

Information Extraction for Learning Expressive Ontologies

Giulio Petrucci^{1,2(✉)}

¹ Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Trento, Italy

² University of Trento, Via Sommarive, 14, 38123 Trento, Italy

petrucci@fbk.eu

Abstract. Ontologies are used to represent knowledge in a formal and unambiguous way, facilitating its reuse and sharing among people and computer systems. A large amount of knowledge is traditionally available in unstructured text sources and manually encoding their content into a formal representation is costly and time-consuming. Several methods have been proposed to support ontology engineers in the ontology building process, but they mostly turned out to be inadequate for building rich and expressive ontologies. We propose some concrete research directions for designing an effective methodology for semi-supervised ontology learning. This methodology will integrate a new axiom extraction technique which exploits several features of the text corpus.

1 Introduction

According to a widely accepted definition, an Ontology is *a formal representation of a shared conceptualization* (see [15]): dealing with a formal and explicit representation of a commonly agreed understanding of a domain, can help to overcome the problem of ambiguity in knowledge representation and sharing. In the Semantic Web scenario, ontologies provide the conceptual scheme for metadata carrying the explicit data semantics, endowing machines with the capability to interpret data unambiguously and perform reasoning over them.

Several approaches, known as *Ontology Learning*, have been proposed along the years to facilitate the encoding of knowledge from large textual sources, which are massively available, into ontologies. Such approaches still experience some severe limitations, specially trying to extract complex and expressive formalizations like axioms.

The intended contribution of our work is twofold. First, we aim to provide a novel automatic technique, capable to abstract the formulation of OWL DL axioms from large text corpora. Second, we intend to combine such technique together with other state-of-the-art methods in a full methodology that will support the ontology engineer in the axiom extraction process, reducing its cost.

The paper is structured as follows. In Sect. 2 we will depict the State of the Art in Ontology Learning. In Sect. 3 we state our research problem and the contribution we want to give. In Sect. 4 we will describe our research methodology. In Sect. 5 we present our preliminary results. In Sect. 6 we present our evaluation plan. Finally, Sect. 7 concludes the paper.

2 State of the Art

Human knowledge is often carried by large unstructured textual sources. Building an ontology from such sources can be resource-intensive and time consuming. To support ontology engineers in such process, several approaches have been proposed, known as *Ontology Learning* (see [7]), which build upon well-established techniques from Natural Language Processing, Machine Learning, Information Retrieval, Knowledge Acquisition and Ontology Engineering. The various tasks relevant in the Ontology Learning process and their mutual dependencies have been organized in the *Ontology Learning Layer Cake* (see Fig. 1), a conceptual sketch of a generic ontology learning layered architecture in which each layer is associated with a task and its output is the input for the one on the top of it.

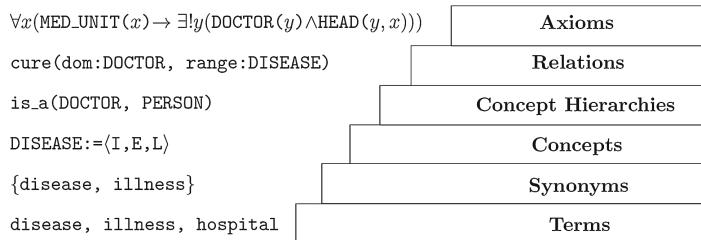


Fig. 1. The Ontology Learning Layer Cake

Recently, the ontology engineering community has been pursuing the ambitious goal to extract increasingly complex information ranging from terms, to relations, to hierarchies, and finally to axioms. Even if the fully automatic acquisition of knowledge by machines is still a long term goal, several automatic and semi-automatic techniques, along with several ready-to-use tools, have been proposed for each layer of the cake.

Term Extraction. The goal of this task is to identify all the relevant terms in the text corpus. Frequency based criteria or metrics from Information Retrieval (TF, TF-IDF, TIF-DRM) are used in [8, 13, 18, 20, 30, 37], together with the C-value/NC-value metric from Computational Linguistics. Linguistic features are exploited in [18, 30, 33] while a combined approach is presented in [44]: extracted terms are connected as nodes in a graph structure, ranked and filtered on the basis of metrics from Graph Theory (Betweenness, Centrality, ecc) or Information Retrieval and finally selected according to some voting schemes (Majority, Intersection, ecc). In [40] a machine learning based approach has been shown capable to adapt to different domains with a reduced training effort.

