Vrije Universiteit Brussel

Proposal bachelor thesis

Title: Parallel “opinion mining” of Twitter
Promotor: Wolfgang De Meuter
Contact: Janwillem Swalens (10F719)
Includes preparation course: No

Context

Twitter contains a vast amount of useful information. “Opinion mining” or “sentiment analysis” refers to the analysis of these data sets to extract opinions, sentiments or emotions. We could for example classify a tweet as positive if it contains words such as “happy” or “excellent”, or negative if it contains “sucks” or “:(“.

Opinion mining can be used to automatically generate reviews of products. We might for example calculate the ratio of positive to negative tweets containing “iPhone” and “battery life”, and compare those to the positive/negative ratio of tweets containing “Galaxy S5” and “battery life”.

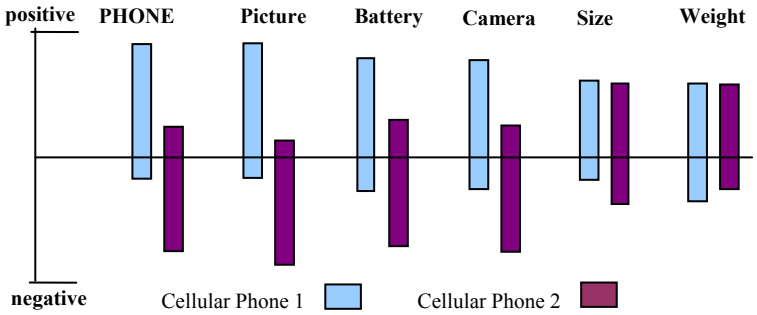

Figure 1: comparison of two cell phones (source: Handbook of Natural Language Processing, Second Edition, N. Indurkhya and F. J. Damerau, 2010)

Because the analysis of such huge amounts of data takes a lot of processing power, you will need to parallelize your program to make use of the multiple cores in your machine.

Proposal bachelor thesis

The aim of this bachelor project is to build a parallel “opinion miner” using the Rust programming language (http://www.rust-lang.org), and validate its scalability.

The main focus of this project will be the parallelization techniques: how do you efficiently structure your program such that the speed-up is maximal? You will need to create and compare a number of parallel implementations of the algorithm.

For this comparison, you will create a set of benchmarks. You can run these on a 64-core server (with 64GB RAM) available to you at the SOFT lab. Your final

report should describe your implementations, followed by a report and interpretation of the benchmark results in a statistically sound way.

Preparatory course bachelor thesis

N/A