ClaimsKG
A Knowledge Graph of Fact-checked Claims

Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland
Malo Gasquet, Matthäus Zloch, Benjamin Zapilko
Stefan Dietze, Konstantin Todorov

Motivation

Investigations into misinformation and the spread of falsehoods and biased discourse are gaining popularity (e.g. [Vousoughi et al. 2018])

- How do false claims spread on social nets?
- What is the stance of a claim-relevant web document?
- How to attribute a source to a web document and discover its type?
- How to assess the veracity of a claim?

Need for ground truth of truth-value labeled claims with metadata

Motivation

Answering specific information needs of (computational) social scientists, journalists, fact-checkers…

- Find all false claims by D. Trump in 2017 that also mention the FBI
- Find the top 5 politicians per month involved in false claims
- Retrieve all claims mentioning journalists

...requires looking up manually a number of different sources, such as fact-checking portals, knowledge bases

Need to enable enhanced retrieval of claims-related information

Examples of Dedicated Resources

Web sources

The **“Liar”** benchmark
- 12.8K claims from Politifact

Clef “**Check that!”** challenge
- 150 claims from Factcheck.org

The “**Emergent**” dataset
- 300 claims from various fact-checking portals

Crowdsourced / manual annotation

The **Open Domain Deception** Dataset
- 7.2K claims from 512 unique contributors

The **SemEval** challenge
- 5.5K claims

The **FEVER** dataset
- >185K claims

Limited in size, static, not regenerable, lack metadata/context, no shared data model

ClaimsKG: A KG of Fact-checked Claims…*and more*

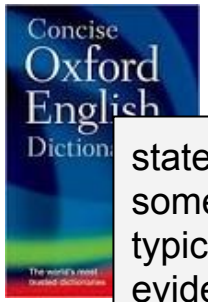
A dataset
- Openly available dynamic KG of claims and associated metadata

A model
- A schema for fact-checked claims based on established vocabularies

Tools for **construction**
- An open source pipeline for crawling, extracting and lifting data

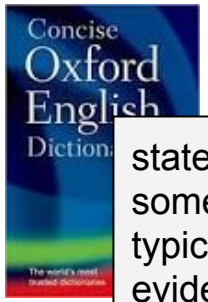
Applications for **search** and **exploration**
- User-friendly open source apps for non-computer scientists

What is a claim?



statement or assertion that something is the case, typically without providing evidence or proof

What is a claim?

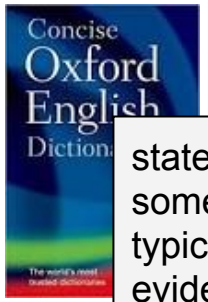


statement or assertion that something is the case, typically without providing evidence or proof

POLITIFACT Snopes
TruthorFiction?
truth and exposing fiction since 1999
Africa Check
Séparer la réalité de la fiction

statement supported by (a group of) people or organizations that appear newsworthy, significant and verifiable

What is a claim?



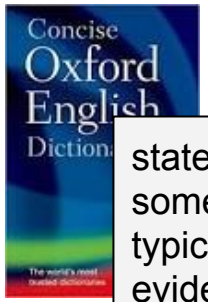
statement or assertion that something is the case, typically without providing evidence or proof

statement supported by (a group of) people or organizations that appear newsworthy, significant and verifiable

AM

the assertion
the argument
aims to prove
or the thesis
to be justified

What is a claim?



statement or assertion that something is the case, typically without providing evidence or proof

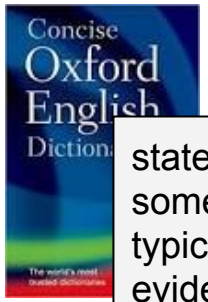
statement supported by (a group of) people or organizations that appear newsworthy, significant and verifiable

AM

the assertion the argument aims to prove or the thesis to be justified

a proposition, an idea which is either true or false, put forward by somebody as true

What is a claim?



statement or assertion that something is the case, typically without providing evidence or proof

