Cartolabe, un système d'information par apprentissage automatique

Anne-Catherine Letournel, Philippe Caillou, Jean-Daniel Fekete, Jonas Renault, Michèle Sebag

▶ **To cite this version:**

Anne-Catherine Letournel, Philippe Caillou, Jean-Daniel Fekete, Jonas Renault, Michèle Sebag. Cartolabe, un système d'information par apprentissage automatique. 2018. hal-03248104

HAL Id: hal-03248104
https://hal.science/hal-03248104

Preprint submitted on 3 Jun 2021

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.

Cartolabe, un système d'information par apprentissage automatique

cartolabe.fr

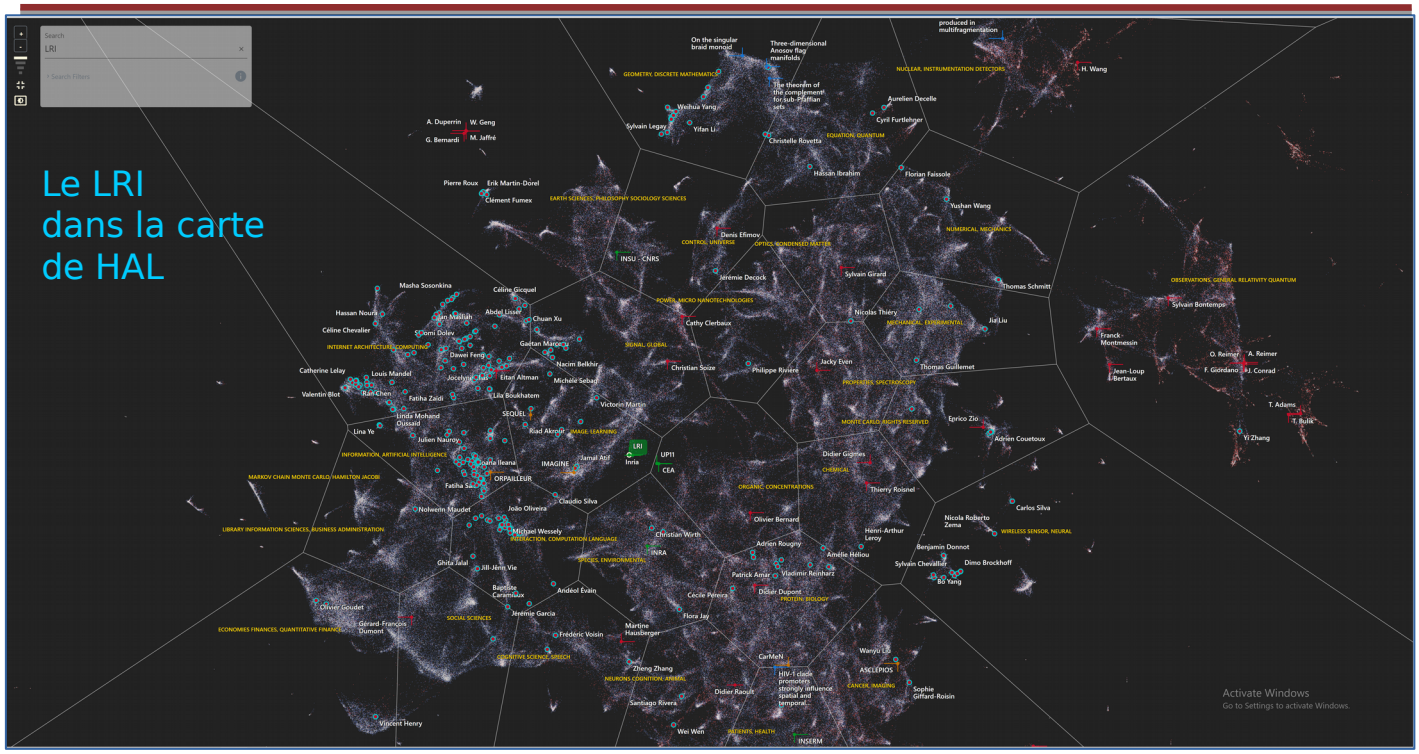
P. Caillou, J-D. Fekete, AC. Letournel, J. Renault, M. Sebag

Décembre 2018

Résumé

Cartolabe permet de visualiser le contenu d'un grand ensemble de documents. Une de ses fonctionalités est de représenter les compétences d'un ensemble de personnes en s'appuyant sur leurs articles scientifiques (ou autres documents disponibles).
Cartolabe crée une distance ou similarité entre documents, permettant de projeter les documents proches sur des points proches de la carte.

Cartolabe intègre aussi des données de plus haut niveau (auteurs, institutions, laboratoires) et les place dans la même carte. Il génère aussi automatiquement un clustering de thématiques afin de nommer des régions.
En donnant une vue d'ensemble, les institutionnels, les industriels ou les scientifiques l'utilisent pour des recherches d'interlocuteurs, d'experts ou des identifications de couvertures thématiques.



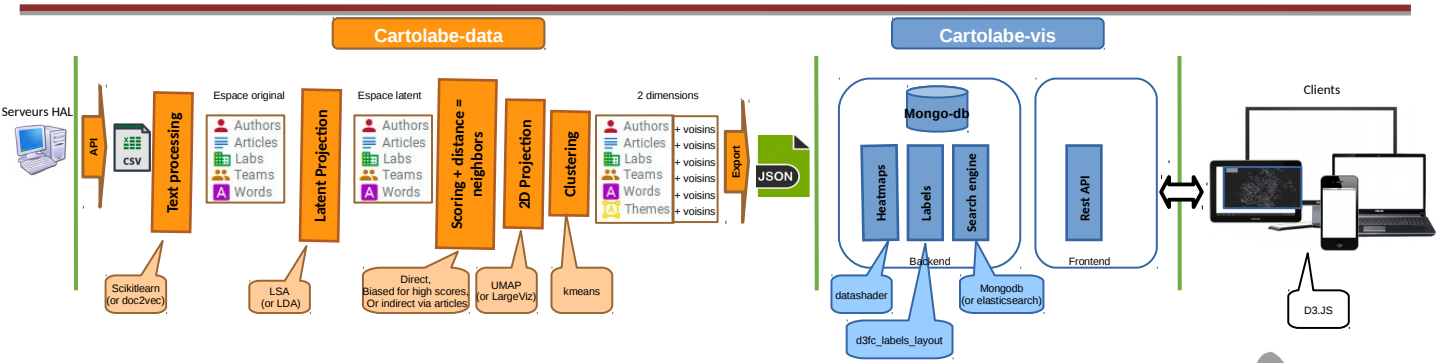
Fonctionalités

- La carte est zoomable ;
- Des interactions visuelles nombreuses : hover, clic, filtres...
- Un moteur de recherche sur les termes exacts ou sur la similarité de sacs de mots ;
- Un clustering sémantique automatique ;
- La superposition de sous-groupes dans la carte.

Innovations

- Un workflow dynamique basé sur des API requétables pour l'extraction des données ;
- Les points projetés sont de natures différentes ;
- La classification utilisée (clusters) est dynamique ;
- Le pipe-line d'analyse de données (python) est conçu pour être transposable à d'autres corpus ;
- La partie visualisation est adaptable à d'autres résultats d'analyse.

Architecture



www.lri.fr