quite surprising side effect: several banks decided to offer additional B2C service to the same citizens, thereby narrowing the time and information gap by providing even more accurate data on the Internet.

After establishing NHSS project, the Fund issued detailed requirements for data exchange with the corresponding banks (Figure 3). These requirements are based on the constructed ontology; thorough understanding of application semantics is a common prerequisite for successful application integration. At first, the amount of data was relatively small and the exchange was technically carried out using simple storage media as well as the Internet services. Here, data protection and privacy issues had to be considered as well.

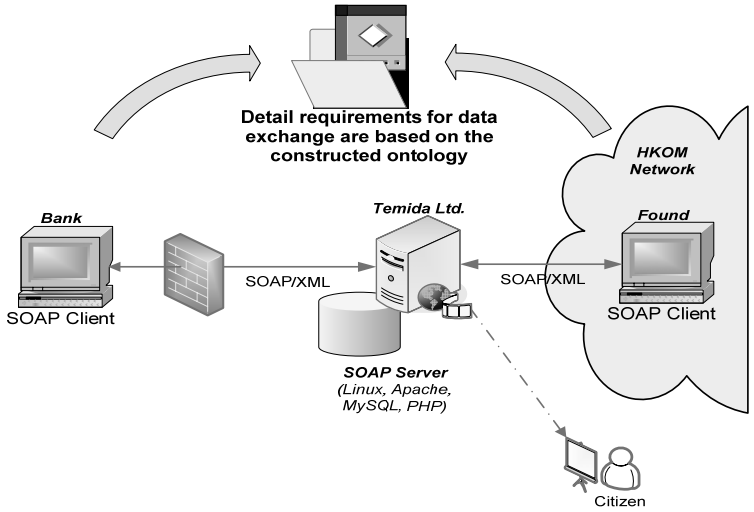

Fig.3: Architecture of solution based on constructed ontology

As always in the large projects like NHSS, change management procedures are mandatory but difficult because of a large number of stakeholders that contribute to the project. A thorough understanding of the problem domain is a necessary prerequisite for successful implementation of change management procedures. Here, the constructed ontology might come in handy. Note also the inherent rigidity towards change of large banks involved in the NHSS project. Our advice, which turned out to be very beneficial in practice, is that all the changes in the communication protocol be backward compatible. In such way, it is not necessary for all the participating actors to switch to the new specifications at the same instant. In NHSS project, different banks upgraded their corresponding software in a few months period following the new requirements document. By obeying the backward compatibility principle, the application at the Fund was able to handle all essential communication, even though the relevant applications in banks were using different versions of the protocol specifications.

CONCLUSIONS AND FUTURE WORK

In the paper we presented the use of a tool OntoGen [3] for the construction of ontology based on a selected legal act. We demonstrated how such technology could help us obtain an overview of the observed domain. The semi-automatically constructed ontology is very similar to manually constructed ontology by the legal expert, although the former required up to 10 times less time for construction. In addition to speed of the

process, the value added to the main an sub-processes is that the process management is even easier and more effective. As a result we can notably speed-up the process of "getting acquainted" with a given problem domain, mostly because the generated birds-eye view quickly reveals the most important top-level domain concepts. We demonstrated that the resulting ontology could serve as a firm basis for a common understanding between several stakeholders as well as for integrating other applications in an interoperable manner.

The constructed domain ontology can be used to support both, process view and data view [5]. The process view represents information system dynamics and is, therefore, crucial for establishing communication protocols for application integration. Common understanding of underlying business processes is crucial for establishing responsibility matrix for handling events generated by the integrated applications. On the other hand, data view incorporates static structure for exchanging information. As a consequence, it can serve as a backbone in the process of designing the integration.

Semantic application integration can be used to achieve application interoperability; therefore, modern technologies like Web services and Service Oriented Architecture have to be considered in the process. Moreover, the new trend of supporting ontologies as underlying data models for machine data interpretation plays a substantial role in the next generation of Web services – Semantic Web services. It is our firm belief that this kind of built-in knowledge will lead to a significant time and money savings in further development, deployment and integration of applications.

REFERENCES

[1] Cestnik, B., A. Kern. IT for supporting the Slovenian National Housing Savings Schema: example in G2B interoperability. 5th Eastern European e|Gov days, Prague, April 2007.

[2] Cestnik B., A. Kern, H. Modrijan. Semi-automatic Ontology Construction for Improving Comprehension of Legal Documents. Electronic Government, 7th International Conference, EGOV 2008, Turin, Italy, August 31 - September 5, 2008. Proceedings, Lecture Notes in Computer Science, Springer, pp. 328-339.

[3] Fortuna B., M. Grobelnik M., D. Mladenić. System for semi-automatic ontology construction. Demo at ESWC 2006. Budva, Črna Gora, June, 2006.

[4] Gennari J., M.A. Musen, R.W. Fergerson, W.E. Grosso, M. Crubezy, H. Eriksson, N.F. Noy, S.W. Tu. The Evolution of Protégé: An Environment for Knowledge-Based Systems Development. 2002.

[5] Kroll, P., P. Kruchten. The Rational Unified Process Made Easy, A Practitioner's Guide to the RUP. Addison-Wesley. 2003.

[6] Linthicum D.S. Next Generation Application Integration, From Simple Information to Web Services. Addison-Wesley. 2004.

[7] Sagev A., A. Gal. Putting Things in Context: A Topological Approach to Mapping Contexts to Ontologies, Journal on Data Semantics IX, pp. 113-140. 2007.

ABOUT THE AUTHORS

Assoc.Prof. Bojan Cestnik, PhD, Temida d.o.o., Ljubljana, Slovenia, Phone: +386 1236 3350, E-mail: bojan.cestnik@temida.si.

Stojan Košti, MSc, Temida d.o.o., Ljubljana, Slovenia, Phone: +386 1236 3350, E-mail: stojan.kosti@temida.si.