1. Task Definition

Given a set of Twitter entries containing an (ambiguous) company name, and given the home page of the company, discriminate entries that do not refer the company. The motivation is to help experts in reputation management and alert services. Nowadays, the ambiguity of names is an important bottleneck for these experts. Twitter has been chosen as target data because it is a critical source for real time reputation management and also because ambiguity resolution is challenging: tweets are minimal and little context is available for resolving name ambiguity.

2. Data

The test and training data will consist of 500 names and 700 tweets for each name. The companies will be manually selected from several resources (such as dbpedia, see http://dbpe dia.org/ontology/Company
) trying to ensure that solving name ambiguity is crucial for the dataset. Thus companies named after common nouns (such as "Amazon") will take preference in the company selection process.

The 700 tweets per name will be in English, Spanish or both. The language of each tweet will be provided as metadata. The system input will include also the home page of the company (html document).

A subset of the 500 names will be provided as training set. The rest of the names will be used as test set.

3. Assessments and System Output

Systems must classify each tweet as positive (it refers to the company) or negative (it refers to something else). Assessment will be three-valued: positive, negative, or unclear. Only positive and negative cases will be used to assess the systems. The ambiguity will be considered at a lexical level: the sense of the name must be derived from the company, even if the sentence does not explicitly talk about the company., as in these examples about the Apple company:

...you can install 3rd-party apps that haven't been approved by Apple... TRUE...RUMOR: Apple Tablet to Have Webcam, 3G... TRUE
...
featuring me on vocals: http://itunes.apple.com/us/album/... TRUE
...Snack Attack: Warm Apple Toast... FALSE
...okay maybe i shouldn't have made that apple crumble... FALSE

4. Submission Format

The system output should be contained in a single tab-separated file. If the team is submitting output for multiple runs each one should be contained in a separate file. The file should be named with the team ID and a numeric suffix in the case of multiple runs (e.g. UNED_1.tsv, UNED_2.tsv, etc). Each line represents a classified tweet and has the following columns: entity name (the name used in the file "weps-3_task-2_test.tsv"), tweet identifier and the assigned label (either TRUE or FALSE).

For example:yamaha 12465638093 TRUEyamaha 12448811836 FALSElufthansa 1246575767
2   TRUE

Each can submit their results until June 21st to the address weps-organizers@lsi.uned.esPleas
e include in the subject of your email your team ID and the words "WePS-3 Task-2 submission".

5. Evaluation

The task will be evaluated as a standard classification task. Given that the degree of ambiguity in Twitter is difficult to predict, the system results can be easily biased to precision or recall. We will use the Unanimous Improvement Ratio (UIR) in order to test the robustness of system improvements against changes in the average ambiguity of the dataset.

6. Important dates

- Release of trial data .... 15 February 2010
- Release of test data .... 7 June 2010
- Submissions due .... 21 June 2010
- Release of official results . 15 July 2010
- Papers due .... 15 August 2010
- Workshop .... At CLEF 2010, Padova, 23 September 2010