

CmpE493 Assignment 3

Eşref Özdemir
2013400144

May 3, 2018

Contents

1	Data Preprocessing	2
2	LexRank	2
3	ROUGE Scores	2

1 Data Preprocessing

We perform the same data preprocessing steps as done in assignment 1 and 2. That is, sentences of each document is tokenized and normalized separately. Normalization process includes case folding, punctuation removal and stemming. We compute tf-idf vectors based on this new document representation.

2 LexRank

After obtaining normalized documents, we use the LexRank method to get the most important 3 sentences. We set the teleportation rate (damping factor) of the underlying Markov Chain to 0.15.

3 ROUGE Scores

ROUGE scores are calculated using the library given in the assignment description. Python script to compute ROUGE scores are given in the scoring directory.

Table 1: ROUGE Scores

Metric	F1-Score	Precision	Recall
Rouge-1	0.301932	0.235734	0.45216
Rouge-2	0.104655	0.078466	0.173119
Rouge-L	0.208407	0.191422	0.402338