

Visualiser des textes en humanités numériques

9 décembre 2021

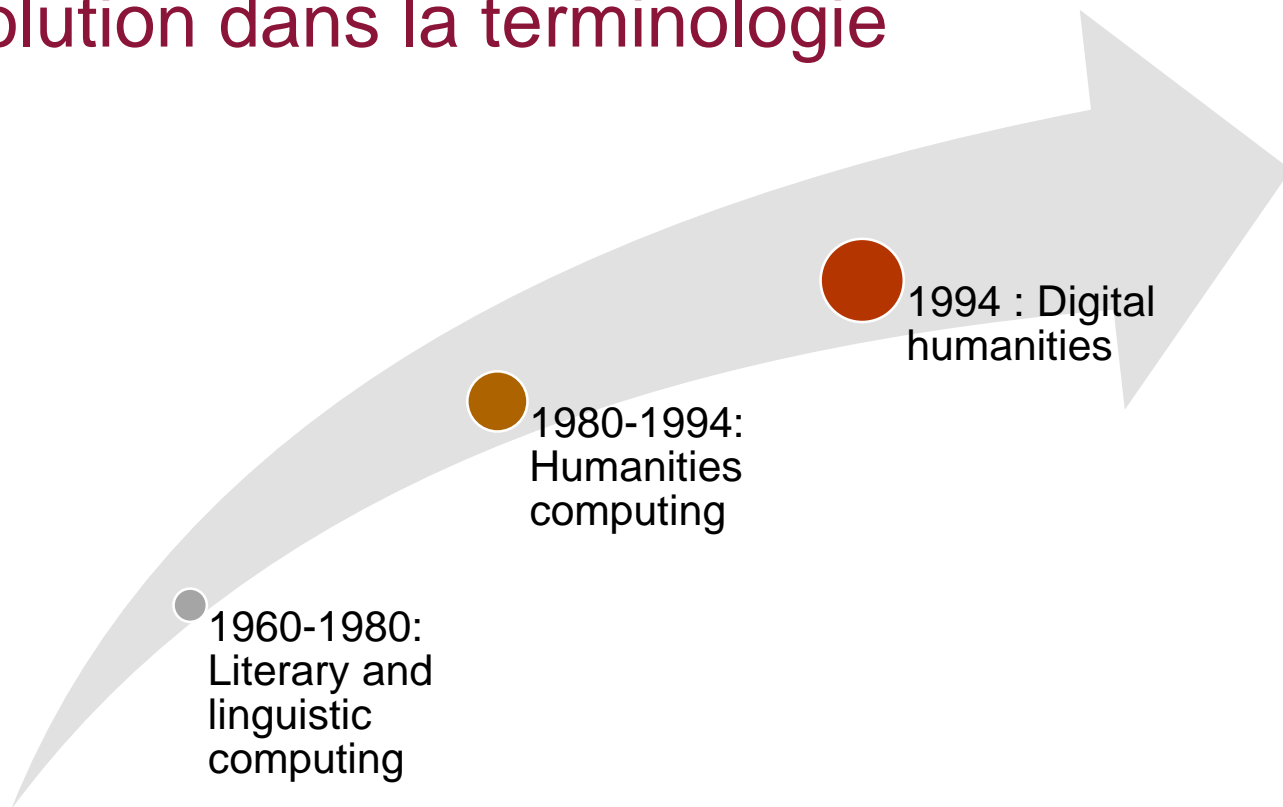
Suzanne Mpouli
Centre des Humanités Numériques
suzanne.mpouli@u-paris.fr

