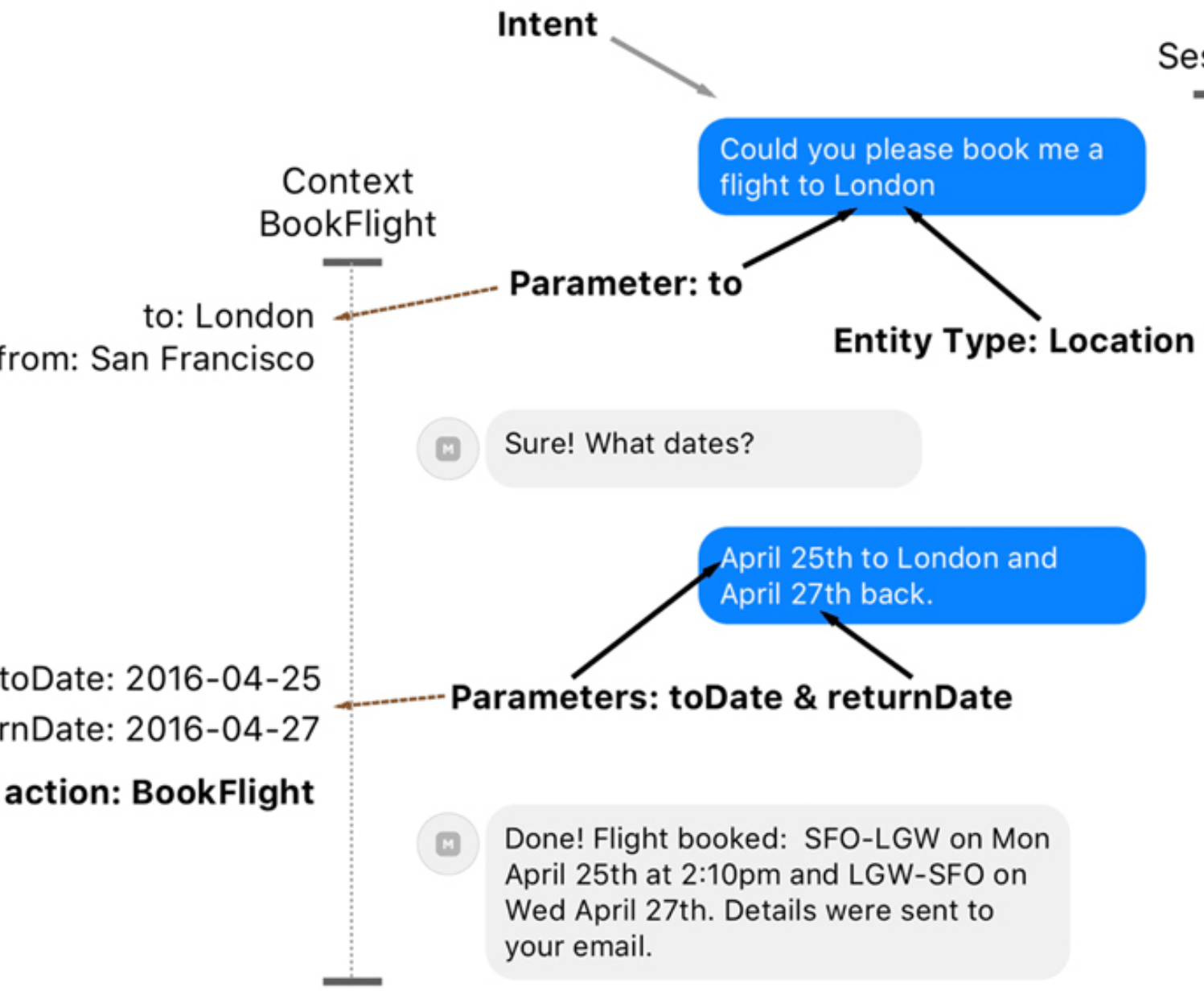
