Developing an Ontology to Capture Documents' Semantics

Elena Lloret, Yoan Gutiérrez and José M. Gómez
Department of Computing and Software Systems, University of Alicante, Apdo. de correos 99, E-03080 Alicante, Spain

Keywords: Ontology Development, Natural Language Processing, Semantic Package, Knowledge Engineering.

Abstract: Ontologies have been shown to be one of the best mechanisms to represent knowledge within a domain for later reasoning and inferring new knowledge that may not be initially explicitly stated. Most of the research works focus on the representation of a particular domain, thus designing and building domain ontologies (e.g. tourism, medical, etc.). However, the development of task-oriented ontologies may be more appropriate, since they can be applied to different domains, avoiding the limitation of the ad-hoc ontologies. Therefore, the goal of this paper is to present a task-oriented ontology, with the purpose of capturing the semantics of a document, in order to be used for Natural Language Processing applications, and more specifically, for the automatic generation of personalized information. The preliminary evaluation and validation of our ontology through a wide range of competence questions clearly shows its potentiality to extract the information according to specific information needs.

1 INTRODUCTION

Currently, there is a wide range of information coming from different sources that increases at an exponential rate. This is posing great challenges to users, who cannot deal with knowledge discovery or management in an efficient and effective manner, thus having to spend a lot of time manually classifying and determining what is of their interest, even though with the help of some automatic tools, such as search engines, or decision-making support systems (Sousa and Oz, 2014). Moreover, personalizing information according to users' needs is another difficult challenge as well (Raju and Babu, 2015). The same user in different moments may be interested in different topics, so intelligent systems able to take into account their preferences and interests, as well as the semantics within documents to understand them are becoming more and more relevant (Tsihrintzis and Watanabe, 2013).

In this sense, Natural Language Processing (NLP) plays a key role to knowledge discovery and management. On the one hand, research in this area normally focuses on a specific independent task, such as opinion mining (Fernández et al., 2010), word sense disambiguation (Gutiérrez et al., 2013) o text summarization (Vodolazova et al., 2013). However, in the present context, it is necessary that different NLP tasks are jointly integrated to design and develop more flexible and adaptative applications that better discover and manage knowledge, as well as deliver personalized information. On the other hand, the development of ontologies enables the development of semantic-oriented applications (Chaves et al., 2012), since by means of this type of representation, one can capture the semantics behind a domain.

One can find ontologies that model and represent the semantics of a broad range of domains, in order to further inferring and/or reasoning knowledge about that domain (e.g., tourism (Chaves et al., 2012), financial (Krieger et al., 2012), education (Ruiz et al., 2012), energy (Linnenberg et al., 2013)), and purposes (e.g., interoperability (Suca and da Silva, 2013), classification (Costa et al., 2013), or summarization (Hípola et al., 2014)). Although there is a huge variety of ontologies and lexical resources for different domains, none of them is focused on the NLP domain. This is an important limitation, because this kind of ontology will allow researchers to use specific knowledge in this field to improve their system results. In this paper, we want to fill in this gap by designing and building an NLP ontology able to capture the semantic of a document (i.e., what is the document about? what is being said in the document?), where its concepts are linked to other existent ontologies and resources, so that we could extract knowledge through the inference from the result of existing and available NLP processes, contributing to the development of more flexible semantic resources. An added value of representing the semantics of a text through an on-

Lloret, E., Gutiérrez, Y. and Gómez, J..
Developing an Ontology to Capture Documents' Semantics.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 2: KEOD*, pages 155-162
ISBN: 978-989-758-158-8
Copyright © 2015 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

tology is that it will provide deep knowledge to help not only humans, but also automatic processes to better understand the text, thus allowing also to infer and deduce new information not explicitly stated.

