GraphLang

Why ?

Reading is too mainstream. What if you could get the important ideas from a text without even reading it ? What about comparing several documents based on their textual content ? Or maybe you just want to visualize the concepts present in a book and their interaction ? GraphLang is the tool you (will) need to boost your texts using graphs.

How ?

This project is all about analysing textual resources using graphs and extracting useful insights. The main idea is to represent a document using the cooccurrences of its words, turning that into a graph and leverage the power of graph analysis tools to make sense of this document.

At first the graph could be built by only considering words directly adjacent to each other and representing this proximity with a link in the graph, where the nodes would be the words themselves. The recipe could then be complexified by considering also words at distance N from each other (N would have to be defined) and defining edge weights as a function of N. Punctuation could also be taken into account and would influence the weight of edges (two words, one at the end of a sentence and the other at the beginning of the next one shouldn't (maybe) have a strong edge between them).

This graph could be extended to take into account multiple documents at once using signals on the edges.

Another idea would be to use the natural graph structure that emerges from the natural language and the relations between words. This could be extracted using a "Natural language processing" tool.

What ?

The techniques mentioned above could be applied to a bunch of textual resources, including news articles, books or even tweets (the latter ones could be used in batch, one tweet alone would certainly not provide enough information).

Who ?

The "Quatre Mousquetaires", to serve you.

Ali Hosseiny

Charles Gallay

Maxime Delisle

Grégoire Clément