Web-pages annotation and adaptability. A semantic portal on the International Space Station

Albanese C. [1], Calabrese A.[2] D'Amico A[3]., Mele F.[2], Minei G.[1], Kostakopoulos L. [1]

[1]°MARS Center, Via E. Gianturco, 31 I-80146 Napoli NA,
{minei,albanese}@marscenter.it
[2]Istituto di Cibernetica Consiglio Nazionale delle Ricerche, Via dei Campi Flegrei 34 Pozzuoli NA,
{f.mele,a.calabrese}@cib.na.cnr.it
[3] Università di Napoli Federico II Facoltà di Science

Abstract. The paper describes a methodological approach to design and develop a semantic portal by retrieving unstructured information from different Web sites on the Net. Such a method is applied to the International Space Station (ISS) knowledge domain, but it proves to be quite general and domain independent. The development of a domain specific ontology, on which the semantic portal is based, allows to annotate and classify the unstructured information available in heterogeneous formats (natural language texts, photos, videos,) on different Web sites, the annotated information becoming instances of the ontology's classes and attributes. The ontology has been developed using Ontoedit. The Portal has a client-server architecture based on the middleware layer of Ontobroker. Queries from Web pages are formulated using FLORID, a Frame Logic syntax based language. PHP language is in charge to manage user requests from the Portal Web page and the replies of Ontobroker server.

1. Introduction

In a well-known article [1], published in May of 2001 on the prestigious magazine *Scientific American*, Tim Berners Lee (one of the "inventors" of the existing World Wide Web) wrote:

"***The Semantic Web is not a separate Web*** *but an extension of the current one, in which information is given* ***well-defined meaning****, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and "understand" the data that they merely display at present*."

Starting from these considerations, we notice that the auspicated Semantic Web is not a completely new entity, but it is a particular extension of the same Web. It has, therefore, the principal objective of increasing the performances of the Net. The existing Web is already considered a very powerful tool. How can these potentialities be increased?

A way out to answer this question is to point out some existing limitations of the Web and in parallel to make evident the advantages related to the existing and future solutions proposed by the Semantic Web.

2. Limitations of the Web at the present

2.1. Heterogeneity of the metadata.

One of the main limitation of the existing web can be defined *the problem of heterogeneity of metadata*. Each user publishes on the Web representations of knowledge domains in structured or partially structured format (database, Web pages, ontologies, services etc.,), but these representations are strongly dependent on the expertise and on the perspective according to which he addresses a specific knowledge domain, so for each knowledge domains there is plenty of different representations.

This problem can be solved according to the paradigm of the Semantic Web, where mechanisms of *integration* and *mediation* are proposed. In this context the ontologies, as reference structures of a knowledge domain, play a fundamental role.

Another wide spread phenomenon in the existing Web is that of the *redundancy*: the same information is presented many times in the Net. We need, therefore, some selection mechanisms to retrieve a small and sufficient number of information. This aspect is also complicated by the fact that a lot of different digital representations exist (texts, images the video) for a same entity.

2.2. Necessity to enrich the metadata.

Another limit of the existing Web is determined by the presence of data and information that are partially and poorly structured, that is, data that have a small number of associated metadata. Also this phenomenon, as in the cases previously discussed, represents an obvious problem to access knowledge sources.

It is necessary to add additional structured metadata to information existing at the moment in the Web, in order to allow:

- to define programs that carry out integration of different metadata
- to build user interfaces that directly allow to access data in a manner more suitable to the informative needs of the user.

2.3. Personalization

In the present mechanisms of interaction between user and computer, the user carries out queries which can be ambiguous, with a consequent extension of the search space in the Web. This ambiguity is due to the fact that a generic user is not represented by a user profile and often his informative needs are not known to the program which is trying to gather the required information.

In this direction, there are an increasing number (even if at the moment small) of user interfaces which allow to differentiate the interaction on the basis of the user peculiarities.