POLITIFACT Snopes TruthorFiction? truth and exposing fiction since 1999 Africa Check Séparer la réalité de la fiction

statement supported by (a group of) people or organizations that appear newsworthy, significant and verifiable

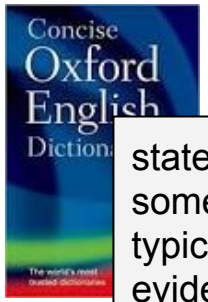
AM

the assertion the argument aims to prove or the thesis to be justified

a proposition, an idea which is either true or false, put forward by somebody as true

general, concise statement that directly supports or contests the topic

What is a claim?



statement or assertion that something is the case, typically without providing evidence or proof

POLITIFACT Snopes TruthorFiction? truth and exposing fiction since 1999 SCAN Africa Check Séparer la réalité de la fiction

statement supported by (a group of) people or organizations that appear newsworthy, significant and verifiable

AM

the assertion the argument aims to prove or the thesis to be justified

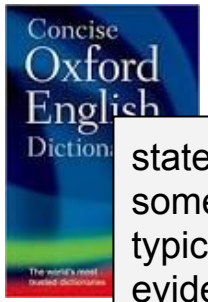
a proposition, an idea which is either true or false, put forward by somebody as true

general, concise statement that directly supports or contests the topic

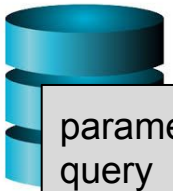
statement formulating a problem together with a concrete solution

QA

What is a claim?



statement or assertion that something is the case, typically without providing evidence or proof



DB

parametrized query over a database

POLITIFACT Snopes

TruthorFiction?

truth and exposing fiction since 1999

Africa Check

Séparer la réalité de la fiction

statement supported by (a group of) people or organizations that appear newsworthy, significant and verifiable

AM

the assertion the argument aims to prove or the thesis to be justified

a proposition, an idea which is either true or false, put forward by somebody as true

