Part 1: Steps followed to complete this assignment

Step 0: Preparation

Step 0.1: Install VirtualBox and Ubuntu 16.04.2 LTS

Step 0.2: Install Apache Solr 6.5.0

Step 0.3: Download assigned HTML files (NYTimesData) and Index HTML files in Solr

Step 1: Build a client application for Solr

Step 1.1: Create a webpage using PHP with solr-php-client library and Apache web server

Step 2: Compute PageRank scores and add scores into Solr

Step 2.1: Create edgelist.txt using Java and JSoup library

Step 2.2: Create external_pageRankFile.txt using python and NetworkX library

(pagerank algorithm initial parameters: alpha=0.85, personalization=None, max_iter=30,

tol=1e-06, nstart=None, weight='weight',dangling=None)

Step 2.3: Place external_pageRankFile.txt into Solr data folder and setup managed-schema.xml
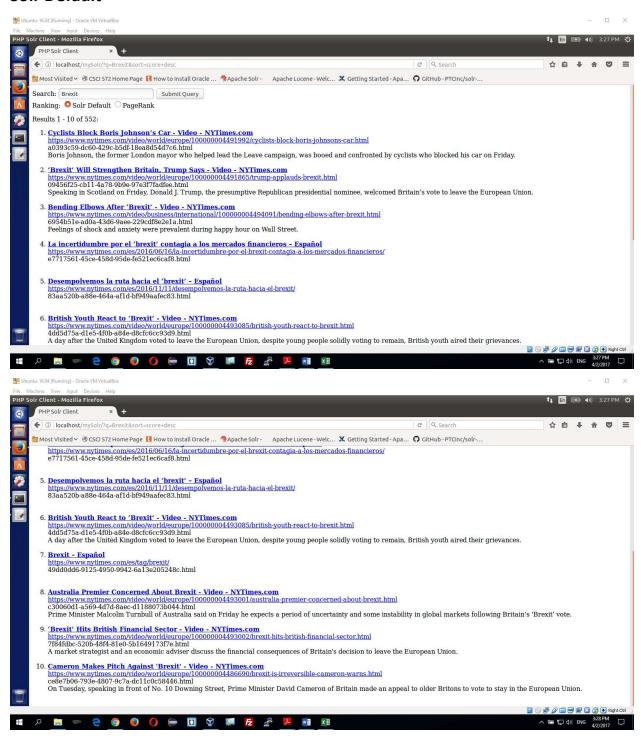
and solrconfig.xml

Step 3: Compare the two ranking algorithms

Step 3.1: Compare the results of the two ranking algorithms with the following eight queries

(Brexit, NASDAQ, NBA, Snapchat, Illegal Immigration, Donald Trump, Russia, NASA)

Part 2: Query results of the two ranking algorithms

Query 1: Brexit

Solr Default

PageRank