THESIS PROPOSAL – 30 CREDITS

Patent Document Summariser

Background

Patent documents are very difficult to understand by people who are not well acquainted with reading and analysing such documents:

- The document is normally written by patent attorneys and the terminology used contain some legal jargons that can be confusing to a normal reader (even if the reader is well acquainted with the technology).

- Sometimes technical terms used are not the same as the ones normally used in technical papers. This can be done intentionally to make it difficult for a non-patent expert to exactly understand what is claimed in the patent.

- The abstract of the patent is not always descriptive of the exact nature of the protection claimed in the patent document or intentionally missing some important details.

On the other hand, patent documents are normally classified into very specific areas and in most cases are very well structured and divided into specific sections, such as background (that describe prior art), detailed description of the invention, specific features that are protected (Claims) and abstract.

Project description

The main goal of this project is to generate one to two paragraph summary text on the essence of the patent and the protected features in an easy to read natural language text.

Customer

IamIP Sverige AB, Stockholm

Contact at NLPLAB

Marco Kuhlmann, marco.kuhlmann@liu.se

Student profile

Knowledge about natural language processing (via courses such as TDDE09) and/or text mining (via courses such as TDDE16, 732A92).