Multilingual and Crosslingual Fact-Checked Claim Retrieval

Atharva Nijasure
Debrup Das
Md Zarif Ul Alam

ACM Reference Format:
Atharva Nijasure, Debrup Das, and Md Zarif Ul Alam. 2018. Multilingual and Crosslingual Fact-Checked Claim Retrieval. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Abstract

In the era of rapid misinformation spread across social media, this research tackles the critical challenge of multilingual fact-checking by addressing two fundamental research questions: (1) How effective are monolingual versus cross-lingual models in retrieving accurate fact-checks for input claims in diverse languages? and (2) What impact does external knowledge augmentation have on fact-check retrieval accuracy? Our novel approach introduces two unique methods of knowledge augmentation: (i) enriching fact-checked documents using parametric knowledge from Large Language Models (LLaMa 3.2), and (ii) extracting and summarizing contextual information from external URLs linked in fact-check documents [16]. Utilizing the MultiClaim dataset spanning 27 languages, we conducted comprehensive experiments across multiple retrieval models, evaluating performance using Precision, Mean Average Precision, and Normalized Discounted Cumulative Gain metrics. Our experimental results reveal significant challenges in cross-lingual fact-checking retrieval, with monolingual models consistently outperforming cross-lingual approaches. While our first approach of LLM-based claim title augmentation did not meet expectations, the second approach of leveraging external URL evidence showed promising improvements in retrieval accuracy. The complete implementation is publicly available in our GitHub repository[15], with the original dataset accessible via our shared Google Drive folder[6].

2 Problem Statement

The objective of this project is to develop a system for **Multilingual Previously Fact-Checked Claim Retrieval**. This task aims to identify the most relevant previously fact-checked document for a given input claim in a social media post (SMP), even when the fact-checked documents are in multiple languages. The goal is to retrieve a fact-check document that has already been validated by experts to determine its relevance to the claim. This setup will allow us to evaluate retrieval performance in both **monolingual** and **cross-lingual** contexts, leveraging the dataset from the MultiClaim paper for experimentation.

Let:

- $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$ represent a set of social media posts (input claims).
- $\mathcal{F} = \{f_1, f_2, \ldots, f_n\}$ denote a set of previously fact-checked documents (fact-checked claims), where each f_i may be in a different language.

For each social media post $s \in \mathcal{S}$, we aim to retrieve the most relevant fact-checked document $f^* \in \mathcal{F}$ that maximizes relevance according to a relevance function $R(s, f)$, where:

$$f^* = \arg\max_{f \in \mathcal{F}} R(s, f)$$

Our research question focuses on evaluating the effectiveness of multilingual retrieval techniques for identifying cross-lingual connections between input claims and fact-checked documents. Specifically, we aim to address:

- How effective are monolingual versus cross-lingual models in retrieving accurate fact-checks for input claims in diverse languages?
- What impact does external knowledge augmentation in fact-check documents have on retrieval accuracy?

This research aims to introduce novel methods to enhance multilingual retrieval, especially in the context of fact-checking across languages.

3 Motivation

The global spread of disinformation poses a significant challenge to the credibility of information consumed by the public. With the advent of social media sites, false claims can spread rapidly, becoming viral and reaching audiences across *diverse linguistic backgrounds*. Addressing misinformation and verifying claims has been particularly challenging for fact-checkers in **low resource language settings**. Previous fact-checking work has been largely limited to monolingual settings, primarily focusing on English. In this project, we aim to reduce this linguistic bias and make fact-checking retrieval systems more advanced and inclusive in multilingual and cross-lingual contexts, in an effort to curb the global spread of misinformation, cutting across language barriers.

4 Related Work

Methods used for PFCR are usually either BM25 (and other similar information retrieval algorithms) or various text embedding-based approaches [20, 23]. Reranking is often used to combine several methods to side-step compute requirements or as a sort of ensembling [19]. PFCR task is also a target of the *CLEF's CheckThat!* challenge, with many teams contributing with their solutions [14].

