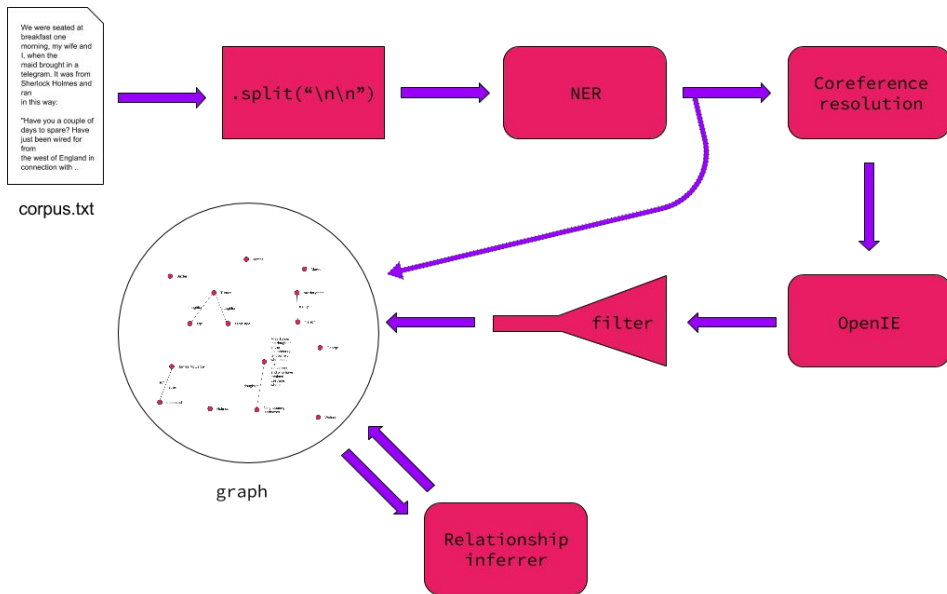# Inferring Relationships from a Corpus

Axel Larsson & Erik Gärtner

Goals

- Extract named entities from a corpus
- Identify relationships between the persons
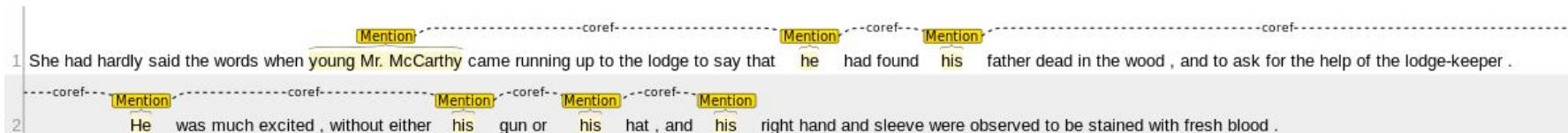- Infer more relationships based on extracted data

Building the Graph

Named Entities Extraction

___

- NER: locate and classify elements in texts into predefined categories; names of persons, locations, organisations etc.
- Stanford CoreNLP NER tags annotator
  - Uses a trained model to detect: names, places and organisations
- We filter for only person names

Resolving co-references

1 She had hardly said the words when young Mr. McCarthy came running up to the lodge to say that he had found his father dead in the wood , and to ask for the help of the lodge-keeper .

2 He was much excited , without either his gun or his hat , and his right hand and sleeve were observed to be stained with fresh blood .

- Mentions not using the primary name, such as:
  - he
  - the president
- Very slow process

Detecting Relationships

___
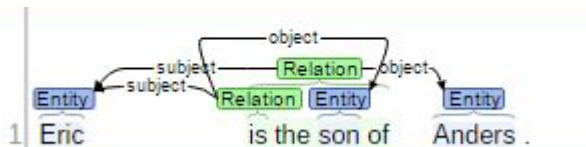
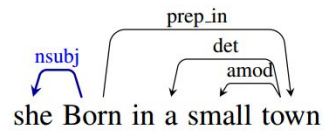- OpenIE triples (subject, relation, object)
- (Eric, is the son of, Anders)
- Stanford OpenIE
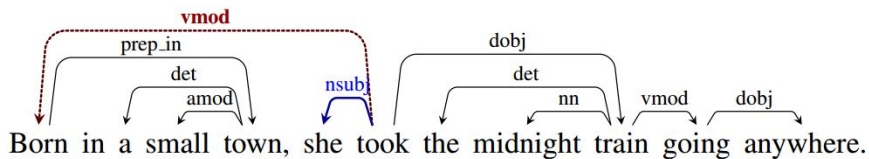
Filtering OpenIE Triples

