ClimatELi: Evaluating Entity Linking on Climate Change Data

Shijia Zhou* **Siyao Peng*** **Barbara Plank**
MaiNLP, Center for Information and Language Processing, LMU Munich, Germany
Munich Center for Machine Learning (MCML), Munich, Germany
zhou.shijia@campus.lmu.de {siyao.peng,b.plank}@lmu.de

Abstract

Climate Change (CC) is a pressing topic of global importance, attracting increasing attention across research fields, from social sciences to Natural Language Processing (NLP). CC is also discussed in various settings and communication platforms, from academic publications to social media forums. Understanding who and what is mentioned in such data is a first critical step to gaining new insights into CC. We present ClimatELi (CLIMATe Entity LInking), the first manually annotated CC dataset that links 3,087 entity spans to Wikipedia. Using ClimatELi, we evaluate existing entity linking (EL) systems on the CC topic across various genres and propose automated filtering methods for CC entities. We find that the performance of EL models notably lags behind humans at both token and entity levels. Testing within the scope of retaining or excluding non-nominal and/or non-CC entities particularly impacts the models' performances.

1 Introduction

Climate change (CC) is a well-established and omnipresent concept influencing daily lives. Natural Language Processing (NLP) tasks such as Entity linking (EL) facilitate knowledge base population and empower individuals to enhance their CC understanding. For example, after reading the news about "young indigenous women from Mexico and Morocco unite for COP27," one could click the following Wikipedia links and explore further about `COP27` or `Climate_of_Mexico`. EL provides easy access to CC-related knowledge and endows downstream applications like extracting information regarding stakeholders: policymakers, scientists, administrators, and etc. (Conde and Lonsdale, 2005).

To gain valuable and broad access to relevant CC information, a major challenge is that EL needs to robustly process different texts, both in terms of the specific CC topic and across a wide range of genres. However, there is a lack of human-labeled evaluation data for EL on CC, so little is known about how well existing EL perform.

> According to the [World Meteorological Organization](http://en.wikipedia.org/wiki/World_Meteorological_Organization), [La Niña](El_Niño-Southern_Oscillation) "refers to the large-scale [cooling](Heat_transfer) of [the ocean surface temperatures](Sea_surface_temperature) in the central and eastern [equatorial](Equator) [Pacific Ocean](Pacific_Ocean)".

Figure 1: Sample ClimatELi annotation.

We present ClimatELi,[1] the first manually annotated CC EL corpus covering five English genres in §3. §4 introduces three EL systems for evaluation and applies an automatic filtering mechanism for CC-related and nominal entities. §5-§6 analyze models' overall performance and on specific genres and threshold conditions. §8 concludes the paper.

2 Related Work

Entity-level information is gaining increasing attention in CC-related research. Different from named entity recognition (NER) applications on CC texts (Maynard and Bontcheva, 2015; Mishra and Mittal, 2021; Piskorski et al., 2022; Vaid et al., 2022; Spezzatti et al., 2022), entity linkers (EL) further disambiguate textual mentions and associate them to knowledge bases (KBs), e.g., Wikipedia (Cucerzan, 2007), DBPedia (Mendes et al., 2011, 2012), Wikidata (Vrandečić and Krötzsch, 2014), via candidate selection and ranking (Rao et al., 2013; Hachey et al., 2013; Moro et al., 2014; Shen et al., 2015). EL evaluations also spread from news, e.g., TAC-

* Equal contribution.

[1] https://github.com/mainlp/ClimatELi

KBP (McNamee and Dang, 2009; Ji et al., 2010), to more genres (Derczynski et al., 2015; Yang and Chang, 2015; Lin and Zeldes, 2021), and topic domains (Klie et al., 2020; Liu et al., 2021).

