Master on Artificial Intelligence

Introduction to Human Language Technologies
9. Coreference

Natural Language Research Group

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona

FIB

Course 2018/19

Outline

Natural
Language
Research
Group

Coreference in
CoreNLP

Coreference in CoreNLP

Natural
Language
Research
Group

Coreference in
CoreNLP
Example

Example:

```python
def show(d):
    crs = d['corefs']
    for k, num in zip(crs, range(1, len(crs) + 1)):
        print('Chain', num, ':')
        l = crs[k]
        for e in l:
            print('\t',e['text'],e['gender'], e['number'],e['isRepresentativeMention'],e['type'])
```

```python
from nltk.parse import CoreNLPParser
tagger = CoreNLPParser(url='http://localhost:9000')
tagger.parser_annotator='tokenize,ssplit,pos,lemma,ner,depparse,coref'
```

```python
text = "Barack Obama was born in Hawaii.  He is the president. Obama was elected in 2008."
res = tagger.api_call(text)
show(res)
```

```
Chain 1 :
         Barack Obama MALE SINGULAR True PROPER
         He MALE SINGULAR False PRONOMINAL
         Obama MALE SINGULAR False PROPER
```

Natural
Language
Research
Group

Coreference in
CoreNLP
Example

Example:

```
In [4]:  1  text = "He is the president. Barack Obama was born in Hawaii. He was elected in 2008."
         2  res2 = tagger.api_call(text)
         3  show(res2)
```

```
Chain 1 :
        He MALE SINGULAR False PRONOMINAL
        Barack Obama MALE SINGULAR True PROPER
        He MALE SINGULAR False PRONOMINAL
```

```
In [5]:  1  text = "Barack Obama was born in Hawaii. Michelle is his wife. He was elected in 2008."
         2  res3 = tagger.api_call(text)
         3  show(res3)
```

```
Chain 1 :
        Michelle FEMALE SINGULAR True PROPER
        his MALE SINGULAR False PRONOMINAL
        He MALE SINGULAR False PRONOMINAL
```

Coreference in CoreNLP

Output of coreNLP:

Natural
Language
Research
Group

Coreference in
CoreNLP

Example

```
In [8]:   1  res.keys()

Out[8]: dict_keys(['sentences', 'corefs'])

In [10]:  1  res['corefs']

Out[10]: {'3': [{'animacy': 'ANIMATE',
            'endIndex': 3,
            'gender': 'MALE',
            'headIndex': 2,
            'id': 0,
            'isRepresentativeMention': True,
            'number': 'SINGULAR',
            'position': [1, 1],
            'sentNum': 1,
            'startIndex': 1,
            'text': 'Barack Obama',
            'type': 'PROPER'},
           {'animacy': 'ANIMATE',
            'endIndex': 2,
            'gender': 'MALE',
            'headIndex': 1,
            'id': 3,
            'isRepresentativeMention': False,
            'number': 'SINGULAR',
            'position': [2, 2],
```

Mandatory Exercise

Natural
Language
Research
Group

Coreference in
CoreNLP
Mandatory
Exercise

Statement:

- Consider the first paragraph in *Alice's Adventures in Wonderland*, by *Lewis Carroll*:

  ```
  Alice was beginning to get very tired of sitting by
  her sister on the bank, and of having nothing to
  do:  once or twice she had peeped into the book her
  sister was reading, but it had no pictures or
  conversations in it, 'and what is the use of a
  book,' thought Alice 'without pictures or
  conversations?'
  ```

- It can be downloaded from:
  `http://www.gutenberg.org/files/11/11-0.txt`
- Apply the CoreNLP coreference solver to the previous paragraph
- Show the coreference chains. What do you think about them? Justify your answer.