

COVID-19 Claim Radar: A Structured Claim Extraction and Tracking System

Manling Li, Revanth Gangi Reddy, Ziqi Wang, Yi-Shyuan Chiang,
Tuan M. Lai, Pengfei Yu, Zixuan Zhang, Heng Ji

Computer Science Department

University of Illinois Urbana-Champaign

{manling2, revanth3, ziqiw9, hengji}@illinois.edu

Abstract

The COVID-19 pandemic has received extensive media coverage, with a vast variety of claims made about different aspects of the virus. In order to track these claims, we present *COVID-19 Claim Radar*¹, a system that automatically extracts claims relating to COVID-19 in news articles. We provide a comprehensive structured view of such claims, with rich attributes (such as claimers and their affiliations) and associated knowledge elements (such as events, relations and entities). Further, we use this knowledge to identify inter-claim connections such as equivalent, supporting, or refuting relations, with shared structural evidence like claimers, similar centroid events and arguments. In order to consolidate claim structures at the corpus-level, we leverage Wikidata² as the hub to merge coreferential knowledge elements, and apply machine translation to aggregate claims from news articles in multiple languages. The system provides users with a comprehensive exposure to COVID-19 related claims, their associated knowledge elements, and related connections to other claims. The system is publicly available on GitHub³ and DockerHub⁴, with complete documentation⁵.

1 Introduction

Claims present in daily news are unfiltered and potentially of great value, but can also have negative effects when misinformation is widespread. The COVID-19 pandemic is a crucial example of when false claims can be particularly harmful, with the torrent of misinformation impacting public perception. For example, a claim such as “*Vaccines are DNA changers*” is likely to discourage vaccinations.

¹Live Demo: <http://18.221.187.153/>

²<https://www.wikidata.org/>

³GitHub: <https://github.com/uiucnlp/covid-claim-radar>

⁴DockerHub: <https://hub.docker.com/repository/docker/blendernlp/covid-claim-radar>

⁵Video: http://blender.cs.illinois.edu/aida/covid_claim_radar.mp4

Further, a study by KFF⁶ COVID-19 Vaccine Monitor project found that 78% of U.S. adults agree with one of eight false claims regarding the pandemic.

In order to distinguish misleading information, a fundamental step is to first identify claims and discover their supporting or refuting relations. Automatic claim detection (Palau and Moens, 2009; Eger et al., 2017; Stab et al., 2018; Li et al., 2019) aims to mine arguments regarding a topic of consideration and has been applied to the COVID-19 scenario (Saakyan et al., 2021; Liu et al., 2020; Reddy et al., 2021). However, existing approaches ignore rich claim structures, or fail to associate claims with structured knowledge elements, thereby being incapable of supporting a more structured analysis. Further, they do not support real-time claim discovery, a feature required to process the rapidly updating COVID-19 pandemic information.

In this paper, we release a claim detection system that aims to automatically mine rich claim structures from news. Different from traditional claim detection systems that discover claims in isolation, we introduce a structured view for claims that consists of:

(1) **Structured Claim Attributes** including claim TOPIC, SUBTOPIC, TEMPLATE, CLAIMOBJECT, CLAIMER, AFFILIATION, LOCATION, and TIME. Our extraction is performed at the corpus-level with entity linking and coreference resolution, which allows for the construction of such comprehensive structures. For example, Table 1 shows a claim related to the topic *Wearing Masks*, where the claimer’s AFFILIATION can not be directly extracted from the local sentence, but it can be derived from the “General Affiliation” of the CLAIMER that is extracted from the corpus.

(2) **Associated Knowledge Elements** namely the entities, relations and events associated with the

⁶Kaiser Family Foundation, an American non-profit organization.

