ClaimLinker
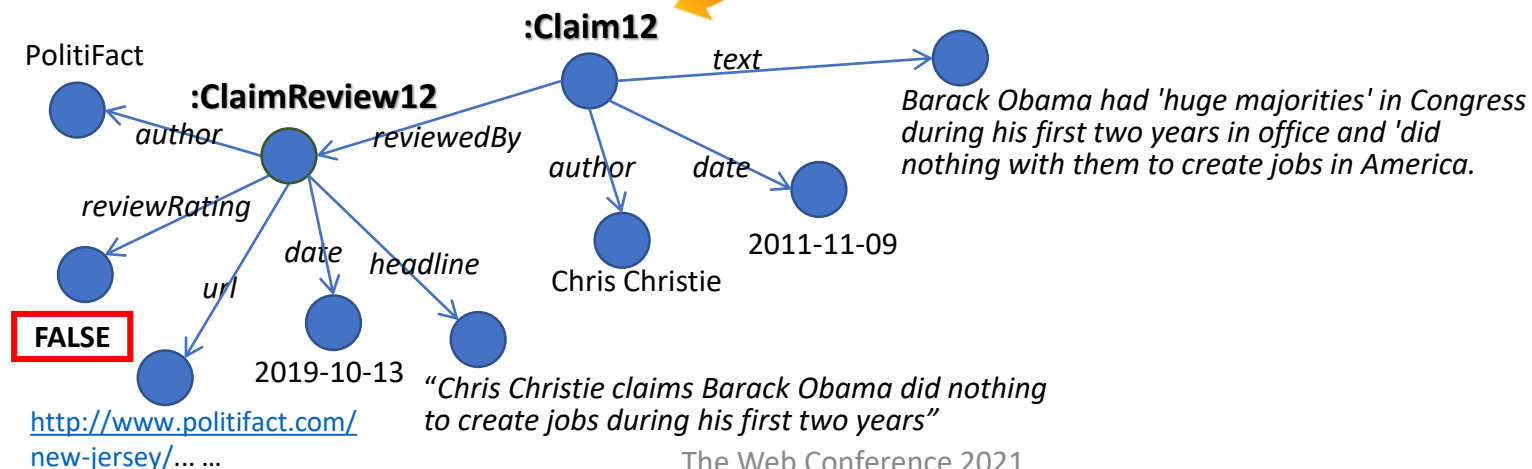## Linking Text to a Knowledge Graph of Fact-checked Claims

Evangelos Maliaroudakis, Katarina Boland, Stefan Dietze,
Konstantin Todorov, Yannis Tzitzikas, Pavlos Fafalios

ClaimLinker in a nutshell

- An **API** and **Web Service** that **annotates** on the fly a text with **fact-checked claims** mined from popular **fact-checking sites** and **integrated into a rich knowledge graph**
  - Offering a **novel kind of semantic annotation** of unstructured content!
  - Fully unsupervised, modular, efficient, scalable, effective (**76% Precision@1**)

> *Obama can complain about Republicans in the House as much as he wants.* <mark>*But in the first two years, he had, you know, huge majorities in the House and Senate, and did nothing with them to create jobs.*</mark>

PolitiFact

:Claim12

:ClaimReview12

author   *reviewedBy*   *text*

Barack Obama had 'huge majorities' in Congress during his first two years in office and 'did nothing with them to create jobs in America.

reviewRating

author   *date*

url   *date*   *headline*

FALSE

2019-10-13

Chris Christie

2011-11-09

"Chris Christie claims Barack Obama did nothing to create jobs during his first two years"

http://www.politifact.com/new-jersey/... ...

Top linked fact-checked claim

Exploration of claim-related Linked Data

verdict

Motivation

Round Table at the CLEF-2020 CheckThat! Lab on September 22, 2020:
David Corney, Full Fact

What do fact checkers want?

1. To know what is the most important claim to check today — *Check-worthiness*
2. To know when someone says something that they've already checked — *Verified claim retrieval*
3. To check claims as quickly as possible — *Evidence retrieval / claim verification*

ArXiv-2021: *Automated Fact-Checking for Assisting Human Fact-Checkers.* Preslav Nakov, David Corney, Maram Hasanain, Firoj Alam, Tamer Elsayed, Alberto Barrón-Cedeño, Paolo Papotti, Shaden Shaar, Giovanni Da San Martino

- Explosive growth of fake news on the Web
  - Potential to affect public opinion towards important decisions (e.g., vaccination)
  - Erosion to democracy and public trust!

- Already, non-profit organizations in many countries fact-check tens of claims every day
  - PolitiFact, FullFact, Snopes, CheckYourFact, …
  - Very time consuming process!

- Need for mechanisms to facilitate the detection of false/misleading information
  - Especially when the same information has already been fact-checked!

