ENHANCED ONTOLOGICAL QUERY EXPANSION MODEL USING BIGRAM COMBINATIONS AND COMBINED SIMILARITY MEASURE FOR IMPROVING INFORMATION RETRIEVAL

MUHAMMAD AHSAN RAZA

DOCTOR OF PHILOSOPHY

UNIVERSITI MALAYSIA PAHANG

SUPERVISOR'S DECLARATION

We hereby declare that we have checked this thesis and in our opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Doctor of Philosophy.

_____

(Supervisor's Signature)

Full Name : DR. RAHMAH BINTI MOKHTAR

Position : SENIOR LECTURER

Date :

_____

(Co-supervisor's Signature)

Full Name : DR. NORAZIAH AHMAD

Position : ASSOCIATE PROFESSOR

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

_____

(Student's Signature)

Full Name : MUHAMMAD AHSAN RAZA

ID Number : PCC16022

Date :

ENHANCED ONTOLOGICAL QUERY EXPANSION MODEL USING BIGRAM COMBINATIONS AND COMBINED SIMILARITY MEASURE FOR IMPROVING INFORMATION RETRIEVAL

MUHAMMAD AHSAN RAZA

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy

Faculty of Computing

UNIVERSITI MALAYSIA PAHANG

JANUARY 2020

ACKNOWLEDGEMENTS

Great **ALLAH,** the most gracious, merciful and beneficent has given me the strength and make me able to fulfil the requirements of this research work successfully in a short period of time within limited resources and knowledge.

I express my gratitude and obligations to my Supervisor Dr. Rahmah Binti Mokhtar, who did her selfless and tireless efforts for providing me all facilities throughout the academic session and her kind and loving attitude always encouraged us to work hard. She guided me in critical times and helped me whenever I needed.

I feel great honour to express my sincere and cordial thanks to my respected Co-Supervisor Associate Professor Dr. Noraziah Binti Ahmad for her guidance, sympathetic attitude and solving problems that I faced in research completion period.

I would like to express my special thanks to Professor Dr. Mazlina Binti Abdul Majid who has been supportive in every way. Her guidance helped me in all the time of research and writing of this thesis.

At the end I feel very respect and happiness towards my family members who gave me lot of encourage and day-by-day prayers with love. Specially my mother prayed for my success and encouraged me to do the given task with full zeal and zest.