2 RELATED WORK

As mentioned before, there are several ontologies which represent different knowledge domains, purposes or tasks, as well as open domain ontologies developed with a very specific goal. For instance, the one developed by (Salim et al., 2010) to help users find the information that they need without worrying about the language; (Jung, 2011), who proposes a multiagent system for building indirect alignment between multilingual ontologies although the case study was realized in tourism business domain; DOLCE[1] a descriptive ontology for linguistic and cognitive engineering; UFO (Guizzardi, 2005) that provides ontological foundations for the most fundamental concepts in conceptual modeling; or SUMO (Niles and Pease, 2003), which is the largest formal public ontology that exists today[2]. Also, DBpedia (Lehmann et al., 2012)[3] is an ambitious project which tries to model Wikipedia[4] information as a machine-readable ontology. More information about different upper-level ontologies and its differences can be found in (Mascardi et al., 2006).

For building ontologies, several methodologies are available, such as BSDM (IBM, 1990), which provides the guidelines developed by IBM for modelling enterprises as a preliminary step to developing IT systems; the one proposed by Uschold and King in 1995 (Uschold and King, 1995), which is one of the most comprehensive methodologies available for building ontologies; KADS (Tansley and Hayball, 1993), a structured way of developing knowledge-based systems (expert systems); IDEF5 (KBSI, 1994), a software engineering method to develop and maintain usable, accurate, domain ontologies; and Tom Gruber's principles for ontology design (Gruber, 1995), an engineering perspective on the ontology development. However, the difficulties on building ontologies lead to the fact that many researchers use other type of knowledge representation derived from lexical and semantic resources as a way of supporting NLP processes. Among them, one of the most well-known is WordNet[5] (WN), a large lexical database of English where nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms called synsets (Miller, 1995). Other derived resources are WordNet Domains (Magnini and Cavaglià, 2000); WordNet Affect (Strapparava and Valitutti, 2004); or Semantic Classes consisting of a set of Base Level Concepts (BLC) (Izquierdo Beviá et al., 2007).

Despite WN serves as a kernel to develop different resources and applications, there are few research works that integrate them with others semantic resources. To the best of our knowledge, the Integration of Semantic Resources based on ISR-WN (Gutiérrez Vázquez et al., 2011) is the only one that takes into account 7 different lexical and semantic resources linked to WN: Level Upper Concepts (SUMO), Domains and Emotion labels, WN Domain, WN Affect, Semantic Classes and Senti WordNet (SentiWN)(Baccianella et al., 2010). ISR-WN consists of an API based on a graph-based structure in which the above mentioned resources are integrated and linked for being used as knowledge base of NLP tools. Therefore, by using these kind of NLP tools we are able to discover and classify text information that will be used for populating the ontology proposed in this work.

[1]http://www.loa.istc.cnr.it/old/DOLCE.html
[2]http://www.adampease.org/OP/
[3]http://wiki.dbpedia.org/
[4]https://www.wikipedia.org/
[5]https://wordnet.princeton.edu/

3 ONTOLOGY DESIGN AND DEVELOPMENT

Our goal is to design and develop a task-oriented ontology that captures the semantics of a document, taking into account linguistic phenomena that a text can include, and automatically populating it using the output of different NLP tools. Moreover, in the design process we also need to consider the specific type of users that will later consume the ontology. We need to conduct a well-defined and precise job in order to capture all the requirements for designing a good ontology, so therefore we rely on the methodology proposed in (Uschold and King, 1995), which includes the following general stages: i) purpose identification; ii) building the ontology (a) Ontology capture; b) Ontology coding; and c) Integrating existing ontologies); iii) evaluation; and iv) documentation.

3.1 Purpose Identification

This ontology aims to capture the semantics of documents through a set of key aspects in texts, such as the temporal dimension, presence of named entities, detection of opinionated information, or conceptual classifications. In addition, the ontology provides a lexical dimension, where the sentence of each document, and a possible summary derived from it, are taken into account. These are determining factors for setting up our own interpretation of possible scenarios (a meta-level specification) and vocabulary. Since our ontology aims to be reused by a large community, we tried to establish basic NLP terminology that was hierarchized by experts in this research field.

Regarding the intended users, we identified and characterized the range of user profiles that could use the ontology. In this manner, the users taken into account can be either NLP experts or data analysts. On the one hand, the former would be interested in making the use of the ontology more extensible by extracting multiple lexical and semantic data included in the documents from which the ontology will be populated. The combination of this information can be reused to generate further information and knowledge, not explicitly stated in the original document.

