An Approach to Manage the Web Knowledge

Filippo Eros Pani, Maria Ilaria Lunesu, Giulio Concas and Gavina Baralla
Department of Electrics and Electronics Engineering, University of Cagliari, Piazza d'Armi, Cagliari, Italy

Keywords: Knowledge Management, Multimedia Content, Semantic Web, Knowledge Base, Taxonomy.

Abstract: The spread of the Social Web is influencing the evolution of Semantic Web: the way of producing and consulting information changes, as well as the way people relate themselves with the Internet and the services it gives. Users will participate at first hand to the developing of the Web which therefore becomes interactive. This study considers this feature, trying to link the worlds of Social Media and Semantic Web, with the aim of proposing a semantic classification of the information coming from the Web, which do not always follow a well-defined order and organization. Starting from a precise analysis of the information of the Web through an accurate and meticulous study on how these are presented and used, in order to give a sorted and easily usable data structure, this approach wants to define a taxonomy able to represent knowledge through an iterative combined approach, where top-down and bottom-up analyses are applied on the knowledge domain we want to represent.

1 INTRODUCTION AND RELATED WORK

Over the last decade a broader knowledge of the Web has strengthened and fostered the developing of new applications: the Web has turned into a multifunctional platform where users no longer get the information passively; in fact, they become authors and makers. This has been mainly possible thanks to the developing of new applications which allow users to add contents without knowing any programming code. The social value which the Web has acquired recently is therefore unquestionable; the Web's structure grows and changes depending on the user's needs, becoming every day more complex.
The new frontier for the Internet is represented by the Web 3.0 (Berners-Lee et al., 2001): with the evolution of the Web into its semantic version, a transition to a more efficient representation of knowledge is a necessary step. Particularly, data are no longer represented just by the description of their structure (syntax) but also by the definition of their meaning (semantics). In fact, a data can have a different meaning depending on the contexts; the use of tools like ontologies and taxonomies helps the classification of information, as shown also in (Decker et al., 2000; Maedche and Staab, 2001; Jacob, 2003; Davies et al., 2003; Strintzis et al., 2004; Jewell et al., 2005; Hepp, 2007; Gruber, 2008) and (Simperi, 2009).

The Web becomes clever and is conceived as a big database in which data are orderly classified. “Information”, therefore, is one of the keywords at the base of the success of both search engines (Google, Yahoo, Bing, …), which become more refined in data retrieval and presentation, and Social Networks (Youtube, Facebook, Twitter, Flickr, …), which allow exchange and sharing, creating an interconnection among users and content makers. However, such data, despite being formally available, are often unreachable as for their semantic meaning and cannot be used as real knowledge.

Various proposals to solve these problems can be found in literature, also to overcome the semantic heterogeneity problem (Euzenat and Shvaiko, 2007) and to facilitate knowledge sharing and reuse (Fensel et al., 2001; Gómez-Pérez and Corcho, 2002). In (Schreiber et al., 2001) an approach based on the use of an ontology to make annotating photos and searching for specific images more intelligent is described; and in (Jaimes and Smith, 2003) a data-driven approach to investigate semi-automatic construction of multimedia ontologies is used. With the emergence of the Semantic Web, a shared vocabulary is necessary to annotate the vast collection of heterogeneous media: in (Jewell et al., 2005) an ontology is proposed to provide a

Eros Pani F., Ilaria Lunesu M., Concas G. and Baralla G..
An Approach to Manage the Web Knowledge.
DOI: 10.5220/0004547900810088
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2013), pages 81-88
ISBN: 978-989-8565-81-5

meaningful set of relationships which may enable this process.

Particularly, in (Lunesu et al., 2011) the problem of representing and managing the knowledge which can be found on the Internet is discussed as for the User Generated Content (UGC), classifying this knowledge through a top-down (TD) and bottom-up (BU) combined approach. To reach such target, an ontology was built as a base to define a repository of multimedia contents, putting a special focus on the georeferencing of multimedia objects. As for the TD approach, the standards used for the multimedia objects (XMP, Exif, etc.) have been defined by selecting data of interest to represent them on the ontology, which in turn was defined through rules of correspondence. As for the BU approach, UGCs of two particularly exemplifying platforms (Flickr and Youtube) have been analyzed, in order to extrapolate some structured tags, folksonomies and attributes of multimedia objects (characteristically Exif, as for Flickr, and proprietary tags, as for Youtube). In conclusion, this ontology allowed for the construction of a repository to store the information extracted from UGC systems, where all the information related to multimedia objects are shown (compatibly with the XMP standard) as well as other tags of general interest, apart from representing also the information which can be found on the as-is folksonomies.