Plan

I. Analyse textuelle et humanités numériques

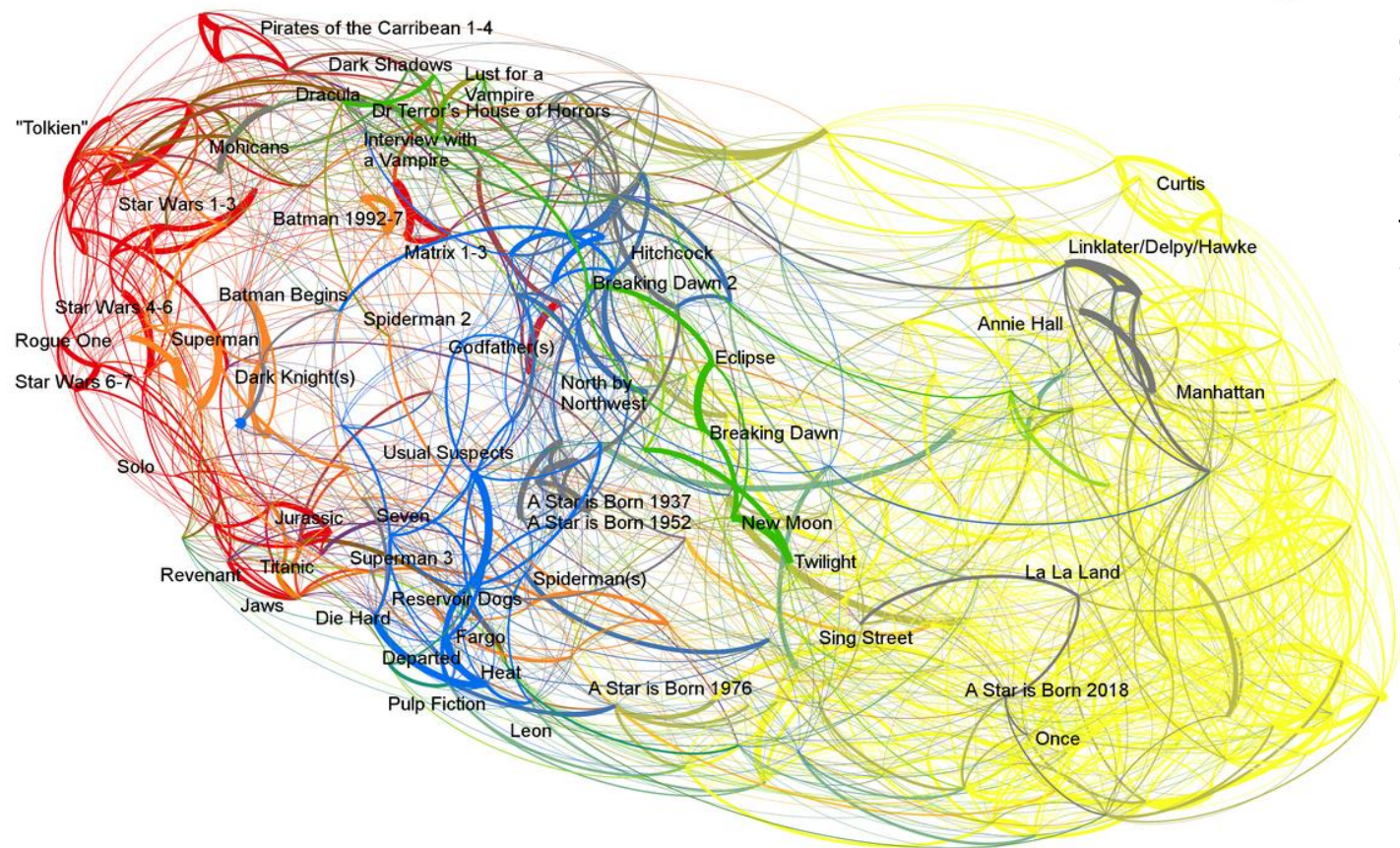
II. Visualisation de données textuelles : l'exemple de Voyant Tools

Evolution dans la terminologie



Intérêt des outils informatiques pour l'étude des textes

- ❖ Possibilité d'analyser de grands volumes de textes
- ❖ Automatisation des tâches
- ❖ Puissance de calcul et précision des ordinateurs
- ❖ Visualisation des résultats



