



HAL
open science

FactCheckBureau: Build Your Own Fact-Check Analysis Pipeline

Oana Balalau, Pablo Bertaud-Velten, Younes El-Fraïhi, Garima Gaur, Oana Goga, Samuel Guimaraes, Ioana Manolescu, Brahim Saadi

► To cite this version:

Oana Balalau, Pablo Bertaud-Velten, Younes El-Fraïhi, Garima Gaur, Oana Goga, et al.. FactCheckBureau: Build Your Own Fact-Check Analysis Pipeline. CIKM 2024 - 33rd ACM International Conference on Information and Knowledge Management, ACM, Oct 2024, Boise Idaho, United States. 10.1145/3627673.3679220 . hal-04684068

HAL Id: hal-04684068

<https://inria.hal.science/hal-04684068>

Submitted on 2 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

FactCheckBureau: Build Your Own Fact-Check Analysis Pipeline

Oana Balalau
INRIA, Institut Polytechnique de Paris
Palaiseau, France

Pablo Bertaud-Velten
INRIA, Institut Polytechnique de Paris
Palaiseau, France

Younes El-Fraïhi
CNRS, Institut Polytechnique de Paris
Palaiseau, France

Garima Gaur
INRIA, Institut Polytechnique de Paris
Palaiseau, France

Oana Goga
INRIA, CNRS, Institut Polytechnique
de Paris
Palaiseau, France

Samuel Guimaraes
Federal University of Minas Gerais,
CNRS, Institut Polytechnique de Paris
Belo Horizonte, Brazil

Ioana Manolescu
INRIA, Institut Polytechnique de Paris
Palaiseau, France

Brahim Saadi
CNRS, Institut Polytechnique de Paris
Palaiseau, France

Abstract

Fact-checkers are overwhelmed by the volume of claims they need to pay attention to fight misinformation. Even once debunked, a claim may still be spread by people unaware that it is false, or it may be recycled as a source of inspiration by malicious users. Hence, the importance of *fact-check (FC) retrieval* as a research problem: given a claim and a database of previous checks, find the checks relevant to the claim. Existing solutions addressing this problem rely on the strategy of *retrieve and re-rank* relevant documents. We have built **FactCheckBureau**, an end-to-end solution that enables researchers to easily and interactively design and evaluate FC retrieval pipelines. We also present a corpus¹ we have built, which can be used in further research to test fact-check retrieval tools. The source code of our tool is available at this link².

CCS Concepts

• **Information systems** → *Learning to rank*.

Keywords

Fact Check Retrieval, Claim Review Datasets, Retrieve-and-rerank

ACM Reference Format:

Oana Balalau, Pablo Bertaud-Velten, Younes El-Fraïhi, Garima Gaur, Oana Goga, Samuel Guimaraes, Ioana Manolescu, and Brahim Saadi. 2024. FactCheckBureau: Build Your Own Fact-Check Analysis Pipeline. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)*, October 21–25, 2024, Boise, ID, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3627673.3679220>

1 Introduction

Social media and the Web are the central communication channels for public figures to disseminate information to many people. This

¹<https://huggingface.co/datasets/NaughtyConstrictor/fact-check-bureau>

²<https://gitlab.inria.fr/cedar/factcheckbureau>



This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

CIKM '24, October 21–25, 2024, Boise, ID, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0436-9/24/10

<https://doi.org/10.1145/3627673.3679220>

way of communicating influential and critical information without veracity evaluation has a far-reaching and adverse impact³. To sanitize the information space, *fact-checkers* assess a given claim for its correctness by evaluating evidence available in the public domain. The result of this investigation is a *fact-checking (FC, in short)* article that includes information about the *claim*, the *evidence*, and the *truthfulness* of the claim. A recent, comprehensive report on challenges raised by online fact-checking has been published by FullFact, a notable FC agency [5, 10]. The report highlights areas where automation can play a revolutionary role – finding check-worthy claims, detecting previously checked claims, retrieving evidence, and automated verification.

