

Response to Referee#1

Manuscript title: **From text to geoinformation – A modular approach for extraction of disaster information from web text data**

RC1: '[Comment on egusphere-2026-2361](#)', Anonymous Referee #1, 24 Jun 2026

Referee comments are written in black italics; author's responses are written in blue and normal font; and proposed changes are highlighted in red italics, Times New Roman.

We would like to extend our sincere thanks to the referee for their time and valuable feedback. The comments provided have played a key role in improving the clarity and academic rigour of the manuscript, thereby helping the research to stand out more clearly. We have revised the manuscript accordingly and have addressed all your comments below. Please find attached our detailed responses.

If you have any further questions, we would be happy to discuss them.

Kind regards,

Vanessa Rittlinger on behalf of all Co-authors

RC1 Comment 1: *This paper presents a workflow for extracting disaster related information from Web text. The authors demonstrated this workflow in a case study on the 2021 flooding event in West Germany. Overall, this paper is more like a project report and lacks a clear research contribution.*

RC1 Answer Comment 1:

We thank the referee for the careful evaluation of our manuscript. We agree that the scientific contribution needed to be highlighted more clearly in the original submission. We will thoroughly revise the manuscript and, in particular, fundamentally rework the introduction to clarify the methodological approach and research contribution, as well as to answer clearer our research questions. In the following, we would like to emphasise once again what we consider to be our scientific contribution. Our key contribution lies in the development of an integrated, domain-specific workflow for extracting georeferenced disaster-related information (DRI) from unstructured web texts, which, to the best of our knowledge, has not yet been published in such detail. Specifically, we are contributing to the advancement of the field by: (1) the integration of established NLP components (NER, geocoding, entity linking) into a coherent, end-to-end pipeline optimised for the extraction of disaster-related information at the finest possible geospatial resolution; (2) focusing on formal web sources via text-level entity extraction, which we position as a complement to document-level classification and data derived from social media; and (3) the provision of a modular, reproducible framework that can be adapted to different types of disasters and shifts the analytical focus from document-level classification by hazard type to entity-level identification across the entire disaster management cycle.

RC1 Comment 2: - *Web text is known to have various biases. Depending on the disaster, some aspects of the disaster may be reported while some other aspects may be ignored. This study specifically used Web news from the GDELT project, and the extracted information will inherit all these biases. This is not to mention the NLP and other methods used to process the text, which have their own algorithmic biases.*

RC1 Answer Comment 2:

We agree with the referee that both web news archives such as GDELT and NLP pipelines introduce selection, sampling and algorithmic biases that compromise the completeness and representativeness of the extracted DRI. The aim of our study is to demonstrate a reproducible framework for processing large-scale text archives in the context of disasters and to show how such data complement, rather than replace, traditional geospatial sources. In the revised manuscript, we will explain this distinction in more detail, along with the selective aspects of the text content and existing information gaps, in sections 5.1 and 5.2 of the discussion, to enable a critical interpretation of the results. Furthermore, we clarify there that the documents were pre-processed using specific relevance criteria, thereby reducing reliance on the raw GDELT classifications.

As an example, we propose the following adjustments in the Section 5.1:

“Nevertheless, we still face certain limitations in our process of using unstructured text data to generate structured geospatial data. Along the modular workflow, several processing steps contribute individually to uncertainties, and their errors are propagated to subsequent steps. For this case study using news data, which is comparatively systematic and formal, our approach of using dictionaries and rule sets is suitable and excels at transparency and speed. However, the evaluation of the dictionary is limited because, to our knowledge, no complete dictionary of disaster-related terms exists. We are aware of the limitations and biases, such as selection and description bias of texts (Hoffmann et al., 2022), nevertheless we want to examine the existing text data and highlight its informational value and potential, as demonstrated by its combination with Satellite data. Naturally, our dictionary-based method is dependent on how comprehensive the dictionary is with respect to the disaster context. If a word variation is not in the list, it is not identified in the text. Similar words, synonyms, or other relevant information, that are not included in the dictionary are also missed (Reveilhac and Morselli, 2022), and cannot subsequently be linked to locations by the rule-based approach. Additionally, it is to mention that negations in the sentence structure are currently not identified. Large Language Model (LLM)-based approaches might be more robust when retrieving information from unstructured data and with less expert involvement.”

