THE INTERPRETATION ON NOUN COMPOUNDS USING PARAPHRASING VERBS AND PREPOSITIONS

Vaibhav, Niraj Kumar
Advisor: Dr. Amitabha Mukerjee
{vaibhavv,nirajkr,amit}@cse.iitk.ac.in
Department of Computer Science and Engineering,
IIT Kanpur, India
March 10, 2013

Proposal:

Our aim is to provide a list of verb phrases that can be used to paraphrase a given noun compound.
Example [1]: malaria mosquito is a compound noun which can be suitably paraphrased as mosquito that carries malaria, or that causes malaria, etc.

Motivation:

The foremost reason for doing such a task is that noun compounds are a major part of any language and today when everything is digital, there must be a smart system which can interpret the semantic meaning of any text correctly, and for doing the semantic analysis of any text, correct interpretation of compounds becomes very important. This task is not so trivial for a machine because the meaning of the compound cannot be directly obtained from the nouns, the system should have some way of interpreting the hidden relations that hold between the nouns. Because noun compounds are so frequent in written text, systems that deal with semantic analysis of text cannot ignore those. And because the meaning of the compound cannot be directly obtained from the nouns, the system should have some kind of way of interpreting it.

Related Works:

Many past works have been done to solve this problem. Broadly there are two strategies to tackle this problem one is top-down and other is bottom-up.
In the top-down strategy, the problem of noun compound interpretation is basically converted into a classification problem. Girju et al. (2005) suggested 21 classes of abstract relations. [2]

The second broad strategy to interpret noun compounds is the bottom-up strategy in which noun compounds are being interpreted through paraphrasing those using suitable verb phrases.

Approach:

For a given noun compound, firstly the verb phrases are extracted from the British National Corpus [B], which is a large text corpus comprised of English written text. Given a noun compound, if we can find an occurrence of one noun with a verb phrase such that it is the verb's object, and an occurrence of the other noun with the same verb phrase such that it is the verb's subject, then, the verb phrase might be suitable for paraphrasing the compound noun. If the individual occurrences of the nouns are hard to find, we can use any of its top two hypernyms(broad meaning constituting a category) found on WordNet [A]. The obtained paraphrases are then validated to produce a ranked list and to eliminate less likely ones. Validation is done by web-based validation followed by mutual information and a final web-based evaluation.[3]

Datasets:

- Fellbaum, C. (1998). *WordNet: An electronic lexical database*, MIT press Cambridge, MA.[A]
- *The British National Corpus*, version 3 (BNC XML Edition). (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/[B]

References:

- Butnariu, C., Kim, S., Nakov, P., Ó Séaghdha, D., Szpakowicz, S. & Veale, T. (2009). SemEval-2010 Task 9: The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions. In *Proceedings of the SemEval-2010 Workshop*.[1]
- Butnariu, C. & Veale, T. (2008). A concept-centered approach to noun-compound interpretation. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-08)*, 81-88.[2]
- Lilit Darbinyan supervised by Stephen Pulman (2010). Interpretation of noun compounds, Oxford University.[3]