LING 573 Project Report

Harita Kannan haritak5@uw.edu
Ben McCready bmccr@uw.edu
Lonny Strunk lstrunk@uw.edu
Xiaopei Wu xw126@uw.edu

Abstract

The following report details the architecture of the extractive multi-document summarization system built for Deliverable 2 of the Spring 2018 offering of LING 573 at the University of Washington. This is a baseline system intended to be improved upon in the coming weeks in the course.

1 Introduction

The task for this assignment is to create a system that

2 System Overview

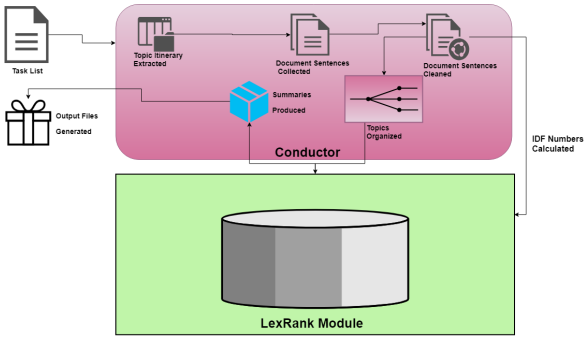

Figure 1: System Architecture

Our system has two major components; the 'conductor', which handles all file-getting and topic-ordering, and a separate LexRank object which does the work of generating summaries from sets of sentences.
The conductor takes the topic list XML file as input: from there, it arranges the XML into a workable itinerary of Topic objects, each of which contains a set of documents associated with it.
The sentences are cleaned up a little to make the output more read-able, and then all of the sentences are fed into the LexRank object for a given trial to create the IDF numbers used to calculate the salience of given sentences.
Once the sentences have been arranged into topics, each set of sentences from a given topic is fed to the LexRank object; a list of summary sentences are returned, which can then be written to output files.

3 Approach

For this baseline system, we implemented an openly available LexRank implementation - this was determined to be the most straightforward way to create as basic a system as possible, so that our team may be able to test out various improvements for the coming deliverables.

3.1 Content Selection

Content for our summarization was selected running the LexRank algorithm, using a degree centrality threshold of 0.1.

3.2 Information Ordering

In this system, information is ordered according to the LexRank score; that is, of all sentences in a given group of documents, the sentence with the highest LexRank score is appended to our summary, until our summary cannot accept the next highest sentence without exceeding the required word length.

3.3 Content Realization

Being that this system produces an extractive summary, all content included in the system output are sentences chosen verbatim from documents within the document cluster for which the summary is being produced.

4 Results

Table 1: ROUGE Evaluation Scores

	PRECISION	RECALL	F1
ROUGE-1	0.22034	0.20016	0.20885
ROUGE-2	0.04889	0.04549	0.04697
ROUGE-3	0.01648	0.01569	0.01603
ROUGE-4	0.00707	0.00679	0.00691

5 Discussion

The primary goal in creating the system under discussion was to implement a baseline text summarization pipeline that can be improved upon; in that sense, this task was successful.
Our ROUGE scores do not compare favorably compared to those of conference-level systems; however, most of the 'summaries' output by our system do a fair job of communicating the 'gist' of the document cluster provided. The summaries resemble the recollections of someone who was half-listening to a news report more closely than they do someone whose paid task it is to create a comprehensive abstractive text summary.

5.1 Error Analysis

Some of our output sentences still contained errors, some more easily catchable than others; abbreviations like "i.e." triggered the creation of a sentence divide on account of the period followed by a space. This could be rectified with a list of common English abbreviations. Some sentences were not broken up at each period that they should have been, requiring some closer debugging.
Other output sentences still contained the header from the newswire piece, such as "LITTLETON, Colo. (AP) –"; these could also be removed with a well-crafted RegEx.
More difficult to cope with are instances in which words were combined (peopleknowledgable, criminalcharges) - this is a more difficult problem, requiring either some pre-treatment of the data if the errors are included in the newswire themselves, or some correction to how sentences are being imported into our system.
The readability of the summaries could be improved by removing certain transition words (although, however) that no longer reliably serve the purpose they were intended to. A more intelligent approach to information ordering, perhaps based on publication date and intra-document sentence ordering, might also help improve the human-judged flow of the generated summaries.

6 Conclusion

Something will conclude here...

References

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

Shostenko, Luka. lexrank (2018) GitHub repository, *https://github.com/wikibusiness/lexrank*