A Survey of Ontology Learning from Text

Kaoutar Belhoucine and Mohammed Mourchid
MISC laboratory, Faculty of Science
Ibn Tofail University
Kenitra, Morocco
e-mail: kaouar.belhoucine@gmail.com, mourchidm@hotmail.com

Abstract—**Ontologies are considered to be a major solution to semantic interoperability in modern information systems. The explosion of textual information on the Web and advanced state in related fields, such as Natural Language Processing (NLP), information retrieval, and data mining, have made (semi-) automatic ontology learning from text a particularly promising research area. This article summarizes the state-of-the-art in ontology learning from text, and discusses the research questions and challenges that remain in this field.**

Keywords-Ontology Learning from Text; Ontology Learning Layer Cake Model; Ontology Evaluation; Trends; Challenges.

I. INTRODUCTION

Ontologies constitute an approach for knowledge representation that defines concepts and their relationships, constraints, axioms, and the vocabulary of a given domain. An ontology should be machine understandable (which excludes natural language), and should capture the consensual knowledge, that is not private to an individual, but accepted by a group as committee of practice.

Ontologies are of great importance to modern knowledge-based systems. By providing a shared schema, they facilitate query answering and reasoning over disparate data sources. However, the manual construction of ontologies is a difficult and expensive task that usually requires a collaboration between domain experts and skilled ontology engineers. Even then, once the ontology has been constructed, our evolving knowledge and updated application requirements demand a process of continuous maintenance on the ontology.

This difficulty in capturing the knowledge required by knowledge-based systems is called "knowledge acquisition bottleneck". To overcome this bottleneck, an automatic or semi-automatic support for ontology construction is desired. This area of research is usually referred to as ontology learning [1]-[3].

We present in this paper a survey of ontology learning from text. Section 2 introduces the ontology concept as it is considered in this discipline. Section 3 discusses the overall process of ontology learning from text: inputs, approaches, techniques, and prominent ontology learning systems. Evaluation methods for ontology learning are discussed in Section 4. Finally, Section 5 concludes with a final discussion on the contemporary trends and remaining challenges in the field.

II. ONTOLOGIES

Before defining the process of ontology learning from text, we must first clarify what we mean by the term "ontology." The term "ontology" comes from the branch of philosophy that is concerned with the study of being or existence. However, within the discipline of Artificial Intelligence, scholars, such as T. Gruber define an ontology as a formal specification of the concepts of the domain of interest, where their relationships, constraints, and axioms are expressed, thus defining a common vocabulary for sharing knowledge [4]. Indeed, these two interdisciplinary definitions are complementary; what must be represented in a knowledge-based system is what exists. In other words, an ontology is composed of two parts; the first part consisting of concepts, taxonomic relations (relations which define a conceptual hierarchy) and of the non-taxonomic relations between them. Further, the other part is constructed of conceptual instances and assertions about them. More formally, an ontology can be defined, according to [5][6], as a tuple:

$$\vartheta := (C, H^C, R, rel, A^\vartheta). \quad (1)$$

Where:

- C is the set of ontology concepts. The concepts represent the entities of the domain being modeled. They are designated by one or more natural language terms and are normally referenced inside the ontology by a unique identifier.
- $H^C \subseteq C \times C$ is a set of taxonomic relationships between the concepts. Such relationships define the concept hierarchy.
- R is the set of non-taxonomic relationships.
- The function rel: $R \rightarrow C \times C$ maps the relation identifiers to the actual relationships.
- A^ϑ is a set of axioms, usually formalized into logic language. These axioms specify additional constraints on the ontology and can be used in ontology consistency checking, as well as inferring new knowledge from the ontology through an inference mechanism.

Besides these elements, there are also the instances of the concepts and relationships, e.g., the instances of the

 ISBN: 978-1-61208-678-1

elements of C, H^C and R. A knowledge base is composed by an ontology ϑ and its instances.

III. Ontology Learning from Text

Ontology learning from text refers to the (semi)-automatic support for identifying concepts, relations, and (optionally) axioms from textual information and using them to first construct and, then, maintain an ontology. Techniques from established fields, such as information retrieval, data mining, and NLP, have all been fundamental in the development of ontology learning systems. This section examines the input used to learn ontologies, their learning approaches, their techniques, and the most prominent ontology learning systems.

A. The input used to learn ontologies

Ontology learning requires input data from which to learn the concepts relevant for any given domain and their definitions, as well as the relationships between them. Dominik Benz [7] defines three different kinds of ontology learning input data:

- Structured data: means data represented according to defined schema such as Database (DB) schemes, existing ontologies and knowledge bases.
- Semi-structured data: designates the use of some mixed structured data with free text, for example: dictionaries such as WordNet [9] or the Wiktionary [10], HTML and XML documents or Wikis and User Tags.
- Unstructured data: consists of natural language texts such as Word and PDF documents, or Web pages.

