Optimising Parameters for ASKNet: A Large Scale Semantic Knowledge Network Creation System

Brian Harrington, Simon Kempner
University of Oxford Department of Computer Science
Keble College Oxford
Oxford, United Kingdom
Email: brian.harrington@cs.ox.ac.uk
Email: simon.kempner@keble.oxon.org

***Abstract*—ASKNet is a system for automatically constructing semantic knowledge networks from natural language text. ASKNet uses existing natural language processing tools to extract entities and relations from text, and then through a combination of lexical information and a novel use of spreading activation, combines that information into a large scale semantic knowledge network. The ASKNet system is large, and quite complex. Historically, users of the system have had to rely on a combination of intuition and empirical evaluation of small sample networks in order to obtain reasonable settings for the various system parameters. In this paper, we develop a testing harness and gold standard that allow us to use simple machine learning methods to find optimal settings for all of the system's parameters. This system also aids future development of internal system algorithms, and can be adapted easily to novel domains.**

***Keywords*-Semantic Networks; Natural Language Processing; Spreading Activation; Knowledge Networks; ASKNet; Parameter Optimisation**

I. Introduction

ASKNet is a system for automatically generating large scale semantic resources using information derived from natural language texts. Using a combination of existing natural language processing tools and a novel application of spreading activation, ASKNet builds semantic networks representing the information contained within a text, and then maps that information onto a larger network representing the sum of its world knowledge.

ASKNet has been in development since 2005, and has been shown to produce good results on a variety of tasks, such as Semantic Relatedness [1], [2], and conceptual knowledge acquisition [3]. The integrated semantic nature of ASKNet also makes it ideal for information management and knowledge discovery [4], [5].

Large systems such as ASKNet necessarily have various parameters which must be optimised in order to obtain the best possible results from the system as a whole. During the development of ASKNet, the parameters controlling elements such as the spreading activation and lexical matching were set by the developers based on their own intuition and unit testing. While the system was being refined, small test networks were built to help developers find appropriate values for these parameters, but due to the large scale nature of the project, it was not feasible for any developer to manually configure all of the parameters. Thus, the values were always set to very rough approximations, and even when running tests on new data sets, the system's configuration was often left in the same state as it had been for previous experiments [6].

This paper details the development of a "gold standard" data set, and testing harness for ASKNet, and the use of an evolutionary based hill-climbing algorithm. The combination of these tools allows us to automatically and find optimal settings for parameters. We then use these improved parameters to repeat a previously published experiment, and find an improvement in both precision and running time.

II. ASKNet

ASKNet uses a combination of natural language processing tools such as the C&C parser [7], and the semantic analysis tool Boxer [8] in order to produce discourse representation structures. These structures are then converted into semantic network fragments as seen in Figure 1. The network fragments are based on an entity relationship paradigm, with nesting to allow entities and relations to be combined to form concepts, which can in turn be combined to form structures of increasing complexity (See Figure 2).

Once the semantic network fragment has been created for a piece of text, ASKNet then uses a Spreading Activation based algorithm [4] in order to determine the appropriate mappings between nodes in the fragment and nodes in the global knowledge network.

A. Spreading Activation

In order to integrate the semantic network fragments into the larger knowledge network, ASKNet uses the *update algorithm*, which is based on the psycholinguistic principles of Spreading Activation [9]. Spreading activation works by considering ASKNet networks as having similar properties to neural networks. By placing an amount of activation in a node, and allowing that node to fire, it can spread the

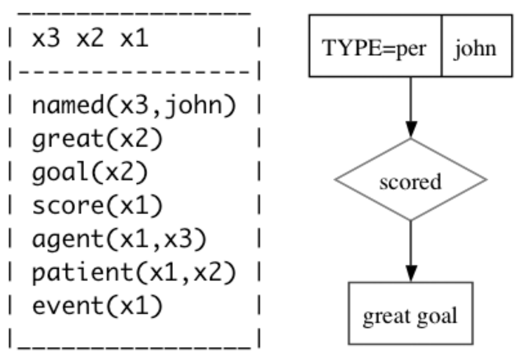

Figure 1. Boxer output, and corresponding semantic network fragment for the sentence "John scored a great goal".