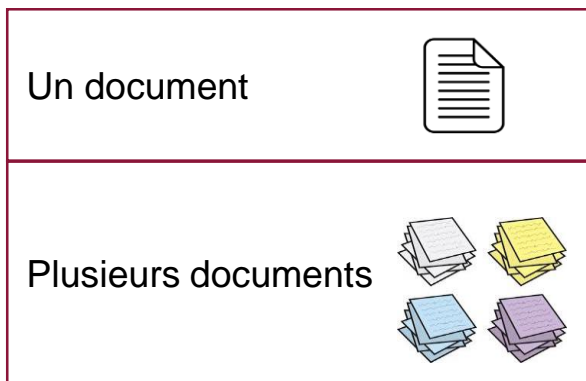
GENRES

- Action → rouge
- Thriller → blue
- Romance → gris

THEMES

- Comédie romantique → jaune
- Super-héros → orange
- Vampire → vert

Niveaux d'analyse



Termes ou
structures
spécifiques

Entités nommées

Ensemble du
contenu

Google Books Ngram Viewer

Q literary and linguistic computing, humanities computing, digital humanities

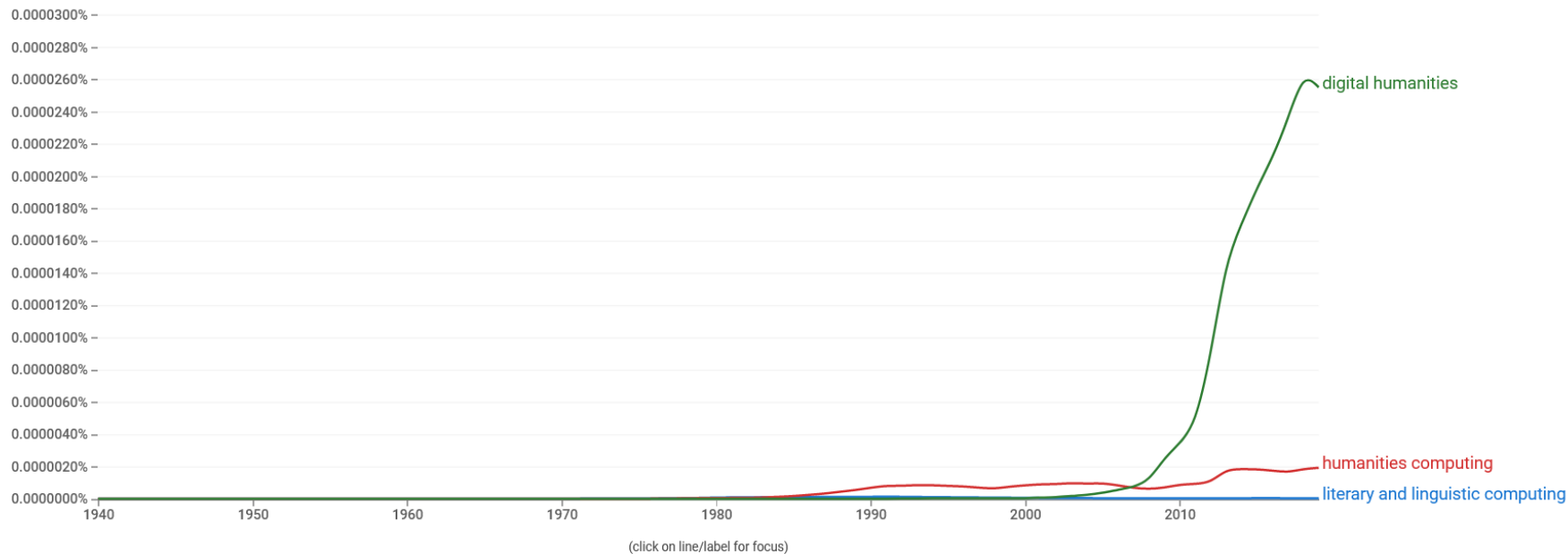


1940 - 2019

English (2019)

