Truth or Dare: Investigating Claims Truthfulness with ClaimsKG

Susmita Gangopadhyay, Katarina Boland, Danilo Dessí, Stefan Dietze, Pavlos Fafalios, Andon Tchechmedjiev, Konstantin Todorov and Hajira Jabeen

Date
Fact-checking websites and Claims

Will Eating Bananas Prevent Coronavirus Infection?
Was COVID-19 Found in Packages of Toilet Paper?
Will Warm Weather Kill Coronavirus?
Did a Chinese Intelligence Officer Reveal the “Truth About the Coronavirus Outbreak”?
Did ‘The Simpsons’ Predict the Coronavirus Outbreak?
What is ClaimsKG?

- A Knowledge Graph of Fact checked Claims.

Claim

- Statement which has been reviewed by a fact-checking organization in order to assess its truthfulness

Objective of ClaimsKG

- To generate ground truth for machine learning models
- Facilitate structured queries about their truth values, authors, dates, journalistic reviews, and other kinds of metadata
# Statistics-ClaimsKG Aug 2022

<table>
<thead>
<tr>
<th>Websites</th>
<th>URL</th>
<th>Total Claims</th>
<th>Total Entities</th>
<th>True Claims</th>
<th>False Claims</th>
<th>Mixture Claims</th>
<th>Other Claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>NA</td>
<td>59,580</td>
<td>1,371,271</td>
<td>7,151</td>
<td>30,858</td>
<td>10,790</td>
<td>10,789</td>
</tr>
<tr>
<td>Politifact</td>
<td><a href="https://politifact.com">https://politifact.com</a></td>
<td>21,450</td>
<td>354,653</td>
<td>2,501</td>
<td>8,353</td>
<td>6,733</td>
<td>3,863</td>
</tr>
<tr>
<td>Snopes</td>
<td><a href="https://snopes.com">https://snopes.com</a></td>
<td>14,031</td>
<td>481,199</td>
<td>2,843</td>
<td>6,803</td>
<td>2,556</td>
<td>1,829</td>
</tr>
<tr>
<td>AFP Factcheck</td>
<td><a href="https://factcheck.afp.com">https://factcheck.afp.com</a></td>
<td>5,058</td>
<td>151,208</td>
<td>3</td>
<td>4,147</td>
<td>97</td>
<td>811</td>
</tr>
<tr>
<td>AFP Factuel(FR)</td>
<td><a href="https://factuel.afp.com">https://factuel.afp.com</a></td>
<td>935</td>
<td>18,739</td>
<td>5</td>
<td>627</td>
<td>94</td>
<td>209</td>
</tr>
<tr>
<td>Checkyourfact</td>
<td><a href="https://checkyourfact.com">https://checkyourfact.com</a></td>
<td>3,971</td>
<td>16,699</td>
<td>233</td>
<td>3,691</td>
<td>4</td>
<td>43</td>
</tr>
<tr>
<td>Vishva news</td>
<td><a href="https://www.vishvasnews.com">https://www.vishvasnews.com</a></td>
<td>3,490</td>
<td>8,930</td>
<td>0</td>
<td>2,933</td>
<td>565</td>
<td>0</td>
</tr>
<tr>
<td>Fullfact</td>
<td><a href="https://fullfact.org">https://fullfact.org</a></td>
<td>2,928</td>
<td>6,870</td>
<td>403</td>
<td>729</td>
<td>152</td>
<td>1,644</td>
</tr>
<tr>
<td>Truth or Fiction</td>
<td><a href="https://truthorfiction.com">https://truthorfiction.com</a></td>
<td>2,908</td>
<td>21,298</td>
<td>853</td>
<td>260</td>
<td>14</td>
<td>1,781</td>
</tr>
<tr>
<td>Africacheck</td>
<td><a href="https://africacheck.org">https://africacheck.org</a></td>
<td>2,854</td>
<td>11,448</td>
<td>197</td>
<td>2,364</td>
<td>258</td>
<td>35</td>
</tr>
<tr>
<td>Fatabyano</td>
<td><a href="https://fatabyano.net">https://fatabyano.net</a></td>
<td>1,379</td>
<td>101</td>
<td>46</td>
<td>820</td>
<td>4</td>
<td>509</td>
</tr>
<tr>
<td>Factograph</td>
<td><a href="https://factograph.info">https://factograph.info</a></td>
<td>255</td>
<td>1,201</td>
<td>19</td>
<td>69</td>
<td>144</td>
<td>23</td>
</tr>
<tr>
<td>Eu Factcheck</td>
<td><a href="https://eufactcheck.eu">https://eufactcheck.eu</a></td>
<td>297</td>
<td>5,699</td>
<td>48</td>
<td>48</td>
<td>159</td>
<td>42</td>
</tr>
<tr>
<td>Polygraph</td>
<td><a href="https://polygraph.info">https://polygraph.info</a></td>
<td>24</td>
<td>293,226</td>
<td>0</td>
<td>14</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