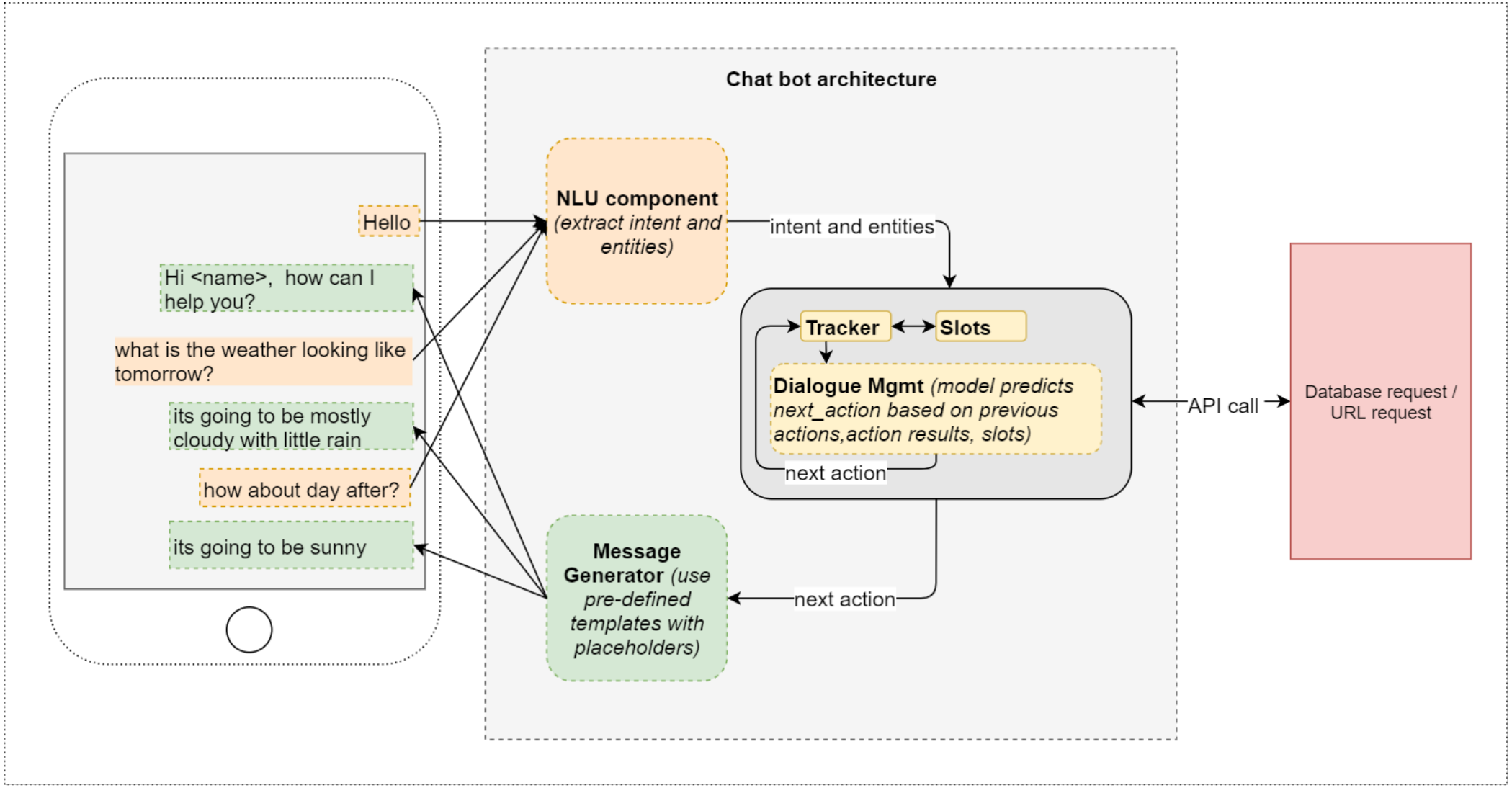
Rule-Based NER + Chatbot