Case-Insensitive

Smoothing



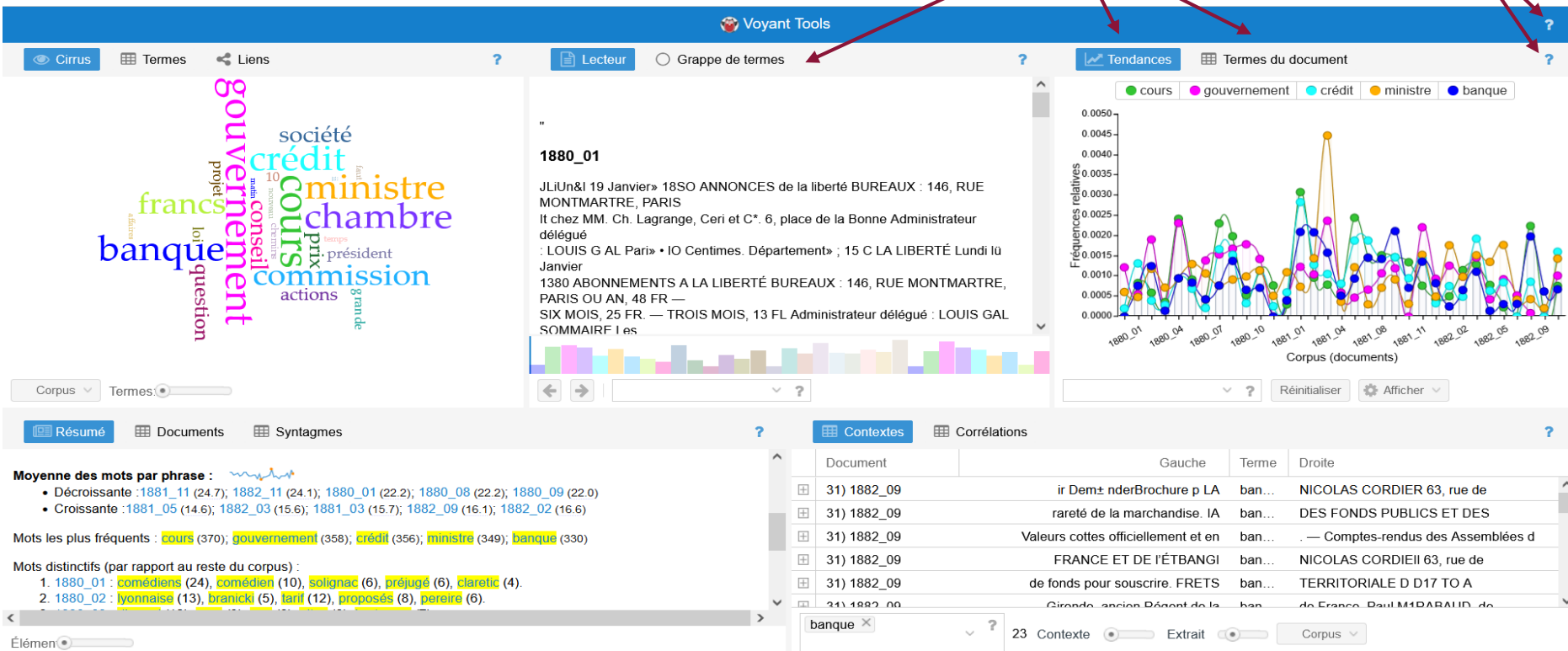


- ❖ Interface en ligne proposant plus d'une vingtaine d'outils d'exploration et de visualisation des textes avec prise en charge de différents types de formats (txt, xml, doc, pdf...).
- ❖ Site principal <https://voyant-tools.org/> + site miroir <https://voyant-tools.huma-num.fr>
- ❖ Documentation en ligne: <https://voyant-tools.huma-num.fr/docs/#!/guide/about>
(anglais uniquement)

Interface par défaut

Noms des outils (l'outil utilisé est surligné en bleu)

Aide et boutons supplémentaires



Aide et boutons supplémentaires

Options de l'outil, par exemple pour modifier la liste de mots-outils (stopwords)

