Rankingdom : une architecture coopérative pour l'analyse à la demande de Wikidata

Hassan Abdallah*, Béatrice Markhoff**, Manon Ovide*, Louise Parkin*, Arnaud Soulet*

*Université de Tours, LIFAT, Blois
**Université de Tours, CITERES UMR 7324, Blois
firstname.lastname@univ-tours.fr

Résumé. Nous présentons Rankingdom, une application Web d'analyse des entités de Wikidata destinée aux travailleurs de la connaissance. Son objectif est de positionner chaque entité considérée dans son domaine, et ce en la comparant à d'autres entités similaires en exploitant plusieurs indicateurs inspirés de l'infométrie. Par soucis de frugalité, nous nous appuyons sur une architecture coopérative qui décentralise le calcul à la demande des analyses sur les clients, mais centralise leurs résultats. Pour rendre possible le calcul à la demande des analyses malgré l'accord d'usage non abusif de Wikidata, les clients bénéficient d'un nouveau moteur de parallélisation des requêtes SPARQL, nommé SParaQL. Nous montrons l'efficacité de SParaQL sur un test de performance et de notre architecture coopérative avec le déploiement réel de Rankingdom.

1 Introduction

Les énormes graphes de connaissances du Web des données comme Wikidata (Vrandečić et Krötzsch, 2014) constituent une source de connaissances inestimable déjà largement utilisée pour la recherche d'information et la recommandation (Zou, 2020). Cependant, plusieurs travaux récents soulignent leur potentiel sous-exploité pour les travailleurs de la connaissance (Weikum, 2021; Zhu et al., 2024). En science des réseaux, l'avènement d'une nouvelle forme de réseaux complexes a entrainé systématiquement la proposition d'outils spécifiques de mesures pour l'analyse des entités décrites par ces données (e.g., la bibliométrie pour les réseaux d'articles scientifiques ou la webométrie pour les pages Web) (Thelwall, 2008). Naturellement, en s'inscrivant dans cette direction, les graphes de connaissances apparaissent donc comme une source potentielle prometteuse, d'autant qu'ils couvrent des domaines où il n'existe pas déjà d'outils d'analyses quantitatives. Prenons par exemple la peinture, il serait intéressant d'utiliser les connaissances de Wikidata pour mesurer et analyser son positionnement par rapport aux autres arts, ses principales périodes, ses peintres influents, etc. Malheureusement, à notre connaissance, il n'existe pas d'outil simple pour les travailleurs de la connaissance pour analyser les entités d'un graphe de connaissances.

Construire un outil d'analyse d'un graphe de connaissances aussi large que Wikidata entraîne deux défis majeurs. Premièrement, l'extrême variété des entités soulève la difficile question d'un schéma d'analyse transdisciplinaire. Par exemple, rien que le domaine de la peinture