NEWSSENSE: Reference-free Verification via Cross-document Comparison

Jeremiah Milbauer[*] and **Ziqi Ding**[†] and **Zhijin Wu**[†] and **Tongshuang Wu**

Carnegie Mellon University, Pittsburgh PA, USA
{ jmilbaue | sherryw }@cs.cmu.edu
{ ziqiding | zhijinw }@andrew.cmu.edu

Abstract

We present NEWSSENSE, a novel sensemaking tool and reading interface designed to collect and **integrate** information from multiple news articles on a central topic. NEWSSENSE provides "reference-free verification," augmenting a central grounding article of the user's choice by: (1) linking it to related articles from different sources; and (2) providing inline highlights on how specific claims are either supported or contradicted by information from other articles. Using NEWSSENSE, users can seamlessly digest and cross-check multiple information sources without disturbing their natural reading flow. Our pilot study shows that NEWSSENSE has the potential to help users identify key information, verify the credibility of news articles, explore different perspectives, and understand what content is supported, contradicted, or missing. NEWSSENSE is available open source at github.com/jmilbauer/NewsSense.

1 Introduction

Why is it so hard, and so exhausting, to read the news? In the quest for knowledge, news readers today must contend with a rapidly evolving 24-hour news cycle, multiple news venues competing for attention and clicks, and the challenge of integrating fact-based reporting, opinion pieces, and social media commentary (Lazer et al., 2018; Benkler et al., 2018; Farkas and Schou, 2019). With news becoming increasingly politicized (Faris et al., 2017) readers also face the challenge of identifying and avoiding misinformation, disinformation, and hyperbolic "clickbait" as they try to remain informed about the world around them.

Various solutions have been proposed to assist with users' news reading. For example, media watchdog companies have created media bias charts to represent political leaning and credibil-
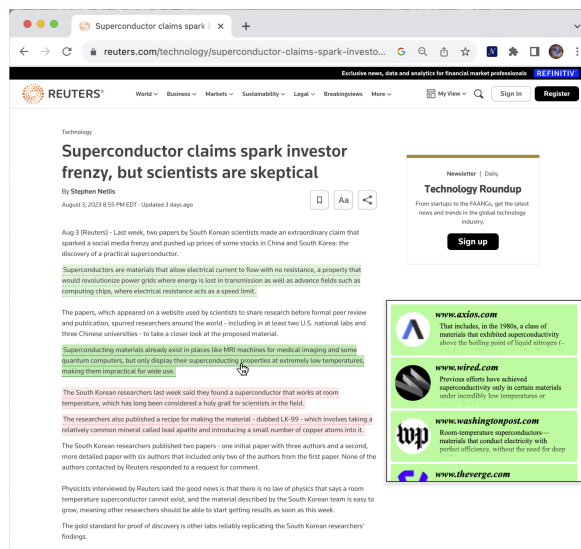

Figure 1: A screenshot from NEWSSENSE browser extension running in Chrome. The extension provides highlights indicated supported and controversial information. When the user clicks on a highlighted sentence, NEWSENSE adds an scrollable overlay box containing snippets of external evidence.

ity of news sources [1] [2]. However, these resources force users to rely on potentially untrustworthy third-party designations of media bias, which treat each *news source* as a whole, without digging into specific articles or topics.

While novel automatic fact checking (Thorne et al., 2018) and fake news detection (Zhou and Zafarani, 2020; Chen et al., 2015) systems can provide verification per-article, these approaches typically rely on a preordained corpus of verified facts which cannot keep up with the always-evolving facts, and may not reflect user preferences or multilateral perspectives. Aggregating articles from heterogeneous sources seem a more promising direction for cross-checking new facts (without predetermined groundtruths) and collecting different perspectives, but existing attempts are still too coarse and over-

---

[*] Corresponding author.
[†] Equal contribution.

[1] https://www.allsides.com/media-bias/media-bias-rating-methods
[2] https://adfontesmedia.com/interactive-media-bias-chart/

whelming. For example, both Google News' "Stories"[3] feature and Ground.news[4] collect articles about the same events but display them in the form of exhaustive lists – users are still forced to read and compare each article on its own.

