Automatic construction of opinion lexicon for Slovene

Opinion mining (or in a more general sense sentiment analysis) refers to the use of natural language processing tools to identify and extract subjective opinions on a certain topic in given collection of textual documents. One of the task in sentiment analysis is polarity detection, where the given text is classified as positive, negative, or neutral. A possible use of this polarity information is to evaluate product reviews, market trends, movie/song/book recommendations, etc. One of the approaches to the polarity detection is to use sentiment lexicons [1].

WordNet is a lexical database for the English language, which contains short definitions of words, their synonyms and usage examples. It is freely available with an intention to aid automatic text analysis and development of artificial intelligence applications. A Slovene version of WordNet is called SloWNet and is linked to the English WordNet.

Unfortunately, for Slovene language no opinion lexicon exists, but we wish to construct one with the help of English lexicon, WordNet and SloWNet.

The goal of the assignment is to translate the Hu & Liu opinion lexicon to Slovene using WordNet and SlowNet. The translated lexicon shall be tested on a small collection of Slovene product reviews.

Supervisor: prof. dr. Marko Robnik Šikonja, marko.robnik@fri.uni-lj.si,

Contact hours: see http://www.fri.uni-lj.si/rmarko/

Literature:

[1] Minqing Hu and Bing Liu. "Mining and summarizing customer reviews." *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD-2004,* 2004.

[2] Minqing Hu and Bing Liu. "Opinion Mining, Sentiment Analysis, and Opinion Spam Detection". List of resources. http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

[3] Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. " O'Reilly Media, Inc.".