

Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web

Fabian Abel, Qi Gao, Geert-Jan Houben, Ke Tao

Web Information Systems, Delft University of Technology
`{f.abel,q.gao,g.j.p.m.houben,k.tao}@tudelft.nl`

Abstract. As the most popular microblogging platform, the vast amount of content on Twitter is constantly growing so that the retrieval of relevant information (streams) is becoming more and more difficult every day. Representing the semantics of individual Twitter activities and modeling the interests of Twitter users would allow for personalization and therewith countervail the information overload. Given the variety and recency of topics people discuss on Twitter, semantic user profiles generated from Twitter posts moreover promise to be beneficial for other applications on the Social Web as well. However, automatically inferring the semantic meaning of Twitter posts is a non-trivial problem.

In this paper we investigate semantic user modeling based on Twitter posts. We introduce and analyze methods for linking Twitter posts with related news articles in order to contextualize Twitter activities. We then propose and compare strategies that exploit the semantics extracted from both tweets and related news articles to represent individual Twitter activities in a semantically meaningful way. A large-scale evaluation validates the benefits of our approach and shows that our methods relate tweets to news articles with high precision and coverage, enrich the semantics of tweets clearly and have strong impact on the construction of semantic user profiles for the Social Web.

Key words: semantic enrichment, twitter, user profile construction, news, linkage

1 Introduction and Motivation

With the advent of social networking, tagging or microblogging that become tangible in Social Web systems like Facebook, Delicious and Twitter, a new culture of participation penetrates the Web. Today, more than 190 million people are using Twitter and together publish more than 65 million messages (*tweets*) per day¹. Recent research shows that the exploitation of tweets allows for valuable applications such as earthquake warning systems [1], opinion mining [2] or discovery and ranking of fresh Web sites [3]. These applications mainly analyze and utilize the wisdom of the crowds as source of information rather than relying on individual tweets. Analogously, previous research in the field of microblogging

¹ <http://techcrunch.com/2010/06/08/twitter-190-million-users/>