We argue that instead of simply collecting and **aggregating** news articles, information and claims from multiple sources should be **integrated** in a way that allows users to identify fine-grained claim-level bias, spin, controversy, or evidence.

We present NEWSSENSE, a novel framework for sensemaking within a cluster of documents, to address the three key problems of news reading – bias, factuality, and article overload – in a single streamlined interface. NEWSSENSE leverages existing modular natural language processing (NLP) techniques to identify and link claims made across *a cluster of news articles*, such that these articles become references for each other. NEWSSENSE also displays the linking information using an *interactive reading interface*, which allows users to easily explore the cross-document connections without being overwhelmed. In pilot user studies, we see that the NEWSSENSE framework has the potential to help users identify key information, verify the credibility of news articles, and explore different perspectives.

While NEWSSENSE is primarily implemented for news articles, our framework can be easily generalized to other assisted reading and cross-checking scenarios (e.g., compare multiple manuscripts in literature reviews). The key contributions of NEWSSENSE are:

1. A pipeline for analyzing the connections between a collection of documents.
2. A two-stage method for efficiently computing cross-document links between claims that support or contradict each other, enabling "reference-free" fact checking.
3. A framework for visualizing cross-document connections, and integrating claims from multiple documents into a single reading experience.

We conclude by discussing the generality and potential social benefits of NEWSSENSE.

2 Related Work

This section covers related research across media analytics, sensemaking, and natural language processing. Though some core ideas of this work have

been explored in the past, to our knowledge they have never been combined in a single system.

Media Bias and Analytics Research on media bias includes academic research to study social media sharing patterns (Roberts et al., 2021; Bakshy et al., 2015) and bias within media publications (Flaxman et al., 2016; Hamborg et al., 2019; Groseclose and Milyo, 2005). Commercial products exist in this area as well, such as the media bias charts of AllSides[5], which classifies political slant into one of five categories, and Ad Fontes Media[6], which models both political slant and factual credibility.

Research on news and social content aggregation has focused primarily on headline detection, timeline construction and clustering (Bouras and Tsogkas, 2012; Laban and Hearst, 2017), and event detection (Atefeh and Khreich, 2015; Kumaran and Allan, 2004). There exist user-oriented products in this space, such as Google News Stories[7], and Ground.news[8]. Some outlets, such as Propublica, aggregate their news stories into timelines[9].

Reading Interfaces and Sensemaking Recent work on reading interfaces has primarily focused on scientific literature, augmenting documents with information about cited papers (Lo et al., 2023; Kang et al., 2022), or enhancing references within documents themselves (Head et al., 2021; Liu et al., 2023a, 2019a, 2023b).

For the News domain specifically, Laban and Hearst (2017) aggregates articles and extracts key quotes to construct a timeline for a given story. We are also aware of an abstract describing work to combine multiple article headlines and ledes into a single digestible form, though no follow-up is available (Glassman et al., 2020).

Fact Verification and NLI Natural Language Inference is a task focused on classifying the relationship between a pair of sentences as either "neutral", "entailment", or "contradiction." Datasets such as SNLI (Bowman et al., 2015) and MNLI (Williams et al., 2017) have become major benchmarks for natural language processing research. Recent work has also considered document-level NLI (Koreeda and Manning, 2021; Chen et al., 2022), as well

---

[3]https://news.google.com/stories/
[4]https://ground.news

[5]https://www.allsides.com/media-bias/media-bias-rating-methods
[6]https://adfontesmedia.com/interactive-media-bias-chart/
[7]https://news.google.com
[8]https://ground.news/
[9]https://www.propublica.org/series

as cross-document reasoning based in NLI (Schuster et al., 2022), and scalable pairwise reasoning (Milbauer et al., 2023).

