

Identification de compatibilités entre descripteurs de lieux et apprentissage automatique

Estelle Delpech^{*,**}, Laurent Candillier^{*,**}, Léa Laporte^{*,***}, Samuel Phan^{*,**}

^{*}Nomao, 1 avenue Jean Rieux, 31 500 Toulouse
{prenom}@nomao.com, www.nomao.com

^{**}Ebuzzing, 97 rue du Cherche-Midi, 75006 Paris
{prenom.nom}@ebuzzing.com, www.ebuzzing.com

^{***}IRIT, Université de Toulouse 3, 118 route de Narbonne, 31062 Toulouse Cedex 9
{nom}@irit.fr, www.irit.fr

Résumé. Les travaux présentés dans cet article s’inscrivent dans le paradigme des recherches visant à acquérir des relations sémantiques à partir de folksonomies (ensemble de tags attribués à des ressources par des utilisateurs). Nous expérimentons plusieurs approches issues de l’état de l’art ainsi que l’apport de l’apprentissage automatique pour l’identification de relations entre tags. Nous obtenons dans le meilleur des cas un taux d’erreur de 23,7 % (relations non reconnues ou fausses), ce qui est encourageant au vu de la difficulté de la tâche (les annotateurs humains ont un taux de désaccord de 12%).

1 Introduction

Notre travail se situe dans le cadre d’un système d’agrégation de descriptifs de commerces. Dans ce système, chaque commerce est décrit (entre autres) par un jeu de tags. Des connaissances sur la compatibilité ou l’incompatibilité de ces tags peuvent aider à prendre la décision d’agréger ou pas deux descriptifs. Par exemple, un descriptif contenant le tag *fitness* ne peut pas être agrégé avec un descriptif contenant le tag *menuisier* car, *a priori*, un même commerce ne peut pas à la fois rendre des services de menuiserie et avoir un lien avec le fitness. Par contre, les tags *peinture* et *carrelage* sont compatibles car ils peuvent, par exemple, être associés à un magasin d’ameublement. Pour acquérir ces informations de compatibilité, nous disposons de données dont la structure est identique à celle d’une *folksonomie*. Les folksonomies sont le produit de systèmes d’annotation collaborative tels Flickr ou Delicious qui permettent à une communauté d’*utilisateurs* d’annoter manuellement des *ressources* (urls, photos) à l’aide de descripteurs (*tags*). Le but de notre étude est de tester diverses approches pour l’acquisition de relations entre tags à partir de folksonomies. L’approche retenue sera ensuite intégrée au système d’agrégation de descriptifs décrit plus haut. En plus de tester les approches état-de-l’art, nous explorons aussi l’apport de l’apprentissage automatique pour la détection de relations sémantiques à partir de folksonomies, ce qui, à notre connaissance, n’a pas encore été proposé.

L’article comprend 5 sections. La section 2 décrit les approches existantes pour l’identification de relations entre tags à partir de folksonomies. Les expériences menées sont décrites