Using Semantic Web Technologies to Improve Web Searching of Human Resources

Goran S. Krstić

Abstract - **Today's web search engines do not rely or rely little on semantics. This paper considers potential possibilities of the Semantic Web, especially the field for improvement and advancement of the present web searching. In the end, it is given a possible scenario for searching human resources in the future semantic environment, which the Semantic Web should offer.**

Keywords - **Web searching, Semantic Web, Ontology, Human resources.**

I. INTRODUCTION

Efficient searching data on the Web mostly depends on the way they are organized and the mechanisms for representation the semantic contents within. The main problem about searching is not in search engines, but in the absence of meta data that would be built in web pages and would closer determine the meaning of those pages contents.

The Internet users still search the web due to keywords and hyperlinks in order to get to desired data. There is need for more efficient getting to searched information, but with spending less time. The Semantic Web should enable more efficient location of these data, when they become intelligible for both people and machines.

The central part in the Semantic Web vision occupies ontologies. The ontologies offer a possibility to work with heterogeneous representations of web resources and their interrelations. The ontologies enable for the web resources' contents and the information within to organize in predefined classes, which in the same time present certain knowledge from that domain [1].

The paper considers possibilities for improving searching the Web by using Semantic Web technologies. At the moment, there is a belief that the first to be created are the Semantic "isles", which would later be linked to the Semantic Web. In the end, it is given a scenario of a Semantic Web for searching human resources on the Web.

II. SEMANTIC WEB

The Semantic Web is a new layer in Internet which enables a semantic representation of the web pages contents [2].

A. Semantic Web Overview

The Semantic Web presents the extension of the existing Web, thus will contain many aspects of the latter, e.g. URI addressing model, protocols with small sets of commands, decentralization etc. The Semantic Web is still however, more a vision than reality [3].

The Semantic Web is based on the idea to create "machine intelligible" data that could be automated, integrated and used by different applications. By using the web language, e.g. RDF and OWL, it is possible to create semantic data models. These models are built on triples *resource-property-value.* A user can create his own classes and properties. Classes in Semantic Web languages are categories or types. Instances of those classes can be created and built up in web pages.

The Semantic Web is an attempt to make data readable to machines, the same way documents are readable to human. The Semantic web will, as well, enable software agents to perform certain actions and procedures automatically, which at present have to be performed manually, i.e. which existing applications cannot execute.

B. Semantic Web Technologies

W3C page on Semantic Web includes a diagram with semantic web architecture (www.w3.org/2000/talks/1206-xml2k-tbl/overview.html). According to this diagram, technologies are hierarchically ranged-from syntactic aspect to semantic meaning of data. The abridged presentation is given in Fig. 1:

Trust
Logic and Proof
Ontology
RDF Schema
RDF
XML Schema
XML

Fig. 1. The Semantic Web levels.

W3C is developing or has developed all the standards except the two at the top. Each level, as one can notice, is built based on the level below. The lowest level does not depend on any level above. Thus the levels can be developed independently. XML and XML Schema have become standards. RDF and OWL are recommendations. The other levels are developing.

Goran S. Krstić is with the Faculty of Economics, Trg kralja Aleksandra Ujedinitelja 11, 18000 Nis, Serbia and Montenegro, E-mail: gorca@ni.ac.yu

III. SEARCHING

Nowadays, the Internet searching is accomplished in the following manner: one inserts the terms being looked for, into the input box on a web page of a certain search location. The result comes out as an address list of the pages that satisfy the condition of searching. Most of the search locations offer also varieties of search aids. Yahoo was the first to define categories. They were not generated by using semantic web technologies, but they present the first usage of semantics on the Web. Many search locations offer additional terms for searching in keeping with a user's request.

No matter which search engine is used, the results that appear, in most cases are irrelevant hits, so that the user is forced to browse a large number of pages in order to get to the desired result.

A. Ttechniques for searching

Internet represents, so far, the greatest challenge for searching. Internet is the environment which is constantly developing and changing. It is inconsistent, and it contains documents of different types and structures. At the moment, there is no standard to index, catalogue, analyze or specify quality and performance of the search engines [3]. Some of the techniques for searching are the following:

Keywords. The most wide-spread approach for indexing on the web is searching resources due to the keywords list. When the user types one or more terms, the search engine tries to find them in its index. Addresses of the pages that contain inserted keywords, return as a result of searching.