Necessity is perceived in Human Computer Interaction, where the researchers are proposing programs in the interface where the users are represented through their knowledge and their personalities (models of users) [6]

The model user is used both to interpret the input of the user (to understand what the informative desires of the user are) both to establish the data and the modality of interaction more useful for the user.

In this work we propose a semantic portal whose interface includes a personalization of the interaction.

3. Traditional and semantic portals

A portal, in his traditional meaning, is substantially an information centralizer that contains addresses of other Web sites.

A traditional portal exposes a series of active links (interactors). They are generally present in the layouts of an interface in the form of URL addresses, sentences in natural language, buttons, etc (Arg1, Arg2, ArgN of the Figure 1). The interactors on a Web page (basic site) are put in correspondence with the information contained in other sites (referenced sites) but they give an access to a whole page of the referenced sites; in other words, they give an access to all the contents of the referenced sites and not to a specific information.

The situation is very similar to the one experienced when, during a discussion on a specific argument, a particular quotation gives reference to an entire book, without pointing to the specific page where the topic, to which the quotation is referring to, is contained.

The solution proposed by the traditional portals imposes to the users to visit the entire referenced sites and only after a complete reading it will be possible to access the relevant information.

The main characteristic of a semantic portal is that the correspondences among the interactors of the basic site (the portal) and the information of the referenced sites are in correspondence with explicit relationships (semantics).

In this work we have adopted a representation,an ontology, [3][4] as reference to connect the information contained in the basic site (in the portal) with related information contained in the referenced sites (figure 1). In the following paragraphs we will give detailed description of how the annotation and the recovery of the information are achieved in our system.

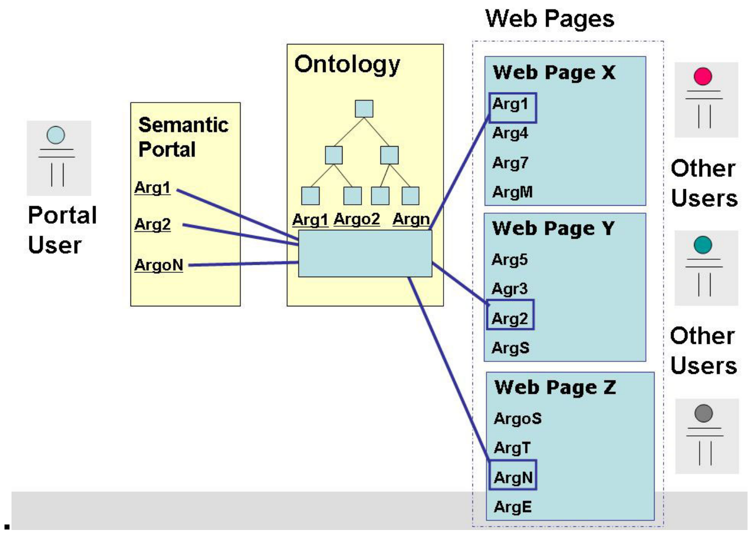

Fig. 1. Semantic Portal Structure

The characterizing aspects of a Semantic Portal are the followings:

– the terms of the Semantic Portal have the same meaning, or similar, to the terms of the sites referenced according to a reference structure (ontology);
– the Semantic Portal has direct access to a selected information contained in the referenced sites;
– a *transparent* access to the information is granted to the users, they are not compelled to explore one by one single Web sites, but, even better, during the information retrieval, they are not aware at all that it is spread through and retrieved from different Web sites.

These main features have been assumed as leading rules during the design and the development of our portal.

The fruition of contents is user oriented, the interfaces are designed in order to be adaptable to different user profiles, in order to avoid the problems related to lack of personalization, discussed above, of the existing Web.

The first step in the work has been a deep understanding of the ISS Knowledge domain, based on the analysis of ISS related Web sites and official technical documentation and on interviews with ISS specialists. A group of experts in space technologies, from MARS Center – Naples, has carried out this activity. It has been the starting point for the classification (i.e. the definition of the relevant categories and relations among them) of information, existing on different sites, and user profiles definition.