The term ontology learning from text is used if ontology learning is based on unstructured data [23]. This type of resources is the most available format as input for ontology learning processes. They reflect mostly the domain knowledge for which the user is building the ontology. In addition, they describe the terminology, concepts and conceptual structures of the given domain. However, some authors, such as M. Rogger et al. [11], consider that processing unstructured data is the most complicated problem because most of the knowledge is implicit and allows conceptualizing it by different people in different ways, even using the same words. For these reasons, this paper focuses especially on ontology learning from unstructured data.

B. Ontology learning approaches and techniques

As we have shown in the previous sections, ontology learning is primarily concerned with definition of concepts, relations, and (optionally) axioms from textual information and using them to construct and maintain an ontology. Although there is no standard regarding this development process, P. Cimiano [13] describes the tasks involved in ontology learning as forming a layer cake. As illustrated in Figure 1, the cake is composed, in ascending order, of terms, synonyms, concepts, taxonomies, relations, and, finally, axioms and rules. We shall now examine this cake layer by layer and present the different approaches and techniques used.

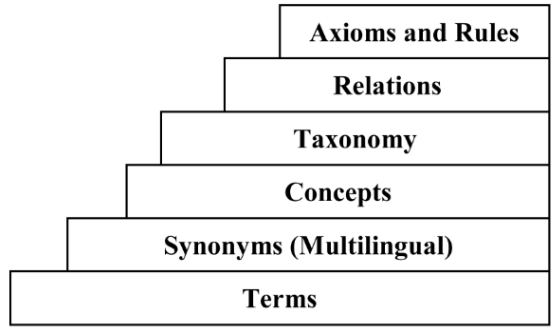

Figure 1. Ontology Learning "Layer Cake" [13].

1) Terms

Terms are the most basic building blocks of the ontology learning cake. Terms can be simple (i.e., single word) or complex (i.e., multi-word) and are considered as linguistic realizations of domain-specific concepts. There are many term extraction methods in ontology learning from text. Most of these extraction methods are based on terminology and NLP research [14]-[16], whilst others are based on information retrieval methods for term indexing [17]. The leading approaches of term extraction use tokenization (or part-of-speech tagging of the domain corpus) to identify terms by manually constructing ad-hoc patterns. Additionally, in order to identify only relevant term candidates, a statistical processing step may be used to compare the distribution of terms between domain specific and general corpora.

2) Synonyms

The synonyms layer addresses the acquisition of semantic term variants in and between languages. It is either based on sets, such as WordNet synsets [18] (after sense disambiguation), on clustering techniques [19]-[22] or other similar methods, including Web-based knowledge acquisition.

3) Concepts

Concepts can be abstract or concrete, real or fictitious. However, the consensus in this field is that concepts should include:

- Intension: formal definition of the set of objects that this concept describes.
- Extension: a set of objects that the definition of this concept describes.
- Lexical realizations: a set of linguistic realizations, (multilingual) terms for this concept.

Most of the research in concept extraction addresses the question from a clustering perspective, regarding concepts as clusters of related terms [13]. Obviously, this approach overlaps almost entirely with that of term and synonym extraction [23] and can be found in [24]-[27].

Alternatively, researchers have also addressed concept formation from an extensional point of view. For example, in the approach of [28][29], they derive hierarchies of

Copyright (c) IARIA, 2018. ISBN: 978-1-61208-678-1

named entities from text whilst also ascertaining concepts from an extensional point of view [13].

4) Concept Hierarchies (Taxonomy)

There are currently three main paradigms to induce concept hierarchies from textual data:

- The first one is the application of lexico-syntactic patterns to detect hyponymy relations, as proposed by [30]. However, it is well known that these patterns occur rarely in corpora. Consequently, though approaches relying on lexico-syntactic patterns have a reasonable degree of precision, their recall is very low.
- The second paradigm is based on Harris's distributional analysis [31]. In this paradigm, researchers have exploited clustering algorithms to automatically derive concept hierarchies from text.
- The third paradigm stems from the information retrieval community and relies on a document-based notion of term subsumption, as proposed for example in [32].

5) Relations (non-hierarchical)

Non-hierarchical relation extraction from text has been addressed primarily within the biomedical field, as there are a large text collections readily available for this area of research (e.g., PubMed [70]). The goal of this work is to discover new relationships between known concepts (i.e., symptoms, drugs, diseases, etc.) by analyzing large quantities of biomedical scientific articles (see e.g., [33]-[35]). Relation extraction through text mining for ontology development was introduced in work on association rules in [36]. Recent efforts in relation extraction from text have been carried on under the Automatic Content Extraction (ACE) program, where entities (i.e., individuals) are distinguished from their mentions. Normalization, the process of establishing links between mentions in a document, and individual entities represented in an ontology, is part of the task for certain kind of mentions (e.g., temporal expressions).