On the other hand, data analysts would exploit the whole ontology for extracting innumerable combinations of semantic queries in order to generate reports based on concurrent evidences.

The scope of our ontology will be determined by a set of competence questions, whose goal is to ensure that the ontology is able to respond to the requirements and needs for which it was developed. These questions were initially made in natural language, for instance, *what type of named entities are mentioned in documents belonging to* sports *domain?*, or *what PERSON-type entities appear in the most relevant sentences of the document (i.e., in the summary)?* (Section 5.2), and they will be checked to ensure that the requirements of our ontology's with respect to users' needs are satisfied, thus offering all the information expected by the users.

3.2 Building the Ontology

3.2.1 Ontology Capture

For identifying the key concepts and relationships in the NLP domain, we did a brain storming to collect all potentially relevant terms and phrases; at this stage, the terms alone represented the concepts of our ontology. This was the list of terms obtained taking in consideration both the inputs and outputs of the NLP tools mentioned in Table 1: *Package, Document, Summary, Sentence, Sorted Sentence, Miscellany, Organization, Person, Place, Temporal Info, Synset, Semantic Class, Sentiment Polarity, Affect , Domain, SUMO, Source, Source Type, label, body, dateTime, gloss, lemma, offset, order, url, wordnet version*. Moreover, with the aim to initially categorize the terms for inclusion, exclusion or borderline, a grouping operation was performed by using *part-of* terms' organization.

After this initial grouping, common terms were detected and some unambiguous text definitions for such concepts and relationships were accurately extracted from the NLP tools, avoiding misunderstanding among concepts names and relations, and providing a higher level of abstraction. This was performed by first carefully considering the concepts and their inter-relationships to generate more generic concepts such as *Linguistic, Lexical, Named Entity, Semantic, Category, Taxonomy Concept, Class Concept* and *Sorted Element*. From these generic concepts, our main work areas were divided into lexical and semantic, being considered the starting point for further place the remaining concepts. The former captures the knowledge about issues that are explicitly stated in the document (e.g., a person named entity, *"Rafa Nadal"*), whereas the latter deals with the ones that are not directly expressed (e.g. a positive polarity, *"Rafa Nadal win..."*).

Then, we continue identifying terms to refer to such concepts and relationships, producing and completing all definitions in all work areas. In this step, the concepts previously mentioned generate hypernymy relations *is-a* with terms of each area. On the other hand, the *part-of* relations generate different types of relations (i.e. object and data properties) mentioned in Figure 1. In addition, these *part-of* relations have been organized semantically in a hierarchy for a better understanding. It is important to remark that we also decided to use hypernymy relations (i.e. *is-a*) between individuals of *Taxonomy Concept* type. This fact allows the creation of dynamic concept trees into our ontology, as it can be seen in the *Object Property* in Figure 1. Several decisions had to be made at this step, since it may happen that different terms seemed to correspond with the same concept definition. In these cases we discarded the use of that name for the term, as it is suggested in the methodology (Uschold and King, 1995).

After several iterations analyzing and discussing the previous issues, we obtained the first version of the ontology design[6], which comprises the terms shown in Figure 1 and where the "part-of" relations can be clearly identified in the owl released.

3.2.2 Ontology Coding

In this stage, we were able to provide an explicit representation of the conceptualization captured in the

[6]http://gplsi.dlsi.ua.es/gplsi13/sites/default/files/resources/semanticpackage.owl

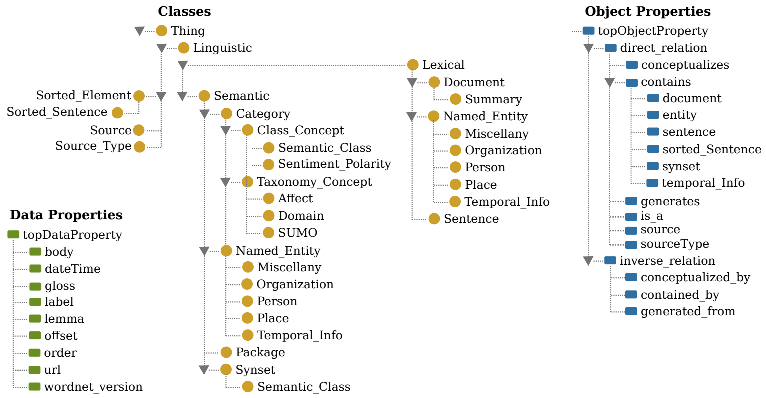

Figure 1: Concept hierarchy, object properties and data properties of our ontology.

previous stage in a formal language. In this case RDF/OWL was chosen. For the process of coding and designing Protégé editor[7] was used, since it provides a comfortable dashboard to design and develop ontologies which can be exported into different ontology languages.