general, concise statement that directly supports or contests the topic

statement formulating a problem together with a concrete solution

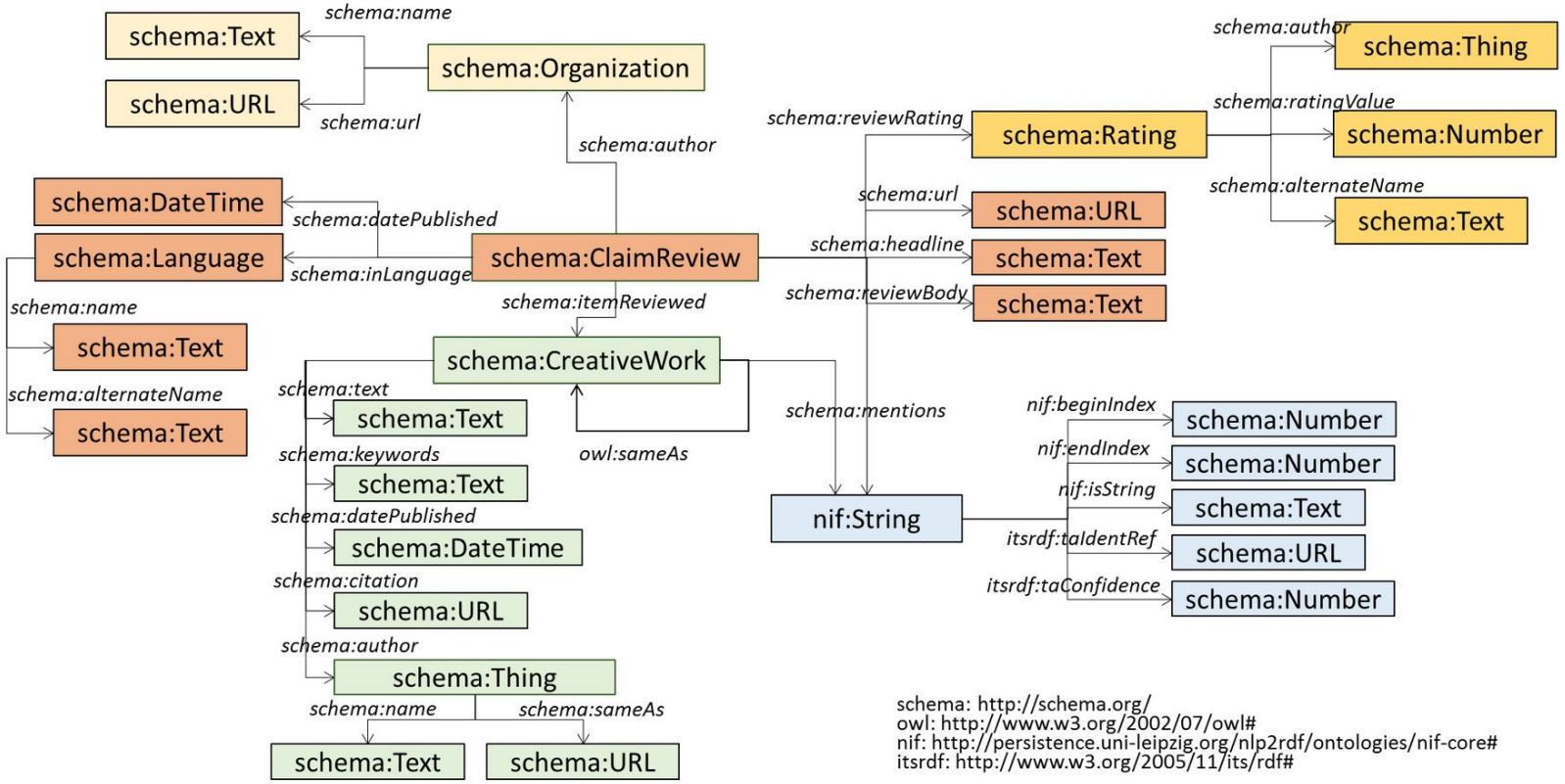
QA

What is a claim?

In this talk:

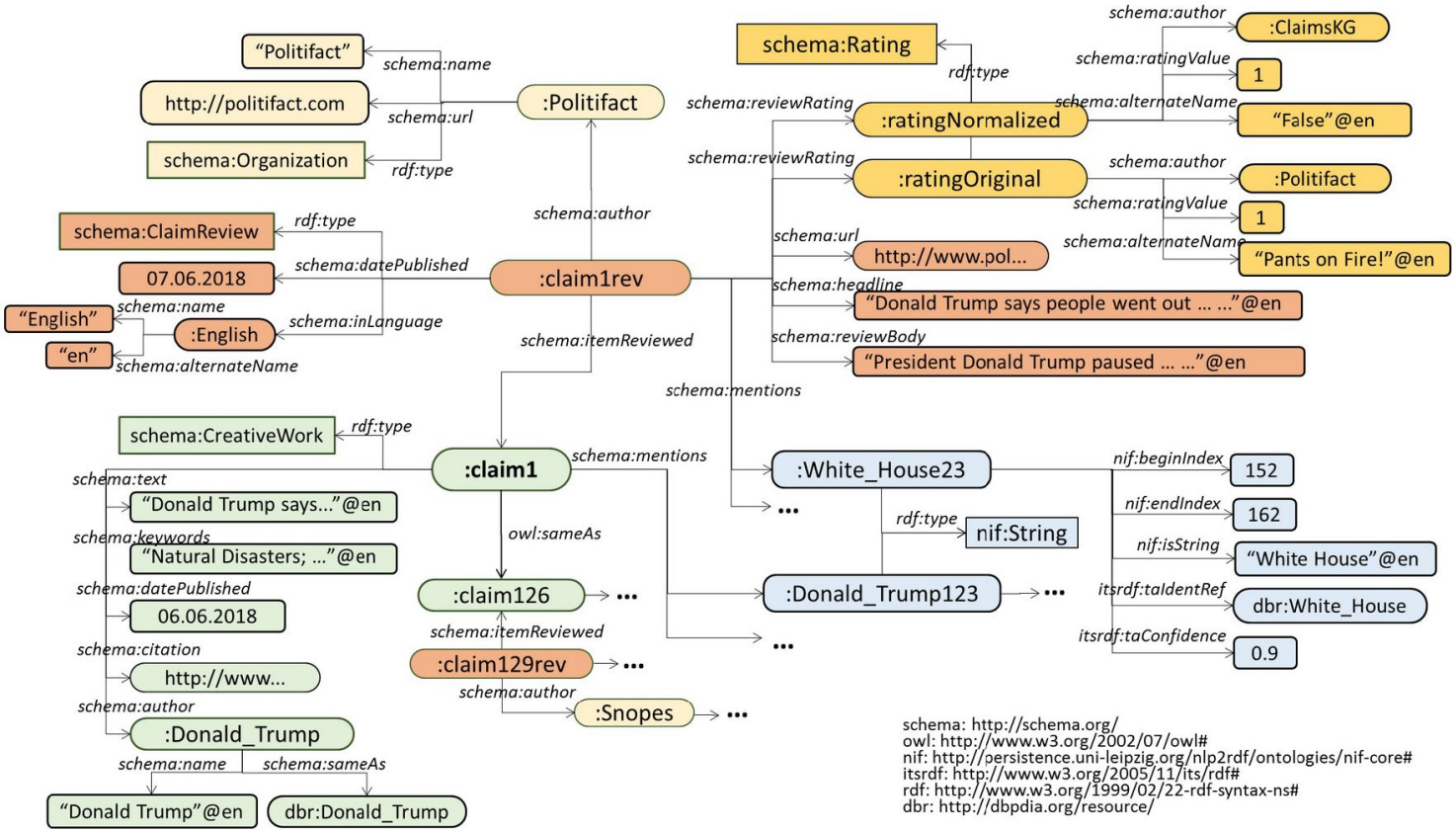
Claim: *a statement reviewed by a fact checking organisation.*