This work focuses on the task of automatically detecting previously checked claims. Often, false claims resurface on the Web in the same form or under linguistic disguises such as paraphrasing or translating to other languages. Not only false news resurface, but they also spread six times faster than correct news [21]. Therefore, it becomes crucial to quickly identify if a claim has already been fact-checked, leading to the FC retrieval (**FCR, in short**) problem: *Given a corpus of FCs C and a query document q (a social media post, an image specifying a claim, or textual description of a claim), retrieve a list of FCs from the corpus, in decreasing order of relevance (likelihood to be have checked this claim).*

Given the nature of claims and FCs, Machine Learning (ML) methods working on text and/or images have been studied for the FCR problem. They are typically harnessed in a *retrieve-and-rerank* paradigm where a model, called retriever, is used to filter out candidate FCs that are further investigated by a second model, the re-ranker, which computes a more precise relevance score [2, 4, 15, 17, 20].

We have developed the interactive system **FactCheckBureau** that supports researchers in *efficiently designing, deploying and analyzing FCR pipelines*. A user can specify, evaluate, and compare pipelines interactively. Further, we equipped FactCheckBureau with a *default FCR pipeline* and a *dataset* that we built to help non-technical users, mainly journalists. Using the querying interface of FactCheckBureau, non-technical users can fetch relevant FCs from our corpus against an input claim. Our system can be used to

³<https://www.weforum.org/press/2024/01/global-risks-report-2024-press-release/>, https://commission.europa.eu/news/boosting-awareness-raising-risks-disinformation-and-information-manipulation-2024-05-08_en

comprehensively test information retrieval techniques proposed in the literature, on a variety of datasets. Our dataset also contributes to the FCR task, in particular in the coverage of French.

2 Related Work

The need to reuse FCs when analyzing new statements has been recognized early on [3]. Accordingly, techniques emerged for retrieving the fact-checked claims most relevant for a given query (tweet or claim) [2, 4, 15, 17, 20]. All these adopt the standard (**retriever, reranker**) architecture proposed for information retrieval tasks. A retriever is used to retrieve efficiently, for a given input, from a potentially very large FCs, a subset considered to closest to it; then, a second, potentially more expensive method is used to re-rank the retrieved results. For both stages, there are *term-based (probabilistic) models*, and *neural* methods. Among the *probabilistic retrievers*, BM25 [16] is the most frequent, and performed well in many studies, e.g., [4, 17]. *Neural retrievers and re-rankers* are often based on Transformer networks, which capture matching signals between words in the FC and the query. In Table 1, we present the datasets and ranking methods used in prior works. We note that there are slight discrepancies on the best performing models. This could be due to the different evaluation settings, hence the importance of a unified evaluation and the relevance of our system.

Table 1: Partial snapshot of the front-runners in the area, adapted from [13]

| Dataset | Type | Languages | Source | # pairs | Evaluation | Best Model |
|---------|--------------------|----------------|----------------------------------|-------------------------------------|---|------------------------------------|
| [17] | Claim-Tweet Pairs | en | X, Snopes, PolitiFact, ClaimsKG | 1,768 | MRR MAP@k Precision@k | BM25 (Top50) + Sentence BERT |
| [7] | Claim Pairs | en,hi,bn,ml,ta | Whatsapp | 2,343 398 (en) | Accuracy Precision Recall F1 score | 1 XLM-R |
| [8] | FC-Tweet Pairs | en,hi,es,pt | X, GFC Tools | 6,533 4,850 (en) | Accuracy F1 score MRR MAP@k | Full Length BM25 |
| [12] | Claim-Thread Pairs | 41 (en,fr,...) | X, GFC tools | 26,048 42,88% (en) 3.46% (fr) | F1 score | Adapted GraphSAGE model |
| [11] | Claim-Tweet Pairs | ar,en | X, Snopes, AraFacts, ClaimsKG | 2,518 1,610 (en) | MRR MAP@k Hit@k | Sentence T5 + GPT Neo |
| [15] | FC-Post Pairs | 27 (en,fr,...) | X, Meta, GFC tools | 31,305 7,307 (en) 2,146 (fr) | Hit@k | GTR-T5 (en) MPNet |
| [18] | FC-Tweet Pairs | en,hi,es,pt | X, Boomlive, AFP,EFE, PolitiFact | 1,600 400 (en) | MRR MAP@k | BM25 (Top 200)+ LaBSE (BERT) |
| [4] | Claim-Tweet Pairs | en | [17] (Snopes) | 1,000 | MRR MAP@k Hit@k | BM25 (Top 100)+ BERT |

Claim normalization is used to improve the linguistic quality of claims and to help retrieve FCs [19]. Other related lines of work focus on identifying check-worthy claims [1]. ClaimKG [6] and CimpleKG [14] are corpora of fact-checks together with associated claims; the recent [15] is multilingual.

