



HAL
open science

La datavisualisation comme outil pour la recherche académique

Clara Galliano, David Reymond, Luc Quoniam

► **To cite this version:**

Clara Galliano, David Reymond, Luc Quoniam. La datavisualisation comme outil pour la recherche académique. Journée Visu 2019, May 2019, Paris, France. . hal-02289068v1

HAL Id: hal-02289068

<https://hal.science/hal-02289068v1>

Submitted on 16 Sep 2019 (v1), last revised 29 May 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La datavisualisation comme outil pour la recherche académique

Clara Galliano, David Reymond, Luc Quoniam

Laboratoire IMSIC – EA 7492

70 Avenue Roger Devoucoux, 83000 Toulon



Introduction

L'essor d'internet et du numérique a permis la démocratisation de l'information grâce à la disponibilité croissante des données rendue possible par le développement des "Nouvelles technologies de l'information et de la communication" (NTIC).

Ces avancées technologiques ont fait émerger des mouvements (« Open Science »), des nouvelles méthodes de recherche (datamining, traitement automatique des langues, text mining, cartographie...), d'organisation des connaissances et de classification de l'information [1].

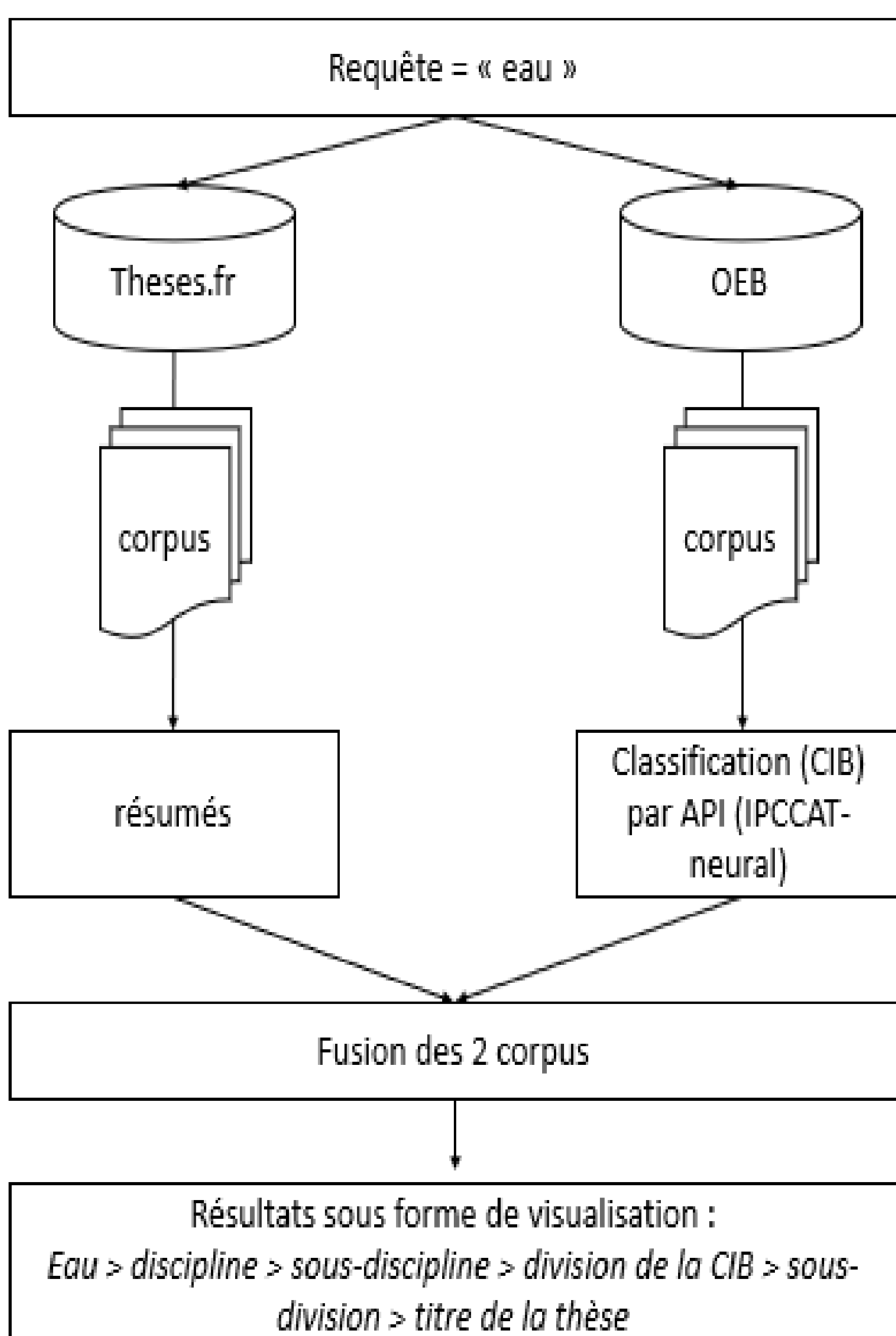
Objectifs

Les objectifs de cette étude sont multiples :

- Dresser une cartographie complète de la discipline des Sciences de l'Information et de la Communication à partir de plusieurs indicateurs bibliométriques (inspirée par des travaux existants [2][3])
- Développer un véritable outil d'aide à la constitution d'un état de l'art sur un champ/objet bien défini
- Montrer le lien entre « information académique » et « information technique » grâce aux ressources publiées par l'OMPI.

Méthodologie

Les bases de données utilisées sont : l'Office Européen des Brevets (OEB) et la base nationale des thèses françaises (theses.fr).



Pour réaliser l'ensemble des points abordés ci-dessus, de nombreux outils sont nécessaires :

- Langages de programmation : Python (pour la collecte et le traitement des données), HTML, CSS, JavaScript (D3.js), JSON
- L'API IPCCAT-neural proposée par la CIB pour une classification automatique de texte
- Logiciel libre de visualisation : Gephi.

Résultats

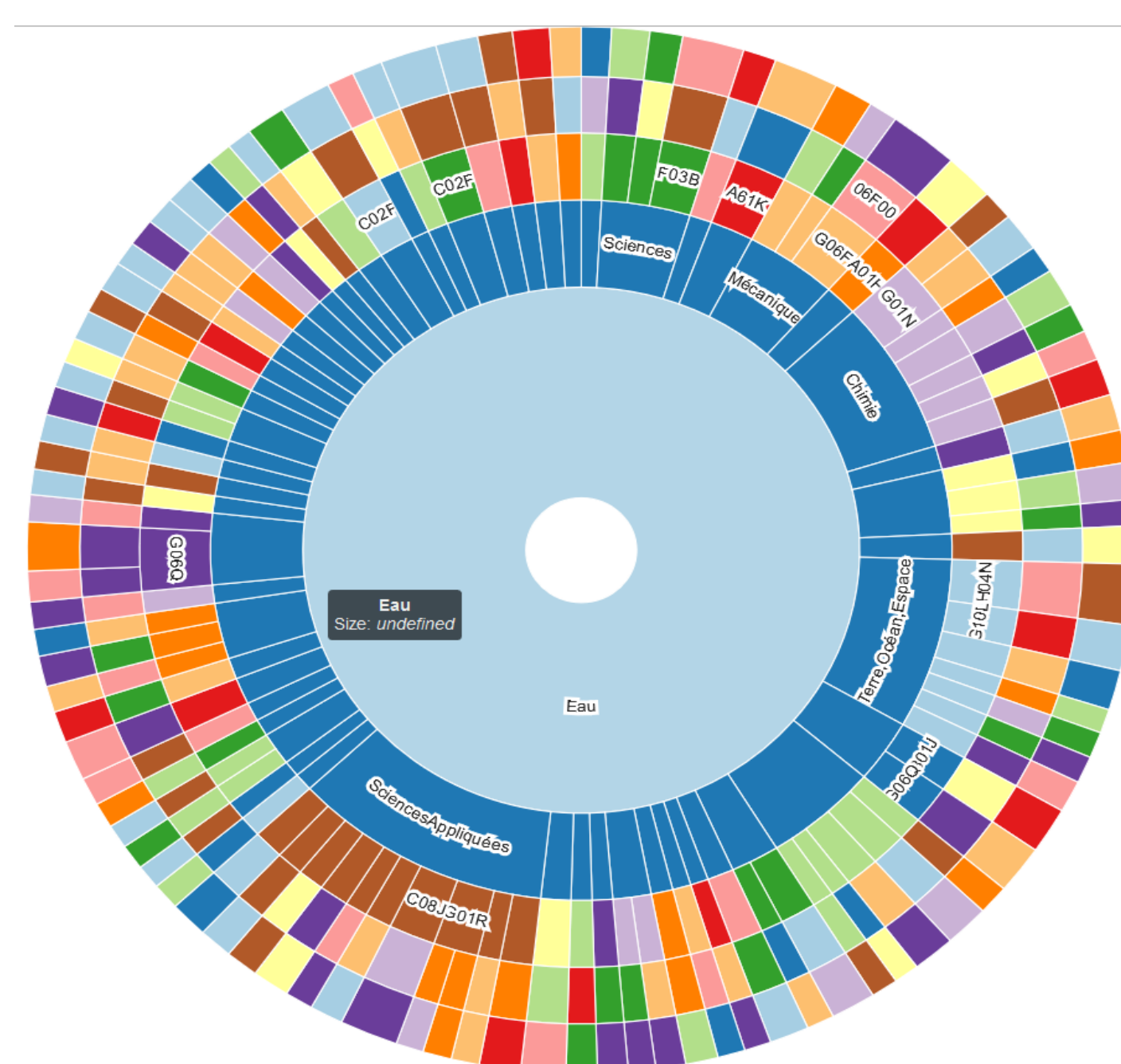
Notre expérimentation s'est appliquée ici sur le domaine de recherche « eau ».