Muhammad Ahsan Raza

ABSTRAK

Ledakan maklumat di Web semakin bertambah dan berubah secara berterusan mengikut keadaan semasa. Oleh itu kaedah dapatan maklumat yang berkesan adalah penting bagi mengatasi fenomena ini. Kebanyakan sistem dapatan maklumat bagi pencarian data web yang besar antaranya enjin pencarian yang popular seperti Google, Bing atau Yahoo, bergantung kepada pertanyaan dan padanan kata kunci dokumen. Permasalahan yang timbul menggunakan padanan kata kunci yang ringkas adalah isu ketidakpadanan istilah pencarian yang ditaip oleh pengguna enjin pencarian dengan perkataan dalam senarai korpus. Sistem dapatan maklumat tidak berupaya menjangka pengguna menaip kata kunci yang tepat dan sama seperti senarai korpus, bagi mendapatkan dokumen yang releven dengan keperluan mereka. Bagi mengatasi masalah ini, beberapa usaha telah diambil antaranya pengembangan istilah kata kunci, di mana kata kunci pertanyaan yang asal ditambah dengan istilah yang releven. Beberapa tahun kebelakangan, Pengembangan Pertanyaan berasaskan Ontologi (OQE) telah digunakan sebagai model pertanyaan pengembangan semantik menggunakan ontologi berasaskan pengetahuan. Namun, terdapat masalah penggunaan model OQE sedia ada termasuk (i) Kekaburan pertanyaan pencarian dari bahasa tabii, (ii) Perkembangan kata kunci berasaskan istilah agar membantu pencarian pertanyaan tidak berstruktur berbanding gabungan pelbagai istilah pertanyaan, (iii) Pengembangan dalam pertanyaan pencarian menggunakan istilah yang kurang tepat. Objektif utama penyelidikan ini adalah untuk penambahbaikan model OQE sedia ada dan mencadangkan model Pengemaskinian Pengembangan Pertanyaan berasaskan Ontologi (EOQE) untuk dapatan maklumat yang lebih berkesan. Model EOQE berupaya mengembangkan konsep semantik pada pertanyaan pencarian bahasa tabii yang tidak berstruktur, agar dokumen berkaitan disiplin sains komputer yang releven dapat dicapai. Model ini dapat mengatasi limitasi model OQE sedia ada dengan tiga langkah utama. Pertama, memperhalusi pertanyaan yang dilakukan sewaktu pemprosesan bahasa pertanyaan pencarian dan menjana istilah carian yang sahih. Kedua, langkah penambahbaikan pengembangan istilah berasaskan ontologi, menapis pertanyaan pencarian dan menjana konsep pengembangan istilah menggunakan teknik gabungan bigram. Ketiga, langkah formulasi pertanyaan yang menapis istilah yang kurang tepat dari set konsep pengembangan dengan menggunakan teknik gabungan pengukuran persamaan. Penilaian prestasi model EOQE dilakukan dengan membandingkan keputusan dapatan dari pertanyaan pengembangan menggunakan model EOQE dengan pertanyaan asal (yang dipanggil sebagai model dasar). Menggunakan Model Ruang Vektor (VSM), yang merupakan piawai sistem Dapatan Maklumat (IR), keputusan pengunaan model EOQE menunjukkan peningkatan sebanyak 32% maksudnya purata kejituan berbanding model dasar dan mencapai 90% nilai dapatan dari keseluruh pertanyaan. Model ini juga menunjukkan pencapaian sebanyak 17% dan 15% penambahan P@20 dan P@40 berbanding model dasar melalui sistem pencarian Google. Selain itu, model ini menunjukkan prestasi kompetitif dalam nilai dapatan, kejituan, purata kejituan dan maksudnya purata kejituan berbanding pelbagai model EOQE berasaskan pengukuran persamaan tunggal. Antara sumbangan utama penyelidikan ini ialah teknik untuk pengembangan istilah semantik tidak berstruktur dan mengatasi pertanyaan bahasa tabii yang kabur menggunakan teknik kombinasi bigram dan teknik pengukuran persamaan kombinasi yang kukuh. Sumbangan ini membolehkan istilah pelbagai pertanyaan dieksploitasi dalam tatacara OQE berbanding menggunakan istilah pertanyaan tunggal dan memformulasikan pertanyaan pengembangan dengan konsep semantik yang lebih tepat.

ABSTRACT

The information explosion over the Web has been increasing and changing rapidly over the time, thus the effective retrieval of information is increasingly gaining in importance. Most Information Retrieval (IR) systems typically rely on query and document keyword matching, in order to search over huge amounts of Web data, examples being famous search engines such as Google, Bing or Yahoo. Problem arising with these simple keyword matching IR systems is vocabulary mismatch issue: the searcher's query terms may not be matched with those of the corpus. IR systems cannot always expect a user to type the exact keyword in query as present in corpus in order to obtain relevant documents. To deal with this issue, several efforts have been made such as query expansion, whereby the search query is expanded with additional relevant terms using original query keywords. In recent years, ontology based query expansion (OQE) emerges as an advance query expansion model to expand search query semantically using ontology knowledgebase. However, common problems with existing OQE model include (i) the inherent ambiguity of natural language search query, (ii) term-based expansion to support unstructured search query rather than considering multiple query terms together (iii) The expansion of search query with irrelevant terms. The main objective of this research is to improve existing OQE model and propose an enhanced ontological query expansion model (EOQE) for effective information retrieval. The EOQE model attempts to semantically expand unstructured natural language search queries in order to retrieve relevant documents for computer science discipline. The model overcomes the limitations of existing OQE model by following three main steps. First, the query refinement step performs linguistic processing of search query and generates valid search term. Second, the enhanced ontology based expansion step disambiguates the search query and generates the additional expansion concepts on the basis of bigram combinations technique. Third, the query formulation step filters irrelevant terms from expansion concepts set using combined similarity measure technique. The performance evaluation EOQE model was based on comparing the retrieval results of queries expanded with EOQE model and the original queries (called as baseline model). On Vector Space Model (VSM) standard IR system, the EOQE model showed 32% improvements in terms of mean average precision against baseline model, and achieved above 90% recall values for most of search queries. The EOQE model also attained a 17% and 15% increase in P@20 and P@40 values, respectively, than baseline model over famous Google search system. Furthermore, the EOQE model demonstrated competitive performance in terms of precision, recall, average precision, and mean average precision values against EOQE model variants based on single similarity measure. The main contributions of this research are to introduce a model to semantically expand unstructured and ambiguous natural language query using bigram combinations and strong combined similarity measure techniques. These contributions enable exploiting multiple query terms in procedure of OQE rather than using individual query terms, and formulating expanded queries with more relevant semantic concepts.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