Despite a lack of human evaluation benchmarks for EL on CC data, there is work on integrating EL as an automatic pipeline for information extraction. Weichselbraun et al. (2015) present Recognyze, an EL platform confined to named entities adopted by two CC web applications on Swiss business news, i.e., the Media Watch on Climate Change and the Climate Resilience Toolkit. Ruiz Fabo et al. (2016) propose an NLP pipeline including EL for identifying supporting and opposing propositions in CC data. Diggelmann et al. (2021) employ EL to extract relevant documents for evidence candidates in CC claim verification. Pérez Ortiz et al. (2022) conduct EL on video transcriptions surrounding two topics—"machine learning" and "climate change"—to speed up video search. Toulet et al. (2022) provide an EL-integrated pipeline and a visualization tool for analyzing scientific articles, including a case study on co-occurring "climate change" and "health" topics. Pita Costa et al. (2024) rely on EL as a semantic annotation to analyze the impact of water-related climatic disasters. However, no evaluation of EL performance on English CC data across genres exists.

3 CLIMATELI: CLIMATe Entity LInking

We present CLIMATELI, the first EL evaluation corpus on CC data. We include ten English documents across five genres—Wikipedia pages, academic articles,[2] web news,[3] United Nations' Intergovernmental Panel on Climate Change reports,[4] and YouTube transcriptions—consisting of 3,087 (1,161 unique) entity links across 12,802 tokens.

Entity Linking Manual EL annotation from scratch is time-consuming and might result in low recall. This is mostly due to annotators' inability to encompass all contents from ever-expanding knowledge bases (KBs). We pick Wikipedia as our target knowledge base due to its enormous size[5] and its integration into our daily life.

Firstly, we manually correct tokenization and POS-tagging from stanza (Qi et al., 2020) predictions. We then create CLIMATELI by opting to manually correct pre-tagged linked entities from Wikifier's (Brank et al., 2017) threshold 1.0 (highest recall) predictions, which provide an extensive baseline for gold annotations while also adding missing entities. We use the markdown style `[Document Tokens](Link_URL)` to ease annotation as shown in Figure 1. Annotations include removing or correcting wrongly detected entity spans and links predicted by Wikifier and adding missing entities not annotated by Wikifier. We also verify whether individual links exist on Wikipedia and resolve various ambiguities.

We next also discuss several CLIMATELI guidelines. Firstly, we follow the flat schema of Wikipedia webpages to annotate only the longest entity without the shorter nested ones. For example, if "climate change mitigation" appears, we only link the entirety to `Climate_change_mitigation`, without annotating the shorter `Climate_change` or `Climate`. Secondly, we determine the associated Wikipedia link based on an entity's contextual reading. For example, when "it" refers to "climate change", we annotate "[it's effect](`Effects_of_Climate_Change`)". Similarly, if the Kyoto Protocol is in context, we annotate the subsequent "[the protocol](`Kyoto_Protocol`)." Lastly, since EL KBs are not tailored for CC, we follow Wikipedia to annotate ELs on generic verbs, adjectives, adverbs, conjunctions, and etc., if such entries exist. These include "refer" (a verbal form of `Reference`), "possible" (an adjectival form of `Logical_possibility`), "successfully" (an adverbial form of `Success`), "while" (`While`), and etc.

	token-level		entity-level		
	accuracy	cohen's κ	precision	recall	F1
untyped	94.31	89.93	89.90	82.63	86.12
typed	92.85	88.94	87.30	80.24	83.62

Table 1: CLIMATELI inter-annotator agreement.

Inter-Annotator Agreement Two authors of this paper, fluent English speakers, split the ten documents, each taking ~3 hours to annotate. We evaluate inter-annotator agreement (IAA) on one Wikipedia article, `Paris_Agreement`, which includes 1,371 tokens, and 334 or 307 entity link annotations by two annotators, respectively. Table 1 presents our IAA using accuracy, Cohen's kappa at the token level, precision, recall, and F1 at the entity level. We also include both untyped and typed results, where the former only matches the entity

[2]https://www.mdpi.com/

[3]https://globalvoices.org/

[4]https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_FullReport.pdf

[5]English Wikipedia contains 6.8M articles as of June 20, 2024, see `Wikipedia:Size_of_Wikipedia`.