A Fact-checked Claims Model



schema: http://schema.org/
owl: http://www.w3.org/2002/07/owl#
nif: http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#
itsrdf: http://www.w3.org/2005/11/its/rdf#

A Fact-checked Claims Model: an Example



ClaimsKG Construction Pipeline

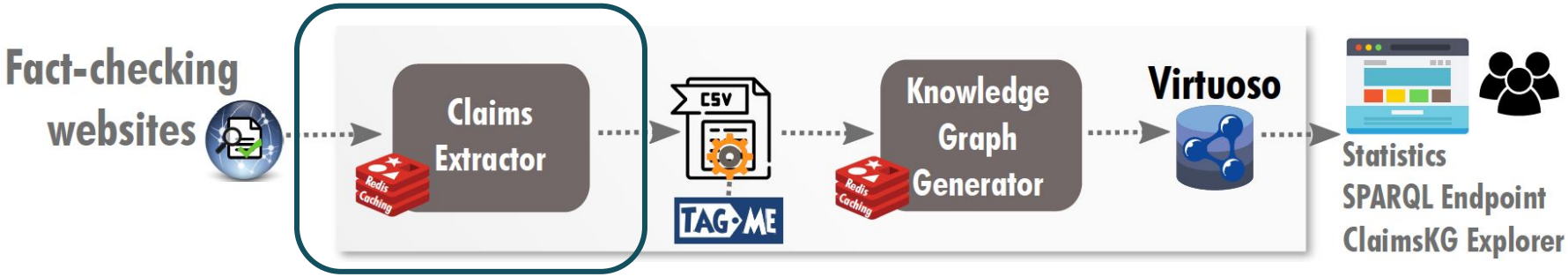
ClaimsKG Construction Pipeline

Snopes - https://www.snopes.com/
Politifact - http://www.politifact.com/
Africa Check - https://africacheck.org/
Truth Or Fiction - http://TruthOrFiction.com
Check Your Fact - http://checkyourfact.com