https://github.com/oksanatkach/UCU_rule_based_2020.git



1. Extract reference data.
2. Preprocess raw text data.
3. Recognize the date.
4. Find the location.
5. Find reference: from or to?
6. Create a bot.
7. Create a booking flight dialog.
8. Catch exceptions.

Natural Language Understanding as a part of chatbot structure



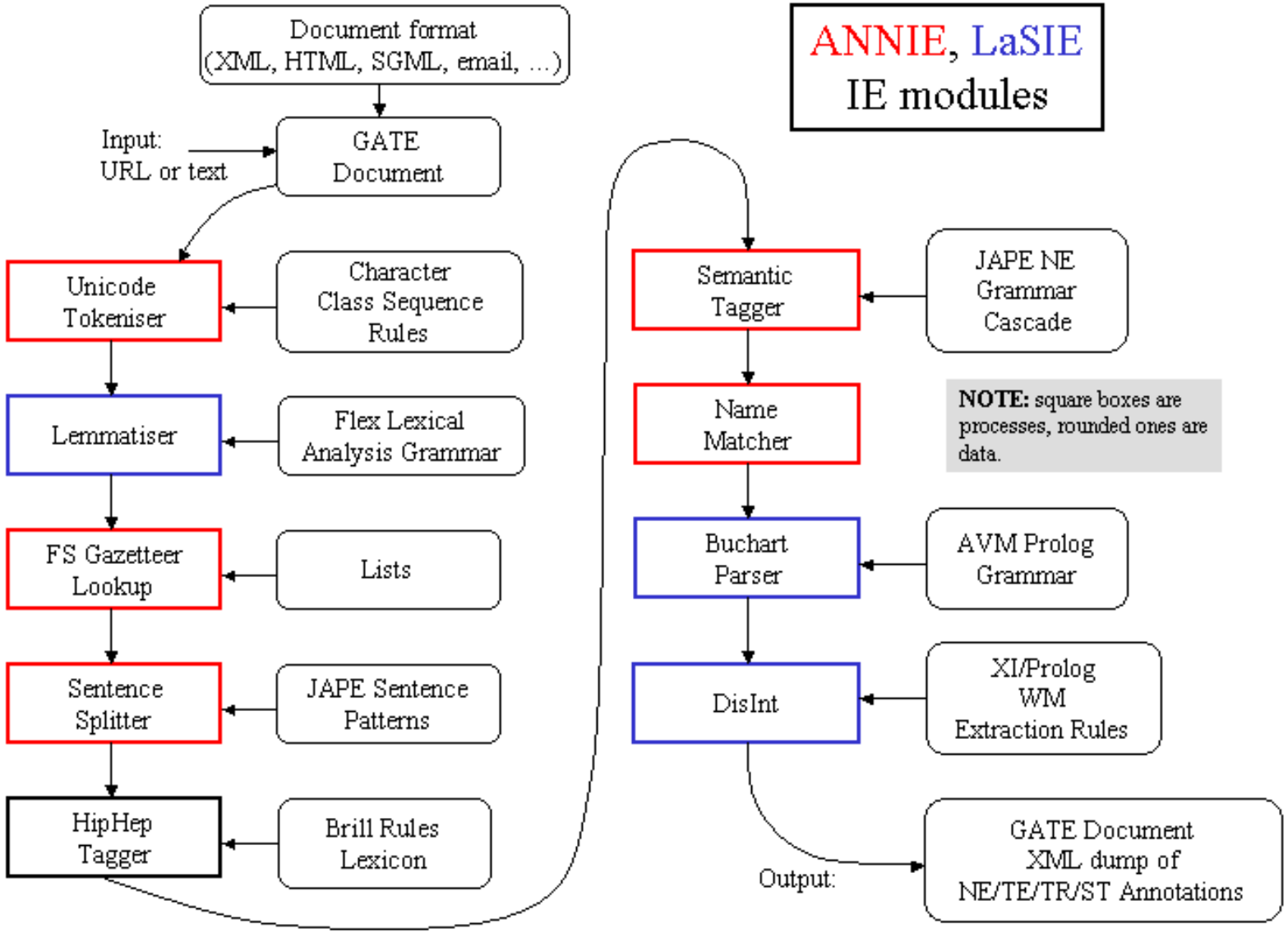
What is NER

In a warm morning of July 15th, Brazil - André, Paulo, Peterson, Ana, Vinicius, Margarito and Rodrigo were all chatting about NER in a Google Hangouts.
Watch here!