This study aims at defining a new approach for the problem of the contents on the Internet, especially semi-structured contents coming from heterogeneous sources referring to a common knowledge domain. Through a combined TD and BU approach, knowledge of a specific domain was extracted defining a common structure through a taxonomy, in order to classify and make the majority of such knowledge available.

With the TD approach the knowledge of interest on the domain was defined, following the specifications and the analysis of the ontologies and other classifications, in order to define a reference taxonomy.

On the other hand, the BU approach started from the selection of some websites concerning the domain of interest, to pinpoint the knowledge in them. Then, these contents were classified with the taxonomy previously defined and the mapping rules between contents and taxonomy.

This taxonomy allowed for the definition of a reference knowledge which may later be managed in terms of really usable and interesting knowledge, fostered by the whole knowledge of all the selected websites.

We chose to test this approach on the knowledge domain of Italian wines reviews. As for the validation, we verified how this KMS allowed such knowledge to become available on systems that were compliant with the Wines ontology as defined as an example of Semantic Web by W3C; then we checked other websites of Italian wines reviews, verifying how their contents of interest could be represented and managed on the KMS through some simple mapping rules.

The paper is structured as follows: in the second section of this paper we present our proposed approach for knowledge management and in the fourth we explain the case study. The next section includes the analysis of results and verification. Finally, the fifth section includes the conclusion and reasoning about the future evolution of the work.

2 THE APPROACH TO MANAGE THE KNOWLEDGE

The proposed approach aims at defining a taxonomy able to represent knowledge through a mixed-iterative approach, where TD and BU analyses of the knowledge domain which has to be represented are applied: these are typical approaches for this kind of problems. In this case, they are applied following an iterative approach which allows, through further refinements, for the efficient definition of the taxonomy able to represent the domain's knowledge of interest.

The knowledge to be represented is the most popular among users of a certain domain. To determine which is the users' knowledge of real interest we chose to select the most used websites by users, the most important and looked up ones. For this definition, websites with a higher ranking on Google among the domain of interest are typically chosen.

2.1 Top-down Phase

When our knowledge or our expectations are influenced by perception, we refer to schema-driven or TD elaboration. A schema is a model formerly created by our experience. More general or abstract contents are indicated as higher level, while concrete details (senses input) are indicated as lower level.

The TD elaboration happens whenever a higher level concept influences the interpretation of lower level sensory data. Generally, the TD process is an information process based on former knowledge or

acquired mental schemes; it allows us to make inferences: to “perceive” or “know” more than what can be found in data. TD methodology starts, therefore, by identifying a target to reach, and then pinpoints the strategy to use in order to reach the established goal.

Our aim is therefore to begin by a formalization of the reference knowledge (ontology, taxonomy or others) to start classifying the information on the reference domain.

The model could be, for instance, a formalization of one or more classifications of the same domain, formerly made in a logic of metadata. Therefore, the output of this phase will be a table with all the elements of knowledge formalized through the definition of the reference metadata.

2.2 Bottom-up Phase

With this phase the knowledge to be represented is analyzed by pinpointing, among the present information, the ones which are to be represented together with a reference terminology for data description.

When an interpretation emerges from data, it is called data-driven or BU elaboration. Perception is mainly data-driven, as it must precisely reflect what happens in the external world. Generally, it is better if the interpretation coming from a system of information is determined by what is effectively transmitted at sensory level rather than what is perceived as an expectation. Applying this concept, we analyzed a set of websites containing the information of the domain of interest; from these websites, both information whose structure needed to be extrapolated and the information in them were pinpointed. Typically, reference websites for that information domain are selected, namely the ones which users mainly use to find information of their interest over the domain itself.