QE	Query Expansion
IR	Information Retrieval
SQE	Semantic Query Expansion
POS	Parts-Of-Speech
OQE	Ontology-based Query Expansion
EOQE	Enhanced Ontological Query Expansion
SS	Semantic Similarity
NS	Neighbour Similarity
P	Precision
Avg. P	Average Precision
VSM	Vector Space Model
EOQE-ont	EOQE Model with Ontology-based Similarity Measure
EOQE-co	EOQE Model with Corpus-based Similarity Measure
EOQE-combine	EOQE Model with Combined Similarity Measure
OWL	Web Ontology Language
SL	Sense Learning
ECI	Expansion Concept Identification
CCS	Computing Classification System
NLP	Natural Language Processing

LIST OF APPENDICES

REFERENCES

Abdulla, AAA., Lin, H., Xu, B., & Banbhrani, SK. (2016) Improving Biomedical Information Retrieval by Linear Combinations of Different Query Expansion Techniques. BMC Bioinformatics 17 (S-7):238. doi:10.1186/s12859-016-1092-8

Acm (2019). The 2012 ACM Computing Classification System. Available at: https://www.acm.org/publications/class-2012, Accessed 20 Jan. 2019

Aggarwal, N., & Buitelaar, P. (2012). Query Expansion Using Wikipedia and Dbpedia. Paper presented at the CLEF (Online Working Notes/Labs/Workshop). http://dblp.uni-trier.de/db/conf/clef/clef2012w.html#AggarwalB12.

Aicha, G., Fatiha, B., Ghalem, B., & Farid, M. (2018). Query Expansion Using Medical Information Extraction for Improving Information Retrieval in French Medical Domain. International Journal of Intelligent Information Technologies (IJIIT), 14(3), 1-17. doi: 10.4018/ijiit.2018070101

Alromima, W., Moawad, I. F., Elgohary, R., & Aref, M. (2016). Ontology-based Query Expansion for Arabic Text Retrieval. International Journal of Advanced Computer Science and Applications(IJACSA), 8(7), 223-231. doi: http://dx.doi.org/10.14569/IJACSA.2016.070830.

Atoum, I., & Otoom, A. (2016). Efficient Hybrid Semantic Text Similarity Using Wordnet and a Corpus. International Journal of Advanced Computer Science & Applications, 7(1). doi: 10.14569/IJACSA.2016.070917.

Azad, H., & Deepak, A. (2017). Query Expansion Techniques for Information Retrieval: A Survey. CoRR, abs/1708.00247.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern Information Retrieval (Vol. 463): ACM press New York.

Bhogal, J., MacFarlane, A., & Smith, P. (2007). A Review of Ontology Based Query Expansion. Information processing & management, 43(4), 866-886. doi: 10.1016/j.ipm.2006.09.003

Bhowmick, P.K., Roy, D., Sarkar, S., & Basu, A. (2010). A Framework for Manual Ontology Engineering for Management of Learning Material Repository. IJCSA, 7, 30-51.

Biancalana, C., Gasparetti, F., Micarelli, A., & Sansonetti, G. (2013). Social Semantic Query Expansion. ACM Transactions on Intelligent Systems and Technology (TIST), 4(4), 60.

Bilotti, M. W., Katz, B., & Lin, J. (2004). What Works Better for Question Answering: Stemming or Morphological Query Expansion. Paper presented at the Proceedings of the Information Retrieval for Question Answering (IR4QA) Workshop at SIGIR.

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia - A Crystallization Point for the Web of Data. Web Semant., 7(3), 154-165. doi: 10.1016/j.websem.2009.07.002

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. Paper presented at the Proceedings of the 2008 ACM SIGMOD international conference on Management of data, Vancouver, Canada.

