Ontological Interaction Using JENA and SPARQL Applied to Onto-AmazonTimber Ontology

Márcio José Moutinho da Ponte[1,2(✉)], Paulo Alves Figueiras[2],
Ricardo Jardim-Gonçalves[2], and Celson Pantoja Lima[1]

[1] Federal University of Western Pará - UFOPA, Santarém, Brazil
{marcio.ponte,celson.lima}@ufopa.edu.br
[2] Institute Development of New Technologies – UNINOVA,
NOVA University of Lisbon, Campus de Caparica, Caparica, Portugal
{paf,gr}@uninova.pt

Abstract. Knowledge representation and use are fundamental processes in many areas. The use of a semantic referential (i.e, a domain ontology and a set of related tools to exploit it) to represent knowledge has allowed the development of new mechanisms of semantic search, inferences, and analysis of complex content, but the development of a semantic referential is still a complex task, time-consuming and fundamentally performed by knowledge holders. Taking that into account this work discusses the development of a semantic referential applied to botanical identification process in the Brazilian Amazon area, mainly focused on the mechanisms of interaction and access to a domain ontology (named Onto-Amazon-Timber) based on JENA API and SPARQL queries. The main aspects of the development of this work are presented and discussed here. Current challenges and open points are also addressed.

Keywords: Ontology · Interaction · JENA · SPARQL

1   Introduction

The Knowledge Management technological platforms frequently include a Semantic Referential (SR) in order to support the knowledge life cycle (i.e, creation, formalisation, sharing, dissemination, acquisition). Essentially a SR includes a controlled vocabulary (e.g. ontology, taxonomy, and thesaurus), semantic vectors, and additional services/tools allowing the proper use of the SR [1].

The use of SR to represent knowledge on the web and in other fields has allowed the development of new inference mechanisms [2]. However, the development of a SR is a complex and lengthy process, since it requires the development of software tools, construction or adaptation of controlled vocabularies involving knowledge experts [3].

The formal representation of knowledge and its complex relationships is the goal of a SR [4]. In our work, the SR is enabled by a domain ontology providing formal and explicit specifications of shared concepts from botanical domain, aiming to capture and

to explain the vocabulary used by experts from the domain. By doing that, the SR aims to ensure a communication free of ambiguities (as much as possible) [5].

Structurally speaking, an ontology contains Concepts, Individuals, Axioms, Properties, and Relations, interrelated by semantic liaisons supporting both management and use of the knowledge hold by the ontology. The semantic liaisons pave the way to: (i) make inferences about acquired knowledge; (ii) create reasoning mechanisms to explore the richness of the existing knowledge; (iii) help detecting structural problems such as inconsistent relations, absence of concepts, individuals, or properties, due to mismanagement of the ontology.

This work aims to develop software artefacts to handle the semantic liaisons previously described based on services from JENA API (Application Programming Interface) and queries SPARQL (Protocol and RDF Query Language). The software artefacts created here are to be assessed using the Onto-AmazonTimber ontology, which holds botanical knowledge and, in our scenarios, specifically targeting the botanical identification process, bringing a collection of features and characteristics of a vast amount of forest species of the Amazon.

The paper is structured as follows. Section 2 highlights the link between the this work and Cyber-Physical Systems (CPS). Section 3 shows the related works relevant to this one. Section 4 gives a short explanation of the case studies selected to assess this work. Finally, Sect. 5 draws some conclusions and discusses the future research.

2   Contribution to Cyber-Physical Systems

Cyber-Physical Systems (CPSs) integrate the dynamics of the physical processes with those of the software and communication, providing abstractions and modelling, design, and analysis techniques for the integrated whole [6]. The interaction of computers, networking, and physical systems happens in multiple ways that require fundamentally new design technologies. The technology depends on the multi-disciplines such as embedded systems. Additionally, it is worth recalling that the software is now embedded into devices with other purposes than computation (e.g. cars, medical devices, scientific instruments, and intelligent transportation systems) [7].

Computers are integrated into the physical world in a transparent way through Embedded Systems, Real-Time Systems, and mobile computing applications. The latter is deeply characterised by highly visual processing software embedded into smartphone platforms, such as apple iOS and Google's Android. Mobile applications, commonly referred as *apps*, allow users from remote areas to have access to the whole world of ubiquous services, ranging from internet banking to health services such as image analysis and diagnosys.

Part of the work developed here is to be assessed in a smartphone-enabled scenario and, as such, both local and remote services will be available to test and validate our SR, which can be considered as a kind of Cyber-Physical System. Having said that, if on the one hand th semantic-based applications are technologically demanding, in terms of memory and performance, on the other hand mobile phones have grown from simple cell phones to highly powerful and sophisticated devices potentially able to host any sort of application in a near future, including semantic-based.

