

Project description

Through the internet, knowledge is now more accessible than it has never been. However, search engines as we know them are not performing any semantic analysis of the input keywords. They only answer with the best documents matching them without analyzing their meaning and how they are related to each other.

Current solutions for semantic analysis all show some limitations. For instance, the traditional approach, based on First Order Logic (FOL), provides accuracy and ability to deal with complex proofs, but the problem of designing a reliable and robust automated translation process of natural language to FOL remains.

Natural Logics overcomes this problem by relying on stylized fragments of natural language where reasoning can be conducted directly by reflecting intuitive reasoning in natural language. In this way, Natural Logics aims to reconcile the robustness and accuracy of theorem prover based on extended syllogistic proof systems while using a language representation whose syntax is closer to natural language, and then, whose translation from and to natural language is easier to process.

The subject of this master thesis is the discussion of Natural Logics as a solution to perform semantic search in scientific texts and its application in a Prolog prototype.

This project starts by theoretically discussing Natural Logics. Specific attention is thus given to ontology based structure knowledge bases to define Natural Logics' concept representations and relations and to traditional syllogistic logic to define inference rules working over them. The approach first focuses on defining a core Natural Logics before extending it in order to successfully handle more and more complex sentences. Graph representation and graph generation with concepts as nodes and their relations as edges will be particularly discussed and evaluated since graph querying will be the main objective of our Prolog prototype. Finally, its results will be confronted with case studies related to bio-science in order to analyse and assess its performances and to discuss the strengths and weaknesses of both our natural language representation and implementation.