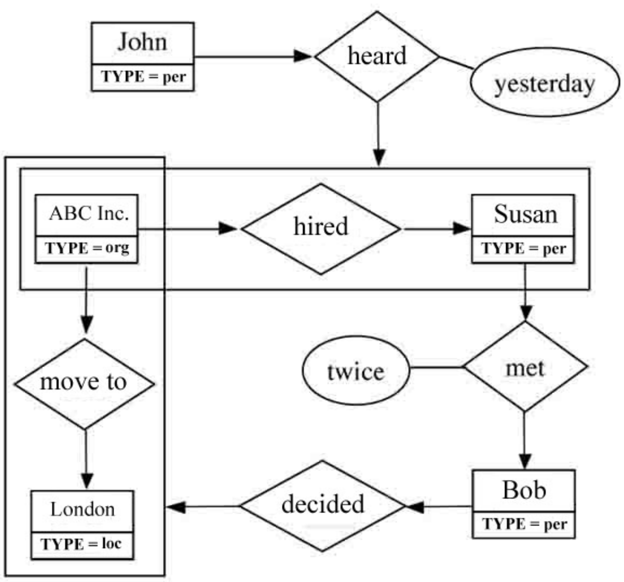

Figure 2. A sample ASKNet network, showing simple elements combining to form more complex concepts.

activation to its neighbouring nodes, the amount passing to each neighbour being relative to the strength of the relation connecting them.

In the update algorithm, ASKNet first uses lexical similarity to get base mapping scores for all named entity node pairs. A small amount of activation is then placed in a *source* node in the network fragment. The activation is allowed to spread through the fragment, settling on various fragment nodes dependant on their relatedness to the source node. The current mappings are then used to transfer that activation from the fragment nodes to corresponding nodes in the main knowledge network, the amount being transferred being dependent on the current mapping score, and the main network is then allowed to fire. The amount of activation received at the end of this process by the main network *target* nodes will determine the update to the (*source*,*target*) mapping score.

Figure 3 shows an example of the update algorithm in progress. An initial mapping score will be created between the pairs (`bu`,`georgebush`) and (`bu`,`johnbush`) based on their lexical similarity (string similarity + named entity type). In order to improve these scores, `bu` is selected as the source node, and given activation which will spread to `go` and `wh` dependant on the strength of the "beat" and "to" relationships. The activation from these nodes will be sent to `whitehouse`, `algore` and `gorevidal` respectively based on their relative mapping score. The main network will then be able to fire, resulting in activation filtering to the `georgebush` node, while the `johnbush` node receives no activation. Thus, the mapping score for (`bu`,`georgebush`) will increase, and the score for (`bu`,`johnbush`) will decrease. This process will continue until the scores reach a stable state, or cross a threshold at which time the nodes will be mapped together.

The update algorithm allows ASKNet to integrate information from a variety of sources into a single cohesive semantic network. Spreading activation has the advantage of being localised, and thus relatively efficient, while at the same time taking into account the relative strength of multiple paths of varying lengths that may connect node pairs.

B. Previous Evaluation

In a previous paper [6], ASKNet networks were created from documents provided in the 2006 Document Understanding Conference [10]. Each of the 5 networks, each containing information from 25 documents was then given to 3 judges in order to evaluate their quality. The judges were asked to evaluate the paths between each pair of named entities in the network, and mark the path as either "entirely correct" (all entities, mappings and relations were correct), or "incorrect" (there was an error of any type).

In the original experiment, it was found that manual evaluation of the entire network was impractical, and therefore the evaluation was performed on the "core" of each network. The core being defined by the named entities that were mentioned in more than 10% of the documents, and the paths connecting them. One of the network cores is shown in Figure 4.

The human evaluators found that an average of 79.1% of the paths were correct, with a Kappa Coefficient [11] of 0.69 indicating a high level of agreement between evaluators. A breakdown of the scores is provided in Table I.

In order to evaluate the work presented in this paper, we will recreate a portion of this experiment.

