Abstract

In recent years, semantic similarity measure has a great interest in Semantic Web and Natural Language Processing (NLP). Several similarity measures have been developed, being given the existence of a structured knowledge representation offered by ontologies and corpus which enable semantic interpretation of terms. Semantic similarity measures compute the similarity between concepts/terms included in knowledge sources in order to perform estimations. This paper discusses the existing semantic similarity methods based on structure, information content and feature approaches. Additionally, we present a critical evaluation of several categories of semantic similarity approaches based on two standard benchmarks. The aim of this paper is to give an efficient evaluation of all these measures which help researcher and practitioners to select the measure that best fit for their requirements.
- Patwardhan, S. 2003. “Incorporating dictionary and corpus information into a context vector measure of semantic relatedness,” Master’s thesis, Univ. Minnesota, Minneapolis.

- Nguyen, H. A. and Al-Mubaid, H. 2006. “New ontology-based semantic similarity measure for the biomedical domain,” in Proc. IEEE GrC, pp. 623–628.

- Jorge, G. , and Eduardo, M. , 2008. “Web-Based Measure of Semantic Relatedness,” 9th international conference on Web Information Systems Engineering, Springer-Verlag Berlin, pp. 136-150.

- Li Y. , Bandar Z. A. , and McLean D. 2003. An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources. IEEE Trans. on Knowledge and Data Engineering, 15(4), 871-882.

- Resnik O. 1999. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity and Natural Language. Journal of Artificial Intelligence Research, 11, 95-130.

- Beckwith, R. and Miller, G. A. 1992. Implementing a Lexical Network. Technical Report 43, Princeton University.

- Miller, G. A, Bechwith, R. , Felbaum, C. , Gross, D. and Miller, K. 1990. Introduction to WordNet: an on-line lexical database. International Journal of Lexicography, 3(4):235-244.

- Knight, K. and Luk, S. 1994. Building a Large-Scale Knowledge Base for Machine Translation. In Proceedings of the National Conference on Artificial Intelligence (AAAI’94), Seattle, WA.

- Reed, S. and Lenat, D. 2002. Mapping Ontologies into Cyc. In Proceedings of the AAAI’02 Conference Workshop on Ontologies For The Semantic Web, Edmonton, Canada, July.

- Nelson, S. J. , Powell, T. and Humphreys, B. L. 2002. The Unified Medical Language System (UMLS) Project. In A. Kent and C. M. Hall, editors, Encyclopedia of library and Information Science, pages 369{378. Marcel Dekker, Inc. , New York.

- Nelson, S. J. Johnston, D. and Humphreys, B. L. 2001 Relationships in Medical Subject Headings. In C. A. Bean and R. Green, editors, Relationships in the Organization of Knowledge, pages 171{184. Kluwer Academic Publishers, New York.

- Hill, D. P. , Blake, J. A, Richardson, J. E and Ringwald, M. 2002. Extension and Integration of the Gene Ontology (GO): Combining GO vocabularies with external vocabularies. Genome Res, 12:1982-1991.

- Rada, R. , Mili, H. , Bicknell, E. , and Blettner, M. 1989. Development and Application of a Metric on Semantic Nets. IEEE Transactions on Systems, Man, and Cybernetics, 19(1):17-30, January/February.

- Richardson, R, Smeaton, A. & Murphy, J. 1994. Using WordNet as a Knowledge Base for Measuring Semantic Similarity Between Words. Technical Report Working paper CA-1294, School of Computer Applications, Dublin City University, Dublin, Ireland.

- Hirst, G. and St-Onge, D. 1998. Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms. In Proceedings of Fellbaum, pages 305-332.

- Choudhari, M. . 2012. Extending the hirst and st-onge measure of semantic relatedness for the unified medical language system. Master Thesis.

- Z. Wu and M. Palmer. 1994. Verb Semantics and Lexical Selection. In Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics (ACL’94), pages 133{138, Las Cruces, New Mexico.
- Slimani, T. Ben Yaghlane, B. and Mellouli, K. 2006. "A New Similarity Measure based on Edge Counting." World Academy of Science, Engineering and Technology, PP 34-38.
- Li, Y., Bandar, Z. A and McLean, D. 2003. An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources. IEEE Transactions on Knowledge and Data Engineering, 15(4):871-882, July/August.
- Leacock, C and Chodorow, M. 1994. Filling in a sparse training space for word sense identification. ms. March.
- Lord, P. W., Stevens, R. D., Brass, A., and Goble, C. A. 2003. Investigating Semantic Similarity Measures across the Gene Ontology: the relationship between Sequence and Annotation. Bioinformatics, 19(10):1275-83.
- Lin, D. 1993. Principle-Based Parsing Without Overgeneration. In Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics (ACL'93), pages 112-120, Columbus, Ohio.
- Jiang, J. J. and Conrath. D. W., 1998. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In Proceedings of the International Conference on Research in Computational Linguistic, Taiwan.
- Tversky, A. 1977. Features of Similarity. Psychological Review, 84(4):327-352.
- Pirró, G. M. , Varelas, G., Hliaoutakis, A., & Raftopoulou, P. 2006. X-Similarity: Computing Semantic Similarity between Concepts from Different Ontologies. Journal of Digital Information Management, 4, 233-237.
- Rodriguez, M. A. and Egenhofer, M. J. 2003. Determining Semantic Similarity Among Entity Classes from Different Ontologies. IEEE Transactions on Knowledge and Data Engineering, 15(2):442-456, March/April.
- Knappe, R., Bulskov, H. and Andreasen, T. 2003. On Similarity Measures for Content-Based Querying. In O. Kaynak, editor, Proceedings of the 10th International Fuzzy Systems Association World Congress (IFSA'03), pages 400-403, Istanbul, Turkey, 29 June - 2 July.
- Zhou, Z., Wang, Y. and Gu, J., 2008. "New Model of Semantic Similarity Measuring in Wordnet." Proceedings of the 3rd International Conference on Intelligent System and Knowledge Engineering, November 17-19, Xiamen, China.
- Rubenstein, H., & Goodenough, J. 1965. Contextual correlates of synonymy. Communications of the ACM, 8, 627-633.
- Miller, G. A., & Charles, W. G. 1991. Contextual correlates of semantic similarity. Language and Cognitive Processes, 6, 1-28.
- Resnik, P. 1995. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In C. S. Mellish (Ed.), 14th International Joint Conference on Artificial Intelligence, IJCAI 1995 (Vol. 1, pp. 448-453). Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc.
- Pirró, G. 2009. A semantic similarity metric combining features and intrinsic information content. Data & Knowledge Engineering, 68, 1289-1308.
- Wan, S., & Angryk, R. A. 2007. Measuring Semantic Similarity Using WordNet-based Context Vectors. In M. El-Hawary (Ed.), IEEE International Conference on Systems, Man and Cybernetics, SMC 2007 (pp. 908 - 913). Montreal, Quebec, Canada: IEEE Computer Society.
- Patwardhan, S., & Pedersen, T. 2006. Using WordNet-based Context Vectors to
Estimate the Semantic Relatedness of Concepts. In EACL 2006 Workshop on Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics Together (pp. 1-8). Trento, Italy.

Index Terms

Computer Science
Applied Mathematics

Keywords
Similarity Measure, structure-based measures, edge-counting, feature-based measures, hybrid measures, Wornet, MeSH ontology