

ISSA : un graphe de connaissances au service de la recherche bibliographique

Anne Toulet*, Franck Michel**, Anna Bobasheva**, Aline Menin **, Sébastien Dupré *, Marie-Claude Deboin *, Marco Winckler **, Andon Tchechmedjiev***

* Cirad Avenue Agropolis 34398 Montpellier Cedex – France
anne.toulet@cirad.fr,

** I3S (Univ. Côte d’Azur, CNRS, Inria) 06900 Sophia Antipolis - France
franck.michel@inria.fr

*** EuroMov Digital Health in Motion (IMT Mines Alès) 30100 Alès - France
andon.tchechmedjiev@mines-ales.fr

Résumé. Face à la multiplication des publications scientifiques, les archives scientifiques ouvertes jouent un rôle central pour aider les utilisateurs à effectuer des recherches bibliographiques. Cependant, les services de recherche classiques basés sur des mots-clés ne parviennent pas toujours à apporter des réponses satisfaisantes à certaines recherches complexes. Dans cet article, nous présentons les méthodes, outils et services mis en œuvre dans le cadre du projet ISSA pour répondre à cette problématique. Le projet vise à (1) fournir un pipeline générique, réutilisable et extensible pour l’analyse des documents d’une archive scientifique ouverte, (2) traduire le résultat en un index sémantique représenté sous la forme d’un graphe de connaissances RDF ; (3) développer des services de recherche et de visualisation qui exploitent cet index sémantique. Le projet ISSA s’inscrit dans la dynamique de la science ouverte et s’appuie sur un cas d’usage, Agritrop, l’archive ouverte du Cirad.

1 Explorer la littérature scientifique

Ces dernières années, l’accélération du rythme des publications et le “tout numérique” ont radicalement transformé la façon d’interagir avec la littérature scientifique, et les utilisateurs ont dû adapter leurs pratiques. Dans ce contexte, les bases de données bibliographiques, les moteurs de recherche et les archives ouvertes occupent une place de premier plan. Cependant, les services de recherche classiques, reposant généralement sur des correspondances de mots clés ou de noms d’auteurs, ne parviennent souvent pas à saisir la richesse des associations sémantiques entre les articles. Certaines recherches complexes trouvent difficilement des réponses et les résultats parfois peu pertinents obligent l’utilisateur à un filtrage manuel fastidieux. Devant ces difficultés, il est important de proposer aux chercheurs et aux professionnels de l’information des outils permettant de s’y retrouver.

Le présent article est un résumé de l’article publié dans la conférence ISWC 2022 «ISSA : Generic Pipeline, Knowledge Model and Visualization Tools to Help Scientists Search and