ASKNet: Leveraging Bio-Cognitive Models in Natural Language Processing

Brian Harrington

University of Wisconsin-Milwaukee
Medical Informatics Group
Milwaukee, WI 53201

Abstract. This presentation will detail the inception, development and evaluation of ASKNet. A system which uses natural language processing tools in order to create psycholinguistically inspired, spreading activation based semantic networks from natural language texts.

Presentation Overview

Until recently, we lacked the tools, corpora and computational power necessary to make use of biologically inspired cognitive models in the field of natural language processing[8]. However, recent advancements in technologies such as dependency parsers[2] and semantic analyzers[1] have greatly improved the semantic content that can be extracted from text, and we now have the resources and techniques necessary to once again revisit the idea of combining computational linguistics with bio-cognitive structures.

ASKNet is a system for automatically generating semantic knowledge networks from English text[3]. Its networks are psycholinguistically based, and rely on spreading activation theory[9] for entity resolution, and semantic disambiguation, which allows ASKNet to combine information from multiple documents into a single cohesive knowledge resource. That resource once built, can further leverage its spreading activation structure in order to complete a wide variety of tasks, producing results generally on par with custom built task-oriented systems.

This presentation will discuss ASKNet's cognitive model, and the way in which it uses spreading activation to determine co-referent entities when combining information from multiple documents. We will show that the use of spreading activation in this context, not only provides a solid underlying model, but also allows the computational complexity of the network integration stage to grow linearly with the size of the network, thus ensuring efficient production of very large scale networks. We will discuss manual evaluation of these networks, which produced a precision score of nearly 80%[7], as well as the production of large scale networks, based on the British National Corpus, Wordnet and Wikipedia that have been used to produce competitive results in standard semantic relatedness tasks[5,10]. Finally, we will discuss the potential for these networks to utilize

their cognitive models in order to accomplish tasks that are difficult or impossible for traditional computational linguistics systems, such as novel fact discovery, world-view based analysis and inter-document question answering[4,6].

ASKNet represents both a system for leveraging biologically inspired cognitive models in the advancement of computational linguistics, as well as a system which produces psycholinguistic inspired models on a scale limited only by the available corpora. Thus, it has the potential to act as a bridge between the two communities, combining the computational benefits and endless innovative abilities of cognitive architectures, with the corpora and information access of computational linguistics.

References

[1] Johan Bos, Stephen Clark, Mark Steedman, James R. Curran, and Julia Hockenmaier. Wide-coverage semantic representations from a CCG parser. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING-04)*, pages 1240–1246, Geneva, Switzerland, 2004.

[2] S. Clark and J. R. Curran. Wide-coverage efficient statistical parsing with CCG and log-linear models. *Computational Linguistics*, 33(4):493–552, 2007.

[3] B. Harrington and S. Clark. Asknet: Automated semantic knowledge network. In *Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI'07)*, pages 889–894, Vancouver, Canada, 2007.

[4] Brian Harrington. Managing uncertainty, importance and differing world-views in asknet semantic networks. In *Proceedings of the fourth IEEE International Conference on Semantic Computing*, Pittsburgh PA, USA, 2010.

[5] Brian Harrington. A semantic network approach to measuring relatedness. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, Beijing, China, 2010.

[6] Brian Harrington. Using asknet semantic networks to discover novel biomedical relations. In *Proceedings of the 4th International Symposium On Applied Sciences In Biomedical And Communication Technologies (ISABEL 2011)*, Barcelona, Spain, 2011.

[7] Brian Harrington and S Clark. Asknet: Creating and evaluating large scale integrated semantic networks. *International Journal of Semantic Computing*, 2(3):343–364, 2009.

[8] S Preece. *A Spreading Activation Model for Information Retrieval*. PhD thesis, University of Illinois, Urbana, IL, 1981.

[9] G. Salton and C. Buckley. On the use of spreading activation methods in automatic information retrieval. In *Proceedings of the 11th annual international ACM SIGIR conference on research and development in information retrieval*, pages 147 – 160, New York, NY, USA, 1988. ACM Press.

[10] Pia-Ramona Wojtinnek and Stephen Pulman. Semantic relatedness from automatically generated semantic networks. In *Proceedings of the 9th International Conference on Computational Semantics*, Oxford, UK, 2011.