Web Search Engines — Project Proposal

Himaja Rachakonda, Anirudhan J. Rajagopalan

N14633788, N18824115

hr970, ajr619

March 21, 2016

1   Title

The project is to be titled "Review-Rehashed" will summarize the reviews of a product with respect to a particular feature.

2   Team Members

1. Himaja Rachakonda, N14633788, hr970@nyu.edu

2. Anirudhan J. Rajagopalan, N18824115, ajr619@nyu.edu

3   Objective

The aim of this project is to extract sentences from online reviews which discuss about the features of a product and present it to the user. For example for a search term, say, "LG G3, Battery Life" will list the excerpts from all reviews that discuss about the battery life of LG G3 phone. Depending on the computational complexity and performance of the search engines, we will try to expand the search feature to multiple features and combination of features.

4   Sketch of Architecture

5   List of web resources

We are planning to scrape data from

1. Amazon.com

2. Bestbuy.com

in very narrow categories of products (say electronics, or books).

6   Technologies used

We are planning to use the following resources for building the search engine.

Programming Language : Java, J2EE

Web server : Apache Tomcat

Index Store : Apache Lucene

NLP library : Apache Mahout