Brusa, G., Caliusco, M.L., & Chiotti, O. (2008). Towards Ontological Engineering: A Process for Building a Domain Ontology from Scratch in Public Administration. Expert Systems, 25, 484-503.

Buckley, C., Salton, G., Allan, J., & Singhal, A. (1995). Automatic Query Expansion Using SMART: Trec 3. Paper presented at the Overview of the 3th Text REtrieval Conference Trec-3, Gaithersburg.

Carpineto, C., & Romano, G. (2012). A Survey of Automatic Query Expansion in Information Retrieval. ACM Computing Surveys (CSUR), 44(1), 1.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. In Advances in neural information processing systems(pp. 288-296)

Chujai, P., Kerdprasop, N., & Kerdprasop, K. (2014). On Transforming the ER Model to Ontology Using Protégé OWL Tool. International Journal of Computer Theory and Engineering, 6, 484-489. doi: 10.7763/IJCTE.2014.V6.914

Conesa, J., Storey, V. C., & Sugumaran, V. (2008). Improving Web-query Processing Through Semantic Knowledge. Data & Knowledge Engineering, 66(1), 18-34.

Crimp, R., & Trotman, A. (2017). Automatic Term Reweighting for Query Expansion. Paper presented at the Proceedings of the 22nd Australasian Document Computing Symposium, Brisbane, QLD, Australia.

Dahab, M. Y., Alnofaie, S., & Kamel, M. (2018). A Tutorial on Information Retrieval Using Query Expansion. In K. Shaalan, A. E. Hassanien & F. Tolba (Eds.), Intelligent Natural Language Processing: Trends and Applications (pp. 761-776). Cham: Springer International Publishing.

Dahir, S., El Qadi, A., & Bennis, H. (2018). Enriching User Queries Using DBpedia Features and Relevance Feedback. Procedia Computer Science, 127, 499-504. doi: https://doi.org/10.1016/j.procs.2018.01.148

Deepak, G., & Priyadarshini, J. S. (2018). Personalized and Enhanced Hybridized Semantic Algorithm for Web Image Retrieval Incorporating Ontology Classification, Strategic Query Expansion, and Content-based Analysis. Computers & Electrical Engineering, 72, 14-25. doi: https://doi.org/10.1016/j.compeleceng.2018.08.020

Efthimiadis, E. N. (1996). Query Expansion. Annual review of information science and technology, 31, 121-187.

El-Midaoui, O., El Ghali, B., El Qadi, A., & Rahmani, M. D. (2018). Geographical Query Reformulation Using a Geographical Taxonomy and WordNet. Procedia Computer Science, 127, 489-498. doi: https://doi.org/10.1016/j.procs.2018.01.147

Ermakova, L., & Mothe, J. (2016). Query Expansion by Local Context Analysis. Paper presented at the CORIA-CIFED.

Evrim, V. & McLeod, D. (2014). Context-based Information Analysis for the Web Environment. Knowl Inf Syst 38: 109. https://doi.org/10.1007/s10115-012-0493-x

Fatima, M., & Pasha, M. (2016). A Non Trivial Approach for Ontology based Query Expansion Using Scoring. Australian Journal of Basic and Applied Sciences, 94-100.

Fox, E. A. (1983). Characterization of Two New Experimental Collections in Computer and Information Science Containing Textual and Bibliographic Concepts. Technical Report 83-561, Cornell University, Department of Computer Science.

Gabrilovich, E., Broder, A., Fontoura, M., Joshi, A., Josifovski, V., Riedel, L., & Zhang, T. (2009). Classifying Search Queries Using the Web as a Source of Knowledge. ACM Trans. Web, 3(2), 1-28. doi: 10.1145/1513876.1513877

Gad W., K., & Kamel, M., S. (2009). New Semantic Similarity Based Model for Text Clustering Using Extended Gloss Overlaps. In: Perner P. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM 2009. Lecture Notes in Computer Science, vol 5632. Springer, Berlin, Heidelberg

Gan, L., & Hong, H. (2015). Improving Query Expansion for Information Retrieval Using Wikipedia. International Journal of Database Theory and Application, 8(3), 27-40.