There is also a growing body of work on NLP systems for fact verification and attribution. Recent datasets include FEVER (Thorne et al., 2018) and VitaminC (Schuster et al., 2021), as well as datasets focused on real-world examples of updating, editing, and citing claims in domains like news and Wikipedia (Petroni et al., 2022; Spangher et al., 2022; Iv et al., 2022).

3 The NEWSSENSE Framework

The core philosophy behind NEWSSENSE is to go beyond article aggregation by integrating the information contained within a cluster of related news articles into the reading experience.

NEWSSENSE starts with a single "focus" article and a set of related "background" articles. The distinction between focus and background article is arbitrary, as any article within the cluster could be designated the focus article. NEWSSENSE then identifies claims within the focus article that are related – either by contradiction or entailment – to claims within the background articles. The claims in the focus article are then highlighted, and linked to the background articles so that users can explore the supporting or contradicting evidence for a given claim without just relying on third-party measurements of bias or credibility.

The NEWSSENSE interface has three primary components: a Focus Article, Sentence Highlights, and External Evidence. Together, these elements display the computed connections between the focus article and the background articles.

3.1 Focus Article

The NEWSSENSE interface features a central panel that displays the focus article, including the entire news article the reader is interested in primarily. The focus article can be presented through a dedicated application, or by adding NEWSSENSE as an overlay on top of the existing web browsing experience.

3.2 Sentence Highlights

The interface highlights claims made in the focus article that have supporting articles with green, while claims with contradicting articles are highlighted with red. By doing so, readers can quickly and easily identify areas of agreement and disagree-

ment across different news sources. However, not all disagreement is conflict; articles may disagree because one simply has more updated information than the other. Disagreements may also be limited to some individual claims, while the content of two articles remains generally consistent. This is one reason why it is essential to link to the external evidence, and allow the user to explore.

3.3 External Evidence

When the reader hovers or clicks on a highlighted claim, an overlay panel appears, containing the supporting or contradicting claim excerpts, as well as their news sources. Although we do not implement it here, NEWSSENSE provides the opportunity to integrate insights from prior analyses of the news-media landscape by annotating links with metadata about the credibility or political slant of the referenced news venue.

For readers' convenience, each supporting or contradicting claim is clickable and directs the reader to the origin of the associated claim. This allows readers to quickly access the relevant claims without having to search through an entire article, and helps them better understand the circumstances of the agreement or the disagreement. For a standalone NEWSSENSE interface not embedded in a web browser, readers would be prompted with a "back" button in the secondary articles to quickly go back to the focus article. By providing this functionality, readers can easily navigate through the focus and secondary articles and compare viewpoints, further enhancing their understanding of the news story.

4 Pilot Study

To gather feedback on the proposed NEWSSENSE interface and provide insights for the actual implementation, we conducted a pilot user study using a NewsReader mockup built with Figma [10]. This section describes the design and results of the study.

4.1 Study Design

We aim to collect feedback on NEWSSENSE's basic functionality, interface design, and content quality.

The participants were assigned a task of reading a news article using NEWSSENSE and answering a set of questions. The questions focused on the content of the news, how and where the user located information, and their level of trust in the
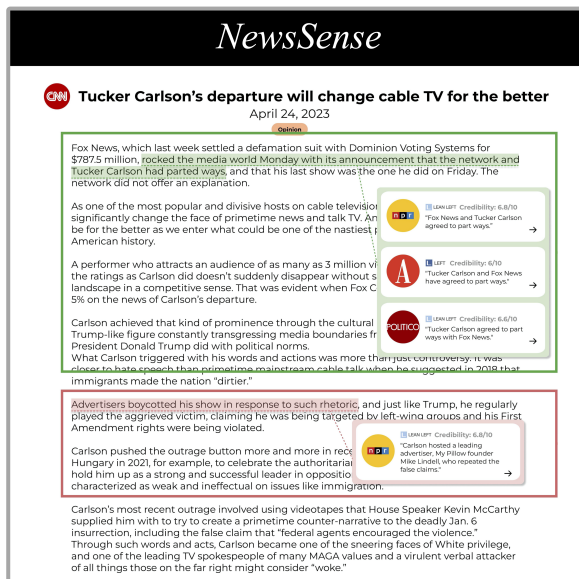
---

[10]https://www.figma.com/

Figure 2: The design layout of the pilot study, prototyped in Figma. The article is presented in a central panel, featuring claims with supporting articles highlighted in green (section boxed in green), and claims with contradicting articles highlighted in red (section boxed in red). Each claim has an associated overlay box of external evidence that appears when the user hovers over the text.

information. These questions aimed to assess the basic functionality of NEWSSENSE in helping readers understand news comprehensively, to motivate further development of the system.