FactsCan - http://factscan.ca/
Fact Check AFP - https://factcheck.afp.com/
Factuel AFP - https://factuel.afp.com/
Full Fact - https://fullfact.org/

→ ***Mainly related to politics, but country-specific and multilingual.***

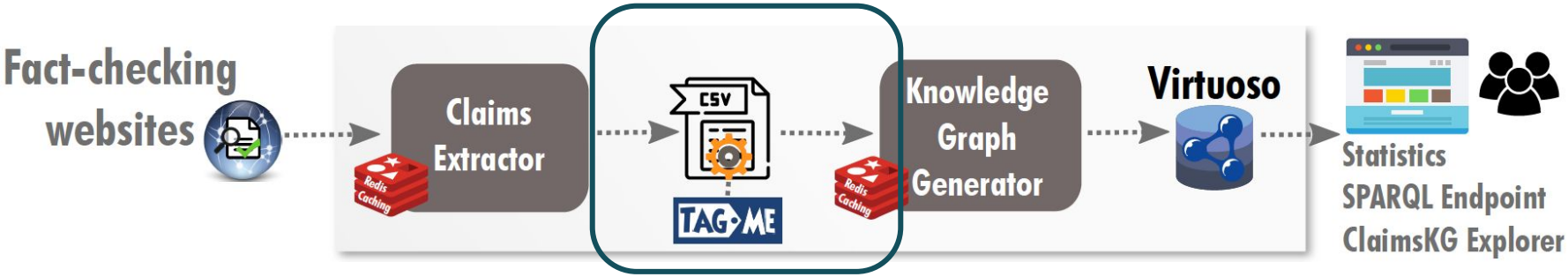
ClaimsKG Construction Pipeline



Website-specific extractors → a multi-sourced dataset (as a CSV file)
https://github.com/claimskg/claimskg-extractor

- the **text of the claim**;
- its **truth value** (both a normalized value and the original one);
- a link to the **claim review** from the fact-checking website;
- the **references** used to verify the claim;
- the **entities** from the claim and its review, their Wikipedia categories;
- the **author** of the claim and of the claim review;
- the **date** of publication of the claim and of the review;
- the **title** of the review article;
- a set of **keywords** that act like **topics** (e.g., "*healthcare*" or "*abortion*")

ClaimsKG Construction Pipeline

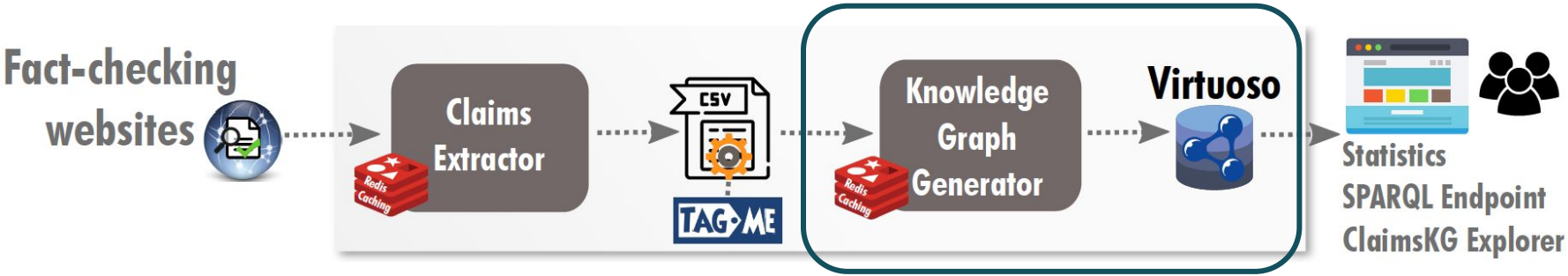


Entities annotation → names of persons, organisations, locations,…

https://github.com/claimskg/tagme

- Entities extracted from the texts of the ***claims***, their ***reviews*** and ***keywords***
- Links to Wikipedia pages and a ***DBpedia*** URI
- Using the TagMe tool (https://tagme.d4science.org/tagme/)