CLAIMTEXT	Cloth face coverings are most likely to reduce the spread of COVID-19 when they are widely used by people in public settings
TOPIC	Wearing Masks
TEMPLATE	Wearing masks is necessary in location [X]
CLAIMOBJECT	public settings [Identity Qnode] Q294440 (public space) [Type Qnode] Q7551384 (social space)
CLAIMER	Reed [Identity Qnode] Q30105757 (Carrie Reed) [Type Qnode] Q1650915 (researcher)
AFFILIATION	Centers for Disease Control and Prevention [Identity Qnode] Q583725 (CDC) [Type Qnode] Q20857065 (United States federal agency)
LOCATION	None
TIME	EarliestStart: 2020-01-01 LatestStart: 2020-07-27 EarliestEnd: 2020-07-27 LatestEnd: None
STANCE	affirm
ASSOCIATED KNOWLEDGE	Cloth face coverings [COM.EQUIPMENT] are most likely to reduce [CONTROL.IMPEDEINTERFERE] the spread [DISASTER.DISEASEOUTBREAK] of COVID-19 [MHI] when they are widely used [SOCIALBEHAVIOR.WEAR] by people in public settings [LOC]
SUPPORTING CLAIMS	33,000 deaths could be avoided by October 1 if 95 percent of people wore masks in public masks can prevent transmission in high-risk scenarios
REFUTING CLAIMS	face masks can be harmful, because they force the wearer to breathe in too much carbon dioxide with an N95 mask, some people have an elevated blood carbon dioxide level, and some also reduced oxygen level Masks can cause carbon dioxide poisoning

Table 1: An example of claim structure.

claims as claim evidence. For example, *reduce* is identified as a CONTROL.IMPEDEINTERFERE event with *COVID-19* as TARGET and *cloth face coverings* as INSTRUMENT. This representation provides a structured perspective of the claim semantics and enables discovery of semantic relatedness across multiple claims via knowledge elements.

(3) Inter-Claim Connections for identifying supporting, refuting and equivalent claims, with complex structured connections via claim attributes

and knowledge elements. For example, Table 1 shows the supporting claims that share the *mask* entity and CONTROL.IMPEDEINTERFERE event, as well as refuting claims about masks having negative effects of elevating blood carbon dioxide level.

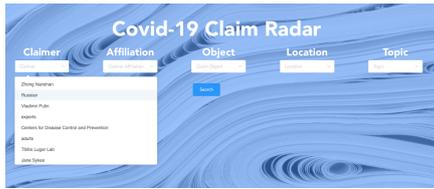
(4) Wikidata Linking for linking claim attributes (including CLAIMER, CLAIMOBJECT, AFFILIATION and LOCATION) and knowledge elements (entities, events and relations) to Wikidata, as shown in Table 1. It enables corpus-level knowledge consolidation and provides external references for users. Note that we use the terms “Qnode” and “Wikidata item” interchangeably.

(5) Structured Search Queries to support multi-dimensional search and analysis. Figure 1a shows our multi-dimensional search interface for searching multiple claim attributes jointly, as well as their associated knowledge elements. Each search dimension also provides some frequent candidates as references, such as *Centers for Disease Control and Prevention* for CLAIMER.

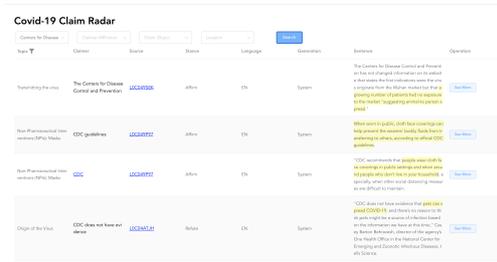
COVID-19 Claim Radar automatically provides users with a comprehensive and structured overview about COVID-19 related claims, allowing an accurate understanding of rapidly emerging claims, their importance, and their interconnections. The structured view enables seamless search with complex queries and discovery of alternative claims over the rich claim structures. The system is particularly useful for tracking current claims, providing alerts, and predicting possible changes, as well as topics related to the ongoing incidents.