It is considered that keywords, although they may be very useful, will not satisfy future growing needs for searching the Web.

Ontologies. Ontology is an explicit specification of the vocabulary for a certain domain; it includes definitions of classes, relations, properties, and limitations. Because of the great variety of data on the Web, the great number of ontologies is domain-directed or personalized in order to express individual or common interest [4]. Ontologies enable the web resource contents and information within, to organize in predefined classes.

The ideal case would be: when the computer itself could browse the contents of the documents, and according to the contents to classify it within a certain category. For example, let one category be "fiction". The computer browses a certain text and grasps that it is "fiction". Possible categories are specified by ontologies. Following that, terms used in queries should automatically map to the categories, provided in ontologies.

Alternatively, terms from ontologies may be presented to the user, who can choose one or more of them for a query. The categories contained in ontologies are presented to the user through corresponding interface, so he does not have to learn by heart all the categories and terms from the ontologies. The second approach is more realistic at the moment. The condition is that web pages be correctly marked, in accordance with ontologies.

Clustering. Search engines usually return a large number of results, which is hard to examine. It is desirable to group these results into adequate categories. Grouping the results into categories is called - *clustering*. One of the top search engines in clustering is Vivisimo (www.vivisimo.com). Vivisimo clusters short quotes from other search engines into *ad hoc* categories.

Ontologies can help in clustering because they enable explicit and clear classification of web resources in predefined categories. Clustering is a good way of presenting a large number of results to the user, and which will be worked on in the future.

B. The Possibilities of Semantic Web for searching improvement

In the present phase of web development is especially interesting the combining of free text searching with a possibility for using RDF meta data for searching [1, 4, 5].

The idea for Semantic Web is to add web pages marks that determine closer the meaning of those pages' contents. In a small domain certain ontology could be standardized, which would be later used by designers to describe the contents of the pages they create. Adding standard ontologies to web pages, enables above all, their better categorization. According to this, there is a need to enlarge the search engines' and agents' software, in order to extract these data. Documents marked in a standard way, using standard ontologies, can be classified better since the designer himself classifies a document into a certain category. In addition, the designer himself can determine the scale of classification, which would enable better ranging of the search results.

One of the advantages of machine readable meta data, as semantic marking, is that search engines can use it for performing additional semantic relations, i.e. they can perform inference.

Although there is a certain dose of doubt that the semantic searching will bring tremendously better results compared to traditional searching, using keywords, it could be expected that the semantic searching will improve the traditional one, offering:

- better categorization,
- possibilities for intelligent query extension and
- possibilities of inference.

In the future will be expected new methods, how the technologies developed for the Semantic Web can improve searching on the Web. One of the desired possibilities would be semantic analysis of the complete web pages' contents.

IV. A SCENARIO FOR SEARCHING HUMAN RESOURCES ON THE WEB

The scenario can be used as a demo or a vision. In both cases, the scenario can be viewed as a tool which helps technological innovations at different levels of a project development. Scenarios are useful for creating awareness on current technological limitations, as well as needs for their overcoming [6].

A. Scenario

The term ontology has roots in philosophy and implies a formal representation of concepts from a certain domain. Above all, they were introduced as ontologies on being. Thus it seems natural to use them for specifying human resources. In addition to this, man is a biological and social being, which implies many descriptions - ontologies. One of the possible scenarios is presented in Fig. 2.