4. The Domain Representation

As a result in this first phase the ISS domain ontology was defined (see Fig.2). For this scope the chosen knowledge domain is the one related to the International Space Station (ISS).

On the basis of the accurate analysis performed, using pre-existing web resources and domain experts direct knowledge, the basic entities of our representation have been chosen, also determining their properties and relations, for the construction of the reference ontology of the Semantic Portal.

During the analysis of the ISS domain related Web Pages a set of homogeneous data (a) have been identified (e.g. information on the space experiments designed for the ISS) related to which more specific data (b) (e.g. particular experiments) have been also identified.

In the Web pages relations (c) between the homogeneous data were present, between experiments and experimental apparatuses (facilities).

In a very natural manner each (a)-type entity has been put in correspondence with the classes of the ontology, each (b)-type entity, relative to the same class, with a sub-class of it.

The (a)-type entities, from which the classes took their name, were also the main constituents of the ontology instances.

In the Web pages we found also relations between homogeneous data; these relations have been used as *non-taxonomic relations* of our ontology.

The class ISS_Domain (representing the complete ISS knowledge domain) is the root of the ontology; the whole domain has been then divided in six classes.

- ***ISS International Space Station***: it contains all the major constituents of the ISS.
- ***Experiment***: It is the class describing the experiments feasible on the ISS, including their typical duration and expected results.
- ***Launch***: It is the class describing the launches of space vehicles, with information on the launch places and dates, contributing to the build-up and utilization of the International Space Station
- ***Equipment*** : This class describes all the ISS related equipments in terms of systems and subsystems, also including the support ground infrastructure.
- ***Technical_characteristic***: It is a class containing all the technical information related to ISS domain.
- ***Purchaser***: It is the class of the nations contributing to ISS Program development

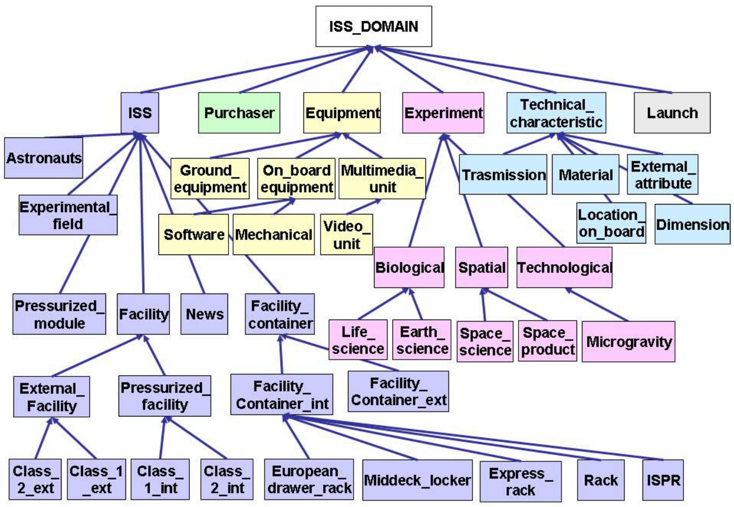

Fig. 2. The ISS domain Reference ontology

4.1. The annotation of the ISS resources.

Annotation is a general method to associate metadata to information resources; the resources which need annotation are the ones relative to un-structured or semi-structured information.

As a general example, data contained in databases are well structured, thanks to the names of the fields into database tables; texts and documents written in natural language, videos and voice streams, on the other side, are unstructured information; a web page (containing Html or XML Tags which give at least an high level structure to the relevant fields in the page) contains a semi-structured information.

Annotation is a very critical step in the development and integration of our Semantic Portal. Through the annotation process we give meaning to the superficial structures (text, videos, pictures), associating a semantics to them through the attribution of one of the categories of our ontology to each syntactic element of the representation language.

