UNIVERSITI PUTRA MALAYSIA

A METHOD TO ENRICH DOMAIN ONTOLOGY USING SYNONYM AND PROBABILITY THEORY

NUR FATIN NABILA BT MOHD RAFEI HENG

FSKTM 2016 15

A METHOD TO ENRICH DOMAIN ONTOLOGY USING SYNONYM AND PROBABILITY THEORY

By

NUR FATIN NABILA BT MOHD RAFEI HENG

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Doctor of Philosophy

September 2016

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the Doctor of Philosophy

A METHOD TO ENRICH DOMAIN ONTOLOGY USING SYNONYM AND PROBABILITY THEORY

By

NUR FATIN NABILA BT MOHD RAFEI HENG

September 2016

Chairman: Assc.Prof. Ali bin Mamat, PhD
Faculty: Computer Science and Information Technology

Ontology has become a popular topic of research for numerous areas of computer science, such as question answering, information retrieval, and use of the semantic web. Considerable efforts have been made in constructing ontologies due to the complexity and time-consuming nature of the task. Concept, taxonomy, and non-taxonomic relations are three important components in the development of ontology. These three components are used to represent the knowledge of the domain texts. Most of the existing techniques focus on extracting the concept, the taxonomic relations, and non-taxonomic relationships within a single sentence. These techniques neglect a sentence when either the subject or object of a sentence is missing or not clear. Thus, the knowledge of domain texts is not properly represented as some relations cannot be identified. This thesis proposes a solution for the enrichment of the knowledge of domain text by finding possible relations. The proposed method suggests the appropriate or the most likely term for an uncertain subject or object of a sentence using the probability theory. In addition, the method can extract the relations between concepts (i.e. subject and object) that appear not only in a single sentence, but also in different sentences by using a synonym of the predicates. The proposed method has been tested and evaluated with three collections of domain texts that describe computers, tourism, and science. Precision, recall, and f-score metrics have been used to evaluate the results of the experiments. The experiment results were compared with the results that were completed manually by the domain experts. For the computer dataset, an F-score value of 62.33% has been achieved using the proposed solution. Additionally, the science dataset achieved an F-score of 78.98%, whereas the tourism dataset achieved an F-score of 81.58%. The result shows that the proposed method has increased and enriched the relationships of domain texts thus providing better results compared to several existing methods. The method is shown to be useful to assist ontology engineer in conceptualization process of ontology engineering.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

KAEDAH UNTUK MEMPERKAYAKAN ONTOLOGY DOMAIN MENGGUNAKAN PERSAMAAN DAN TEORI KEBARANGKALIAN

Oleh

NUR FATIN NABILA BT MOHD RAFEI HENG

September 2016

Pengerusi: Prof. Madya Ali bin Mamat, PhD

Fakulti: Sains Komputer and Teknologi Maklumat

Ontologi telah menjadi suatu topik kajian yang popular untuk pelbagai bidang dalam sains komputer, seperti penjawaban soalan, capaian maklumat, dan penggunaan web semantik. Banyak usaha telah dibuat dalam pembinaan ontologi berikutan kerja-kerjanya yang rumit dan memakan masa. Konsep, taksonomi, dan hubungan bukan taksonomi adalah tiga komponen penting dalam pembangunan ontologi. Ketiga-tiga komponen ini digunakan untuk mewakili pengetahuan tentang teks-teks domain. Kebanyakan teknik-teknik yang sedia ada bertumpu kepada mengekstrak konsep, hubungan taksonomi, dan hubungan bukan taksonomi dalam satu ayat. Teknik-teknik tersebut mengabaikan sesuatu ayat apabila sama ada subjek atau objek di dalamnya tiada atau tidak jelas. Oleh itu, pengetahuan tentang teks domain tidak dapat dipersembahkan dengan betul lantaran beberapa hubungan tidak dapat dikenalpasti. Tesis ini mencadangkan satu penyelesaian untuk memperkayakan pengetahuan teks domain dengan mencari hubungan yang berkemungkinan. Kaedah yang dikemukakan mencadangkan terma yang sesuai atau yang paling berkemungkinan untuk subjek atau objek yang tidak pasti dalam sesuatu ayat dengan menggunakan teori kebarangkalian. Di samping itu, kaedah ini boleh mengekstrak hubungan antara konsep-konsep (iaitu subjek dan objek) yang bukan sahaja muncul dalam satu ayat, tetapi juga dalam ayat yang berbeza dengan menggunakan persamaan predikat. Metrik ketepatan, kebolehdapatan dan F-score digunakan untuk menilai hasil eksperimen. Hasil eksperimen yang dihasilkan oleh kaedah yang dikemukakan akan dibanding dengan hasil yang diperoleh secara manual oleh pakar domain. Bagi set data komputer, kaedah penyelesaian mendapat nilai F-score sebanyak 62.33%. Manakala nilai F-score bagi set data sains and set data pelancongan masing-masing adalah sebanyak 78.98%, dan 81.58%. Hasil eksperimen menunjukkan bahawa kaedah penyelesaian ini telah meningkatkan dan memperkayakan pengetahuan teks domain dan seterusnya menghasilkan keputusan yang lebih baik berbanding beberapa kaedah yang sedia ada.