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

	Input claims	FC claims	Pairs	Languages
Kazemi et al. [12]	*NA*	150,000	258	5
Jiang et al. [10]	*NA*	90	1,573	1
Shaar et al. [21]	2,259	44,164	2,440	2
Hardalov et al. [9]	316,564	10,340	332,660	1
MultiClaim [16]	28,092	205,751	31,305	27/39

Table 1: PFCR datasets. FC claims are *fact-checked. NA* means that we were not able to identify the correct number of input claims. The number should be similar to the number of pairs in most cases.

Other methods use visual information from images [13, 24], abstractive summarization [7], or key sentence identification [22] to improve the results.

PFCR datasets are summarized in Table 1. MultiClaim [16] dataset has the highest number of fact-checked claims. It also has the second-highest number of input claims and pairs after CrowdChecked, but that dataset is significantly noisier. Finally, this dataset has by far the most languages, while the second biggest dataset in this regard has 5 language with only 50 samples per language.

5 Approach

5.1 Main Approaches

Social media post claims often lack important *contextual* details, making it challenging to retrieve the most relevant and up-to-date fact-check documents. Our main approach to solve this limitation, is to augment the claim and the fact checked documents with **external knowledge**. External knowledge provides essential context and supplementary information, enabling the system to bridge these gaps by enhancing **contextual understanding**. We use two approaches in this regard:

5.1.1 ***Enriching fact-checked claims with LLM parametric knowledge***. In the dataset, we observed that, for the baseline experiments, the fact-checked documents were encoded using information consisting of the previously fact-checked claim sentence denoted as $factchecked\_claim$ and the title of the fact-checked claim document denoted as $title$. Our retrieval models lacked deeper contextual knowledge associated with the fact-checked sources which resulted in a drop in performance. To improve the quality of our retrieval models, we proposed expanding the fact-checked documents by utilizing an LLM such as LLaMa3.2 1B[1] LLM to generate enriched textual representations after processing the claim and title.

Let $F = \{f_1, f_2, \ldots, f_n\}$ represent the original set of fact-checked documents, where each document f_i is composed of:

$$f_i = \text{concat}(\text{factchecked\_claim}_i, \text{title}_i)$$

Here:

- titles_i: Extracted titles of the fact-checked documents.
- $\text{factchecked\_claims}_i$: A single sentence representing the verified previously fact-checked claim present in the document.

[1]For more information, see https://huggingface.co/meta-llama/Llama-3.2-1B

To expand F, we introduce additional context (denoted as context_i), generated using the LLaMa3.2 1B[2] LLM. . The expanded document f_i' is defined as:

$$f_i' = \text{concat}(f_i, \text{context}_i)$$

where:

$$\text{context}_i = \text{LLM}(\text{Prompt} : \text{factchecked\_claim}_i, \text{title}_i)$$

This expanded dataset F' is represented as:

$$F' = \{f_1', f_2', \ldots, f_n'\}$$

The enriched dataset F' leverages the additional context generated by the LLM, enabling the retrieval models to perform better by incorporating supplementary information that was not available in the original dataset. The choice of prompt and LLM was finalized based on a few experiments conducted to evaluate the quality of the augmented documents, as well as resource availability like total time for augmenting all documents, etc.

5.1.2 ***Using external urls linked in fact-check documents***. All fact-check documents include *URLs* leading to external sources containing vital *textual* and *visual* evidence as shown in Figure 1. We use text scraping tools such as *BeautifulSoup* to extract the raw

