RDF QUERY AND PROTOCOLS LANGUAGE USING FOR DESCRIPTION AND REPRESENTATION OF WEB ONTOLOGIES

1 PIÑERES-MELO, MARLON, 2 ARIZA-COLPAS, PAOLA, 3 MORALES-ORTEGA, ROBERTO, 4 AYALA-MANTILLA, CRISTIAN, 5 PELUFFO-MARTINEZ, GABRIEL, 6 MENDOZA-PALECHOR, FABIO, 7 COMBITA-NIÑO, HAROLD, 8 HERRERA-TAPIAS, BELIÑA, 9 DIAZ-MARTINEZ, JORGE.

1 Department of System Engineering, Universidad del Norte, Colombia
2,3,6,7,8,9 Assoc. Prof, Universidad de la Costa CUC, Department of Computer Science and Electronic.
4,5 Naval School ARC, Barranquilla, Colombia

E-mail: 1 pineresm@uninorte.edu.co, 2 pariza1@cuc.edu.co, 3 rmorales1@cuc.edu.co,
4 gimac.ensb@armada.mil, 5 gabriel.peluffo@armada.mil, 6 hcombital@cuc.edu.co,
7 fmendoza1@cuc.edu.co, 8 bherrera1@cuc.edu.co, 9 jdiaz5@cuc.edu.co

ABSTRACT

The purpose of this article is to expose the metadata structure based on RDF (Resource Description Framework) and the way in which queries can be made using SPARQL (Protocol and RDF Query Language), as a principle for searching the Semantic Web. It also describes what must be considered to build a Web Ontology and the tools that can help the Software developer to make queries using SPARQL.

Keywords: SPARQL, RDF, OWL, Semantic Web, metadata.

1. INTRODUCTION

With the current technological advancement, contact with the Internet is almost unavoidable for any individual, this was slowly incorporating among the other legendary media such as radio, television, the written press and telephone, exceedingly exaggerated thanks to the many benefits offered, among which it is possible to highlight the low cost of interaction between its users such as merchants, the press, etc. But it is important to highlight that this has been possible in large part to the evolution of the applications and platforms that are available within this great web.

Since its inception with the very popular HTML, which was available through the use of the HTTP protocol, to countless new applications provided by different software development firms, with different ideologies, they have managed to convert web applications into a rich mix of technologies such as: XML, Java, ASP, JavaScript, PHP, JSP, J2EE, Flash. Not counting the numerous development Frameworks, which have emerged from these, which have exponentially potentialized these benefits. Adding the dynamism of the web pages and the personalization by the end users, in the already very common access to database of different organizations remotely, they are considered one of the characteristics that show the constant evolution of the applications that reside in Internet.

Currently when a user does a search, you can do it on a database or a search engine using some keywords, based on this search engine often show results with data that have nothing to do with what you want to find. Suppose there is a writer who has a book published in an editorial and wants to ask a question to ask which books an author has published in a web browser, it is possible to show the information related to authors whose name matches a keyword.

To avoid this type of problems, the World Wide Web Consortium (W3C) has proposed new technologies that facilitate the incorporation of semantics to Web pages such as the Resource Description Framework (RDF) and its extension, such as the Resource Description Framework...
Schema (RDF-S) and the Ontology Web Language (OWL) allowing through this to model areas of knowledge in a well-structured format, thus allowing the creation of Web Ontologies.

The representation and validation of knowledge has become a highly important aspect in the processes of automatic validation of knowledge. This area of work has intensified efforts to achieve unified standards that allow for a unified form of representation. The purpose of this research article is to provide the reader with a sample of the different standards that are used for the process of representation and knowledge extraction.

A moment, but this is not the whole story as far as the evolution of the Web is concerned, because although all the technologies although they have collaborated widely in achieving a better website there are still many aspects to improve. One of the current trends that seeks to achieve this objective is the Semantic Web, which has been devised by the same promoter and creator of the Web Tim Berners-Lee, (who currently also serves as the president of the W3C consortium) this has as The purpose is to provide machines with the ability to understand and thus use what is currently available on the Web. One of the interesting approaches of this "new website" is the incorporation of intelligent agents that have the ability to navigate and save work in terms of obtaining search results on the Web, all this is achieved thanks to the creation of representations that can be understood by both users and machines, which reduces unwanted results and adds to the current web cohesion, to facilitate the location and integration of available services.

