

First collection phase

Sample API

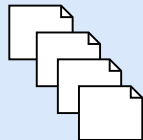
Tweet collection C_0



Objective: build a collection of random French tweets emitted during a long period of time.

Pre-processing

Tweet collection C_0



Co-occurrence matrix



Extract French tweets
Build vocabulary V
Build co-occurrence matrix M

Clustering

Co-occurrence matrix



Clusters



Objective: build K clusters of words, to group together terms that frequently co-occur.

Second collection phase

Sample API

Tweet collection C_1



Filter API
keywords: c_1 , lang: fr

Tweet collection $C_{K,N}$



Filter API
keywords: c_i , lang: fr



Filter API
keywords: c_K , lang: fr



Objective: Find the optimal parameters K and N that maximize the structure similarity between tweet collections C_1 and $C_{K,N}$.