

Using a Keynes Metric for Single and Multi Document Summarisation

Mahmoud El-Haj

m.el-haj@lancaster.ac.uk

Paul Rayson

p.rayson@lancaster.ac.uk

School of Computing and Communications

Lancaster University

UK

Multiling 2013 Participation



We participated with single-document and multi-document corpus-based summarisers for both Arabic and English languages.

Coordinated the creation of the Arabic version of Multiling 2013 dataset.

Our Summarisers



The summarisers used word frequency lists and log likelihood calculations to generate single and multi document Arabic and English summaries.

Corpus-based Summarisation

-
- Comparing a sample corpus with a larger standard corpus (Scott, 2000).
 - The first word list is the frequency list of all words in the document(s)
 - The second list is a much larger standard corpus.

Why Word Frequency?

- 1) Words with high frequency in the input documents are very likely to appear in the human summaries.
 - 2) The automatic summarisers include less of these high frequency words.
- Overlap can be improved by including more of the high frequency words. Nenkova et al. (2005), (Li et al., 2006).

Standard Word Frequency Lists

- Most frequent 5,000 words for both Arabic and English using:
 1. The frequency dictionary of Arabic (Buckwalter and Parkinson, 2011).
 2. The Corpus of Contemporary American English (COCA) top 5,000 words (Davies, 2010).

Summarisation Methodology

- For each word in the dataset (Arabic and English) we calculated the log likelihood scores using the word frequency lists.
- We summed up the log likelihood scores for each sentence
- We picked sentences (≤ 250 words) with the highest sum of log likelihood scores.

Results Single-document (Auto)



System	AutoSumm	MeMoG	NPowER
ID2 (Lancs)	0.136	0.136	1.685
ID41	0.129	0.129	1.661
ID42	0.127	0.127	1.656
ID3	0.127	0.127	1.654
ID1	0.124	0.124	1.647
ID4	0.123	0.123	1.641
ID5	0.040	0.040	1.367

English Automatic Evaluation Scores (single-document)

System	AutoSumm	MeMoG	NPowER
ID3	0.092	0.092	1.538
ID2 (Lancs)	0.087	0.087	1.524
ID41	0.055	0.055	1.418
ID42	0.055	0.055	1.416
ID4	0.053	0.053	1.411
ID5	0.025	0.025	1.317

Arabic Automatic Evaluation Scores (single-document)

Results Multi-document (Manual)



System	Score
ID4	3.547
ID11	3.013
ID6	2.776
ID21	2.639
ID51	2.571
ID61	2.388
ID5	2.245
ID1	2.244
ID3	2.208
ID2	1.893

English Manual Evaluation Scores (multi- document)

System	Score
ID6	3.711
ID2	3.578
ID3	3.578
ID4	3.489
ID1	3.467
ID11	3.333
ID21	3.111
ID51	2.778
ID5	2.711
ID61	2.489

Arabic Manual Evaluation Scores (multi- document)

Results Multi-document (Auto)

Group	SysID	Avg Perf
a	ID4	0.2220
a	ID11	0.2129
a	ID61	0.2103
ab	ID1	0.2085
ab	ID21	0.1903
ab	ID6	0.1798
ab	ID2	0.1751
ab	ID5	0.1728
b	ID3	0.1590
b	ID51	0.1588

English: Tukey's HSD test MeMoG groups

Group	SysID	Avg Perf
a	ID61	0.2488
ab	ID4	0.2235
abc	ID1	0.2190
abc	ID11	0.2054
abc	ID21	0.1875
abc	ID2	0.1587
abc	ID5	0.1520
bc	ID51	0.1450
bc	ID6	0.1376
c	ID3	0.1230

Arabic: Tukey's HSD test MeMoG groups

Enhance Multi-document Scores



- First, we treated the set of related documents (multi-documents) as a single big document.
- This can be solved by running the summariser on each then selecting the top sentence(s) of each document.
- Second, we did not work on eliminating redundancies.

Thanks



-
- Questions?

References



-
- W. Li, B. Li, and M. Wu. 2006. Query Focus Guided Sentence Selection Strategy.
 - A. Nenkova and L. Vanderwende. 2005. The impact of frequency on summarization. Microsoft Research, Redmond, Washington, Tech. Rep. MSR-TR-2005-101 .
 - T. Buckwalter and D. Parkinson. 2011. A Frequency Dictionary of Arabic: Core Vocabulary for Learners. Routledge, London, United Kingdom.
 - M. Davies. 2010. The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English. *Literary and Linguistic Computing* , 25:447–464.

Calculating Log-Likelihood

	Corpus One	Corpus Two	Total
Frequency of Word	a	b	a+b
Frequency of other words	c-a	d-b	c+d-a-b
Total	c	d	c+d

The values **c** and **d** correspond to the number of words in corpus one and corpus two respectively. Where **a** and **b** are the observed values (**O**). For each corpus we calculated the expected value E using the following formula:

$$E_i = \frac{N_i \sum_j O_j}{\sum_j N_j}$$

N_i is the total frequency in corpus **i** (**i** in our case takes the values 1 (**c**) and 2 (**d**) for the Multi-Ling Arabic summaries dataset and the frequency dictionary of Arabic (or MultiLing English Summaries dataset and COCA corpus) respectively. The log-likelihood can be calculated as follows:

$$E1 = c*(a+b) / (c+d) \text{ and } E2 = d*(a+b) / (c+d)$$