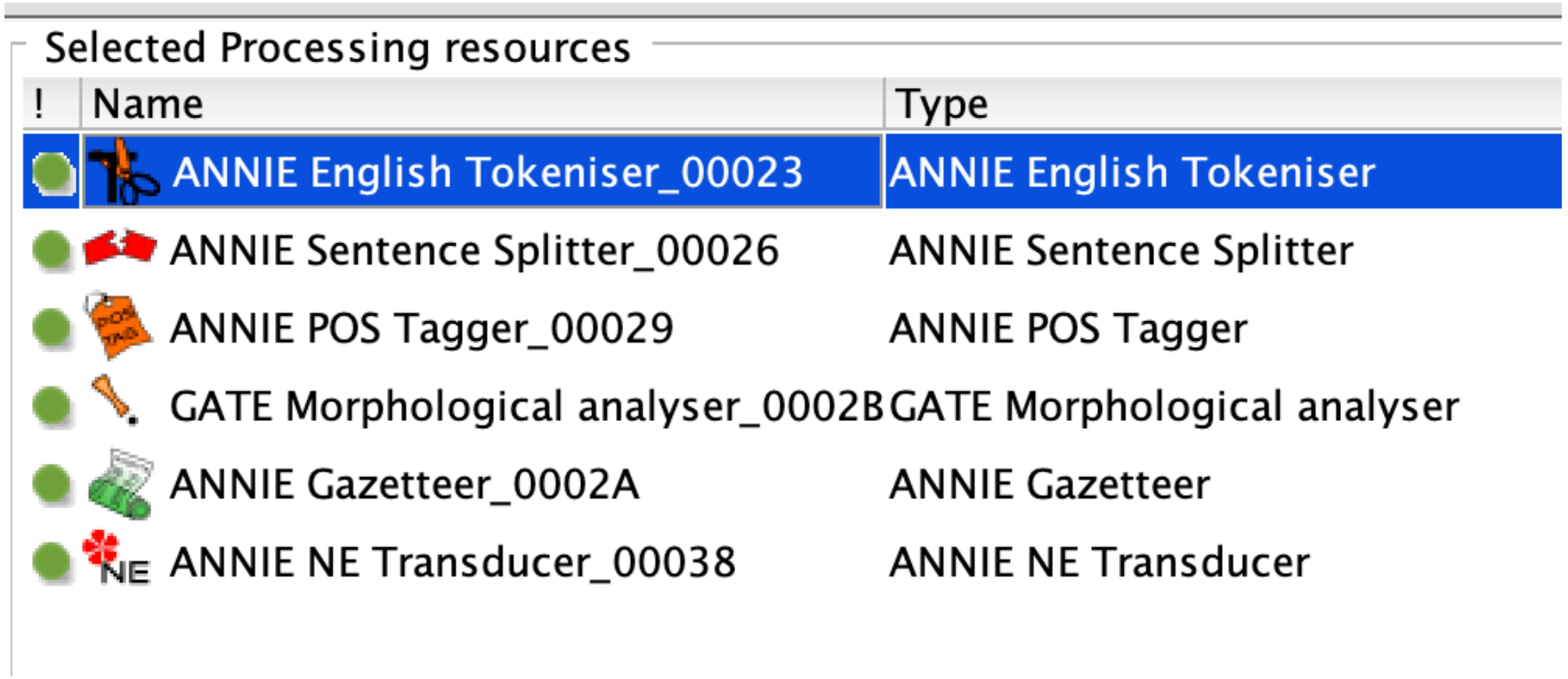
Potential tags:

- LOCATION
- ORGANIZATION
- DATE
- MONEY
- PERSON
- PERCENT
- TIME

Rule-Based NER Pipeline



GATE as a Classic Example of Rule-Based NER



What is a gazetteer

a list of known named entities or entity keys used for dictionary-based NLP

Companies	Company keys
20th Century Fox	company
Allianz	Inc.
Associated Press	Incorporated
AT&T	CEO
Benz	ministry
Commerzbank	bank
EasyJet	airline
Heinz	office

Facts in DBpedia and SPARQL

SPARQL is a script for making queries to knowledge bases.