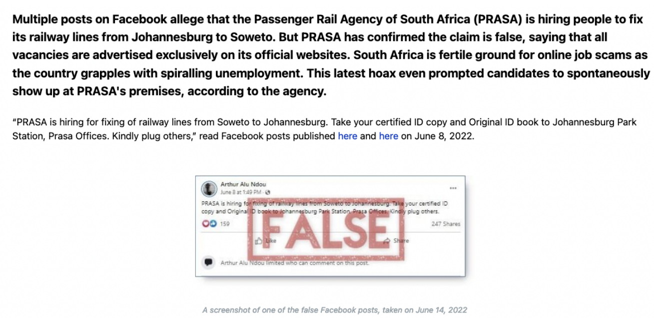

Figure 1: An example fact-checked document from the dataset.

text from the linked web pages. Using LLMs such as **Llama-3.2**, the scraped text is converted into concise summaries, focusing on details directly relevant to the claims. The new summaries along with original claim and title are then encoded and are indexed for retrieval. Like previous approach, the expanded document f_i' is defined as:

$$f_i' = \text{concat}(f_i, \text{summarized}_i)$$

where:

$$\text{summarized}_i = \text{LLM}(\text{Prompt} : \text{factchecked\_claim}_i, \text{title}_i, \text{URL page:}\langle\text{Scraped text content}\rangle)$$

Model	P@3	P@5	P@10	MAP@3	MAP@5	MAP@10	NDCG@3	NDCG@5	NDCG@10
BM25 [1]. (Monolingual)	0.1171	0.0834	0.0503	0.2198	0.2325	0.2429	0.2441	0.2641	0.2860
BM25 (Crosslingual)	0.0695	0.0471	0.0265	0.1428	0.1482	0.1518	0.1549	0.1637	0.1716
LaBSE [4] (Monolingual)	0.1013	0.0744	0.0449	0.1792	0.1926	0.2016	0.2029	0.2244	0.2439
LaBSE (Crosslingual)	0.0811	0.0581	0.0349	0.1617	0.1714	0.1786	0.1783	0.1946	0.2105
msmarco-bert-codensor [5] (Monolingual)	0.1460	0.1122	0.0681	0.2653	0.2896	0.3046	0.2986	0.3368	0.3678
msmarco-bert-codensor (Crosslingual)	0.0614	0.0434	0.0253	0.1210	0.1279	0.1324	0.1335	0.1449	0.1548
Facebook DPR [3] (Monolingual)	0.0634	0.0472	0.0295	0.1186	0.1276	0.1348	0.1332	0.1483	0.1644
Facebook DPR (Crosslingual)	0.0129	0.0100	0.0061	0.0259	0.0285	0.0298	0.0287	0.0331	0.0362
distiluse_ml_v2 [2] (Monolingual)	0.1424	0.1068	0.0639	0.2606	0.2821	0.2953	0.2922	0.3254	0.3523
distiluse_ml_v2 (Crosslingual)	0.0840	0.0616	0.0376	0.1667	0.1781	0.1865	0.1840	0.2037	0.2226

Table 2: Baseline evaluations for different models on monolingual and crosslingual splits.

5.2 Baselines

We will adopt baseline methods from related work, particularly those outlined in the MultiClaim paper, as they are well-suited to our task of Multilingual Previously Fact-Checked Claim Retrieval. Specifically, we plan to also use machine translation to convert non-English fact-checks and claims into English, enabling the application of widely used English-based retrieval models.

We will apply two retrieval models as baselines:

(1) BM25: [18] a traditional sparse retrieval method based on term frequency and inverse document frequency, which has shown robust performance in various retrieval tasks.

(2) Dense Retrieval Models: a dense retrieval model using neural embeddings to improve semantic matching, particularly useful for matching claims with fact-checks even when wording differs.