Just as this incursion of technologies has changed the way of performing many daily activities, it has also done so in terms of the way of investigating, since the Web offers immense extraordinary advantages when consulting information on a subject envisioned from different cultures around the world, but since these searches do not have cohesion, the investment of time necessary to find information relevant to the research that is intended to be carried out is high, and precisely this problem is the focus of this research.

This paper presents a description of what should be considered to build a Web Ontology and the tools that can help a Software Developer to make queries using SPARQL. It is distributed as follows: first, the features of the Resource Description Framework (RDF) are shown, second the benefits of using the Web Ontology Language (OWL), and finally SPARQL.

2. BRIEF REVIEW OF LITERATURE.

Based on the systematic review of the literature, different works have been found that strengthen the areas of knowledge processing and representation. The branch that is responsible for the processes of knowledge representation is Artificial Intelligence, to achieve this it is necessary to group situations that have similar characteristics or properties instead of making individual representations.

These types of representations must have specific characteristics among which can be detailed:

1) Have the ability to be understood by all those who supply it.
2) Facilitate the process of correction and quick visualization of these corrections.
3) Have the ability to be able to be used in different situations without being directly accurate or complete.

Among the branches of artificial intelligence focused on the analysis and representation of knowledge, the semantic web, expert systems, robotics, computer vision and machine learning can be highlighted.

The Semantic Web consists of: “An extended web, which is endowed with a greater meaning. It is developed with universal languages that will allow users to find answers to their questions in a faster and easier way thanks to the better structuring of information” [1]. That is why, with the incorporation of the semantic Web, it will allow delegating to the web applications the processing of the information resulting from user searches. Although there is a massive number of internet users today, for many the term semantic web is unknown, due to ignorance of the W3C standards that are involved in the development of it [2].

Another authors said that, “The Semantic Web is an extension of the World Wide Web in which the contents of the Web can be expressed much more than in a natural language, and also in a format that can be understood, interpreted and used by different software, allowing them to search, share and integrate information easier.” [3] Its origin emanates from WWW creator Tim Berners Lee (which has formed a consortium called W3C).
This consortium has as a goal the improvement and standardization of all languages and tools developed for the Web in May 2001, together with James Hendler and Ora Lassila, when publishing a research article called “The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities”. [4]

The massive information that is currently on the Internet has significantly hindered the fact of performing specific searches on the Web, because search engines rely on the use of keywords, to find the resources requested by different users, which throw irrelevant results [5]. That is why the Semantic Web takes advantage of all the resources of the current Web, adding automatic processing through software tools giving it greater potential.

The expert systems used in artificial intelligence are software that emulates the behavior of a human expert in solving a problem [6]. Expert systems work in ways that store specific knowledge for a given field and solve problems, using that knowledge, by logical deduction of conclusions. With them, an improvement in quality and speed of responses is sought, resulting in an improvement in the productivity of the expert.

Expert systems can be based on rules, that is, they have predefined knowledge that is used to make all decisions (applying heuristics), or based on cases (CBR, Case Based Reasoning), applying case-based reasoning, where The solution to a similar problem presented above is adapted to a new problem [7].

For true human experts it is a great support that reduces time and sometimes can perform tasks on its own. But, for an expert system to be an effective tool it must have two capabilities: on the one hand it must be possible to explain the reasoning of the expert system, on the other it must be able to integrate new knowledge as well as modify its knowledge obtained by improved ones [8].

Expert systems have been one of the great advances in artificial intelligence [9]. Among its many advantages, the fact of being able to work with a large amount of information stands out, which is more limited in human experts and can negatively affect problem solving by the latter. In addition, today, expert systems can work at very high speeds.

But not everything are advantages. There are numerous limitations or disadvantages that occur in expert systems. First, the high cost involved and the time that must be spent in programming them. In addition, to update an expert system it is necessary to reprogram it [10]. On the other hand, AI has failed to develop systems that can solve problems in a general way, of applying common sense to solve complex situations or of controlling ambiguous situations.

Robotic is a science or branch of technology, which studies the analysis, design, construction and application of robots, which can perform tasks performed by humans, mainly laborious, repetitive or dangerous [11]. The sciences and technologies supported by robotics are mainly: Mechanics, Electronics, Automatic Control and Computer Systems, which together integrate robotic systems.