Kaedah ini terbukti berguna untuk membantu jurutera ontologi semasa proses pembentukan konsep dalam pembinaan ontologi.

ACKNOWLEDGEMENTS

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

I would like to express sincere gratitude and appreciation to my supervisor Associate Professor Dr. Ali Mamat, for his invaluable help, guidance and sharing his opinions on the thesis. I would also like to thank my supervisor committee Associate Professor Dr. Norwati Mustafa and Associate Professor Dr. Masrah Azrifah Azmi Murad, sharing the thought and invaluable assistance.

Special thanks also to Professor Mustafa Mat Deris, Dr. Nurlida Basir and Dr. Nurzi Juana Zaizi, who were abundantly helpful with their invaluable assistance, support, and guidance. My thanks also go to my dear Dean and ex-Deans (Professor Dr. Norita, Professor Dr. Bachok and Professor Dr. Jalani Sukaimi), my sponsors (KPT and USIM), friends and colleagues (although cannot list all names here, as every single one of them contributed to the success of the work presented) for their help and encouragement.

I would like to thank my family, in particular my mum and dad for their love and support has given me so much strength. Last but not least, to my husband (Mohd Nizam bin Mohd Noh) and kids (Muhammad Naufal Hazeem, Muhammad Aysar Faheem and Nur Naura Kauthar) for their love, support, help, and patience through the duration of the study.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Ali Mamat, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Norwati Mustafa, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Masrah Azrifah Azmi, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

_____
BUJANG KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No.: Nur Fatin Nabila Bt Mohd Rafei Heng (GS17455)

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____
Name of Chairman of Supervisory Committee: _____

Signature: _____
Name of Member of Supervisory Committee: _____

Signature: _____
Name of Member of Supervisory Committee: _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TERMINOLOGY

Terminology	Definition
Concept	A related term in text that appears as subject or object of sentence.
Non-taxonomic relation	A relation that shows non-hierarchical relation between concepts.
Complete sentence	A sentence that have subject, predicate and object in a sentence.
irregular sentence	A sentence that have missing or uncertain subject or object in a sentence.
Uncertain value of concept	Subject or object of a sentence is missing or not clear, denoted as '*'.

CHAPTER 1

INTRODUCTION

This chapter presents the motivation of the research, research problem, objective and scope of the research.

1.1 Motivation

Recently ontology has become a popular topic of research in many areas of computer science. These include artificial intelligence, information retrieval, and Semantic Web. The semantic web (Berners-Lee et al., 2001), inspired by Tim Berners Lee, is an extended form of the current web, which aims to provide the textual content with the semantics or meaning into ontology to enable machines to facilitate text understanding. This allows both humans and machines to find information, while obtaining precise answers to specific user queries.

