

Extended Project Report (D2)  
For github this goes in `.../doc/D2.pdf`

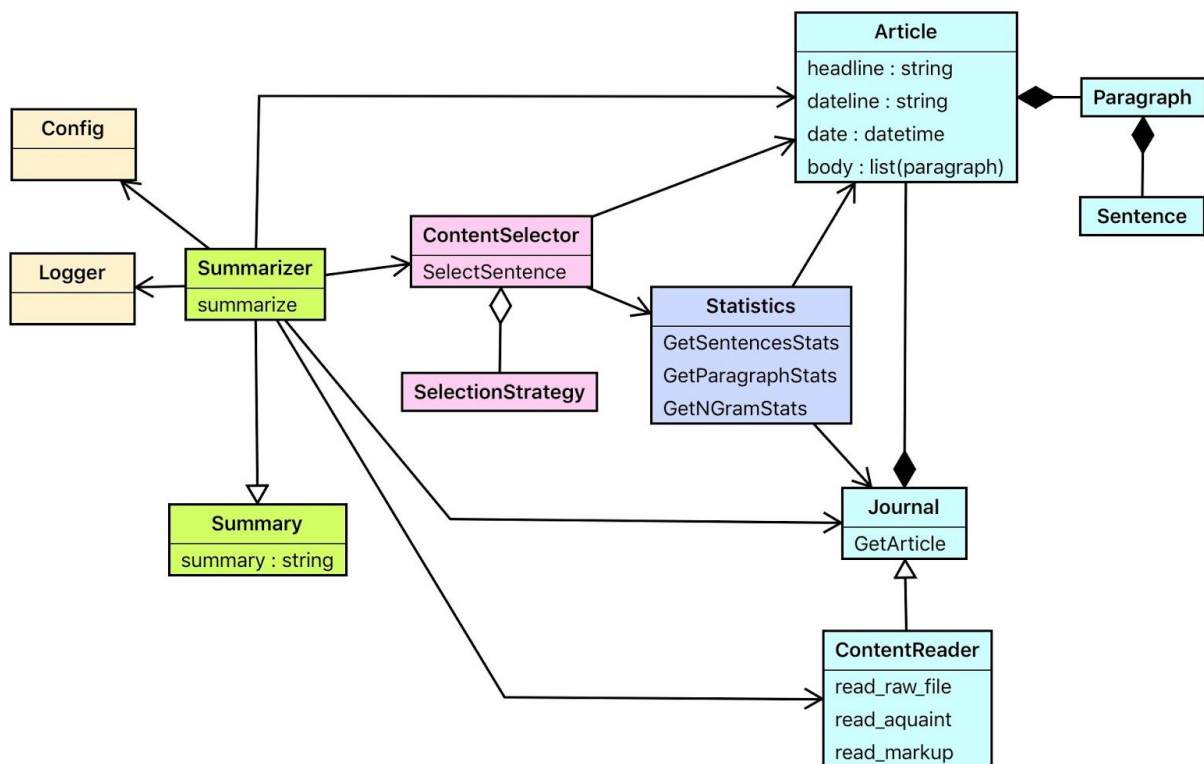
This extended version should include content for all the sections of the report (no more lorem ipsums), though some of it will not be very detailed yet. You should specifically focus on the following:

## System overview

This is the D2 system for Ling573 which establishes an **end-to-end baseline summarization** system. We read groups of stories and write to the outputs directory as specified by the D2 requirements. Then we generate Rouge scores (1-4) against provided gold standard summaries and discuss our baseline findings.

Operational Definitions: The requirements specified well-formed English sentences, we will leave this to the level of syntax and punctuation. Semantic composition is outside of our scope.

## System architecture



# Approach

The system is capable of single- and multi-document summarization and can handle batch processing.

For each document group in a batch, we read the data from the specified source, as detailed in a configuration file; select sentences to extract for the summary; compile the sentences into the realized summary; and write summary to output files.

Then scoring is applied using the reference implementation for this project.

## Content Selection

Our current system features two methods for summarization. The first is a first sentence summarization and the second is a pivoted QR matrix decomposition.

First sentence summarization means that the first sentence of each article in a docset is selected. Given the nature of news articles, this approach produces a fair summary with very little processing.

The pivoted QR matrix decomposition approach (from Conroy and O'Leary), uses feature vectors, based on the sentences in a docset. Said feature vectors comprise a matrix, where each row represents a sentence and each column represents a feature. In the Conroy paper, these features are meant to capture ideas, which are represented by the words in each sentence.

The sum of each feature vector is weighted, using a weighting function from the paper. This weighted sum is the score for each sentence. The sentence with the highest score is selected. Then, the features from that sentence are removed from the matrix. This reduces redundancy in future sentences that will be selected.

Once all features from the selected sentence have been removed from the remaining feature vectors, the sentence scores are recalculated and this process is repeated until no more sentences will fit in the 100-word summary.

If the best scoring sentence at any recursion of the matrix is too long to add to the summary, the next best scoring sentence that is not too long is used.

## Information Ordering

For our baseline first sentence summarization method sentences are simply used in the same order that each article is processed.

In our QR matrix method, sentences that appear earlier in their article appear earlier in our summary. This means that a selected sentence that has a high score may appear lower in the summary while a lower score may appear higher. The goal of this feature is to allow descriptive sentences near the beginning of a news article to appear before sentences that explain or expound later in their own articles.

## Content Realization

To generate the summary, we look at each selected sentence sequentially. If the sum of the words in the summary and the words in the current selected sentence is less than 100, the sentence is added to the summary. This realization strategy isn't particularly clever, but it does ensure that the summary adheres to the length requirement and contains only complete sentences. We chose to exclude sentences of less than six words.

Each summary is written to a summary file formatted as described in the Delivery 2 specifications.

## Results

Base results

*D2 Req: this subsection should describe the results of your base system. You should include a table (like the following) presenting the ROUGE-1, ROUGE-2, ROUGE-3, and ROUGE-4 scores of your system. You should present ROUGE recall values: the "R" column in the ROUGE evaluation output.*

	PRECISION	RECALL	F1
ROUGE-1			
ROUGE-2			
ROUGE-3			
ROUGE-4			

# Discussion

The baseline works surprisingly well.

*D2 Req: You should also present some error analysis help to motivate future improvements.*

Our system is not particularly clever.

- Selection is more or less as shallow as a Seattle rain puddle
- Ordering is non-existent
- Realization includes minimal formatting (removing breaking characters “\n” and bylines)
- Nltk sentence parser is not perfect. Some changes need to be made to parser to avoid sentence parsing on words like “Mt.”