The annotation of resources gives the opportunity to get a useful enrichment in terms of metadata of the existing Web structures; it's one of the powerful tools on the way of the Semantic Web.

This process is accomplished by establishing relations among one or more elements that are present in a structure published on web (for example a web page) with a class of an ontology. Due to this annotation process the knowledge can be shared not only with humans (for whom the Existing web is designed) but also with S/W agents.

In our annotation process we adopted two principal annotation methods which use *frame* based ontologies. Two main annotation methods have been established:

- To associate an un-structured entity x present in the Web to an instance of a class A in the reference ontology (in this case x becomes an instance of the class A)
- To associate an un-structured entity x present in the Web to the value of a *Slot* of the ontology.

We can find in literature different annotation systems. The system that we used is OntoMat Annotizer [9], the reference annotation tool of Cream (Creating Relational Annotation-based Metadata) [7].

In figure 3 we show a screenshot of OntoMat, where the reference ontology developed for the ISS knowledge domain is present in the upper left corner, and a text excerpt of a Web page on the ISS is selected, the text excerpt becomes an instance of the *Facility* class in the reference ontology.

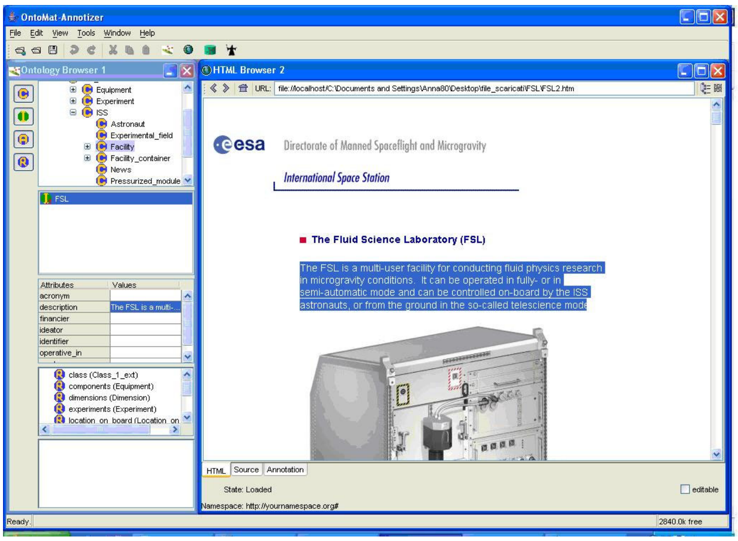

Fig. 3. A screenshot of OntoMat, where the reference ontology developed for the ISS knowledge domain is present in the upper left corner

5. The software architecture of the Semantic Portal

Ontoprise has developed a series of tools that constitute a complete environment of middleware for the Semantic Web. The semantic portal, related to the knowledge on the ISS, has been realized using many software components of such environment (figure 4).

Ontoedit [8] is a tool of the architecture of Ontoprise through which it is possible to create and edit ontologies by using graphical interfaces. OntoEdit is an interactive editor to define ontologies, describe instances, define rules etc.

Ontobroker [11] is another component of the architecture, it is an integrated system to extract and to reason about metadata of the specific domain. The information recovery in the annotated Web pages is an important function carried out by Ontobroker in the architecture of the ISS semantic portal.

Ontobroker integrates the access to different information sources like databases, keyword based search engines etc. It reads various input formats like XML, OXML, RDF(S), F-Logic [5] and Prolog. Thus it permits a homogeneous access to a not homogeneous set of information sources. Ontobroker has two principal components: Webcrawler and Inference Engine.

The Webcrawler extracts formal knowledge from HTLM pages; it picks up the existing information in the Web pages, it extracts their annotations and analyzes them in the internal representation of Ontobroker.

Florid [2], [10][11][12] (a dialect of Frame Logic) it is one of the principal representation languages of of Ontobroker. Frame Logic [5] is a deductive, object oriented database language which combines the declarative semantics and expressiveness of deductive database languages with the rich data modelling capabilities supported by the object oriented data model.