3.2.3 Integrating Existing Ontologies

During either the capture and coding processes or both, there is the question of how and whether to use ontologies that already exist. In terms of reusing other shared ontologies, we would carry out an exhaustive search for further versions of the ontology so that other concepts (e.g. extracted from DBPedia.org or Schema.org) can be reused.

3.3 Evaluation

To make a technical judgement of the ontology, our evaluation has been focused on validating if the ontology meets the requirements specifications outlined by different competence questions (please see Section 5).

3.4 Documentation

The final stage in the methodology employed concerns the documentation of the ontology. For this, we assigned annotations to each concept (Classes and Properties) of the top model by using the following tags: *rdfs:label* and *rdfs:comment*.

[7]http://protege.stanford.edu/

4 TOOLS AND RESOURCES FOR TEXT PROCESSING

We selected different NLP tools that were able to detect and extract the information needed for populating our ontology. For each of the required tasks, competitive available tools with high re-using potentials were chosen as our main premise to minimize the probability of errors made by the tool, and, to avoid the time-consuming task that would be to obtain all this information manually. Table 1 summarizes the NLP tools employed (all of them working for English, and some of them also for other languages, such as Spanish).

5 EVALUATION AND RESULTS

5.1 Application Scenario

Our scenario is formed by a preliminary document, divided into two smaller subdocuments, reporting sport news, and more specifically, a news from a recent tennis match between *Rafael Nadal* and *Fabio Fognini* in the *Barcelona Open 2015* competition extracted from the BBC news Website[8].

The reason for selecting this scenario at this stage was due to the fact that this type of news are informative enough (it normally provides dates, named entities, key information of the match, etc.) to check whether our ontology could capture all their semantics, or determine what important information could be missing, and therefore, important to be considered in refined versions of the ontology. Once the document was selected, we used the output of the NLP

[8]http://www.bbc.com/sport/0/tennis/32436695

Table 1: NLP tools employed for identifying and extracting the instances for the ontology.

NLP task	Tool name	Input and Output
Semantic Analysis (Gutiérrez et al., 2011)	ISR-WN	**Input:** Text (i.e. Documents, Sentences) **Output:** Disambiguated word senses, relevant semantic classes of WN, relevant domains of WND, relevant emotions of WN-Affects, relevant categories of SUMO
Sentiment Analysis (Fernández et al., 2013)	Sentiment	**Input:** Text (i.e. Documents, Sentences) **Output:** Polarity (pos, neg, neutral)
Text Summarization (Lloret and Palomar, 2013)	GPLSI Compendium	**Input:** Text (i.e. Documents) **Output:** Most relevant sentences
Named Entity Recognizer (Finkel and Manning, 2010)	Standford NER	**Input:** Text (i.e. Documents, Sentences) **Output:** Person, location, organization, and misc named entities
Temporal Expression Recognition (Llorens et al., 2013)	TipSem	**Input:** Text (i.e. Documents, Sentences) **Output:** Date and time involved into text

