Extractive summarization

Dataset Preparation:

* Extract all the given input files data
* Separate the Passages and Summaries from the extracted data using delimiter
* Preprocess the dataset by removing symbols, non alphanumerics from the text
* Remove the stopwords from the text

Model Making

* Merge the sentences to create the corpus to train Word2Vec Model.
* Make adustments of iter, min\_count, window and vector size to train Word2Vec Model.
* Save the Model

Inference

* Load the saved Model
* Receive the input passage
* Preprocess the input and split the sentences.
* Fetch the vectors of words using model
* Make a similarity matrix with zeros of required size.
* Calculate the cosine distance across all the sentences with each other and replace the similarity matrix values.
* Fetch the scores of each sentence using Pagerank algorithm and sort it.
* Select the required top\_n important sentences using scores.
* Join all the fetched top\_n sentences to create a summary.