In this context, the interest of FactCheckBureau is to enable researchers and technicians working in fact-checking organizations to build and personalize their pipelines, experiment, and analyze with different modules.

3 FactCheckBureau at a glance

We describe the data sources we work on (Sec. 3.1), and the user interaction modes with our tool (Sec. 3.2).

3.1 Inputs: fact-checks and claims

Conceptually, there are two main entities: `fact-checks`, and `claims`. In principle, a claim can be a social media post, an image specifying a claim, or a simple text phrase. In the corpus we built for our demo (detailed in Section 4), claims are tweets. Therefore, our `claim` is characterized by its `accountHandle` (the Twitter account having published the tweet), `text`, `date`, `language`, `hashTags`, and `URLtoEmbeddedMediaContent`. The attributes of `fact-check` are derived from the `ClaimReview` schema that many FC agencies adopt in their FC articles (<https://schema.org/ClaimReview>); `ClaimReview` was promoted by Google, which used to show, next to search results, related FCs⁴. Specifically, the attributes of an FC are: `title`, `claimant`, `publisher`, `dateOfPublication`, `URLtoArticle`, `claimText`, `language` and `rating`. The relationship between these two key entities is captured by a many-to-many relation `claimAboutFC(claim-id, FC-id)`: a claim and an FC are paired in this way, if (according to a specific automated or manual decision method) the FC is about the claim. We also say the claim and FC are **aligned**.

3.2 Users' interactions with FactCheckBureau

The core task to be solved in FC retrieval pipelines is: given a claim (also called query) and an FC corpus, find the FCs most relevant to the claim. Specifically, each FC retrieval pipeline contains a *candidate FC retrieval* module and a *candidate ranking* module. If the claim is text, it can be used as such, but other formats may require some *pre-processing*, based on the query type, before being entered into the text-based retrieval pipeline. For instance, if the claim is an image, the text needs to be extracted by OCR, or the image can be captioned; if the claim is a tweet, pre-processing may remove or split hashtags in individual words, normalize numerical data, transcribe emojis to text, etc. To be comprehensive, FactCheckBureau models FC retrieval pipelines as consisting of three stages: pre-processing, retrieval, and ranking.

FactCheckBureau has **two main operation modes: deployment and development**, shown in Figure 1, where dark navy modules are used in deployment, whereas in development, all the modules (both navy and light blue) may be involved. In development mode, it supports designing, inspecting, and comparing FC retrieval pipelines; in deployment mode, a retrieval pipeline can be deployed and used to query the FC corpus. As explained below, our demonstration will showcase the four use cases (design, inspect, compare, and deploy).

Inspect. A user builds a retrieval pipeline by choosing or loading: pre-processing modules; a retrieval module; and a ranking module. The user also supplies aligned pairs, and chooses the metric(s) to use to evaluate the quality of the pipeline. Since relevant FC retrieval is a ranked-list search problem, we support the familiar Mean Average Precision (`MAP@k`), Mean Reciprocal Ranking (`MRR@k`), Normalized Discounted Cumulative Gain (`NDCG@k`), and `Hits@k`. The user triggers the *evaluation* of the pipeline (top snippet in Figure 2); for each query, this leads to a list of FCs, ordered by their relevance, the former as computed by the pipeline.

⁴That had been discontinued, among others, because some of the shown FC were not semantically close enough to the respective search results [9]. This highlights the importance of the FC retrieval problem.

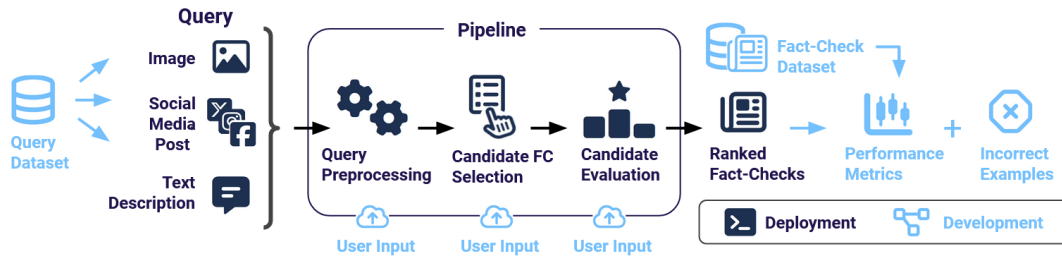


Figure 1: FactCheckBureau architecture in the two use modes (Development and Deployment)

FactCheckBureau also presents the values of the chosen metric(s) for different cut-off values k .