Primary information, important ones, already emerge during the phase of websites analysis and gathering: during a first skimming phase, the minimum, basic information necessary to well describe our domain can be noticed. Then, important information are extrapolated by choosing fields or keywords which best represent the knowledge, in order to create a knowledge base (KB). In this phase, one of the limits could be the creation of the KB itself, because each website is likely to show a different structure and a different way of presenting the same information. Therefore, it will be necessary to pinpoint the present information of interest, defining and outlining them. After this analysis of gathering of information, a classification is made and it has to reflect, in the most faithful way, the structure of the knowledge proposed by every single website, respecting both its contents and hierarchy.

To analyse data, we chose to build a tabular system for each website coming from a precise identification of each information area existing in every website taken as a knowledge base.

For each website we created a table which accurately gathers and describes the information that can be found in it, with a detailed field of descriptions. With this stage we obtained a complete representation of the knowledge which can be found on the chosen websites, but not a usable one because it had not been classified yet.

2.3 Integration Phase

In these phases we will try to reconcile these two representations of knowledge of the domain, as represented in the former phases.

Thus, we want to pinpoint, for each single TD's metadata, where the information can be found in the table's fields representing the knowledge of each website (which, for us, represents the knowledge we want to represent, considering the semantic concept and not the way to represent it, absolutely subjective for every website).

At this point we check if, in every table, the information of our representation of knowledge coming from the TD can be found in the tables coming from the BU, verifying if it exists as a field or can be found in a field or is missing.

Then, we will create a mapping macro-table of knowledge containing, for each item of the taxonomy, the correspondence if and where that information exists in the various websites and also the information of the websites which are not represented by the taxonomy.

From the macrosystem a KB originates, which is able to represent both the formalization of knowledge and the present knowledge.

2.4 Formalization of Knowledge

Starting from this KB, further iterative refining can be made by re-analyzing the information in different phases: 1) with a TD approach, checking if the information which are not represented by the chosen formalization can be formalized; 2) with a BU approach, analyzing if some information of the websites can be connected to formalized items; 3) with the mixed phase by which these concepts are

reconciled. This is obviously made only for the information to be represented.

The knowledge we want to represent is the one considered of interest by the users for the domain: for this reason, the most important and looked up websites are chosen. At the end of this analysis we will define a taxonomy able to represent the knowledge of interest for this domain, which may also not have items from the taxonomy (or ontology from which we started in the TD analysis), but may have items which did not exist in it, emerged from the BU analysis. The final result of this phase will be a reference taxonomy, where, for each item, there is a linked information about where the knowledge of interest can be found on each website.

3 CASE STUDY ABOUT WINES

In this study we chose as a case study the domain of wines and, particularly, the one belonging to the technical files and/or descriptions of "Italian wines": the choice was not made randomly as the world of wines is rich in contents and complete enough to give a good starting point for our study. In fact, there are thousands of contents which can be found on the Internet; also, there are different studies on the classification of wines from which we can draw on.

3.1 Knowledge Base of Interest

Contents on wine available on the Web are thousands, offering a significant KB.

Our study takes into consideration a subdomain of wine, represented by all the most important reviews which can be found on the Internet. From the analyses of the domain on the web and the Google Ranking of these websites, we chose a list of suitable and representative websites, having considered the popularity and the reliability given by the Web. The websites we took into consideration are the following: Decanter.com; DiWineTaste.com; Lavinium.com; GamberoRosso.it; introspective.com Snooth.com; Vinix.com. These websites are considered as representative for our study also because of their own information structures, particularly various and differentiated. Each website has its own structure and a different representation of the information. To correctly define our domain it was therefore necessary to precisely analyse the contents in each of them and the layouts. The structure of the page showing the review is useful to understand if the same website always uses the same structure and the same items for every review. Unfortunately we saw that some of them show the same information differently depending on the review, using, for instance, different tags for the same data. This, obviously, is a limit in the process of classification of contents. It is thus necessary to align the different items for the same website, used to represent the same data.