———

- Wikidata - a free knowledge base
- List of properties on Person:s of type Relationship
  - father
  - mother
  - brother
  - sister
  - spouse
  - partner
  - child
  - stepfather
  - stepmother
  - relative
  - godparent

```
[
  {
    "title": "brother",
    "id": "P7",
    "description": "male sibling",
    "english": ["bro", "brother"],
    "swedish": ["broder", "bror", "brorsa"]
  },
  {
    "title": "father",
    "id": "P22",
    "description": "the male parent",
    "english": ["father", "dad", "daddy"],
    "swedish": ["far", "fader"]
  },
```

Inferring Relationships

- Rule-based engine
- Iterates on the graph
- Add new inferred relationships such as:
  - "Abel is the son of Adam"
  - Son -> Father

Experimental Setup

———

- Sherlock - The Boscombe Valley Mystery
  - ~9600 words
  - ~10 family relationships
  - ~3 minutes to extract relationships
  - Manually annotated for scores
  - Training + testing
- CoreNLP
  - Opensource
  - Cutting edge
- Scala

The Graph
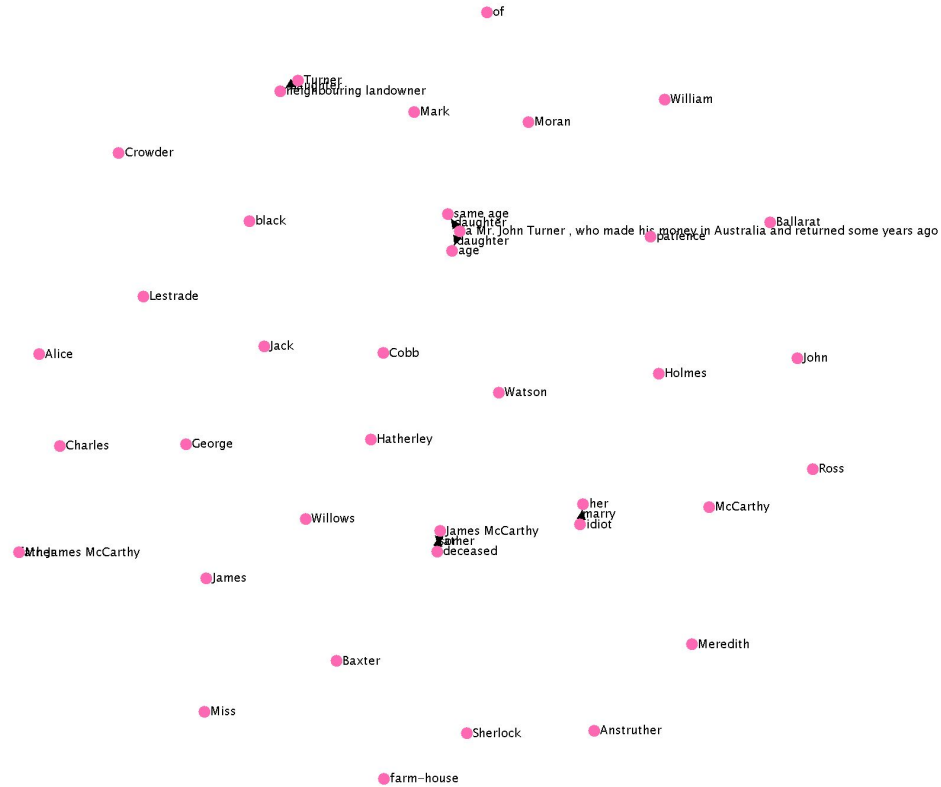
(idiot, marry, her)

— — —

This fellow is madly, insanely, in love with her, but some two years ago, when he was only a lad, and before he really knew her, for she had been away five years at a boarding-school, what does the idiot do but get into the clutches of a barmaid in Bristol and marry her at a registry office?

Results and Evaluation

———

- Managed to extract relationships from a novel
- Promising but further work needed
- Evaluation scores
  - True positives: 2
  - False positives: 3
  - False negatives: 7
  - Recall: 0.22
  - Precision: 0.40
  - F-score: 0.29

Future Work

———

- Improve relationship extraction; more important than NER.
- Add more languages
- Improve rules engine

Questions?