6) Axioms and rules

The extraction of rules from text occurs at an early stage [37]. Initial blueprints for this task can be found in the work of [38]. This work used an unsupervised method for discovering inference rules from the text, which was based on an extended version of the Harris' distributional hypothesis. Furthermore, the European Union-funded project Pascal [39] on textual entailment challenge has strongly increased the awareness of the problem of deriving lexical entailment rules. The focus of Pascal, therefore, was to learn lexical entailments for application in question answering systems.

C. Prominent systems

Several ontology learning systems have been proposed with the goal of reducing both the time and cost for ontology development. We present in this section the most prominent ontology learning systems according to the following criteria: broad adoption or popularity, completeness in the number of ontology learning tasks and outcomes, or recency of work.

- ASIUM [40] is a semi-automated ontology learning system that learns subcategorization frames of verbs and ontologies from syntactic parsing of technical texts in natural language (French). ASIUM successively aggregates the clusters to form new concepts in the form of a generality graph that represents the ontology of the domain.
- Text-to-Onto [21] is a framework for semi-automatic ontology learning from texts which implements a variety of algorithms for diverse ontology learning subtask. It leverages data mining and NLP techniques in the ontology development and maintenance task. It proceeds through ontology import, extraction, pruning, and refinement.
- SYNDIKATE [42] is a system for automatically acquiring knowledge from real-world texts, and for transferring their content to formal representation structures which constitute a corresponding text knowledge base. SYNDIKATE uses only linguistics-based techniques to perform its ontology learning tasks.
- OntoLearn [43] is a system for (semi-)automated ontology learning from domain texts. OntoLearn uses text mining techniques and existing linguistic resources, such as WordNet [9] and SemCor [69] to learn, from available document warehouses and dedicated Web sites, domain concepts and taxonomic relations among them.
- CRCTOL [44], known as Concept-Relation-Concept Tuple-based Ontology Learning, is a system to mine ontologies automatically from domain specific documents. CRCTOL uses linguistics and statistics-based techniques to perform its ontology learning tasks.
- OntoGain [22] is a system for unsupervised ontology acquisition from unstructured text which relies on multi-word term extraction. OntoGain uses linguistics and statistics-based techniques to perform its ontology learning tasks.
- OntoCmaps [58] is a domain-independent and ontology learning tool that extracts deep semantic representations from corpora. OntoCmaps generates rich conceptual representations in the form of concept maps and proposes an innovative filtering mechanism based on Degree (number of edges from and to a given term), Betweenness (number of shortest paths that pass through a term), PageRank (fraction of time spent visiting a term) and Hits (ranks terms according to the importance of hubs and authorities) metrics from graph theory.
- LexOnt [59] is a semi-automatic ontology creation tool that uses the Programmable Web directory of services. Its algorithm generates and ranks frequent terms and significant phrases by comparing them to external domain knowledge such as Wikipedia, WordNet and

the current state of the ontology. LexOnt constructs the ontology iteratively, by interacting with the user. The user can choose, add these terms to the ontology and rank terms.

Table I provides a comparison of the inputs used, outputs supported, and techniques employed by the prominent ontology learning systems from text.

TABLE I. SUMMARY OF PROMINENT ONTOLOGY LEARNING SYSTEMS FROM TEXT

System	Input Language	Input Type	Output	Technique		
				Linguistics-based	*Statistics-based*	*Logic-based*
ASIUM (2000)	French	Unstructured (corpora)	Terms	Sentence parsing, Syntactic Structure analysis, Subcategorization frames		
			Concepts			
			Taxonomic relations		Agglomerative Clustering	
Text-to-Onto (2000)	German, XML, HTML, Document Type Definition (DTD)	Natural language texts, Web docs, semi-structured (XML, DTD) and structured (DB schema, ontology) data	Terms	Part-of-speech tagging, Sentence parsing, Syntactic Structure analysis		
			Concepts	Concepts from domain lexicon	Co-occurrence analysis	
			Taxonomic relations	Hypernyms from WordNet, Lexico-syntactic patterns	Agglomerative Clustering	
			Non-taxonomic relations		Association rule mining	
SYNDIKATE (2001)	German	Unstructured text	Terms	Syntactic Structure analysis, Anaphora resolution		
			Concepts	Use of semantic templates and domain knowledge		Inference engine
			Taxonomic relations			
			Non-taxonomic relations			
OntoLearn (2002)	French	Unstructured/semi structured text	Terms	Part-of-speech tagging, Sentence parsing	Relevance analysis	
			Concepts	Concepts and glossary from WordNet		
			Taxonomic relations	Hypernyms from WordNet		
CRCTOL (2005)	English	Unstructured/semi structured text (WordNet)	Terms and Concepts	Part-of-speech tagging, Sentence parsing, use of domain lexicon, Word sense disambiguation	Relevance analysis	
			Taxonomic and Non-taxonomic relations	Lexico-syntactic patterns, Syntactic Structure analysis		
OntoGain (2010)	English	Unstructured/semi structured text (WordNet)	Terms and Concepts	Part-of-speech tagging, Shallow parsing, Relevance analysis		
			Taxonomic relations		Agglomerative Clustering, Formal concept analysis	
			Non-taxonomic relations		Association rule mining	
OntoCmaps (2011)	English	Unstructured/semi structured text	Terms	Part-of-speech tagging and syntactic patterns based on dependency grammar formalism	Relevance analysis	
			Concepts			
			Taxonomic relations			
			Non-taxonomic relations			
LexOnt (2012)	English	Unstructured, semi-structured (Wikipedia, WordNet) and structured (ontology) data	Terms	Linguistic patterns to determinate collocations	Relevance analysis	
			Taxonomic relations			