> ▪ *How to quickly check if a statement's veracity has been already examined by a trustworthy fact-checking portal?*
>   ▪ *How to check if there are fact-checked claims that can provide information about a statement's veracity?*

Background – ClaimsKG

- RDF knowledge graph of fact-checked claims
  - Generated through a pipeline which periodically harvest data from highly reputable **fact-checking portals**
  - Claims and review articles are annotated with **DBpedia entities**
  - **RDFS model** based on established vocabularies such as *schema.org*
  - It currently contains metadata for around **35K** fact-checked claims

- Enables running complex structured (SPARQL) queries
  - About truth values, authors, dates, related entities, topics, and other contextual information
  - *E.g., "Give me false claims of 2017 mentioning both Donald Trump and the FBI"*

Claim Linking – Problem Definition

- Given:
  - A knowledge graph of fact-checked claims (e.g., ClaimsKG)
- Input:
  - A text (e.g., a sentence in a news article, a tweet, etc.)
- Output:
  - A set of claim annotations of the form:

```
<text snippet, position, ranked list of <claim URI, score> pairs>
```

Obama can complain about Republicans in the House as much as he wants. But in the first two years, he had, you know, huge majorities in the House and Senate, and did nothing with them to create jobs.

- http://data.gesis.org/claimskg/claim12 (score: 0.85)
- http://data.gesis.org/claimskg/claim17 (score: 0.60)

System Overview

Pre-processing

Claims KG — Indexing — Search Service Provision — Search API

Real Time

INPUT: text ▸ Detection of check-worthy text ▸ Candidate Generation ▸ Candidate Ranking ▸ OUTPUT: Claim annotations

Use of **Elasticsearch**

- Efficiency, scalability
- Indexed fields for each claim:
  - Claim text, headline of fact-checking article, claim URI, truth value, author name, URL of fact-checking article

Consider **all text sentences** as check-worthy

- Limiting false negative cases
- Compatible with our use case scenarios

Query **Elasticsearch**

- Considering the whole sentence as the query keywords
- Using the default similarity model of Elasticsearch (Okapi BM25)
- Retrieving the top-30 results and filtering out claims with low Elasticsearch relevance score (<5)

Combination of **Elasticsearch score** and 8 **textual similarity measures**

- *Common words, lemm. words, named entities, disambig. entities, POS tags, n-grams, n-chargrams, cosine similarity*
- **Lightweight** and **fully unsupervised**

Evaluation results

- Ground truth:
  - Test set of CLEF2020 Check-That! 2020 Task 2 (*verified claim retrieval*)
    - 200 tweets to be matched against 10,373 fact-checked claims (v.3.0 of the dataset)
  - No training, no tuning! Tweet pre-processing: removal of URLs

Ranking method	Correct claim in top position (P@1)	Correct claim in top-3 list	Correct claim in top-5 list
Elasticsearch score (100%)	71.5%	79.0%	80.5%
Textual Similarity Measures (100%)	62.0%	77.5%	82.5%
Combination (50% / 50%)	**76.0%**	**80.0%**	**83.5%**

- Our combined method **outperforms all unsupervised methods** participated in the challenge
- P@1 can reach ~**90%** if we consider a higher Elasticsearch score threshold (20 instead of 5)
  - However, then recall drops to ~63%

System components & use cases

- Java Library
  - It communicates with an Elasticsearch service and offers the claim linking functionality

- Web Service
  - It accepts HTTP requests and returns the results in JSON

- Web application
  - It offers a form where the user can give some text and check if there are fact-checked claims linked to that text
  - Demo at: http://users.ics.forth.gr/~fafalios/claimlinker/

- Bookmarklet
  - It allows a user to select a piece of text in any web page (e.g., in an online news article) and check if there are fact-checked claims linked to the selected text
  - Usage guidelines at http://users.ics.forth.gr/~fafalios/claimlinker/

All available as open source: https://github.com/malvag/ClaimLinker

Web Application

http://users.ics.forth.gr/~fafalios/claimlinker/

You know, interest on debt will soon exceed security spending.