Granados, B. M., Luis, G. Á., Escudero S., Trillo R., Ilarri S., & Mena E. (2014) SQX-Lib: Developing a Semantic Query Expansion System in a Media Group. In: de Rijke M. et al. (eds) Advances in Information Retrieval. ECIR 2014. Lecture Notes in Computer Science, vol 8416. Springer, Cham

Gruber. T. R. (1995). Toward Principles for ohe Design Of Ontologies Used for Knowledge Sharing? International Journal of Human-Computer Studies 43 (5):907-928. doi:http://dx.doi.org/10.1006/ijhc.1995.1081

Guessoum D., Miraoui M., & Tadj C. (2016). A Modification of Wu and Palmer Semantic Similarity Measure. Paper presented at the UBICOMM 2016: The Tenth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies.

Guisado-Gámez, J., & Prat-Pérez, A. (2015). Understanding Graph Structure of Wikipedia for Query Expansion. Paper presented at the Proceedings of the GRADES'15.

Gulzar, Z., Leema, A. A., & Deepak, G. (2018). PCRS: Personalized Course Recommender System Based on Hybrid Approach. Procedia Computer Science, 125, 518-524. doi: https://doi.org/10.1016/j.procs.2017.12.067

Guo L., Su X., Zhang L., Huang G., Gao X., Ding Z. (2018) Query Expansion Based on Semantic Related Network. In: Geng X., Kang BH. (eds) PRICAI 2018: Trends in Artificial Intelligence. PRICAI 2018. Lecture Notes in Computer Science, vol 11013. Springer, Cham

He, B., & Ounis, I. (2009). Studying Query Expansion Effectiveness. Paper presented at the Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval, Toulouse, France.

He, Y., Li, Y., Lei, J., & Leung, C. H. C. (2016). A Framework of Query Expansion for Image Retrieval Based on Knowledge Base and Concept Similarity. Neurocomput., 204(C), 26-32. doi: 10.1016/j.neucom.2015.11.102

Ingason A. K., Helgadóttir, S., & Loftsson, H. (2008). A Mixed Method Lemmatization Algorithm Using a Hierarchy of Linguistic Identities (HOLI). Advances in Natural Language Processing, 205–216, 2008.

Ingrosso, F., & Polguère, A. (2015). How Terms Meet in Small-World Lexical Networks: The Case of Chemistry Terminology. Paper presented at the Terminology and Artificial Intelligence (TIA 2015).

Jansen, B. J., Spink, A., & Pedersen, J. (2005). A Temporal Comparison of Altavista Web Searching: Research Articles. J. Am. Soc. Inf. Sci. Technol. 56, 6, 559–570.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web. Inf. Process. Manage. 36, 2, 207–227.

Jena A. (2019). Jena Ontology API. Available at: https://jena.apache.org/index.html, Accessed 20 Jan., 2019.

Järvelin, K., Kekäläinen, J., & Niemi, T. (2001). ExpansionTool: Concept-based query expansion and construction. Information Retrieval, 4(3), 231-255.

Jing, Y., & Croft, W. B. (1994). An Association Thesaurus for Information Retrieval. Paper presented at the Proceedings of the 4th International Conference on Computer Assisted Information Retrieval -- Recherche d'Informations Assistee par Ordinateur (RIAO 94).

Jiyeon, C., Youkyoung, P., & Mun, Y. (2016, 18-20 Jan. 2016). A Hybrid Method for Retrieving Medical Documents with Query Expansion. Paper presented at the 2016 International Conference on Big Data and Smart Computing (BigComp).

Kalender, M., Dang, J., & Uskudarli, S. (2010). Unipedia: A Unified Ontological Knowledge Platform for Semantic Content Tagging and Search. Paper presented at the Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on.

Kamel, M., Lee, A., & Powers, E. (2007). Ontology-aided vs. Keyword-based Web Searches: a Comparative User Study. Paper presented at the Large Scale Semantic Access to Content (Text, Image, Video, and Sound).

Kapidakis, S., Mastora, A., & Peponakis, M. (2012). Query Expansion of Zero-Hit Subject Searches: Using a Thesaurus in Conjunction with NLP Techniques. Paper presented at the International Conference on Theory and Practice of Digital Libraries.

