2009 Ninth International Conference on Intelligent Systems Design and Applications

Using Ontologies to achieve Semantic Interoperability in the Web: an approach based on the Semantic Triangle Model

Angelo Chianese, Annarita Fasolino, Vincenzo Moscato, Porfirio Tramontana
Dipartimento di Informatica e Sistemistica
University of Naples
via Claudio 21, 80125, Naples, Italy
{*angchian,vmoscato,fasolino,ptramont*}*@unina.it*

***Abstract*—In this paper, we propose an architecture that, exploiting the Semantic Web technologies, has the objective of allowing semantic interoperability among software agents in the Web. Such an architecture takes advantage by the Semantic Triangle model in which communication agents share the referents (real world objects) and not the references (mental image of a real object of the sender agent), thus ensuring an effective semantic interoperability in the information exchange process. We have carried a case study in order to assess the appropriateness and the feasibility of the process for the semantic information exchange by realizing and testing an instantiation of the related architecture.**

I. Introduction

In the Internet era, large-scale computer networks and the pervasive World Wide Web infrastructure have largely solved the problem of providing ubiquitous access to any kind of information. However, while information can now be easily retrieved and accessed, the problem of processing and interpreting its meaning by automatic approaches has not yet been solved adequately and it remains an important research topic.

In such a context the *Semantic Interoperability* is often referred to the ability of two or more computer systems of exchanging information and having the meaning of sent information automatically and correctly interpreted by the receiving system. As a consequence, a transferred message must include, in its expressive form, all the information required by the receiving system to interpret its meaning correctly whatever the algorithms used by the receiving systems (which may be unknown to the sending system). Presently, the challenge of realising semantic interoperability among systems cannot be completely met, since current document management applications have limited capabilities for structuring and interpreting documents.

With the advent of the *Semantic Web* [2], approaches exploiting metadata or based on semantic annotations from shared ontologies [1] provide the mostly used solution to the problem of extracting some kind of knowledge from documents. The result is a number of Web resources with machine interpretable mark-up that can be easily managed by software agents. The introduction of a shared ontology solves the problem of a unique interpretation of the meaning, but it still does not keep the "subjectivity" of data sources in the communication process. Moreover, the use of shared ontologies is affected by another fundamental problem, i.e. their *maintenance*: maintaining an ontology requires trusted centralized entities that periodically up-to-date concepts from the ontology. Such a problem has limited the use of ontologies to restricted and well-defined application domains. A further problem with software infrastructures based on shared ontologies is the scalability one that is a well-known open issue of the Semantic Web.

This paper presents a novel communication approach and a possible implementation of it that, exploiting the Semantic Web technologies, allows semantic interoperability among software agents in the Web, preserving not only the semantics but also the subjectivity of the agent's world vision in the communication. The approach is inspired to the *Semantic Triangle* communication model presented by Odgen and Richards [3], where communication agents share the *referents* (real world objects) and not the *references* (mental image or impression about a real object from the sender agent point of view), thus ensuring an effective semantic interoperability in the information exchange process.

For evaluating strengths and weaknesses of the proposed communication model, an instantiation of the communication process based on *semantic machines* deployed on a distributed software architecture has been developed and a case study has been reported.

II. The Semantic Triangle Model

Odgen and Richards [3] proposed a model of communication between agents that characterizes each message by three distinct entities, i.e., mental thought (*reference*), sign (*symbol*), and real object (*referent*). A reference and a symbol are linked by a symbolization relationship, reference and referent are linked by a reference relationship, while symbol and referent are not linked by any direct relationship. The relationship between symbol and referent is mediated by the subjective mind of the person who encodes or decodes the message, and thus this relationship is variable and

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.82

subjective. These relationships were modeled by Odgen and Richards by the so-called *Semantic Triangle Model*.

For better understanding the introduced concepts at the base of our work, the current section will be concluded with a simple example explaining the fundamentals of communication in the semantic triangle model.

Let us suppose that Bob asks his friend John if he still has the mouse (computer device) that yesterday he gave him, by composing the message: "Do you have a mouse?". Because the "mouse" word (symbol) has two meanings related to two different classes of real objects (referents) - i.e. computer device or animal -, John may not understand the exact sense of the word. Moreover, even if John understands the right sense of "mouse", he has to remember that Bob does not refer to a generic mouse, but to the particular one that yesterday he gave him. Thus, the correct communication outcome depends on the image that the "mouse" word creates in John's mind (reference). In order to preserve the semantics, Bob has to attach useful information (reference) to the message - e.g. by composing the message "Do you have the mouse that yesterday I gave you?" - that permits John to understand what is the real object (referent) that Bob asked him, although John has a different personal concept of "mouse". In fact, the reference "that yesterday I gave you" can be seen as a sort of pointer to a local but sharable concept of Bob's memory.