Considerable efforts have been made in constructing ontologies due to the fact that it is a highly complex and time-consuming task (Shamsfard & Barforoush, 2004). Manual construction of ontology is difficult, expensive, and time-consuming (Shamsfard & Barforoush, 2004). Therefore, several works (Byrd & Ravin, 1999; Faure & Nedellec, 1998; Hahn & Schnattinger, 1998; Morin, 1999; Maedche & Volz, 2001; Maedche & Staab, 2000; Navigli et al. 2003) have introduced an automatic or semi-automatic ontology construction using textual data in order to reduce the time and effort required for manual ontology construction.

Even though the number of approaches for constructing ontology is increasing, most of these approaches have only focused on extracting concept (Pantel & Lin, 2002; Tomokiyo & Hurst, 2003; Punuru & Chen, 2006) and taxonomic (is-a) relationship related components (Hearst, 1992; Caraballo, 1999; Cederberg & Widdows, 2003; Cimiano & Staab, 2004). Often neglecting the importance of relationships other than is-a relations, also known as non-taxonomic relationships (Liu et al., 2005).

Many existing techniques (Maedche & Staab, 2000; Kavalec et al., 2004; Akbik & Brob, 2009; Imsombut, 2009; Villaverde et al., 2009; Punuru & Chen, 2007, 2012; Serra et al., 2013) on extracting relationships between two concepts focus on terms that appear as subject and object in a single sentence. Although these techniques are able to extract the relations between concepts, i.e. subject and object that appear in the same sentence from domain texts, the relation cannot be extracted if the two related concepts appear in different sentences or the sentence does not fulfill subject-predicate-object pattern. Hence, the potential relations might be missing. As a consequence, the domain

texts can be considered as not properly presented, as some relations cannot be identified.

This thesis presents a technique to enrich the knowledge of domain text and overcome the limitations of existing techniques by finding all potential relations, which might be overlooked.

1.2 Problem Statement

Various works have focused on ontologies as they have potential in many application areas, such as text mining, information retrieval, knowledge management and the Semantic Web. Ontology provides a description of a certain domain of concern that consists of several components, such as axioms, instance, concept and relation. However, most of the ontologies are constructed manually, which is a difficult task, costly and time-consuming (Shamsfard & Barforoush, 2004).

Considerable works have been completed to construct ontology from domain texts. Nevertheless, most of these works have only focused on extracting concepts and taxonomic relationships such as is-a relation only, with very little work on extracting non-taxonomic relationship. The extraction of non-taxonomic relationships is considered as one of the most challenging and important tasks (Maedche & Staab, 2001; Kavalec et al., 2004; Sanchez & Moreno, 2008). Some works have been proposed for identifying non-taxonomic relations focusing on identifying a given specific relationship such as part-whole (Girju et al., 2003; Berland & Charniak, 1996) and cause-effect (Girju & Moldovan, 2002). However, those works are not able to identify other relationships that are crucial for the domain.

Existing research (Maedche & Staab, 2000; Kavalec et al., 2004; Villaverde et al., 2009; Imsombut, 2009; Akbik & Brob, 2009; Punuru & Chen, 2007, 2012) on extracting non-taxonomic relations between two concepts (terms) focus on terms that appear as subject and object in a single sentence. For example, in the sentence, "*The company produces 50 billion paper every year*". The terms *company* and *paper* are identified as subject and object of a sentence and *produce* is a relationship that relates subject company and object paper. A problem arises if the two concepts (*company* and *paper*) do not exist in the same sentence, then these two concepts will not be considered in the construction of ontology. For instance, if the sentence is irregular: either an object or a subject is missing or not clear, then the relationship between concepts is not extracted. For example, the sentence, "*The company produces it every year*" is considered as an irregular sentence as it does not have a clear object. It is not clear what object "*it*" is referring to in the sentence. In this case, it is assumed that the missing value is "uncertain" or as an "unknown" condition. The object "*it*" may have been described in the previous sentence, but in the existing techniques, relations from sentences that have an uncertain value are not extracted. As a result, the domain texts can be considered as not properly represented, as some relations cannot be identified.

Therefore, the research described in this thesis is to investigate the feasibility of developing an alternative technique to overcome the issue of missing potential relations, which are not handled by the previous methods.