2 Overview

The architecture of our structured claim extraction system is illustrated in Figure 2. The system pipeline consists of different components with two main modules, namely, Claim Extraction (CE) (Section 3) and Knowledge Extraction (KE) (Section 4). Each module creates a separate knowledge base, using the document corpus as input. The corpus-level knowledge base is then associated to claims according to the justifications, as well as coreferential entities and events. Inter-claim relations such as equivalent, supporting or refuting are then identified based on their structural connections (Section 5).



(a) Home page with a multi-dimensional search interface.



(b) List of claims returned corresponding to a search for claimer “Center for Disease Control and Prevention.”



(c) Structured claim view with associated knowledge elements and equivalent claims shown. Hovering over a knowledge element shows its corresponding arguments.

Figure 1: Screenshots of the demo corresponding to (a) main page, (b) list of claims returned from search, and (c) the structure claim view.

3 Claim Extraction

3.1 Core Claim Extraction

We employ a zero-shot claim detection framework that identifies claims relating to COVID-19 in addition to background attributes such as the CLAIMER and CLAIMOBJECT. Specifically, the system consists of a claim-spotting model to identify sentences that contain claims, with additional modules for filtering topics, and detecting the claimer and claim objects.

For the claim-spotting model, we use ClaimBuster⁷ (Hassan et al., 2017) to identify sentences which contain claims. Next, we leverage an extractive Question Answering (QA) system (Alberti et al., 2019) in a zero-shot setting for topic filtering, claimer detection and claim object detection. We use a QA model that is trained on SQuAD 2.0 (Rajpurkar et al., 2018) and Natural Questions (Kwiatkowski et al., 2019).

For topic filtering, we hand-craft questions corresponding to each topic. Then, we use the claim sentence as context and pass these questions as input to the QA model. The answer score for each question is used as the corresponding topic score and a threshold is set on the highest topic score in order to select the claim. Table 2 shows the examples of individual questions used to select claims relating to specific topics about COVID-19.

For claim object detection, we use the answer span for the question corresponding to the claim topic as the CLAIMOBJECT. For identifying the claim span, we use the claim boundary detection service released as part of the Project Debater (Bar-Haim et al., 2021). Next, we leverage the same QA model for claimer detection, by using the answer corresponding the question “Who said that <claim span>?”, with the entire news article as context.

3.2 Knowledge-Enhanced Claim Extraction

To identify the knowledge elements associated to the extracted claims, we leverage entities, relations and events that are extracted from the Knowledge Extraction module (detailed in Section 4). We extract knowledge elements within each claim span and within the sentences before and after the claim span. To provide a comprehensive understanding of claim attributes such as the AFFILIATION of the claimer, we extract entity-entity relations of types “General Affiliation” and “Organization Affiliation”

