

# Team Inigo Montoya

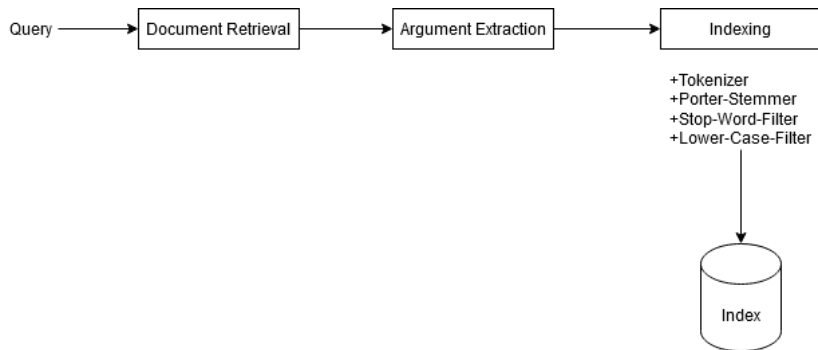
## Approach Presentation

Johannes Huck

Martin-Luther Universität Halle-Wittenberg

Touché Lab on Argument Retrieval at CLEF 2020  
September 23, 2020

# Overview



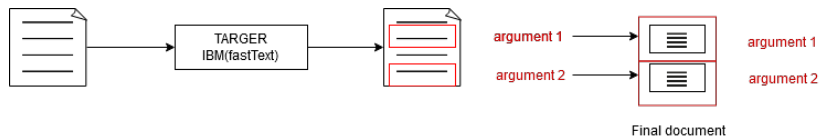
## Preprocessing web pages

- Sending queries from the provided file to *ChatNoir*
- Using *BoilerPipe* to extract main body of retrieved web pages
- Sending cleaned web pages to *TARGER* to mine arguments



## Argument Extraction

- *TARGER* uses classifier *IBM(fastText)*
- Output: premise and claim encoded using the IOB-format
- Arguments: premise and claim together
- For each web page: one document with all the arguments



## Index Construction

- Index will be constructed over the aforementioned documents
- Index contains the mined arguments and the ClueWeb12 ID
- Index extended with the Stemming-Analyzer by *whoosh*
  - Tokenizer
  - Porter-Stemmer
  - Stop-Word-Filter
  - Lower-Case-Filter

## Searching and Ranking

- Ranking function: Okapi BM25
  - $k1=1.2$
  - $b=0.75$
- Disjunctive OR-query will be searched on the index
- Higher rank: documents with more words from the query
- Approach retrieves first 20 results
- Web pages with the corresponding IDs will be shown

## Searching and Ranking

- Ranking function: Okapi BM25
  - $k1=1.2$
  - $b=0.75$
- Disjunctive OR-query will be searched on the index
- Higher rank: documents with more words from the query
- Approach retrieves first 20 results
- Web pages with the corresponding IDs will be shown

Thank you for your attention.



# An Open-Domain Web Search Engine for Answering Comparative Questions

Lightning Talk for Touché at CLEF 2020 - Task 2





# “Is X better than Y for Z?”

→ State-of-the-art search engines do not perform very well on those kind of questions

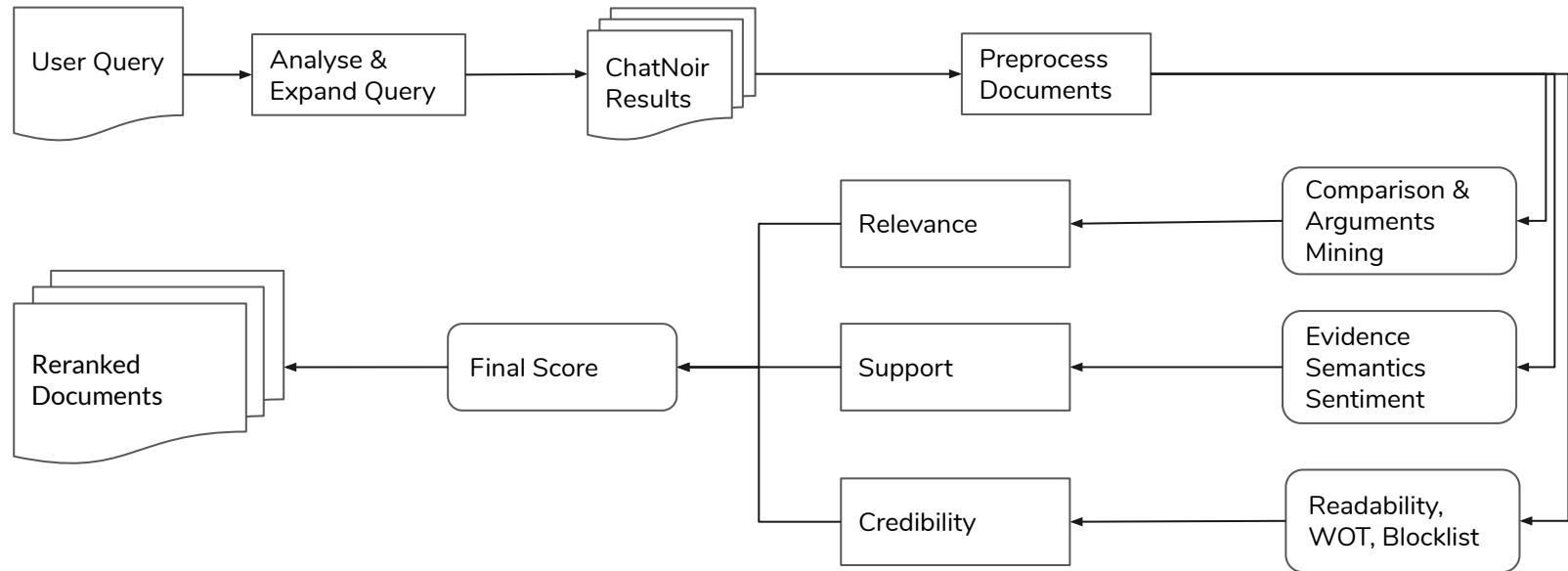
## Challenges:

- Extract the comparative relation included in the user query
- Identify unique and diverse arguments in web documents
- Verify credibility of the source

**Goal:** Providing documents from credible sources including diverse arguments



# Comparison Retrieval Model





**Thank you!**

**Tinsaye Abye**

ta83jyga@studserv.uni-leipzig.de

**Tilmann Sager**

ts99nami@studserv.uni-leipzig.de

**Anna Juliane Triebel**

wir13ljt@studserv.uni-leipzig.de



# Evaluation

- nDCG@5 score of 0.580
- Still not usable for production: one query needs ~20 minutes on a middle class desktop
- Combining different techniques seems to be promising