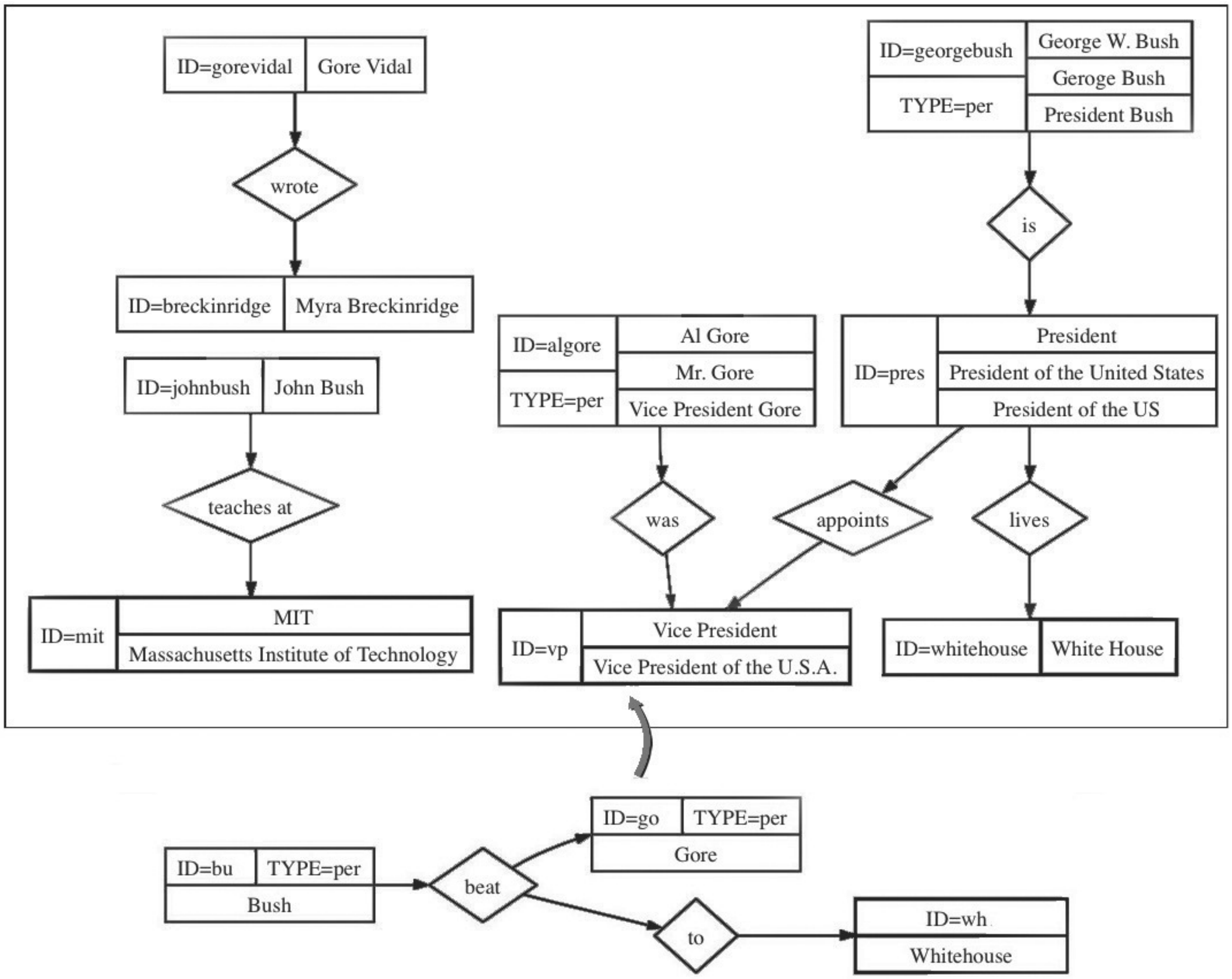

Figure 3. An example ASKNet semantic network fragment being added to a knowledge network relating to U.S. politics

Topic	Eval 1	Eval 2	Eval 3	Avg
Elian Gonzalez	88.2%	70.1%	75.0%	77.6%
Galileo Probe	82.6%	87.0%	91.3%	87.0%
Viruses	68.4%	73.7%	73.7%	71.9%
Vladimir Putin	90.3%	82.8%	94.7%	89.9%
West Bank	68.2%	77.3%	70.0%	72.3%
Average Precision:				79.1%

Table I
EVALUATION RESULTS FOR THE 2008 EXPERIMENT

III. ESTIMATING PARAMETERS

A. Developing a Gold Standard

The first step in developing a method for automated parameter refinement is to create a gold standard evaluation. In order to build such a resource, 742 lines of text were processed from the BBC News Business edition RSS feed (*http://feeds.bbci.co.uk/news/business/rss.xml?edition=int*). A GUI tool (See Figure 5) was created that allowed users to select, for each potential mapping that ASKNet would consider, whether that mapping was correct. 2 evaluators were asked to complete the mappings using the tools, and in the case of disagreements, a third evaluator was asked to break ties. A total of 1306 mappings were produced in under 30 minutes per evaluator, with an inter-rater Kappa Coefficient of 0.989, indicating an extremely high level of agreement.

B. An Evolutionary Hill Climbing Search

For our experiments, we attempted to optimise the settings for 5 parameters.

All parameters were initially set to their default values provided by the system developers. Then ASKNet was run on the BBC data, and the mappings were compared against those in the gold standard. A weighted harmonic mean was used to calculate an F-Score of 0.436. A weighting of $\frac{3}{4}precision$ to $\frac{1}{4}recall$ was chosen to emphasise the

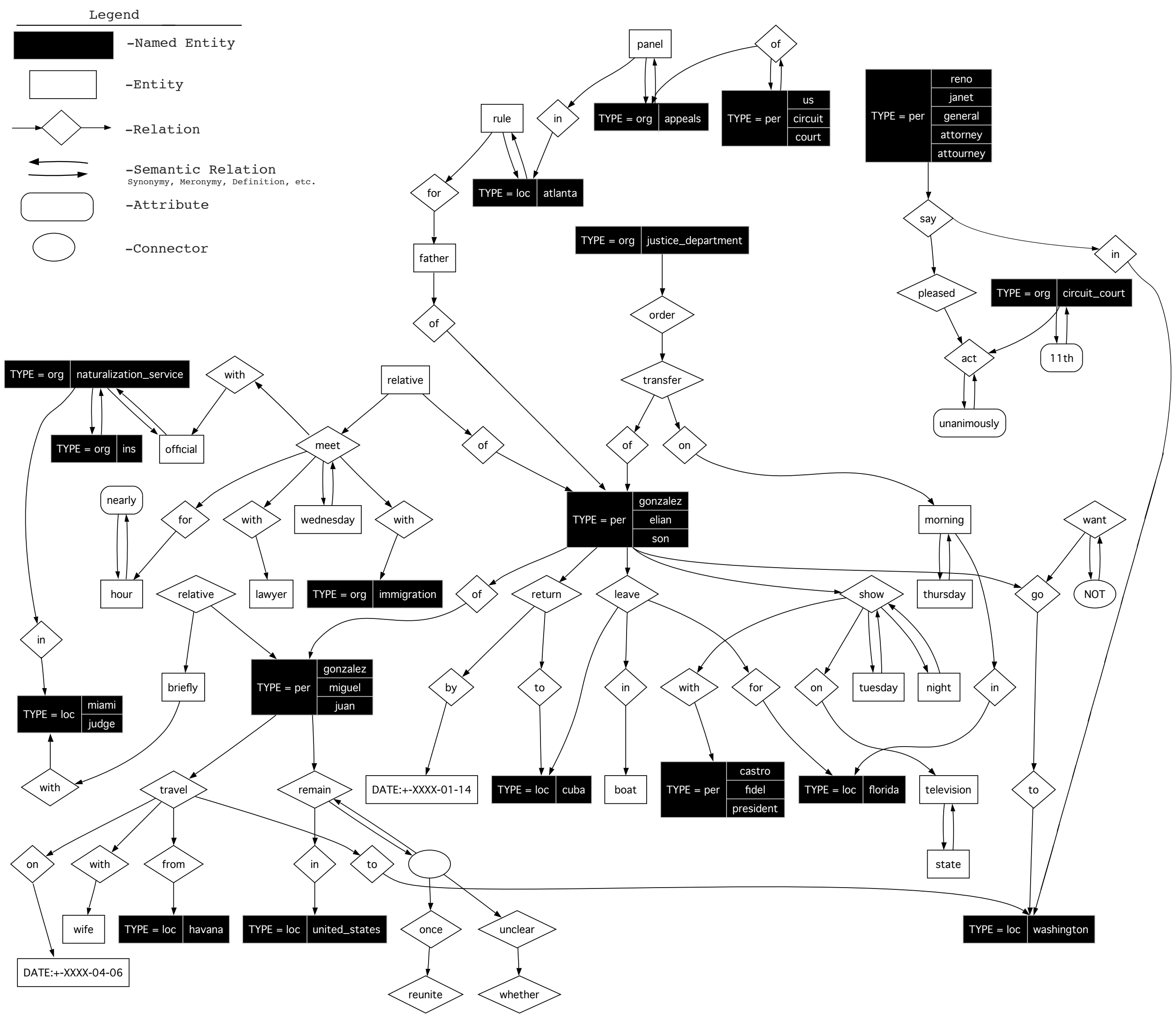

Figure 4. The core of the ASKNet network containing information on the Elian Gonzalez custody trial.

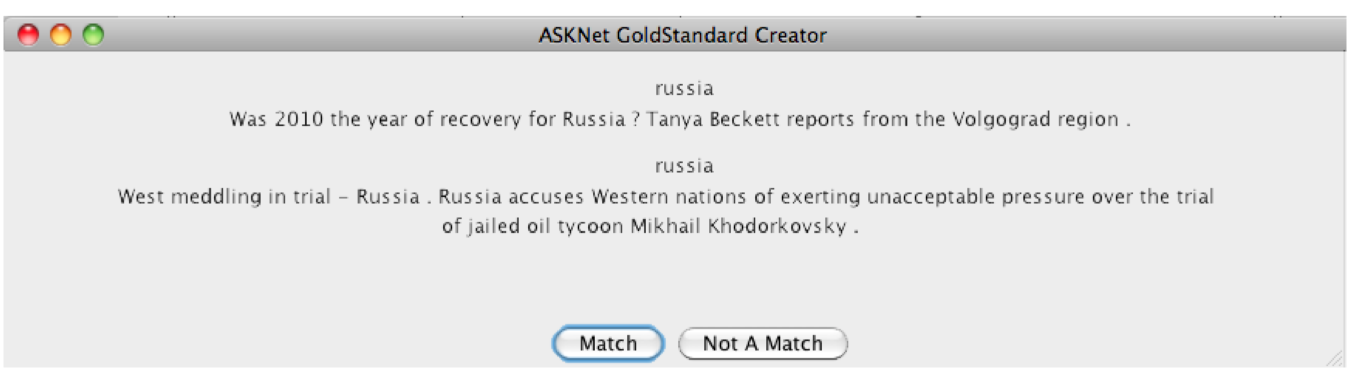

Figure 5. The user interface of the gold standard maker.

mapThreshold	The mapping score level above which we map a node pair together
initialActivation	The amount of activation initially given to the source node
iterations	The number of iterations of the firing algorithm to be run
signalAttenuation	The amount of signal that is lost with each firing (controls the maximum distance that activation can spread from its source)
firingThreshold	The minimum activation required to cause a node to fire

importance of precision in this context.

An evolutionary hill climbing algorithm based on machine learning techniques was then implemented in which parameters were adjusted in turn by small delta values, repeating the gold standard test until the F-Score was optimised for a local maximum. Once a local maximum was reached, a parameter was selected at random to be "mutated" to a random value.

This hill climbing search was allowed to run for 2 hours on a 2.4GHz processor, and eventually resulted in a maximum weighted F-Score of 0.510. The parameters that achieved these values were then saved for use in the experiment.

IV. Evaluation

In order to evaluate the performance improvement that our optimised parameters generated, ASKNet was run on the same data set as was used in the 2008 experiment detailed in Section II-B. However, since we are only concerned with improving the mapping, and have not affected the parsing or semantic analysis, we chose to modify the experiment to focus on the precision of the mappings, as opposed to the overall network.

The experiment was repeated, but with evaluators only being asked to judge whether the mappings were correct. They were asked to evaluate each named entity in the network, and provide a score of "correct" (The entity corresponds to a single real world entity, and all instances of that entity have been correctly mapped onto a single node) or "incorrect" (Two or more real world entities have been mapped to a single node, or one real world entity has been split between multiple network nodes).

The experiment was first run with the original settings, yielding an overall precision of 71.6%. It should be made clear that these results are lower than those presented in Table I, as the new experiment is focusing solely on the mappings, which is the most difficult element of network creation, and thus would have produced a higher proportion of errors than the parsing and semantic analysis phase.