⁷<https://idir.uta.edu/claimbuster/api/>

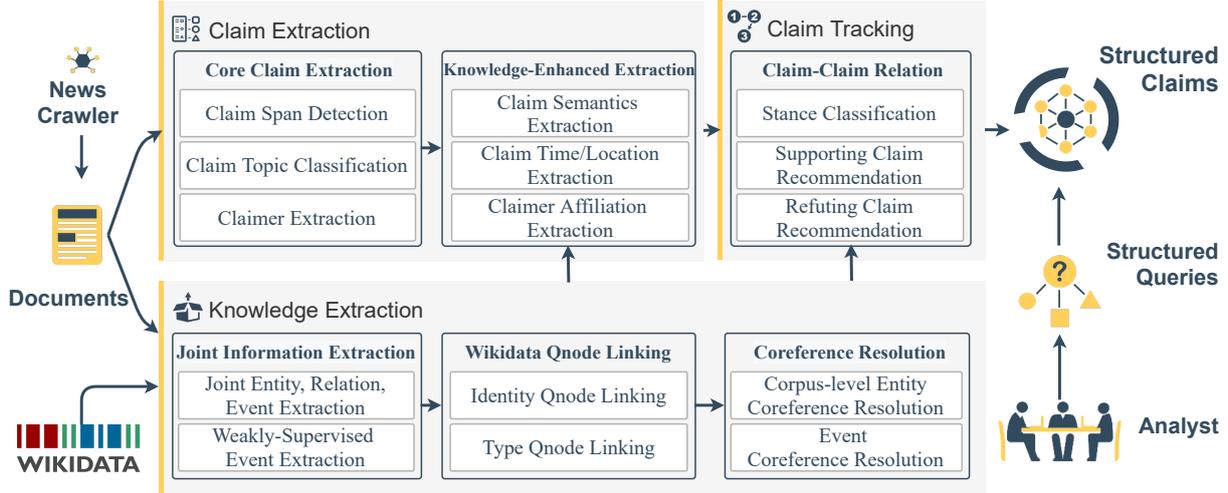


Figure 2: Architecture of the structured claim and knowledge extraction system.

Topic	Question
Transmission of COVID-19	What transmits the virus?
Contraction of COVID-19	Who can contract the virus?
Protection from COVID-19	What can protect from the virus?
Origin of COVID-19	What animal is associated with the origin of the virus?
Origin of COVID-19	Where did the first case of the virus occur?
Wearing Masks	What are the harmful effects of wearing masks?
Wearing Masks	Where is it necessary to wear masks to prevent the virus?
Cure for COVID-19	What can cure the virus?

Table 2: Examples of questions corresponding to individual topics about COVID-19.

from the entire corpus and perform corpus-level entity conference resolution. We also fill in each claim’s LOCATION and TIME according to the spatial and temporal attributes of the events mentioned in the claim span.

4 Knowledge Extraction

4.1 Joint Information Extraction

We first perform joint extraction of events of 144 types, entities of 7 types and relations of 38 types using the state-of-the-art supervised Information Extraction system (Lin et al., 2020)⁸. To extract event types and entity types newly emerging in the COVID-19 pandemic scenario, we employ a keyword-based event detection system. Specifically, we manually collected a list of keywords for each new event type, and compute keyword rep-

⁸We use the extended version (Li et al., 2020) that supports the most comprehensive DARPA AIDA ontology. The ontology is attached to the Appendix.

resentations by averaging the contextualized representations from BERT (Devlin et al., 2019) of keyword occurrences in an unlabeled pandemic-related corpus. Then we aggregate keyword representations for the same event type to get the event type representation. For event trigger detection, we first compute BERT representations of all the tokens in a sentence, and consider a token as an event trigger if its cosine similarity with an event type representation is larger than a threshold.

4.2 WikiData Qnode Linking

Wikidata is the most extensive crowdsourced knowledge graph. As such, it allows us to tie claimers, claim objects, and knowledge elements (e.g., entities) together to consolidate claim structures at the corpus level.

The massive number of entities in Wikidata (i.e., QNodes) makes entity linking challenging. To effectively narrow down the search space, we propose a candidate retrieval paradigm based on entity profiling. Wikidata entities and their textual fields are first indexed into a Elasticsearch. During inference, given a mention and its context, we use a trained sequence-to-sequence (seq2seq) model to generate the profile of the target entity, which consists of its title and description. We use the profile to query the indexed search engine to retrieve candidate entities. We use Wikipedia anchor texts and their corresponding Wikidata entities as the supervision signals for training the framework. In addition to instance-level linking, we also perform Qnode linking on all entity types in our ontology.

4.3 Coreference Resolution

We conduct entity coreference resolution within each document (Lai et al., 2021b) by employing SpanBERT (large) (Joshi et al., 2020) as the base Transformer encoder and train the entire neural model on ACE 2005 (Walker et al., 2006), NIST TAC-KBP EDL 2016⁹ (Ji et al., 2015), EDL 2017¹⁰ (Ji et al., 2017), and OntoNotes (English) (Pradhan et al., 2012). After that, we utilize the Wikidata entity linking results to refine the predictions of the neural model. We prevent two entity mentions from being directly merged if they are linked to different entities (i.e., Qnodes) with high confidence. To construct a corpus-level knowledge graph, all entities that are linked to the same Qnode will be merged into the same cluster (even if the entities are from different documents).

