

Controlling Robots with Tweets

10 July 2019

© 2019 The University of Sheffield

This material is licenced under the Creative Commons

Attribution-NonCommercial-ShareAlike Licence

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

Acknowledgement

This work is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654024, the SoBigData project.

<http://sobigdata.eu/>



Unpacking files now (to save time)

Instructions in more detail—first section of the handout

- Make a new `twitter` folder on your Desktop
- Right-click on each of these files
 - `gate-developer-8.6.zip`
 - `headstart-2019.zip`

and *extract all* to the `Desktop\twitter` folder (this can take a few minutes).

- The handout shows what the destination folder should end up with

What is Natural Language Processing?

natural language the kind humans use with each other, as opposed to *formal languages* such as

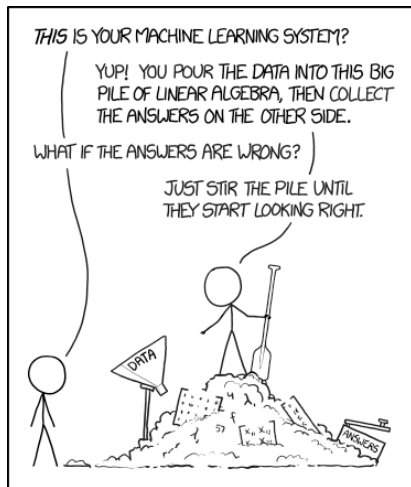
- programming languages
- mathematical languages (logic, algebra)

NLP using computer programs to analyse natural language

applications information retrieval, information extraction, machine translation, question answering, . . . controlling robots

Machine learning

- Requires a lot of training data
- Produces a *trained model* that is not usually human-readable
- Not transparent; hard to correct except by retraining



Rule-based systems

- Easy to get started without a lot of data
- You can see how the system works

Determiner? Adjective * Noun \longrightarrow NounPhrase (1)

the big dog, the dog, a dog, big dogs

Capitalized * Surname \longrightarrow Person (2)

Forename + Capitalized+ \longrightarrow Person (3)

Eric Bloom, Ignatius J Reilly

What is GATE?

- General Architecture for Text Engineering
- open-source software, widely used
- runs in Java on Windows, Mac OS, GNU/Linux
- GUI environment for developing applications
- applications can be run headlessly (embedded, in a server, “behind the scenes”) in various ways
- easy to get started with rule-based systems (but it also has machine learning)

ANNIE demonstration

...

How does this work?

- ANNIE has *gazetteers* (lists) of personal names (Fyodor, Gabe, Gabriel, Gahiji, Gaius...), company suffixes (PLC, PrC, Prp. Ltd., PT, Pty...), etc.
- and JAPE rules for combining them

```
( ({Lookup.majorType == person_male})+  
  ({Token.orth == upperInitial})+  
):match  
-->  
:match.Person = {gender = male}
```

which means wherever you find *one or more known male forenames followed by one or more unknown but capitalized words*, create a *Person* annotation with the feature *gender=male*.

The Robot application

- Follow the handout instructions to run GATE, load the starter application, and set your robot's name.
- Create a document that contains the phrase “turn left”. The easiest way to do this is to create a document without setting any parameters, then type “turn left” in the document content.
- Create a corpus for the document and run the application over it. The robot should turn slightly left.
- Now look at the gazetteer and JAPE grammar to see how it works. The right-hand side of a JAPE rule can run Java code as well as creating annotations.

The Robot application

- The gazetteer creates *Lookup* annotations on the document with *majorType* features.
- The JAPE rules match patterns of annotations in the document; usually they create more annotations, but they can also run Java code.
- The gazetteer can be edited in the GATE GUI. Be sure to right-click on it and *Save and Reinitialize* to keep the changes.
- It can also be edited in jEdit—in this case use *Reinitialise* to load the changes from the files.
- Use JEdit to edit the JAPE grammar; *Reinitialize* it in the GATE GUI to load the changed file.

GATE Cloud Twitter Collector

- Normally runs on GATE Cloud¹ to collect tweets for later analysis
- Success stories: election analysis, abuse, climate change
- We are using a version modified to run a local GATE application: every tweet that it collects is turned into a GATE document and run through the `application.xgapp`

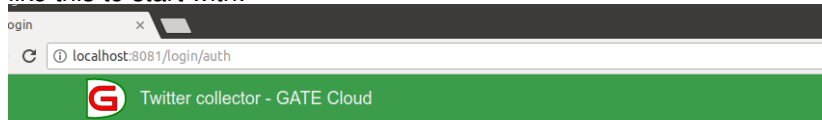
¹<https://cloud.gate.ac.uk/>

The Twitter Collector takes a while to start up, so you probably want to develop your gazetteers and rules extensively in the GATE GUI before using it. When you are ready, go the directory where you unpacked it and run this command:

```
java -jar twitter.war
```

GATE Cloud Twitter Collector

The web interface is at `http://localhost:8081/` and looks like this to start with:



admin →

gatecloud →

Please log into your Twitter collector

 Remember me