We report the basic syntax of Florid:

• *c*[] (class definition) defines a class with name *c*;
• *c*[*a*=>> {*c*1,...,*cn*}] (attribute definition) implies that the attribute *a* can be applied to the elements of *c* (it is also possible to define attributes applied to classes) and an attribute value must be member of all classes *c*1,...,*cn*;
• *c*1:: *c*2 (is-a relationship) defines *c*1 as a subclass of *c*2 which implies that all elements of *c*1 are also elements of *c*2, all attributes and their value restrictions defined for *c*2 are also define for *c*1, and multiple attribute inheritance exists, i.e. *c* :: *c*1[*a* =>> {*c*3}] and *c* :: *c*2[*a* =>> {*c*4}] implies *c*[*a* =>> {*c*3,*c*4}];
• *e* : *c* (is-element-of relationship) defines *e* as an element of the class *c*.

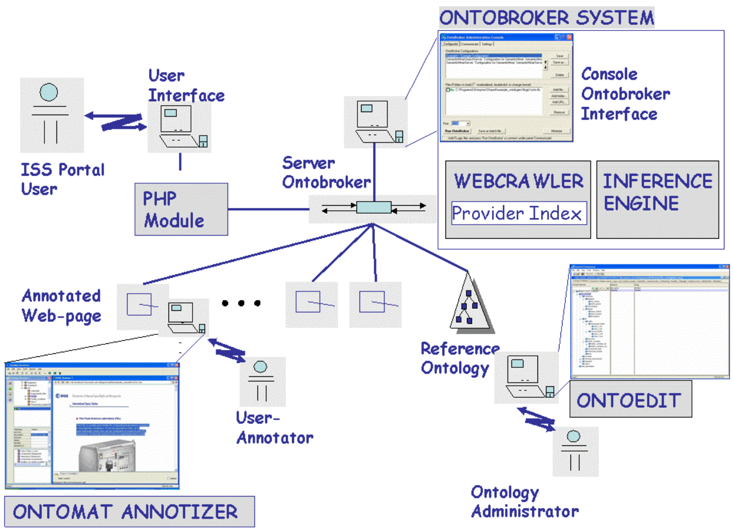

Fig. 4. The software architecture of the semantic portal

In Florid it is possible to define rules to perform inferences of the type:

- *FORALL y x[a ->> y] <- y[a ->> x].*
- *FORALL x,y x:c1[a1 ->> y] <-> y:c2[a2 ->> x].*

Ontobroker has an interactive interface (Ontobroker Console) that answers to query in Florid (figure 5).

For instance:

FORALL X, Y<- X:#News [#description->>Y] AND
EXISTS Z X:#News[#title->>Z].

This query returns the instances of the class *News* that has a value Z for the attribute *title* showing the value *Y* of the attribute *description*.

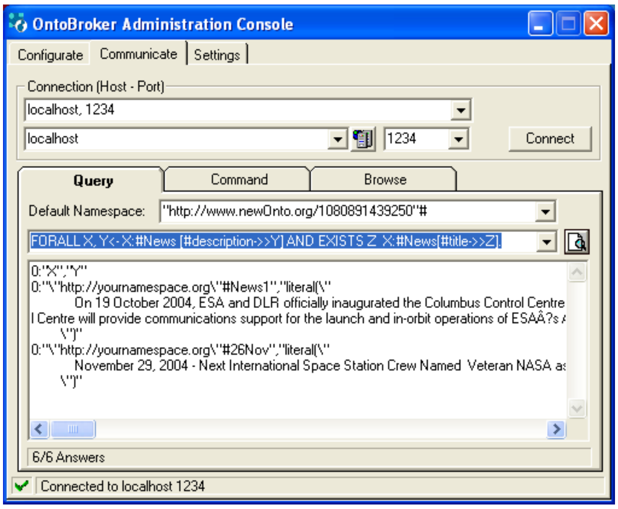

Fig. 5. Ontobroker Console

In the context of the whole architecture, Ontobroker can also play the role of server; it reads the file that contains the definition of the ontology, evaluates the queries formulated by the application program (of the semantic portal ISS) and returns the related answers.