As an example, we propose the following adjustments in the Section 5.2:

“The substantial data volume generated by aggregating news content from the web presents significant challenges for systematic organization and quality assurance. The GDELT Project serves as a major aggregator of global news data, however, its data processing methods are not

fully disclosed, putting constraints on the interpretability of derived analyses in research contexts (Hoffmann et al., 2022). In our study, we use the GDELT-GKG database as a starting point for compiling a list of websites, to which we then apply a processing pipeline specifically tailored to the context of the case study. To this end, we also download the text content of the webpages ourselves. At present, we are focusing exclusively on the date extracted for each website from the GDELT metadata. Specific dates and times mentioned in the text that relate directly to the toponym and the DRI are not currently incorporated into our approach. Accordingly, we can use the time stamps in this analysis to show, at document level, when the web page containing the text was recorded in the database, thereby enabling us to track the media timeline of news coverage. Given the documented risk of misinformation in online content (Abraham et al., 2025; Lemoine-Rodríguez et al., 2024), we are aware of potential challenges regarding data reliability. This analysis focuses on evaluating the full range of available text data to determine where, when and for which disaster cycle classes information is available. Our data analysis is currently being validated using expert-tagged toponym-DRI relationships and by cross-referencing the results with published geospatial datasets. In future, more complex data analyses of web texts could be carried out by comparing the data with various other event-related datasets, e.g. with on-site data containing physio geographical information (Madruga de Brito et al., 2025). This underscores the relevance of a fundamental study in the field of web texts and NLP methods for a domain-specific workflow aimed at generating geospatial data.”

RC1 Comment 3: - *How long is the time delay in the Web news in GDELT? If the news in GDELT is largely delayed, then the extracted information is unlikely going to be useful for disaster response.*

RC1 Answer Comment 3:

We thank the referee for this critical point on the GDELT data with regard to its operational use. The GDELT Global Knowledge Graph (GKG) database is updated every 15 minutes (Owuor et al., 2020). Another factor influencing the timely availability of data for disaster relief depends primarily on the publication delay of news sources. In an operational mode where the database is accessed every 15 minutes, we consider near-real-time analysis to be feasible in an operational setting. In this study, however, we focus on a retrospective approach to assess the feasibility of the workflow and the possibility of data extraction, rather than testing operational real-time performance. We will explicitly highlight this retrospective approach and address further potential applications in the Conclusion. We will explicitly emphasise this retrospective approach. *For example, we propose to add in the introduction section: “We illustrate the proof-of-concept approach using a retrospective analysis of a past flood event in Germany, as well as its spatial and temporal development based on its footprint derived from web texts.”*

RC1 Comment 4: - *Related to the previous points, how would the information extracted from news be useful for disaster managers? Can the authors provide some concrete examples in which the information extracted from the news is something unknown to disaster managers?*

RC1 Answer Comment 4:

We agree with the referee that demonstrating the practical benefits for disaster management is of crucial importance and should be included in the revised manuscript. We see the benefit in integrating the information we have gathered, which may precede or complement conventional data streams, and that this data can therefore serve as an additional data source for analysing the event, providing us with spatial information that cannot be derived from conventional geospatial data sources, such as flood layers or building damage data. We will thoroughly revise the manuscript to make this intention and research aim clearer. The results are not intended to provide standalone, fundamental data on the event; we have merely investigated how detailed the information is and which levels of detail we can extract from text data relating to an event, alongside other data sources. In the Results section, we will therefore highlight clearer examples of the information obtained. Examples include how, amongst other things, the deployment of volunteers can be identified, or how media reports can be used to identify the emergency services involved and their geographical deployment. Or how one can filter information on infrastructure damage, such as to railways or bridges, or the mention of measures such as retention basins and their geographical location. In the revised manuscript, we will outline concrete examples of a deployment scenario involving extracted text-based insights to the section 5.4. *“Based on our findings regarding potential ways of analysing web text data in comparison with other geospatial data sources in a disaster context, we see the following potential applications that can be derived from web-based information: for disaster managers and other areas of research, by detecting early signs of infrastructure failures, identifying local access problems, highlighting public perception or misinformation, or uncovering secondary hazards. In this way, information from this perspective can be incorporated as an additional data source from the media sector.”*

RC1 Comment 5: - *This paper proposes a general workflow for extracting multiple types of information from Web text. However, there is no comparison with previous methods. A concrete research paper would focus on extracting one or two types of information (e.g., topics or locations) and compare the proposed new method with existing methods.*

RC1 Answer Comment 5:

We thank the referee for emphasising the importance of benchmarking. We acknowledge that component-wise comparison against state-of-the-art methods is the standard evaluation protocol in algorithmically focused research. Our contribution, however, lies not in individual algorithmic advances but in the architectural integration and domain

adaptation of existing methods into a coherent pipeline for extracting georeferenced disaster-relevant information at variations of spatial detail. Accordingly, we evaluate the pipeline end-to-end against a manually annotated reference dataset and refer, for the individual components (NER, geocoding, relation extraction), to the benchmark results published in their respective methodological papers (e.g., FlairNER: Akbik et al., 2018; Hu et al., 2024). In the revised manuscript we make this evaluation strategy more explicit. **We propose to add for example a part of a sentence in section 3.2, line 193:** *“We use the FlairNER model, which has already proven its worth in the successful integration of toponym identification into German texts (Akbik et al., 2019), and benchmarked for toponym identification (Hu et al., 2024).”*

RC1 Comment 6: - *The methods used in the workflow seem to be outdated, such as TF-IDF which is an old information retrieval technique. It is unclear what the methodological innovation of this paper is.*

RC1 Answer Comment 6:

We wish to clarify that the innovative contribution of this study lies in the architectural integration of NER and spatial resolution for a specific disaster domain, and not in the introduction of novel machine learning architectural models. We made a conscious decision to integrate the TF-IDF approach into only a small part of the pipeline, namely the initial creation of the flood terminology dictionary, as, to the best of our knowledge, no existing, linguistically optimised German flood terminology dictionary is available. In combination with expert curation, this approach ensured a high degree of interpretability and minimised the risk of misinterpretations and a lack of traceability, as can occur, for example, with LLM models. Where practicable, we have integrated established methods (e.g. FlairNER) into the subsequent analysis. We would like to point out that TF-IDF was used exclusively for the one-off creation of the keyword dictionary and not as a recurring component of the pipeline. **We propose making adjustments in Section 3.1 to clarify our methodological intent and to present the application of this method more clearly:** *“As the identification of relevant key terms is standard practice in the context of crisis management, we have adopted a pragmatic approach to create a robust and practical keyword dictionary. As part of our dictionary compilation process, we utilise a corpus compiled by our subject matter experts, consisting of German reports and documents on flooding, flood events and heavy rainfall sourced from the official websites of the BKG or regional authorities. To identify relevant terms and keywords from these documents, as well as from additional expert knowledge, we apply a TF-IDF (Term Frequency–Inverse Document Frequency) approach (Senaratne et al., 2023) , which filters out relevant words and keywords. As our database consists of documents in German, we must take into account variations in word suffixes depending on gender, case and singular or plural form (Kharis et al., 2021). We have opted for an approach in which we include variations of the keywords in the dictionary based on their lemma, rather than lemmatising all words and the entire text (Table 1).”*

RC1 Comment 7: - The performance of the workflow as reported in Figure 5 is low, with F1-score being about 0.4. It is unclear whether this workflow can extract information accurately.

RC1 Answer Comment 7:

We would like to thank the referee for raising this point regarding our evaluation results. We are aware that, compared with other domains, this F1 score indicates a moderate level of performance. Our method integrates several advanced text-processing tasks: domain-specific entity recognition, context-based disambiguation, spatial inference and precise coordinate alignment; to the best of our knowledge, this has not yet been achieved on a comparable scale in the literature. We have thus laid the foundations for this extremely demanding, multi-faceted challenge. Whilst the aggregated F1 score indicates moderate performance, the component-level evaluation reveals higher accuracy in the recognition of topographical entities and the identification of DRI keywords. We will present these additional results in the appendix. The relational linking phase represents the primary bottleneck, a well-documented challenge in relation extraction tasks (Wadhwa et al., 2023). Although addressing this challenge through algorithmic improvements would go beyond the scope of this article, our approach provides a foundation for potential methodological advances in domain-specific areas. Based on our comparisons with the geographical data relating to the event, as well as with further information from other documents (Section 5.3), we can conclude that we are able to extract information that corresponds with these details. We are, however, aware that there is a margin of error, which we will also highlight more clearly in the revised manuscript. **We propose to modify the last paragraph in section 5.1:**

“In the final stage of the pipeline, the linking of toponyms and DRIs, the rule-based entity linking process is constrained by its static heuristic rules. We acknowledge the associated high false-positive rate and moderate precision values; accordingly, we employ the statistical distribution of sentence distances between place names and DRIs, derived from our reference data, as a proxy to define matching thresholds. Furthermore, relationships between multiple place names and multiple DRIs within a single sentence are not currently resolved, which may lead to incorrect assignments. However, comparative studies on relation extraction indicate that performance is highly sensitive to method, dataset, and domain specificity, with F1 scores ranging from 38% to 86% across different contexts (Detroja et al., 2023; Nasar et al., 2021), aligning with our results. Given the domain-specific nature of disaster text and the proof-of-concept scope of this study, we consider this performance acceptable for establishing a baseline. While the potential for integrating Large Language Models (LLMs) into structured information extraction and entity linking processes is evident (Dagdelen et al., 2024), our approach achieves a straightforward, simple, and rapid method for contextualizing information. Thus, this study provides a reproducible foundation, with future research directed toward integrating advanced models to resolve complex multi-entity relationships.”