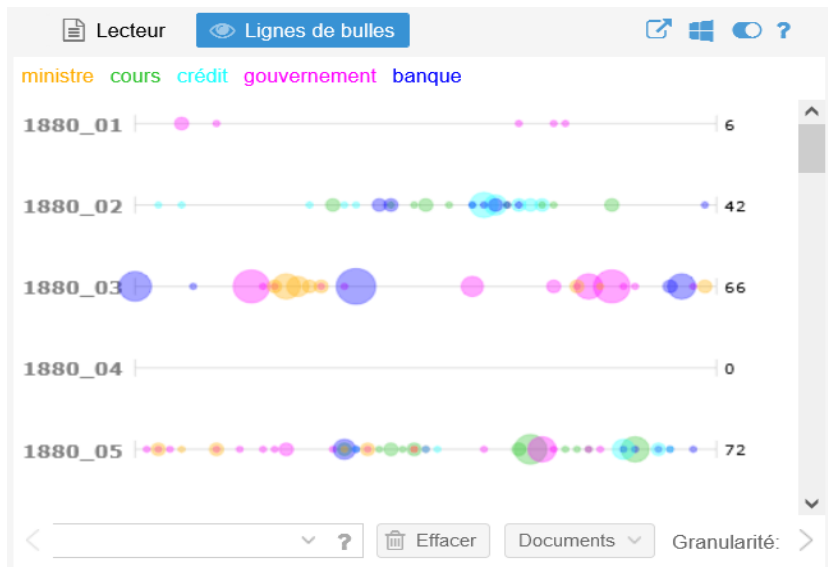
Définir les options de cet outil

Choix d'un nouvel outil (les outils sont classés par type)

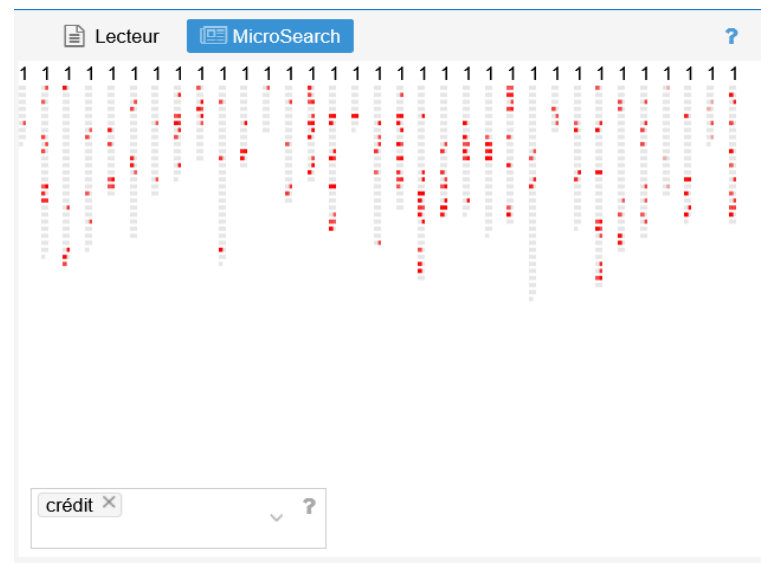
The image displays the 'Termes du document' menu and a line graph. The menu is open, showing various tool categories: Liens, Collocations, Outils de corpus, Outils de documents, Outils de visualisation, Outils de tableaux, and Autres outils. The graph shows the relative frequencies of three terms (émissions, changements, climatique) across five time periods (AR1fr-1992, AR2fr-1995, AR3fr-2001, AR4fr-2007, AR5fr-2014). The y-axis is labeled 'Fréquences relatives' and ranges from 0.000 to 0.011. The x-axis is labeled 'Corpus (documents)'.

Corpus (documents)	émissions	changements	climatique
AR1fr-1992	0.0040	0.0018	0.0058
AR2fr-1995	0.0055	0.0022	0.0040
AR3fr-2001	0.0038	0.0095	0.0035
AR4fr-2007	0.0035	0.0025	0.0012
AR5fr-2014	0.0045	0.0032	0.0018

