



UNIVERSITY OF CAPE TOWN

INFORMATION RETRIEVAL

CSC4000W

Assignment 1b Report

Author:

Alon Bresler
Andreas von Holy
Osher Shuman

Student Number:

BRSALO001
VHLAND002
SHMOSH001

May 18, 2016

1 System Design

Included:

- stemming (porter)
- uses tf-idf

Own:

- stopword exclusion (can still add more stop words if needed. can do when testing system as a whole.)
- title similarity (modify ranking formula once full system developed.)
- Thesaurus \Rightarrow Natural language toolkit (nltk) with WordNet package to get synonyms for search terms. Synonyms stemmed using porter if stemming parameter is set true.

2 Performance Results

2.1 Mean Average Precision (MAP)

2.2 Normalized Discounted Cumulative Gain (NDCG)