Synonym Extraction. In order to group together terms with similar meaning, the well known *distributional hypothesis* is largely exploited as in [8, 29, 33]. Other

Text Mining clustering techniques are used in [13] while external resources as WordNet¹ are used in [30].

Concept Learning. At this layer, concepts must be induced in an intensional way. Clustering algorithms and Latent Semantic Indexing (LSI) are used in [13] while in [33] sets of terms considered synonyms are casted into new concepts named after the more occurring term. WordNet Domains and other resources are used in [36] to associate the newly extracted keyphrases with existing concepts. Context similarity in [8, 29] induces concepts definition over sets of terms.

Concept Hierarchy. The goal of this task is to learn taxonomic relations over extracted concepts. Lexico-syntactic patterns from [17] are used in [8, 44]. The approach presented in [35] uses syntactic features to train a binary classifier to predict if two nouns are in a taxonomic relation. In [33] taxonomy is induced with a hierarchical clustering process while in [29] a complex context similarity measure is used to add new concepts in an existing ontology. In [21, 37] a graph with concepts as vertices is built via pattern-driven web search operations, while in [14] a similarity metric over a vector space model is used to evaluate term relatedness. Task-specific algorithms are then used in order to turn these graphs into taxonomic trees.

Relation Learning. The output of this layer is the set of all relations among concepts and individuals, eventually organized in a hierarchical order. Linguistic patterns and metrics from Graph Theory are used in [44]. Linguistic patterns are used also in [18] in order to detect verb-based relations. In [29], context similarity is taken as an evidence of a generic conceptual relation. Statistical significance of co-occurrence is used in [32] in order to predict a relation between two terms. In [9], patterns expressed in an *ad hoc* formalism are used to detect instances of known relations. Syntactic features are used in [8] in order to detect general relations, while a combination of a set of patterns together with WordNet is used for mereological relations. The assumption of an intrinsic redundancy in large corpora is exploited in [16] in order to apply graph mutual reinforcement between a set of relation lexico-syntactic patterns and corresponding matching instances. Iterative approaches are followed in [1, 5, 42] starting from a small hand-crafted set of patterns, matching textual relation instances are used to extend the pattern set. Parse tree feature spaces are used in [6] in order to train tree kernel based classifiers which can predict occurrences of predefined relations. Hand-crafted rules are used in [2] in order to label a set of examples which will be used to train a classifier. This semi-supervised approach has been extended in [11] with POS-tags patterns acting as syntactic constrains and a large dictionary of relation acting as a lexical constraint, in order to improve the extractions quality. A pairwise vector space clustering is used in [34] in order to detect recurring patterns and identify relation instances. A clustering technique is used also in [26] in order to detect generic relations for given type signatures. Ontological

¹ <http://wordnet.princeton.edu/>.

resources like Freebase,² YAGO³ and Wikipedia⁴ infoboxes are exploited in [25, 27, 28, 41] as sources of evidence in order to match textual relation instances from which different types of features are extracted to train classifiers or cluster instances.

Axiom Learning. The last layer addresses the problem of axioms learning. The usage of lexico-syntactic patterns has been exploited in [39] for generating formal class description from definitional sentences. This approach has been followed in [38] in order to detect disjointness among classes using external lexical resources, underlying ontologies and Pointwise Mutual Information (PMI) relying on the Web as a source of evidence. A radically different approach is presented in [31] where Discourse Representation Theory (DRT, see [19]) is used in order to represent the content of a single document in formal structures called Discourse Representation Structures (DRSs). Those structures are then mapped to OWL constructs applying a set of translation rules and exploiting several external lexical (FrameNet⁵ and VerbNet⁶) and ontological (FOAF,⁷ DBpedia,⁸ Dolce+DnS Ultralite⁹) resources. The translation from DRSs to OWL constructs relies on a set of *Ontology Design Patterns* (ODPs). An ODP is “*a reusable successful solution to a recurrent modeling problem*¹⁰” and can be seen as a sort of template to be used in some particular and recurring situations. Acting as constraints, they are supposed to ensure quality for the final ontology construction. In [22], a system for the extraction of \mathcal{EL}^{++} concepts definitions from text is presented. Text fragments involving concepts from the SNOMED CT¹¹ ontology are matched and their lexical and ontological features are used to train a maximum entropy classifier in order to predict the axiom describing the involved entities. Users can provide their feedback, helping the system to correct the underlying model.

