# BackEnd

In the Backend module, we make the research. This is done by three step :

- fetching articles

- preprocessing articles text

- clustering of the articles

**Fetching articles**

The goal : get in local all article and meta-data that match our search.

**Preprocessing articles**

The goal : keep only the words the most pertinant of the article and create tf-idf file to give to our clustering process.

This step can be make on the abstract or the full-text or the both. Now, we let for abstract beacause this is faster and there are too parasite words in full-text. For the both, this is a simple concatenation between them. We add he title to the abstract too.

**Clustering articles**

The goal : discover by calcul the clusters among the articles and classify the articles in them.

In this step, we create an object for clustering and bind it to our database. It will keep all meta-data of the process. We launch a number of trial and when this is finish, we attribute to each article of the research, his coordinates x-y and the topics name of his cluster. The topic name of a cluster is simply a string with the words that appears the most in the articles of the cluster. Each word i separated by comma.

For each step, we save intermediate data so if there is a bug, we can restart where the app failed. We check in which step we were and go directly there. We try to not make the job twice.