In the realization of the portal we have used the language PHP. The software components of the portal ISS communicate with the server Ontrobroker through scripts PHP which contain the queries that will be evaluated by Ontobroker.

The preceding query is sent by the application ISS to the server Ontobroker, through the script:

```
<?
$command1 = "isalive";
 $query ="FORALL X,Y,Z<-X :News[#description->>Y] AND Exists Z X:#News[#title->>Z]“.
$Ontobroker  = new COM ("Ontoprise.Client");
$Ontobroker->Query($query);
for($i=1; $i<$Ontobroker->Rows(); $i++)
{
  $temp = ($Ontobroker->Row($i));
  print "$temp1";
 }
?>
```

6. The personalization of the Semantic Portal information contents

The fruition of contents in the Semantic Portal is user oriented, the interfaces are designed in order to be adaptable to four different user profiles.

- Curious (Somebody looking for broad-spectrum information on ISS).
- Student (Somebody who has basic knowledge on ISS and wants to deepen it).
- Teacher (Somebody asking for detailed information on ISS technical aspects).
- Researcher (Somebody who is interested in past experiments on ISS and in making new ones by himself).

In figure 6 the Home page of the Portal is shown, in the table on the left side of the screenshot the user profiles are listed, while on the right side the foreseen functionalities for accessing information common to all the user profiles are shown.

Fig. 6. The Semantic Portal Home page

Additional knowledge access functionalities associated to specific user profiles have also been defined, for example the Curious user information services are the following:

- General information on ISS;
- What are the facilities;
- Astronaut's and material's launch;

- Which are past and future experiments on board;
- Why go so far to experiment something;
- How astronauts live a on board;

In the figure 7 the Home page for the Curious user profiles are shown with the user specific functionalities for knowledge retrieval.

For each user profile specific knowledge access functionality has been chosen to annotate the relevant information as shown in the table below.

Specific knowledge access functionalities	**User Profile**
What are facilities	Curious
Which are facilities' experiments on board	Student
Which are the facilities' technical and physical characteristics	Teacher
Which is the power available for different facilities' container	Researcher

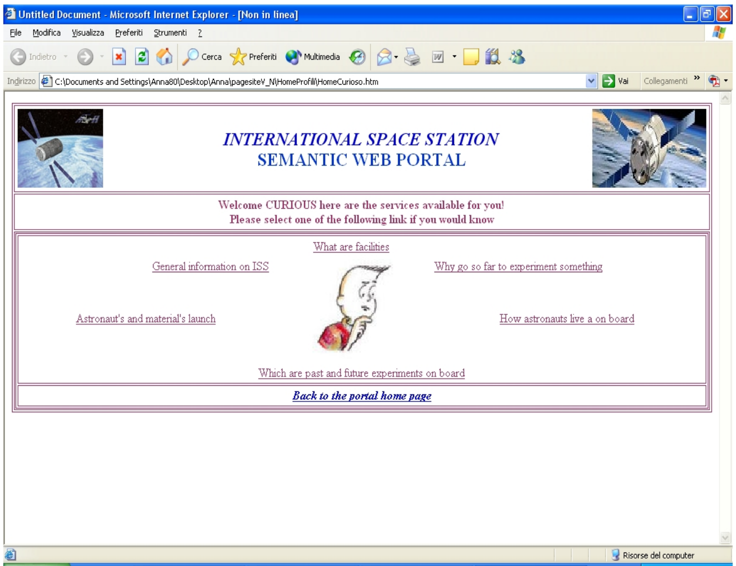

Fig. 7. Home page for the Curious user profile

7. Conclusions

This paper describes the development of a Semantic Portal, in which through the layouts of the user interface the information on the ISS, that exist on various Web sites, are presented in a transparent and user oriented manner.