DBpedia is a large open source knowledge base. It has its own classes of objects, subjects an attributes defined as resources.

A resource is compiled of **facts** such as:

subject	verb	object
(Barack_Obama	spouse	Michelle_Obama)

These can be called using the ontology query language, SPARQL.

To test SPARQL requests:
https://dbpedia.org/sparql

Queries in DBpedia

1) define a PREFIX:

PREFIX dbo: <http://dbpedia.org/ontology/>

2) use SQL-style syntax (like data base requests) to start the query:

SELECT ?
WHERE {}

3) use facts three-tuple syntax to specify what to extract from the ontology:

prefix:object_class prefix:label_class ?

This will return the subject.

SPARQL Example

All cities of a particular country

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbr: <http://dbpedia.org/resource/>
SELECT ?s WHERE {
 ?s a dbo:City ;
    dbo:country dbr:Ukraine
}
```

SPARQL Example

All cities with population more than 100000

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbpprop: <http://dbpedia.org/property/>

SELECT ?name WHERE {
 ?city rdf:type dbo:City ;
       rdfs:label ?enName ;
       dbo:populationTotal ?population

FILTER (langMatches(lang(?enName), "en"))
BIND (str(?enName) AS ?name)
FILTER (?population>100000)
}
```

SPARQL Example

All English names of all countries

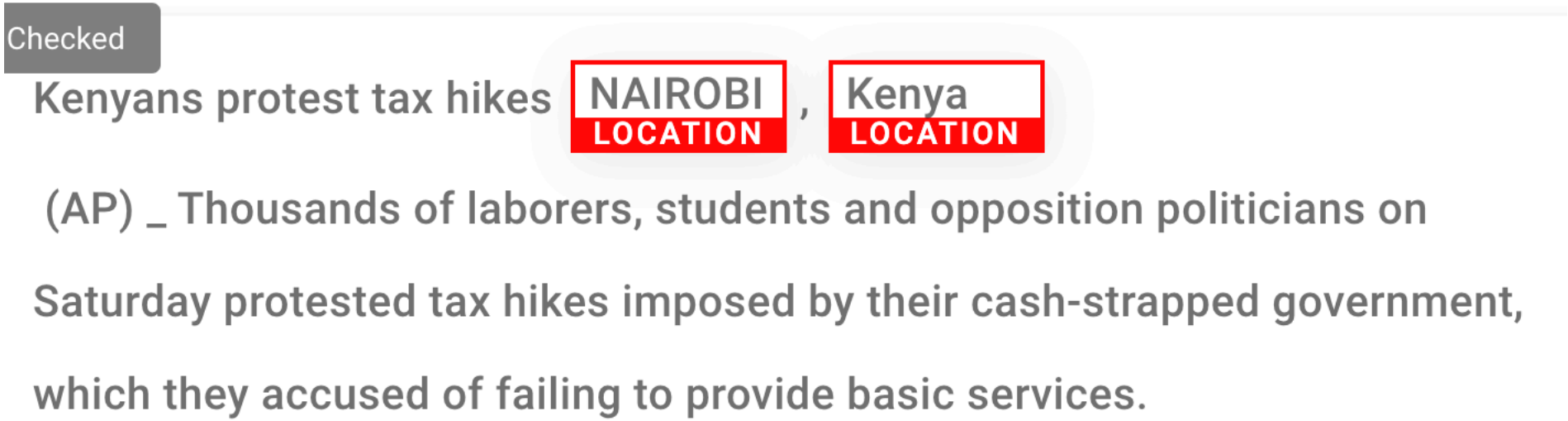
```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbpprop: <http://dbpedia.org/property/>

SELECT ?name WHERE {
 ?city rdf:type dbo:Country ;
       rdfs:label ?enName ;

FILTER (langMatches(lang(?enName), "en"))
BIND (str(?enName) AS ?name)
}
```