Kuzi S, Carmel D, Libov A, Raviv A (2017) Query Expansion for Email Search. Paper presented at the Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan

Lauro N.C., Verde R., Palumbo F. (2000) Factorial Methods with Cohesion Constraints on Symbolic Objects. In: Kiers H.A.L., Rasson JP., Groenen P.J.F., Schader M. (eds) Data Analysis, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg.

Lawrence, D. W., et al. (2006). Developing the Injury Prevention and Safety Promotion Thesaurus, International English Edition: An Interdisciplinary Tool for Indexing and Searching for Research Literature. Progress report 1." Safety Science 44(4): 279-296.

Leung, C. H., Li, Y., Milani, A., & Franzoni, V. (2014). Collective Evolutionary Concept Distance Based Query Expansion for Effective Web Document Retrieval. Paper presented at the International Conference on Computational Science and Its Applications.

Lexico (2019). English thesaurus. Available at: https://www.lexico.com/en/english-thesaurus, Accessed 03 Feb. 2019

Li H., & Xu J. (2014). Semantic Matching in Search. Foundations and Trends® in Information Retrieval 7 (5):343-469. doi:10.1561/1500000035

Li, W., Ganguly, D., & Jones, G. J. F. (2016). Using WordNet for Query Expansion: ADAPT @ FIRE 2016 Microblog Track. Paper presented at the FIRE.

Li, Y., Yang, Y., Wang, J., & Xu, W. (2018). Zero-Shot Transfer VQA Dataset. arXiv preprint arXiv:1811.00692.

Lo, R. T. W., He, B., & Ounis, I. (2005, January). Automatically Building a Stopword List for an Information Retrieval System. In Journal on Digital Information Management: Special Issue on the 5th Dutch-Belgian Information Retrieval Workshop (DIR) (Vol. 5, pp. 17-24).

Lu, M., Sun, X., Wang, S., Lo, D., & Duan, Y. (2015). Query Expansion via Wordnet for Effective Code Search. Paper presented at the Software Analysis, Evolution and Reengineering (SANER), 2015 IEEE 22nd International Conference.

Mahdabi, P., & Popescu-Belis, A. (2016). Comparing Two Strategies for Query Expansion in a News Monitoring System. In E. Métais, F. Meziane, M. Saraee, V. Sugumaran & S. Vadera (Eds.), Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings (pp. 267-275). Cham: Springer International Publishing.

Manning C., D. (2011). Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?. In: Gelbukh A.F. (eds) Computational Linguistics and Intelligent Text Processing. CICLing 2011. Lecture Notes in Computer Science, vol 6608. Springer, Berlin, Heidelberg

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60

Markov, I., & Rijke, M. d. (2019). What Should We Teach in Information Retrieval? SIGIR Forum, 52(2), 19-39. doi: 10.1145/3308774.3308780

Matuszek, C., Cabral, J., Witbrock, M. J., & DeOliveira, J. (2006). An Introduction to the Syntax and Content of Cyc. Paper presented at the AAAI Spring Symposium: Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering.

Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-based and Knowledge-based Measures of Text Semantic Similarity. Paper presented at the Proceedings of the 21st national conference on Artificial intelligence - Volume 1, Boston, Massachusetts.

Miller, G. A. (1995). WordNet: a Lexical Database for English. Commun. ACM, 38(11), 39-41. doi: 10.1145/219717.219748

Moawad, I., Alromima, W. & Elgohary, R. (2018). Bi-Gram Term Collocations-based Query Expansion Approach for Improving Arabic Information Retrieval, Arabian Journal for Science and Engineering, 2018, Volume 43, Issue 12, pp 7705–7718, https://doi.org/10.1007/s13369-018-3145-y

Moral, C., de Antonio, A., Imbert, R., & Ramírez, J. (2014). A Survey of Stemming Algorithms in Information Retrieval. Information Research: An International Electronic Journal, 19(1), n1.

Moreau, F., Claveau, V., & Sébillot, P. (2007). Automatic Morphological Query Expansion Using Analogy-based Machine Learning. Paper presented at the European Conference on Information Retrieval.

Musen, M. A., & Protégé Team (2015). The Protégé Project: A Look Back and a Look Forward. AI matters, 1(4), 4–12. doi:10.1145/2757001.2757003

Nakade, V., Musaev, A., & Atkison, T. (2018). Preliminary Research on Thesaurus-based Query Expansion for Twitter Data Extraction. Paper presented at the Proceedings of the ACMSE 2018 Conference, Richmond, Kentucky.

