ACL 2025 VIENNA

Task-Oriented Automatic Fact-Checking with Frame-Semantics

Jacob Devasier, Rishabh Mediratta, Akshith Putta, Chengkai Li

IDIR Department of Computer Science and Engineering
The University of Texas at Arlington
https://idir.uta.edu/

GitHub

Overview

Problem

- **Misinformation** poses serious **threats** to informed decision-making
- Most research relies on **pre-curated data**, e.g., Wikipedia tables
- Previous work focuses on **low-volume tables** (<100 rows per table)
- LLM-based fact-checking methods often **lack interpretability**
- Different claims require **varying fact-checking processes**

Contributions

- Novel **task-oriented fact-checking paradigm** using **frame semantics**
- Two **high-volume structured** datasets with annotated **real-world claims** from reputable fact-checking organizations
- **FE-based retrieval** improves performance and **interpretability**
- Publicly released **demo**, **code**, and **datasets**

Case Studies

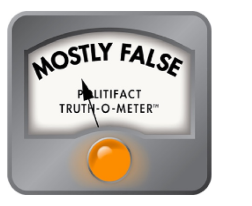
Case Study 1: Voting Records

Demo

Datasets (congress.gov):

- **342,466 Bills** from 1973-2023
- **7,195,798 Votes** since 1989
- **4,230** Avg. votes per member
- **79 Fact-checks** annotated with relevant bills and votes

While serving in Congress, Transportation Secretary Sean Duffy “voted against upgrading air traffic control systems.”

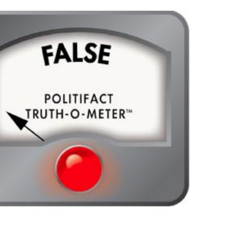


Case Study 2: Country Statistics (OECD)

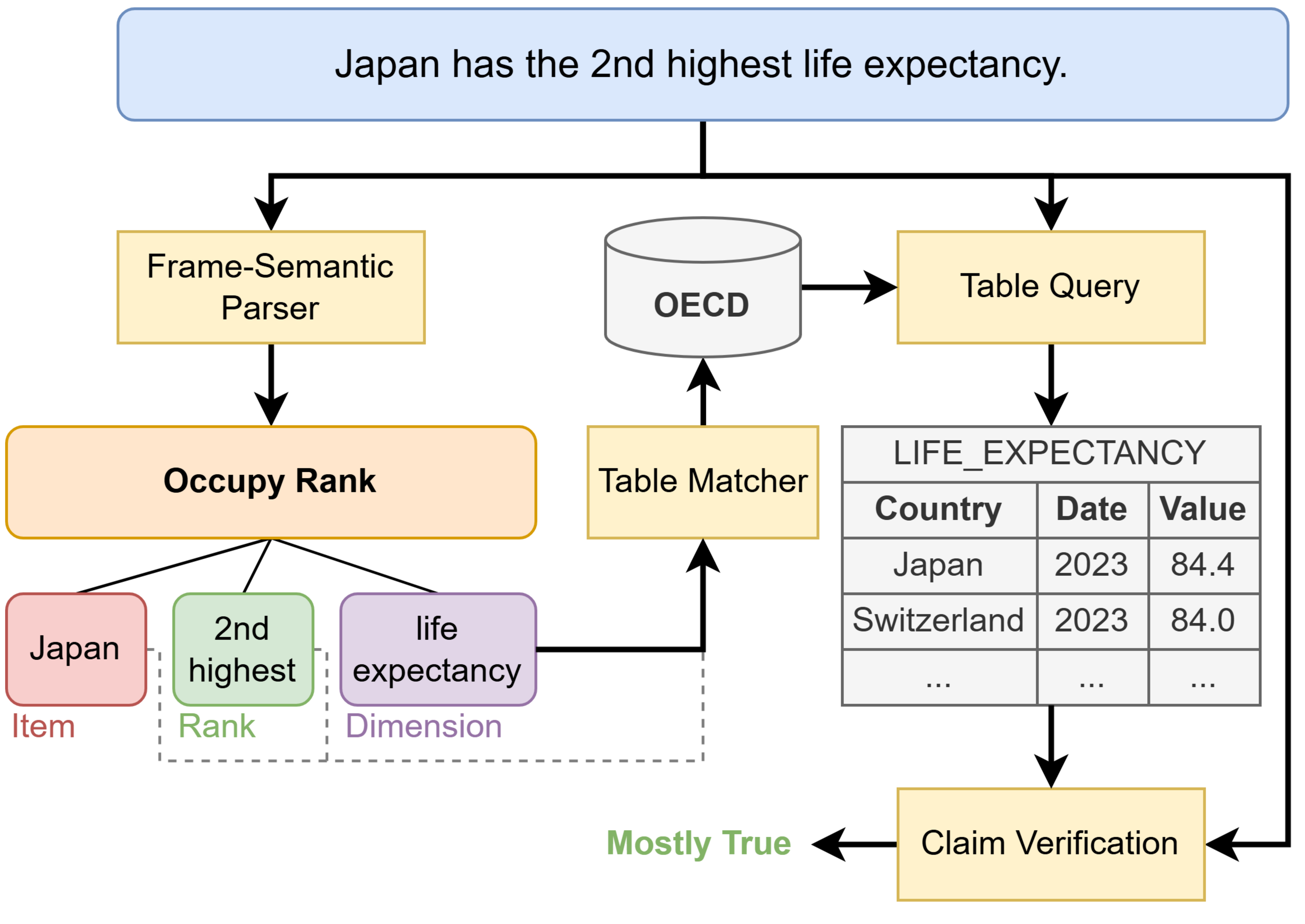
Datasets (oecd.gov):

