

# Recommandation et cartographie d'articles scientifiques

## Supervision

- [Romain Rouvoy](#) (Maître de conférences – Spirals) – [prendre contact](#)
- [Filip Krikava](#) (Chercheur – Spirals) – [prendre contact](#)

## Lieu

[Spirals Research Group](#)

Inria Lille - Nord Europe

Parc Scientifique de la Haute Borne

40, avenue Halley - Bat. B, Park Plaza

59650 Villeneuve d'Ascq - FRANCE

## Contexte

Dans le domaine de la recherche et de l'innovation, un grand nombre de résultats sont publiés dans des revues scientifiques de manière à pouvoir être consultés *a posteriori* par d'autres chercheurs. Cependant, il est parfois difficile de trouver tous les articles en lien avec un sujet de recherche particulier. Certes, des moteurs de recherche comme DBLP [1] ou Google Scholar [2] permettent de trouver des articles similaires par recherche de mots clés dans le titre ou le contenu, cependant des mots clés trop généraux tendent à retourner de longues listes d'articles tandis que des mots clés trop précis peuvent donner lieu à l'omission de nombreux résultats pertinents.

Dans de nombreux cas, les chercheurs travaillent à partir d'un petit nombre d'articles qu'ils jugent très pertinents au regard des travaux qu'ils mènent. Ce sont ces articles qui leurs servent notamment de point de départ pour trouver d'autres articles connexes au sujet qu'ils traitent. À l'image de la recommandation de livres proposée par Amazon sur son site en ligne, nous souhaitons donc mettre en place un service capable de recommander des publications scientifiques à partir d'un jeu d'article.

## Projet

Dans le cadre de ce projet technique, l'objectif est de mettre en œuvre dans un premier temps un service de recommandation d'articles scientifiques qui est capable d'extraire les principaux sujets abordés par un jeu d'articles sélectionnés par un utilisateur en utilisant l'algorithme LDA (*Latent Dirichlet Allocation*) mis à disposition par la librairie Mahout [3] ou MLlib [4]. Ces sujets sont extraits sous la forme d'une séquence de mots clés qui pourront être ensuite soumis à des moteurs comme Google Scholar pour trouver d'autres articles pertinents sur le sujet.

Dans un deuxième temps, ce service pourra aussi être utilisé pour produire automatiquement une cartographie des articles collectés (e.g., sous la forme d'un graphe) et ainsi mieux comprendre la nature de leurs liens (mots clés en commun, auteurs en commun, article référant un autre, etc.).

## References

1. **DBLP.** <http://www.dblp.org>
2. **Google Scholar.** <http://scholar.google.fr>
3. **Apache Mahout.** <https://mahout.apache.org>

4. **Spark MLlib.** <https://spark.apache.org/docs/latest/mllib-guide.html>,  
<https://github.com/apache/spark/pull/476>