3   Related Work

The identification of botanic species is an integral part of any forest inventory, essential for forest management plan and, therefore, mandatory for commercialisation of wood. However, the usual process of botanical identification has been traditionally based on empirical knowledge coming from native experts of the forest area (Bushmen), who adopt popular names to identify and classify the species. However, such terminology normally does not match to the scientific names cataloged by taxonomists [8].

In this context, the Onto-Amazon Timber ontology (Fig. 1) is a domain ontology essentially focused on the botanic identification process of Amazonian species, exploiting pattern recognition concepts and technologies, which allow increased accuracy of the botanical identification process, thus reducing the differences of knowledge representation between Bushmen and taxonomists.
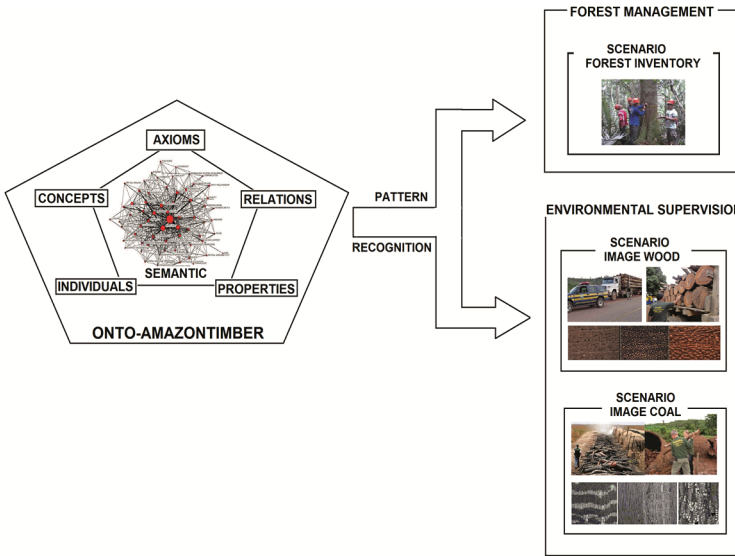

Fig. 1.   Scenarios of assessment of Onto-Amazon Timber ontology

The application scenarios (Fig. 1) are used for recognition of patterns and features extracted from the images (stored in the Onto-Amazon Timber). Those patterns and features are obtained through the use of axioms and image processing on the stored images. Such application scenarios provide inputs deemed relevant to support the decision-making process in botanic identification process, in the cases depicted in Fig. 1, namely forest inventory, inspection of transportation of legal wood, and species recognition using coal-based images.

4 Case Study Onto-AmazonTimber: Semantic Structure and Ontology Interaction

It is worth recalling that the main focus of this work is to create a common ground, semantically speaking, allowing bushmen and taxonomists to exchange knowledge without ambiguities. In other words, the Onto-Amazon Timber aimis at offering an equivalence between popular and scientific names in the botanical identification process of the amazonian species, in order to increase the accuracy of results.

The ontological Relations are formed by regular expressions that are organised in Objects and Data properties. The former links the related ontological entities, providing the required degree of semantics to build knowledge. The latter has the function to connect typed data to ontological entities in order to characterise them.

The Onto-AmazonTimber offers the following Objects Properties: *Classified by, Composed by, Formed by, Included in, It has, It has popular name, It has scope, It has synonymy*. The following expressions represent the set of Data properties in Onto-AmazonTimber: *Has image, It has function, It has measured*. The axioms are represented by semantic relationships consist of expressions, entities, literal data.

The axioms define ontological entities and the appropriate restrictions on their interpretation, allowing to infer new knowledge from the existing knowledge. The development of axioms requires the participation of experts with good knowledge from the domain of work, in order to help creating the sets of logic sentences to be inferred when handling the ontological entities. These axioms are necessary and sufficient to express these issues and characterize their solutions. Moreover, any solution to an issue of competence should be described by the axioms of ontology and should be consistent.

For illustrative purposes only, the axiom with high semantic value in the ontology, which describes the characteristics of each botanical species (Fig. 2). The entity *Specie* contains several individuals and, among them, the *Dipteryx odorata* which contains various semantic relations restrictions grouped in *Objects properties* defined by (*Dipteryx odorata included in list most marketed, Dipteryx odorata it has popular name Cumaru*) and *Data properties* defined by (*Dipteryx odorata has image "c:\imagemontologia/…"*). The latter is the link to access the image of that species, allowing the ontology to offer image processing capabilities. These axioms make it possible to infer what is the botanical species holding the characteristics expressed by the restrictions of the axioms.
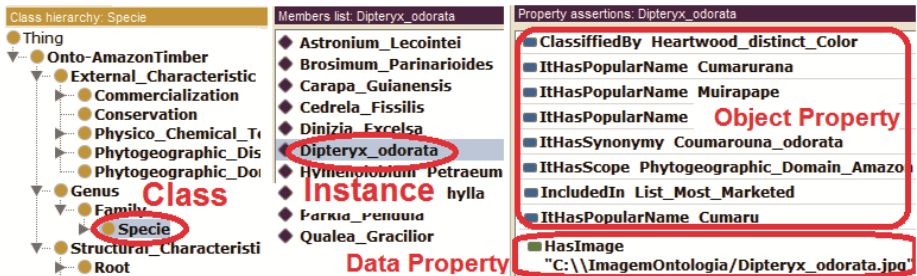

Fig. 2. The semantic structure of Onto-AmazonTimber ontology.

