

L'utilisation des entités nommées pour l'expansion sémantique des requêtes Web

Bissan Audeh, Philippe Beaune, Michel Beigbeder

École Nationale Supérieure des Mines de Saint-Étienne-ENSM.SE/FAYOL
158 Cours Fauriel
42023 Saint-Etienne cedex 2,
{audeh, beaune, mbeig}@emse.fr

Résumé. Les entités nommées sont des éléments intéressants pour les applications fondées sur le Traitement du Langage Naturel. Dans le cas de la recherche d'information, les entités nommées sont largement employées par les utilisateurs du web dans les requêtes de recherche, soit pour définir un concept de base, soit pour décrire un autre concept dans la requête. Du côté du modèle de recherche, les entités nommées sont des éléments riches en information qui aident à mieux cibler les documents pertinents. Dans cet article, nous étudions l'avantage d'étendre les entités nommées dans la requête. L'idée est d'utiliser une technique d'expansion sémantique sur une ontologie générale (Yago) pour désambiguïser les entités nommées et pour trouver leurs différentes appellations que l'on intègre dans la requête en utilisant 3 approches : sac de mots, dépendance séquentielle, et concept clé. Nous mesurons l'efficacité de ces expériences en termes de précision et rappel, et nous étudions l'effet du rôle des entités nommées sur l'expansion. Nous concluons que l'expansion des entités nommées est une méthode simple qui améliore significativement la qualité de la recherche quand elle est comparée à un modèle de référence sans expansion. De plus, cette méthode est assez compétitive par rapport à l'approche pseudo retour de pertinence souvent utilisée pour l'expansion de la requête.

1 Introduction

L'expansion de la requête est une approche souvent utilisée en recherche d'information pour aider le modèle de recherche à mieux identifier les documents pertinents. Le succès de cette technique dépend du bon choix des termes d'expansion et la façon d'ajouter ces nouveaux termes à la requête. D'un côté, si les nouveaux mots clés apportés par l'approche d'expansion ne sont pas pertinents pour le besoin d'information, la possibilité de trouver des documents pertinents baisse, ce qui a un effet négatif sur le rappel et la précision. Pour cette raison, le contrôle de la qualité des termes d'expansion est une étape indispensable de l'expansion de la requête. D'un autre côté, un bon choix des termes d'expansion ne suffit pas pour garantir le succès de l'expansion si ces termes n'ont pas été intégrés correctement dans la requête. Dans cet article, nous nous concentrons sur les entités nommées, c'est à dire les termes qui