1.3 Objectives

The objectives of this thesis are as follows:

1. To design a technique to determine and predict the missing subject or object in irregular sentences;

2. To develop a method for extracting possible non-taxonomic relations from domain texts; and

3. To measure the completeness and relevancy of the proposed method.

1.4 Research Contribution

This thesis provides three main contributions to knowledge:

1. The technique to find a suitable subject or object for filling up the missing subject or object in irregular sentence.

2. The method improved knowledge extraction from domain texts by extracting the relations between concepts that appear not only in a single sentence, but also in different sentences by using synonym relations.

3. The method improved nearly 80%- 96% of completeness and 40%-94% of relevancy compared to relation extraction from experts and better than existing method (i.e. Punuru and Chen (2007, 2012) and Serra et al. (2013)).

1.5 Scope of the thesis

This thesis focuses on the extraction of ontology components from a collection of domain texts. Current methods mostly concentrate on obtaining possible relationships among concepts (i.e. subject and object) which appear in a sentence. Nevertheless, the potential relationships between concepts may not be extracted whenever there is a sentence without a subject or an object, called irregular subject-predicate-object (S-P-O) pattern. Therefore, the domain texts may be presented improperly. This thesis aims to retrieve as much knowledge as possible, especially that related to non-taxonomic relationships from domain sets. These domain texts, such as articles, news, and texts in natural language are collected from the Internet. For evaluation purposes, the

Computer, *Science*, and *Tourism* domain texts collected from websites were used.

1.6 Organization of the thesis

The remainder of the thesis is organized as follows:

Chapter 2, *Literature Review*, provides the literature review of the research areas such as ontology and ontology engineering. This chapter also provides critical reviews of related approaches in line with the research interests.

Chapter 3, *Methodology*, describes the research design and the steps involved in conducting this research. The research was conducted in three main steps. The first step involved identifying problems and the advantages of existing relation extraction techniques. The second step involves designing and implementing of the proposed approach. The third step involves the experiments and evaluation process, in which the proposed approach was compared against the existing approaches.

Chapter 4, *The Proposed Method of Extracting Non-taxonomic Relations*, presents a detailed description of the proposed non-taxonomic relation extraction solution. This solution is to improve the knowledge of domain text using the probability theory.

Chapter 5, *Results and Discussion*, describes the experiment, which was carried out to evaluate the proposed model. Two experiments were conducted to evaluate the proposed approach against the existing methods and domain expert. The evaluation process for each experiment was based on the standard information extraction measurements, namely precision and recall.

Finally, Chapter 6, *Conclusions and Future Work*, summarizes the thesis and provides the conclusion. It also presents some suggestions for future work, which would be interesting to solve, however, is outside of the current research scope.

REFERENCES

Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *ACM SIGMOD Record* (Vol. 22, No. 2, pp. 207-216). ACM.

Akbik, A., & Brob, J. (2009, April). Wanderlust: Extracting semantic relations from natural language text using dependency grammar patterns. In *WWW Workshop*.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature genetics, 25(1), 25-29.

Brank, J., Grobelnik, M. & Mladenić, D (2005). A survey of ontology evaluation techniques. In In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005). Citeseer.

Berland, M., & Charniak, E. (1999, June). Finding parts in very large corpora. In*Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 57-64). Association for Computational Linguistics.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, *284*(5), 28-37.

Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004). Data driven ontology evaluation.

Byrd, R., & Ravin, Y. (1999, June). Identifying and extracting relations in text. In *Proceedings of NLDB* (Vol. 99).

Caraballo, S. A. (1999, June). Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 120-126). Association for Computational Linguistics.

Cederberg, S., & Widdows, D. (2003, May). Using LSA and noun coordination information to improve the precision and recall of automatic hyponymy extraction. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4* (pp. 111-118). Association for Computational Linguistics.

Ciaramita, M., Gangemi, A., Ratsch, E., Saric, J., & Rojas, I. (2005, July). Unsupervised Learning of Semantic Relations between Concepts of a Molecular Biology Ontology. In *IJCAI* (pp. 659-664).

