

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

## Marvin - semantic annotator

by Nikola Milošević - Thursday, June 25, 2015

<https://inspiratron.org/blog/2015/06/25/marvin-semantic-annotator/>

We created Marvin, an semantic text annotation tool that uses external resources such as DBPedia and WordNet to annotate semantically text. Marvin is built in java and can be used as a standalone application or as a library.

Marvin semantic annotator already has a lot of knowledge, which will make anyone probably depressed and therefore we gave him a name with reference to "Hitchhikers guide to the Galaxy" depressed robot.

Basically, Marvin queries the sources he can for the definitions about words in inputted text. He uses some NLP transformations in order to obtain as much knowledge as possible. For DBPedia he uses unigram and bigrams, while wordnet queries only with single words.

Instruction how to download, install and setup Marvin can be found here: <http://nikolamilosevic86.github.io/Marvin/>

Probably next release will include some new data sources such as MetaMap for biomedical tagging, possibly SNOMED, ICD11 and some ontologies. Also, improvements will be made on querying current data sources.

Since the project is developed as part of my PhD and my PhD is on biomedical text mining, there is a huge possibility that apart from DBPedia and WordNet, Marvin will focus mainly on biomedical data. However, we will tend to make it as simple as possible, and that's why for [UMLS](#) we will be using [MetaMap](#). However, we will try to make it as independent as possible from installing all resources. Currently user needs to install WordNet on his machine.

For the end here is compilation of Marvin's quotes. Hope you'll enjoy it:

<https://www.youtube.com/watch?v=YKd6ZxCEJdI>

---

All rights reserved and copyrighted by [inspiratron.org](https://inspiratron.org) and Nikola Milosevic