Named Entity Resolution

Rodrigo Agerri

UPV/EHU

General Objective

The disappearance of York University chef Claudia Lawrence is now being treated as suspected murder, North Yorkshire Police said. However detectives said they had not found any proof that the 35-year-old, who went missing on 18 March, was dead. Her father Peter Lawrence made a direct appeal to his daughter to contact him five weeks after she disappeared. His plea came at a news conference held shortly after a 10,000 reward was offered to help find Miss Lawrence. Crimestoppers said the sum they were offering was significantly higher than usual because of public interest in the case.

News

[[The disappearance of [York University chef Claudia Lawrence]] is now being treated as suspected murder, North Yorkshire Police said. However detectives said they had not found any proof that the 35-year-old, who went missing on 18 March, was dead. [Her father Peter Lawrence] made a direct appeal to his daughter to contact him five weeks after she disappeared. His plea came at a news conference held shortly after a 10,000 reward was offered to help find Miss Lawrence. Crimestoppers said the sum they were offering was significantly higher than usual because of public interest in the case.

Different Domains

The elegant Hotel Bologna is set in the historic centre of Pisa, 5 minutes' walk from the train station. It offers free Wi-Fi and a shuttle to the airport.

All rooms have minibar, and flat-screen TV with satellite channels. They feature floor-to-ceiling windows, wooden floors and independent heating.

Breakfast at the Bologna is a varied buffet including fruit, cold meats, and fresh bread and cakes. It can be enjoyed on the courtyard terrace, or provided to take away for early check-outs. The bar offers free tastings of traditional Tuscan products every evening.

The Campo dei Miracoli and the famous Leaning Tower are a 12-minute walk from the hotel. The highway that links Pisa with Florence and Livorno is 3km away.

Tourist Domain

[The elegant Hotel Bologna] is set in [the historic centre of Pisa], 5 minutes' walk from the train station. It offers free Wi-Fi and a shuttle to the airport.

[All rooms] have minibar, and flat-screen TV with satellite channels. [They] feature floor-to-ceiling windows, wooden floors and independent heating.

[Breakfast] at [the Bologna] is a varied buffet including fruit, cold meats, and fresh bread and cakes. [It] can be enjoyed on the courtyard terrace, or provided to take away for early check-outs.

The bar offers free tastings of traditional Tuscan products every evening.

The Campo dei Miracoli and the famous Leaning Tower of Pisa are a 12-minute walk from [the hotel]. The highway that links Pisa with Florence and Livorno is 3km away.

Named Entity Resolution

- Named Entity Recognition and Classification (NERC)
  - Robust, coarse-grained
- Coreference Resolution
  - Robust, across domains and languages
- Named Entity Disambiguation (NED)
  - DPpedia Spotlight
  - Post-processing
- Domain Adaptation unsolved!!

OpeNER Challenge

- **Two domains + six languages = 12 systems in 2 years!!**
- **Open/Free** software easy to **commercialize**
- **No NERC** already available for every language and domain
- **No coreference system** available for every language and domain
- Lack of readily available training data
- Strategy: one solution, one tool, one methodology

General Approach

- **Named Entity Recognition and Classification (NERC)**
  - Statistical and external knowledge
- **Coreference Resolution**
  - Rule-based but leveraging linguistic annotations from statistical NLP (http://ixa2.si.ehu.es/ixa-pipes)
- **Named Entity Disambiguation (NED)**
  - Basada en DBpedia Spotlight
- **Semi-supervised domain adaptation**

Evaluation

- **Selected:** true positives + false positives
- **Target:** true positives + false negatives
- **Precision:** true positive / selected
- **Recall:** true positives / target
- **F measure**: 2 * precision * recall / (precision + recall)

Examples

- http://dbpedia-spotlight.github.io/demo/

- http://nlp.stanford.edu:8080/corenlp/

- http://www.newsreader-project.eu/results/demos/

- http://svn.ask.it.usyd.edu.au/trac/candc/wiki/Demo

- http://demo.languagecomputer.com/cicerolite/

- http://www.opener-project.eu/

- http://www.olery.com/reputation/

- **Hackathons!!**

Hackathons

- A day full of real users hacking together real products in an informal and fun atmosphere.

- Multifunctional:
    - Promotion
    - Evaluation
    - Focus
    - Team Spirit

Hackathons

- Open up registration on Facebook, Meetup.com, LinkedIn
- Use all channels you can find to tell about the event
    - 09:00 - Opening Presentation & explanation of the technology
    - 09:45 – Shape the groups
    - 10:45 – Initial Idea presentation and group reorientation
    - 13:00 - Lunch best served as buffet or finger food
    - 16:30 – Presentations of work done
    - 17:30 – Voting and Prizes

- Prices don't have to be huge, it's symbolic
- T-shirts give a nice team feeling
- Build upon previous hackathons

Hackathon Review

- **Demos presentation**

- **JapaNER**:
  - Implementation of Japanese toolkit (komoji)
  - Added Japanese tokenization and pos-tagging within a day
  - Managed to plug resulting NAF straight into the Opinion Detector Basic and obtain reasonable results

- **III Hackathon Video**