4.2 Results

Following the pilot user study with over 10 users, we identified several key findings. First, all users found NEWSSENSE to be useful in locating important information and verifying the credibility of news articles, aligning with our initial goal. The user-friendly interface of NEWSSENSE was well-received, though participants suggested enhancing interactivity to set it apart from other solutions. For instance, displaying real-time feedback like *"NewsSense is analyzing the article"* during loading.

Regarding content quality, some users found NEWSSENSE limited and suggested increased labeling or categorization within articles. One user noted *Two highlighted sentences per page are insufficient for in-depth analysis."* User preferences varied for article summarization, with some wanting more key points and others preferring brevity. Contradicting previous feedback, one user preferred *"Summarizing key points only, rather than selecting sentences with unclear relevance."* Addressing this,

NEWSSENSE could allow customization, letting users choose key point count and filter supporting/contradicting data.

4.3 Study Takeaways

We found that users liked how NEWSSENSE highlighted important sentences from an article. We realized that the claims which are consistent across multiple articles (ie, those which are supported at least once) are likely to be the most important aspects to a given story. NEWSSENSE could thus help inform readers when there are key claims from across the article cluster missing from the article they are reading.

We also found that the bias labels for news venues could be overwhelming, and including them ran counter to our aim of reference-free verification; we eliminated these labels.

Users also appreciated how highlighted sentences functioned as summaries. Consequently we enhance the visibility of text highlights and further emphasize the alignment or contradiction of specific source by making the External Evidence cards colored accordingly.

5 System Overview

Following our user study, we implemented the NEWSSENSE framework as a browser plugin, which adds augmentations to news articles encountered on the web. Figure 1 shows the final appearance of the browser plugin. Open source code for the system and plugin, as well as a demo video, can be accessed at github.com/jmilbauer/NewsSense.

This section provides a description of the natural language processing system which powers NEWSSENSE. Figure 3 illustrates the four general steps of the pipeline: Collection, Selection, Filtering, and Linking.

5.1 Article Collection

First, we must collect a cluster of news articles that are all about the same news event. Our implementation scrapes data from Google News Stories, a website that collects many articles about the same events across news venues. After collecting article URLs via Google News Stories, we then collect the content of each article. A typical story contains over 50 articles.

5.2 Claim Selection

The next phase of the pipeline is to select the claims within each article cluster. We initially assumed

Figure 3: The four stages of the NEWSREADER linking pipeline: Article collection, Claim detection, claim filtering, and claim linking

a 1-to-1 mapping between sentences and claims, but quickly found that news articles often contain complex multi-clause sentences, which are not suitable for natural language inference. To address this issue, we few-shot prompting to generate a list of claims from sentences using a large language model (LLM). In our experiments, prompt exemplars are drawn from the PROPSEGMENT dataset (Chen et al., 2022), and the LLM used is OpenAI `text-davinci-003`. Full prompt details are provided in Appendix A. We also note that the authors of PROPSEGMENT report that `T5-Large` performs reasonably well on a similar proposition-level task, suggesting the possibility for further pipeline improvements, or one that does not rely on APIs.

5.3 Claim Filtering

Articles often contain over 30 sentences. For a cluster of 50 articles, a pairwise comparison of the full cartesian product of sentences has $\mathcal{O}((NL)^2)$, which is in practice well over 1,000,000 comparisons. Performing this level of computation at scale, even if we are pre-computing results for each article cluster, is simply not feasible. To address this, we perform an initial filtering step with leverages the fact that the vast majority of claims across any two articles are unrelated. We consider two approaches for claim filtering: Embedding Similarity filtering (ES) and Lexical Overlap filtering (LeO).