4.1   Ontology Interaction with JENA

JENA consists of a Java framework allowing to work on programming environment with dynamic handling of Resource Description Framework (RDF) models, represented by resources, properties and literals, forming tuples (*predicate, [subject] [object]*) that originate the objects created by Java. It presents a set of features to support application development in the context of ontologies. In addition to the features for manipulating the OWL language and use the Simple Protocol and RDF Query Language (SPARQL) [9].

The API JENA offers a set of methods giving access to the elements of a given ontology (classes, properties and individuals). Examples of methods are *listClasses ()*, *listIndividuals (), or listSubClasses (),* as shown in Table 1. Those methods call the *toList ()* method in order to get the elements through an instance of the *java.util.List* class. Additionally to that, JENA offer two basic methods allowing to identify which class or instance is being manipulated in the iterations, namely *getURI ()* and *getLocal-Name ()*. Whilst the former returns the full name or URI (prefix + name) of the object, the latter returns only the name of the given object.

Table 1.   Few methods from JENA API

Methods	Source
listClasses	OntClass concept : newM.listClasses().toList()) concept.getLocalName()
listIndividuals	Individual intance : newM.listIndividuals().toList()) intance.getLocalName()
listSubclasses	OntClass concept = (OntClass) newM.getOntClass(uri); OntClass subCon-cept : concept.listSubClasses().toList()) subConcept.getLocalName()

JENA offer a larger set of methods to access the ontological structure. For instance, the *getObjectsFromObjectTriple* method gets the list objects from a concept A that are related through a specific property with another object of concept B. The excerpt of Java code below illustrates the invocation of that method.

```
OntologyInteraction listCaracteristics = new OntologyInteraction() ;

ArrayList<String> objects = listCaracteristics.
getObjectsFromObjectTriple("Dipteryx_odorata", "ClassifiedBy");
```

In the given example, the invoked method lists the botanical characteristics of *Dipteryx_odorata* botanical species that are interconnected by *Classified By* property. Such semantic relationship may be observed in Fig. 3, where the concept Species has a series of objects including the *Dipteryx_odorata*, which presents some property of objects that create relationships with other objects, such as *Heartwood_Distinct_Color instance of Heartwood_Color class*.
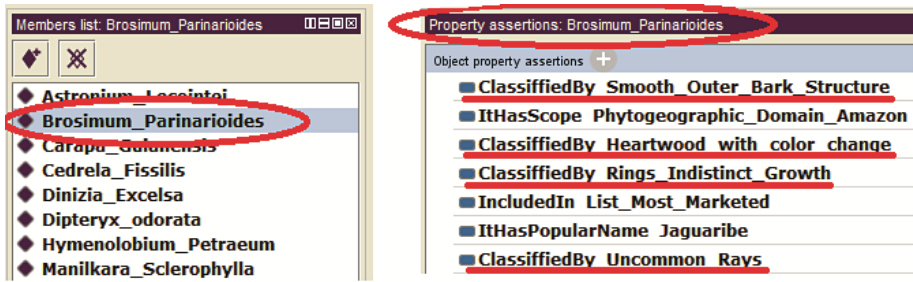
Fig. 3. Semantic relationships obtained by the *getObjectFromObjectTriple* method.

4.2   Ontology Interaction with SPARQL

Several knowledge management software applications require the integration of data from distributed, autonomous data sources. Until recently it was rather difficult to access and query data in such a setting because there was no standard query language or interface. With SPARQL [10], a W3C recommendation for an RDF query language and protocol, this situation has changed. It is now possible to make RDF data available through a standard interface and query it using a standard query language.

The RDF created from the semantic modeling presented in ONTO-AMAZON-TIMBER ontology stands as environment for SPARQL queries of this work, for illustrative purposes only, one relevant SPARQL query is showed in the source below, which aims to identify a botanical species using as criteria some features supplied by the user.

```
PREFIX rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#

PREFIX mm:<http://www.semanticweb.org/2016/ontology8#>

SELECT DISTINCT ?Especie WHERE {

        ?Species mm:IncludedIn mm:List_Most_Marketed.

        ?Species mm:itHasSynonymy mm:Coumarouna_odorata.

        ?Species mm:itHasPopularName mm:Cumaru.

        ?Species mm:itHasScope mm:Amazonia.

        ?Species mm:classifiedBy mm:Cener_Blue.}
```

Most forms of SPARQL query contain a set of triple patterns called a *basic graph pattern*. Triple patterns are like RDF triples except that subject, predicate, or object may be a variable. A basic graph pattern *matches* a subgraph of the RDF data when RDF terms from that subgraph may be replaced by the variables and the result is the RDF graph equivalent to the subgraph [11].

