

Classification de documents en réseaux petits mondes en vue d'apprentissage

Khazri Mohamed*, Tmar Mohamed**, Mohand Boughanem***, Abid Mohamed*

*Ecole Nationale d'Ingénieurs de Sfax, Route de Soukra, 3038, Sfax

Mohamed.khazri@yahoo.fr

mohamed.Abid@enis.rnu.tn

**Institut Supérieur d'Informatique et du Multimédia de Sfax, 3018, Sfax

mohamed.tmar@isimsf.rnu.tn

***Institut de Recherche en Informatique de Toulouse, route de Narbonne, 31000, Toulouse, France

bougha@irit.fr

1 Introduction

Les systèmes de recherche d'information préconisent une fonctionnalité très intéressante voire indispensable lors de tout processus de recherche : il s'agit de la reformulation automatique de la requête. Cette fonctionnalité permet de rétablir les choix de l'utilisateur dans la perspective de retrouver plus de documents qui répondent à son besoin en information. Il est à noter à ce niveau que le besoin en information de l'utilisateur est très vague : l'utilisateur ne sait en général pas ce qu'il cherche. Par ailleurs, il peut tolérer un résultat initial imprécis sous réserve de l'améliorer par feedback Rocchio (1971).

Faire recours à de nouvelles méthodes d'apprentissage est alors devenu une nécessité. Plusieurs modèles qui ont été auparavant délaissés, tels que la classification, sont repris en vue d'améliorer l'apprentissage en recherche d'information. Nous proposons dans ce papier une méthode d'apprentissage en faisant appel aux réseaux petits mondes (*small worlds en anglais*, Watts (1999)).

2 Notre Approche

Les propriétés des réseaux petits mondes paraissent intéressantes dans les problèmes de classification. D'autant plus que ces propriétés sont valorisées. Comme application à la recherche d'information, nous présumons qu'un ensemble de documents peut constituer des réseaux petits mondes pour moins qu'ils parlent du même sujet, et qu'une idée peut être transmise d'un document à un autre document si les auteurs partagent le même intérêt.

Nos objectifs pour l'intégration des *small worlds* en recherche d'information ont deux effets : un effet de construction des *small worlds* par le biais de la classification; et un effet d'estimation de pertinence sur d'autres documents.

En partant de l'hypothèse suivante : «une classe est raisonnable si elle admet certaines propriétés : celles des *small worlds*». Le premier effet va simplement faire une construction de *small worlds* de documents homogènes (pertinents ou non pertinents). Pour ce faire, nous proposons trois stratégies : une stratégie de construction de graphes de documents (1), une stratégie de propagation des liens (2), et une stratégie de construction des classes des documents (3). Pour la stratégie (3) nous utiliserons une méthode de classification hiérarchique, et l'identification du nombre de classes dépend de la qualité de classification et de la nature de