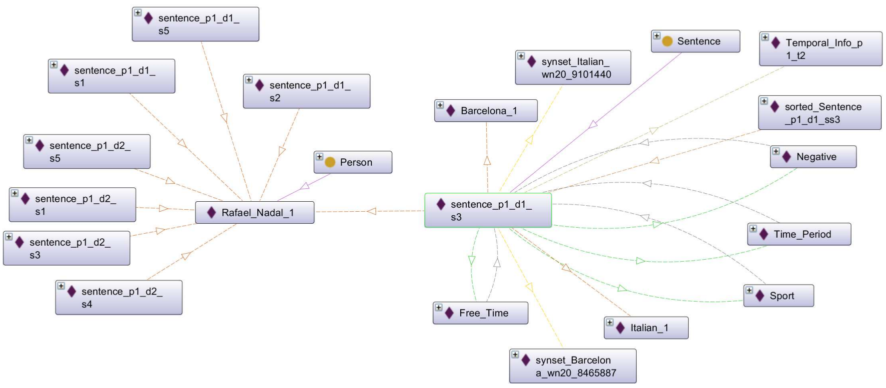

Figure 2: Example of instances in the ontology (*p: package; d: document; s: sentence*).

tools described in Section 4 to determine the instances for populating the ontology.

The population of the ontology was manually performed from the output of the NLP tools, although we plan to develop an automatic module able to directly populate the ontology from the information provided by the NLP tools.

An example of the instances of the ontology after processing the third sentence in the document through the NLP tools is shown in Figure 2. Please note that the numbers were not identified as quantities by the NLP tools used, and therefore not represented in the ontology. This sentence corresponds to: *The Italian, seeded 13th, won 6-4 7-6 (8-6) in the third round - Nadal's worst result in Barcelona since 2003, when he was 16 years old.* As it can be seen, this sentence contains several named entities (e.g. *Barcelona*), as well as additional information, such as temporal expression (i.e., *since 2003*); a negative polarity (*Nadal's worst result in Barcelona since 2003*); two WN synsets (e.g. *Italian* and *Barcelona*), being the sentence classified as belonging to the sport or freetime domains. Moreover, we can go deeper in the ontology and obtain, for instance, that named entity *Rafael Nadal*, also appears in other sentences (e.g., in the second sentence of document 1, or the fifth sentence of document 2, among other), as well as it is indicated that this named entity is classified as a person named entity, thus we can deduce that *Rafael Nadal* is a person.

5.2 Validating the Competence Questions

To verify that the ontology is able to extract the information for which it was designed and developed, a set of 30 competence questions were defined in order to determine whether the ontology could provide a correct response to these questions, thus validating its correctness. The competence questions had different degrees of difficulty, ranging from simple questions (e.g. *what PLACE named entities are in the documents?*) to more complicated ones (e.g. *which are the positive and negative sentences that talk about the sports domain?*), or even *which PERSON named entities appear in the relevant sentences of the document (i.e., in its summary)?* Moreover, they were defined taking into account the two type of users that could benefit from this ontology (data analyst and NLP expert). Our purpose here was to translate the competence questions in natural language to SPARQL questions to be executed in the ontology and assess if the

Table 2: Example of competence questions for validating the ontology.

Query:	Could I know which other types of entities appear in the same sentences as the ones mentioning Rafa Nadal?
User type:	Data Analyst
SPARQL:	Select DISTINCT ?entityExtra ?type ?polarity ?body WHERE { ?sentence rdf:type sem:Sentence; sem:body ?body; sem:entity ?entity; sem:entity ?entityExtra. ?entity rdf:type ?type. ?type ?p sem:Named_Entity. ?sentence sem:conceptualized_by ?polarity. ?polarity rdf:type sem:Sentiment_Polarity. FILTER (regex(str(?polarity), 'Negative')). FILTER (regex(str(?entity), 'Nadal') && (?entity != ?entityExtra))} GROUP BY ?entityExtra ?type ?polarity ?body ORDER BY ASC (?entity)
Result:	?entityExtra: Italian_1 ?type: Miscellany ?polarity: Negative ?entityExtra: Fabio_Fognini_1 ?type: Person ?polarity: Negative ?entityExtra: Barceona_1 ?type: Place ?polarity: Negative
Query:	Which entities of the documents *document_p1_d1* are not mentioned in the summary?
User type:	NLP expert
SPARQL:	Select Distinct ?document ?entity WHERE { ?entity ?i ?o. ?o rdfs:subClassOf sem:Named_Entity. ?package sem:document sem_inst:document_p1_d1. MINUS { ?summary ?s sem:Summary; ?p ?sorted_sentenceS. ?sorted_sentenceS sem:sentence ?sentence. ?sentence sem:entity ?entity. } } Order by ?document ?entity
Result:	?entity: Barcelona_1, ?type: Place ?entity: David_Ferrer_1, ?type: Person ?entity: Italian_1, ?type: Miscellany ?entity: Key_Nishikori_1, ?type: Person ?entity: Santiago_Giraldo_1, ?type: Person ?entity: Sao_Paulo_1, ?type: Place ?entity: Spaniard_1, ?type: Miscellany ?entity: Swede_Elias_1, ?type: Person ?entity: Barcelona_1, ?type: Place ?entity: Barcelona_1, ?type: Place ?entity: Barcelona_1, ?type: Place ?entity: 2015-04-23T00:00:00, ?type: Temporal_Info ?entity: 2003-01-01T00:00:00, ?type: Temporal_Info ?entity: 2015-01-01T00:00:00, ?type: Temporal_Info