III. The semantic interoperability in the perspective of the Semantic Triangle Model

Different architectural patterns for achieving semantic interoperability in distributed environments have been proposed [10], [11] and the *point-to-point semantic integration pattern*, and the *Semantic Web pattern* could be considered the most diffused approaches to allow interoperability.

In the point-to-point semantic interoperability pattern, the communication is based on messages which directly embed a complete description of their semantics. This aspect implies that the message minimum semantic unit includes both symbols and descriptions of referents, which instances could be maintained in sender local knowledge sources.

This approach represents a *subjective* communication approach where the subjective vision of the sender knowledge is preserved, but *information redundancy* and *incoherency* problems may arise.

Vice-versa in the Semantic Web, based on the *Semantic Web Layer Cake* [2] model, communication relies on the use of ontologies that are conceptualizations of specific domains in order to formalize semantics of data. The Semantic Web links and relates elements of a message to a common ontology, using the *Resource Description Framework* and the *Web Ontology Language* that allow data to be shared and reused on the Web. A symbol will be directly linked to its referent by means of an ontology which provides the correct semantics of real world objects, thus solving the problem of message ambiguity. This kind of communication approach can be classified as an *objectivistic* one, where the knowledge (that is formalized by the ontology) is independent on the agents involved in the communication. This model obviously simplifies the communication problem and the implementation of systems based on such an approach, but some new problems arise: *ontology acceptance* [5], [7], *ontology building and maintenance* [6] and *ontology expressiveness*.

In order to cope with weaknesses of both communication approaches, it is possible to propose a sort of *hybrid model* that exploits the interoperability mechanisms of both the subjective and objective model. In particular, in the hybrid model each sender agent has its own subjective knowledge that may be either mapped into shared objective knowledge sources (such as an ontology), or directly included in a coded message in order to preserve its personal interpretation of transmitted concept, coherently with Odgen and Richards model.

IV. An implementation of the Semantic Triangle Communication Model in the Semantic Web

Internet and the Semantic Web offer the necessary infrastructure for implementing the hybrid communication model presented so far, and obtaining a semantic interoperability among software agents in the Web [4]. In this section, the software requirements of a possible implementation of the communication model will be presented.

A. The Message Conceptual model

Messages exchanged in the hybrid communication model are considered as the aggregation of *digital assets* (e.g., images, textual documents, audio, etc...) containing a set of *information concepts* that constitute the message semantic content that an agent is interested to transmit; these concepts refer to the agent knowledge that can be mapped either into a local ontology, or a universal ontology that is a-priori accepted by all the agents involved in the communications.

The information concepts can be classified in two distinct types: *entities* and *facts*. An *entity* is a noun, verb or other part of a speech that can be retrieved on any language dictionary. A *fact* is an expression corresponding to peoples, places, events or any other thing that could not be retrieved in a language dictionary. Facts can further be classified into: *Encyclopedic fact* and *Non encyclopedic fact*.

Encyclopedic facts have a general relevance such that they could be contained in a general encyclopedia or in a domain encyclopedia. Referents to encyclopedic facts can be called *Universal Fact Referents*, since they have to be universally shared and accepted by any communication agent (or at least, they have to be shared by communication agents belonging to a specific domain). Non-Encyclopedic facts are relevant in the internal world of the agent but could be not universally relevant or there could not be a universally accepted semantics. Referents for non-encyclopedic facts are named *Local Fact Referents* and are maintained in the Semantic Machine of the agent.

To make explicit the binding of information concepts with real word objects referents, we propose to use *semantic tags* for providing a symbolic representation of information concepts, and references for binding together semantic tags and related referents. Eventually, digital assets can be combined and associated to *information objects* (e.g. Web pages).