 ISBN: 978-1-61208-678-1

As shown in Table I, most of the existing ontology learning systems focus only on concept and relation extraction. They generally rely on shallow NLP techniques and statistical methods. Though these systems are able to address the requirements of constructing small ‘toy’ ontologies, in time, the need for researchers to return to the basics and address more fundamental issues about knowledge acquisition bottleneck is revealed. This explains the reduction in the number of complete ontology learning systems developed in the last few years.

IV. Ontology Evaluation

“Ontology evaluation is defined in the context of two interesting concepts; verification and validation. The definition is interesting because it also offers a way to categorize current ontology evaluation endeavors. Ontology verification is concerned with building an ontology correctly, while ontology validation on the other hand is concerned with building the correct ontology” [61].

A. Evaluation approaches

A variety of approaches to ontology evaluation have been proposed in the literature [61][62][47]. Depending on the kind of ontology and the purpose of the evaluation, these approaches can be grouped into the following categories.

1) Gold Standard-based evaluation

Attempts to compare the learned ontology with a predefined gold standard ontology that represents an idealized outcome of the learning algorithm. However, having a suitable gold ontology can be challenging, since it should be one that was created under similar conditions with similar goals to the learned ontology [62].

2) Task-based evaluation

Examines how the results of the ontology-based application are affected by using the ontology [45]. For example, in the case of an ontology designed to improve the performance of document retrieval, users may collect some sample queries and determine if the documents retrieved are more relevant when the ontology is used.

3) Corpus-based evaluation

Evaluates how far an ontology is able to cover any given domain [45]. This type of approach compares the learned ontology with the content of a text corpus that significantly covers the corresponding domain. Techniques from natural language processing or information extraction are used to analyze the content of the corpus.

4) Criteria-based evaluation

Measures to what extent an ontology adheres to certain desirable criteria. We can distinguish between measures related to the structure of an ontology and more sophisticated measures [62].

B. Evaluation tools

Since the OntoWeb 2 position statement stressed the insufficient research on ontology evaluation and the lack of evaluation tools [48], several ontology evaluation tools have been developed. They differ according to the context of the evaluation. We present the most important examples below [12]:

- Swoogle [52] is an ontology search engine that offers a limited search facility that can be interpreted as topic coverage. Given a search keyword, Swoogle can retrieve ontologies that contain a class or a relation that (lexically) matches the given keyword.
- OntoKhoj [53] is an ontology search engine that extends the traditional (keyword-based search) approach to consider word senses when ranking ontologies covering any given topic. It accommodates a manual sense disambiguation process, then, according to the sense chosen by the user, hypernyms and synonyms are selected from WordNet.
- OntoQA [54] is a tool that measures the quality of ontology from the consumer perspective, using schema and instance metrics. It takes as an input a crawled populated ontology or a set of user supplied search terms, and ranks them according to metrics related to various aspects of an ontology.
- OntoCAT [55] provides a comprehensive set of metrics for use by the ontology consumer or knowledge engineer to assist in ontology evaluation for re-use. This evaluation process is focused on the ontology summaries that are based on size, structural, hub, and root properties.
- AKTiveRank [56] is a tool that ranks ontologies using a set of ontology structure-based metrics. It processes keywords as an input, and queries Swoogle for the given keywords in order to extract candidate ontologies. After that, it then applies measures based on the coverage and the structure of the ontologies to rank them accordingly. Its shortcoming is that its measures are at the “class level” only.
- OS_Rank [57] is an ontology evaluation system that evaluates ontologies and ranks them based on class name, the degree of detail for each searched class, the number of semantic relations of searched classes, and the interest domain based on WordNet to resolve different semantic problems.
- OOPS! (OntOlogy Pitfall Scanner!) [60] is a tool that scans ontologies looking for potential pitfalls that could led to modeling errors. OOPS! is very useful for ontology developers during the ontology validation activity, concretely during the diagnosis phase. The tool operates independently of any ontology development platform.