- **434 Tables** of diverse topics
- **596,552** Avg. rows per table
- **38** different countries (and sometimes more)
- **68 Fact-checks** annotated with relevant table and frame elements

The U.S. is “taking in almost $2 billion a day in tariffs.”



System Architecture



Claim Understanding

- Identifies **semantic frames** and **frame elements**
- Fine-tuned RoBERTa-based **joint frame-semantic parsing** model

Evidence Retrieval

- Frame elements used to find relevant **database tables**
  - **Voting**: matches Agent and Issue FEs to senators and bills
  - **OECD**: matches frame-specific FEs to tables

Claim Verification

- Retrieved **evidence** and **claim** are given to LLM to **predict veracity**
- Outputs one of: True, Mostly True, Half-True, Mostly False, False
- LLM is prompted to consider the ***spirit of the claim*** in prediction

Experiments

Semantic Frame Survey

Predicted distribution of frames across **21K fact-check articles**

- *Vote*, *Occupy Rank*, and *Comparing Entities* among the most common frames

Frame	Samples (%)
Taking_sides	7,152 (34.0%)
Speech	6,010 (28.6%)
Change_position_on_a_scale	5,547 (26.4%)
Comparing_two_entities	5,530 (26.3%)
Cause_change_of_position_on_a_scale	4,675 (22.2%)
Vote	3,229 (15.4%)
Comparing_at_two_different_points_in_time	2,436 (11.6%)
Conditional_occurrence	2,355 (11.2%)
Creating	2,194 (10.4%)
Occupy_rank	1,106 (5.3%)
Oppose_and_support_consistency	1,010 (4.8%)
Recurrent_action_in_Frequency	935 (4.4%)
Ratio	932 (4.4%)
Capability	869 (4.1%)

Evidence Retrieval

Evaluated **similarity models** and **querying strategies**

- **Voting:** DistilBERT-TAS-B
- **OECD:** RoBERTa v2

Model	Query	Data	R@K
`distilbert-tas-b`	Full claim	Vote	0.143
`distilbert-tas-b`	Issue FE	Vote	**0.165**
`Max Possible`	-	Vote	*0.568*
`RoBERTa (v2)`	Full claim	OECD	0.653
`RoBERTa (v2)`	FE	OECD	**0.726**
`Max Possible`	-	OECD	*0.910*

Frame Element semantic search **query outperforms full claim** query

Frame-Semantic Parsing

- RoBERTa **joint frame-semantic parsing model** drastically outperforms in-context learning baseline on FE identification
- Similar performance on **frames**

Model	Frames	Frame Acc	FE Acc
Random	Vote	0.488	0.254
GPT-4o-mini	Vote	0.974	0.618
Vote FSP	Vote	**0.990**	**0.889**
Random	OECD	0.602	0.000
GPT-4o-mini	OECD	0.537	0.372
GPT-4o-mini*	OECD	0.713	0.461
OECD FSP	OECD	**0.742**	**0.873**

End-to-End Fact Verification

Benchmarked end-to-end system against PolitiFact verdicts

- **Naïve** – No evidence
- **Irrelevant** – Samples where no useful evidence is found

Model	Dataset	Accuracy
GPT-4o Naive	Vote	0.044
Ours w/ Irrelevant	Vote	0.076
Ours w/o Irrelevant	Vote	**0.207**
GPT-4o Naive	OECD	0.073
Ours w/ Irrelevant	OECD	0.214
Ours w/o Irrelevant	OECD	**0.429**