FIT-IT SEMANTIC SYSTEMS
WWW.IDIOM.AT

idiom

UNDERGRADUATE THESIS (DIPLOMARBEIT)

SEQUENTIAL PROBABILITY MODELS FOR RELATIONSHIP EXTRACTION

We are looking for a student with sound programming skills, preferably in Python and/or Java, and a background in statistics interested in writing an undergraduate thesis in the area of natural language processing and text mining.

To be supervised either at Vienna University of Economics and Business Administration or Graz University of Technology, the thesis would be part of a larger project aimed at analyzing the production and consumption of electronic content (IDIOM = Information Diffusion across Interactive Online Media; www.idiom.at). For generating and structuring its underlying knowledge base, the project requires algorithms to automatically identify and tag relations between information entities.

The aims of this thesis are:

- Identification of weaknesses in the current approach based on trigger phrases,
- Improvement of the existing approach by applying a Sequential Probability Model,
- Evaluation of the performance of the refined architecture.

The developed methodology will help to improve crucial parts of IDIOM's core architecture. Thus, final results will be needed around August 2007. An excellent thesis completed within this timeframe will be rewarded with a bonus between 300 and 1,000 EUR.

FURTHER INFORMATION ON THIS PROJECT

Dr. Albert Weichselbraun
Wirtschaftsuniversität Wien
Institut für Informationswirtschaft
Augasse 2-6, A-1090 Wien
albert.weichselbraun@wu-wien.ac.at
Tel: 43-1-31336-5229