There are several classifications of robots, depending on the authors, associations or institutes; Some of the main ones are: for generations, based on their intelligence, for their level of control, and others for their level of programming language [12,13].

Some of the main applications regardless of the classification are agriculture and forestry, help for the disabled, construction, domestic, dangerous environments, space, medicine and health, mining, submarine, surveillance and security, service, nuclear, medicine and construction [14,15].

Computer vision or artificial vision is the set of tools and methods that allow obtaining, processing and analyzing real-world images so that they can be treated by a computer. This allows you to automate a wide range of tasks by providing the machines with the information they need to make the right decisions in each of the tasks in which they have been assigned [16].

One of the advantages that are obtained from machine learning is that it manages to automatically distinguish patterns using mathematical algorithms [17]. These types of techniques are used for the classification of images or decision-making and, broadly speaking, they can be divided into two main types, which are supervised and those that are not supervised.

Supervised learning is a type of learning that allows the computer to learn through the patterns that have been previously provided. This is done by offering the computer patterns that have been previously labeled, that is, it offers information that it can interpret for itself [18].
On the contrary, in the case of unsupervised learning, what is carried out is a computer training using patterns that have not been previously labeled or classified. Which implies more effort but allows the computer to identify the patterns by itself in the future [19,20]. These learnings allow the detection of objects since, when the computer receives the visual information, it can relate it to the previously learned patterns and, in this way, identify the objects and differentiate them [21].

Also, having the ability to identify each of the objects, this ability can be used to carry out a video analysis [22]. This is especially useful when applying it to security and control strategies, such as the identification of specific patterns and their search within a video system.

Artificial 3D vision systems are concerned with providing a computer with a system that works in a similar way as human vision would, but in a digitalized process. In fact, these systems can mimic human vision by creating a three-dimensional model, whether it is an isolated object or a scene that is displayed together [23].

In the automotive sector, the use of industrial artificial vision systems corresponds mainly to inspections in the manufacture and assembly of all vehicle parts [24]. In addition, it should be considered that tasks such as stamping, machining, painting and burrs inspection depend directly on artificial vision equipment that supervises the entire process.

Industrial artificial vision is a fundamental element in the quality control of food products, especially in the final stages. Both the selection of fruits and the correct closure of the canned cans, or the elaboration of processed meat products, depend directly on artificial vision systems that monitor the development in all its phases [25].

3. APPLICATIONS USING WEB ONTOLOGIES

Among the software applications that make use of Web Semantic, we can highlight:

- **AKT (Advanced Knowledge Technologies) [26]**: This project was carried out by the Engineering and Physical Sciences Research Council (EPSRC) with the intervention of 5 universities in the United Kingdom. The main objective of this is to achieve integrated methods that allow the capture, modeling, publication, reuse, sharing and management of knowledge, focusing on developments in the field of AI, internet technologies, psychology, multimedia and linguistics.

- **ASG (Adaptive Services Grid) [27]**: is a solution to implement the agility and adaptability promised by service-oriented architectures. Based on the available standards, a solution for the entire life cycle provisioning service has been built. A key concept of this solution is the use of semantic information about services to automatically plan, promulgate and supervise service works to meet user requests.

- **DIP (Data, Information, and Process Integration with Semantic Web Services) [28]**: the objective has been to develop and extend the Semantic Web and Web services technologies in order to produce a new Semantic Web Services (SWS) technology infrastructure - an environment in which different web services can discover and cooperate with each other automatically. DIP's long-term vision is to deliver the enormous potential benefits of Semantic Web services for e-Work and e-Commerce.

- **KW (Knowledge Web) [29]**: The main objective of the dissemination of Web knowledge in the area of industry is to promote greater awareness and faster adoption of Semantic Web technology in Europe in full synergy with the research activity. This disclosure will help reduce the time needed to transfer the technology to the industry and the market. Dissemination to Education. Knowledge website aims to work towards the establishment of a European Association for Semantic Web Education (EASE), which will act as the focus for the educational activities of the Semantic Web.
• IMS Global: IMS GLC: [30] is a nonprofit collaboration between the world's leading educational technology providers, content providers, educational institutions, school districts and government organizations dedicated to improving education and learning through the strategic application of the technology.

