ISSN 1870-4069

A Proposal for Domain Ontological Learning

Yuridiana Alemán, María J. Somodevilla, and Darnes Vilariño

Benemérita Universidad Autónoma de Puebla, Faculty of Computer Science, Puebla, Mexico
{yuridiana.aleman, mariajsomodevilla, dvilarinoayala}@gmail.com

Abstract. In this paper, the population ontology problem is addressed and a semiautomatic methodology for ontology learning is proposed. In this work a study of advances in this research area is presented, which is divided according with the aim of the ontology and the class extraction, creation, or population. Based on this study, an initial proposal for ontology semiautomatic population is introduced. This initial proposal consists of four general steps: class extraction, creation, population and evaluation, where the population step is the main objective of this research. As future work, this methodology is intended to be applied to pedagogic domain, specially to classroom learning tools.

Keywords: Ontology learning, Semiautomatic population, Pedagogy domain, Ontology creation, NLP.

1 Introduction

In recent years, the available information has increased exponentially; thus, information science researchers propose strategies to develop processes and generate answers according to the user requirements in Information Processing Systems (IPS) [1]. The classic techniques of information retrieval cannot resolve problems like heterogeneity and ambiguity of web information. It is necessary to develop new semantics approaches to improve actual research, for example ontologies.

Ontologies can be used for purposes such as structure knowledge in taxonomies, vocabulary manage, natural language processing applications, searches, recommendation systems, and e-learning among others [2, 3, 4]. Ontologies can model interaction systems between users and their environment, since to its property to manage complex knowledge in reusable formal representations.

Ontology is a formal, explicit specification of a shared conceptualization. Their classes, relationships, constraints and axioms define a common vocabulary to share knowledge [5]. The ontology construction process can be manual or automatic, if it is automatic, the process is called "Ontology learning" and include: relevant terminology acquisition, synonyms extraction, concepts formation, hierarchical organization of elements, relationships learning, properties, attributes, with their rank and domain, hierarchical organization of relationships, instantiation of schema axioms, and arbitrary axioms definition [6]. In the manual construction process, a domain expert is necessary for model formalization; but generally, it is difficult to transmit their knowledge and the proper way to formalize it [7]. Ontology population

is carried out in the previously defined last steps; these steps look for identifying instances of non-taxonomic relationships and properties of ontology classes. A generic methodology for automatic population include candidate instances identification using natural language processing techniques, classifier construction with machine learning and information extraction and new instances classification [8].

In this paper the population ontology problem is addressed and a semiautomatic methodology for ontology learning is outlined; for experiments, the pedagogic domain will be used. The article is organized in 4 sections following described. Section 2 introduces the population problem as well as the relevance to the selected domain. Section 3 presents an analysis about the related works, this section is divided in extraction features, creation and population subsections. Finally, section 4 presents conclusions and a general description about the proposed methodology.

2 Research Problem

The ontology construction process can be performed in a semiautomatic or automatic way from a set of initial data, which can be a text corpus in natural language. An automatic process involves artificial intelligence techniques and not human intervention, the semi-automatic process involves the human intervention in any process step. The idea is to start from a set of texts and relate their grammatical features with ontological entities. This relationship can be not flexible; however, there are exceptions where Natural Language Processing (NLP) techniques allow achieving a good texts representation through ontologies.

Currently, the ontology creation process is a critical process, which can be done manually but involves much time and resources, on the other hand, works about population process use specific domains and manual evaluation. Thus, in the present research a methodology for semi-automatic population of ontologies is outlined.

The main objective is to design an ontology in the pedagogical domain and then, to propose a methodology for its semi-automatic population using NLP techniques in unstructured texts. In the creation and population stages, it is necessary to use a specific domain for implementing the methodology and carrying out evaluation metrics; therefore, the pedagogical domain is used for initial experiments.