13 Websites  
59k Claims  
1M Entities  
14 Languages
Updates for Latest Release – ClaimsKG Aug 2022

- Claims in various languages like English, French, Russian, Urdu, Hindi, Punjabi, Assamese, Tamil, Malayalam, Gujarati, Telegu, Marathi, Odia, and Bengali


- Removal of factscan.ca (due to unavailability)

- Time frame: 1996 to 2022 (August)

ClaimsKG-Pipeline and Modules

Extractor

Web Scraping

Input CSV

Fact Checking Websites

Python Module

API

Entity fishing

Knowledge graph

Generator

NERD

RDF

Data Dump

SowiDataNet

virtuoso

SPARQL Endpoint

Triple Store

Statistics

ClaimsKG
ClaimsKG-Pipeline and Modules

Generator

- Entity Annotation and Linking
- Rating Normalization
- Lifting and Serialization

Extractor

Input CSV

NERD

API

Web Scraping

Fact Checking Websites

Python Module

Entity fishing

Knowledge graph

ClaimsKG

SowiDataNet

Data Dump

RDF

virtuoso

Triple Store

SPARQL Endpoint

Statistics
ClaimsKG-Pipeline and Modules

Generator
- Entity Annotation and Linking
- Rating Normalization
- Lifting and Serialization

Extractor
- Web Scraping
- Input CSV

Python Module
- Fact Checking Websites

API
- Entity fishing
- Knowledge graph

NERD

RDF

SowiDataNet
datautorium
- Data Dump

ClaimsKG

Virtualo
- Triple Store

SPARQL Endpoint
- Statistics
ClaimsKG-Pipeline and Modules

Generator

- Entity Annotation and Linking
- Rating Normalization
- Lifting and Serialization

Extractor

- Web Scraping
- Input CSV

Fact Checking Websites

Python Module

API

Entity fishing

Knowledge graph

ClaimsKG

SowiDataNet+Datorium

Data Dump

RDF

Triple Store

Statistics

SPARQL Endpoint
Final Outcome

The **Claims data model** for representing fact-checked claims and associated information

[https://data.gesis.org/claimskg/site/](https://data.gesis.org/claimskg/site/)

An **open-source pipeline** for crawling and extracting data from fact-checking websites


A **large-scale KG** of claims and associated metadata

[https://doi.org/10.7802/2469](https://doi.org/10.7802/2469)

A **Web interface** for search and exploration of the resource.

[https://data.gesis.org/claimskg/explorer](https://data.gesis.org/claimskg/explorer)
Claim

@prefix ckg: http://data.gesis.org/claimskg

<ckg/creative_work/709de33f-d764-557c-9992-e31075fb3237> a schema:CreativeWork ;
  schema:author <ckg/creative_work_author/00d0476e-0ba3-5544-88be-7c0303ce856c> ;
  schema:datePublished "2021-03-17"^^xsd:date ;
  schema:mentions <ckg/mention/3d755852-2309-5364-9c47-82b385f9d18a> ;
  schema: text "Biden faked an interaction with reporters" @en .

<ckg/claim_review/0974e149-fd11-5073-affb-248850076bb6> a schema:ClaimReview ;
  schema:author <ckg/organization/factcheck_afp> ;
  schema:datePublished "2021-03-19"^^xsd:date ;
  schema:headline "Biden did not fake comments to reporters on White House lawn" @en ;
  schema:inLanguage <ckg/language/English> ;
  schema:itemReviewed <ckg/creative_work/709de33f-d764-557c-9992-e31075fb3237> ;
  schema:reviewBody "Joe Biden’s delay in holding a news conference has spurred criticism online,... ." @en ;
  schema:reviewRating <ckg/rating/normalized/claimskg_FALSE>,<ckg/rating/original/factcheck_afp_false_1> ;

A sample claim in Knowledge Graph from the Website Fullfact
Use-Cases

1. Requesting number of claims year wise mentioning Coronavirus
Use-Cases

2. Requesting all false claims mentioning “Coronavirus” year-wise.

"Genuine news report says Pope Francis announced coronavirus vaccine is ‘required to enter heaven’" 2020-12-09

"Germany has sent a bill to China for “coronavirus damages”." 2020-04-20

"Flu vaccine will give you coronavirus" 2020-10-08

"Genuine message about COVIDSafe, Australia’s coronavirus contact-tracing app" 2020-04-29