ontology is able to provide a correct answer for each of them.

Table 2 shows two examples of questions in natural language, the user type to whom the query would be more appropriate, their SPARQL translation, and the result obtained after querying the ontology.

Concerning the results, we obtained that 96.6% of the competence questions (i.e., 29 out of 30) were correctly answered by the ontology, thus meaning that it is reliable enough for extracting personalized information depending on the users' needs. There was only one question for which the information required was not represented in our ontology. This was related to the type of questions asking for the evaluation of an element at a global level, for instance, when one wants to ask which documents the entity *X* (e.g., *Rafa Nadal*) is positively and negatively considered. To be able to respond to this type of question, a change in the ontology design would be needed, as it is analyzed and discussed in the next section (Section 6).

6 POTENTIALS AND LIMITATIONS OF THE ONTOLOGY

Although we showed that the ontology is able to capture and provide the information for which it was designed, from the analysis of the competence questions, we also realised that our ontology may have limitations for a specific type of questions, as it was previously mentioned. In this respect, the ontology is not able to directly answer questions like *what is the polarity for the entity X?* or *which documents negatively refer to the entity Y?*. This is because at this state we cannot capture multi-aspect polarity for the entities involved in a document, although we could obtain the sentences in which a specific entity is considered positive, negative or neutral and deduce the polarity of the entity from this information. To overcome this limitation, the initial ontology design should be slightly modified, introducing a new concept that would store the information regarding its evaluation (e.g., polarity evaluation). This concept should be at the top level of the ontology, in a similar way as the concept *Sorted_Element* that was introduced to be able to store the position of the sentences in the summary with respect to the original document. Nevertheless, this limitation does not affect to the expressiveness we initally wanted to reach with the ontology, and thus it could be considered as an issue for improving it.

Regarding the potentials of the ontology, we would like to stress upon the fact that despite it is not a big or complex ontology, it is able to easily determine and infer information that can be personalized depending on the users' needs, for instance, in our illustrative scenario, one may be interested in obtaining only information about the performance of *Rafa Nadal*, whereas other user could be more interested in knowing what other facts also happened in that match. Moreover, information obtained from different sources could be also related and deduced using this ontology. For example, if more documents had been tested, we could have obtained a series of facts and sentences all of them related to a specific entity, polarity, domain, etc. Note that the competency questions developed for this ontology design are generics and respond adequately to the scenario selected about a Rafa Nadal news report. In it, the specific entities involved act as variables inside SPARQL queries. So that, any other scenario can be used if linguistic elements such as document, sentence, named entity, temporal information (date references), words (for getting word sense), identification of conceptualizations (semantic classes, sentiment polarity, emotions (affects), domain, SUMO categories), and so on can be found.

One of the advantages of our proposed ontology is that, differently from other existing ontologies, this is a task-oriented ontology that captures the semantic of documents. Given that this information can be obtained independently by different NLP tools, all these outputs can be integrated in a single-ontology to maximize the exploitation and allow better reasoning processes. Since our medium-term goal is that the ontology could be also automatically populated from the output of these NLP tools, the ontology will then have another added-value, allowing that both humans or automatic processes can use the information contain to easily obtain and generate the type of information more suitable to their interests.

