Story Visualization via NLP and Word Clouds

This paper describes a software application that helps users visualize a story by making use of natural language processing (NLP) software and data visualization. They make use of two open source software libraries for extracting meaning from a story like textual narrative and displaying it in an interactive manner to users. The first library is Stanford’s Core NLP, a Java API that provides natural language processing services such as sentence recognition, tokenizing, parts-of-speech identification, dependency parsing, named entity recognition, and coreference resolution. CoreNLP is used to extract the subject/verb/object/place/time/context associations of a textual story. They consist of suite of tools called annotators that can perform many NLP functions, some of them are:

• Breaking a text document into individual sentences

• Tokenizing a sentence (breaking it into individual “words”)

• Identifying parts of speech (POS) within a sentence (nouns, verbs, adjectives, adverbs, etc.)

• Named entity recognition – recognizing names of people, places, organizations, dates/times, etc.

• Constituency parsing – constructing taxonomies of noun phrases and verb phrases of a sentence

• Dependency parsing – constructing the graph of dependency relationships between terms in a sentence

• Co-reference resolution – finding all expressions that refer to the same entity in a text

The second is D3, a JavaScript API built on scalable vector graphics which provides powerful data visualization capabilities.

Limitation:

It is limited to a narrow branch of text understanding; it applies only to understanding a story. The application is not well suited for extracting meaning from other types of text (e.g. ascertaining taxonomic hierarchies, understanding philosophical readings, etc.).