"Quatre hommes contaminés au coronavirus se sont échappés d’un hôpital au Cameroun" 2020-03-25
Federated Queries with External Databases

2. Requesting all claims mentioning President of the United States from the Democratic Party

"Barack Obama began his presidency 'with an apology tour.'”,
"http://dbpedia.org/resource/Barack_Obama"

"‘America is dead’ — Joe Biden actually said this at his Iowa rally.”,
"http://dbpedia.org/resource/Joe_Biden"

"Jimmy Carter said that medical marijuana cured his cancer.“
"http://dbpedia.org/resource/Jimmy_Carter"
Use-Cases

1. Requesting all claims related to “LGBTQ” community year wise

- Sydney Morning Herald
  - [Link](https://www.smh.com.au)
  - Title: Same-sex marriage legalised in Australia as Parliament ...
  - Date: 7 Dec 2017
  - Summary: Australia has become the 26th country to legalise same-sex marriage after the law passed on Thursday with the overwhelming backing of the ... (Click for full article)

- Australian Parliament House
  - [Link](https://www.aph.gov.au)
  - Title: Chronology of same-sex marriage bills introduced into the ...
  - Summary: On 15 November 2017 Senator Dean Smith (LIB, WA) introduced, on behalf of eight cross-party co-sponsors, a bill to amend the Marriage Act 1961 (Cth) so as to ... (Click for full article)

- CNN.com
  - [Link](https://www.cnn.com)
  - Title: Australia legalizes same-sex marriage
  - Date: 7 Dec 2017
  - Summary: Australia's parliament voted to legalize same-sex marriage Thursday after years of political jockeying and fierce public debate. (Click for full article)
1. Veracity labels of all claims mentioning diseases

Truth Values

Owning a chihuahua can help cure a person of asthma. **FALSE**

"Rubbing hydrogen peroxide on the body “once, preferably twice a day” treats cancer. **FALSE**

Malaria kills 756,000 under five Nigerian children each year **FALSE**

"Dandelion root extracts can cure can forms of cancer" **FALSE**

Sunscreen causes cancer **FALSE**

Paracetamol painkillers like Panadol cause addiction **FALSE**
Future Goals

- Strengthen the NERD tool
- Monitoring sparql endpoint
- Scrape data at regular intervals
- Add sources with diverse languages
- Analyze similar claims
- As a corpus for research
Links to ClaimsKG data and tools

- ClaimsKG Website: https://data.gesis.org/claimskg/site/
- Dataset Download: https://doi.org/10.7802/2469
- Previous Versions: https://zenodo.org/record/3518960
- SPARQL Endpoint: https://data.gesis.org/claimskg/sparql
- Claims Ontology: https://data.gesis.org/claimskg/site/#model
- Source Code: https://github.com/claimskg
- ClaimsKGExplorer: https://data.gesis.org/claimskg/explorer


ADVICE - $0.50
GOOD ADVICE - $2.00

Fragen?
Thank You!

If you would like to know more please contact:
Susmita.gangopadhyay@gesis.org
## Entity fishing Performance*

<table>
<thead>
<tr>
<th></th>
<th>Priors</th>
<th>entity-fishing</th>
<th>Wikifier</th>
<th>DoSeR</th>
<th>AIDA</th>
<th>Spotlight</th>
<th>Babelfy</th>
<th>WAT</th>
<th>(Ganea &amp; Hofmann, 2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE2004</td>
<td>83.1</td>
<td>83.5</td>
<td>83.4</td>
<td>90.7</td>
<td>81.5</td>
<td>71.3</td>
<td>56.1</td>
<td>80.0</td>
<td>88.5</td>
</tr>
<tr>
<td>AIDA-CONLL-testb</td>
<td>66.1</td>
<td>76.5</td>
<td>77.7</td>
<td>78.4</td>
<td>77.4</td>
<td>59.3</td>
<td>59.2</td>
<td>84.3</td>
<td>92.2</td>
</tr>
<tr>
<td>AQUAINT</td>
<td>80.3</td>
<td><strong>89.1</strong></td>
<td>86.2</td>
<td>84.2</td>
<td>53.2</td>
<td>71.3</td>
<td>65.2</td>
<td>76.8</td>
<td>88.5</td>
</tr>
<tr>
<td>MSNBC</td>
<td>71.1</td>
<td>86.7</td>
<td>85.1</td>
<td>91.1</td>
<td>78.2</td>
<td>51.1</td>
<td>60.7</td>
<td>77.7</td>
<td>93.7</td>
</tr>
</tbody>
</table>

F score for entity-fishing tool for multiple datasets

* [https://github.com/kermitt2/entity-fishing](https://github.com/kermitt2/entity-fishing)