

Analyse géographique de séries de publications : application aux conférences EGC

Eric Kergosien*, Marie-Noëlle Bessagnet**,
Christian Sallaberry**, Annig Le Parc - Lacayrelle **, Albert Royer **

*GERiiCO, Université de Lille 3, 59653, Villeneuve d'Ascq Cedex
eric.kergosien@univ-lille3.fr

**LIUPPA, Université de Pau, 64013 Pau cedex
prénom.nom@univ-pau.fr

Résumé. Dans cet article, nous présentons une méthodologie originale permettant de faire des analyses scientométriques basées sur trois dimensions (spatiale, temporelle et thématique) à partir d'un corpus de publications. Cette méthodologie comporte 3 étapes : (1) la préparation et la validation des données pour compléter les critères usuels tels que les noms d'auteurs, affiliation, ... par des critères spatiaux, temporels et thématiques ; (2) l'indexation des contenus des publications et métadonnées associées ; (3) l'analyse et/ou la recherche d'information multidimensionnelle. Les expérimentations sont menées sur la série de publications des conférences EGC de 2004 à 2015.

1 Introduction

Nous assistons à un accroissement prodigieux des publications scientifiques disponibles au format numérique, que ce soit à l'échelle nationale ou internationale. Ce que promet la société du numérique est une toute autre façon de représenter et de concevoir l'espace et le temps : c'est notamment le cas des travaux en scientométrie.

Nos travaux s'inscrivent pleinement dans cette démarche en proposant une méthodologie semi-automatique pour l'analyse de l'évolution dans le temps et dans l'espace (1) d'un ensemble de publications scientifiques et (2) des thématiques concernées. L'intérêt de ces travaux est notamment d'appuyer les scientifiques dans leur travail de veille en mettant en avant l'évolution des thématiques au fil du temps, selon les lieux des conférences et les lieux des laboratoires d'affiliation des auteurs. Bien que de nombreux travaux en scientométrie présentent des méthodes pour analyser des communautés à partir de publications scientifiques (que ce soit en revues ou en conférences), il n'existe pas à notre connaissance de travaux proposant une analyse géographique d'un corpus de publications, i.e. combinant les dimensions spatiale, temporelle et thématique. Notre méthode pour l'analyse du corpus se décompose en 3 étapes : (1) préparation et validation du corpus pour le marquage géographique, (2) indexation du contenu des publications et des métadonnées associées, (3) analyse semi-automatique du corpus et recherche d'information (RI) multidimensionnelle. En fonction du corpus, le travail de préparation est plus ou moins conséquent et peut impliquer une action de saisie manuelle. Nous expérimenterons notre méthode sur le corpus de 1103 publications présentées à EGC