Combos	Filters				Gold		Wikifier		TagMe		Cao et al.	
	Valid	*Nom*	*CText*	*CLink*	Total	Unique	Total	Unique	Total	Unique	Total	Unique
Orig					3,087	1,161	4,823	1,730	4,165	2,129	1,399	576
/	✓				3,061	1,141	4,783	1,714	4,072	2,082	1,304	511
N-only	✓	✓			2,346	881	2,587	1,008	2,764	1,333	1,106	437
C-only	✓		✓		1,831	558	2,030	605	1,875	683	977	338
/	✓			✓	958	230	1,009	211	855	220	479	139
NC-only	✓	✓	✓		1,586	491	1,554	481	1,557	565	896	316
/	✓	✓		✓	872	209	830	185	772	197	457	129

Table 2: Frequencies of all versus unique entity links in human and model annotations under different filters.

span, and the latter requires annotating the same Wikipedia link. We achieve high (80%+) token and entity-level IAAs with precision scoring higher than recall on both untyped and typed entities.

4 Evaluation Setup

This section presents the experiment setups for evaluating entity-linking models on CLIMATELI data.

Entity Linkers We employ three Wikipedia linking models frequently used in NLP and social science to generate EL predictions. Wikifier (Brank et al., 2017) and TagMe (Ferragina and Scaiella, 2010) are easy to use and allow users to configure the confidence thresholds on predicted entity links. Additionally, we include a BART-based (Lewis et al., 2020) generative sequence-to-sequence EL model, Cao et al. (2021).[6] Since entity spans differ vastly across ELs, we conduct a post-processing step to normalize predicted entities: removing leading determiners from nouns and dropping nested entities following Wikipedia's style, which displays no nested or overlapping ELs.

Entity Filters We design filters to focus our EL evaluation on valid, nominal, and climate entities.

- *Valid links (Valid)*: we discard disambiguation and content-less pages, e.g., `Reduction` and `Climate_overshoot`, and invalid URLs;
- *Nominal (Nom)*: Since ELs such as Wikifier provide links to verbs, e.g., *'he [thinks](Thought) ...'*, we remove non-nominal entities whose entirety are verbs, adjectives, adverbs, etc.;
- *Climate Text or Link*: we only retain Wikipedia links that either include the word "climate" (*CText*) or the link `Climate_change` (*CLink*).

Table 2 presents the frequency of linked entities and unique ones from gold human annotations and model predictions in different filtering scenarios.

[6]We use Cao et al. rather than GENRE (Generative ENtity REtrieval) to refer to the third EL model to avoid confusion between the model and CLIMATELI's text genres.

We observe that *Nom* reduces valid gold entities from 3,061 to 2,346, *CText* to 1,831, and *CLink* to 958. The number of unique gold entities is halved after *CText*-filtering and quartered by *CLink*, and drops more dramatically on model predictions. Since *CLink* is more restrictive and has a lower recall than *CText*, we use *CText* as the climate filter in evaluations. Moreover, Wikifier and TagMe generate more predictions than humans, whereas Cao et al. is more conservative. In addition to the annotations on 12K tokens, we release a list of 1,251 CC-related Wikipedia links for future research.

5 Results

This section evaluates four filtering scenarios, comparing *Valid Nom* and *Valid CText* versions (*N-only* and *C-only*) and their intersections (*NC-only)* to unfiltered (*Orig*) entities. For Wikifier and TagMe, we use the default confidence threshold with the highest recall. Table 3 presents the overall token- and entity-level performances on untyped and typed entities under four filtering scenarios.

Token-level As the filtering conditions become more stringent, the accuracy increases for all three models. Namely, *NC-only* achieves the highest accuracy, followed by *C-only* and *N-only*, and *Orig* scores the lowest. TagMe exhibits the largest disparity between typed and untyped token accuracy among the three models, but this difference decreases after adding filters. Nevertheless, comparing token-level accuracy between unfiltered and filtered versions is unfair since the latter has more non-entity tokens and raises chance agreement.

Entity-level Entity-level results reveal difficulties in EL, with all typed F1s below 60%. Wikifier remains the winner for both untyped and typed F1s. TagMe achieves satisfying performance on retrieving entity spans but deteriorates largely on typed scores. Cao et al. perform the worst on untyped scores, but its degradation to typed is relatively small. Besides, due to fewer predicted entities (cf.