Our event coreference resolution is performed within each document and adopts a similar method as entity coreference resolution, while incorporating additional symbolic features such as the event type information (Lai et al., 2021a). We use the multilingual XLM-RoBERTa (XLM-R) (Conneau et al., 2020) as the base Transformer encoder. We train the model on ACE 2005 (Walker et al., 2006) and ERE (Song et al., 2015a).

5 Claim-Claim Relation Extraction

We consolidate the claims from the entire corpus according to the Wikidata Qnode linking results and claim attributes.

5.1 Stance Classification

We identify the stance from the perspective of each claimer, namely whether the claimer *affirms* or *refutes* a claim. This is different from prior stance detection tasks (Hardalov et al., 2021), which define stance with respect to target-context pairs, such as claim-evidence or headline-article.

In this setting, we follow Reddy et al. (2021) to use pre-trained Natural Language Inference (NLI) models for stance detection. Specifically, we formulate hypotheses for both of the *affirm* and *refute* labels, using the claim’s corresponding topic. Then, the claim sentence is used as the premise as input to the NLI model, with the hypothesis corresponding to higher entailment score considered as the stance. We use a Bart-large (Lewis et al., 2020)

⁹LDC2017E03

¹⁰LDC2017E52

model trained on MultiNLI (Williams et al., 2018) as our pre-trained NLI model.

5.2 Equivalent Claims

We use the structured claim information to identify claims that are equivalent. Specifically, we consider claims that share the same SUBTOPIC, CLAIMOBJECT and STANCE as equivalent. For CLAIMOBJECT, we use the corresponding Wikipedia QNode to account for diversity in the object mentions.

5.3 Supporting and Refuting Claims

We also identify claims that are supporting or refuting each other. We formulate this as an NLI task where the claims are corresponding premise-hypothesis pairs. We use high entailment or contradiction scores as an indication of whether two claims are supporting or refuting each other respectively. We leverage the same pre-trained NLI model as used in Section 5.1.

6 Experiment

6.1 Dataset

The system can take any set of news articles to extract claims and perform visualization. The live demo ¹¹ supports two functions: (1) **Real-time Extraction**: Users are able to copy a piece of news content and extract claims; (2) **Periodical Update**: To track claims in this rapidly evolving pandemic, we periodically collect newly emerging COVID-19 related news articles from Google News ¹², and perform claim extraction and knowledge extraction to update the *COVID-19 Claim Radar*.

6.2 System Performance

The performance of each component is shown in Table 4. We evaluate the end-to-end performance of our system on 1,139 COVID-19 news articles released by the Linguistic Data Consortium (LDC2021E11). We translated the Spanish and Russian news into English and perform end-to-end extraction on the entire corpus. More analysis on the extraction results are detailed in the Appendix.

6.3 Case Study

In the context of comprehensive claim structures, our system can perform explainable and reliable predictions in terms of supporting and refuting

¹¹<http://18.221.187.153/>

¹²<https://news.google.com/rss/search?q=xx>

	#doc	#claim	#claimer	#affiliation	#location	#start _{earliest}	#start _{latest}	#end _{earliest}	#end _{latest}	#entity	#event
English	484	905	581	133	166	714	693	661	336	11,302	1,718
Spanish	385	427	285	94	76	324	318	309	114	5,812	722
Russian	234	566	362	73	135	466	457	442	237	6,751	1,179

Table 3: Results of structured claim extraction.