• OpenKnowledge [31]: It is a system that allows peers in an arbitrarily large peer-to-peer network to interact productively with each other without any global agreements or prior knowledge at WHO runtime to interact with or how the interactions will proceed. Any type of service (for example, a WSDL service) can become a partner or what is offered in services to users by easily creating their own partners, by sharing existing code or writing their own.

4. CONCEPTUAL INFORMATION

4.1 RESOURCE DESCRIPTION FRAMEWORK (RDF)

Resource description framework, developed by w3c, whose purpose is to express the declarations [26] of resources with the form of Subject - Predicate - Object. Its syntax is based on XML.

Subject: this is what is being described, ie the resource, for example the URI http://www.w3.org/Icons/WWW/w3c_main returns the W3C logo in PNG or GIF format [32].

Predicate: is the property or relation that you want to establish about the resource. Consider that the properties can be defined and used independently of the classes [33].

Object: is the value of the property or • the other resource with which the relationship is maintained. It can be a literal or an Object. The literal includes a specific data and the object refers to another subject [34].

The graphical notation for the graphs of RDF can be summarized by the graph [1].

Figure 1 is interpreted as: the object or literal is the value of the predicate for the subject [35]. This definition can also be visualized in the form of a graph as shown in figure 6, taking into account that the arcs are the predicate, the subject is represented by means of circular nodes and the rectangular nodes correspond to the value of the object. Below are the three ways to represent RDF documents:

1. In figure 2, we can observe the data model in the form of Triplet:

   ![RDF Triplet](image)

   The interpretation of Figure 2 is: "SPARQL for Web Ontologies" is the title of the resource http://www.universidad.edu.co/ponencias

2. An example of an RDF document is illustrated in Figure 3

   ```xml
   <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
            xmlns:dc="http://purl.org/dc/elements/1.1#"
            xmlns:dc:title="http://www.universidad.edu.co/ponencias">
     <rdf:RDF rdf:about="http://www.universidad.edu.co/ponencias">
         <dc:title>SPARQL para las Ontologías Web</dc:title>
     </rdf:RDF>
   </rdf:RDF>
   ```

   ![RDF Document](image)

3. In Figure 4 its respective representation as a graph:
In Figure 5 you can see an example of model based on RDF, generated by the validation service of the W3C.

### 3.2 WEB ONTOLOGY LANGUAGE (OWL)

The language of the Web Ontologies is used to represent a domain of knowledge, additionally allows adding vocabulary for the description of classes and properties such as cardinality between classes, equality between classes, characteristics of properties such as symmetry, relations between Classes such as disjunction, primitives describe the concepts of a domain and the relationship between these concepts, based on RDF and RDF Schema [36,37].

In short, it is used to provide more vocabulary to describe classes and properties taking into account relationships between classes and characteristics of properties [38,39]. This language of Web Ontologies is derived from DAML + OIL (Darpa Agent Markup Language + Ontology Inference Layer) providing maximum expressivity, conserving computationality and resolubility. Here is an example of OWL (see figure 6):

```xml
<rdf:Description>
  <rdf:type rdf:resource="&owl:AllDisjointClasses;">
    <owl:members rdf:parseType="Collection">
      <rdf:Description rdf:about="#Articulo#"/>
      <rdf:Description rdf:about="#Editorial#"/>
      <rdf:Description rdf:about="#Persona#"/>
    </owl:members>
  </rdf:Description>
</rdf:Description>
```

Fig 6. OWL Representation
3.3 SPARQL.

Protocol and RDF Query Language [40], define a query language for RDF / RDFS, used by the engine API of inferences to search in a Web Ontology. Make queries in The Semantic Web without SPARQL is like make a query in a database without the SQL structured language. With regarding the structure of a query SPARQL [41], is based on comparison of patterns in the RDF graph. The boss’s triples are the RDF triples, but with the option of a query variable instead of an RDF term in the positions of the subject, predicate or object[42]. Combining triple patterns, in this case the subject-predicate-object, is achieved a basic triplet pattern, where it is necessary an exact comparison between the triplets. The basic syntax of a SPARQL query [6,16] is shown in the table 1.

