

AXIS - Plateforme d'Extraction d'Article et Analyses Statistiques

Brett Becker*, Matthieu Rochard*, Thi Lan Huong Nguyen*, H  l  ne de Ribaupierre**,
Laurent d'Orazio*

*Univ Rennes, CNRS, IRISA, Lannion, France

laurent.dorazio@univ-rennes1.fr

<http://www.iut-lannion.fr/>

****Cardiff University, Cardiff CF10, Wales**

deRibaupierreH@cardiff.ac.uk

Résumé. Une bibliographie est un élément fondamental dans un projet de recherche, en particulier pour définir un problème et motiver une proposition. Cet article de positionnement illustre le fonctionnement d'un outil permettant d'extraire des bibliographies. Il permet également de visualiser des résultats d'analyses statistiques sur des articles scientifiques. Cet outil propose un affichage récursif des références des articles et permet aux utilisateurs de parcourir toutes les sources de l'article. Le sexe des auteurs est déterminé automatiquement. Cela permettra de développer des statistiques précises sur un grand nombre d'articles.

1 Introduction

Effectuer des analyses sur les articles scientifiques est une tâche longue et complexe de par la grande quantité d'informations à collecter et à traiter. Les types de documents analysables sont très variés, il peut s'agir d'articles d'atelier, de colloques, de conférences, de revues, ou de rapports tels que des thèses de master ou de doctorat, sur un sujet écrit par un auteur donné. La quantité d'informations à traiter connaît également une croissance rapide. En effet, ces dernières années, le nombre de possibilités de soumettre des articles de recherche a fortement augmenté. à titre d'exemple, le nombre de documents de recherche référencés par DBLP était d'environ 25 000 en 1990, 76 000 en 2000, 223 000 en 2010, et était d'un peu moins de 379 000 en 2019. Dans un tel contexte, comment est-il possible d'aider les chercheurs à analyser une telle quantité d'informations, d'en savoir plus sur l'évolution de la recherche dans le monde et en particulier comment les femmes¹ chercheuses ont été prises en considération ? Le manque de diversité, qu'il soit genré ou ethnique, dans la recherche scientifique, est une problématique de plus en plus prise en compte et analysée. Une application permettant la simplification de la récolte de données et d'analyses statistiques de ces données permettrait de mettre ces ressources à la disposition d'un public plus large et d'étudier ces différents composants de façon plus aisée.

1. Dans cet article, nous prenons une approche binaire pour le genre