Component		Benchmark	Metric	Score
Claim Extraction	Claim	NewsClaims	F ₁	36.0%
	Claim Object	NewsClaims	F ₁	57.0%
	Claimer	NewsClaims	F ₁	50.1%
	Stance	NewsClaims	Acc.	87.5%
Knowledge Extraction	Entity	ACE	F ₁	89.6%
	Relation	ACE	F ₁	58.6%
	Event Trigger	ACE	F ₁	72.8%
	Event Argument	ACE	F ₁	54.8%
Wikidata Qnode Linking	TACKBP-2010	Acc.	90.9%	
Coreference	Entity	OntoNotes	CoNLL	92.4%
	Event	ACE	CoNLL	84.8%

Table 4: Performance of each component. The benchmark references are: NewsClaims (Reddy et al., 2021), ACE (Walker et al., 2006), ERE (Song et al., 2015b), TACKBP-2010 (Ji et al., 2010), OntoNotes (Pradhan et al., 2012).

claims. For example, for the claim “*masks should be carefully taken off after getting inside a car or room*”, we are able to discover its refuting claim as “*wear them in your car, your bed, the shower, wear three of them if you want just leave it to the rest of us to decide when it is necessary*”, since they share the entities *mask* and *car*, but their STANCE is conflicting, i.e., *refute* and *affirm* respectively.

In addition, we compare the claims extracted from multiple languages. The claims from different languages can be refuting. For example, regarding the TOPIC about “*transmitting the virus*”, the claim extracted from a Spanish document “*...small mammals might have transmitted coronavirus to a worker...*” (STANCE = *affirm*) is refuting with the claim extracted from Russian document “*domestic animals cannot be infected with COVID-19 coronavirus and spread it*” (STANCE = *refuting*).

7 Related Work

Claim detection is a central task in argumentation mining (Palau and Moens, 2009; Eger et al., 2017; Stab et al., 2018). It can be framed as a classification problem (Does a document or a sentence contain a claim?) or a sequence labeling problem

(Which tokens make up the claim?). The classification setting has been explored as a retrieval task of online comments made by public stakeholders on pending governmental regulations (Kwon et al., 2007), for sentence detection in essays (Lippi and Torroni, 2015), and Wikipedia (Roitman et al., 2016; Levy et al., 2017). The sequence labeling setting has been tackled on Wikipedia (Levy et al., 2014) and news articles (Goudas et al., 2014; Sardianos et al., 2015).

COVID-19 related claim detection and argument mining are generally still limited. The majority of other argument mining approaches for the biomedical domain focus on research literature (Li et al., 2019; Saakyan et al., 2021; Liu et al., 2020). The work by Reddy et al. (2021) is one of the few exceptions that tackle this challenge and propose a pipeline to extract health-related claims with claim attributes from news articles. However, it does not attempt associating claims and their attributes with structured knowledge elements. To the best of our knowledge, detecting structured COVID-19 claims associated with structured knowledge elements has not been approached yet. Our system leverages the state-of-the-art information extraction and Wikidata entity linking techniques to dynamically construct a COVID-19 claim knowledge base.

8 Conclusions and Future Work

We present our *COVID-19 Claim Radar* system, to automatically extract claims in real time from rapidly updating information on the COVID-19 pandemic. We provide users with an in-depth structured view of claims, along with associated knowledge elements. Our system enables exploring various inter-claim connections, including supporting and refuting relations, shared claimers and claim objects, along with related events and entities. In future work, we plan to validate claims from multiple modalities, languages, and sources, as well as support information surgery to correct false claims automatically. In addition, we aim to track claims so as to predict changes in perspectives of claimers and facilitate generating alerts for such changes.

Ethical Considerations

Usage Requirements

COVID-19 Claim Radar provides investigative leads rather than final results, so it should not be used as direct conclusions or be applied to any human subjects directly. Research involving human subjects should first be approved by the stakeholder’s IRB (Institutional Review Board) who will ensure the safety of the studies.