B. The Communication Process

The proposed communication process is supposed to be decomposed into four sequential activities:

1) *Information encoding* - a sender agent composes a message in a particular format that is understandable by other semantic machines (in this step the binding between information concepts and referents is performed using semantic tags and the reference mechanism);
2) *Information transmission* - sender agent publishes the message (e.g. by Web pages) on the Web, or transmits the message to a receiver agent (e.g. by e-mail);
3) *Information acquisition* - a receiver agent performs the message acquisition (download);
4) *Information decoding* - receiver agent binds the received information with the related referent by means of sender references.

With respect to the example discussed in section 2, the communication could be outlined as follows:

1) Bob uses an e-mail client (semantic machine) to compose the message "Do you have a mouse?" that is represented in a particular format (e.g. XML) and in which the personal concept of a"mouse" is bound to the related referent by the tag "mouse" and a reference (e.g. an hypertextual link pointing to a Web page deployed in a Web server containing the description of the mouse);
2) the e-mail is sent to the pop server of receiver;
3) the email is downloaded by John;
4) the semantic machine of John decodes the message and binds the concept of "mouse' to the corresponding referent using Bob's reference; after the communication, John may decide to add it to his local ontology for ordering a new mouse of the same kind to a vendor.

The Message Encoding and Decoding activities are specified by the UML Activity Diagrams reported in Figures 1 e 2.

The *Tag Extraction* activity performs the retrieval of information concepts from a message to be transmitted and their symbolic representation by semantic tags. The set of these semantic tags can be indicated directly by the message author or it can be retrieved automatically by apposite tag/information extraction algorithms. For each tag, the *Reference Binding* activity is performed in order to retrieve from

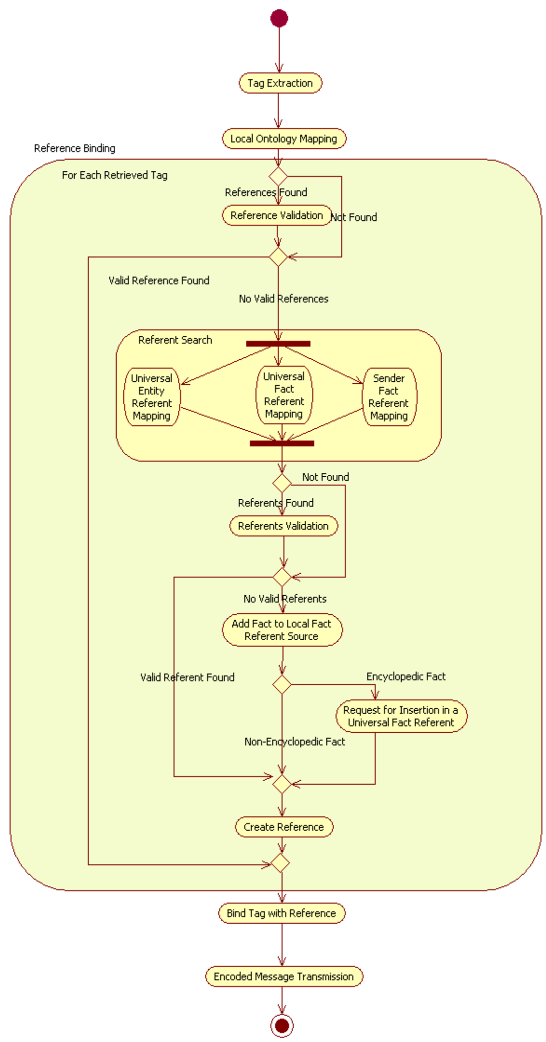

Figure 1. Message Encoding and Transmission Activity

the local ontology a reference to the information concept represented by the semantic tag. More precisely, for each tag the *Local Ontology Mapping* activity is executed, in order to search in the local ontology for possible references associated with the tag. If more than one possible reference is found, then a *Validation* activity is carried out, where a single reference correctly representing the tag has to be selected. After the eventual disambiguation, two cases may happen: (i) the reference correctly representing the tag has been found in the local ontology (*Mapping Hit*) and it is possible to bind the message tag with this reference (*Bind Tag with Reference* activity); (ii) the tag has not a corresponding reference in the local ontology (*Mapping Miss*), thus, a *Referent Search* activity is entered, where a referent reporting the semantics of the information concept