<table>
<thead>
<tr>
<th>Basic syntax of a SPARQL query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prologue (optional)</td>
</tr>
<tr>
<td>(optional)</td>
</tr>
<tr>
<td>(repeatable)</td>
</tr>
<tr>
<td>Query Result forms (required)</td>
</tr>
<tr>
<td>of ?variable</td>
</tr>
<tr>
<td>DESCRIBE sequence of ?variable or</td>
</tr>
<tr>
<td>&lt;iri&gt;</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Query Dataset Sources (optional)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>GraphPattern (optional, required for</td>
</tr>
<tr>
<td>ASK)</td>
</tr>
<tr>
<td>Query Results Ordering (optional)</td>
</tr>
</tbody>
</table>

Next in Figure 7 an example of a query using SPARQL.

In the query of figure 7, it is to find the articles that exist in a Web ontology and based in it, it shows the title of the article, authors, keywords, date of publication, language and the name of the journal editor The query consists of formed by the following parts [35]:

- The PREFIX clause that is used to define the namespaces to use, this facilitates that you do not have to write the URI several times in the query [36]
- The SELECT clause that identifies the variables that appear in the query results.
- The WHERE clause that provides the basic graphic pattern to compare with the data chart [38].
- The ORDER BY clause, which facilitates the ascending ordering or descending.
- The LIMIT clause, which limits the maximum amount of results a to show.

The basic graphic pattern in this example consists of a simple triple pattern with the variables "? title? Author? keyword? date? language? editor "in the position of the object. The result of the query is seen below in the figure 8 and table 2:
Table 2: Pattern Definition.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation model</td>
<td>Paola Ariza</td>
<td>RDF, RDF</td>
<td>04/05/2018</td>
<td>English</td>
<td>EDUCOS TA</td>
</tr>
<tr>
<td>Web 3.0</td>
<td>Marlon Alberto Piñeres Melo</td>
<td>OWL, RDF</td>
<td>04/05/2018</td>
<td>English</td>
<td>EDUCOS TA</td>
</tr>
</tbody>
</table>

3.4 TOOL FOR THE DESIGN OF METADATA BASED IN RDF

For the design of RDF documents, Protégé was used (http://protects.stanford.edu/) as Framework open source, which facilitates the editing of ontologies and generating this kind of documents, including also RDFS, OWL, XML Schema among others. In the figure 8 you can illustrate the use of Protégé for the edition of the Ontology Web:

![Fig 7. Protegé](image)

However, there are different tools that have been developed to process and manage data from the semantic web, which are shown in Figures 10 and 11, which were detailed in the article.

![Fig 9. Query en Twinkle](image)

Fig 9. Query en Twinkle

To test the queries in the Ontology Web, Twinkle1 was used which is an application that provides an interface user graph, for motor use SPARQL query. It is used for people who want to try their queries and developers of the Semantic Web.

This tool was inspired by another called XQuisitor, which was used to make queries in XQuery (used to make queries to an XML document). Figure 9 illustrates the way in which that a query is made in the ontology Web:
This article showed the importance of the use of different tools for the representation of knowledge, which are very useful for venturing into the use of the semantic Web.

4 CONCLUSIONS

Currently, RDF, RDF-Schema, OWL and SPARQL are used significantly for the implementation of Semantic Web projects. RDF documents can be combined with standards like FOAF, Dublin Core, among others and that are accessed through queries SPARQL independent of the platform with Tools like Protégé, Twinkle among others; evidencing the importance that the Semantic Web has.

The Protégé tool is very useful for the design of Web Ontologies, because it has built-in algorithms to validate if the document of the Ontology is inconsistent or not, facilitating in this way the formal technical reviews of OWL document. For developers it will be advantageous test the SPARQL queries, if they use Twinkle, since this will allow you to evaluate the results and isolate a query from the code, allowing know if it is consistent with the expected results.

Finally, it can be said that the use of these technologies can be applied in other disciplines such as Engineering of Software, social networks, knowledge management, systems geographical, Law, among others, so the investigation is inevitable in this area, which without place doubts supports the trend of the Web 2.0 towards Web 3.0.

With this article you can conceptualize about the different tools that can be used for the use and transformation of data into information.

REFERENCES:


[34] Stanford Center for Biomedical Informatics Research. Recuperado en Octubre 14 de 2010 de: http://protege.stanford.edu/