For further inspection, a user can choose the *deep-dive* option, where FactCheckBureau enables to inspect test samples where the specified pipeline performed poorly. It also reports the performance of each model in the pipeline in isolation.

- **Input:** Chose models, aligned pairs, metric(s);
- **Output:** Computed metric (plots and tables), the performance of candidate selector only, identify top 5 badly performing examples.

Compare. While developing a retrieval solution, evaluating and comparing different models is essential to find the best possible combination of models for candidate retrieval and ranking. The *compare* mode (middle in Figure 2) enables a user to compare previously *saved* pipelines and/or newly specified ones. The user obtains a consolidated performance report comparing all the specified pipelines under a set of chosen metrics. It also provides a *deep-dive* option, to compare models' performance on selected test samples.

- **Input:** Choose pipelines form a list of pipelines;
- **Output:** Single plot of overall performance, plot of candidate identifier performance, 5 worst performance instances.

Design. This option is for users who do not intend to develop a pipeline but need to use one. The user can supply an FC corpus, or a default one (the one we prepare for the demo, Section 4) can be used. The user specifies the claim language (or we can auto-detect it), and claim type (post, image, or text).

- **Input:** Specify query type (post, image, text), and dataset language;
- **Output:** A recommended pipeline based on (i) the most frequently used components for these inputs, or comparable inputs (same language, same query type) if there is no history of running on the same inputs; (ii) simple rules to chose the necessary pre-processing models based on the input type.

Deploy. FactCheckBureau can be used as a search interface (bottom in Figure 2) for finding the relevant FCs for an input query, or, alternatively, for a specific *topic*, specified as a short phrase, e.g., "Covid". The user chooses a previously specified retrieval pipeline already present in the system, and configure the number of relevant documents she wants to retrieve. Then, FactCheckBureau returns a list of the FCs relevant for the claim, respectively, FCs about the given topic.

- **Input:** Choose a pipeline, or select "auto-design"; specify query or topic;

- **Output:** Querying interface – querying through post, image, text, or topic (will use FC tags available).

4 Dataset and FC pipeline

Dataset. We built a corpus of 218K claim reviews in 14 languages published by 83 fact-checking agencies recognized as verified signatories by IFCN (International Fact-Checking Network). Further, we have collected 9.1K tweets mentioned in various FC articles and 8K recent tweets from prominent Members of the European Parliament.

We used Google Fact-Check API⁵ for collecting FCs. The data returned follows the ClaimReview schema but with some fields omitted as described in Google's documentation⁶. The returned data also includes the URL to the FC article. We also collected the FC article text using these URLs to enrich our dataset. Some of the reputed FC agencies, like Le Monde, do not publish their FCs via Google FC Explorer, or stopped at some point, therefore, we crawled their web pages to collect FCs. For social media post collection, we used a paid subscription of X to gather the 8K tweets from 402 MEP. We focused on social media posts in English and French FCs articles for the aligned pair collection. We collected tweets mentioned in 4.7K English FCs, 1.2K French FCs and 3.2K FC in other languages. This resulted in 9.1K aligned pairs of social media posts and relevant FC articles, as many FC has two tweets related to it. For some of our FC retrieval experiments described next, we will consider these pairs to be the ground truth, allowing the search of a FC by given its paired tweet. For others, the text of the claim described on the ClaimReview schema of the FC is used as the its ground-truth pair.