3.2 Top-down Phase

In this phase we analyzed the existing formalizations for the representation of knowledge of this domain. A very interesting formalization which we pinpointed was the one by the Associazione Italiana Sommelier (AIS), The Italian Sommelier Association, providing a detailed description of all the terms associated with wine. Another important formalization was the one by the European law defining the reference features of a certain wine, such as type, colour, grape variety, etc. From these two, a reference taxonomy for those features was created. As an additional formalization, we chose a reference scheme, represented by an ontology already existing on the Web and made by W3C: wine ontology [http://www.w3.org/TR/owl-guide/wine.rdf]. An ontology is more complex than a taxonomy. It has, apart from class hierarchies, property hierarchies with cardinalities for the assignable values. It offers a general view of the world of wines, with a less detailed description for certain fields as stated on the reviews found on the Web. Moreover, from this ontology we took into consideration only the areas of interest existing in our classification, omitting those ones representing elements not of interest (such as, for instance, each winemaker's property).

Starting from these reference formalizations, a first taxonomy was built in which we pinpointed the items to create the reference table. After choosing the items of interest in the reference ontology, we analyzed the direct correspondence among tags of the two representations, directly extracting the ontology ones from the OWL code. To standardise our taxonomy we decided to take into consideration the RDF standard indicating, just for the items with a correspondence, its URI. The RDF standard allows to associate a URI also to the properties. website, used to represent the same data.

3.3 Bottom-up Phase

The BU analysis required a detailed analysis of the contents of these websites, trying to pinpoint the information we considered as important; then we

studied the structure of each single source, useful to see the existing data and their position in the layout of the page.

Once the KB for the domain of interest composed by the websites was defined, the next step was classifying all the chosen information. Such classification is made by creating a classification of the BU contents because it was built from the bottom: information on the websites are thus accurately analyzed.

We start from the analysis of the specific to reach a general classification of data. One of the initial steps of our project contemplated the study of the structure of each source, useful to see the existing data and their position on the page's layout.

This procedure happens to be important also at this point of the study, because allows for the evaluation of the classification of information. Both the item “maturazione”, but also the organoleptic analysis (visual, olfactory and gustatory test) if existing, are systematically shown on the websites taken into consideration, into the area which we identified as “tasting notes”. For this reason, to build the hierarchy we tried to respect the original, already existing one.

The type of classification was also revealed during the data analysis phase, during the study of semantics and uniformation.

With the creation of the tables we tried to represent the knowledge in the shape of fields as faithful as possible to those ones already existing in the samples taken into consideration.

The evaluation of this phase is subjective and left to the intuition of the analyst, which freely interprets the information at their disposal, intuitively obtaining the taxonomic tree. This step happens to be very tricky, because is susceptible to accidental mistakes. However, we could say that the various structures found in the domain which we considered, apart from the caption used to define each field, are not so different, thus the classification did not raise any big doubt, as for the representation.

Thus, the macrosystem made 7 tables, one for each website. Every table has the list of information of the website it represents.

3.4 Mixed Phase

During this phase, the items of the fields existing in the taxonomy defined in the TD phase were compared to the fields of the tables created in the BU phase. To do this, we built a mapping macro-table of knowledge containing, for each item of the taxonomy, the correspondence if and where that information exists on the various websites and also the information existing on the websites which were not represented by the taxonomy.

To each item we thus assigned a numerical value to represent this mapping: 1) existing and extractable information; 2) existing but not extractable information; 3) sometimes existing and extractable information; 4) sometimes existing but not extractable information; 5) always missing information.

For the fields with values 1 and 3, the corresponding field and the mapping rule to extrapolate the information are also indicated.

The information with value 2 and 4 is embedded (hidden in the text) and, therefore, should be specifically looked for with tools of semantic analysis. Anyway, the field in which it exists is indicated.

With this analysis and classification of every single data we managed to solve the inhomogeneity of the information existing in the Web, as for the domain of interest. This allowed to study both its structure and the type of information existing, giving us the chance to examine how data are presented and the classification given for each website.

When creating the taxonomy, which wants to be a semantic classification, we also tried to represent the structure of data and the existing hierarchies of the sample websites.