For Embedding Similarity filtering, we encode each claim in each article using a Transformer-based sentence encoder. Then, for each claim we retain only the k most similar other claims for comparison. In our implementation, we use the Sentence Transformers (Reimers and Gurevych, 2019)

	Prec.	Rec.	Macro-F1	TNR
Embedding Similarity	0.99	0.95	0.97	0.99
Lexical Overlap	0.91	0.89	0.90	0.91

Table 1: The positive-class precision, recall, macro-averaged F1, and true negative rate for the two filtering methods. Embedding Similarity outperforms lexical overlap on every metric.

model `all-Mini-LM-L6-v2`.

For Lexical Overlap filtering, we compare each sentence only with sentences that have overlapping words, as these sentences are likely to discuss similar topics. In our implementation, we process claims by first remove stopwords, then stemming using the NLTK (Loper and Bird, 2002) implementation of the Porter Stemmer (Porter, 1980), and compute overlap scores using the Jaccard Index.

We evaluated each filtering method on the MNLI (Williams et al., 2017) validation data, treating pairs of randomly sampled sentences as negative examples, and labeled "entailment" and "contradiction" sentence pairs as positive examples. For ES, we set a threshold of 0.3 cosine similarity; for LeO, we set a threshold of 0.1 overlap. We note that `all-Mini-LM-L6-v2` included MNLI in its training data.

We include a summary of the results of these experiments in Table 1, which indicates that the ES method outperforms the LeO method. Of particular interest is the true negative rate, as this indicates the percentage of non-related sentences we expect to filter out.

5.4 Claim Linking

Once claim pairs have been filtered, we classify each pair according to the Natural Language Inference (NLI) framework, as "entailment," "contradiction," or "neutral." We employ a pretrained language model, RoBERTa (Liu et al., 2019b), which was then fine-tuned on MNLI (Williams et al., 2017), a popular dataset for NLI. We download this fine-tuned version of RoBERTa from the Hugging Face model library [11].

To avoid clutter, we keep fewer than 100 of the most confident predictions for each positive class (entailment or contradiction) within the article cluster. Claims are then assigned back to the sentences from which they were generated, and the sentence pairs are linked.

6 Discussion

NEWSSENSE provides an intuitive and effective interface for integrating information from a large cluster of news articles into a single, focused reading experience. Although applied in this demo to news articles, the NEWSSENSE framework could just as easily be applied to the analysis of other types of document clusters as well. The pipeline itself is highly modular, and can easily adopt advancements in NLP technologies to increase the accuracy or decrease processing time.

6.1 Future Work

The generality of the NEWSSENSE also introduces a number of opportunities for future development.

Expanding the Scope of NewsReader Often, articles contain references to past events. In the future, we would like to explore the possibility of extending the NEWSSENSE framework beyond the immediately article clusters to include all relevant articles in a timeline of events.

Additionally, as we explored the NEWSSENSE framework, we noticed that the clustering approach we used – the Google News Stories – sometimes established associations between source *news* articles, and background *primary source* articles. As a result, we would encourage further exploration of the NEWSSENSE framework when applied to heterogenous and primary-source document collections, which might include primary scholarly literature.

Beyond the news, we also anticipate that the NEWSSENSE framework could prove useful for the analysis of scientific or other scholarly articles, by tracing ideas and providing attribution across documents.

NLP Pipeline Improvements Because the NEWSSENSE pipeline is modular, a number of improvements can be explored. For sentence segmentation, methods that use a fine-tuned language model could improve the speed of segmentation. For sentence filtering, we relied on a pretrained sentence retrieval model – but future iterations of NEWSSENSE could use a sentence retrieval model fine-tuned on "unrelated" pairs of MNLI sentences.