RC1 Comment 8: - *This paper is based on a single case study on a flooding event. A stronger paper would have another case study, ideally on another type of disaster, to demonstrate the generalizability of this workflow.*

RC1 Answer Comment 8:

We acknowledge the referee's valuable suggestion. We recognise that validation based on several case studies underpins the generalisability of the findings. The 2021 flood disaster in western Germany was selected because extensive digital datasets are available for this event, enabling big data analysis and comparison with other geodata. The resulting architecture is specifically designed to enable reconfiguration for different types of hazards. Encouraged by the referee's suggestion, we have carried out a second validation case study and included it in the paper: the flood event in southern Germany in May 2024. We propose presenting this second study area in a new Section 3.4, 'Transferability Study', and presenting the results in Section 4.5, 'Transferability Assessment'. The current results confirm the applicability of the approach to a further event across different time periods and regions. In this study, we carry out a validation based on two flood events to demonstrate the temporal and spatial generalisability within the context of flooding. The extension to other types of hazards (e.g. earthquakes, forest fires) will be added in Section 5.4 'Discussion' as potential future applications; this primarily requires the replacement of the domain-specific vocabulary and the reference corpus.

We propose to add a section 3.4 Transferability study

To assess the transferability of our pipeline, we have also applied it to another flood event: The 2024 floods in southern Germany (Bayerisches Landesamt für Umwelt (LfU), 2025). To this end, we have likewise extracted web text data from a collection of GDELT news data covering the period 29th May 2024 to 13th June 2024 of the event and presented the results. We therefore analyse 1125 documents for this period and focus area.

We propose to add a section 4.5 Transferability assessment

Our method is applied to another flood event in Germany. To this end, we extracted data from the GDELT database, processed it using the same procedure, and fed it into the pipeline. For this event we can extract in total 41590 DRI points. As can be seen from Figure 12, information on damage reports and the disaster itself can also be extracted from the texts relating to this event, providing details of the damage and the course of events in the affected area as well as during the flood period. These are particularly high in areas with large inundated areas, but also show correlations in areas with smaller inundated areas. However, not all areas (grey/reddish) are mentioned in our text data in connection with the flood.

Aggregation of the absolute number of media mentions of infrastructure damage per municipality between 29 May 2024 and 13 June 2024