3.5 Formalization of Knowledge

This activity was iteratively repeated to best represent the knowledge and its connections described in the macro-system mentioned above. As expected, not all the fields were taken into consideration, neither among those existing in the initial taxonomy nor among the extrapolated ones, and those ones which appear just once in the whole macro-system were rejected (evaluation made considering the field value = 5), such as, for instance, “Bicchiere consigliato” or “Temperatura di servizio consigliata”.

The inhomogeneity among the information existing in the different websites was analyzed by looking for the semantic correspondences represented in the macrosystem with the column ‘field details’. The same principle was used to uniform fields with numerical values. The final range takes into account the classification used by the majority of websites.

A simplifying table summarizing the procedure of classification described above is shown below.

The result of these phases was the knowledge

Table 1: Classification.

Macrosystem Items	Final Tags
Wine's identification name	Wine
<Produttore> <Winery> <Producer>	Winery: address, telephone, fax, e-mail, web, map, other wine, other info winery
<classification> <denominazione> <tipologia>	Classification: Vino da tavola, IGT, DOC, DOCG
<tipologia> <type>	Colour: white, rose, red
<type> <tipologia>	specification
Qualification: embedded	Qualification: classic, reserve, superior
<typical grape composition> <Varietal> <vitigni> <uve>	grape
<titolo alcolometrico> <alcohol> <alcol>	Alcohol
Label	Label
<origin> <region> <zona>	State/Region
<tasting notes> <reviews> <overview>	Tasting notes
<prezzo enoteca> <prezzo> <starting at> <$> <average bottle price>	Price
<abbinamento> <suggested recipe pairings> <food pairing suggestions>	Food pairing suggestion
<posted by> <source>	Author
<posted on> <inserito> <degustazione in data>	Date
<decanter rating>: max 5 stelle <rated>: max 5 bicchieri <valutazione>: max 5 chiocciole <punteggio>: max 5 diamanti <voto>: max 5 chiocciole Punteggio: max 3 bicchieri	Rate: 60-70; 71-75; 76-80; 81-85; 86-90; 91-100

base formalized through the taxonomy. The table shows some items of it, with a field of value 1 or 3 and expressed in textual form (for instance, those ones directly extractable through tags or metadata). Other fields, represented by an icon, were rejected, though their presence was considered.

4 ANALYSIS OF RESULTS AND VERIFICATION

During the validation phase we verified how our KMS made the acquired knowledge usable for the systems compliant with other ontology of wines and for other websites on Italian wines reviews. We went on verifying how the contents of interest of these websites could be represented and managed on the KMS through some simple mapping rules.

Then, we tried to solve the clear inhomogeneity by paying more attention to the semantic meaning and not to the notation used to represent those contents. In fact, the purpose of the study was not to describe the whole world of wines, but just the part of it represented by the information which can be found on the Web.

After matching the two systems, Ontology and Taxonomy, the information were generalized and made coherent. This allowed us to verify that our system is able to represent and combine specific information, and at the same time understands the main variances between the two systems, namely the difference of some considered information.

This kind of study can also be used to enrich an already existing ontology with fields coming from a general classification, evaluating a possible integration of such information without damaging the existing hierarchy, so that we can have a broader and more accurate view over the analyzed domain.

4.1 Choice of Samples

To continue with the phase of verification of the created taxonomy, we decided to take into consideration another set of samples – again, wine reviews which can be found on the Web.

The choice of the websites for the testing phase followed the same criteria used during the analysis of the domain. The main obstacle we found was due to the popularity of the product and the large amount of followers who have a very subjective way of representing the information about wine and the acquired knowledge. Here comes the need of pinpointing sources with clear, easily extractable and objective information.

One of the main features which these sources needed to have was the presence of differentiated fields with a single notation rather than a broad textual field. So, also in this case, all the websites gathering a large quantity of information in just a macro-textual area were rejected. In fact, these kind of websites, though full of contents, were not suitable for the testing phase. The embedded

information, though fostering the acquisition of a general knowledge, do not facilitate its own structured classification. Similarly, some apparently suitable sources happened to have very few contents, with a database so poor that it did not mention the most appreciated wines.