3 Problem Statement and Contribution

Reasoning-based applications can infer new knowledge beyond what explicitly stated in the domain representation they rely on. Their power of reasoning depends on the expressivity of such representation: an ontology provided with complex TBox axioms can act as a valuable support for the representation and the evaluation of a deep knowledge about the domain it represents.

Automatic learning of expressive TBox axioms is a complex task. From a linguistic point of view, conjunctions, negations and disjunctions are difficult

² <http://www.freebase.com>.

³ www.mpi-inf.mpg.de/yago-naga/yago/.

⁴ <https://www.wikipedia.org/>.

⁵ <https://framenet.icsi.berkeley.edu/fndrupal/>.

⁶ <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>.

⁷ <http://www.foaf-project.org/>.

⁸ <http://dbpedia.org>.

⁹ <http://www.ontologydesignpatterns.org/ont/dul/DUL.owl>.

¹⁰ http://ontologydesignpatterns.org/wiki/Main_Page.

¹¹ <http://www.ihtsdo.org/snomed-ct/>.

to parse and interpret. The same is for detecting local contexts for universal quantifications. Statistical relevance based metrics could be misleading as definitional sentences can appear infrequently in the corpus and the knowledge to be encoded in a single axiom can be spread across several sentences.

State-of-the-art methods and tools capable to handle large text corpora still appear to be more suitable to support the construction of light-weight ontologies. On the other hand, axiom extraction methods are intended to work in a sentence-by-sentence translation modality, with constraints on the structure of the sentence or the particular domain of interest. As a consequence, ontology learning from large text corpora still remains a heavily manual process which can end up having unsustainable costs. Therefore, we state our research problem in the form of the following question:

How can semi-supervised techniques significantly help to extract relevant expressive axioms from a large domain text corpus and therefore minimize human intervention in the process of ontology building?

Answering this question, the first contribution we aim to give is the design of **an automatic technique to extract TBox axioms from large text corpora**. Our target axioms are those that can be expressed through OWL DL constructs. Each text corpus is assumed to have some inner coherence about some domain. As in [43], we focus on corpora from which a significant amount of ontological statements (definitions of concepts, their relations, ecc) rather than factual statements can be extracted – like encyclopedic texts, textbooks or specifications. Setting this research goal, we hypothesize that statistical, linguistic and semantic features of such text corpora can be exploited to train a system capable to detect textual occurrences of axioms.

The second contribution is the organization of our new axiom extraction technique and other state-of-the-art methods together with human activities into a fully-fledged **methodology**, outlined in Fig. 2. The initial input is a text corpus and the final output is a set of axioms which are intended to be expressive conceptualization of the knowledge originally contained in the input corpus. The four phases are:

1. **Light-Weight Extraction (LE)**: state-of-the-art methods are used to extract concepts, relations and concept hierarchies from the corpus.
2. **User Selection (US)**: the ontology engineer can select a subset of the ontology elements produced by the previous phase or add new ones that may have been missed. This interaction validates the knowledge extracted so far and highlights what the engineer considers as more relevant.
3. **Axioms Extraction (AE)**: the user can select a relation, concept or a pair of concepts and the system will provide all the hypotheses of axioms involving them, along with some textual evidence if possible. Hypotheses will be ranked according to how much the system is *confident* that the single axiom actually holds in the corpus.
4. **User Feedback (UF)**: the user can express a judgment about the correctness of the axiom hypotheses suggested by the previous phase, in order to remove uncertainty.

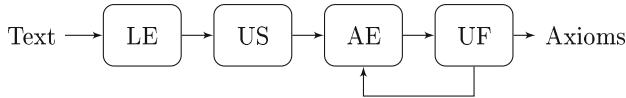


Fig. 2. An outline for the proposed methodology

4 Research Methodology and Approach

In order to train a system capable to learn axioms from text, we first need to identify which features of the text corpus must be exploited. As shown in Sect. 2, several approaches relying on lexico-syntactic features and deep linguistic analysis of sentences have been proposed. With some ongoing experiments we are investigating the possibility to combine them with statistical relevance based ones as in [30, 44]. As definitory sentences could be strong points of evidence for axioms holding in the text, we want to exploit them as well with deeper linguistic analysis techniques, as in [31, 37]. Some preliminary studies have been performed using the BPMN 1.1 specification¹² as the input text corpus and a corresponding ontology, manually developed by an ontology engineer.