In researchers about this domain, the authors present manual ontologies, and the focus is in creation, not in population process. Figure 1 shows a Venn diagram including main works revised being divided according to the ontology focus: applied to e-learning or classroom classes. Also the figure presents concepts in a specific topic or as a tool to facilitate the learning process among lecturer and students.

The proposed work is in the dark area (intersection between face to face domain and ontology building as a tool). This last area is addressed in most of the works, but always using a manual construction, thus, the research takes relevance in the proposal of a semiautomatic approach. In section 3, scientific literature about semiautomatic population is discussed.

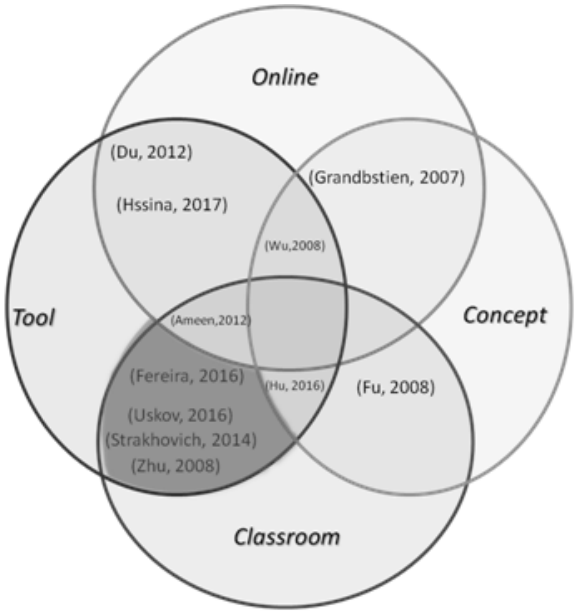

Fig. 1. Researches in pedagogy domain

3 Related Work

Figure 2 relates research topics, besides the techniques and methods used in each stage. The works are presented as follows, and the evaluation analysis is done in each paper of the others subsections.

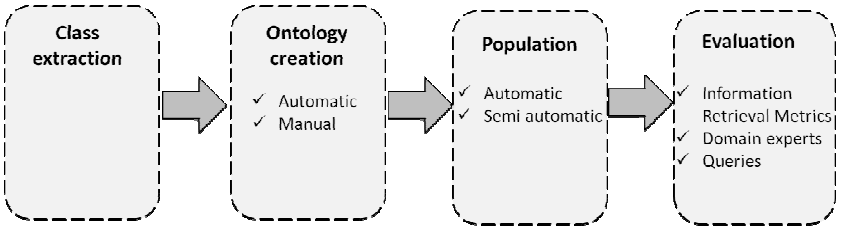

Fig. 2. Art state workflow instead

3.1 Class Extraction

Ontologies for the Use of digital learning Resources and semantic Annotations on Line (OURAL) project is present in [9], the project includes people from several disciplines (educational science, computer science, and cognitive psychology) building e-learning services. The authors present the extracted class using NLP tech-

niques in unstructured texts about learning situation. Educational domain was also analyzed in [10], but its application was into Chinese language.

Others works like [11] present methods for semi-automatic class extraction using a database of Spanish verbs, diathesis alternations and syntactic-semantic schemes (ADESSE tool) [12], where the semantic extracted patterns are the classes. This methodology was applied in educational domain and replicated in financial domain in [13]; in both works, the class extraction was completed with the domain expert opinion. A method for class extraction using linguistic patterns and NLP metrics such as morphological labeling is presented in a recent research [14].

3.2 Ontology Creation

Table 1 summarizes the work done, in conjunction with the domain used in addition to the construction method.

