Fact Checking

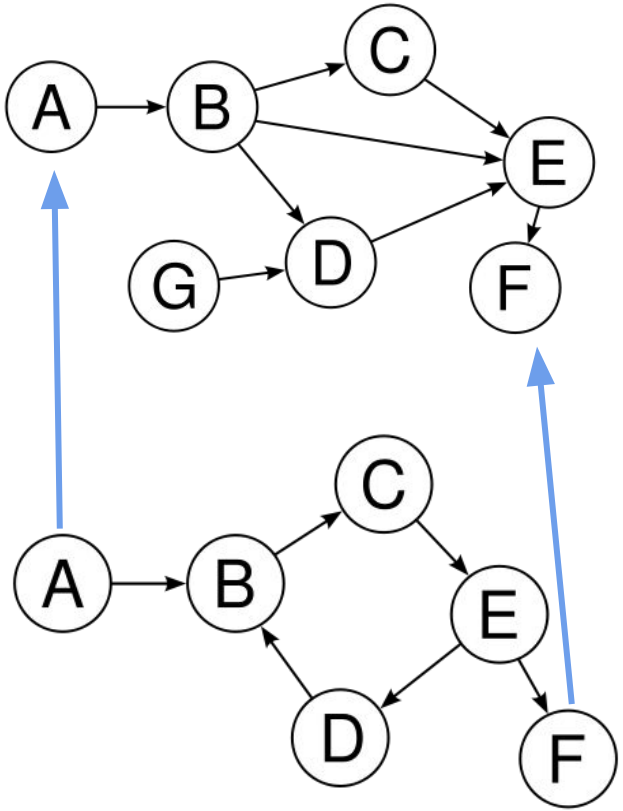
SNLP Miniproject

Confirmatio Ex Machina

Clemens Damke & Lukas Brandt

PADERBORN UNIVERSITY

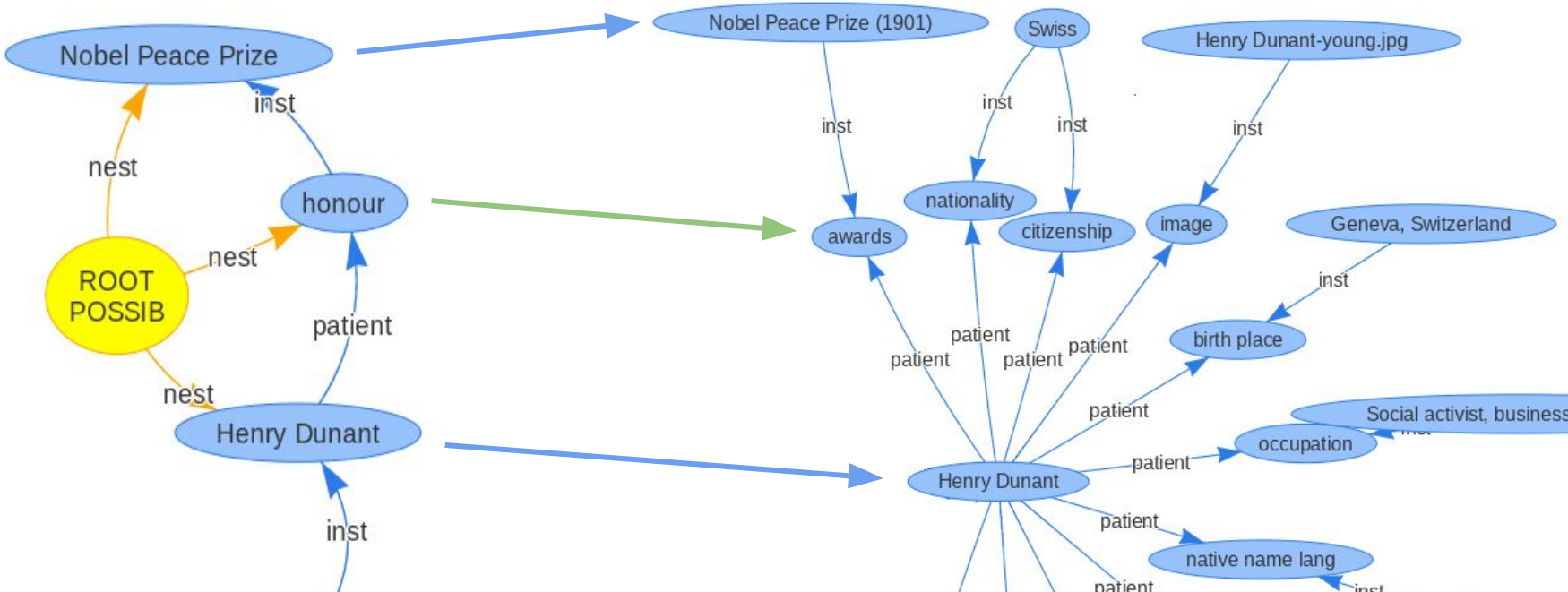
Approach



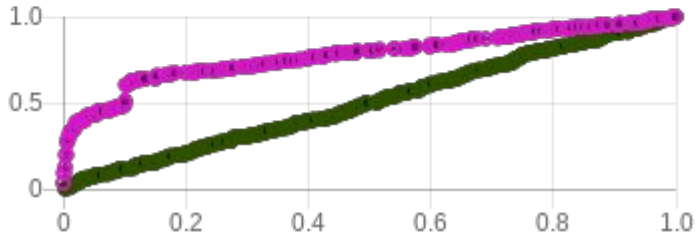
1. Construct a big Concept graph from the Wikipedia corpus and a small Concept graph from the fact
2. Find matching start and end points via the String similarity
3. Search for undirected paths between starts and ends and compare the connecting edges and nodes as well
4. Calculate Score based on the existence/absence of starts and ends and the results of the string similarity

Approach

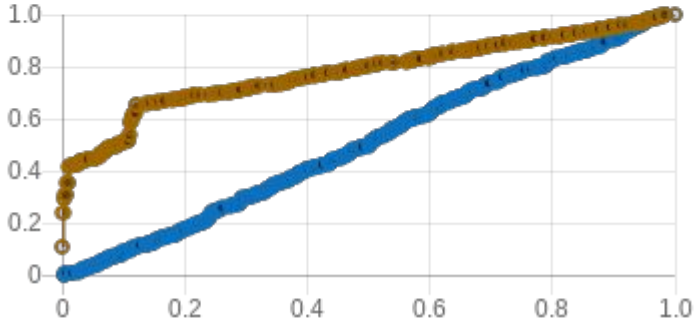
Nobel Peace Prize is Henry Dunant's honour. (True)



Evaluation Results



Train: 0.775



Test: 0.779

Discussion of the approach

Strengths

- Concept graphs matching allows to include facts from several corpus documents for checking and scoring
- Matching of undirected paths allows fact checking with loosely coupled text appearances
- Flexible in regard to the structure of the given fact

Weaknesses

- Only compares grammatical and syntactical similarity and not semantics
- Path matching for all potentially relevant paths is infeasible because subgraph matching is NP-hard

Possible Improvements of the approach

- Add more domain specific corpus sources other than Wikipedia
- Write further transformation rules to simplify the resulting graphs for better matchings
- Integrate Word2Vec to include comparison of paths via their semantics
- Experiment with several threshold parameters
  - Max length
  - Amount of considered paths
  - Weights of considered paths

Code and Documentation

Repository:

https://github.com/ConfirmatioExMachina

Documentation:

- Code commentary
- Explenations in README.md
- docs-Repository with Presentations and generated documentation with Codox

Task distribution

Clemens		Lukas
	Approach planning	
CoreNLP		Wikipedia Corpus
Elasticsearch		Word2Vec
Neo4j		
	Presentation and Documentation	