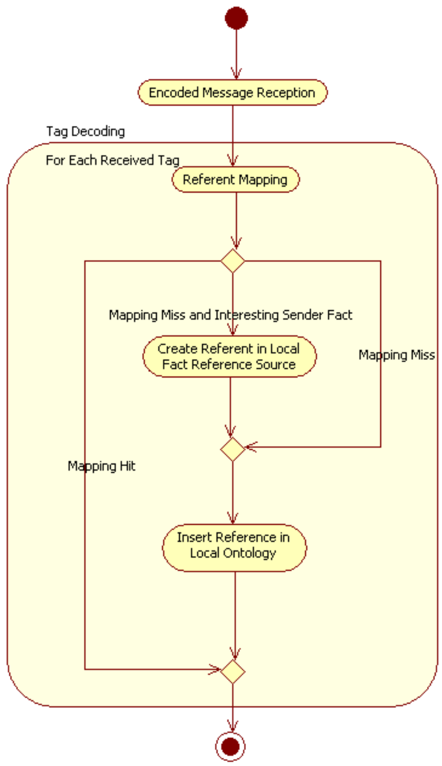

Figure 2. Message Reception and Decoding Activity

must be found in an available Referent Source, and therefore a corresponding reference (*Create Reference*) has to be built.

In the Referent Search activity, different sources can be queried depending on the type of information concept to be represented and on the available Referent Sources. Entity Referents can be retrieved in Universal Entity Referent Sources, such as *Wordnet* or other on-line dictionaries. Encyclopedic relevant facts can be retrieved in Universal Fact Referent Sources, such as *Wikipedia*. Encyclopedic relevant facts related to a specific domain can be retrieved in sources containing domain ontologies, e.g. a song may be retrieved in *last.fm* database, while a movie may be retrieved in *imdb.com* database. Non-Encyclopedic relevant facts, i.e. facts that are relevant only in the sender context, can be retrieved just in a Fact Referent Source published by the sender itself.

Facts for which the sender declares a semantics that is different, or more specific, from the one proposed by a universal fact source, can also be retrieved in the Local Fact Referent source of the sender.

In our architecture, each universal referent is a directly addressable resource. It can be accessed either by a *HTTP GET* service request, or by a Web service that performs the wrapping of the related referent source [8]. For accessing local referents (also coded in *XML/RDF*) a possible solution is offered by the *REST* technology. In this case, a referent can be directly stored into the Local Fact Referent Source of the semantic machine via a *HTTP PUT* or *POST* request, and directly accessed via a *HTTP GET* request.

At the end of the Referent Search activity, it is possible that more than one possible referent has been found. In this case, another *Validation* activity is needed, in order to state if any of the referents provide a satisfying semantics for the information concept, or to select the more suitable one, if more than one referent were found. If a satisfying referent has been found, then the *Create Reference* and *Bind Tag* activities are performed, and a new reference to the chosen referent is added to the local ontology and will be bound with the tag. Elsewhere, a new referent has to be inserted in the Local Fact Referent source. Eventually, a *Request for Insertion* activity for adding the new created referent to the Universal Fact Referent Source can be performed. This possibility is granted by some sources, such as Wikipedia, but the referent will be inserted only if it will be accepted by the referent source managers, and only in an asynchronous way.

At the end of the Reference Binding activity, the message minimum semantic unit will be composed of a set of couples $\langle tag, reference \rangle$. In our architecture it is considered as a Web resource characterized by an *URI* with naming and addressing functionalities, and will be encoded by *XML/RDF* languages. Thus, using its *URI*, each information object can be accessed by *HTTP* protocol via a *GET* request, and the result of such a request will be an *XML/RDF* representation that separates the physical view from the content resource view, thus reflecting the semantic Web layer cake schematization, and providing interoperability with applications of the Semantic Web.

As to the receiving activity, its purpose is to map each received couple $\langle tag, reference \rangle$ into any concept from the receiver agent local ontology, in order to reconstruct the correct semantics (*Referent Mapping* Activity). If a reference is found in the local ontology (*Mapping Hit*), the information concept is known to the receiver and it must not be decoded. Otherwise, if no reference is found, then a new reference has to be created and inserted in the local ontology. Moreover, if a received concept points out from the sender fact referent source, then the receiver agent can choose to accept the referent and to add it to its Local Referent Fact source.

C. The Semantic Machine

The communication process described in the previous sections can be implemented in a distributed environment including several *Semantic Agents*, each one equipped with a *Semantic Machine* that executes message sending and receiving activities.

Figure 3 shows the architecture of the Semantic Machine that provides both message sending and message receiving functionalities in the musical context. It has been developed using Java-based technologies and a brief description of the main component is reported in the following.

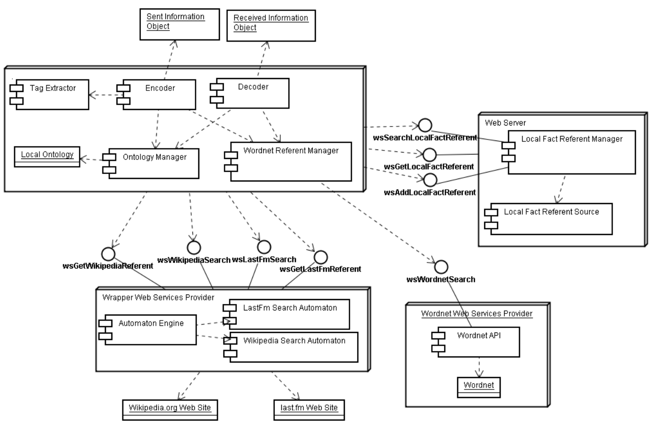

Figure 3. System Architecture

The *Encoder* is the component responsible for implementing the information encoding and transmission activities. It asks the *Ontology Manager* component for retrieving/inserting references in the local ontology and asks the *Referent Manager* component for referents in the available Entity or Fact sources, and for inserting referents into the Local Fact source. The *Decoder* is the component responsible for the information acquisition and decoding activities. It queries the Ontology Manager to retrieve/insert references in the local ontology and asks the Referent Manager for inserting new referents into the local Fact source.

The *Tag Extractor* component has been implemented using the API provided by *NLP library* of Stanford University that offers functionalities for text processing and analysis. The *Local Ontology* and the *Local Fact Referent Source* components were implemented by a relational DBMS (*MySQL*). Ontology Manager and Local Fact Referent Manager exploit the JDBC APIs for providing data access. In particular, the Ontology Manager is deployed on the same computational node of Encoder and Decoder components, while the Referent Manager is deployed on a distinct node and runs on a Web Server, being its functionalities exposed as REST resources that are reachable by the HTTP primitives.

The contents of Wordnet, that implements the *Universal Entity Referent Source*, are accessed using the WordNet API component, deployed in the context of a Web Service Application Server. The *Wordnet Referent Manager* interacts with this component to translate Wordnet lemmas into the referent format required by our model. Eventually, the selected Universal Fact Referent Sources (Wikipedia.org and LastFm.com) are accessed by Web services realised with wrapping technologies that, interacting with the Web applications, execute their functionalities, thus obtaining the requested information in the desired format. The wrapper architectural components include the Automaton Engine, the LastFm Search Automaton, and the Wikipedia Search Automaton [9].

V. A Case Study

In this section, the implementation of a semantic machine that supports the execution of the hybrid communication process in a specific context will be presented.

The considered scenario presents two agents that exchange messages concerning the italian musical festival domain, and containing information about musical events, such as program, event dates, and international and local artists taking part to the events.

For concepts related to the music domain, it is a commonly accepted opinion that the *last.fm* database (accessible via the *www.last.fm* website) is the larger and reliable source of information available on the Web. Thus, in the case study last.fm has been selected as a referent source for facts such as songs, singers, musical genres, radio, musical events and so on. For all the other facts, i.e. facts that are encyclopedically relevant but not directly related to the musical domain, the English version of the *Wikipedia* Web site has been selected as referent source.

As to the entities, the *Wordnet* repository has been considered as the unique Referent Source for entities in English language. Finally, for all the facts that are not encyclopedically relevant (i.e. for which no valid Referents can be retrieved in the considered Referent Sources) or for which the referent source is considered not reliable or not detailed enough, a Local Fact Referent Source deployed in the semantic agent was considered.

A. Working Example Execution

We assume that a semantic agent *A* wants to send to the semantic agent *B* the digital asset consisting of the following text fragment: "*Since 2000, the 'Roma Rock festival annually takes place at Ippodromo delle Capannelle in Rome during the summer period. It is a very important event in Italy for rock and pop music. It has brought international singers and musical groups, as well as young and promising artists, critics, journalist, and young people from all parts of the world to one of the most beautiful city that a person could visit, where music, art and history mix together such as in an imaginary world. In the last edition famous artists such as Duran Duran, Pino Daniele and Ben Harper and young local artists as Subsonica (rock band), Simone Crisiticchi and Fabrizio Moro (singers) played and over than 200.000 persons attended the festival*".