Table 1. Researches on ontologies creation

Method	Domain
Automatic	[15] Painters biographies [16] Technical and medical texts [17] FIFA [7] EOLSS Collection
Manual	[18] News about the basic level education [19] Material for teaching English [20] Online Course Manuals [2] Computing [21] Online education [22] Software engineering courses [23] Courses offered by a university [24] Intelligence Levels [25] Books for level K12 [26] E-learning

Artequackt project, a system of biographies according to the user parameters, is presented in [15]; it is a pilot test where the ontology was created automatically using WordNet tool (https://wordnet.princeton.edu/). Others works such as [16] and [17] show building process using episode extraction in an unstructured texts domain, the authors use International Federation of Association Football (FIFA) news and also use information retrieval techniques (precision and recall) for evaluation. A new methodology to automatically build ontologies from Spanish unstructured texts is proposed in [13], the authors describe three stages: concepts search, relations extractions and ontology construction process. Finally, [7] presents a similar approach than [13] but with a deep semantic analysis (multilingual scenarios); the authors use anaphora resolution techniques, clustering and extraction of lexical-syntactic patterns.

The research which report ontologies manually constructed is based in the domain and evaluation, since the domain used for this proposal, mainly analyzes the works in pedagogical domain. [18] and [4] present an ontology for event recognition using 2000 papers about news in the academic life; the authors use pattern extraction and information retrieval metrics.

For improving the classroom learning, authors such as [19] and [25] have proposed techniques based on ontologies. In [19] an ontology for interaction between students and teachers for English language teaching is introduced. On the other hand, [25] proposes an ontology for the internet learning process. In both works is defined an ontology for each entity in the learning process, and the evaluation is conducted with a manual process supervised for domain experts. Other researchers are focused on online education ([20], [2], [21] and recently [26]) where ontologies are manually defined from XML resources available in the Internet, and the evaluation is a manual process too. An ontology created from CASE diagrams for on-line education is presented in [22]; its evaluation is addressed by experts in a manual process.

There are works such as [24] focused on automatic learning; in this paper, an ontology based on the Internet of Things used in a classroom is created, considering the student intelligences. The ontology creation process from the courses information offered in advanced levels is explained in [23], where students can choose courses according with their academic background. Both works present the structure, information, and hierarchy of the classes in a manual way.

3.3 Population

Works reported in literature used an automatic or semi-automatic approach. Table 2 shows them and domains where they were applied.

Table 2. Works on ontologies population

Method	Domain
Automatic	[27] Google text snippets [28] Biomedical [29] Independent [8] Legal and Tourism [6] Academic Profiles
Semi automatic	[4] Scientific papers [18] Scientific papers [30] Geographical locations [31] E-learning SCORM [3] Independent

A method for ontologies population using Google snippets is proposed in [27], the research is based on manual patterns for class assignment and relations, these patterns are consulted at Google where they are analyzed for obtaining new instances. Another method for automated population is proposed in [28], the methodology combines

traditional linguistic analysis and technologies for the extraction of textual knowledge; it is based on contextual distance and knowledge gained with semantic roles.

A different work in automatic population is present in [29], where an automatic unsupervised model is created, and for implementation the authors use texts from the Wikipedia tourism domain. The tourism domain and legal corpus were used in [8] for introducing a generic process as automatic population. In these works are used techniques such as grammatical features extraction, named entities, morphological tagger and queries evaluation using WordNet with precision and recall. The academic profiles are explored in [6], where the authors carry out automatic population using curriculums and Spanish papers abstracts; they use a gold built for domain experts in evaluation step.

For semiautomatic approaches, some works uses domain experts too and NLP techniques. A system for semiautomatic population of unstructured text is presented in [4] and [18], in this paper supervised learning with Marmot, Crystal and Badger tools is applied. A weakly supervised approach is described in [30], where a manual analysis is used to generate a syntactic model, and information retrieval metrics for evaluation step. A related research is worked in [31], where hyponymy and hypernym are evaluated by validating and verifying ontological relationships in a restricted domain. A question answering Web system is described in [3], this system combines multiple knowledge bases and a NLP parser for transforming queries in Protocol and RDF Query Language (SPARQL).

The most of research carried out in 3.1, 3.2 and 3.3 is based on domain experts in creation or evaluation steps, the works related to automatic approach focus on creation but not population. This analysis is relevant for the proposed research, since any method for population task is feasible in pedagogical domain and the creation step is manual. For evaluation step, it is necessary to analyze different metrics, such as, accuracy, precision and recall, which have been used by several researchers.