ClaimsKG Construction Pipeline



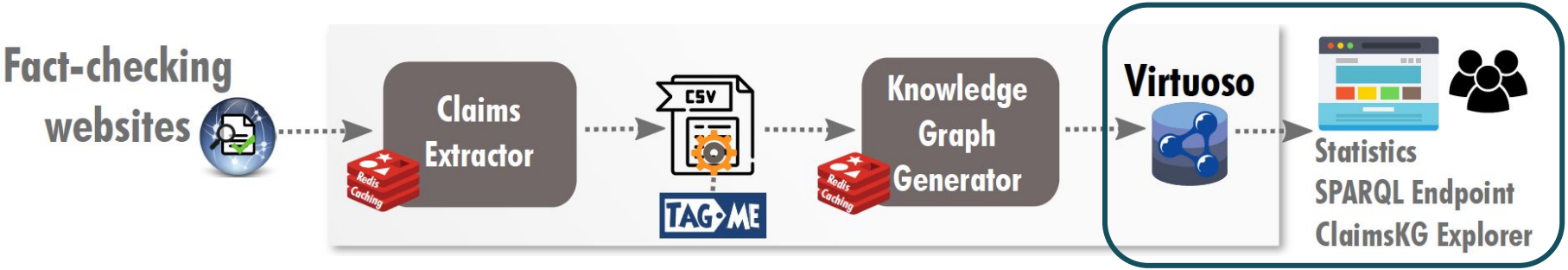
Populating the KG following our data model

https://github.com/claimskg/claimskg_generator

- Read the extracted information from the CSV file
- Generate unique URI identifiers as UUIDs using key attributes as seeds
- Handling ***simple claims co-references*** via exact matching

- Regular updates (bi-monthly)
- Maintenance

ClaimsKG Construction Pipeline



Tools for access, information retrieval and exploration

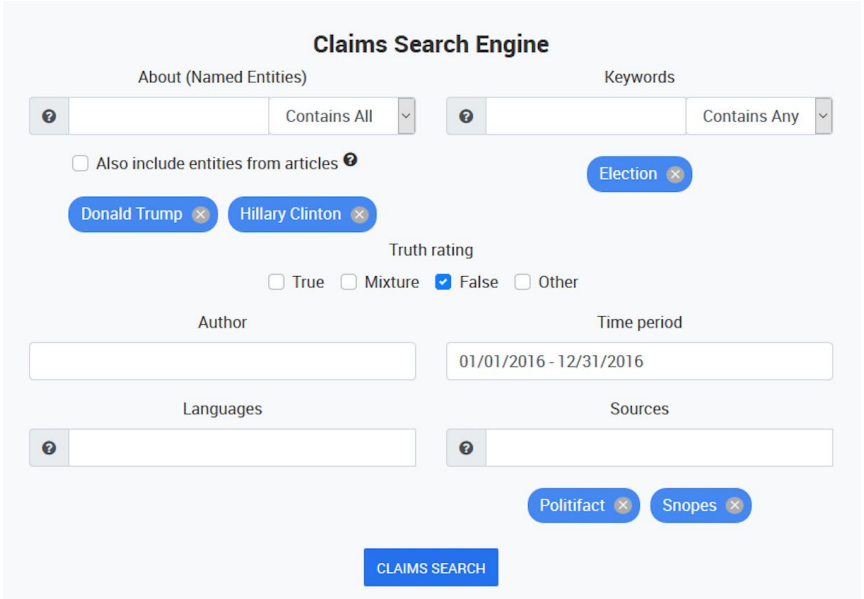
→ public SPARQL endpoint: https://data.gesis.org/claimskg/sparql

→ Open source user apps for exploring fact-checked claims and their descriptive statistics built on top of ClaimsKG and its SPARQL endpoint