Starting from this, we plan to build a set of training examples: the inputs is a text corpus and the expected output is a set of axioms actually holding according to the corpus content. We have started investigating some available ontologies to extract axioms from them and build our expected outputs. Usually, they are not supplied with a corresponding text corpus so that we will have to build it by ourselves harvesting the web or some other large document repositories. Our idea is to train our system with these examples assuming that the system can *scale* to other corpora maintaining an acceptable performance level, in a sort of domain adaptation fashion. The approaches presented in [24] for keyphrases extraction and in [12] for taxonomy induction exploited the same intuition. We want to extend this investigation to the axiom extraction problem.

Our target ontology language is OWL DL. We plan to develop our system in an iterative way, having different type of constructs as target for each step. So far, we have planned to proceed according to the following hypothesis of work:

1. Atomic Negation, Concept Intersection, Inverse Properties, Nominals
2. Complex Concept Negation, Role Hierarchy, Cardinality Restrictions
3. Universal Restrictions, Limited Existential Quantification

5 Preliminary Results

During the first year of our work we focused on reviewing the State of the Art and performing some exploratory activities. We outlined the preliminary version of our methodology on the main idea of alternating manual and automatic tasks. In

¹² <http://www.omg.org/spec/BPMN/1.1/>.

order to decide which techniques to use in the first phase of our methodology, we analyzed several state-of-the-art methods. After some comparative experiments, we opted for KX, a tool implementing the approach presented in [30] for the term extraction task, as it turned out giving the best performance. We still have to investigate further in order to decide which relation extraction and concept hierarchy induction techniques to be adopted.

We tested several NLP tools over the BPMN 1.1 specification in order to get a rough idea of their behavior. We performed deep syntactic analysis with Mate-tools (see [3]) and with the Stanford Core NLP toolkit (see [23]) which has been extended with a plugin (still under development) that annotates the shortest path parse tree fragment (as in [6, 41]) between pairs of relevant concepts spotted by KX: such paths could be easily analyzed in order to check how they can be useful for our work. We used SEMAFOR (see [10]) in order to extract all the occurrences of the semantic frames collected in FrameNet. We also tried to use Boxer (see [4]) to extract a deep semantic representation of the text based on DRSs, which have the interesting property of being equivalent to First-Order Logic formulae.

Recently we started building our training set according to analogous experiences in literature. In [29], the OpenCyc ontology is exploited for the domains of *Fisheries and Aquaculture* and *Economy and Finance*. The Aquatic Sciences and Fisheries Abstract (ASFA) thesaurus¹³ has been used as a corpus for the former, while the Harvey Glossary¹⁴ for the latter. In [37] the Economy and Finance corpus has been extended with scientific journal or conference papers about the topic. The Economist also provide a related glossary.¹⁵ In the same work, authors followed the same procedure to build a corpus for the domain of *Artificial Intelligence*, but no valuable ontology seems to be available. The BioPortal¹⁶ project provides a large repository of ontologies for the biomedical domain. Several corpora for such domain¹⁷ are available as it has been explored in depth by the NLP research community. However, as it comprises a large variety of different topics, finding suitable corpus/ontology pairs may not be straightforward.

During the last weeks, further explorations suggested us the opportunity to deepen some topics from the distributed compositional semantics field in order to deal with linguistic issues in axiom detection.

6 Evaluation Plan

Our axiom extraction technique will be evaluated according to its capacity of correctly identify axioms actually holding in the input text corpus. We plan to continuously evaluate it using the axioms contained in our training set ontologies

¹³ <http://www4.fao.org/asfa/asfa.htm>.

¹⁴ <http://biz.yahoo.com/glossary>.

¹⁵ <http://www.economist.com/economics-a-to-z>.

¹⁶ <http://bioportal.bioontology.org/ontologies>.

¹⁷ e.g. <http://www2.informatik.hu-berlin.de/~hakenber/links/benchmarks.html>.

as the gold standard. Well-known metrics like Precision and Recall could be used in a quite straightforward way in this scenario.

Later, we plan to involve some human assessors in order to judge some axioms, maybe verbalized, as actually holding in the text corpus or not. Our idea is to involve more human assessors and measure the quality of the judgment using some inter-annotator agreement metric.

The final evaluation phase will focus on the effectiveness of our methodology in term of cost reduction. Starting from the same specification we will build again the BPMN ontology using our methodology. This will allow us to compare both the correctness of the extracted axioms and the time spent for this task with the ones of the corresponding hand-made ontology.