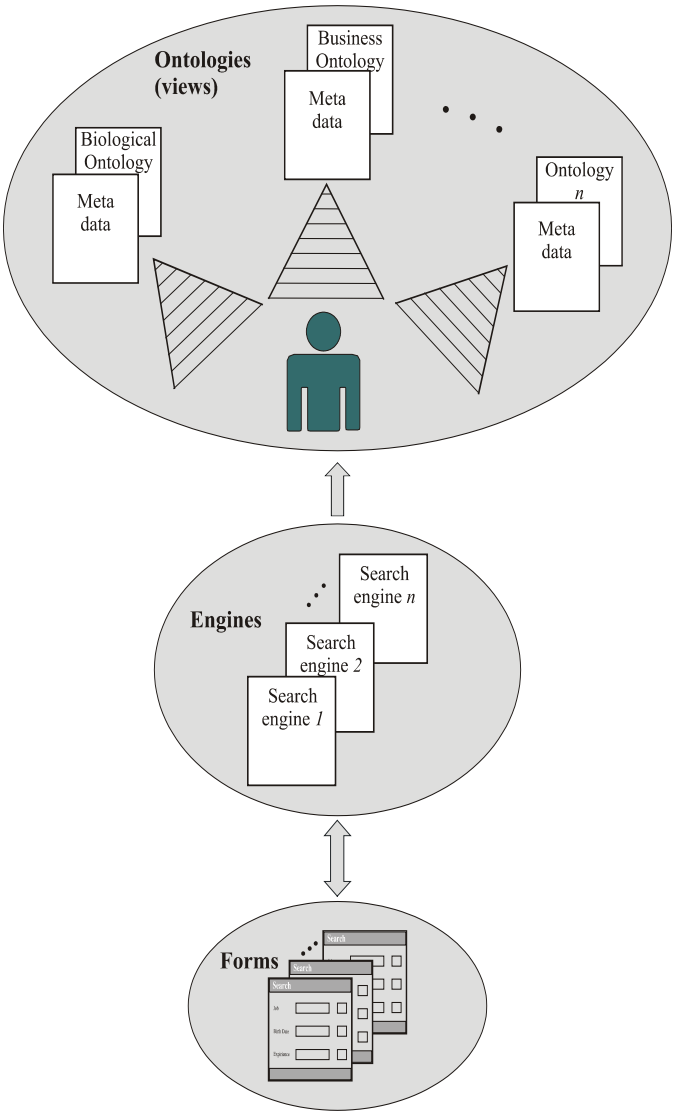

Fig. 2. A scenario for searching human resources.

Ontologies (biological, business, hobby etc.) represent various views of human being from various expects of observation. Using these ontologies, the web pages designers mark they web (home) pages. An example of embedding meta data into web pages is presented in Fig. 3. In the centralized environment, search engines cash out pages with meta data and then perform analyzing and indexing, in order to respond the user's queries. The system can be also distributed, where the role of searcher would take on specialized web agents. Meta data do not have to be known or visible to the user. They perform the searching by predefined forms, where each form corresponds on the structure of ontology.

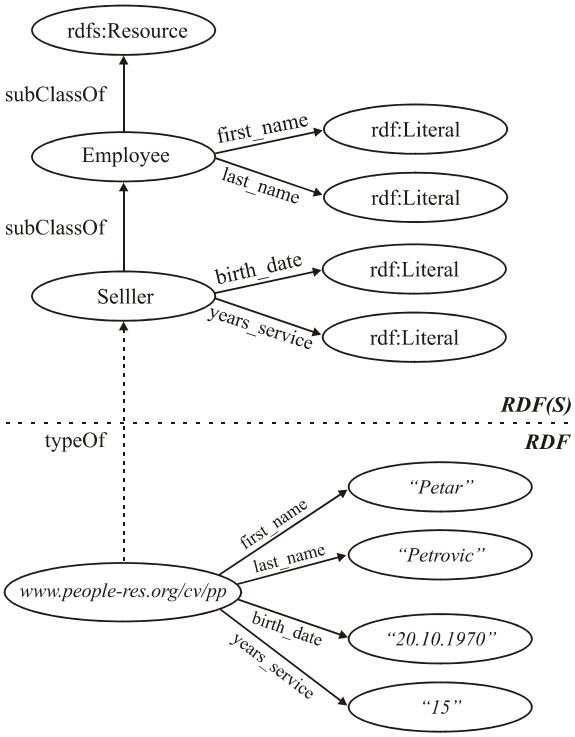

Fig. 3. Embedding meta data into web pages.

The *Employee* class has *first_name* and *last_name* properties. They are inherited in subclass *Seller*, which beside the mentioned has properties *birth_date* and *years_service*. It has to be emphasized that the approach here is to search employee categories before all, i.e. profile of employee is required, instead of person with a name, which exists anyway in the full text of a page, and it can be indexed in a classic manner. One part of RDF code which corresponds with the given example is:

```
...
xmlns=http://www.people-res.org/rdf/business#>
       <Seller rdf:about="http://www.people-
res.org/cv/pp.htm">
              <first_name>Petar</first_name>
              <last_name>Petrovic</last_name>
              <birth_date>20.10.1970</birth_date>
              <years_service>15</years_service>
              ...
       </Seller>
</rdf:RDF>
```

The user who does the searching does not have to know all the details, i.e. while searching he would be offered forms containing the list of classes (categories) and the properties from ontologies. While searching, at the same time user can use more then one ontology. In addition, the user can specify the coefficient of weight for each property, according to its significance for searching. As a result, one gets the matrix of weights:

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{m1} & w_{m2} & \dots & w_{mn} \end{bmatrix}$$

Each row of the matrix presents weights of the properties from an ontology. Using weights, the user can emphasize how much is a certain property important for its searching (for example, from 1 to 10). The property not being requested from the user or not corresponding while being searched, automatically obtains the coefficient of 0 weights. The results are ranged according to the sum of matrix weights elements:

$$S = \sum w_{ij} \quad i = 1, m;\ j = 1, n$$

, so they range from the highest to the lowest.