Combos	Models	untyped				typed			
		accuracy	precision	recall	F1	accuracy	precision	recall	F1
Orig	Wikifier	**76.02**	49.55	**77.42**	**60.43**	68.64	38.36	**59.93**	**46.78**
	Tagme	75.39	47.71	64.37	54.80	57.03	16.47	22.22	18.92
	Cao et al.	72.98	**55.90**	25.33	34.86	**70.15**	**45.46**	20.60	28.35
N-only	Wikifier	**87.95**	**68.07**	**75.06**	**71.40**	**82.67**	52.34	**57.72**	**54.90**
	Tagme	84.82	57.20	67.39	61.88	70.30	20.12	23.70	21.76
	Cao et al.	79.94	63.56	29.97	40.73	77.67	**53.35**	25.15	34.18
C-only	Wikifier	**87.92**	**60.20**	**66.74**	**63.30**	**84.96**	**51.63**	**57.24**	**54.29**
	Tagme	86.73	54.35	55.65	54.99	76.61	22.13	22.67	22.40
	Cao et al.	84.12	58.34	31.13	40.60	82.14	50.56	26.98	35.19
NC-only	Wikifier	**90.27**	**68.02**	**66.65**	**67.32**	**87.62**	**57.79**	**56.62**	**57.20**
	Tagme	88.88	58.83	57.76	58.29	79.43	23.31	22.89	23.10
	Cao et al.	85.92	60.71	34.30	43.84	84.09	52.90	29.89	38.20

Table 3: Typed and untyped token-level accuracy and entity-level precision, recall, and F1 scores.

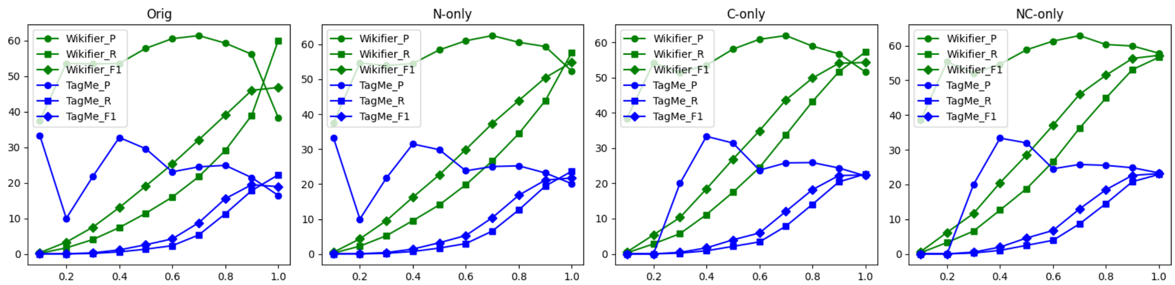

Figure 2: Typed entity precision, recall, and F1 of Wikifier and TagMe on thresholds 0.1 to 1.0 (min to max recall).

Table 2), Cao et al. demonstrates higher precision than recall, whereas TagMe and Wikifier prefer recall. However, on *NC-only* entities, Wikifier and TagMe's privilege in recall diminishes, and all three models exhibit higher precision than recall.

6 Quantitative Analysis

We examine two impacting factors on EL model scores: confidence thresholds and text genres.

Wikifier and TagMe Thresholds Figure 2 visualizes how different Wikifier and TagMe confidence thresholds affect model performances. Unsurprisingly, Wikifier beats TagMe on all thresholds. On the more filtered data, the recall of both models increases with the threshold. In contrast, precision fluctuates: Wikifier dips around 0.2 and peaks around 0.7; TagMe reaches its best precision around 0.4 and decreases thereafter. Moreover, it is intriguing that the precision and recall of both models converge at threshold 1.0 under the *NC-only* filter, which means the classifiers are balanced in terms of sensitivity and specificity.

Genres ClimatELi contains documents from 5 genres with different linguistic styles and discourse structures. Given our interest in CC, Table 4 presents genre performances on *NC-only* data. All three models achieve the highest performance in the vlog genre. Though TagMe and Cao et al. perform inferior to Wikifier, their advantages are remarkable in vlogs compared to other genres.

Figure 3 visualizes the distribution of entity lengths across various genres and illustrates that most (68.39%) entity spans in vlogs include only one token. This reduces the likelihood of entity nesting and decreases the difficulty of determining the positions of entity spans. Meanwhile, academic, ipcc, and wiki have considerably longer entities, possibly explaining TagMe and Cao et al.' sub-optimal performances in these three genres.

