[Inspiratron.org - Natural language processing, machine learning and cybersecurity](#)

Marvin - A tool for semantic annotation released

by Nikola Miloševi? - Monday, February 08, 2016

https://inspiratron.org/blog/2016/02/08/marvin-a-tool-for-semantic-annotation-released/

During the last week I have released a version of Marvin - a tool for semantic annotations, that is able to annotate text using various sources, such as UMLS (using MetaMap), DBPedia, using some SPARQL interface, WordNet and probably most importantly SKOS (Simple Knowledge Organization System) format for representing lexicons, dictionaries and terminologies. Primarily, the tool is supposed to be helpful in data labeling and normalization of biomedical texts, however, with the help of SKOS, WordNet and DBPedia it can be helpful in any domain.

When I mentioned normalization and labeling, for some readers not familiar with text mining and some aspects of semantic web, I better briefly explain. Basically, usual natural language text that comes in certain articles is not normalized. Concepts may come in many different forms, both syntactic (different word suffixes, prefixes, syntactic forms of the word) and lexicological (for example some authors would say cancer, while other will call same disease tumor). In order to semantically process data and for example extract the information of interest easier, text should be normalized. This means that all forms of the words that have same meaning should be the same. So all synonyms should be normalized to the one specific preferred term.

However, it is not probably

obvious how this works with Marvin. Basically Marvin can perform normalization using SKOS and WordNet, since both WordNet and dictionaries in SKOS may contain a list of synonyms, word inflections and preferred term. However, DBPedia and UMLS have different purpose. They tag the text with broader concept identifiers, their descriptions and relationships with other concepts. With this tags it is possible to perform more advanced queries over the text or extracted data, since these tags add semantics into the text. Queries such as finding all mentions of adverse events or diseases in article could be performed even if adverse event or disease as word is nowhere in article mentioned.

The advantage of Marvin is that it can be used as a library with other mining tools. Marvin is written in Java, so any other Java tool may use Marvin for annotating text. It also contains data about provenance. With SKOS, it also labels data with all higher hierarchy concept names, not only one.

Marvin - A tool for semantic annotation released - 02-08-2016
by Nikola Miloševi? - https://inspiratron.org

The documentation about how to run and use Marvin can be found on the GitHub page of the project:

https://github.com/nikolamilosevic86/Marvin

Source code of the project can be found here as well. I also published a paper on ArXiv where I included much more details on how Marvin works and how to set it up. The paper can be found on the following address:

http://arxiv.org/abs/1602.00515

or PDF at

http://arxiv.org/ftp/arxiv/papers/1602/1602.00515.pdf

Release of the tool can be found here:

https://github.com/nikolamilosevic86/Marvin/releases/tag/v1.0

_____