The consultation is a two-step process, as follows: the SELECT clause identifies the variables selected in the query represented by a question mark, another function shown is the DISTINCT whose function is to exclude clauses repeated in the answer. Other function is the WHERE acting as restrictive filter using the semantic relations of the ontology.

The basic pattern graphic in the case study is composed of several patterns of three tuples with a single variable *(?Species)* as the object. Tuples make such references as the characteristics (*List_Most_Marketed; Coumarouna_odorata; Cumaru; Amazonia; Heartwood_Bluish*) selected criterion for selection of Species, integrated by their property objects (*Included In; It Has Synonym, It Has Popular Name, It has Scope, Classified By*).

The semantic relationships referred to in this SPARQL query are depicted in Fig. 3, observing the peculiarities and distinctions of access methods, which use the botanical characteristics as variable for selection in *getObjectsFromObjectTriple* method or as selection criteria for the SPARQL method.

5   Conclusions and Future Work

This paper presented a set of software mechanisms allowing interaction and access of the domain ontology named Onto-Amazon Timbe ontology, using the JENA API integrated with SPARQL queries.

The Onto-Amazon Timber ontology has fundamentally been conceived to support the botanical identification process of Amazonian species. Therefore, a short illustrative example of semantic and conceptual structure in the domain of botany was given. The work presented here is part of the assessment of the Onto-Amazon Timber ontology which belongs to a broader context, represented by a semantic framework supported by a solid conceptual model and application scenarios, using pattern recognition and images for identification of botanical species in the Amazon.

Additionally, this work (among others) help reinforcing the potential of JENA API as a powerful tool to handle ontologies. This API offers numerous features that, combined with the Java language, accelerate and facilitate the implementation on new software tools due to the fact that Java already offer packages containing various classes and implemented interfaces and flexibility to interact with other languages such as the queries SPARQL used in this research.

Finally, it is worth emphasising the relevant role of ontologies in the current development of the semantic web. Therefore, the evolution of technologies, mechanisms, and tools to handle semantic resources is a must in this quest, such as the combination of the JENA API and SPARQL queries, presented in this paper.

Future work covers the assessment and validation of the SR here presented, including the development of new application scenarios, as well as integration with mobile devices and technologies of artificial intelligence in order to optimize the botanical identification algorithms to increase the accuracy of the results produced.

References

1. Lima, C., Zarli, A., Storer, G.: Controlled vocabularies in the European construction sector: evolution, current developments, and future trends. In: Loureiro, G., Curran, R. (eds.) Complex Systems Concurrent Engineering, pp. 565–574. Springer, London (2007)
2. Bittencourt, I.I., Isotani, S., Costa, E. Mizoguchi, R.: Research directions on semantic web and education. J Scientia **19**(1), 59–66 (2008)
3. Guarino, N.: Formal ontology and information systems. In: Proceedings of the 1st International Conference on Formal Ontologies in Information Systems. IOS Press, Trento (1998)
4. Legg, C.: Ontologies on the semantic web. Ann. Rev. Inf. Sci. Technol. **41**, 407–451 (2007)
5. Breitman, K.K.: Web semântica: a internet do futuro. LTC, Rio de Janeiro (2005). CNPq. Sala de Imprensa, 19 September 2008. http://www.cnpq.br/saladeimprensa/noticias/2008/0919e.htm. Accessed April 2009
6. http://chess.eecs.berkeley.edu/cps/
7. Zhang, F.M., Szwaykowska, K., Wolf, W., Mooney, V.: Task scheduling for control oriented requirements for Cyber-Physical Systems. In: Proceedings of the 2008 Real-Time Systems Symposium, pp. 47–56 (2005)
8. Procópio, L.C., Secco, R.S.: A importância da identificação botânica nos inventários florestais: o exemplo do—tauari (*Couratari* spp. e *Cariniana* spp. - Lecythidaceae) em duas áreas manejadas no estado do Pará, pp. 31–44. Acta Amazonica, Manaus (2008)
9. Apache.org. Apache JENA (2015). http://jena.apache.org/. Accessed November 2015
10. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF. W3C Recommendation, January 2008. http://www.w3.org/TR/rdf-sparql-query/
11. W3C – World Wide WEB Consortium. eXtensible Markup Language (XML) (2014). http://www.w3.org/XML/