We also found that in many cases, the NLI algorithm used for claim linking made mistakes, perhaps owing to the fact that news articles may not perfectly match models trained on MNLI. To address this, more robust approaches to NLI such as VITAMINC (Schuster et al., 2021) could be explored. Other NLI approaches that are particularly applicable to NEWSSENSE would be those designed for both premises and sentences, such as SeNtLI (Schuster et al., 2022), or for many-to-many reasoning, such as LAIT (Milbauer et al., 2023), which speeds up inference through late interaction. Depending on how NEWSREADER is deployed, models of different sizes could be used: larger state of the art models for a centralized server, or smaller models for performing fast inference on the user's device.

More Useful Information Our final version of NEWSSENSE focused on a relatively paired-down and streamlined interface. However, users did suggest that they would like to see article summaries, and in fact used the highlighted claims as pseudo-summaries – indeed, good information is likely repeated across multiple articles. We would consider adding a way for NEWSSENSE to gather and convey the highlights – the key, supported, claims from across the article cluster – when a user is reading an article. We noticed other forms of unintended but incredibly useful functionality: For example, as stories develop, new facts emerge that may contradict old ones. This means that newer articles might supersede older ones. Future iterations of NEWSSENSE should help readers understand when a contradiction may be due to evolving stories.

Deployment A larger-scale user study would help determine what further improvements could

be made to the framework. Our fully interactive interface would help us run a study at larger scale.

7 Conclusion

We presented a novel framework for sensemaking within a cluster of documents. We applied this framework to news articles, building NEWSSENSE, an interactive tool that links claims within one document to supporting or contradicting evidence across the entire document cluster. NEWSSENSE assists readers by helping them to understand the connections and perspectives across many documents. Readers can thus attain a more comprehensive understanding of a given subject, while avoiding the dangers of information overload. Crucially, NEWSSENSE provides a framework for *reference-free* fact verification, which is essential in domains such as the news where events evolve in real time, because a knowledge source for factual grounding may not be available.

Our work expands the growing body of literature on natural language processing applications to document-level sensemaking by demonstrating the utility of automatically generated cross-document links, as well as the application of sensemaking tools to the news reading experience.

Limitations and Ethics

NEWSSENSE falls within the genre of computer science literature that aims to solve problems such as misinformation. A broad critique of this literature is that it may be considered a form of *techno-solutionism*, in the sense that we seek to develop technological solutions to problems that are potentially social in origin, and perhaps better addressed with a social approach. However, we posit that because the problem of misinformation propagation and newsmedia overload are both enabled by technology, we do have a responsibility to explore the ability of technological systems to address these challenges. Unlike techniques that involve traditional fact verification, the reference-free approach of NEWSSENSE does not take on the role of deciding what is true and what is not; it simply helps users understand the context of each claim, and make their own decisions.

Beyond this broad critique, NEWSSENSE might be limited in its application: it presupposes a diverse and free media environment, it does not distinguish between venue quality, might suffer from false balance, and could give undue credence to ideas repeated across low-quality outlets. These issues could result in NEWSSENSE providing a false sense of factuality. Additionally, there may be obstacles to its adoption, as the people who choose to use a system such as NEWSSENSE may already be predisposed to consider and critically evaluate diverse perspectives in the news; not those who need it most. We also consider that the highlighted links may clutter the reading experience, but we believe this concern is mitigated by the fact that news websites are already quite cluttered (by ads, sponsored links, and article thumbnails) and that users found the highlights helpful in identifying the key components of the articles.

Acknowledgements

References

Farzindar Atefeh and Wael Khreich. 2015. A survey of techniques for event detection in twitter. *Computational Intelligence*, 31(1):132–164.

Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132.

Yochai Benkler, Robert Faris, and Hal Roberts. 2018. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Christos Bouras and Vassilis Tsogkas. 2012. A clustering technique for news articles using wordnet. *Knowledge-Based Systems*, 36:115–128.

Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. 2015. A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*.

Sihao Chen, Senaka Buthpitiya, Alex Fabrikant, Dan Roth, and Tal Schuster. 2022. Propsegment: A large-scale corpus for proposition-level segmentation and entailment recognition. *arXiv preprint arXiv:2212.10750*.