Distribution d'un terme

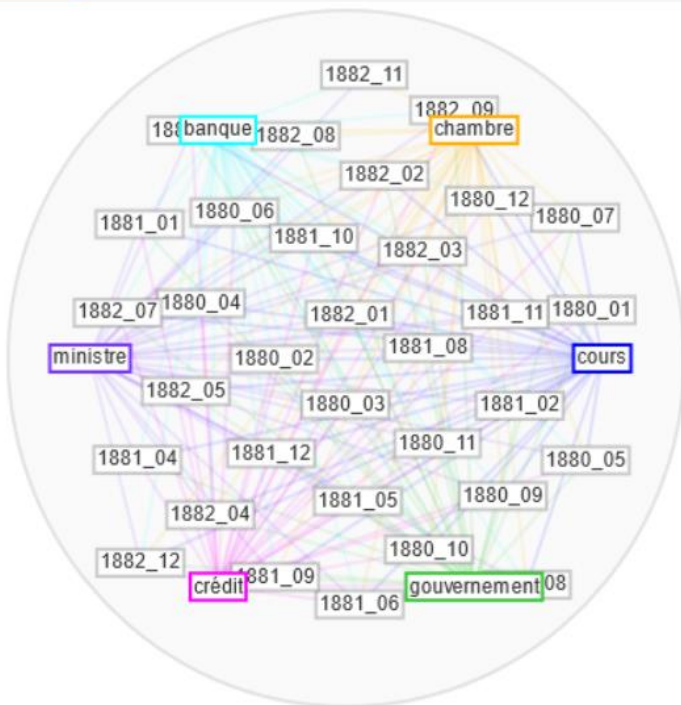


Les grosses bulles correspondent à des passages de forte concentration du terme.

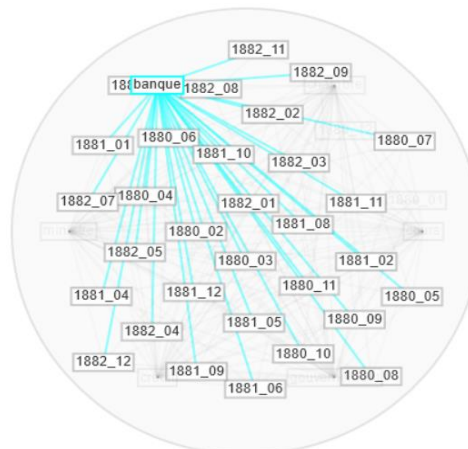
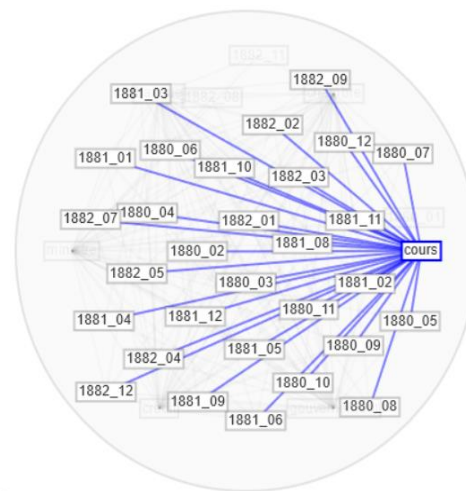


Les zones rouges correspondent aux parties du texte où figure(nt) le(s) mot(s) recherchés. Le bloc gris indique la longueur du document.

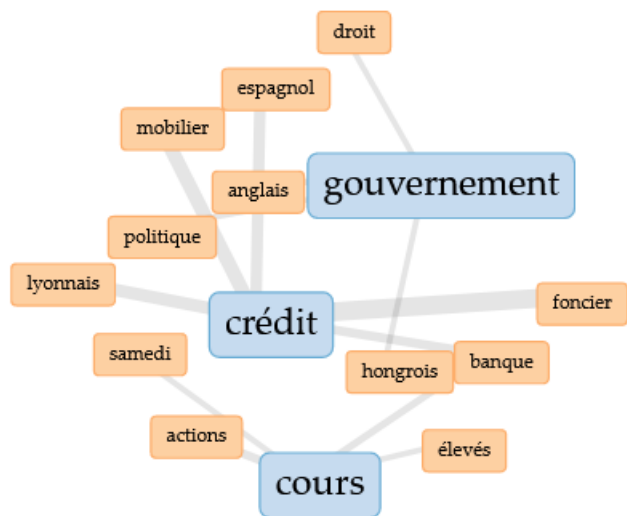
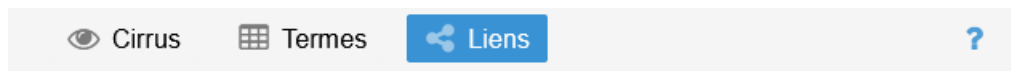
Relations termes-documents

☐ Mandala


Un terme agit comme un aimant et attire vers lui tout texte dans lequel il est très fréquent.



