WordNetContext: Information Retrieval-friendly Access to WordNet Senses

Chumki Basu
Perspecta Labs
cbasu@perspectalabs.com

Laura Dietz
University of New Hampshire
dietz@cs.unh.edu

Christiane Fellbaum
Princeton University
fellbaum@Princeton.edu

Knowledge graphs have shown to be effective in improving information retrieval effectiveness, in particular together with entity linking [3, 6, 13, 10], which sets a new standard for the Robust 2004. When utilizing knowledge graphs and semantic annotations in information retrieval, two of the most useful features are the full text of the Wikipedia article and the textual context surrounding entity links [3, 6]. For a given free-text query, both Wiki-text and entity link contexts effectively support the retrieval of relevant entities; they constitute a rich source for query-specific expansion terms and entity-aware text relevance features.

We are now adjusting this approach to better utilize WordNet for information retrieval. WordNet is a lexical database that has been curated manually over several decades according to psycholinguistic and computational theories of human lexical memory [7]. The major hurdle is that WordNet is a "vertical" resource, describing a taxonomic hierarchy of terms, where for information retrieval we also require "horizontal" information, i.e., access to other contextually related words for the same word sense. Currently, the only horizontal information available in WordNet are short glosses. Princeton's SemCor [8] constitutes an early attempt to link text tokens to the appropriate WordNet synsets. However, this resource is small and the annotated text is somewhat outdated.

While manually selected synsets show improvements for retrieval [12], fully automated approaches either expand with all synsets or include expensive word sense disambiguation into the retrieval step [9]. Here we are investigating a third approach by building a "horizontal" resource: We apply word sense disambiguation [9] to large corpora and extract contexts surrounding disambiguated word senses. We construct WordNetContext, an auxiliary text resource to accompany WordNet, by associating each word sense with (1) the gloss and (2) all sense contexts. This new resource enables fast and efficient identification of the WordNet sense that is relevant to a keyword query, simply by indexing and retrieving from this resource. As a result, we obtain a reliable means for fully automated query expansion through disambiguated synonyms.

We use this approach to cross-reference knowledge graphs with relevant WordNet senses. As depicted in Figure 1, these cross-references are based on the similarity between Wiki-text and entries in our WordNetContext resource. Finally, the WordNetContext resource text will be overlaid with annotations of WordNet's morphosyntactic, and semantic relations [4]. Since many queries, corpora, WordNet and Wikipedia are multi-lingual, we also envision various feedback mechanisms relevant for cross-language information retrieval.

At its core, the new WordNetContext resource provides an ecosystem for the exchange of sense mappings and relations, including "horizontal" information about co-occurring terms, phrases, and Wikipedia entities. Therefore, we believe that the availability of WordNetContext will crucially increase the usefulness of the WordNet resource for information retrieval and text understanding. To the best of our knowledge, previous works [5, 11, 1, 2] have not explored such a ressource for disambiguation and expansion in retrieval.

References

[1] Cao, G., Nie, J.Y., Bai, J.: Integrating word relationships into language models. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM (2005) 298–305

[2] Collins-Thompson, K., Callan, J.: Query expansion using random walk models. In: Proceedings of the 14th ACM international conference on Information and knowledge management. (2005) 704–711

In: Joint Proceedings of the First International Workshop on Professional Search (ProfS2018); the Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR); and the International Workshop on Data Search (DATA:SEARCH'18). Co-located with SIGIR 2018, Ann Arbor, Michigan, USA – 12 July 2018, published at http://ceur-ws.org

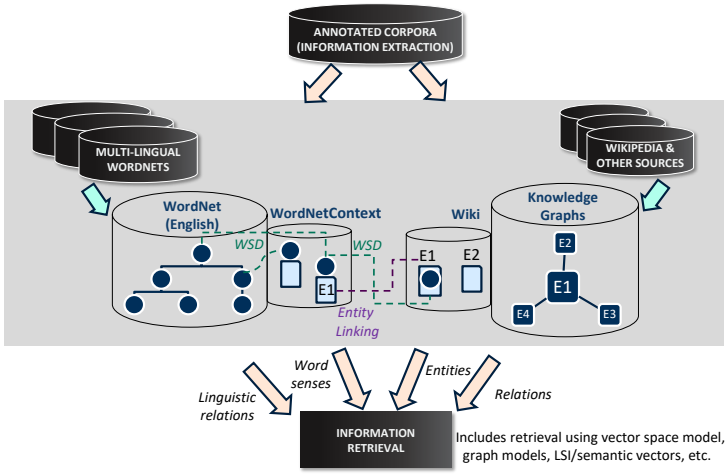

Figure 1: Cross-referencing knowledge graphs and WordNet through WordNetContext, word sense disambiguation (WSD), and entity linking for KG-aware information retrieval.

[3] Dalton, J., Dietz, L., Allan, J.: Entity query feature expansion using knowledge base links. In: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. (2014) 365–374

[4] Fellbaum, C., Osherson, A., Clark, P.E.: Putting semantics into wordnet's" morphosemantic" links. In: Language and Technology Conference, Springer (2007) 350–358

[5] Kotov, A., Zhai, C.: Tapping into knowledge base for concept feedback: leveraging conceptnet to improve search results for difficult queries. In: Proceedings of the fifth ACM international conference on Web search and data mining. (2012) 403–412

[6] Liu, X., Fang, H.: Latent entity space: a novel retrieval approach for entity-bearing queries. Information Retrieval Journal **18**(6) (2015) 473–503

[7] Miller, G., Fellbaum, C.: Wordnet: An electronic lexical database (1998)

[8] Miller, G.A., Chodorow, M., Landes, S., Leacock, C., Thomas, R.G.: Using a semantic concordance for sense identification. In: Proceedings of the workshop on Human Language Technology, ACL (1994) 240–243

[9] Navigli, R.: Word sense disambiguation: A survey. ACM Computing Surveys (CSUR) **41**(2) (2009) 10

[10] Raviv, H., Kurland, O., Carmel, D.: Document retrieval using entity-based language models. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. (2016) 65–74

[11] Shah, C., Croft, W.B.: Evaluating high accuracy retrieval techniques. In: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. (2004) 2–9

[12] Voorhees, E.M.: Query expansion using lexical-semantic relations. In: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. (1994) 61–69

[13] Xiong, C., Callan, J.: Esdrank: Connecting query and documents through external semi-structured data. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. (2015) 951–960