V. Ontology Learning Trends and Problems

To summarize the progress and trends that the ontology learning community has witnessed over the past years, we sent queries to Google Scholar, relating to ontology learning and compared the number of returned publications from 2007 to 2017. Some of our results are shown in Figure 2.

 ISBN: 978-1-61208-678-1

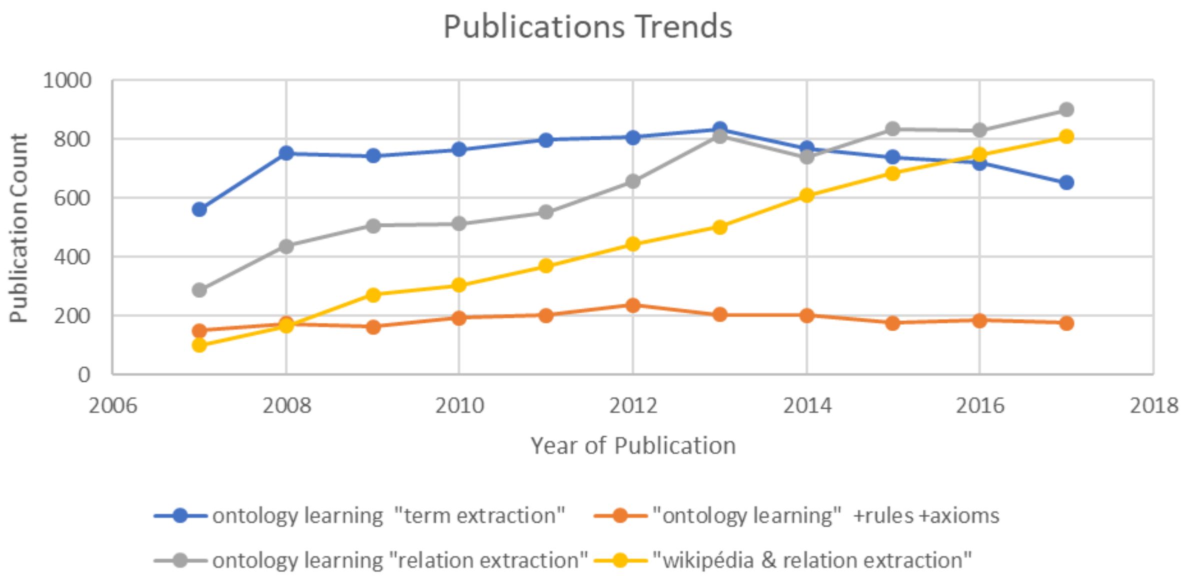

Figure 2. Publications Trends.

We browsed a large number of research papers that were returned. As a result of our research, we have observed the following trends:

- The most recent literature points to an increase in interest in using Web data to address the knowledge acquisition bottleneck and to make ontology learning operational on a Web scale.
- Current research efforts are focused on either enhancing existing term recognition techniques or moving to the more advanced phase of relation discovery.
- The measures of terms extraction from texts have more or less stabilized, with an F-measure generally above 90%. The current state-of-the-art techniques are based mainly on statistical semantics, and paradigmatic and syntagmatic relations [63] - that is to say, the relevance of search terms is determined through general observations in very large samples of data and through the way the constituent parts of the search term are constructed.
- There is a noticeable trend of increased application of lexico-syntactic patterns [64], machine learning methods [65], or hybrid approach that combines lexico-syntactic pattern analysis with supervised classification [66][67] for taxonomic and non-taxonomic relation discovery on very large datasets from the Web. The relative redundancy of Web data has allowed this group of techniques that rely on repetitions and regularities to be revived and flourish.
- (Semi)-structured Web data, such as Wikipedia [68] and Freebase [41], have become a necessary part of emerging work for relations discovery.
- Efforts are not being towards the development of new ontology learning tools, but instead towards the improvement of existing ones: increasing in automation, precision, recall and F-measure.

We have also identified the following open issues:

- The fully automatic learning of ontologies may not be possible, considering that an ontology is, after all, a shared conceptualization of a domain.
- The results for discovery of relations between concepts is less than satisfactory.
- The axiom learning from text is currently in the early stages of development.
- There is a lack of reusable services for ontology learning. Many proposed ontology learning methods and approaches highly depend on their specific environment consisting of language, domain, application and input.
- A common evaluation platform for ontologies is currently absent, but is needed.