7 Error Types

This section further exemplifies common error types of the EL models' predictions.

Misinterpreting polysemous words as false nominals Polysemous tokens occurring in non-nominal positions were sometimes wrongly linked to their nominal interpretations. For example, the adjective "current" (i.e., belonging to the present time) receives a wrong link to `Ocean_current`; the coordinating conjunction "both" gets wrongly linked to a Trap song `Both_(song)`; the modal verb "will" gets falsely associated with `Free_will` (i.e., capacity to make decisions independently).

Models	Genre	precision	recall	F1
Wikifier	aca.	58.35	57.47	57.91
	ipcc	55.93	55.59	55.76
	news	57.77	57.19	57.48
	vlog	**64.85**	**69.03**	**66.88**
	wiki	55.73	51.48	53.52
TagMe	aca.	19.95	20.51	20.22
	ipcc	24.84	23.56	24.19
	news	26.51	26.42	26.47
	vlog	**43.71**	**42.58**	**43.14**
	wiki	15.21	14.53	14.86
Cao et al.	aca.	47.26	24.05	31.88
	ipcc	51.20	19.34	28.07
	news	61.02	36.12	45.38
	vlog	**66.40**	**53.55**	**59.29**
	wiki	46.27	30.54	36.80

Table 4: Typed entity scores on 5 *NC-only* genres.

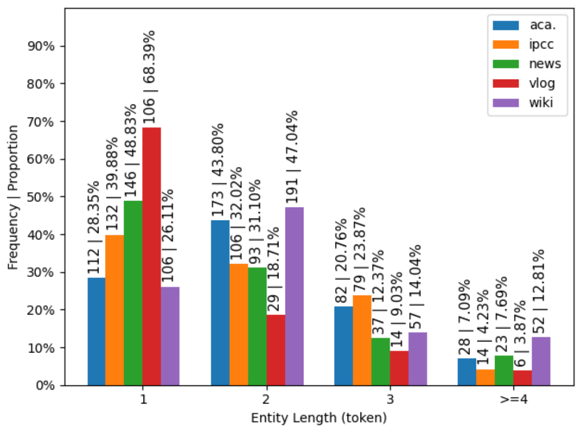

Figure 3: *NC-only* entity length distributions per genre.

Overly specifying generic nouns to particular readings Entity linkers, particularly Wikifier with threshold 1.0, are inclined to magnify the specificity of nominal terms and impose or enforce an association to some Wikipedia page. The "meeting" of governmental parties is distorted to `Confluence` (i.e., the joining of two watercourses into one); "organizations" in general gets overspecified to `Non-governmental_organization`. Particularly when the more general or common interpretation of a noun is absent in Wikipedia, models tend to impose an association with an overly specified EL. For example, a "step" as part of a plan is misrepresented as a `Step_dance`, a stair step (`Stairs`), or the mathematical `Step_function`. Similarly, most "growth"s are forcibly linked to `Population_growth` or `Economic_growth` even when referring to other objects' increase in size.

Unable to capture contextual readings In some cases, models succeed in annotating "[its effects](`Effects_of_climate_change`)" when the pronoun "it" refers to `Climate_change` or "Both [the EU](`European_Union`) and [its member states](`Member_state_of_the_European_Union`)" when "it" refers `European_Union`. However, the coreference of "the"-headed definite common nouns is difficult. For example, when "the Accord" refers to `Copenhagen_Accord`, it gets falsely linked to `Prices_and_Incomes_Accord` in Australia. Similarly, a coreferring "the agreement" to `Paris_Agreement` is always mis-interpreted as `Joint_Comprehensive_Plan_of_Action`, i.e., the Iran Nuclear Deal. The more tricky situation is when some ELs occur nearby but are irrelevant to the current entity. Still, EL models are falsely influenced by these contexts. In "the 195 UNFCCC participating member states and the European Union", TagMe annotates "member states" wrongly as `Member_state_of_the_European_Union`, and Wikifier annotates them as `European_Union`, while the gold label should be `List_of_parties_to_the_United_Nations_Framework_Convention_on_Climate_Change`.

