

Recherche d'information et analyse factorielle des correspondances : étude des publications récentes sur la malaria

Annie Morin, IRISA, Université de Rennes 1, France

amorin@irisa.fr

Résumé

Il existe un certain nombre de méthodes statistiques disponibles en recherche d'information. Une des plus connues est l'analyse factorielle des correspondances (AFC) d'un tableau lexical croisant mots et documents. Dans ce papier, nous voulons illustrer l'utilisation intensive de l'AFC et de représentations visuelles appropriées des résultats pour traiter un grand nombre de documents. Nous avons choisi d'étudier des publications récentes (moins de 5 ans) sur la malaria, maladie qui provoque chaque année plus d'un million de décès dans le monde. L'objectif final est de déterminer les grandes tendances actuelles de la recherche à partir des publications sur cette maladie. On retrouve sur les plans factoriels les grands domaines de la recherche sur la malaria : biologie moléculaire, immunologie, épidémiologie, politique de santé publique, entomologie, biologie, pharmacologie. Pour chaque thème, on peut examiner les équipes de recherche participantes et leurs pays d'origine. Dans une deuxième étape, on sélectionne les mots caractéristiques d'un domaine pour sélectionner les documents concernés et les étudier plus en détail. Tout ce travail utilise exclusivement les propriétés de l'analyse des correspondances.

Références bibliographiques :

A. Morin, P. Bosc, G. Hebrail, L. Lebart , *Bases de données et statistiques*, Dunod 2002

M.W. Berry, *Survey of text mining*, Springer 2003

L.Lebart, A. Salem, *Statistique textuelle*, Dunod, 1994

M. Kerbaol., J.-Y. Bansard Sélection de la bibliographie des maladies rares par la technique du vocabulaire commun minimum, *JADT2000* vol1 Ed. Rajman, J-C Le Chapelier,ED EPFL 2000

J.-P. Benzécri, *L'analyse des correspondances* ,Dunod, Paris : 1973

A. Morin. Intensive use of correspondence analysis for information retrieval. *26th International Conference INFORMATION TECHNOLOGY INTERFACES ITI*, pages 1-4, Cavtat, Croatia, 2004

J. Rouillier, J.Y Bansard et M. Kerbaol, Application de l'analyse des données textuelles à une revue bibliographique de la littérature médicale, *JADT'2002*, Ed. A. Morin, P. Sébillot