The role played by the personalization of the knowledge access functionalities are stressed as one of the most relevant aspect of the evolution of the existing Web towards the Semantic Web, this could be an idea to reduce unanswerable question not minimizing the set of rules for making inference from the set of data, but giving additional rules related to the context of the final user.

The development of the Semantic Portal has been possible due to a preliminary annotation of resources (terms, sentences, photos, etc.-) present in various Web sites, thanks to this annotation a definite meaning has been given to them, having as a reference the classes of a specific developed ontology for the ISS domain.

The development of a domain specific ontology, on which the semantic portal is based, allows to annotate and classify the unstructured information available in heterogeneous formats on different Web sites, the annotated data becoming instances of the ontology's classes and attributes.

The reference ontology has been developed using Ontoedit, a tool for the design and editing of ontologies, OntoMat has been used for annotating the web pages. The Portal has a client-server architecture based on the middleware layer of Ontobroker, to develop the communication interface with the ontology. Queries for information retrieval from Web pages are formulated using FLORID, a Frame Logic syntax based language. PHP language is in charge to manage user requests from the Portal Web page and replies of Ontobroker server.

The reference ontology has been developed using Ontoedit, a tool for the design and editing of ontologies, OntoMat has been used for annotating the web pages. The Portal has a client-server architecture based on the middleware layer of Ontobroker, to develop the communication interface with the ontology. Queries for information retrieval from Web pages are formulated using FLORID, a Frame Logic syntax based language. PHP language is in charge to manage user requests from the Portal Web page and replies of Ontobroker server.

The strength of the proposed approach, from the architectural as well as from the methodological approach point of view, is the possibility to merge the information available in the Web sites with the one present in databases (DBs). The access to the information stored in DBs is even easier due to the fact that it is already structured; the merging is achieved by associating the DBs tables to the ontology's classes and the fields' values to the slots of these classes.

In addition, using some features of the Ontobroker system (AltaVista loc, AltaVistasm) which uses some AltaVista web search engine functionalities, it's also possible to access unstructured information from the whole Net. This last features, together with the possibility to enlarge as possible the annotated Web pages once the reference ontology has been established, makes the Semantic Portal a really "open system".

References

1. Berners-Lee T., Hendler J., Lassila O. The Semantic Web, Scientific American, May 2001 http://www.scientificamerican.com/2001/0501issue/0501bernerslee.HTML
2. J. Frohn, R. Himmeröder, P. Kandzia, C. Schlepphorst. How to Write F-logic Programs in FLORID - A Tutorial for the Database Language F-logic. http://www.informatik.uni-freiburg.de/~dbis/florid/
3. Guarino N. Formal Ontology and Information Systems. National Research Council, LADSEB-CNR, Corso Stati Uniti 4, I-35127 Padova, Italy
4. Guarino N.(1997) Understanding, building, and using ontologies. "Using Explicit Ontologies in KBS Development", by van Heijst, Schreiber, and Wielinga." International Journal of Human and Computer Studies 46.
5. Kifer M., Lausen G., Wu J. Logical Foundations of Object-Oriented and Frame-Based languages. In journal of the Association fo Computing Machinery, May 1995.
6. Tasso C., Omero P, La personalizzazione dei contenuti WEB, Franco Angeli, 2002"
7. Cream, http://www.aifb.uni-karlsruhe.de/WBS/sha/papers/aa_cream_www11.pdf/
8. Ontoedit Project, http://www.ontoprise.de/products/ontoedit_en
9. OntoMat, http://annotation.semanticweb.org/tools/ontomat
10. Ontoprise GmbH - How to write F-Logic Programs, www.ontoprise.de
11. Ontoprise GmbH - How to Use Ontobroker, www.ontoprise.de
12. OntoAnnotate, http://www.ontoprise.de/products/ontoannotate
13. Fensel D., Ontologies: A silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag Berlin Heidelberg New York, 2001