

ASSIST: Outil pour l'extraction et l'analyse statistique d'articles

Justine Fouillé^{*,**}, Thi Lan Huong Nguyen^{*,**}, Baptiste Alix^{*,**}, Brett Becker^{*,**},
Matthieu Rochard^{*,**}, H  l  ne de Ribaupierre^{*}, Laurent d'Orazio^{**}

^{*}Cardiff University, Wales
deRibaupierreH@cardiff.ac.uk,
^{**}Univ Rennes, CNRS, IRISA, France
laurent.dorazio@univ-rennes1.fr

R  sum  . Il y a moins d'auteurs que d'auteurs dans la recherche scientifique. Cependant, il n'existe pas encore,    notre connaissance, de syst  me permettant diff  rentes analyses sur les donn  es disponibles et de confirmer cette hypoth  se. Ce travail propose une extension d'un outil pr  c  demment r  alis  , afin de le rendre plus performant et d'ajouter de nouvelles fonctionnalit  s. Ces nouvelles fonctionnalit  s sont le nuage de mots-cl  s ou la nouvelle fonctionnalit   statistique. Les sources, r  f  rences et autres informations sur l'article seront affich  es pour chaque article r  cup  r  . Les sexes des auteurs seront d  termin  s    l'aide d'une base de donn  es reliant les pr  noms aux sexes, afin de pouvoir obtenir des statistiques sur un grand nombre d'articles collect  s.

1 Introduction

Publier ou p  rir illustre assez bien la pression que ressentent les chercheurs pour r  ussir une carri  re universitaire. Les chercheurs sont ainsi encourag  s    promouvoir leurs progr  s en r  digeant des articles en les soumettant    des ateliers, des symposiums, des conf  rences ou des revues, au niveau national ou international.

Des   tudes ont r  cemment d  montr   qu'il y a moins d'auteurs que d'auteurs dans le domaine de la recherche scientifique¹. Cependant, il n'existe pas encore de syst  me fournissant des donn  es concr  tes pour le prouver, permettant par exemple de conna  tre la proportion de femmes pour un domaine donn   (par exemple la m  decine ou l'informatique), un sous-domaine (les bases de donn  es ou l'interaction humaine par exemple), un domaine de recherche (comme l'int  gration de donn  es ou l'optimisation des requ  tes), ou selon les diff  rents   v  nements et selon leur classement (A*, A, B, C ou non class  ). Il est donc impossible de voir l'  volution de la place des femmes au cours du temps.

1. Dans cet article, nous ne consid  rons que les femmes et les hommes, et des recherches suppl  mentaires devraient   tre men  es pour inclure tout le monde.