Pipeline. We preload FactCheckBureau with our proposed FC retrieval pipeline that supports the default query interface for non-technical users. In our FC retrieval setting, the collection of FC *articles* serves as the document corpus and tweets serve as the input query. We experiment with around 41K articles in English and French. Our pipeline starts with pre-processing a tweet by *removing links, emojis, escape control and special characters, standarizing Unicode presentation with more than one representation, normalizing numbers and dates* using `num2words`⁷ and `dateparser`⁸ libraries, and *tokenizing text* using `Spacy`⁹ tokenizer. We employed the well-established BM25

⁵<https://toolbox.google.com/factcheck/apis>

⁶<https://tinyurl.com/25t28phf>

⁷<https://pypi.org/project/num2words/>

⁸<https://pypi.org/project/dateparser/>

⁹<https://spacy.io/api/tokenizer>

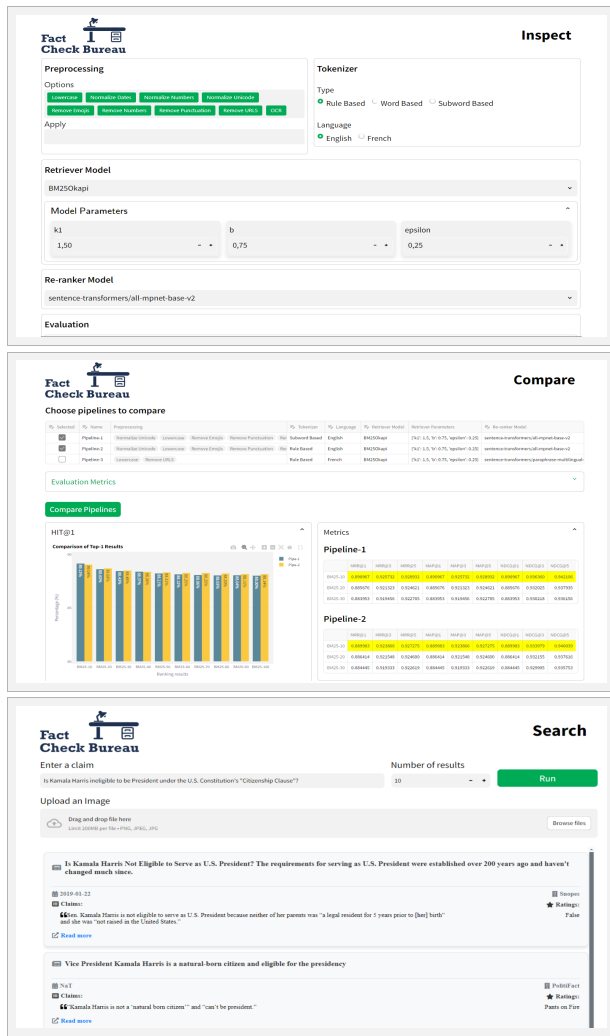


Figure 2: FactCheckBureau in Development mode

[16] as our retriever model and used all-mpnet-base-v2 and paraphrase-multilingual-mpnet-base-v2 for re-ranking documents in English and French respectively.

Discussion. We discuss some insights we had while developing our pipeline highlighting the non-trivial nature of FC retrieval problem:

- I1** Retrieving FC *articles* is better than *claims* specified in ClaimReview: Using our pipeline, we evaluated the performance for two retrieval tasks: first, given a post, retrieve claims specified in ClaimReview (P2C) and, second, for an input post, retrieve the corresponding FC article (P2A), on our datasets and a recent multilingual FC dataset MultiClaim [15]. We computed the $Hit@k$ metric that indicates the percentage of samples for which the correct answer is present in the top- k answers of the re-ranker model. For a given k , the $Hit@k$ metric is computed on different sizes of candidate FCs list retrieved by our retriever model BM25. A candidate list of size J is referred to as BM25- J . We report $Hit@5$ at different BM25- J for $J = 10, 20, 50, 100$ on both datasets in

Table 2: Comparison of two tasks P2C (post to claim) and P2A (post to article) on top-5 results of our pipeline for different candidate list size.

| Dataset | Task | BM25-10 | BM25-20 | BM25-50 | BM25-100 |
|-------------|------|---------|---------|---------|----------|
| MultiClaim | P2C | 67.94 | 71.14 | 73.53 | 74.10 |
| | P2A | 86.67 | 86.02 | 85.69 | 84.94 |
| Our Dataset | P2C | 63.95 | 67.30 | 70.19 | 71.23 |
| | P2A | 85.75 | 85.27 | 84.21 | 83.41 |