7 Conclusions

Several approaches have been proposed to support ontology engineers in the tasks of building ontologies from large textual corpora. Looking at state-of-the-art techniques for semi-automatic ontology engineering, despite the progress made so far, they are not able to give significant support in rich and expressive ontology building so that their cost can become unaffordable. To overcome this limitation, we want to design a new axiom extraction technique to be integrated into a semi-automatic methodology together with other state-of-the-art methods. We hypothesize that available large textual and knowledge sources can be used to train a domain-independent system capable of providing a human actor some hypotheses about axioms holding among concepts in the knowledge contained in a text corpus. We also outline an evaluation process, involving both automatic evaluation against a gold standard and human assessment in order to evaluate relevance and correctness of the extracted axioms together with the effectiveness of our methodology in term of human effort reduction.

References

1. Agichtein, E., Gravano, L.: Snowball: extracting relations from large plain-text collections. In: Proceedings of the Fifth ACM Conference on Digital Libraries (2000)
2. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M., Etzioni, O.: Open information extraction from the web. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence, pp. 2670–2676 (2007)
3. Bohnet, B.: Very high accuracy and fast dependency parsing is not a contradiction. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 89–97 (2010)
4. Bos, J.: Wide-coverage semantic analysis with boxer. In: Proceedings of the 2008 Conference on Semantics in Text Processing, pp. 277–286 (2008)
5. Brin, S.: Extracting patterns and relations from the World Wide Web. In: Atzeni, P., Mendelzon, A.O., Mecca, G. (eds.) WebDB 1998. LNCS, vol. 1590, pp. 172–183. Springer, Heidelberg (1999)

6. Bunescu, R.C., Mooney, R.J.: A shortest path dependency kernel for relation extraction. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 724–731 (2005)
7. Cimiano, P., Mädche, A., Staab, S., Völker, J.: Handbook on ontologies. In: Staab, S., Studer, R. (eds.) *Ontology learning*. International Handbooks on Information Systems, pp. 245–267. Springer, Heidelberg (2009)
8. Cimiano, P., Völker, J.: Text2Onto. In: Montoyo, A., Muñoz, R., Métais, E. (eds.) *NLDB 2005*. LNCS, vol. 3513, pp. 227–238. Springer, Heidelberg (2005)
9. Dahab, M.Y., Hassan, H.A., Rafea, A.: TextOntoEx: automatic ontology construction from natural english text. *Expert Syst. Appl.* **34**(2), 1474–1480 (2008)
10. Das, D., Schneider, N., Chen, D., Smith, N.A.: Probabilistic frame-semantic parsing. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, pp. 948–956 (2010)
11. Etzioni, O., Fader, A., Christensen, J., Soderland, S., Mausam, M.: Open information extraction: the second generation. In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, vol. 11, pp. 3–10 (2011)
12. Fallucchi, F., Pazienza, M.T., Zanzotto, F.M.: Generic ontology learners on application domains. In: *Proceedings of the International Conference on Language Resources and Evaluation* (2010)
13. Fortuna, B., Grobelnik, M., Mladenic, D.: OntoGen: semi-automatic ontology editor. In: Smith, M.J., Salvendy, G. (eds.) *HCII 2007*. LNCS, vol. 4558, pp. 309–318. Springer, Heidelberg (2007)
14. Fountain, T., Lapata, M.: Taxonomy induction using hierarchical random graphs. In: *Proceedings of the 2012 Conference of the North American Chapter of the ACL: Human Language Technologies*, pp. 466–476 (2012)
15. Guarino, N., Oberle, D., Staab, S.: What is an ontology? In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*. International Handbooks on Information Systems, pp. 1–17. Springer, Heidelberg (2009)
16. Hassan, H., Hassan, A., Emam, O.: Unsupervised information extraction approach using graph mutual reinforcement. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 501–508 (2006)
17. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: *Proceedings of the 14th Conference on Computational linguistics*, vol. 2, pp. 539–545 (1992)
18. Jiang, X., Tan, A.H.: Mining ontological knowledge from domain-specific text documents. In: *Proceedings of the Fifth IEEE International Conference on Data Mining*, pp. 665–668 (2005)
19. Kamp, H., Reyle, U.: *From Discourse to Logic. Studies in Linguistics and Philosophy*, vol. 42. Springer, Netherlands (1993)
20. Kang, Y.B., Haghghi, P.D., Burstein, F.: CFinder: an intelligent key concept finder from text for ontology development. *Expert Syst. Appl.* **41**(9), 4494–4504 (2014)
21. Kozareva, Z., Hovy, E.: A semi-supervised method to learn and construct taxonomies using the web. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 1110–1118 (2010)
22. Ma, Y., Syamsiyah, A.: A hybrid approach to learn description logic based biomedical ontology from texts. In: *ISWC 2014 Proceeding* (2014)
23. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The stanford coreNLP natural language processing toolkit. In: *Proceedings of the 52nd Annual Meeting of the ACL*, pp. 55–60 (2014)

