

The **objective** is to deploy a web platform which generates tag clouds with meaningful information extracted from a collection of documents. This work focuses on such an insightful *visualization* of the topics in the documents.

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*

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

Rather than focusing on the artistic side of tag clouds, like most tag cloud tools do, an approach on semantic similarity between topics was taken. As such, the position of each topic is determined by the semantic similarity of itself and its surroundings.

A word cloud visualization of the lyrics "do re mi k-ching! galletas pastel pastel galletas". The words are arranged in a non-uniform, overlapping manner. "do re mi" is at the top in green. "k-ching!" is in the middle in blue. "galletas" appears twice, once in blue and once in teal. "pastel" appears twice, once in purple and once in red. The words are of varying sizes, with "do re mi" and "galletas" being larger than "k-ching!". The background is white with a thin black border on the left and bottom.

The first step to allow interaction with the tag cloud was through the functionality of clicking a word in the tag cloud, and firing an event for that particular word. This was done by looking for a match of the cursor's coordinates and all of the words' coordinates.

It might be of interest knowing how active or inactive the topics have been throughout the years. For such cases, a two-color gradient is used to represent how active each of the topics have been, or to know if its use has declined after some time. The brighter the color, the more active it is.

A word cloud containing the following terms: 'problem', 'communication', 'programming', 'genetic', 'theory', 'optimization', and 'evolutionary'. The words are arranged in a scattered pattern, with 'programming' and 'genetic' being the most prominent.

In order to quickly search through the documents by typing a keyword, a structure known as an index was used. 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. The data is stored and structured as XML.

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. Key points to advance development are:

- Data retrieval from tag cloud. Through user interaction, more information about the selected topic can be obtained, such as researchers involved.
- Given a document from the index in XML, a tag cloud should be generated.

- Porter, Martin; Boulton, Richard. *Stemming Language Snowball*. <http://snowball.tartarus.org/>
- Mcavallo. *Tag cloud Java library* <http://opencloud.mcavallo.org/>
- Apache Software Foundation. *Apache Solr* <http://lucene.apache.org/solr/>