Table 3: Evaluation of the impact of different data augmentation features, OCR, and image captioning (IC) used for both datasets. The $Hit@k$ metric is computed over the empirically best FC candidate list size 10, i.e., BM25-10.

| Dataset | Features | Hit@1 | Hit@3 | Hit@5 | Hit@10 |
|------------------|--------------|-------|-------|-------|--------|
| Our Dataset (EN) | T | 73.67 | 83.65 | 86.35 | 88.67 |
| | T + OCR | 76.00 | 86.33 | 88.69 | 90.82 |
| | T + IC | 76.55 | 85.92 | 88.42 | 90.90 |
| | T + OCR + IC | 73.40 | 82.79 | 85.94 | 88.52 |
| Our Dataset (FR) | T | 72.41 | 85.22 | 88.57 | 92.33 |
| | T + OCR | 74.04 | 86.53 | 90.29 | 93.71 |
| MultiClaim (EN) | T | 75.05 | 84.43 | 86.67 | 88.53 |
| | T + OCR | 85.93 | 94.41 | 96.21 | 97.28 |
| MultiClaim (FR) | T | 74.82 | 86.84 | 90.09 | 93.18 |
| | T + OCR | 78.93 | 91.0 | 94.08 | 97.04 |

Table 2. The performance of P2C retrieval tasks improves as more candidates FC are considered for re-ranking, whereas this trend is inverse for the P2A task. This implies retrieving FC articles is not only better performing (improvement of 12 points), it is also more efficient as the computationally expensive re-ranker model needs to re-rank a significantly smaller (by a factor of $X10$) candidate list.

- I2** Careful data augmentation helps in retrieval: We experimented with augmenting the tweets with the features extracted from the images embedded within the tweets – OCR and image caption. We evaluated the impact of different augmentation over English and French-aligned pairs of our dataset and that of MultiClaim [15]. The results are reported in Table 3 where different setups are indicated as tweet text (T), text appended with OCR text (T+OCR), tweet text appended with image caption (T+IC), and a combination of both the features (T+OCR+IC). For space consideration, we have reported the best augmentation feature for the French subset of our dataset and the MultiClaim dataset. Our main takeaway from this experiment is that tweets need to be carefully augmented, and not necessarily more features ensure better performance.

5 Conclusion

In this work, we present FactCheckBureau, an interactive tool for designing and deploying configurable FC retrieval pipelines. The designing task supports researchers in specifying and comparing the performance of different pipelines following the retrieve-and-rerank paradigm. Along with the tool, we proposed an FC retrieval pipeline and constructed an FC dataset that backs the search interface for non-technical users.

Acknowledgments. This work is supported in parts by the EU project ELIAS (101120237) grant, EU 101041223 grant, French AI Chair SourcesSay (ANR-20-CHIA-0015), and Hi!PARIS Center.