The improved settings were then tried, yielding a result of 79.5%. An improvement of nearly 8%. This means that simply optimising 5 of the parameters across the system removed nearly 8% of the errors made by the system. While this may not seem like a vast improvement at first, in a large scale network such as those built in previous experiments

Topic	Eval1	Eval2	Eval3	Avg
Elian Gonzalez	61.3%	58.0%	64.6%	61.3%
Galileo Probe	78.2%	72.3%	80.1%	76.9%
Viruses	73.5%	68.2%	74.7%	72.1%
Vladimir Putin	81.2%	84.4%	89.0%	84.9%
West Bank	61.2%	62.3%	64.2%	62.6%
Average Precision				71.6%

Table II
EVALUATION RESULTS WITH DEFAULT SETTINGS

Topic	Eval1	Eval2	Eval3	Avg
Elian Gonzalez	70.3%	69.1%	75.2%	71.5%
Galileo Probe	86.4%	78.9%	82.0%	82.4%
Viruses	73.1%	69.3%	72.2%	71.6%
Vladimir Putin	84.4%	88.9%	94.7%	89.3%
West Bank	80.2%	82.1%	85.3%	82.5%
Average Precision				79.5%

Table III
EVALUATION RESULTS WITH OPTIMISED SETTINGS

[3], this could remove tens of thousands of possible errors. In this experiment, the inter-rater Kappa coefficient was 0.72, once again indicating a high level of agreement between all three evaluators, and confirming that the improved score shown is due to an actual improvement in the mappings, and not due to evaluator bias.

V. Conclusion

In this paper, we have developed an automatic system to optimise the parameters of ASKNet using a gold standard annotated by human evaluators, and a hill climbing algorithm with genetic mutations. With the parameters optimised by this system we are able to improve the precision of the system's mapping ability, one of the core functionalities of ASKNet, by almost 8%, as shown by a manual evaluation.

These techniques can be useful both in improving the quality of the networks produced by ASKNet, but also allow researchers to efficiently tune parameters to new data sources and types of information. In order to build an ASKNet network on a new type of information, such as scientific text or narratives, it is only necessary for researchers to develop a new gold standard markup, using the tools already provided, and use the same hill-climbing algorithm to find optimised parameters.

In this experiment, we only chose to optimise the 5 most important parameters in ASKNet. However, these techniques could be used to evaluate more fundamental changes, such as re-designing of the underlying algorithms and data structures. By providing ASKNet developers with a simple, efficient, automated tool to adjust settings, and a gold standard against which to test, developers can efficiently evaluate changes they are making to ASKNet without having to rely on intuition, or undergo the relatively time-consuming task of developing large scale networks.

REFERENCES

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

[2] P.-R. Wojtinnek and S. Pulman, "Semantic relatedness from automatically generated semantic networks," in *Proceedings of the 9th International Conference on Computational Semantics*, Oxford, UK, 2011.

[3] P.-R. Wojtinnek, B. Harrington, S. Rudolph, and S. Pulman, "Conceptual knowledge acquisition using automatically generated largescale semantic networks," in *Proceedings of the 18th International Conference on Conceptual Structures*, Kuching, Sarawak, Malaysia, 2010.

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

[5] B. 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.

[6] B. Harrington and S. Clark, "Asknet: Creating and evaluating large scale integrated semantic networks," *International Journal of Semantic Computing*, vol. 2, no. 3, pp. 343–364, 2009.

[7] S. Clark and J. R. Curran, "Wide-coverage efficient statistical parsing with CCG and log-linear models," *Computational Linguistics*, vol. 33, no. 4, pp. 493–552, 2007.

[8] J. Bos, S. Clark, M. Steedman, J. R. Curran, and J. Hockenmaier, "Wide-coverage semantic representations from a CCG parser," in *Proceedings of the 20th International Conference on Computational Linguistics (COLING-04)*, Geneva, Switzerland, 2004, pp. 1240–1246.

[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*. New York, NY, USA: ACM Press, 1988, pp. 147 – 160.

[10] H. T. Dang, "Overview of duc 2006," in *In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL)*, New York NY, USA, 2006.

[11] J. Carletta, "Assessing agreement on classification tasks: the Kappa statistic," *Computational Linguistics*, vol. 22, no. 2, pp. 249–254, 1996.