 ISBN: 978-1-61208-678-1

VI. Conclusion and Future Challenges

This work presented a survey of ontology learning from text. For this intent, we have identified the ontology learning tasks, and introduced, the most used techniques to perform each task. Further, we have provided a comparison table of ontology learning systems, a brief overview of ontology evaluation, and summarized the current trends and open problems in this field. In addition to these problems, the growing use of Web data will introduce new challenges. Firstly, research efforts increasingly be dedicated to creating new, or adapting existing techniques to work with the noise, richness, and diversity of Web data. Secondly, the amount of Web data, which is growing exponentially, will be a significant challenge which merits further attention in the future. Questions of efficiency and robustness in processing data will be at the forefront of this challenge. Thirdly, as more communities of different cultural and linguistic backgrounds contribute to the Web, the availability of textual resources required for multilingual ontology learning will improve. Lastly, as the availability of ontologies increases, ontology alignment will become more pertinent.

References

[1] P. Cimiano, J. Volker, and R. Studer, “Ontologies on demand? - a description of the state-of-the-art, applications, challenges and trends for ontology learning from text”, Information, Wissenschaft und Praxis, pp. 315-320, 2006.

[2] B. Omelayenko, “Learning of ontologies for the Web: the analysis of existent approaches”, In: Proceedings of the International Workshop on Web Dynamics, pp. 16-25, 2001.

[3] V. Novacek, “Ontology learning”. Master’s thesis, Faculty of Informatics, Masaryk University, Czech Republic, is.muni.cz, 2006.

[4] T. Gruber, “Toward principles for the design of ontologies used for knowledge sharing”, International Journal of Human-Computer Studies, pp. 907-928, 1995.

[5] A. Maedche and S. Staab, “Ontology Learning for the Semantic Web”, Intelligent Systems, IEEE, pp.72-79 , 2002.

[6] L. Drumond, R. Girardi, “A Survey of Ontology Learning Procedures”, Proceedings of the 3rd Workshop on Ontologies and their Applications, vol. 427 of CEUR Workshop Proceedings, Salvador, Bahia, Brazil, 2008.

[7] D. Benz, “Collaborative ontology learning”, Master’s thesis, University of Freiburg, 2007.

[8] V. Kashyap, “Design and creation of ontologies for environmental information retrieval”, Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW), Banff, Alberta, Canada ,1999.

[9] G. Miller, “Wordnet: a lexical database for english”, Commun, ACM 38(11), pp.39-41, 1995.

[10] Wiktionary, accessible from en.wiktionary.org.

[11] M. Rogger and S .Thaler, “Ontology Learning”, Seminar paper, Applied Ontology Engineering, Leopold–Franzens–University Innsbruck, 2010.

[12] A. B. Bouiadjra and S. M. Benslimane, “FOEval: Full Ontology Evaluation, Model and Perspectives”, In Proceedings of the 7th International Conference on Natural Language Processing and Knowledge Engineering, pp. 464-468, IEEE, Nov. 2011.

[13] P. Cimiano, “Ontology Learning and Population from Text”, ISBN: 978-0- 387-30632-2, Springer, 2006.

[14] D. Borigault, C. Jacquemin, and M.-C. L’Homme, editors, “Recent Advances in Computational Terminology”, Natural language processing series, vol. 2, pp. 328-332 John Benjamins Publishing Company, 2001.

[15] K. Frantzi and S .Ananiadou, “The c-value / nc-value domain independent method for multiword term extraction”, Journal of Natural Language Processing, pp. 145-179, 1999.

[16] P. Pantel and D. Lin, “A statistical corpus-based term extractor”, In E. Stroulia and S. Matwin, editors, AI 2001, Lecture Notes in Artificial Intellingence, pp. 36-46, Springer Verlag, 2001.

[17] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval Information Processing & Management 24(5), pp. 515–523, 1988.

[18] G. Miller, R. Beckwith, C. Fellbaum , D. Gross , and K. J. Miller, “Introduction to WordNet: An On-line Lexical Database”, International Journal of Lexicography 3, pp. 235-244, Dec. 1990.

[19] D. Bourigault and C. Jacquemin,“Term extraction+ term clustering: An integrated platform for computer-aided terminology”, In Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics, pp. 15-22, 1999.

[20] D. Faure and C. Nedellec, “Knowledge acquisition of predicate argument structures from technical texts using machine learning: The system ASIUM”. Knowledge Acquisition, Modeling and Management, pp. 329-334, 1999.

[21] A. Maedche and S. Staab, “The text-to-onto ontology learning environment”, In Software Demonstration at ICCS-2000-Eight International Conference on Conceptual Structures, Aug., 2000.

[22] E. Drymonas, K. Zervanou, and E. Petrakis, “Unsupervised ontology acquisition from plain texts: The OntoGain system”, Natural Language Processing and Information Systems, pp. 277-287, 2010.