Cimiano, P., & Staab, S. (2004). Learning by googling. *ACM SIGKDD explorations newsletter*, *6*(2), 24-33.

Cimiano, P., Völker, J., & Studer, R. (2006). Ontologies on demand?-a description of the state-of-the-art, applications, challenges and trends for ontology learning from text. Information, Wissenschaft und Praxis, 57(6-7).

Cimiano, P., & Völker, J. (2005). Text2Onto. In *Natural language processing and information systems* (pp. 227-238). Springer Berlin Heidelberg.

Dahab, M. Y., Hassan, H. A., & Rafea, A. (2008). TextOntoEx: Automatic ontology construction from natural English text. *Expert Systems with Applications*, *34*(2), 1474-1480.

Dellschaft, K., & Staab, S. (2006). On how to perform a gold standard based evaluation of ontology learning. In *The Semantic Web-ISWC 2006* (pp. 228-241). Springer Berlin Heidelberg.

Ding, Y., & Foo, S. (2002). Ontology research and development. Part 1-a review of ontology generation. *Journal of information science*, *28*(2), 123-136.

Dukes, Kais. (2011). The Quranic Arabic Corpus. School of computing, university of leeds, UK http://corpus. Quran.com/ontology.jsp

Faure, D., & Nédellec, C. (1998, May). A corpus-based conceptual clustering method for verb frames and ontology acquisition. In *LREC workshop on adapting lexical and corpus resources to sublanguages and applications* (Vol. 707, No. 728, p. 30).

Faure, D., Nédellec, C., & Rouveirol, C. (1998). Acquisition of Semantic Knowledge using Machine learning methods: The System" ASIUM". In*Universite Paris Sud*.

Girju, R., Badulescu, A., & Moldovan, D. (2003, May). Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 1-8). Association for Computational Linguistics.

Girju, R., & Moldovan, D. I. (2002, May). Text Mining for Causal Relations. In *FLAIRS Conference* (pp. 360-364).

Greenwood, M. A., & Stevenson, M. (2006, July). Improving semi-supervised acquisition of relation extraction patterns. In Proceedings of the Workshop on Information Extraction Beyond The Document (pp. 29-35). Association for Computational Linguistics.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International journal of human-computer studies*, *43*(5), 907-928.

Giaretta, P., & Guarino, N. (1995). Ontologies and knowledge bases towards a terminological clarification. *Towards very large knowledge bases: knowledge building & knowledge sharing*, *25*, 32.

Gómez-Pérez, A., Corcho, O., & Fernandez-Lopez, M. (2004). Ontological engineering: with examples from the areas of knowledge management, e-Commerce and the Semantic Web. (advanced information and knowledge processing).

Gruninger, M., and Fox, M.S. (1995), "Methodology for the Design and Evaluation of Ontologies",. Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal.

Guarino, N. (1998). Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy (Vol. 46). IOS press.

Hahn, U., & Schnattinger, K. (1998). Towards text knowledge engineering.Hypothesis, 1(2).

Hamon, T., & Nazarenko, A. (2001). Detection of synonymy links between terms: experiment and results. Recent Advances in Computational Terminology, 2, 185-208, page 13.

Hasegawa, T., Sekine, S., & Grishman, R. (2004, July). Discovering relations among named entities from large corpora. In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (p. 415). Association for Computational Linguistics.

Hearst, M. A. (1992, August). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2* (pp. 539-545). Association for Computational Linguistics.

Imsombut, A. (2009, October). A statistical approach for semantic relation extraction. In *Natural Language Processing, 2009. SNLP'09. Eighth International Symposium on* (pp. 54-58). IEEE.

Kavalec, M., Maedche, A., & Svátek, V. (2004). Discovery of lexical entries for non-taxonomic relations in ontology learning. In *SOFSEM 2004: Theory and Practice of Computer Science* (pp. 249-256). Springer Berlin Heidelberg.

Khan, L., & Luo, F. (2002). Ontology construction for information selection. InTools with Artificial Intelligence, 2002.(ICTAI 2002). Proceedings. 14th IEEE International Conference on (pp. 122-127). IEEE.

