WordNetGraph: Structuring WordNet Natural Language Definitions

Vivian Silva, Santos , Jelena Mitrovic[*1], and Siegfried Handschuh

[1]Faculty of Computer Science and Mathematics, University of Passau – Germany

Abstract

WordNetGraph: Structuring WordNet Natural Language Definitions
WordNet is largely used as a linguistic resource in a number of semantic tasks, such as Question Answering, Information Retrieval, Text Entailment, etc., but systems usually query only the links between terms, such as synonym, hypernym or derivational form relationships. The synsets' definitions are usually left aside, although they contain a large amount of relevant information. These natural language definitions can serve as a rich source of knowledge, but structuring them into a comprehensible semantic model is essential for making them useful in semantic interpretation tasks.

In order to allow the use of WordNet's natural language definitions as a structured knowledge source in NLP tasks, we developed the WordNetGraph, a graph knowledge base built according to the methodology described in [1]. WordNetGraph builds upon a conceptual model based on entity-centered semantic roles for definitions [2], that is, roles that express the part played by an expression in a definition, showing how it relates to the definiendum, i.e., the entity being defined. This model extends the classic Aristotle's genus-differentia definition pattern [3, 4, 5]: the genus concepts is replaced by the supertype role (the definiendum's superclass, immediate or not); the essential properties represented by the differentia concept is split into the differentia quality and differentia event roles; and other roles, such as associated fact, purpose or accessory quality, among others, represent the definiendum's non-essential attributes.

For building the graph, a small sample of WordNet definitions was first automatically pre-annotated, using the syntactic patterns described in [2] to assign the suitable semantic roles to each segment in a definition, and then manually curated to create a training dataset. This dataset was used to train a machine learning classifier [6], which was later used to label all WordNet noun and verb definitions. After a post-processing phase to fix minor errors in the sequence of labels, the classified data was then serialized in RDF format. Figure 1 shows an example of labeled definition (for the WordNet synset "lake poets"). The same labeled definition is depicted in the final graph format in Figure 2.

WordNetGraph was primarily designed for and successfully used in an interpretable text entailment recognition approach for providing human-readable justifications for the entailment decision. Using an algorithm based on distributional semantics [7] to navigate the graph, we look for a path linking the entailing text T to the entailed hypothesis H. If we succeed, then the entailment is confirmed, and the contents of the nodes in the retrieved path are used to build a natural language justification that explains why the entailment is true and what exactly the semantic relationship between T and H is. The complete description of the text entailment recognition approach, including evaluation results and justification examples can

---

[*]Speaker

be found in [8].

Figure 1. Example of role labeling for the definition of "lake poets"

Figure 2. RDF representation for the definition of "lake poets"

In future work, this methodology will be applied to GermaNet, and it will also include the adjective synsets because they are organized hierarchically in the lexico-semantic network for the German language.

References

Silva, V. S., Freitas, A., and Handschuh, S. (2018). Building a Knowledge Graph from Natural Language Definitions for Interpretable Text Entailment Recognition. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).

Silva, V. S., Handschuh, S., and Freitas, A. (2016). Categorization of semantic roles for dictionary definitions. In Cognitive Aspects of the Lexicon (CogALex-V), Workshop at COLING 2016, pages 176–184.

Berg, J. (1982). Aristotle's theory of definition. ATTI del Convegno Internazionale di Storia della Logica, pages 19–30.

Granger, E. H. (1984). Aristotle on genus and differentia. Journal of the History of Philosophy, 22(1):1–23.

Lloyd, A. C. (1962). Genus, species and ordered series in Aristotle. Phronesis, pages 67–90.

Mesnil, G., Dauphin, Y., Yao, K., Bengio, Y., Deng, L., Hakkani-Tur, D., He, X., Heck, L., Tur, G., Yu, D., et al. (2015). Using recurrent neural networks for slot filling in spoken language understanding. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 23(3):530–539.

Freitas, A., da Silva, J. a. C. P., Curry, E., and Buitelaar, P. (2014). A distributional semantics approach for selective reasoning on commonsense graph knowledge bases. In International Conference on Applications of Natural Language to Data Bases/Information Systems, pages 21– 32. Springer.

Silva, V. S., Freitas, A., and Handschuh, S. (2018). Recognizing and justifying text entailment through distributional navigation on definition graphs. In AAAI.