Yimin Chen, Niall J Conroy, and Victoria L Rubin. 2015. Misleading online content: recognizing clickbait as"

false news". In *Proceedings of the 2015 ACM on workshop on multimodal deception detection*, pages 15–19.

Robert Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. Partisanship, propaganda, and disinformation: Online media and the 2016 us presidential election. *Berkman Klein Center Research Publication*, 6.

Johan Farkas and Jannick Schou. 2019. *Post-truth, fake news and democracy: Mapping the politics of falsehood*. Routledge.

Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298–320.

Elena L. Glassman, Janet Sung, Katherine Qian, Yuri Vishnevsky, and Amy Zhang. 2020. Triangulating the news: Visualizing commonality and variation across many news stories on the same event.

Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *The quarterly journal of economics*, 120(4):1191–1237.

Felix Hamborg, Karsten Donnay, and Bela Gipp. 2019. Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, 20(4):391–415.

Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, and Marti A. Hearst. 2021. Augmenting scientific papers with just-in-time, position-sensitive definitions of terms and symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery.

Robert Iv, Alexandre Passos, Sameer Singh, and Ming-Wei Chang. 2022. Fruit: Faithfully reflecting updated information in text. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3670–3686.

Hyeonsu Kang, Joseph Chee Chang, Yongsung Kim, and Aniket Kittur. 2022. Threddy: An interactive system for personalized thread-based exploration and organization of scientific literature. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, New York, NY, USA. Association for Computing Machinery.

Yuta Koreeda and Christopher D Manning. 2021. Contractnli: A dataset for document-level natural language inference for contracts. *arXiv preprint arXiv:2110.01799*.

Giridhar Kumaran and James Allan. 2004. Text classification and named entities for new event detection. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 297–304.

Philippe Laban and Marti A Hearst. 2017. newslens: building and visualizing long-ranging news stories. In *Proceedings of the Events and Stories in the News Workshop*, pages 1–9.

David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science*, 359(6380):1094–1096.

Michael Xieyang Liu, Jane Hsieh, Nathan Hahn, Angelina Zhou, Emily Deng, Shaun Burley, Cynthia Taylor, Aniket Kittur, and Brad A. Myers. 2019a. Unakite: Scaffolding developers' decision-making using the web. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, page 67–80, New York, NY, USA. Association for Computing Machinery.

Michael Xieyang Liu, Tongshuang Wu, Tianying Chen, Franklin Mingzhe Li, Aniket Kittur, and Brad A. Myers. 2023a. Selenite: Scaffolding decision making with comprehensive overviews elicited from large language models.

Nelson F Liu, Kenton Lee, and Kristina Toutanova. 2023b. Anchor prediction: Automatic refinement of internet links. *arXiv preprint arXiv:2305.14337*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Kyle Lo, Joseph Chee Chang, Andrew Head, Jonathan Bragg, Amy X. Zhang, Cassidy Trier, Chloe Anastasiades, Tal August, Russell Authur, Danielle Bragg, Erin Bransom, Isabel Cachola, Stefan Candra, Yoganand Chandrasekhar, Yen-Sung Chen, Evie Yu-Yen Cheng, Yvonne Chou, Doug Downey, Rob Evans, Raymond Fok, Fangzhou Hu, Regan Huff, Dongyeop Kang, Tae Soo Kim, Rodney Kinney, Aniket Kittur, Hyeonsu Kang, Egor Klevak, Bailey Kuehl, Michael Langan, Matt Latzke, Jaron Lochner, Kelsey MacMillan, Eric Marsh, Tyler Murray, Aakanksha Naik, Ngoc-Uyen Nguyen, Srishti Palani, Soya Park, Caroline Paulic, Napol Rachatasumrit, Smita Rao, Paul Sayre, Zejiang Shen, Pao Siangliulue, Luca Soldaini, Huy Tran, Madeleine van Zuylen, Lucy Lu Wang, Christopher Wilhelm, Caroline Wu, Jiangjiang Yang, Angele Zamarron, Marti A. Hearst, and Daniel S. Weld. 2023. The semantic reader project: Augmenting scholarly documents through ai-powered interactive reading interfaces.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.