References

- [1] Fatma Arslan, Naeemul Hassan, Chengkai Li, and Mark Tremayne. 2020. A Benchmark Dataset of Check-worthy Factual Claims. In *14th International AAAI Conference on Web and Social Media*. AAAI.
- [2] Alberto Barrón-Cedeño, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari, Nikolay Babulkov, Bayan Hamdan, Alex Nikolov, Shaden Shaar, and Zien Sheikh Ali. 2020. Overview of CheckThat! 2020: Automatic Identification and Verification of Claims in Social Media. In *CLEF*. Springer. https://doi.org/10.1007/978-3-030-58219-7_17
- [3] Sylvie Cazalens, Philippe Lamarre, Julien Leblay, Ioana Manolescu, and Xavier Tannier. 2018. A Content Management Perspective on Fact-Checking. In *The Web Conference*. ACM. <https://doi.org/10.1145/3184558.3188727>
- [4] Tanmoy Chakraborty, Valerio La Gatta, Vincenzo Moscato, and Giancarlo Sperl'i. 2023. Information retrieval algorithms and neural ranking models to detect previously fact-checked information. *Neurocomputing* 557 (2023), 126680. <https://doi.org/10.1016/j.neucom.2023.126680>
- [5] FullFact. 2020. The challenges of online fact checking. <https://fullfact.org/media/uploads/coof-2020.pdf>.
- [6] Susmita Gangopadhyay, Katarina Boland, Danilo Dessi, Stefan Dietze, Pavlos Fafalios, Andon Tchechmedjiev, Konstantin Todorov, and Hajira Jabeen. 2023. Truth or Dare: Investigating Claims Truthfulness with ClaimsKG. In *Linked Data-driven Resilience Research*, Vol. 3401. <https://ceur-ws.org/Vol-3401/paper7.pdf>
- [7] Ashkan Kazemi, Kiran Garimella, Devin Gaffney, and Scott Hale. 2021. Claim Matching Beyond English to Scale Global Fact-Checking. In *IJCNLP*. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.acl-long.347>
- [8] Ashkan Kazemi, Zehua Li, Verónica Pérez-Rosas, Scott A Hale, and Rada Mihalcea. 2022. Matching tweets with applicable fact-checks across languages. In *De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection (with AAAI)*.
- [9] Emma Lurie and Eni Mustafaraj. 2020. Highly Partisan and Blatantly Wrong: Analyzing News Publishers' Critiques of Google's Reviewed Claims. In *Truth and Trust Online Conference*. Hacks Hackers. <https://truthandtrustonline.com/wp-content/uploads/2020/10/TTO07.pdf>
- [10] Preslav Nakov, David Corney, Maram Hasanain, Firoj Alam, Tamer Elsayed, Alberto Barrón-Cedeño, Paolo Papotti, Shaden Shaar, and Giovanni Da San Martino. 2021. Automated Fact-Checking for Assisting Human Fact-Checkers. In *IJCAI*. <https://doi.org/10.24963/ijcai.2021/619>
- [11] Preslav Nakov, Giovanni Da San Martino, Firoj Alam, Shaden Shaar, Hamdy Mubarak, and Nikolay Babulkov. 2022. Overview of the CLEF-2022 CheckThat! lab task 2 on detecting previously fact-checked claims. (2022).
- [12] Dan S Nielsen and Ryan McConville. 2022. Mumin: A large-scale multilingual multimodal fact-checked misinformation social network dataset. In *SIGIR*.
- [13] Rubaa Panchendrarajan and Arkaitz Zubiaga. 2024. Claim detection for automated fact-checking: A survey on monolingual, multilingual and cross-lingual research. *Natural Language Processing Journal* 7 (2024), 100066. <https://doi.org/10.1016/j.nlp.2024.100066>
- [14] Youri Peskine, Raphaël Troncy, and Paolo Papotti. 2024. CimpleKG: a Continuously Updated Knowledge Graph of Fact-Checks and Related Misinformation. In *Infoc sur Seine workshop*.
- [15] Mátús Pikuliak, Ivan Srba, Róbert Móro, Timo Hromadka, Timotej Smolen, Martin Melisek, Ivan Vykopal, Jakub Simko, Juraj Podrouzek, and Mária Bieliková. 2023. Multilingual Previously Fact-Checked Claim Retrieval. In *EMNLP*. <https://doi.org/10.18653/V1/2023.EMNLP-MAIN.1027>
- [16] Stephen Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval* 3, 4 (2009). <https://doi.org/10.1561/15000000019>
- [17] Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, and Preslav Nakov. 2020. That is a Known Lie: Detecting Previously Fact-Checked Claims. In *ACL*. <https://doi.org/10.18653/V1/2020.ACL-MAIN.332>
- [18] Iknor Singh, Carolina Scarton, Xingyi Song, and Kalina Bontcheva. 2023. Finding Already Debunked Narratives via Multistage Retrieval: Enabling Cross-Lingual, Cross-Dataset and Zero-Shot Learning. *arXiv:2308.05680* (2023).
- [19] Megha Sundriyal, Tanmoy Chakraborty, and Preslav Nakov. 2023. From Chaos to Clarity: Claim Normalization to Empower Fact-Checking. In *EMNLP Findings*. <https://doi.org/10.18653/V1/2023.FINDINGS-EMNLP.439>
- [20] Nguyen Vo and Kyumin Lee. 2020. Where Are the Facts? Searching for Fact-checked Information to Alleviate the Spread of Fake News. In *EMNLP*. <https://doi.org/10.18653/V1/2020.EMNLP-MAIN.621>
- [21] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359 (2018). <https://doi.org/10.1126/science.aap9559> arXiv:<https://www.science.org/doi/pdf/10.1126/science.aap9559>