[23] P. Buitelaar, P. Cimiano, and B. Magnini, “Ontology Learning from Text: An Overview”, Ontology Learning from Text: Methods, Evaluation and Applications, IOS Press, 2005.

[24] D. Hindle, “Noun Classification from Predicate-Argument Structures”, In Proceedings of the Annual Meeting of the Association for Computational Linguistic, pp. 268-275, 1990.

[25] D. Lin and P. Pantel, “Induction of Semantic Classes from Natural Language Text”, In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 317-322, 2001.

[26] D. Lin and P. Pantel, “Concept Discovery from Text”, In Proceedings of the International Conference on Computational Linguistics (COLING), pp. 577-583, 2002.

[27] M. Reinberger and P. Spyns, “Unsupervised Text Mining for the Learning of DOGMA-inspired Ontologies”, Ontology Learning from Text: Methods, Evaluation and Applications, IOS Press, 2005.

[28] R. Evans, ”A Framework for Named Entity Recognition in the Open Domain”, In Proceedings of the Recent Advances in Natural Language Processing (RANLP-2003), pp. 137-144, 2003.

[29] O. Etzioni et al., ”Web-Scale Information Extraction in KnowItAll (Preliminary Results)”, In Proceedings of the 13th World Wide Web Conference (WWW), pp. 100-109, 2004.

[30] M. A. Hearst, “Automatic acquisition of hyponyms from large text corpora”, In Proceedings of the 14th conference on Computational linguistics, vol. 2, pp. 539-545, Association for Computational Linguistic, 1992.

[31] Z. Harris, Mathematical Structures of Language, John Wiley & Sons, 1968.

[32] M. Sanderson and B. Croft, “Deriving concept hierarchies from text”, In Research and Development in Information Retrieval, pp. 206-213. 1999.

[33] T. Rindflesch, L. Tanabe, J. Weinstein, and L. Hunter. Edgar, “Extraction of drugs, genes, and relations from biomedical literature”, In Pacific Symposium on Biocomputing, 2000.

[34] J. Pustejovsky, J. Castano, J. Zhang, B. Cochran, and M. Kotecki, “Robust relational parsing over biomedical literature: Extracting inhibit relations”, In Pacific Symposium on Biocomputing, 2002.

[35] S. Vintar, L. Todorovski, D. Sonntag, and P. Buitelaar, “Evaluating context features for medical relation mining”, In ECML/PKDD Workshop on Data Mining and Text Mining for Bioinformatics, 2003.

[36] A. Maedche and S. Staab, “Discovering conceptual relations from text”, In W. Horn, editor, Proceedings of the 14th European Conference on Artificial Intellignece (ECAI’2000), 2000.

[37] D. Lin and P. Pantel, “DIRT - Discovery of Inference Rules from Text”, In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 323-328, 2001.

[38] D. Lin and P. Pantel, “Induction of Semantic Classes from Natural Language Text”, In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 317-322, 2001.

[39] I. Dagan, O. Glickman, and B. Magnini, “The PASCAL Recognising Textual Entailment Challenge”. Lecture Notes in Computer Science, vol. 3944, pp.177-190, Jan. 2006.

[40] D. Faure and C. Nedellec, “Knowledge acquisition of predicate argument structures from technical texts using machine learning: The system ASIUM”. Knowledge Acquisition, Modeling and Management, pp. 329-334, 1999.

[41] K. Xu, S. Reddy, Y. Feng, S. Huang, and D. Zhao, “Question Answering on Freebase via Relation Extraction and Textual Evidence”, arXiv.org, 2016.

[42] U. Hahn, M. Romacker, and S. Schulz, “MedSynDiKATe--design considerations for an ontologybased medical text understanding system”, In Proceedings of the AMIA Symposium, pp. 330-334, American Medical Informatics Association, 2000.

[43] R. Navigli and P. Velardi, “Learning Domain Ontologies from Document Warehouses and Dedicated Websites”, Computational Linguistics, 30(2), MIT Press, pp. 151-179, 2004.

[44] X. Jiang and A. H. Tan, “CRCTOL: A semantic- based domain ontology learning system”, Journal of the American Society for Information Science and Technology, 61(1), pp. 150-168, 2009.

[45] K. Dellschaft and S. Staab, “Strategies for the evaluation of ontology learning”, In Ontology Learning and Population: Bridging the Gap between Text and Knowledge, P. Buitelaar and P. Cimiano, Eds. IOS Press, pp. 253-273, 2008, Amsterdam.

[46] R. Porzel and R. Malaka, “A task-based approach for ontology evaluation”, In Proceedings of the 16th European Conference on Artificial Intelligence,Valencia, Spain, Citeseer, 2004.

[47] W. Wong, W. Liu, and M. Bennamoun, “Ontology Learning from Text: A Look Back and into the Future”, Article in ACM Computing Surveys, Article 20, Jan. 2011.

