The University of Amsterdam at WePS3

Manos Tsagkias, Krisztian Balog

ISLA, University of Amsterdam, The Netherlands
e.tsagkias@uva.nl, k.balog@uva.nl

Abstract. In this paper we describe our participation in the Third Web People Search (WePS3) evaluation campaign. We took part in the Online Reputation Management (ORM) task. Ambiguity of organization names (e.g., "Amazon" or "Apple") raises obvious difficulties for systems that attempt to trace mentions of and opinions about a specific company in Web data, in an unsupervised manner. Problems are further amplified in the context of user generated content, where proper capitalization of named entities is often absent. The ORM task, introduced this year, addresses this very problem, by setting out the following challenge: given a set of Twitter entries containing an (ambiguous) company name and given the homepage of the company, discriminate entries that do not refer the company. Given the above definition, it is natural to formulate the problem as a binary classification task. Our focus was on building a general organization classifier that predicts, for each tweet, whether it is about a company. Our goal is to assess how a system without external aid from other sources (the company's homepage, Wikipedia entry, etc.) can perform. We, therefore, focus on extracting features that are organization-independent and build on the characteristics of Twitter, such as noisy text, abbreviations and Twitter-specific language.

Specifically, we trained a J48 decision tree classifier using the following groups of features: (i) company name (matching based on character 3-grams), (ii) content value (whether the tweet contains URLs, hashtags or is part of a conversation), (iii) content quality (ratio of punctuation and capital characters), (iv) organizational context (ratio of words found in tweets labelled as positive).

We submitted a single run that performed around the median of all submitted systems. One interesting observation that requires further investigation is that our F-score for the negative class was substantially higher than for the positive class (0.55 vs. 0.36); for other teams it was usually the other way around.

In future work we plan to build company-specific models by exploiting content both from Twitter and from external sources.