After these considerations, the websites we decided to take into consideration for the tests were the following: guida-vino.com; vinogusto.com; kenswineguide.com; buyingguide.winemag.com.

4.2 Testing Phase

For each sample website, in this testing phase we verified whether the information in them could be found in the classification proposed by us, and whether our taxonomy could be able to represent them. For each website, therefore, the following table was built, representing the specific fields of information which was the same for every review that we analyzed.

Table 2: Testing phase.

Existing Information	Field Details	Taxonomy Item
Label	Label's image	Wine.label
Producer	About the producer	Wine.winery
Classification	IGT, DOC, DOCG	Wine.classification
Grape variety	Grape variety	Wine.grape
Range of prices	Price	Wine.prices
Others years	Others years	Wine.winery.info Winery.otherWines
Presentation/ comments	Wine tasting	Wine.tastingNotes
Rate: max 5 stars	Rate	Wine.rate

In the light of the results obtained in this testing phase, we are satisfied with the taxonomy which we created. In fact, with this testing phase, we saw that the classification defined in our study reflects the type of contents needed. Such classification, therefore, is usable, re-usable and possibly extendible to the domain of interest of wine.

5 CONCLUSIONS

The spread of the Social Web is significantly influencing the evolution of Semantic Web: users themselves are creating rules for the representation of information. The structure of the Web grows and changes giving the user the chance to actively participate in the developing of the Web. For this reason, our study took into consideration this feature with the uniformation of UGCs, trying to link the two worlds: Social Media and Semantic Web. Also the main search engines (Google, Yahoo, Bing, …) and the main Social Network (Youtube, Facebook, Twitter, Flickr, …) are evolving, specializing and interconnecting themselves on data retrieval, presentation, exchange and sharing.

That being so, the basic idea of our study was to propose a solution to the problem of the different contents of the Web, coming from different sources but belonging to the same domain of knowledge.

Our proposal is to define a taxonomy able to represent knowledge through a mixed iterative approach, articulated in a top-down analysis and a bottom-up one of the domain of knowledge which is to be represented. Thus, first we tried to define the knowledge of interest on the domain, depending on the specifications, and through the analysis of the existing ontologies, in order to define a reference taxonomy. Then, the knowledge we considered as important (and as an element of common interest) was extracted from a selection of websites belonging to the domain of interest. These contents are to be classified in the taxonomy mentioned before, also using mapping rules made ad-hoc. The taxonomy created allowed for a definition of the reference knowledge which could then be managed as an actual usable knowledge, fostered by all the information existing on the selected websites. Due to the large amount of the information available, we chose as domain of knowledge a sub-domain of wine, represented by the reviews which can be found on the Web.

From the analysis of the domain on the Web and the Google Ranking of many websites, we chose a list of some suitable and representative ones after considering popularity and reliability given from the Web.

We chose to validate the proposed approach by verifying how the KMS allowed to make the acquired knowledge usable and accessible to the systems compliant with the Ontology of Wines taken into consideration along with other websites of Italian wine reviews, underlining how, also in this case, the collected information could be represented and managed on the KMS through some simple mapping rules. Such a system could be enriched by deducing an ontology of information existing on the Web to be compared with another ontology representing the same domain. A similar comparison

has the advantage to be simple, less disorganized and surely less susceptible to mistakes than the one proposed in our project.

A further, interesting development could be the creation of repositories able to collect the information previously classified and, through a system made ad-hoc, they would be presented to the final user in a structured and customized way, depending on the requests, and possibly developing a graphic interface which could be able to draw the curiosity and the interest of the user.

REFERENCES

Barton, J., Currier, S., Hey, J. M. N. (2003). Building Quality Assurance into Metadata Creation: an Analysis based on the Learning Objects and e-Prints Communities of Practice. In Sutton, S., Greenberg, J. and Tennis, J., Eds. *Proceedings 2003 Dublin Core Conference*: Supporting Communities of Discourse and Practice - Metadata Research and Applications, Seattle, Washington (USA).

Berners-Lee, T., Hendler, J., Lassila, O., 2001. In The Semantic Web, Scientific American, pp. 28-37. http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html

Davies, J., Fensel, D., van Harmelen, F., 2003. Towards the Semantic Web: Ontology-driven Knowledge Management, *John Wiley & Sons*, ISBN: 9780470848678.

Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., Horrocks, I., 2000. The Semantic Web: the roles of XML and RDF. In: *Internet Computing, IEEE*, Vol. 4, Issue 5, pp. 63-73.

Euzenat, J., Shvaiko, P., 2007. Ontology matching, *Springer-Verlag*, Berlin Heidelberg (DE), ISBN: 978-3-540-49612-0.

Fensel, D., Van Harmelen, F., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., 2001 OIL: an ontology infrastructure for the Semantic Web. *In Intelligent Systems, IEEE*, Vol. 16, Issue 2, pp. 38-45.

Gómez-Pérez, A., Corcho, O., 2002. Ontology languages for the Semantic Web. *In Intelligent Systems*, Vol. 17, Issue 1, pp. 54-60.

Gruber, T., 1993. A Translation Approach to Portable Ontology Specification. *In Knowledge Acquisition*, Vol. 5, pp. 199-220.

Gruber, T., 2008. Ontology. In Liu L., Özsu, M. T. (Eds.), Encyclopedia of Database Systems, *Springer-Verlag*.

Hepp, M., 2007. Ontologies: State of the Art, Business Potential, and Grand Challenges. In Hepp, M., De Leenheer, P., de Moor, A., Sure, Y. (Eds.), Ontology Management: Semantic Web, *Semantic Web Services, and Business Applications*, ISBN 978-0-387-69899-1, Springer, pp. 3-22.

Horrocks, I., 2008. Ontologies and the semantic web. In Communications of the ACM, 51(12):58-67.

Jacob, E. K., 2003. Ontologies and the Semantic Web. In Bulletin of the American Society for Information Science and Technology, Wiley Periodicals, Vol. 29, Issue 4, pp. 19–22.

Jaimes, A., Smith, J., 2003. Semi-automatic, data-driven construction of multimedia ontologies. *In Proceedings of IEEE International Conference on Multimedia and Expo* (ICME), Vol. 2.

Jewell, M. O., Lawrence, K. F., Tuffield, M. M., Prugel-Bennett, A., Millard, D. E., Nixon, M. S., Schraefel, M. C., Shadbolt N. R., 2005. OntoMedia: An Ontology for the Representation of Heterogeneous Media. *In Multimedia Information Retrieval Workshop*, ACM SIGIR.

Lassila, O., Swick, R., 1999. Resource Description Framework (RDF): Model and Syntax Specification, Recommendation W3C. http://www.w3.org/TR/REC-rdf-syntax

Lunesu, M. I., Pani, F. E., Concas, G., 2011. An approach to manage semantic informations from UGC. In International Conference on Knowledge Engineering and Ontology Development , Paris, France.

Lunesu, M. I., Pani, F. E., Concas, G., 2011. Using a standards-based approach for a multimedia knowledge-base. *In International Conference on Knowledge Management and Information Sharing Paris*, France.

Maedche, A., Staab, S., 2001. Ontology learning for the Semantic Web. *In Intelligent Systems, IEEE*, Vol. 16, Issue 2, pp. 72 – 79.

Noy, N. F., McGuinness, D. L., 2001. Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems, Laboratory Technical Report KSL-01-05, 2001.

Schreiber, A. Th., Dubbeldam, B., Wielemaker, J., Wielinga, B., 2001. Ontology-Based Photo Annotation. *In IEEE Intelligent Systems*, Vol. 16, pp. 66-74.

Simperi, E., 2009. Reusing ontologies on the Semantic Web: A feasibility study. *In Data & Knowledge Engineering*, Vol. 68, Issue 10, pp. 905-925.

Strintzis, J., Bloehdom, S., Handschuh, S., Staab, S., Simou, N., Tzouvatras, V., Petridis, K., Kompatsiaris, I., Avrithis, Y., 2004. Knowledge representation for semantic multimedia content analysis and reasoning. *In Proceedings of the European Workshop on the Integration of Knowledge*, Semantics and Digital Media technology.

W3C, Wine ontology. http://www.w3.org/TR/owl-guide/wine.rdf