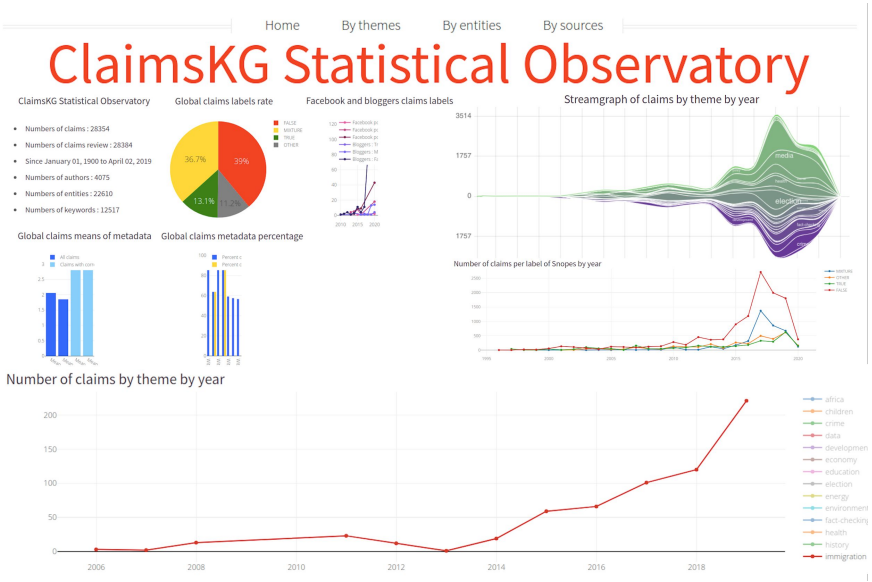
- ClaimsKG Explorer: https://data.gesis.org/claimskg/explorer/home
- ClaimsKG Statistical Observatory: https://data.gesis.org/claimskg/observatory

→ Presentation of these tools at the P&D session tonight.

Web apps for claims exploration and search



For more information, come see me at my stand tonight at the P&D session.



Statistics per property

Claim data coverage (as of October 2019)

Property	Global	Snopes	Politifact	Africa Check	Truth or Fiction	Check your Fact	FactsCan	Fullfact	AFP Factcheck	AFP Factuel (FR)
Number of claims	33,261	12,812	16,476	520	1,311	823	125	250	678	266
Claim text	99.10%	99.98%	100.00%	100.00%	100.00%	100.00%	100.00%	0.00%	100.00%	100.00%
Claim author	44.57%	0.00%	100.00%	0.00%	0.00%	0.00%	100.00%	39.54%	82.32%	62.40%
Claim date published	44.74%	0.00%	99.91%	0.00%	0.00%	0.00%	98.40%	0.00%	100.00%	100.00%
Claim with citation	81.00%	82.32%	76.95%	96.97%	100.00%	99.76%	100.00%	0.00%	99.55%	95.48%
Claim keywords	77.10%	67.47%	100.00%	99.88%	0.00%	0.00%	100.00%	100.00%	0.00%	0.00%
Claim entity mention	98.82%	99.66%	99.98%	98.56%	100.00%	100.00%	100.00%	0.00%	100.00%	100.00%

Statistics per rating type

Truth values and claim exact matches

Property	Global	Snopes	Politifact	Africa Check	Truth or Fiction	Check your Fact	FactsCan	Fullfact	AFP Factcheck	AFP Factuel (FR)
Number of claims	33,261	12,812	16,476	520	1,311	823	125	250	678	266
Identical claims	87	38	49	0	0	0	0	0	0	0
True claims	4,404	1,846	2,334	60	125	0	39	0	0	0
False claims	12,350	6,809	5,036	211	246	0	48	0	0	0
Mixture claims	10,834	1,925	8,844	0	63	2	0	0	0	0
Other claims	5,706	2,232	262	282	877	821	38	250	678	266

What can we do with ClaimsKG - *some examples*

- **Advanced, entity-centric search**, exploration and information discovery, exploiting data from various sources via federated SPARQL queries
  - the top-3 journalists mentioned in claim reviews of 2018;
  - the number of claims per month mentioning Obamacare in 2016;
  - the top-5 persons mentioned in false claims together with “*abortion*”;
  - all false claims of 2017 by Donald Trump about climate change...