- DPR (*Dense Passage Retrieval*) [11] with pre-trained BERT or multilingual LM to handle cross-lingual retrieval for fact verification. We use different Dense retrieval models as shown in Table 2.
- LaBSE (*Language-Agnostic BERT Sentence Embedding*) [8]: It is designed for robust multilingual semantic similarity tasks, supporting over 100 languages. LaBSE is particularly effective for tasks like cross-lingual retrieval and multilingual sentence similarity, making it suitable for claim-fact-check applications across languages.
- *Msmarco-bert-co-condensor* [5]: It utilizes the *Condenser pre-training architecture*, which learns to condense information into the dense vector through LM pre-training. On top of it, coCondenser adds an unsupervised corpus-level contrastive loss to warm up the passage embedding space.
- *Distiluse-base-multilingual-cased-v2* [17]: Multilingual knowledge distilled version of multilingual Universal Sentence Encoder (USE). Supports 15 languages: Arabic, Chinese, Dutch, English, French, German, Italian, Korean, Polish, Portuguese, Russian, Spanish, Turkish.

These baselines allows us to evaluate the effectiveness of our proposed approaches against established methods in both monolingual and cross-lingual retrieval setups. For the baseline experiments, we computed MAP, Precision, and NDCG at thresholds of 3, 5, and 10. The results are summarized in Table 2.

6 Experiments

6.1 Dataset

Our project utilizes an enhanced version of the **MultiClaim** dataset, which was originally introduced as a comprehensive multilingual benchmark for cross-lingual *previously fact-checked claim retrieval* (**PFCR**). MultiClaim stands as the largest and most linguistically diverse dataset of its kind, designed to support tasks related to fact-checking and claim retrieval across multiple languages.
The *MultiClaim* dataset comprises three main components:

- **Fact-checks:** A total of 205K professionally verified fact-checks written by expert fact-checkers, spanning 39 languages. These fact-checks cover a broad spectrum of topics and linguistic diversity, providing a rich resource for multilingual PFCR tasks.
- **Social Media Posts:** A collection of 28K posts sourced from social media platforms, covering 27 different languages. These posts represent user-generated text, containing text making an input claim.
- To establish the PFCR framework, MultiClaim includes **31K verified connections** between the social media posts and fact-checks, linking user-generated claims to corresponding fact-checks.

6.2 Evaluation Protocol

Performance was measured using three standard IR metrics at different cut-off points ($k = 3, 5, 10$):

- Precision@k (P@k)
- Mean Average Precision@k (MAP@k)
- Normalized Discounted Cumulative Gain@k (NDCG@k)

6.3 Experimental Details

We evaluated five baseline models - *BM25* , *LaBSE (Language-agnostic BERT Sentence Embedding)*, *Msmarco-bert-codensor*, *Facebook DPR*, and *Distiluse-multilingual-v2* model - in both monolingual and cross-lingual settings. The experiments were designed to assess retrieval performance using standard information retrieval metrics. We implemented the pipelines from scratch, and didn't use existing codes. The *pyterec-eval* library was used for calculating the evaluation metrics. We have made the GitHub repository public during the final submission period. Almost all of the code in the repository [15] was implemented by us. We did use the LLMs made

Model	P@3	P@5	P@10	MAP@3	MAP@5	MAP@10	NDCG@3	NDCG@5	NDCG@10
distiluse_ml_v2 [2] (Crosslingual)	0.079	0.058	0.036	0.158	0.169	0.177	0.175	0.193	0.213
msmarco-bert-codensor [5] (Crosslingual)	0.057	0.040	0.023	0.120	0.1263	0.130	0.131	0.140	0.150
Facebook DPR [3] (Crosslingual)	0.010	0.007	0.005	0.020	0.022	0.023	0.022	0.025	0.029
Facebook DPR [3] (Monolingual)	0.061	0.045	0.028	0.117	0.126	0.131	0.130	0.145	0.158
msmarco-bert-codensor [5] (Monolingual)	0.141	0.106	0.064	0.261	0.283	0.297	0.292	0.326	0.356
distiluse_ml_v2 [2] (Monolingual)	0.006	0.006	0.004	0.010	0.012	0.013	0.012	0.015	0.019

Table 3: Approach by augmenting Fact-check claim titles with external knowledge from Llama-3.2-1B-instruct