[48] Y. Kalfoglou, “Evaluating ontologies during deployment in applications, position statement”, OntoWeb 2 meeting, Amsterdam, Dec. 2002.

[49] A. Maedche and S. Staab, “Measuring similarity between ontologies”. In Proceedings of the European Conference on Knowledge Acquisition and Management (EKAW), pp. 251-263, 2002.

[50] M. Sabou, C. Wroe, C. Goble and G. Mishne, “Learning domain ontologies for Web service descriptions: An experiment in bioinformatics”, In Proceedings of the 14th International Conference on the World Wide Web, pp. 190-198, 2005.

[51] K. Dellschaft and S. Staab, “On how to perform a gold standard based evaluation of ontology learning”, In Proceedings of the 5th International Semantic Web Conference (ISWC), pp. 228-241, 2006.

[52] L. Ding et al., “Swoogle: A Search and Metadata Engine for the Semantic Web”. In Proceedings of the 13th CIKM, pp. 652-659, 2004.

[53] C. Patel, K. Supekar, Y. Lee, and E. K. Park. OntoKhoj, “A Semantic Web Portal for Ontology Searching, Ranking and Classification”, In Proceeding of the Workshop On Web Information And Data Management, pp. 58-61, ACM, 2003.

[54] S. Tartir, I. B. Arpinar, M. Moore, A. P. Sheth, and B. Aleman-Meza, “OntoQA Metric-Based Ontology Quality Analysis”, IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, Houston, TX, USA, 2005.

[55] V. Cross and A. Pal, “OntoCAT: An Ontology Consumer Analysis Tool and Its Use on Product Services Categorization Standards”, In Proceedings of the First International Workshop on Applications and Business Aspects of the Semantic Web, pp. 44-58, 2006.

[56] M. Jones and H. Alani, “Content-based ontology ranking”, In Proceedings of the 9th International Protege Conference, pp. 92-96, 2006.

[57] Y. Wei and J. Chen, “Ranking Ontology based on Structure Analysis”, Second International Symposium on Knowledge Acquisition and Modeling IEEE, pp. 119-122, 2009.

[58] A. Zouaq, D. Gasevic, and M. Hatala, “Towards open ontology learning and filtering”, Information Systems, vol. 36, no.7, pp. 1064–1081, 2011.

[59] K. Arabshian, P. Danielsen and S. Afroz, “LexOnt: A Semi-Automatic Ontology Creation Tool for Programmable Web”, In : AAAI Spring Symposium: Intelligent Web Services Meet Social Computing, pp. 2-8, 2012.

[60] M. Poveda Villalon and MC. Suárez-Figueroa, “OOPS!–OntOlogy Pitfalls Scanner!”, oa.upm.es, 2012.

[61] H. Hlomani and D .Stacey, “Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey”, Semantic Web Journal, pp. 1-11, 2014.

[62] J. Raad and C. Cruz, “A Survey on Ontology Evaluation Methods”, Proceedings of the International Conference on Knowledge Engineering and Ontology Development, pp. 179-186, Nov 2015.

[63] F. Sun, J. Guo, Y. Lan, J. Xu, and X. Cheng, “Learning Word Representations by Jointly Modeling Syntagmatic and Paradigmatic Relations”, Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, pp. 136-145, 2015.

[64] S. Faralli and S. P. Ponzetto, “DWS at the 2016 Open Knowledge Extraction Challenge: A Hearst-Like Pattern-Based Approach to Hypernym Extraction and Class Induction”, Springer: Communications in Computer and Information Science, vol. 641, pp. 48-60, 2016.

[65] J. Gao and S. Mazumda, “Exploiting Linked Open Data to Uncover Entity Types”, Springer: Communications in Computer and Information Science, vol. 548, pp. 51-62, 2016.

[66] S. Consoli, D. Reforgiato Recupero, “Using FRED for Named Entity Resolution, Linking and Typing for Knowledge Base Population”, Springer: Communications in Computer and Information Science, vol. 548, pp. 40-50, 2016.

[67] T. Kliegr and O. Zamazal, “LHD 2.0: A text mining approach to typing entities in knowledge graphs”, Web Semantics: Science, Services and Agents on the World Wide Web, 2016.

[68] P. Arnold and E .Rahm, “Automatic extraction of semantic relations from wikipedia”, International Journal on Artificial Intelligence Tools vol. 24, No. 2 , 2015.

[69] S. Landes, C. Leacock and R.I. Tengi, “Building Semantic Concordances”, In Christiane Fellbaum (editor), WordNet: an Electronic Lexical Database, pp. 199– 216, MIT Press, Cambridge,1999.

[70] PubMed, accessible from pubmedcentral.nih.gov.