You can also use the ClaimLinker service as you browse the Web throug
How to use it:
1. Drag and drop the above link in your bookmarks
2. Select a piece of text in a webpage and click the bookmark

Submit

(1) Give some text and click on Submit

cfe-34c2-542b-9ffa-41a4c8677ced at ClaimsKG

http://data.gesis.org/claimskg/creative_work/16076cfe-34c2-542b-9ffa-41a4c8677ced

Property	Value
?:author	• claimskg:creative_work_author/ba5bfcad-3102-55dc-9e6a-42dc03c4a6db
?:datePublished	• 2011-02-11 (xsd:date)
is ?:itemReviewed of	• claimskg:claim_review/000026e1-eb23-588c-83cd-8d8cdc2bca21
?:keywords	• Economy,Federal Budget (en)
?:text	• 'Within a few years, we will be spending more on interest payments than on national security.' (en)
rdf:type	• <http://schema.org/CreativeWork>

This page shows information obtained from the SPARQL endpoint at https://data.gesis.org/claimskg/sparql.
As Turtle | As RDF/XML | Browse in Disco | Browse in Tabulator | Browse in OpenLink Browser

(3) Start browsing ClaimsKG

1. You know, interest on debt will soon exceed security spending.

'Within a few years, we will be spending more on interest payments than on national security.' `false` `ClaimsKG`

'In just 17 years, spending for Social Security, federal health care and interest on the debt will exceed ALL tax revenue!' `mostly true` `ClaimsKG`

'By 2022, just the interest payment on our debt will be greater than the defense of our country.' `mostly true` `ClaimsKG`

'The debt will soon eclipse our entire economy.' `true` `ClaimsKG`

'Our spending has caught up with us, and our debt soon will eclipse the entire size of our national economy.' `true` `ClaimsKG`
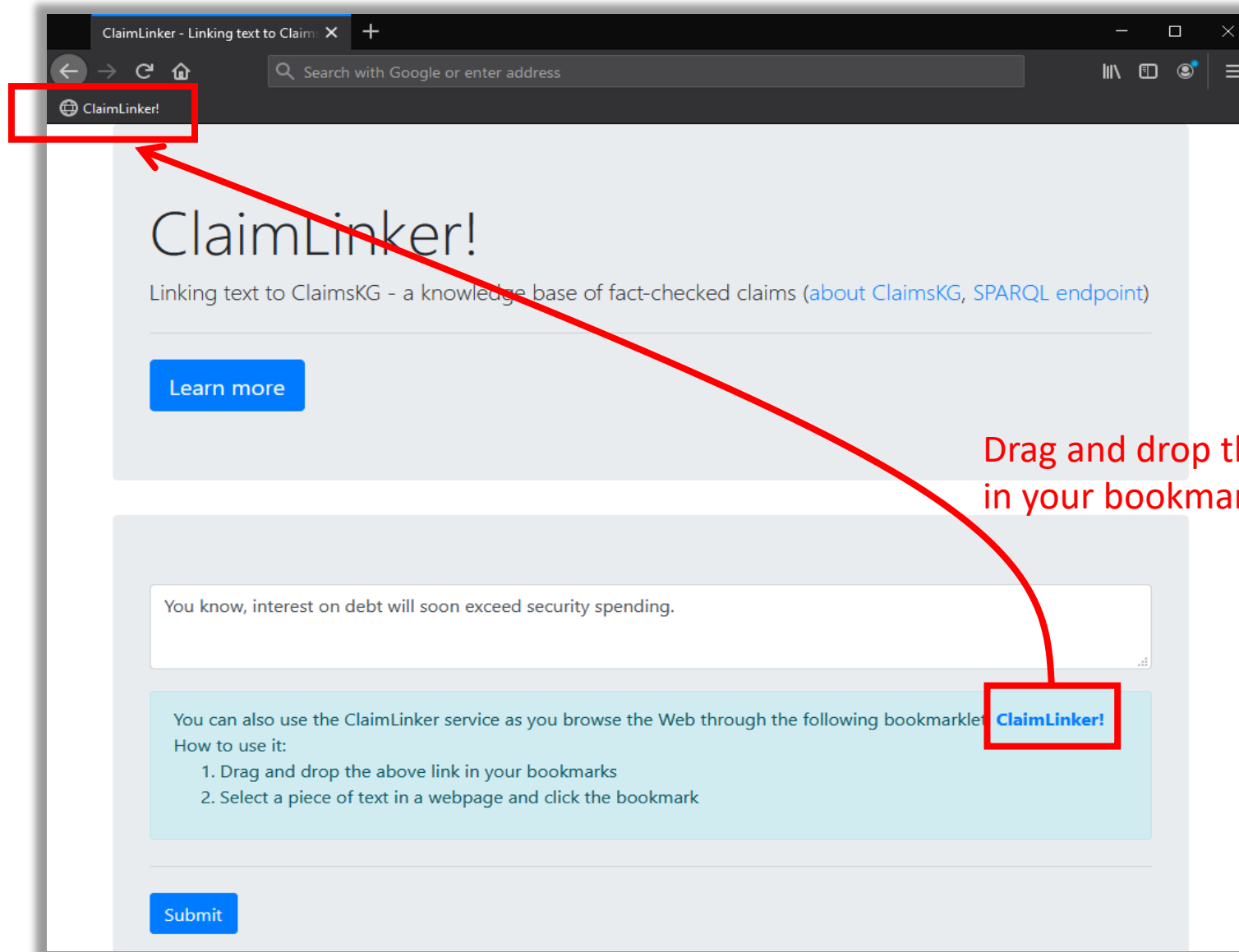
(2) Check the linked fact-checked claims and their verdicts