Lin, D. (2003). Dependency-based evaluation of MINIPAR. In *Treebanks* (pp. 317-329). Springer Netherlands.

Liu, W., Weichselbraun, A., Scharl, A., & Chang, E. (2005). Semi-automatic ontology extension using spreading activation. *Journal of Universal Knowledge Management*, *1*, 50-58.

López, M. F., Gómez-Pérez, A., Sierra, J. P., & Sierra, A. P. (1999). Building a chemical ontology using methontology and the ontology design environment.IEEE intelligent Systems, (1), 37-46.

Lozano-Tello, A., & Gómez-Pérez, A. (2004). Ontometric: A method to choose the appropriate ontology. *Journal of Database Management*, *2*(15), 1-18.

Maedche, A. (2002). *Ontology learning for the semantic web*. Springer Science & Business Media.

Maedche, A., & Staab, S. (2000, July). Semi-automatic engineering of ontologies from text. In *Proceedings of the 12th international conference on software engineering and knowledge engineering* (pp. 231-239).

Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web.*IEEE Intelligent systems*, *16*(2), 72-79.

Maedche, A., & Volz, R. (2001, November). The ontology extraction & maintenance framework Text-To-Onto. In *Proc. Workshop on Integrating Data Mining and Knowledge Management, USA*.

Mann, G. S. (2002, September). Fine-grained proper noun ontologies for question answering. In Proceedings of the 2002 workshop on Building and using semantic networks-Volume 11 (pp. 1-7). Association for Computational Linguistics.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database*. *International journal of lexicography*, *3*(4), 235-244.

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

Morin, E. (1999, August). Automatic acquisition of semantic relations between terms from technical corpora. In *Proc. of the Fifth International Congress on Terminology and Knowledge Engineering-TKE'99*.

Muzaffar, A. W., Azam, F., & Qamar, U. (2015). A Relation Extraction Framework for Biomedical Text Using Hybrid Feature Set. Computational and mathematical methods in medicine, 2015.

Navigli, R., Velardi, P., & Gangemi, A. (2003). Ontology learning and its application to automated terminology translation. *Intelligent Systems, IEEE*,*18*(1), 22-31.

Navigli, R., & Velardi, P. (2004). Learning domain ontologies from document warehouses and dedicated web sites. *Computational Linguistics*, *30*(2), 151-179.

Nedellec, C. (2000). Corpus-based learning of semantic relations by the ILP system, Asium. In *Learning language in logic* (pp. 259-278). Springer Berlin Heidelberg.

Nédellec, C., & Nazarenko, A. (2006). Ontologies and information extraction.*arXiv preprint cs/0609137*.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.

Pantel, P., & Lin, D. (2002, July). Discovering word senses from text. In*Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 613-619). ACM.Pantel, P., & Ravichandran, D. (2004, May). Automatically Labeling Semantic Classes. In *HLT-NAACL* (Vol. 4, pp. 321-328).

Philpot, A., Fleischman, M., & Hovy, E. H. (2003, October). Semi-automatic construction of a general purpose ontology. In Proceedings of the International Lisp Conference. New York, NY. Invited.

Punuru, J., & Chen, J. (2006, May). Automatic acquisition of concepts from domain texts. In *GrC* (pp. 424-427).

Punuru, J., & Chen, J.. (2007). Extracting of non-hierarchical relations from domain texts. In *Computational Intelligence and data mining, CIDM, IEEE Symposium.* pp. 444-449.

Punuru, J., & Chen, J. (2012). Learning non-taxonomical semantic relations from domain texts. *Journal of Intelligent Information Systems*, *38*(1), 191-207.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program: electronic library and information systems*, *14*(3), 130-137.

Porzel, R., & Malaka, R. (2004, August). A task-based approach for ontology evaluation. In *ECAI Workshop on Ontology Learning and Population, Valencia, Spain*.

Ribeiro, M. (2014). Extraction of non-taxonomic relations from texts to enrich a basic ontology.

Riloff, E. (1996, August). Automatically generating extraction patterns from untagged text. In *Proceedings of the national conference on artificial intelligence* (pp. 1044-1049).