Model	P@3	P@5	P@10	MAP@3	MAP@5	MAP@10	NDCG@3	NDCG@5	NDCG@10	Gain
LaBSE [4]	0.171	0.107	0.057	0.460	0.465	0.470	0.472	0.483	0.492	–
LaBSE [4] (**augmented**)	0.203	0.128	0.067	0.556	0.563	0.567	0.569	0.581	0.590	**+19.92%**
msmarco-bert-co-condensor [5]	0.178	0.114	0.061	0.479	0.487	0.492	0.492	0.506	0.518	–
msmarco-bert-co-condensor (**augmented**)	0.207	0.130	0.069	0.560	0.567	0.571	0.574	0.587	0.598	**+15.44%**
distiluse-base-multilingual-cased-v2 [17]	0.187	0.117	0.062	0.495	0.501	0.505	0.511	0.521	0.531	–
distiluse-base-multilingual-cased-v2 (**augmented**)	0.205	0.129	0.068	0.544	0.551	0.556	0.561	0.574	0.586	**+10.35%**
Facebook DPR [3]	0.076	0.052	0.031	0.196	0.203	0.210	0.204	0.218	0.234	–
Facebook DPR [3] (**augmented**)	0.080	0.055	0.032	0.203	0.210	0.217	0.212	0.226	0.243	**+3.84%**

Table 4: Augmentation approach using external URLs linked in fact-check documents, summarized by Llama-3.1-8B-it .Gains in NDCG@10 are shown for models with augmented scraped content compared to their baselines. Evaluations on smaller subset.

available for free by OpenAI and Anthropic to debug and modify certain parts of our code.

7 Results

The experimental results from Table 2 demonstrates varying performance across different models and settings for our baselines. In the monolingual setting, **msmacro-bert-codensor** emerged as the top performer, achieving the highest scores across all metrics (P@3 = 0.146, MAP@10 = 0.306, NDCG@10 = 0.368). This was followed by **BM25**, which showed strong performance particularly in precision metrics (P@3 = 0.117, MAP@10 = 0.243, NDCG@10 = 0.286). In the cross-lingual scenario, **Distiluse-multilingual-v2** model demonstrated the strongest performance (P@3 = 0.084, MAP@10 = 0.1865, NDCG@10 = 0.222), closely followed by the **LaBSE** model. Both of these models notably outperformed other models including BM25 and Facebook DPR. The performance gap between monolingual and cross-lingual settings is substantial, with cross-lingual scores being consistently lower across all metrics and models. This difference highlights the inherent challenges in cross-lingual information retrieval.

Augmentation with LLMs using titles of Fact checked docs. In this setting, we present the results obtained by augmenting *external knowledge* from an LLM by prompting it with the title of the fact-checked document. The **Llama-3.2-1B-instruct** model was utilized for this experiment. The results for both monolingual and cross-lingual splits of the dataset are presented in Table 3.

Augmentation Approach by scraping URLs of fact-checking documents. In this approach, we augment the document corpus of previously fact-checked claims by scraping the Web URLs accompanying the documents. We scraped the contents of the web page using the *Beautiful Soup* library, which resulted in noisy information content. Hence, we summarized the the scraped contents using **Llama-3.1-8B-instruct** to extract the most relevant information related to the document. The results obtained on a smaller subset of the dataset (with and without augmentation) are presented in Table 4.

8 Analysis

Using external urls linked in fact-check documents Our experiments are performed on a smaller cross-lingual subset of the dataset. For every model evaluated (**LaBSE, msmarco-bert-co-condensor, distiluse-base-multilingual-cased-v2, and Facebook DPR**), the augmented settings consistently outperform their non-augmented counterparts, in terms of *NDCG@10* and other metrics such as *Precision* and *MAP*. The largest improvements in NDCG@10 were observed for **LaBSE** and **msmarco-bert-co-condensor** models, showing gains of **+19.92%** and **+15.44%**, respectively. This indicates that external augmentation had the most substantial positive impact on these models. Facebook DPR had the smallest gain, which can be partly explained by the of evaluating an English-based model on a cross-lingual split. The results demonstrate that augmenting the document corpus with externally scraped information significantly enhances retrieval performance. The improvements in retrieval accuracy indicate that external knowledge augmentation provides *critical context* and additional relevant information to the models, enabling them to better rank factually accurate documents. Despite the noisy nature of scraped web content, summarizing the content with Llama-3.1-8B-instruct successfully distills the relevant facts and improves document relevance in retrieval tasks.