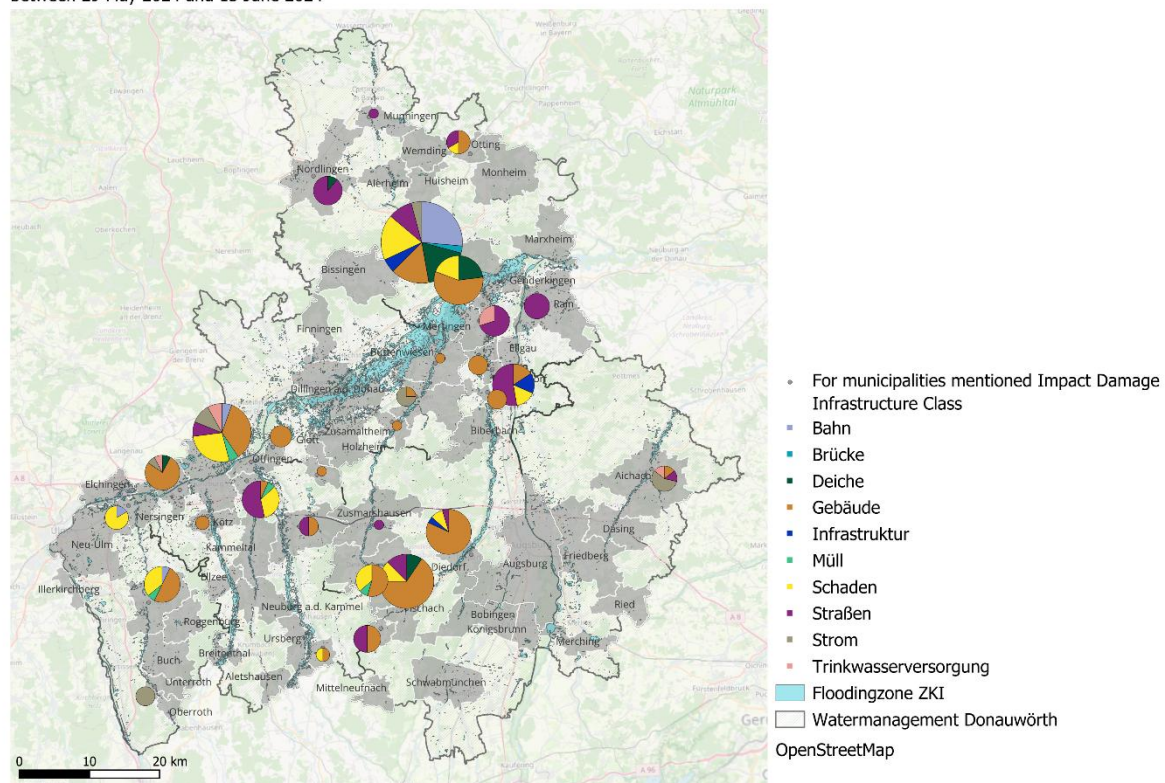


Figure 12: Map extract showing Donauwörth during the flood event in June 2024, illustrating the results of the workflow in a further region. At local authority level, the aggregated mentions of damage to infrastructure in web news reports for the event period are presented; these are broken down by the DRI damage data and the number of mentions in the pie chart, and visualised according to scale.

References

- Abraham, T. M., Wen, T., Wu, T., and Chen, Y.: Leveraging data analytics for detection and impact evaluation of fake news and deepfakes in social networks, *Humanit. Soc. Sci. Commun.*, 12, 1040, <https://doi.org/10.1057/s41599-025-05389-4>, 2025.
- Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., and Vollgraf, R.: FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP, in: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, Minneapolis, Minnesota, 54–59, <https://doi.org/10.18653/v1/N19-4010>, 2019.
- Bayerisches Landesamt für Umwelt (LfU): Hochwasser Mai/Juni 2024 - Wasserwirtschaftlicher Bericht, 2025.
- Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A. S., Ceder, G., Persson, K. A., and Jain, A.: Structured information extraction from scientific text with large language models, *Nat. Commun.*, 15, 1418, <https://doi.org/10.1038/s41467-024-45563-x>, 2024.
- Detroja, K., Bhensdadia, C. K., and Bhatt, B. S.: A survey on Relation Extraction, *Intell. Syst. Appl.*, 19, 200244, <https://doi.org/10.1016/j.iswa.2023.200244>, 2023.
- Hoffmann, M., Santos, F. G., Neumayer, C., and Mercea, D.: Lifting the Veil on the Use of Big Data News Repositories: A Documentation and Critical Discussion of A Protest Event Analysis, *Commun. Methods Meas.*, 16, 283–302, <https://doi.org/10.1080/19312458.2022.2128099>, 2022.
- Kharis, M., Kisyani, Suhartono, Pairin, U., and Darni: How to Lemmatize German Words with NLP-Spacy Lemmatizer?: *International Seminar on Language, Education, and Culture (ISoLEC 2021)*, Malang, Indonesia, <https://doi.org/10.2991/assehr.k.211212.036>, 2021.

- Lemoine-Rodríguez, R., Mast, J., Mühlbauer, M., Mandery, N., Biewer, C., and Taubenböck, H.: The voices of the displaced: Mobility and Twitter conversations of migrants of Ukraine in 2022, *Inf. Process. Manag.*, 61, 103670, <https://doi.org/10.1016/j.ipm.2024.103670>, 2024.
- Madruga de Brito, M., Sodoge, J., Kreibich, H., and Kuhlicke, C.: Comprehensive Assessment of Flood Socioeconomic Impacts Through Text-Mining, *Water Resour. Res.*, 61, e2024WR037813, <https://doi.org/10.1029/2024WR037813>, 2025.
- Nasar, Z., Jaffry, S. W., and Malik, M. K.: Named Entity Recognition and Relation Extraction: State-of-the-Art, *ACM Comput Surv*, 54, 20:1-20:39, <https://doi.org/10.1145/3445965>, 2021.
- Reveilhac, M. and Morselli, D.: Dictionary-based and machine learning classification approaches: a comparison for tonality and frame detection on Twitter data, *Polit. Res. Exch.*, 4, 2029217, <https://doi.org/10.1080/2474736X.2022.2029217>, 2022.
- Senaratne, H., Mühlbauer, M., Götzer, S., Riedlinger, T., and Taubenböck, H.: Detecting crisis events from unstructured text data using signal words as crisis determinants, *Int. J. Digit. Earth*, 16, 4601–4620, <https://doi.org/10.1080/17538947.2023.2278714>, 2023.