Corrélation des termes

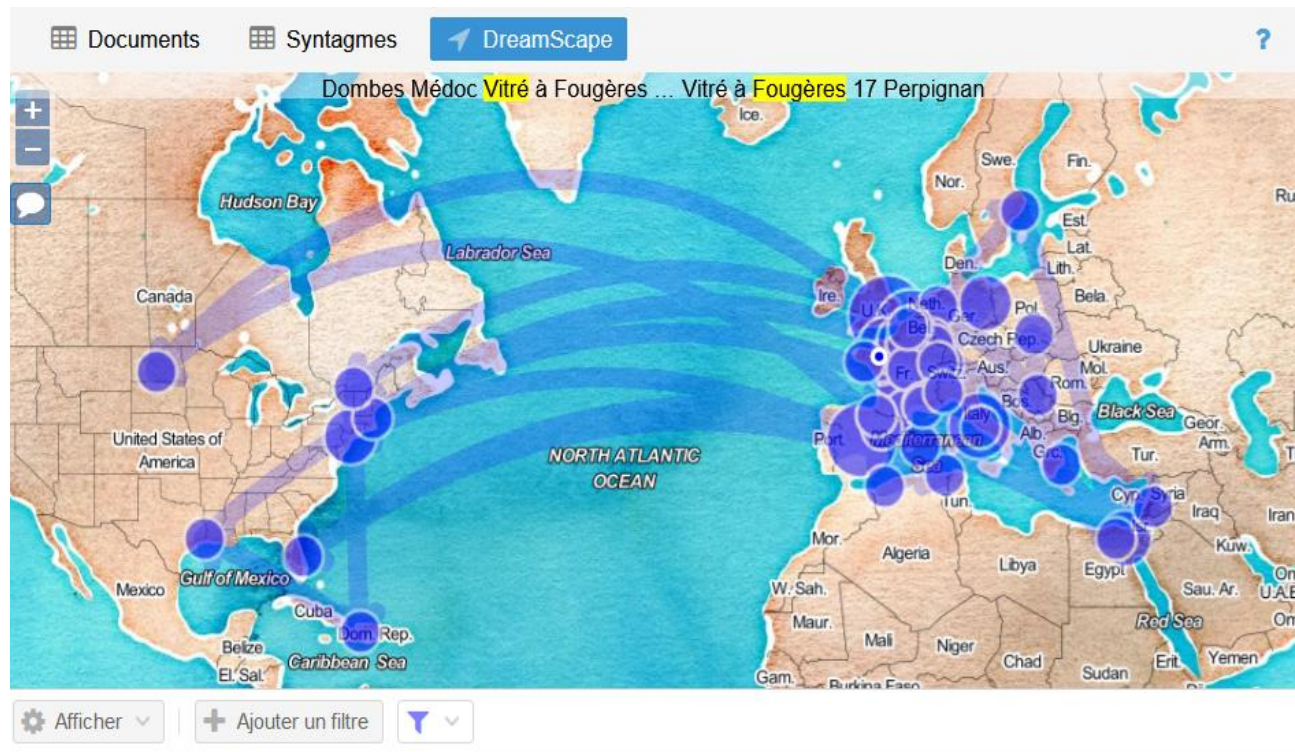


Mise en relation des mots-clés (bleu) et des termes avec lesquels ils sont le plus souvent utilisés.

LOUIS G AL Pari» • IO Centimes. Département» ; 15 C LA LIBERTÉ

Lecture dynamique du texte; le mot concerné est en bleu et ses occurrences dans le texte sont indiquées par les traits verts. On peut aussi faire apparaître les occurrences d'autres termes (ici, « banque » en rouge).

Géolocalisation



Aperçu dynamique des déplacements spatiaux dans le texte.

Conclusion

- Visualiser des données et expliciter les liens qui les unissent est fait partie intégrante des humanités numériques.
- Les outils de visualisation peuvent permettre de se familiariser avec des données.
- Il est important de choisir le type de visualisation qui convient le mieux à ce qu'on veut mettre en valeur.

Merci de votre attention