Using External Parametric Knowledge of LLMs to Augment Titles and Claims of Fact-Checked Documents Our experiments were conducted on the complete dataset of both cross-lingual and multilingual fact-checked documents. The performance observed with this approach did not surpass the baseline across all evaluation metrics. We believe several factors contributed to the unexpected results.

Firstly, the choice of the LLaMa 3.2 1B model was made to effectively augment over 150,000 documents in a relatively short time. However, a more powerful parametric LLM, with 8B or 7B parameters, may yield better results, which warrants further analysis.

Secondly, we could have considered using a multilingual LLM for this task. However, due to time constraints, we opted to proceed with the LLaMa 3.2 1B model.

Lastly, more in-depth research is needed in prompt engineering. While deciding on the appropriate prompt structure, we encountered various unexpected issues. Such as the repeated or multiple generation of the same keywords, or the generation of very generic terms, analyzing quality of document augmentation in languages we (authors) don't use, etc. Designing a single prompt that effectively addresses all these issues proves to be a challenging task.

References

[1] [n. d.]. BM25 Baseline Implementation. https://github.com/dorianbrown/rank_bm25 Accessed: 2024-11-22.

[2] [n. d.]. Distiluse Multilingual v2 Model. https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v2 Accessed: 2024-11-22.

[3] [n. d.]. Facebook DPR (Dense Passage Retrieval) Model. https://huggingface.co/facebook/dpr-question_encoder-single-nq-base Accessed: 2024-11-22.

[4] [n. d.]. LaBSE Model. https://huggingface.co/sentence-transformers/LaBSE Accessed: 2024-11-22.

[5] [n. d.]. msmarco-bert-codensor Model. https://huggingface.co/sebastian-hofstaetter/colbert-distilbert-margin_msmarco Accessed: 2024-11-22.

[6] 2024. MultiClaim Dataset. https://drive.google.com/drive/folders/1cg5VC4ULDGNzicUMVD2chxzMh7l9A6IE.

[7] Varad Bhatnagar, Diptesh Kanojia, and Kameswari Chebrolu. 2022. Harnessing Abstractive Summarization for Fact-Checked Claim Detection. In *Proceedings of the 29th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2934–2945. https://aclanthology.org/2022.coling-1.259

[8] Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT Sentence Embedding. arXiv:2007.01852 [cs.CL] https://arxiv.org/abs/2007.01852

[9] Momchil Hardalov, Anton Chernyavskiy, Ivan Koychev, Dmitry Ilvovsky, and Preslav Nakov. 2022. CrowdChecked: Detecting Previously Fact-Checked Claims in Social Media. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online only, 266–285. https://aclanthology.org/2022.aacl-main.22

[10] Ye Jiang, Xingyi Song, Carolina Scarton, Ahmet Aker, and Kalina Bontcheva. 2021. Categorising Fine-to-Coarse Grained Misinformation: An Empirical Study of COVID-19 Infodemic. https://doi.org/10.48550/arXiv.2106.11702 arXiv:2106.11702 [cs].