A partir de la précision du classement hiérarchique du schéma de la CIB [4], nous obtenons le résultat suivant :

«eau» > discipline > classe CIB > sous-classe > titre de la thèse

Grâce à l'indexation des résumés de thèse permise par la classification de la CIB à partir d'une requête précise, nous avons proposé les visualisations suivantes :

- Cercle et cercle zoomable
- Nested Treemap (treemap imbriqué)
- Sunburst
- TidyTree.



Conclusion et perspectives

Cette expérimentation vise à amorcer l'utilisation de la CIB comme pivot à la recherche documentaire et à l'indexation de connaissance.

En perspective, l'étude s'est appropriée la discipline des SIC à laquelle nous appartenons. Mais notre recherche se veut interopérable et transposable à d'autres domaines des SHS et d'autres disciplines, avec un caractère interdisciplinaire.

Références

- [1] Hudon, M. & El Hadi, W. (2010). Organisation des connaissances et des ressources documentaires : De l'organisation hiérarchique centralisée à l'organisation sociale distribuée. *Les Cahiers du numérique*, vol. 6(3), 9-38. <https://www.cairn.info/revue-les-cahiers-du-numerique-2010-3-page-9.htm>.
- [2] Lrhoul, H., Chartron, G., Bachr, A. & Benammar, O. (2015). La datavisualisation comme outil de pilotage de la recherche scientifique médicale au sein de la Faculté de Médecine et de Pharmacie de Casablanca. In : Évelyne Broudoux éd., *Big Data - Open Data : Quelles valeurs ? Quels enjeux : Actes du colloque « Document numérique et société », Rabat, 2015* (pp. 165-181). Louvain-la-Neuve, Belgique: De Boeck Supérieur. doi:[10.3917/dbu.chron.2015.01.0165](https://doi.org/10.3917/dbu.chron.2015.01.0165).
- [3] Gallot, S. (2014) « Les enjeux d'une cartographie des SIC pour la discipline et les unités de recherche », *Revue française des sciences de l'information et de la communication* [En ligne], 5 | URL : <http://journals.openedition.org/rfsic/1191>
- [4] Fiévet, P., & Guyot, F. (2018). *Automatic Categorization of Patent Documents in the International Patent Classification (IPCCAT)*. Présenté à The International Conference on Search, Data and Text Mining and Visualization. (IC-SDV), Nice. Consulté à l'adresse <https://haxel.com/ii-sdv/2018/Programme/monday-23-april-2018>