What can we do with ClaimsKG - *some examples*

- **Advanced, entity-centric search**, exploration and information discovery, by exploiting data from various sources via federated SPARQL queries
  - the top-3 journalists mentioned in claim reviews of 2018.
  - the number of claims per month mentioning Obamacare in 2016.
  - the top-5 persons mentioned in false claims together with "*abortion*".
- Support **social science research** and use-case studies on particular topics
  - Agenda-setting studies on the influence of mass media on the public's focus of attention
  - Study the evolution of the political discourse about *immigration* over the last years

What can we do with ClaimsKG - *some examples*

- **Advanced, entity-centric search**, exploration and information discovery, by exploiting data from various sources via federated SPARQL queries
  - Requesting the top-3 journalists mentioned in claim reviews of 2018.
  - Requesting the number of claims per month mentioning Donald Trump in 2016.
  - Requesting the top-5 persons mentioned in false claims together with "*abortion*".
- Support **social science research** and use-case studies on particular topics
  - Agenda-setting studies on the influence of mass media on the public's focus of attention
  - Study the evolution of the political discourse about *immigration* over the last years
- Support **CS / AI research** by providing readily balanced training datasets
  - Topic-wise filtering, customized set of attributes
  - Define classification tasks: [{true} vs. {false}], [{true, false} vs. {mixture}], etc.

Challenges / Ongoing Work

Enriching & curating ClaimsKG

- Diversify fields/topics, sources and languages, discover missing links, etc.
- Linking claims to web documents and to authors (when missing)

Extracting a topic structure from the ClaimsKG's keywords; linking it to established vocabularies in political and social sciences

- Improve interoperability and claim discovery across fact-checking portals

Understand and analyse the biases of journalistic fact-checked data

- Fact-checked claims widely used for automatic fact-checking algorithms and a variety of use-cases in social sciences
- What are the limitations of studies based on such data?

Challenges / Ongoing Work

Matching / clustering claims according to various criteria

- Contexts, topics, sources, proposition, utterances,...

Identifying complex claim structures from certain fact-checking portals

- Colin Kaepernick says *Winston Churchill said, "A lie gets halfway around the world before the truth has a chance to get its pants on.":* the claim is true, while the subclaim is false.

Learning **claim embeddings** from ClaimsKG enriched with DBpedia context

Towards an **extended and generalized claims model** *(cf. talk at CKG WS)*

- What are the dimensions that define a claim, outside of the fact-checking context?
- How do definitions from different fields generalize to a common understanding of that notion?
- Go **beyond facts**, the current center of interest in traditional KGs

Thank you for listening.

https://data.gesis.org/claimskg/site/

We thank

POLITIFACT TruthorFiction? Seeking truth and exposing fiction since 1999 Snopes FullFact Africa Check FACTSCAN AFP CHECKYOURFACT

First level agenda-setting: inserting topics, events or entities into the public discourse, thereby regulating societal priorities;
Second-level agenda-setting: increasing the salience of specific features or attributes of entities in the discourse.

Through the Web, citizens are now able to play a more active role in influencing the public discourse.
Using ClaimsKG, an exploratory search on a topic and related entities may be performed in order to gain insights on relevant viewpoints, attributes and actors

The query given in Fig. 5 retrieves all claims mentioning Trayvon Martin or George Zimmerman, yielding 68 claims in total with 8 claims rated true, 33 false, and 24 mixture. The distribution of truth values hints at this being a highly controversial topic with potentially highly polarized Viewpoints.

Retrieve all entities mentioned in claims together with “stand your ground law”

ClaimsKG General Statistics *(as of September 2019)*

	total	true	false	mixture	other
Global	33261	4404	12350	10834	5706
Politifact	16476	2334	4841	8729	262
Snopes	12477	1768	6737	1843	2129
Truth or Fiction	1060	110	241	63	646
Africa Check	590	66	216	0	308
Factscan	125	39	48	0	38
other	1681	0	0	2	1679

The resource contains all claims published since the foundation of each portal, first claim from 1996.