Roulin, C., & Cooper, C. (1993). Bringing Thesaurus Together: Terminological Problems and Interests. In *Terminology and Knowledge Engineering* (pp. 251-259).

Sánchez, D., & Moreno, A. (2008). Learning non-taxonomic relationships from web documents for domain ontology construction. *Data & Knowledge Engineering*, *64*(3), 600-623.

Schreiber, G. (2000). Knowledge engineering and management: the CommonKADS methodology. MIT press.

Schutz, A., & Buitelaar, P. (2005). Relext: A tool for relation extraction from text in ontology extension. In *The Semantic Web–ISWC 2005* (pp. 593-606). Springer Berlin Heidelberg.

Serra, I., & Girardi, R. (2011). A Process for Extracting Non-Taxonomic Relationships of Ontologies from Text. *Intelligent Information Management*,*3*(4).

Serra, I., Girardi, R., & Novais, P. (2013, April). PARNT: A Statistic based Approach to Extract Non-Taxonomic Relationships of Ontologies from Text. In*Information Technology: New Generations (ITNG), 2013 Tenth International Conference on* (pp. 561-566). IEEE.

Seung-bin Yim (2008). Semi-automatic generation of domain specific ontologies. Thesis. Technical University of Vienna.

Shamsfard, M., & Barforoush, A. A. (2004). Learning ontologies from natural language texts. *International journal of human-computer studies*, *60*(1), 17-63.

Stewart, T. R. (2001). Improving reliability of judgmental forecasts. InPrinciples of forecasting (pp. 81-106). Springer US.

Subhashini, R., & Akilandeswari, J. (2011). A survey on ontology construction methodologies. *International Journal of Enterprise Computing and Business Systems*, *1*(1), 60-72.

Sure, Y., Staab, S., & Studer, R. (2009). Ontology engineering methodology. InHandbook on ontologies (pp. 135-152). Springer Berlin Heidelberg.

Swartout, B., Patil, R., Knight, K., & Russ, T. (1996, November). Toward distributed use of large-scale ontologies. In Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems (pp. 138-148).

Tan, K. W., Han, H., & Elmasri, R. (2000). Web data cleansing and preparation for ontology extraction using WordNet. In Web Information Systems Engineering, 2000. Proceedings of the First International Conference on (Vol. 2, pp. 11-18). IEEE.

Tomokiyo, T., & Hurst, M. (2003, July). A language model approach to keyphrase extraction. In *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment-Volume 18* (pp. 33-40). Association for Computational Linguistics.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review*, *11*(02), 93-136.

Uschold, M. (2000, August). Creating, integrating and maintaining local and global ontologies. In Proceedings of the First Workshop on Ontology Learning (OL-2000) in conjunction with the 14th European Conference on Artificial Intelligence (ECAI-2000).

Uzun, Y. (2005). Keyword Extraction Using Naïve Bayes. In Bilkent University, Department of Computer Science, Turkey www. cs. bilkent. edu. tr/~ guvenir/courses/CS550/Workshop/Yasin_Uzun. pdf.

Villaverde, J., Persson, A., Godoy, D., & Amandi, A. (2009). Supporting the discovery and labeling of non-taxonomic relationships in ontology learning.*Expert Systems with Applications*, *36*(7), 10288-10294.

Walpole, D.E., Myers, R.H. & Myers S.L. (2012) Essentials of Probability and Statistics for Engineers and Scientists (1st ed): Prentice Hall

Yamaguchi, T. (2001, August). Acquiring Conceptual Relationships from Domain-Specific Texts. In *Workshop on Ontology Learning* (Vol. 38).

Yangarber, R., Grishman, R., Tapanainen, P., & Huttunen, S. (2000, July). Automatic acquisition of domain knowledge for information extraction. In*Proceedings of the 18th conference on Computational linguistics-Volume 2* (pp. 940-946). Association for Computational Linguistics.

Yildiz, B. (2007). Ontology-driven information extraction. PhD thesis, Ph.D. Thesis, Vienna University of Technology, 7-8