7 CONCLUSION AND FUTURE WORK

This paper presented the design, development and validation of a novel ontology, with the purpose of capturing the semantics of documents. The ontology was evaluated in a particular scenario (sport news) through a set of 30 competence questions translated into SPARQL, in order to determine whether the ontology was able to respond or not. The results obtained showed that all the questions, except one were correctly answered. The non-answered question was due to the fact that the ontology did not explicitly store information about the evaluation of an entity (e.g. the polarity of an entity); instead we focused on the polarity of the sentence, based on the content provided in it. After the verification performed with the competence questions, the ontology showed to be appropriate for capturing the semantics of a document, as well as it has great potential for generating personalized information, adapting the type of information to the users' or information needs.

In the future, we will first analyze and determine whether it would be necessary to modify the current ontology design to also capture some information for the entities at a global level. In the medium-term we plan to develop an automatic module to support annotations made by NLP tools, and create the instances of the ontology.

ACKNOWLEDGEMENTS

This research work has been partially funded by the University of Alicante, Generalitat Valenciana, Spanish Government and the European Commission through the projects"Explotación y tratamiento de la información disponible en Internet para la anotación y generación de textos adaptados al usuario" (GRE13-15), DIIM2.0 (PROMETEOII/2014/001), ATTOS (TIN2012-38536-C03-03), LEGOLANGUAGE (TIN2012-31224), and SAM (FP7-611312).

REFERENCES

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *LREC*. European Language Resources Association.

Chaves, M. S., de Freitas, L. A., and Vieira, R. (2012). Hontology: A multilingual ontology for the accommodation sector in the tourism industry. In *KEOD 2012 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Barcelona, Spain, 4 - 7 October, 2012*, pages 149–154.

Costa, R., Figueiras, P., Maló, P. M. N., and Lima, C. (2013). Classification of knowledge representations using an ontology-based approach. In *KEOD 2013 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Vilamoura, Algarve, Portugal, 19-22 September, 2013*, pages 184–191.

Fernández, J., Gómez, J. M., and Martínez-Barco, P. (2010). Evaluación de sistemas de recuperación de información web sobre dominios restringidos. *Procesamiento del lenguaje natural*, 45:273–276.

Fernández, J., Gutiérrez, Y., Gómez, J. M., Martınez-Barco, P., Montoyo, A., and Munoz, R. (2013). Sentiment analysis of spanish tweets using a ranking algorithm and skipgrams. *Proc. of the TASS workshop at XXIX Conference of the Spanish Society for Natural Language Processing (SEPLN 2013)*, pages 133–142.

Finkel, J. R. and Manning, C. D. (2010). Hierarchical joint learning: Improving joint parsing and named entity recognition with non-jointly labeled data. In *Proceedings of ACL 2010*.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, 43(5-6):907–928.

Guizzardi, G. (2005). *Ontological Foundations for Structural Conceptual Models*. PhD thesis, CTIT, TelematicaInstituut, Enschede, Holanda.

Gutiérrez, Y., Castaneda, Y., González, A., Estrada, R., Piug, D. D., Abreu, J. I., Pérez, R., Orquın, A. F., Montoyo, A., Muñoz, R., and Camara, F. (2013).

UMCC DLSI: Reinforcing a ranking algorithm with sense frequencies and multidimensional semantic resources to solve multilingual word sense disambiguation. *Proc. of SemEval*, pages 241–249.

Gutiérrez, Y., Orquín, A. F., Vázquez, S., and Montoyo, A. (2011). Enriching the integration of semantic resources based on wordnet. *Procesamiento del Lenguaje Natural*, 47:249–257.

Gutiérrez Vázquez, Y., Fernández Orquín, A., Montoyo Guijarro, A., and Vázquez Pérez, S. (2011). Enriching the integration of semantic resources based on WordNet. *Procesamiento del Lenguaje Natural*, 47:249–257.

Hípola, P., Senso, J. A., Leiva-Mederos, A., and Domínguez-Velasco, S. (2014). Ontology-based text summarization. the case of texminer. *Library Hi Tech*, 32(2):229–248.

IBM (1990). Introduction to business system development method. Technical report, IBM.