During implementation of searching, it can be used the ontological indexing [1], i.e. the terms from ontologies are being indexed, similarly to indexing of keywords. Another method and more promising - is using some of the RDF query language. For example, RDF query language is SPARQL which is being developed by W3C.

The exposed scenario demands considerable verses of the web page designer in the ontology's contents. While users make queries trough offered forms, without need for knowing the ontologies themselves. Due to a relatively complex organization and a large number of possible ontologies, it should be expected that the role of ontology choice, and thus the domain of searching, undertake web services that would be specialized for certain domains of searching.

B. Approximate searching

The triples *object-attribute-value* can be observed as a separated case of the semantic nets [7]. In RDF we analogously have the triples *resource-property-value* (Fig. 4).

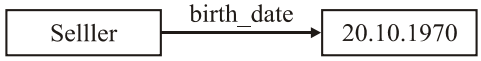

Fig. 4. An RDF triple.

A r*esource-property* relation is of the "*has"* type, and a *property-value* of the "*is"* type. This implies: Seller "has" birth_date and birth_date "is" 20.10.1970.

According to this, we can compare people. For this purpose, it can be used semantic nets metrics, where is applied the technique of conceptual distance. In this technique, the distance is calculated according to the number of arcs that are placed between the two concepts in the net. If the conceptual distance is small, the concepts are similar and vice versa (Fig. 5).

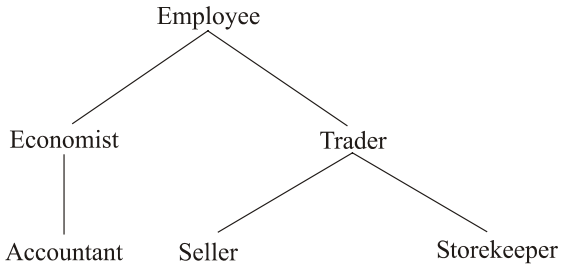

Fig. 5. A possible hierarchy of categories.

A person who is a seller is closer to a person who is a storekeeper, i.e. there is more similarity between them than with a person who works in the accountancy. The search software which is aware of this hierarchy, when there are not exact results, can retrieve approximate results. The user can specify the conceptual distance to which he wants to go, when there aren't exact hits.

V. CONCLUSION

The paper considers potential possibilities for applying the Semantic Web technologies for improving searching on the Web. The starting premises for realization of advanced semantic environment for retrieving human resources on the Web are given.

The Semantic Web is still more a vision than a real system. The result of implementation of presented scenario would be a distributed "database" of human resources. The advantage of this distributed database vs. centralized one is that every person would update the data on his Home page, without considering his geographic location. This would primarily contribute to search data more efficiently, providing approximate results when there are not exact hits. Further, it would provide more accurate data, better use of computer resources, and save cost of maintenance.

REFERENCES

[1] J. Davies, R. Weeks, "QuizRDF: Search Technology for the Semantic Web", *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004.

[2] T. *Berners*-Lee, J. Hendler, O. Lassila, "A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities", *Scientific American*, May 2001.

[3] T. B. *Pasin*, *Explorer guide to the Semantic Web*, Manning, 2004.

[4] J. *Pokorny*, "Web Searching and Information Retrieval", *Computing in Science & Engineering*, July/August 2004.

[5] A. *Fernandes*, A. C. Moura, F. Porto, "An Ontology-Based Approach for Organizing, Sharing, and Queering Knowledge Objects on the Web", *Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03)*, 2003.

[6] M. *Visciola*, "Search Types and Context of Use in the Semantic Web", *Proceedings of the Fourth International Conference on Web Information Systems Engineering (WISE'03)*, 2003.

[7] Ž. Ristić, Z. Bošnjak, N. Balaban, *Ekspertni sistemi*, Savremena administracija, Beograd, 1993.