



GENERATION OF INTERACTIVE TAG CLOUDS ON SCIENTIFIC ARTICLES

Jesús Antonio Soto Velázquez



INTRO

Having a large collection of scientific articles grouped by topic, it was thought that the best way to visualize the main topics of these groups was through the use of tag clouds. At first glance, one is able to recognize what the articles in a certain group are about.



A system that enables users to search through topics to find the corresponding group of articles and the involved artists is the **objective**. The *key* is the interaction between the user and the tag cloud.

TACKLING THE PROBLEM

Several aspects were taken into account when choosing the right information to put in the tag cloud, such as:

Stopword filtering Unimportant words in the given context were discarded. Resolved with *String matching algorithms*.

Word stemming Words with the same root were grouped together. Resolved with the *Snowball library*.

Language detection Articles in other languages were to be excluded from the tag clouds. Resolved with the *language-detection library*.

The tag cloud Manage structure and appearance of a tag cloud. Resolved with the *OpenCloud java library*.

Portability As the intention was to reach as many users as possible, a web environment was chosen. Technologies used: *HTML5*, *CSS*, *Javascript*, *Servlets*

EVOLVING APPROACH

Having in mind the objective of the project, it was thought of a better way to position the items in the tag cloud to illustrate the relationship between similar and important topics in the group of articles.



INTERACTING WITH USERS

The key to get more insight on the groups of articles was to allow the interaction with the tag cloud. An user can click an item in the tag cloud to find out who are the active researchers in that area, as well as discovering other topics they are involved in. HTML5 and its canvas was used to retrieve information on the clicked item.

INDEXING ARTICLES

To successfully deploy a system where topics inside the groups of articles were to be quickly searched, a structure known as an index was chosen. Each document is indexed so that the topics of the articles can be queried. The tool used to index the groups of articles, Solr, provides a simple interface between the data stored and the means of returning the desired information in a web environment.

FUTURE WORK

Although the project focused on the many components involved, there is much work to be done to integrate these parts into a system for use in the web. Future work will focus mostly on deploying an architecture where the user inputs topics as queries, and the server processes these queries to return the tag clouds where such topics have been seen in certain groups of articles.

REFERENCES

Charlie the unicorn for being the source of my inspiration.

INITIAL SOLUTION

The initial solution made use of 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 presentation.

[algorithms](#) [analysis](#) [approach](#) [computational](#)

[crossover](#) [data](#) [dynamics](#) [editorial](#) [evolutionary](#)

[genetic](#) [images](#) [issue](#) [metabolic](#) [models](#) [network](#)

[optimization](#) [part](#) [particle](#) [pathways](#) [problem](#)

[programming](#) [protein](#) [quantitative](#) [schema](#)

[special](#) [structure](#) [swarm](#) [systems](#) [theory](#) [using](#)