Izquierdo Beviá, R., Suárez Cueto, A., and Rigau Claramunt, G. (2007). A proposal of automatic selection of coarse-grained semantic classes for WSD. *Procesamiento del Lenguaje Natural*, 39:189–196.

Jung, J. J. (2011). Exploiting multi-agent platform for indirect alignment between multilingual ontologies: A case study on tourism business. *Expert Systems with Applications*, 38(5):5774–5780.

KBSI (1994). Knowledge based systems incorporated. Technical report, Wright-Patterson Air Force Base, Ohio.

Krieger, H., Declerck, T., and Nedunchezhian, A. K. (2012). Mfo-the federated financial ontology for the MONNET project. In *KEOD 2012 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Barcelona, Spain, 4 - 7 October, 2012.*, pages 327–330.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., and Bizer, C. (2012). DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia — www.semantic-web-journal.net. *Semantic Web Interoperability, Usability, Applicability an IOS Press Journal Search form*, 1(5):1–29.

Linnenberg, T., Mueller, A. W., Christiansen, L., Seitz, C., and Fay, A. (2013). Ontoenergy - A lightweight ontology for supporting energy-efficiency tasks - enabling generic evaluation of energy efficiency in the engineering phase of automated manufacturing plants. In *KEOD 2013 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Vilamoura, Algarve, Portugal, 19-22 September, 2013*, pages 337–344.

Llorens, H., Saquete, E., and Navarro-Colorado, B. (2013). Applying semantic knowledge to the automatic processing of temporal expressions and events in natural language. *Information Processing & Management*, 49(1):179 – 197.

Lloret, E. and Palomar, M. (2013). COMPENDIUM: a text summarisation tool for generating summaries of multiple purposes, domains, and genres. *Natural Language Engineering*, 19(2):147–186.

Magnini, B. and Cavaglià, G. (2000). Integrating subject field codes into wordnet. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece.

Mascardi, V., Cord, V., and Rosso, P. (2006). A comparison of upper ontologies. Technical report, Dipartimento di Informatica e Scienze dellInformazione.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

Niles, I. and Pease, a. (2003). Mapping WordNet to the SUMO ontology. *Proc of the IEEE International Knowledge Engineering conference*, pages 1–7.

Raju, Y. and Babu, D. S. (2015). A novel approaches in web mining techniques in case of web personalization. *International Journal of Research in Computer Applications and Robotics*, 3(2):6–12.

Ruiz, L. M. G., Hermida, J. M., and Montoyo-Guijarro, A. (2012). A new proposal for learning objects quality evaluation in learning strategies based on ontology for education. In *KEOD 2012 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Barcelona, Spain, 4 - 7 October, 2012.*, pages 373–376.

Salim, J., Hashim, S. F. M., and Aris, A. (2010). A framework for building multilingual ontologies for Islamic portal. In *2010 International Symposium on Information Technology*, volume 3, pages 1302–1307. IEEE.

Sousa, K. and Oz, E. (2014). *Management Information Systems*. Cengage Learning.

Strapparava, C. and Valitutti, A. (2004). WordNet-Affect: An affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086.

Suca, E. G. and da Silva, F. S. C. (2013). An ontology for portability and interoperability digital documents - an approach in document engineering using ontologies. In *KDIR/KMIS 2013 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval and the International Conference on Knowledge Management and Information Sharing, Vilamoura, Algarve, Portugal, 19 - 22 September, 2013*, pages 373–380.

Tansley, S. and Hayball, C. C. (1993). Knowledge-based systems analysis and design - a kads developer's handbook. pages I–XV, 1–528.

Tsihrintzis, G. and Watanabe, T. (2013). *Intelligent Interactive Multimedia Systems and Services: Proceedings of the 6th International Conference on Intelligent Interactive Multimedia Systems and Services (IIMSS2013)*. Frontiers in Artificial Intelligence and Applications. IOS Press.

Uschold, M. and King, M. (1995). *Towards a methodology for building ontologies*. Citeseer.

Vodolazova, T., Lloret, E., Muñoz, R., and Palomar, M. (2013). Extractive text summarization: Can we use the same techniques for any text? In *Natural Language Processing and Information Systems*, pages 164–175.