Ambiguity caused by the flat schema Annotating the longest entity and ignoring the nested ones is common in named entity linking and recognition (McNamee and Dang, 2009; Ji et al., 2010; Tjong Kim Sang and De Meulder, 2003). However, this results in human label variation in annotating overlapping ELs. For example, "adapt to climate change effects" can be annotated as "[adapt to climate change effects](`Climate_change_adaptation`)", but can also be annotated as "[adapt to climate change](`Climate_change_adaptation`) [effects](`Effects_of_climate_change`)". Both interpretations are equally reasonable and converge with the annotation guidelines.

8 Conclusion

This paper presents ClimateEli, an entity linking corpus for English climate change data on five genres. We analyze existing EL systems and propose filters to focus the evaluation on nominal climate entities. We show that ELs struggle to detect long entity spans and link them to corresponding Wikipedia pages. Future work includes nested EL annotations since CC-related terminologies exhibit nesting and overlap, expanding annotated texts and Wikipedia links to more languages, and training a CC-adapted EL model for downstream NLP tasks.

Limitation

There are a few limitations in our work that we plan to improve in future research. Firstly, we only annotated flat entity linking without nesting; thus, embedded ELs and co-occurrences between nested entities are not fully captured. Secondly, our annotated texts and Wikipedia links are limited to English, and we only evaluate the performances of English EL models. Future expansion of multilingual texts and Wikipedia entries would benefit cross-lingual and cross-national comparison studies. Thirdly, although EL models benefit from being domain-generic, researchers are interested in evaluating them on specific domains and could follow different logical approaches. This paper employs simple rule-based filtering on manual annotations and model predictions to assess EL on CC-related data. However, with more CC-specific EL data available, we could finetune EL models on domain-specific data for direct evaluation.

Acknowledgements

This work belongs to the KLIMA-MEMES project funded by the Bavarian Research Institute for Digital Transformation (bidt), an institute of the Bavarian Academy of Sciences and Humanities. The authors are responsible for the content of this publication.

References

Janez Brank, Gregor Leban, and Marko Grobelnik. 2017. Annotating documents with relevant wikipedia concepts. In *Proceedings of the Slovenian Conference on Data Mining and Data Warehouses (SiKDD 2017)*, Ljubljana, Slovenia.

Nicola De Cao, Gautier Izacard, Sebastian Riedel, and Fabio Petroni. 2021. Autoregressive entity retrieval. In *International Conference on Learning Representations*.

Cecilia Conde and Kate Lonsdale. 2005. Engaging stakeholders in the adaptation process. pages 47–66. Cambridge University Press, Cambridge and New York.

Silviu Cucerzan. 2007. Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 708–716, Prague, Czech Republic. Association for Computational Linguistics.

Leon Derczynski, Diana Maynard, Giuseppe Rizzo, Marieke van Erp, Genevieve Gorrell, Raphaël Troncy, Johann Petrak, and Kalina Bontcheva. 2015. Analysis of named entity recognition and linking for tweets. *Information Processing & Management*, 51(2):32–49.

Thomas Diggelmann, Jordan Boyd-Graber, Jannis Bulian, Massimiliano Ciaramita, and Markus Leippold. 2021. CLIMATE-FEVER: A Dataset for Verification of Real-World Climate Claims. ArXiv:2012.00614 [cs].

Paolo Ferragina and Ugo Scaiella. 2010. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, page 1625–1628, New York, NY, USA. Association for Computing Machinery.

Ben Hachey, Will Radford, Joel Nothman, Matthew Honnibal, and James R. Curran. 2013. Evaluating entity linking with wikipedia. *Artificial Intelligence*, 194:130–150. Artificial Intelligence, Wikipedia and Semi-Structured Resources.

Heng Ji, Ralph Grishman, HT Dang, K Griffit, and J Ellis. 2010. Overview of the tac 2010 knowledge base population track. In *Proceedings of the 2010 Text Analysis Conference*.

Jan-Christoph Klie, Richard Eckart de Castilho, and Iryna Gurevych. 2020. From Zero to Hero: Human-In-The-Loop Entity Linking in Low Resource Domains. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6982–6993, Online. Association for Computational Linguistics.

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, Online. Association for Computational Linguistics.