4 Conclusions

This paper addressed the ontology population problem and recent topics related to. The population is carried out in manual or automatic process, but only a few papers deal with automatic or semiautomatic creation and population issue. A four steps proposal in semi-automatic population is ongoing:

-Preprocessing: Considers the corpus creation from pedagogical documents follows by the classes extraction using NLP and information retrieval techniques.

-Creation: The ontology system creation by a semiautomatic approach comprises three single ontologies: learning styles, multiple intelligences and teaching learning strategies.

-Population: Use of NLP techniques and semantic analysis.

-Evaluation: Time, consistency and domain will be evaluated.

As a future work, the proposed methodology will be implemented in the initial domain, but the idea is to extend it to obtain a general methodology applicable to other pedagogy domains.

References

1. López Bolaño, F., Castillo Pérez, J.: Esquema metodológico para la construcción automática de ontologías. Revista Vínculos 10, 20–30 (2013)
2. Dai, X., Li, X.: Study of Learning Source Ontology Modeling in Remote Education. In : 2010 International Conference on Multimedia Technology, pp. 1–4 (2010)
3. El-Ansari, A., Beni-Hssane, A., Saadi, M.: A Multiple Ontologies Based System for Answering Natural Language Questions. In Rocha, Á., Serrhini, M., Felgueiras, C., eds. : Europe and MENA Cooperation Advances in Information and Communication Technologies. Springer International Publishing, pp. 177–186 (2016)
4. Celjuska, D., Vargas-vera, D.: Semi-Automatic Population of Ontologies from Text. In : In: Workshop on Data Analysis WDA-2004, pp. 33–49 (2004)
5. Guarino, N., Masolo, C., Vetere, G.: OntoSeek: content-based access to the Web. IEEE Intelligent Systems and their Applications 14(3), 70–80 (1999)
6. Reyes-Ortiz, J., Bravo, M., Herrera-Alcántara, O., Gudiño, A.: Poblado automático de ontologías de perfiles académicos a partir de textos en español. Research in Computing Science 95, 159–170 (2015)
7. De la Villa Moreno, M.: Método para la Construcción Automática de Ontologías Basadas en Patrones Lingüisticos. Ph.D. dissertation (2016)
8. Faria, C., Girardi, R.: A domain-independent process for automatic ontology population from text. Science of Computer Programming 95, Part 1, 26–43 (2014).
9. Grandbastien, M., Azouaou, F., Desmoulins, C., Faerber, R., Leclet, D., Quenu-Joiron, C.: Sharing an ontology in Education: Lessons learnt from the OURAL project. In : Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), pp.694–698 (2007)
10. Fu, J., Jia, K., Xu, J.: Domain Ontology Learning for Question Answering System in Network Education. In : 2008 The 9th International Conference for Young Computer Scientists, pp.2647–2652 (2008)
11. Ochoa Hernández, J.: Desarrollo de una metodología para la construcción automática de ontologías en español a partir de texto libre. Ph.D. dissertation (2011)
12. Garcíal, J., Vaamonde, G., González Dominguez, F.: ADESSE, a Database with Syntactic and Semantic Annotation of a Corpus of Spanish. Proceedings of the International Conference on Language Resources and Evaluation LREC 2010, pp. 1903–1910 (2010)
13. Ochoa, J., Hernández-Alcaraz, M., Valencia-García, R., Almela, A.: Learning semantic relations from Spanish natural language documents in the financial domain. In : Proceedings of the 3rd International Conference on Computer Modeling and Simulation, held at Mumbai, India. Chengdu: Institute of Electrical and Electronics Engineers, Inc, pp.104–108 (2011)
14. Kang, Y.-B., Haghighi, P., Burstein, F.: CFinder: An intelligent key concept finder from text for ontology development. Expert Systems with Applications 41, 4494–4504 (2014)
15. Alani, H., Kim, S., Millard, D., Weal, M., Hall, W., Lewis, P., Shadbolt, N.: Automatic Ontology-Based Knowledge Extraction from Web Documents. IEEE Intelligent Systems 18, 14–21 (2003)

