



WikiFactFind: Semi-automated fact-checking based on Wikipedia

Mykola Trokhymovych

Ukrainian Catholic University
trokhymovych@ucu.edu.ua

Diego Saez Trumper

Wikimedia Foundation
diego@wikimedia.org

Introduction. Research questions

RQ1: How do manual search strategies impact the fact-checking process?

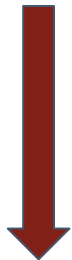
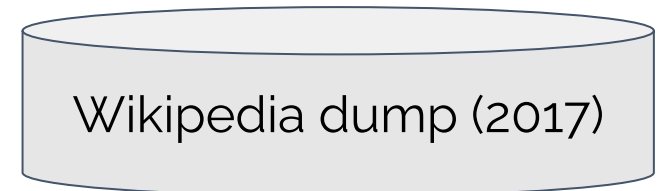
RQ2: Does the claim label influence the results of evidence search?

RQ3: What is the relation of Wikipedia article quality and SERP results?

Data preparation

Original FEVER. Data sample

```
{"id": 75397,  
  "verifiable": "VERIFIABLE",  
  "label": "SUPPORTS",  
  "claim": "Nikolaj Coster-Waldau worked with the Fox Broadcasting Company.",  
  "evidence": [[[92206, 104971, "Nikolaj_Coster-Waldau", 7],  
                [92206, 104971, "Fox_Broadcasting_Company", 0]]]}
```



- Leave only SUPPORTS (S), REFUTES (R) labeled samples
- Actualize article names (approx. 7.4% changed)

Customized FEVER data sample used for search

Claim	Evidence Articles	Label
Nikolaj Coster-Waldau worked with the Fox Broadcasting Company.	Fox_Broadcasting_Company, Nikolaj_Coster-Waldau	SUPPORTS
Hermit crabs are arachnids.	Arachnid, Hermit_crab, Decapoda	REFUTES
There is a capital called Mogadishu.	Mogadishu	SUPPORTS

Metrics

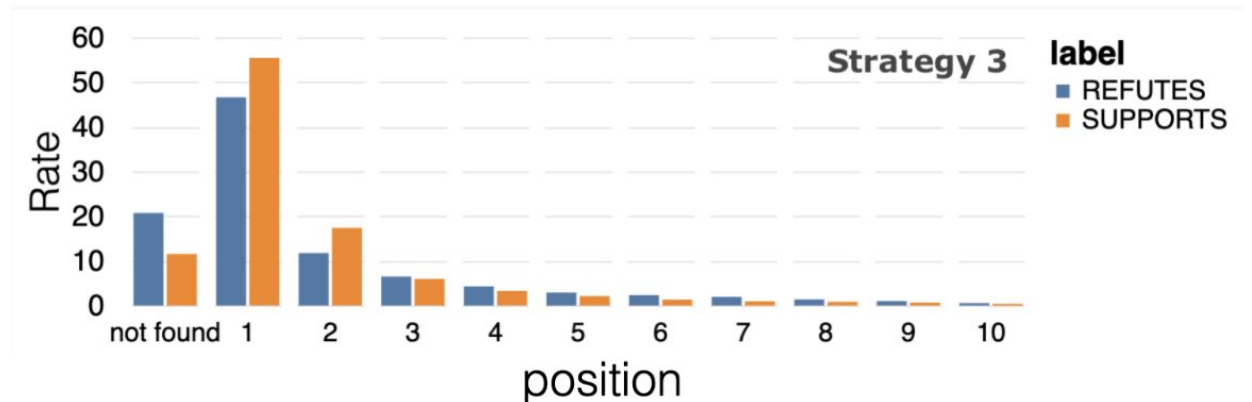
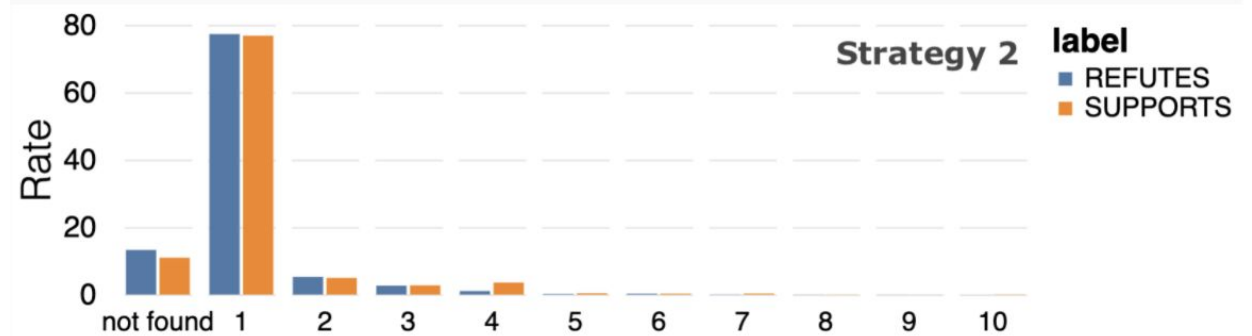
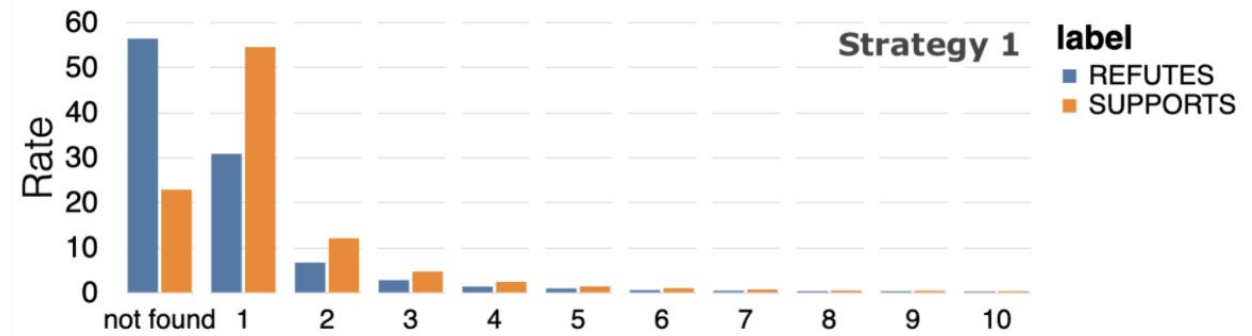
Evaluation criterias:

- Rate of found items (**RFI**) for top-10 results (equal to Recall@10)
- Rate of correctly placed items on the first position (**RCPI**)
- Distribution of desired evidence on top-10 positions of SERP

Comparing performance for different strategies

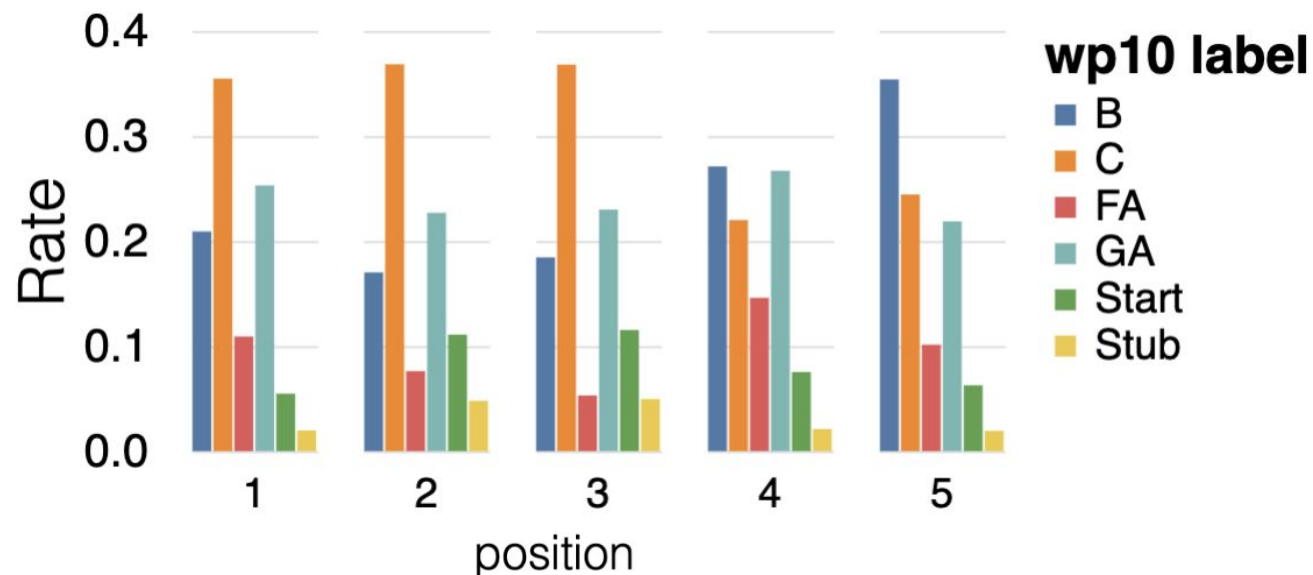
Strategy	RFI (train/test)	RCPI (train/test)
Strategy 1. Using raw Wikipedia search	0.681 / 0.539	0.705 / 0.773
Strategy 2. Enhanced Wikipedia Search	0.885 / 0.827	0.870 / 0.847
Strategy 3. Google search engine **	0.843	0.749

** Not fully comparable to other strategies due to differences in data



Influence of evidence quality

- Using Objective Revision Evaluation Service (**ORES**) to evaluate quality of Wikipedia articles with **WP10** model
- Calculate the scores for specific page revision that was up-to-date for the time of Wikipedia dump used in FEVER
- Analyzed the WP10 label across SERP position distribution



Efficient facts search tool

- Enhance the data with ORES features
- Train Learning-to-rank (LTR) model for fact checking search
- Results: increased Recall@1 from **0.847** to **0.875**



Conclusions

- Discovered that strategy selection has a significant impact on recall of search.
- Found out that searching for sources to refute incorrect claims can be more complicated than looking for correct statement evidence.
- Observed how the page quality differs across positions in search
- Built basic learning to rank model that shows that using page quality features can increase the recall for first positions, increasing the chance to be observed.

Thank you for attention

Contact:

