Generating Interactive Tag Clouds on Scientific Documents

Visualizing clusters as tag clouds

Having a repository of clusters, which store relevant information about scientific articles that are thought to be alike, it was thought that an intuitive and friendly way of visualizing this data could be accomplished through the creation tag clouds.

The **objective** is to develop a system with the appropriate technologies that will present the data found in the clusters as tag clouds, in such a way that at first sight, a person would get to know which are the most important topics in alike articles. Someone interested in a certain topic would be able to interact with the tag cloud in order to get more information regarding the articles within that topic, and the authors involved.

Tackling the problem

Several aspects were taken in consideration when creating the cloud cloud, such as:

- Stopword filtering. Words not important in the context were discarded.
- Stemming. Words with the same root and meaning were given the corresponding importance.
- Language detection. Articles in other languages were to be excluded from the tag clouds.

Initial solution

The initial solution involved OpenCloud, a library in Java that facilitates the creation of tag clouds for the web. Using HTML and CSS, the tag cloud was given the desired styling and presentaion.

Evolving approach

Having in mind the importance of users understanding at first sight the content of the cluster of articles, it was decided that there was a better, simpler, and visually pleasing way to display the tag cloud. Alike terms could be grouped in a section of the tag cloud, choosing the appropriate size, depending on the importance of such word.

Interacting with the Users

Architecture

Future Work