Navigli, R., & Velardi, P. (2017). An Analysis of Ontology-based Query Expansion Strategies. Paper presented at the Proceedings of the 14th European Conference on Machine Learning, Workshop on Adaptive Text Extraction and Mining, Cavtat-Dubrovnik, Croatia.

Ngom, A. N., Kamara-Sangare, F., & Lo, M. (2017, 14-16 Dec. 2017). Proposition of a Method for Assessing Similarity Between Two Ontologies. Paper presented at the 2017 International Conference on Computational Science and Computational Intelligence (CSCI).

Nguyen, H.-M., Tran, K.-N., & Vo, X.-V. (2015). GeTFIRST: Ontology-based Keyword Search Towards Semantic Disambiguation. International Journal of Web Information Systems, 11(4), 442-467. doi: 10.1108/ijwis-06-2015-0019

Oh, H. S., & Jung, Y. (2015). Cluster-based Query Expansion Using External Collections in Medical Information Retrieval. Journal of biomedical informatics, 58, 70-79.

OWL (2012). Web Ontology Language (OWL). Available at: http://www.w3.org/2001/sw/#owl, Accessed 20 Feb. 2019.

Pramukantoro, E. S., & Fauzi, M. A. (2016, 15-16 Oct. 2016). Comparative Analysis of String Similarity and Corpus-based Similarity for Automatic Essay Scoring System on E-Learning Gamification. Paper presented at the 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS).

Ptáček, T. (2012). Advanced Methods for Sentence Semantic Similarity. (Master), University of West Bohemia, Pilsen. Retrieved from http://www.itspy.cz/wp-content/uploads/2014/05/acmspy2012_submission_35.pdf

Raza, M. A., Mokhtar, R., & Ahmad, N. (2018). A Survey of Statistical Approaches for Query Expansion. Knowledge and Information Systems. doi: https://doi.org/10.1007/s10115-018-1269-8

Raza, M. A., Mokhtar, R., Ahmad, N., Pasha, M., & Pasha, U. (2019). A Taxonomy and Survey of Semantic Approaches for Query Expansion. IEEE Access, 7, 17823-17833. doi: 10.1109/ACCESS.2019.2894679

Remi, S., & Varghese, S. C. (2015). Domain Ontology Driven Fuzzy Semantic Information Retrieval. Procedia Computer Science, 46, 676-681. doi: https://doi.org/10.1016/j.procs.2015.02.122

Rivas, A. R., Iglesias, E. L., & Borrajo, L. (2014). Study of Query Expansion Techniques and Their Application in the Biomedical Information Retrieval. The Scientific World Journal, 2014.

Roder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. Paper presented at the Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China.

Ruban, S., Vanitha, T., & Behin Sam, S. (2016). Design and Implementation of an Interactive Query Expansion Methodology for Information Retrieval. ARPN Journal of Engineering and Applied Sciences, 11, 1106-1110.

Salton, G., & Buckley, C. (1988). Term-weighting Approaches in Automatic Text Retrieval. Information Processing & Management, 24(5), 513-523. doi: https://doi.org/10.1016/0306-4573(88)90021-0

Sam, B., & Shetty, A. (2015). A Hybrid Framework to Refine Queries Using Ontology. Indian Journal of Science and Technology, 8(24), 1.

Sodanil, M., Phonarin, P., & Porrawatpreyakorn, N. (2014). An Ontology-Based Query Expansion for an Agricultural Expert Retrieval System. Paper presented at the Proceedings of International Conference on Information Integration and Web-based Applications & Services, Vienna, Austria.

Song, M., Song, I.-Y., Hu, X., & Allen, R. B. (2017). Integration of Association Rules and Ontologies for Semantic Query Expansion. Data & Knowledge Engineering, 63(1), 63-75.

Stevenson, M. (2003). Word Sense Disambiguation: The Case for Combinations of Knowledge Sources. Stanford, CA.: CSLI Publications.

Storey, V. C., Burton-Jones, A., Sugumaran, V., & Purao, S. (2008). CONQUER: A Methodology for Context-aware Query Processing on the World Wide Web. Information Systems Research, 19(1), 3-25.

Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: A Core of Semantic Knowledge. Paper presented at the Proceedings of the 16th international conference on World Wide Web, Banff, Alberta, Canada.

Ullrich, M., Maier, A., & Angele, J. (2014). Taxonomie, Thesaurus, Topic Map, Ontologie – ein Vergleich White Paper.

Uma D. M. & Meera, G. G. (2014). Wordnet and Ontology Based Query Expansion for Semantic Information Retrieval in Sports Domain (Vol. 2).

Voorhees, E. M. (1994). Query Expansion Using Lexical-Semantic Relations. Paper presented at the Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, Dublin, Ireland.

Wang, H., Fan, H., Li, J., & Wang, J. (2014). Research on the Method of Semantic Query Expansion in Civil Aviation Emergency Domain Ontology. International Journal of Digital Content Technology and its Applications (JDCTA), 8(5).

Wang, H., Zhang, Q., & Yuan, J. (2017). Semantically Enhanced Medical Information Retrieval System: A Tensor Factorization Based Approach. IEEE Access, 5, 7584-7593. doi: 10.1109/ACCESS.2017.2698142.

Weber, R. A. (2002). Ontological Issues in Accounting Information Systems. In V. Arnold & S. Sutton (Eds.), Researching Accounting as an Information Systems Discipline (pp. 21). Sarasota, Florida: American Accounting Association.

Wu, H., Li, J., Kang, Y., & Zhong, T. (2018). Exploring Noise Control Strategies for UMLS-based Query Expansion in Health and Biomedical Information Retrieval. Journal of Ambient Intelligence and Humanized Computing. doi: 10.1007/s12652-018-0836-x

Wu, J., Ilyas, I., & Weddell, G. (2011). A Study of Ontology-based Query Expansion. Technical report CS-2011–04

Xiong, C., & Callan, J. (2015). Query Expansion with Freebase. Paper presented at the Proceedings of the 2015 International Conference on The Theory of Information Retrieval, Northampton, Massachusetts, USA.

Xu, Y., Jones, G. J. F., & Wang, B. (2009). Query Dependent Pseudo-relevance Feedback Based on Wikipedia. Paper presented at the Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, Boston, MA, USA.

Yunzhi, C., Huijuan, L., Shapiro, L., Travillian, R. S., & Lanjuan, L. (2016). An Approach to Semantic Query Expansion System Based on Hepatitis Ontology. Journal of Biological Research-Thessaloniki, 23(1), 11.

Zarrouk, M., Lafourcade, M., & Joubert, A. (2014). About Inferences in a Crowdsourced Lexical-Semantic Network. EACL 2014, 174.

Zhang, J., & Rasmussen, E. M. (2001). Developing a New Similarity Measure from Two Different Perspectives. Information Processing & Management, 37(2), 279-294. doi: https://doi.org/10.1016/S0306-4573(00)00027-3

Zhao, L., & Callan, J. (2010). Term Necessity Prediction. Paper presented at the Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada.

Zhao, Y., Fesharaki, N.J., Li, X. et al. (2018). Semantic-Enhanced Query Expansion System for Retrieving Medical Image Notes. J Med Syst, 42: 105. https://doi.org/10.1007/s10916-018-0954-1

Zhu D., Wu S., Carterette B., & Liu H. (2014). Using Large Clinical Corpora for Query Expansion in Text-based Cohort Identification. Journal of Biomedical Informatics, 49 (Supplement C):275-281. doi:https://doi.org/10.1016/j.jbi.2014.03.010

Zingla, M. A., Chiraz, L., & Slimani, Y. (2016). Short Query Expansion for Microblog Retrieval. Procedia Computer Science, 96, 225-234. doi: https://doi.org/10.1016/j.procs.2016.08.135

Zingla, M. A., Latiri, C., Mulhem, P., Berrut, C., & Slimani, Y. (2018). Hybrid Query Expansion Model for Text and Microblog Information Retrieval. Information Retrieval Journal, 21(4), 337–367. doi: https://doi.org/10.1007/s10791-017-9326-6.

Zohar, H., Liebeskind, C., Schler, J., & Dagan, I. (2013). Automatic Thesaurus Construction for Cross Generation Corpus. Journal on Computing and Cultural Heritage (JOCCH), 6(1), 4.