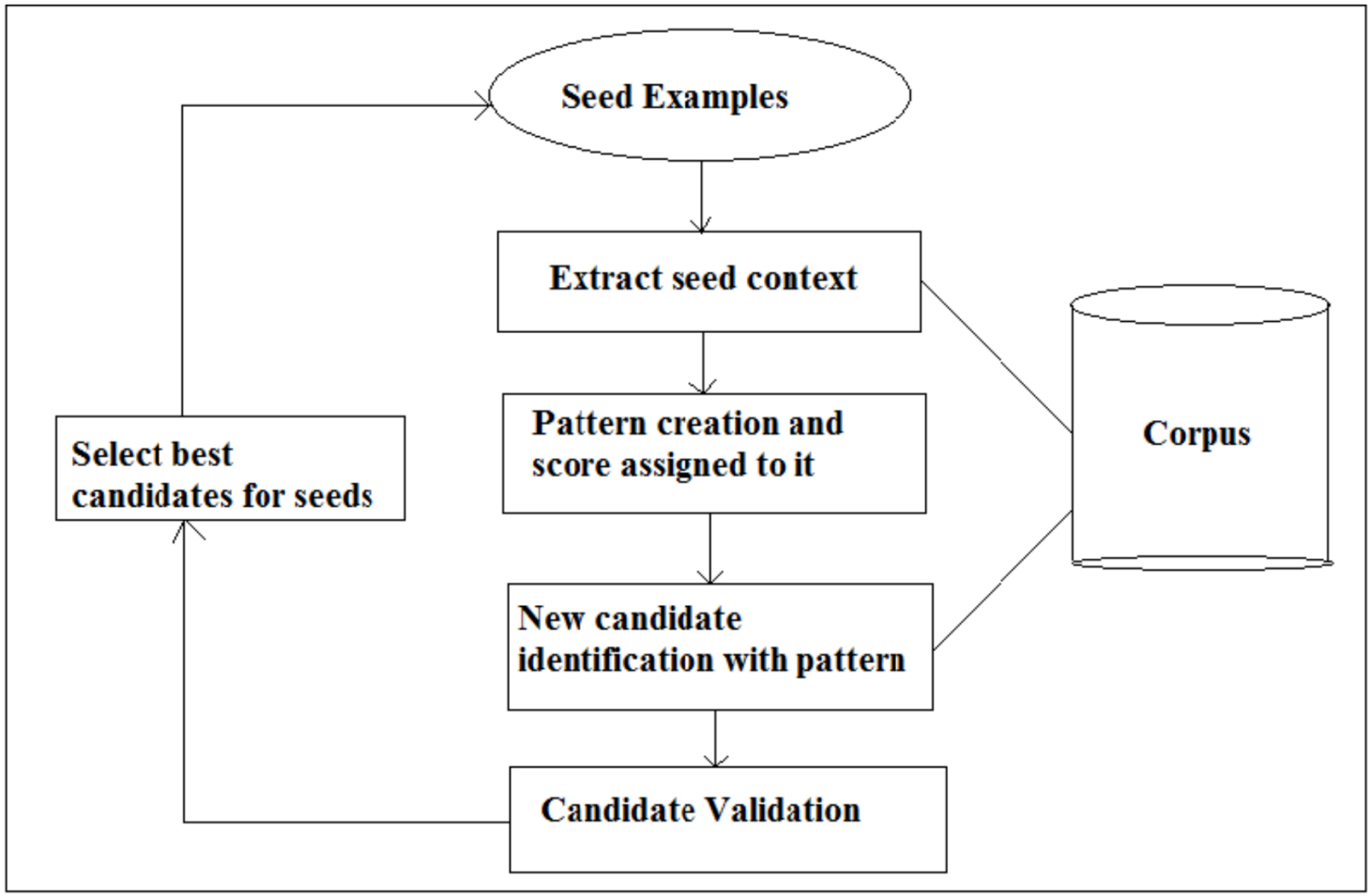
Annotation

how it happens



Check out the Doccano annotation tool:
http://ec2-3-15-25-86.us-east-2.compute.amazonaws.com/

Semi-supervised Bootstrapping



https://arxiv.org/pdf/1511.06833.pdf