

Guillaume Le Garrec
guillaume.lgarrec@etu.univ-paris-diderot.fr

Claude Cunha
dacostacunhaclaude@gmail.com

This project was supported by
IdEx Université de Paris
ANR-18-IDEX-000101



The Text Summarizer

Abstract

Even today, narrative summarization remains a challenge. Its different issues have been presented and discussed in a great number of papers (eg Mani 2001) problematic aspects have been found to be constructing the timeline of a story along with keeping track of the many different plots and subplots happening at the same time. Previous works described representations for time relations in narrative, but none of those presented a complete procedure to do so. This is mainly due to the complexity of processing implicit information. More generally, the problem lies in trying to create a comprehensive way to model what happens in a narrative text. With those studies in mind, we tried to use another approach to simplify the summarizing process. Instead of creating an algorithm to build a model and derive a summary from it, we tried to create a set of grammatical features to select the most important sentences from the story itself. Using extractive methods relying on those features, we aimed at producing a summary meant to be a quick glance inside a book.

Data & Methods

The summary is created by weighting sentences chapter per chapter and the top sentence is selected for each chapter. To weigh the sentences we start by ranking the synsets appearing in the book using wordnet from the nltk.corpus module. The words contained in the stopwords corpus from the nltk.corpus module are not taken into account because they would hinder the ranking of the synsets. We also rank the proper nouns presents in the book, which are detected by the pos_tag

tagger from the nltk.tag module. Sentences that contain more than 20 words are filtered out because they do not fit in a summary. The remaining sentences are then ranked using the ranks of their words' synsets. A sentence also receives a bonus for the number of proper nouns it contains as well as their rank. The last criteria is the position of a sentence in a chapter. Sentences towards the beginning or the end of a chapter are preferred as they are more likely to contain introductive or conclusive information.

As an example, we used the book *Oliver Twist* by Charles Dickens.

Results : sentences selected

Chapter I
Oliver cried lustily.

Chapter II
'Make a bow to the gentleman, Oliver,' said Mrs. Mann.

Chapter III
'It's a nasty trade,' said Mr. Limbkins, when Gamfield had again Stated his wish.

Chapter IV
'My dear,' said Mr. Sowerberry, deferentially, 'this is the boy From the workhouse that I told you of.' Oliver bowed again.

Chapter V
Oliver wondered, in his own mind, whether it had taken a very long time to get Mr. Sowerberry used to it.

Chapter VI
You can hold a knife to that black eye, as you run along.

Chapter VII
'Oh, Mr. Bumble, sir!' said Noah: 'Oliver, sir,-- Oliver has--'

Conclusion

If we expect the sentences selected to depict what is happening in a book, we can say that our method does not give great results (the sentences might even contain spoilers). Also, choosing not to keep long sentences might hide the real style of the writing. Therefore, a book which mainly contains long descriptive sentences and only a few short dialogue sentences could be judged as a dialogue centered book to someone reading the summary. Limiting the sentences selection to exactly one sentence per chapter might also be irrelevant when the amount of information carried by each chapter varies greatly. Although the sentences selected are not the most relevant, we think that, considering the simplicity of the methods used, it yields satisfying results. Additional criteria, for example regarding the predominant kind of sentences, could also be used to produce a more precise set of sentences but are more complicated to handle.

Bibliography

- Mani I. *Automatic Summarization*, John Benjamins Publishing Co., 2001

-Hirschman, Lynette, and Guy Story.
"Representing Implicit AndExplicit Time Relations In Narrative." *IJCAI*. Vol. 81. 1981.

We wish to thank the University of Paris Diderot for allowing us to produce this work, as well as Nicolas Ballier and Jean-Baptiste Yunès for their help.