Jessica Lin and Amir Zeldes. 2021. WikiGUM: Exhaustive entity linking for wikification in 12 genres. In *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 170–175, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Fangyu Liu, Ivan Vulić, Anna Korhonen, and Nigel Collier. 2021. Learning domain-specialised representations for cross-lingual biomedical entity linking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 565–574, Online. Association for Computational Linguistics.

Diana Maynard and Kalina Bontcheva. 2015. Understanding climate change tweets: an open source toolkit for social media analysis. In *Proceedings of EnviroInfo and ICT for Sustainability 2015*, pages 242–250. Atlantis Press.

Paul McNamee and Hoa Trang Dang. 2009. Overview of the tac 2009 knowledge base population track. In *Text analysis conference (TAC)*, volume 17, pages 111–113.

Pablo Mendes, Max Jakob, and Christian Bizer. 2012. DBpedia: A multilingual cross-domain knowledge base. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1813–1817, Istanbul, Turkey. European Language Resources Association (ELRA).

Pablo N. Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. Dbpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems*, I-Semantics '11, page 1–8, New York, NY, USA. Association for Computing Machinery.

Prakamya Mishra and Rohan Mittal. 2021. Neuralnere: Neural named entity relationship extraction for end-to-end climate change knowledge graph construction. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244.

Maria Pérez Ortiz, Sahan Bulathwela, Claire Dormann, Meghana Verma, Stefan Kreitmayer, Richard Noss, John Shawe-Taylor, Yvonne Rogers, and Emine Yilmaz. 2022. Watch less and uncover more: Could navigation tools help users search and explore videos? In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, CHIIR '22, page 90–101, New York, NY, USA. Association for Computing Machinery.

Jakub Piskorski, Nikolaos Nikolaidis, Nicolas Stefanovitch, Bonka Kotseva, Irene Vianini, Sopho Kharazi, and Jens P Linge. 2022. Exploring Data Augmentation for Classification of Climate Change Denial: Preliminary Study.

Joao Pita Costa, Luis Rei, Nejc Bezak, Matjaž Mikoš, M. Besher Massri, Inna Novalija, and Gregor Leban. 2024. Towards improved knowledge about water-related extremes based on news media information captured using artificial intelligence. *International Journal of Disaster Risk Reduction*, 100:104172.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Delip Rao, Paul McNamee, and Mark Dredze. 2013. *Entity Linking: Finding Extracted Entities in a Knowledge Base*, pages 93–115. Springer Berlin Heidelberg, Berlin, Heidelberg.

Pablo Ruiz Fabo, Clément Plancq, and Thierry Poibeau. 2016. More than word cooccurrence: Exploring support and opposition in international climate negotiations with semantic parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1902–1907, Portorož, Slovenia. European Language Resources Association (ELRA).

Wei Shen, Jianyong Wang, and Jiawei Han. 2015. Entity linking with a knowledge base: Issues, techniques, and solutions. *IEEE Transactions on Knowledge and Data Engineering*, 27(2):443–460.

Andy Spezzatti, Elham Kheradmand, Kartik Gupta, Marie Peras, and Roxaneh Zaminpeyma. 2022. Note: Leveraging artificial intelligence to build a data catalog and support research on the sustainable development goals. In *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, COMPASS '22, page 579–584, New York, NY, USA. Association for Computing Machinery.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

Anne Toulet, Franck Michel, Anna Bobasheva, Aline Menin, Sébastien Dupré, Marie-Claude Deboin, Marco Winckler, and Andon Tchechmedjiev. 2022. Issa: Generic pipeline, knowledge model and visualization tools to help scientists search and make sense of a scientific archive. In *The Semantic Web – ISWC 2022*, pages 660–677, Cham. Springer International Publishing.

Roopal Vaid, Kartikey Pant, and Manish Shrivastava. 2022. Towards Fine-grained Classification of Climate Change related Social Media Text. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 434–443, Dublin, Ireland. Association for Computational Linguistics.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Albert Weichselbraun, Daniel Streiff, and Arno Scharl. 2015. Consolidating heterogeneous enterprise data for named entity linking and web intelligence. *International Journal on Artificial Intelligence Tools*, 24(02):1540008.

Yi Yang and Ming-Wei Chang. 2015. S-MART: Novel tree-based structured learning algorithms applied to tweet entity linking. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 504–513, Beijing, China. Association for Computational Linguistics.