The editor provides the following set of tags : 'summer' - 'Ippodromo delle Capannelle' - 'Roma Rock Festival' - 'singer' - 'musical group' - 'artist' - 'critic' - 'journalist' - 'Rome' - 'Italy' - 'city' - 'music' - 'art' - 'history' - 'imaginary world' - 'festival' - 'edition' - 'event' - 'rock' - 'pop' - 'Duran Duran' - 'Pino Daniele' - 'Ben Harper' - 'Simone Cristicchi' - 'Fabrizio Moro' - 'rock band'.

During the encoding process, corresponding referents for each tag have been searched in the known referent sources, obtaining the following results:

1) added to the local fact source - 'Roma Rock Festival', 'Ippodromo delle Capannelle';
2) retrieved on wordnet and wikipedia: 'Italy', 'Rome', 'city', 'music', 'artist', 'critic', 'art', 'festival', 'event', 'edition', 'history', 'rock', 'pop', 'summer';
3) retrieved on wordnet: 'musical group', 'singer', 'rock band';
4) retrieved on wikipedia: 'imaginary world'; retrieved on wikipedia and last.fm: 'Duran Duran', 'Pino Daniele', 'Simone Cristicchi', 'Fabrizio Moro', 'Ben Harper', 'Subsonica'.

The following issues were addressed during the encoding task:

1) many tags were associated with common entities and they were retrieved both in wordnet vocabulary and in wikipedia encyclopedia. (for the disambiguation aims, the wordnet referent was preferred because of its greater trustability);
2) tags associated with singer name facts were usually retrieved either in Lastfm or in wikipedia web sites (for the disambiguation aims, Lastfm referent source was chosen because of its greater suitability in the musical domain);
3) tags that were not mapped with any referent (such as 'Roma Rock festival' and 'Ippodromo delle Capannelle') were added to the Local Fact Referent Source (in the objectivistic approach such tags should not be considered).

Some composite tags (such as 'imaginary world') were retrieved just from encyclopaedic sources. The result of the encoding process was a XML/RDF document reporting the list of tags with the corresponding URIs of chosen referents.

VI. Conclusions

We proposed an architecture that allows semantic interoperability among software agents in the Web. Our system tries to cope with some of the problems that are typical of knowledge engineering:

- the *ontology building* - our approach allows the building in an automatic way of domain ontologies by exploiting a bottom-up strategy;
- the *ontology validation* - our approach allows to validate quality of the generated ontologies.

Future works will be devoted to implement wrapping modules for other referents sources for and realize a complete experimentation in order to validate the efficiency/effectiveness and scalability of the approach with respect to other different architectural patterns.

References

[1] T. Gruber, *A translation approach to portable ontology specifications*, Knowledge Acquisition, vol. 5, n.2, pp. 199-200, 1993.

[2] T. Berners-Lee, J. Hendler, and O. Lassila, *The Semantic Web*,Scientific American, vol. 284, pp. 34-43, 2001.

[3] C. K. Ogden, and I. A. Richards, *The Meaning of Meaning*, Ed. Harcourt Brace Jovanovich, 1989.

[4] J. Hendler, *Agents and the Semantic Web*, IEEE Intelligent Systems, pp. 30-37, 2001.

[5] V. Devedzic, *Understanding ontological development*, Communicaions of the ACM, vol. 45, pp. 136-144, 2002.

[6] J. Harris, *Judging the likely Success of an Ontology*, 2004.

[7] A. Gangemi, *Ontology Design Patterns for Semantic Web Content*, ISWC 2005 (LNCS 3729), pp. 262-276, 2005.

[8] G. Canfora, A. R. Fasolino, G. Frattolillo, and P. Tramontana, *A wrapping approach for migrating legacy system interactive functionalities to Service Oriented Architectures*, Journal of Systems and Software, vol. 81, n.4, pp. 436-480. 2008.

[9] G. Di Lorenzo, A. R. Fasolino, L. Melcarne, P. Tramontana, and V. Vittorini, *Turning Web Applications into Web Services by Wrapping Techniques*, WCRE 2007, pp. 199-208, 2007.

[10] M.Y. Selvage, *Achieve semantic interoperability in a SOA - Patterns and best practices*, IBM report, 2006.

[11] D. Briola, A. Locoro, and V. Mascardi, *Ontology Agents in FIPA-compliant Platforms: a Survey and a New Proposal*, WOA 08, 2008.