Bookmarklet

http://users.ics.forth.gr/~fafalios/claimlinker/

Bookmarklet

http://users.ics.forth.gr/~fafalios/claimlinker/

While browsing…

Select a text and click the bookmark

Check the linked fact-checked claims

Dog-owners face 78% higher risk of catching Covid-19

#	Claim	ClaimsKG
1	'Walking your pet dog may raise your risk of catching coronavirus by 78 per cent'	Link

Exploration of claim-related Linked Data

verdict
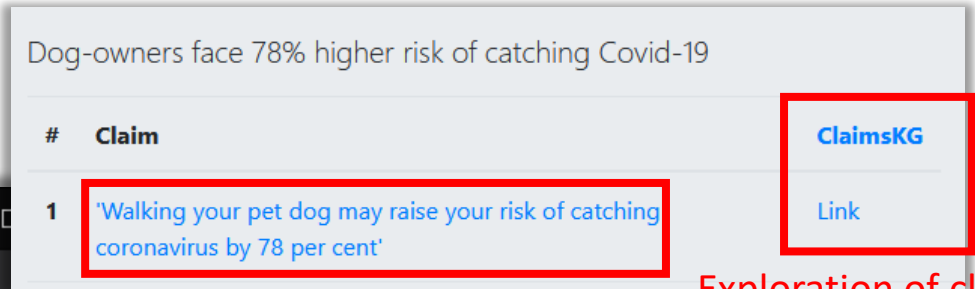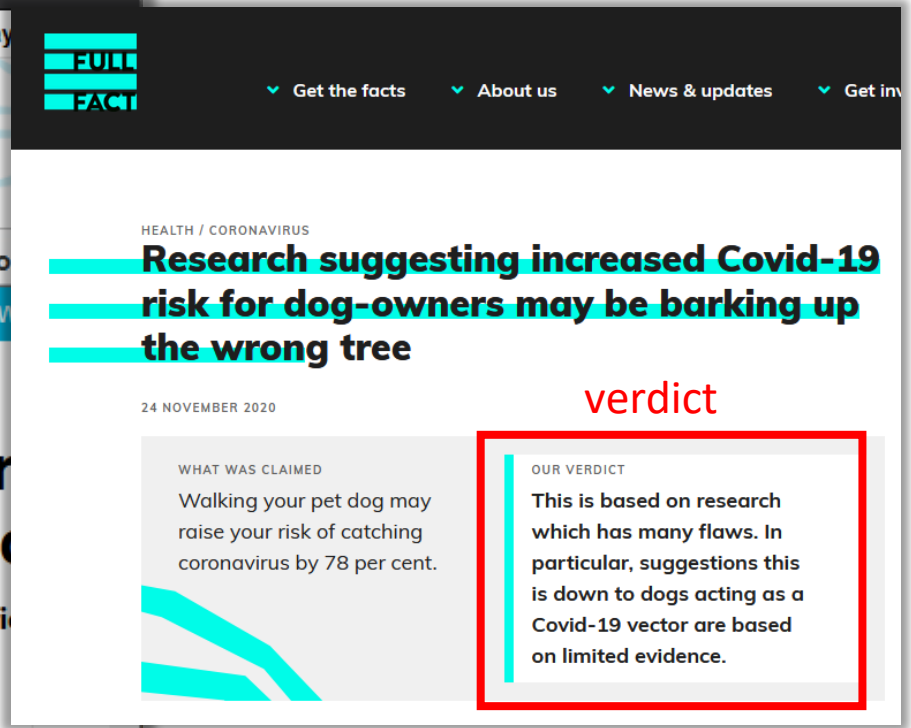
Conclusions

- **ClaimLinker**: linking text to fact-checked claims of a knowledge graph
  - Offering a **novel kind of semantic annotation** of unstructured content!
  - Fully unsupervised, modular, efficient, scalable, effective (**76% Precision@1**)
    - Outperforming existing unsupervised methods

- Through the claim URIs of an annotation, one can browse the knowledge graph and explore the claim-related **Linked Data**
  - For example, we can find other claims related to the same or similar entities, or claims originating from the same source

- Future Work:
  - Consider more advanced approaches to candidate generation and ranking (in order to better cater for the **vocabulary mismatch problem**)
    - E.g., Self-supervised (transformer-based) language models

Thank you!

Try **ClaimLinker** and give us feedback: http://users.ics.forth.gr/~fafalios/claimlinker/

Contact: **Pavlos Fafalios** (fafalios@ics.forth.gr)