[11] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 6769–6781. https://doi.org/10.18653/v1/2020.emnlp-main.550

[12] Ashkan Kazemi, Kiran Garimella, Devin Gaffney, and Scott Hale. 2021. Claim Matching Beyond English to Scale Global Fact-Checking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, 4504–4517. https://doi.org/10.18653/v1/2021.acl-long.347

[13] Watheq Mansour, Tamer Elsayed, and Abdulaziz Al-Ali. 2022. Did I See It Before? Detecting Previously-Checked Claims over Twitter. In *Advances in Information Retrieval (Lecture Notes in Computer Science)*, Matthias Hagen, Suzan Verberne, Craig Macdonald, Christin Seifert, Krisztian Balog, Kjetil Nørvåg, and Vinay Setty (Eds.). Springer International Publishing, Cham, 367–381. https://doi.org/10.1007/978-3-030-99736-6_25

[14] Preslav Nakov, Alberto Barrón-Cedeño, Giovanni da San Martino, Firoj Alam, Julia Maria Struß, Thomas Mandl, Rubén Míguez, Tommaso Caselli, Mucahid Kutlu, Wajdi Zaghouani, Chengkai Li, Shaden Shaar, Gautam Kishore Shahi, Hamdy Mubarak, Alex Nikolov, Nikolay Babulkov, Yavuz Selim Kartal, Michael Wiegand, Melanie Siegel, and Juliane Köhler. 2022. Overview of the CLEF–2022 CheckThat! Lab on Fighting the COVID-19 Infodemic and Fake News Detection. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, Alberto Barrón-Cedeño, Giovanni Da San Martino, Mirko Degli Esposti, Fabrizio Sebastiani, Craig Macdonald, Gabriella Pasi, Allan Hanbury, Martin Potthast, Guglielmo Faggioli, and Nicola Ferro (Eds.). Springer International Publishing, Cham, 495–520. https://doi.org/10.1007/978-3-031-13643-6_29

[15] Atharva Nijasure and Others. 2024. Multilingual Previously Fact-Checked Claim Retrieval. https://github.com/AtharvaNijasure/IRMultiLingualSemEval.

[16] Matúš Pikuliak, Ivan Srba, Robert Moro, Timo Hromadka, Timotej Smolen, Martin Melisek, Ivan Vykopal, Jakub Simko, Juraj Podrouzek, and Maria Bielikova. 2023. Multilingual previously fact-checked claim retrieval. *arXiv preprint arXiv:2305.07991* (2023).

[17] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. http://arxiv.org/abs/1908.10084

[18] Stephen Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Found. Trends Inf. Retr.* 3, 4 (apr 2009), 333–389. https://doi.org/10.1561/1500000019

[19] Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, and Preslav Nakov. 2020. That is a Known Lie: Detecting Previously Fact-Checked Claims. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 3607–3618. https://doi.org/10.18653/v1/2020.acl-main.332

[20] Shaden Shaar, Nikola Georgiev, Firoj Alam, Giovanni Da San Martino, Aisha Mohamed, and Preslav Nakov. 2022. Assisting the Human Fact-Checkers: Detecting All Previously Fact-Checked Claims in a Document. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2069–2080. https://aclanthology.org/2022.findings-emnlp.151

[21] Shaden Shaar, Fatima Haouari, Watheq Mansour, Maram Hasanain, Nikolay Babulkov, Firoj Alam, Giovanni Da San Martino, Tamer Elsayed, and Preslav Nakov. 2021. Overview of the CLEF-2021 CheckThat! Lab Task 2 on Detecting Previously Fact-Checked Claims in Tweets and Political Debates. In *Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum*, Vol. 2936. CEUR-WS, 13 pages. http://ceur-ws.org/Vol-2936/paper-29.pdf

[22] Qiang Sheng, Juan Cao, Xueyao Zhang, Xirong Li, and Lei Zhong. 2021. Article Reranking by Memory-Enhanced Key Sentence Matching for Detecting Previously Fact-Checked Claims. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.425

[23] Nguyen Vo and Kyumin Lee. 2018. The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 275–284. https://doi.org/10.1145/3209978.3210037

[24] Nguyen Vo and Kyumin Lee. 2020. Where Are the Facts? Searching for Fact-checked Information to Alleviate the Spread of Fake News. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, 7717–7731. https://doi.org/10.18653/v1/2020.emnlp-main.621

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009