Jeremiah Milbauer, Annie Louis, Mohammad Javad Hosseini, Alex Fabrikant, Donald Metzler, and Tal Schuster. 2023. LAIT: Efficient multi-segment encoding in transformers with layer-adjustable interaction. In *Proceedings of the 61st Annual Meeting of*

the Association for Computational Linguistics (Volume 1: Long Papers), pages 10251–10269, Toronto, Canada. Association for Computational Linguistics.

Fabio Petroni, Samuel Broscheit, Aleksandra Piktus, Patrick Lewis, Gautier Izacard, Lucas Hosseini, Jane Dwivedi-Yu, Maria Lomeli, Timo Schick, Pierre-Emmanuel Mazaré, et al. 2022. Improving wikipedia verifiability with ai. *arXiv preprint arXiv:2207.06220*.

Martin F Porter. 1980. An algorithm for suffix stripping. *Program*, 14(3):130–137.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Hal Roberts, Rahul Bhargava, Linas Valiukas, Dennis Jen, Momin M Malik, Cindy Sherman Bishop, Emily B Ndulue, Aashka Dave, Justin Clark, Bruce Etling, et al. 2021. Media cloud: Massive open source collection of global news on the open web.

Tal Schuster, Sihao Chen, Senaka Buthpitiya, Alex Fabrikant, and Donald Metzler. 2022. Stretching sentence-pair nli models to reason over long documents and clusters. *arXiv preprint arXiv:2204.07447*.

Tal Schuster, Adam Fisch, and Regina Barzilay. 2021. Get your vitamin c! robust fact verification with contrastive evidence. *arXiv preprint arXiv:2103.08541*.

Alexander Spangher, Xiang Ren, Jonathan May, and Nanyun Peng. 2022. Newsedits: A news article revision dataset and a novel document-level reasoning challenge. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 127–157.

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. Fever: a large-scale dataset for fact extraction and verification. *arXiv preprint arXiv:1803.05355*.

Adina Williams, Nikita Nangia, and Samuel R Bowman. 2017. A broad-coverage challenge corpus for sentence understanding through inference. *arXiv preprint arXiv:1704.05426*.

Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.

A Appendix: Prompt for Claim Extraction

This is the prompt used for claim extraction from news article sentences:

```
Extract all the claims from a sentence, ignoring
    extraneous words such as unimportant
    adverbs. A sentence may contain multiple
    claims. Each claim should be of the form <
    subject> <predicate> <object>, and should
    have the first occurrence of any pronouns
    replaced by their antecedents.

Sentence: "The 3rd and 4th stations all
    announced that they would be postponed, and
    the Monaco station was subsequently
    cancelled."
Claim: Monaco station was cancelled.
Claim: 4th stations announced they would be
    postponed.
Claim: The 3rd stations announced they would be
    postponed.
Claim: The 4th stations postponed.
Claim: The 3rd stations postponed.

Sentence: "Lewis Hamilton and Mercedes have once
    again confirmed themselves as drivers and
    constructors world champions."
Claim: Mercedes confirmed themselves as
    constructors world champions.
Claim: Lewis Hamilton confirmed themselves as
    drivers world champions.

Sentence: "Local organizers in East Palestine,
    Ohio on Monday said their activism has
    successfully pressured rail company Norfolk
    Southern to agree to a limited relocation
    plan for some residents affected by last
    month's train derailment, but added they
    have no intention of backing down from their
    demand for justice for thousands of people
    in the area who are struggling in the
    aftermath of the accident."
Claim: Local organizers said their activism has
    pressured rail company Norfolk Southern to
    agree to a limited relocation plan.
Claim: Local organizers have no intention of
    backing down from their demand for justice.
Claim: Rail company Norfolk Southern agree to a
    limited relocation plan.

Sentence: <INSERT SENTENCE HERE>
```