We are making our URLs prettier -- Claim your personal URL now!

UMASS AMHERST

Completed • Knowledge • 41 teams

UMass Twitter NER

Thu 6 Nov 2014 – Mon 15 Dec 2014 (8 months ago)

Dashboard

Home
Data
Make a submission

Information
Description
Evaluation
Rules

Forum

Leaderboard

My Team

My Submissions

Leaderboard

1. Skeuomorph
2. Aysha
3. @JustinBieber
4. Schattie
5. pighead
6. NGU
7. CrazyBird
8. Aha
9. catscatsBORSCHTcats
10. ominousmango

Forum (0 topics)

41 teams

49 players

104 entries

This competition is private-entry. You've been invited to participate.

Find named entities in tweets!

The final project will be to construct an NER tagger for Twitter.

See the instructions on the course webpage for additional details.

The task of *named entity recognition* is to take tokenized sentences as input, then recognize *spans* of text that correspond to a name. We will not require entity types to be predicted.

Groups of 1 or 2, and collaboration policy

You can either do this project individually, or in a group of two. Please decide this by the milestone report date.

If you decide to do this as a group, when you submit the milestone and final reports/code, please submit a single copy of the materials, and make sure the names of both team members are on them. You both will receive the same grade.

This project is more open-ended than the problem sets. After building the base system, you are free to explore different types of features and analyses. Feel free to discuss the project with anyone else in the class. Discussions can help spur creativity for features or experiments to try. However, all code should be written individually or within your two-person group. The project reports should also be written only within the groups.

Software

You can use whatever NLP/ML software or resources you like. We will provide a small bit of starter code to use [CRFSuite](http://www.chokkan.org/software/crfsuite/), a software package that does first-order CRF sequence tagging. It requires you to run your own script to extract observation features as a text file. Then you tell it to train and predict with these feature files.

We provide [starter code](starter_code.zip) for

- A very simple feature extractor.

- The evaluation script to evaluate accuracy locally. It computes precision and recall of name spans.

More details on the [milestone page](milestone.html).

Deliverables

There are several points to turn things in.

1. (Due Sunday, Nov 16) [Milestone](milestone.html): create a very bare-bones tagger for the NER task, with the training and development set, and submit predictions to the Kaggle board, plus submit a short document about it.
2. (DATE TBD, week of 12/1) Final test set evaluation: we will provide the final test set, and it will be blind --- no labels given! You have to run your system to create predictions. **Only submit predictions from your system.**
3. (DATE TBD, few days after that) Final project reports and code turn-ins are due.

Grading

The project is, in total, 20% of your grade.
The milestone is 5%. It's designed to help the success of the final result.
The rest is derived from both:

1. System building.
- Required: core system with lexical, character affix, and shape features, plus positional offset versions.
- Optional: One or more possible extensions of new features to try. (At least one is required if you are working in a group of two.)

2. Analysis and exploration in your project report, such as analyzing the model weights, or doing ablation tests.

Finally, there will be extra credit for the top performers on the final test set.
Extra credit will also be available for doing additional extensions, or doing some sort of additional analysis project.

Started: 4:27 pm, Thursday 6 November 2014 UTC
Ended: 11:59 pm, Monday 15 December 2014 UTC (39 total days)
Points: this competition did not award ranking points
Tiers: this competition did not count towards tiers

© 2015 Kaggle Inc