16. Valencia, R.: Un entorno para la extracción incremental de conocimiento desde texto en lenguaje natural. Universidad de Murcia, Murcia (2005)
17. Lee, C.S., Kao, Y.F., Kuo, Y.H., Wang, M.H.: Automated ontology construction for unstructured text documents. Data & Knowledge Engineering 60, 547–566 (2007)
18. Vargas-Vera, M., Celjuska, D.: Event Recognition on News Stories and Semi-Automatic Population of an Ontology. In : Web Intelligence, 2004. WI 2004. Proceedings. IEEE/WIC/ACM International Conference on, pp.615–618 (Sept 2004)
19. Zhu, F., Fok, A., Ip, H., Cao, J.: ENGOnto: Integrated Multiple English Learning Ontology for Personalized Education. In : 2008 International Conference on Computer Science and Software Engineering, vol. 5, pp.210–213 (Dec 2008)
20. Wu, H.: Research of Internet Education System Based on Ontology. In : 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 4, pp.602–605 (2008)
21. Du, L., Zheng, G., You, B., Bai, L., Zhang, X.: Research of Online Education Ontology Model. In : 2012 Fourth International Conference on Computational and Information Sciences, pp.780–783 (2012)
22. Bagiampou, M., Kameas, A.: A Use Case Diagrams ontology that can be used as common reference for Software Engineering education. In : 2012 6th IEEE International Conference Intelligent Systems, pp.35–40 (2012)
23. Ameen, A., Khan, K., Rani, B.: Creation of Ontology in Education Domain. In : 2012 IEEE Fourth International Conference on Technology for Education, pp.237–238 (2012)
24. Uskov, V., Pandey, A., Bakken, J., Margapuri, V.: Smart engineering education: The ontology of Internet-of-Things applications. In : 2016 IEEE Global Engineering Education Conference (EDUCON), pp.476–481 (2016)
25. Hu, J., Li, Z., Xu, B.: An Approach of Ontology Based Knowledge Base Construction for Chinese K12 Education. In : 2016 First International Conference on Multimedia and Image Processing (ICMIP), pp.83–88 (2016)
26. Hssina, B., Bouikhalene, B., Merbouha, A.: An Ontology to Assess the Performances of Learners in an e-Learning Platform Based on Semantic Web Technology: Moodle Case Study. In : Europe and MENA Cooperation Advances in Information and Communication Technologies. Springer, pp. 103–112 (2017)
27. Geleijnse, G., Korst, J.: Automatic Ontology Population by Googling. In : In: Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence BNAIC, pp.120 (2005)
28. Ruiz-Martínez, J., Valencia-García, R., Martínez-Bájar, R., Hoffmann, A.: BioOntoVerb: A top level ontology based framework to populate biomedical ontologies from texts. Knowledge-Based Systems 36, 68–80 (2012)
29. Vicient, C., Sánchez, D., Moreno, A.: An automatic approach for ontology-based feature extraction from heterogeneous textualresources. Engineering Applications of Artificial Intelligence 26, 1092-1106 (2013)
30. Tanev, H., Magnini, B.: Weakly Supervised Approaches for Ontology Population. In : Proceedings of the 2008 Conference on Ontology Learning and Population: Bridging the Gap Between Text and Knowledge, Amsterdam, The Netherlands, The Netherlands, pp.129–143 (2008)
31. Vidal, M., Pinto, D., Montes, A., Serna, G., Ayala, D.: Identification of Ontological Relations Using Formal Concept Analysis. In : Proceedings of the Ninth Latin American Workshop on Logic/Languages, Algorithms and New Methods of Reasoning, pp.1–9 (2014)