Required workflow Our system is designed to facilitate the understanding of rapidly updating and expanding news articles regarding COVID-19 pandemic, which is difficult for human to keep track of newly emerging claims and to discern false claims from the true ones. Our claim extraction tool (and all claim discovery tools for biomedical applications) is not intended to be used for direct applications involving decisions or human subjects. Instead, our tool aims to highlight structures of claims from a large amount of news text data, which would be too time-consuming for humans to digest. As a result, the tool would be useful to identify claims and analyze the inter-connections between claims. It allows users to narrow down concerned claims from the claimers or affiliations, and then followed by a careful evidence checking to validate claims before making further decisions. Our system does not perform claim verification, which we leave as future work. Failure to follow this workflow, and use of the system without the required human validation, could lead to undesired experimental design wasting time and resources.

Evidence checking We provide evidence in the form of structured output in the surrounding contexts with confidence values, as well as the original news article and raw text content as justification. In addition, we provide Wikidata as external knowledge for the user’s reference. In order to minimize potential harm caused by extraction errors, consumers of the extracted claims and knowledge elements should double-check the source information and verify the accuracy of the discovered leads prior to undertaking expensive or time-consuming experimental studies.

Limitations of System Performance

COVID-19 Claim Radar is capable of converting a large number of news articles into structured claims. However, none of our extraction components is perfect, as reported in the experiments. However, as

we described in the workflow, the output of our system is intended to be interpreted by humans. Without human validation, incorporating the system output into a decision-making application could result in undesirable results.

Limitations of Data Collection

The system output might cause harm if it is used in a manner that magnifies the errors or bias in its training data or source input data.

Bias in training and development data The performance of our system components as reported is based on the specific benchmark datasets, which could be affected by such data biases. Thus questions concerning generalizability and fairness should be carefully considered. In our paper, most components rely on weak distant supervision such as external knowledge base Wikidata or manually selected keywords. In order to ensure proper application, we recommend: ethical considerations are expected to be included in every step of the system design, the system ensures high transparency and interpretability of data, algorithms, models, and functionalities.

Bias in source data Proper use of the technology requires that input documents are legally and ethically obtained. Our goal is to automatically process unstructured text from diverse sources to obtain structured claims, and highlight the complex connections across claims to better identify refuting and supporting claims. The input should not disclose personally identifiable health information, and is expected to have countermeasures for protecting vulnerable groups.

References

- Chris Alberti, Kenton Lee, and Michael Collins. 2019. A bert baseline for the natural questions. *arXiv preprint arXiv:1901.08634*.
- Roy Bar-Haim, Yoav Kantor, Elad Venezian, Yoav Katz, and Noam Slonim. 2021. [Project Debater APIs: Decomposing the AI grand challenge](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 267–274, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. [Unsupervised](#)

- cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. Neural end-to-end learning for computational argumentation mining. *arXiv preprint arXiv:1704.06104*.
- Theodosios Goudas, Christos Louizos, Georgios Petasis, and Vangelis Karkaletsis. 2014. Argument extraction from news, blogs, and social media. In *Hellenic Conference on Artificial Intelligence*, pages 287–299. Springer.
- Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021. A survey on stance detection for mis- and disinformation identification. *arXiv preprint arXiv:2103.00242*.
- Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarni, Anil Kumar Nayak, et al. 2017. Claimbuster: The first-ever end-to-end fact-checking system. *Proceedings of the VLDB Endowment*, 10(12):1945–1948.
- Heng Ji, Heng Ji, Ralph Grishman, Hoa Trang Dang, Kira Griffitt, and Joe Ellis. 2010. Overview of the tac 2010 knowledge base population track.
- Heng Ji, Joel Nothman, Ben Hachey, and Radu Florian. 2015. Overview of TAC-KBP2015 tri-lingual entity discovery and linking. In *Proc. Text Analysis Conference (TAC2015)*.
- Heng Ji, Xiaoman Pan, Boliang Zhang, Joel Nothman, James Mayfield, Paul McNamee, and Cash Costello. 2017. Overview of TAC-KBP2017 13 languages entity discovery and linking. In *Proc. Text Analysis Conference (TAC2017)*.
- Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77.
- Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, et al. 2019. Natural questions: A benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7.
- Namhee Kwon, Liang Zhou, Eduard Hovy, and Stuart W Shulman. 2007. Identifying and classifying subjective claims. In *DG. O.*, pages 76–81. Citeseer.
- Tuan Lai, Heng Ji, Trung Bui, Quan Hung Tran, Franck Dernoncourt, and Walter Chang. 2021a. [A context-dependent gated module for incorporating symbolic semantics into event coreference resolution](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3491–3499, Online. Association for Computational Linguistics.
- Tuan Manh Lai, Trung Bui, and Doo Soon Kim. 2021b. End-to-end neural coreference resolution revisited: A simple yet effective baseline. *arXiv preprint arXiv:2107.01700*.
- Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1489–1500.
- Ran Levy, Shai Gretz, Benjamin Sznajder, Shay Hummel, Ranit Aharonov, and Noam Slonim. 2017. Unsupervised corpus-wide claim detection. In *Proceedings of the 4th Workshop on Argument Mining*, pages 79–84.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880.
- Manling Li, Alireza Zareian, Ying Lin, Xiaoman Pan, Spencer Whitehead, Brian Chen, Bo Wu, Heng Ji, Shih-Fu Chang, Clare Voss, et al. 2020. Gaia: A fine-grained multimedia knowledge extraction system. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 77–86.
- Xiangci Li, Gully Burns, and Nanyun Peng. 2019. Scientific discourse tagging for evidence extraction. *arXiv preprint arXiv:1909.04758*.
- Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7999–8009.
- Marco Lippi and Paolo Torrioni. 2015. Context-independent claim detection for argument mining. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

- Zhenghao Liu, Chenyan Xiong, Zhuyun Dai, Si Sun, Maosong Sun, and Zhiyuan Liu. 2020. Generalizing open domain fact extraction and verification to covid-fact thorough in-domain language modeling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 2395–2400.
- Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*, pages 98–107.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. [Conll-2012 shared task: Modeling multilingual unrestricted coreference in ontonotes](#). In *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning - Proceedings of the Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes, EMNLP-CoNLL 2012, July 13, 2012, Jeju Island, Korea*, pages 1–40. ACL.
- Pranav Rajpurkar, Robin Jia, and Percy Liang. 2018. Know what you don’t know: Unanswerable questions for squad. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 784–789.
- Revanth Gangi Reddy, Sai Chinthakindi, Zhenhailong Wang, Yi R Fung, Kathryn S Conger, Ahmed S Elsayed, Martha Palmer, and Heng Ji. 2021. Newsclaims: A new benchmark for claim detection from news with background knowledge. *arXiv preprint arXiv:2112.08544*.
- Haggai Roitman, Shay Hummel, Ella Rabinovich, Benjamin Sznajder, Noam Slonim, and Ehud Aharoni. 2016. On the retrieval of wikipedia articles containing claims on controversial topics. In *Proceedings of the 25th international conference companion on world wide web*, pages 991–996.
- Arkadiy Saakyan, Tuhin Chakrabarty, and Smaranda Muresan. 2021. Covid-fact: Fact extraction and verification of real-world claims on covid-19 pandemic. *arXiv preprint arXiv:2106.03794*.
- Christos Sardianos, Ioannis Manousos Katakis, Georgios Petasis, and Vangelis Karkaletsis. 2015. Argument extraction from news. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 56–66.
- Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, S. Kulick, Neville Ryant, and Xiaoyi Ma. 2015a. From light to rich ere: Annotation of entities, relations, and events. In *EVENTS@HLP-NAACL*.
- Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, and Xiaoyi Ma. 2015b. From light to rich ere: annotation of entities, relations, and events. In *Proceedings of the the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 89–98.
- Christian Stab, Tristan Miller, and Iryna Gurevych. 2018. Cross-topic argument mining from heterogeneous sources using attention-based neural networks. *arXiv preprint arXiv:1802.05758*.
- Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. Ace 2005 multilingual training corpus. *Linguistic Data Consortium, Philadelphia*, 57.
- Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122.