24. Medelyan, O., Witten, I.H.: Domain-independent automatic keyphrase indexing with small training sets. *JASIST* **59**(7), 1026–1040 (2008)
25. Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, vol. 2, pp. 1003–1011 (2009)
26. Mohamed, T.P., Hruschka, Jr., E.R., Mitchell, T.M.: Discovering relations between noun categories. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1447–1455 (2011)
27. Moro, A., Navigli, R.: Integrating syntactic and semantic analysis into the open information extraction paradigm. In: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, pp. 2148–2154 (2013)
28. Nakashole, N., Weikum, G., Suchanek, F.: Patty: A taxonomy of relational patterns with semantic types. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1135–1145 (2012)
29. Novalija, I., Mladenic, D., Bradesko, L.: Ontoplus: text-driven ontology extension using ontology content, structure and co-occurrence information. *Knowl.-Based Syst.* **24**(8), 1261–1276 (2011)
30. Pianta, E., Tonelli, S.: KX: a flexible system for keyphrase extraction. In: Proceedings of the 5th International Workshop on Semantic Evaluation. pp. 170–173 (2010)
31. Presutti, V., Draicchio, F., Gangemi, A.: Knowledge extraction based on discourse representation theory and linguistic frames. In: ten Teije, A., Völker, J., Handschuh, S., Stuckenschmidt, H., d’Acquin, M., Nikolov, A., Aussenac-Gilles, N., Hernandez, N. (eds.) EKAW 2012. LNCS, vol. 7603, pp. 114–129. Springer, Heidelberg (2012)
32. Schütz, A., Buitelaar, P.: *RelExt*: A tool for relation extraction from text in ontology extension. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 593–606. Springer, Heidelberg (2005)
33. Shih, C.W., Chen, M.Y., Chu, H.C., Chen, Y.M.: Enhancement of domain ontology construction using a crystallizing approach. *Expert Syst. Appl.* **38**(6), 7544–7557 (2011)
34. Shinyama, Y., Sekine, S.: Preemptive information extraction using unrestricted relation discovery. In: Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the ACL, pp. 304–311 (2006)
35. Snow, R., Jurafsky, D., Ng, A.Y.: Semantic taxonomy induction from heterogeneous evidence. In: Proceedings of 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL (2006)
36. Tonelli, S., Rossocher, M., Pianta, E., Serafini, L.: Boosting collaborative ontology building with key-concept extraction. In: 2011 Fifth IEEE International Conference on Semantic Computing (ICSC), pp. 316–319 (2011)
37. Velardi, P., Faralli, S., Navigli, R.: Ontolearn reloaded: a graph-based algorithm for taxonomy induction. *Comput. Linguist.* **39**(3), 665–707 (2013)
38. Völker, J., Haase, P., Hitzler, P.: Learning expressive ontologies. In: Proceedings of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge, pp. 45–69. IOS Press, Amsterdam (2008)
39. Völker, J., Hitzler, P., Cimiano, P.: Acquisition of OWL DL Axioms from Lexical Resources. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 670–685. Springer, Heidelberg (2007)

40. Witten, I.H., Paynter, G.W., Frank, E., Gutwin, C., Nevill-Manning, C.G.: KEA: Practical automatic keyphrase extraction. In: Proceedings of the Fourth ACM Conference on Digital Libraries, pp. 254–255 (1999)
41. Wu, F., Weld, D.S.: Open information extraction using wikipedia. In: Proceedings of the 48th Annual Meeting of the ACL, pp. 118–127 (2010)
42. Zhu, J., Nie, Z., Liu, X., Zhang, B., Wen, J.R.: Statsnowball: a statistical approach to extracting entity relationships. In: Proceedings of the 18th International Conference on World Wide Web, pp. 101–110 (2009)
43. Zouaq, A., Gasevic, D., Hatala, M.: Towards open ontology learning and filtering. *Inf. Syst.* **36**(7), 1064–1081 (2011)
44. Zouaq, A., Gasevic, D., Hatala, M.: Linguistic patterns for information extraction in ontocmaps. In: Proceedings of the 3rd Workshop on Ontology Patterns (2012)