Displaying Evolving Events Via Hierarchical Information Threads for Sensitivity Review

Hitarth Narvala(✉), Graham McDonald, and Iadh Ounis

University of Glasgow, Glasgow, UK
h.narvala.1@research.gla.ac.uk,
{graham.mcdonald,iadh.ounis}@glasgow.ac.uk

Abstract. Many government documents contain sensitive (e.g. personal or confidential) information that must be protected before the documents can be released to the public. However, reviewing documents to identify sensitive information is a complex task, which often requires analysing multiple related documents that mention a particular context of sensitivity. For example, coherent information about evolving events, such as legal proceedings, is often dispersed across documents produced at different times. In this paper, we present a novel system for sensitivity review, which automatically identifies hierarchical information threads to capture diverse aspects of an event. In particular, our system aims to assist sensitivity reviewers in making accurate sensitivity judgements efficiently by presenting hierarchical information threads that provide coherent and chronological information about an event's evolution. Through a user study, we demonstrate our system's effectiveness in improving the sensitivity reviewers' reviewing speed and accuracy compared to the traditional document-by-document review process.

Keywords: Sensitivity Review · Information Threading

1 Introduction

In many countries, government documents are legislated to be released to the public to comply with Freedom of Information (FOI) laws [12]. However, government documents often contain sensitive information (e.g. personal or confidential information). Therefore, the documents must undergo an exhaustive sensitivity review to identify and protect sensitive information before the documents can be released to the public. However, identifying sensitive information is a complex task that often requires analysing hidden connections between documents, such as coherent information about an event or a discussion. For example, a chronological evolution of a legal proceeding, which mentions sensitive information about the alleged crimes or the victims' personal details, could be spread across different documents. Sensitivity reviewers need to quickly find and understand the

N. Goharian et al. (Eds.): ECIR 2024, LNCS 14612, pp. 261–266, 2024.
https://doi.org/10.1007/978-3-031-56069-9_29

complete context of an event or discussion when they make sensitivity judgements. A coherent and chronological presentation of the information thread, which captures the event's evolution, can help the reviewers achieve this goal.

For general document review tasks (e.g. e-Discovery), various systems have been proposed (e.g. [1,13]) to perform complex exploratory search tasks and efficiently assess the relevance of a document for review. However, these systems do not principally focus on sensitivity review, where all documents that should be released to the public are relevant for sensitivity review. Recently, a few sensitivity review systems (e.g. [4,5]) have been proposed for searching and navigating a collection of sensitive documents, or to assist the reviewers with automatic sensitivity classification prediction. However, these systems are not focused on presenting coherent information to the reviewers, which can help them accurately and efficiently identify sensitivities from multiple related documents. More recently, Narvala et al. [8] proposed a system for sensitivity review that can identify and prioritise coherent groups of documents, i.e., high-level semantic categories [7] and finer-grained information threads [10] to improve the efficiency of the sensitivity review process. In particular, the system by Narvala et al. [8] leveraged *sequential* information threads [10] that present a coherent and chronological *sequence* of information about an event. However, sequential information threads may not effectively capture the diverse aspects pertaining to an event's evolution [9]. In particular, an event's evolution typically forms a *hierarchical* structure [9], where each branch of the hierarchy contains a chronologically evolving sequence of documents that describe a specific aspect relating to the event.

In this paper, we present a novel system for assisting sensitivity reviewers by identifying hierarchical information threads that capture diverse aspects of an event from multiple documents. By presenting hierarchical information threads, our system enables the reviewers to quickly provide accurate sensitivity judgements through the collective inspection of coherent information about an event's evolution across multiple documents. Moreover, we present a user study to show the effectiveness of our system in improving the reviewers' review speed and accuracy compared to the traditional document-by-document review process. You can try out our system at: http://demos.terrier.org/sensreview/thread-demo.

2 Hierarchical Threading for Sensitivity Review

We now provide details about effectively leveraging hierarchical information threads for accurate and efficient sensitivity reviews. In a traditional document-by-document sensitivity review, the reviewers are presented with different documents in a *sequence*, e.g., based on the documents' creation timestamp or using the semantic relatedness between the documents [7]. In contrast, our system enables a *collective* review of coherent information from multiple documents by identifying and presenting hierarchical information threads [9] to the reviewers. In particular, our system splits the task of sensitivity reviewing documents into

two stages: (1) **Thread Review**, where the reviewers can select a thread to review multiple passages from different documents that mention a particular event, and (2) **Document Review**, where the reviewers can review the entire document by taking references from the reviewed passages that were presented during the Thread Review stage.

Figure 1 shows the Thread Review interface, which presents a hierarchical information thread and allows the reviewers to provide sensitivity judgements (i.e., sensitive or non-sensitive) for each of the document passages in the thread. The reviewers can also highlight (i.e., annotate) the specific portion of text that contains any sensitive information in a passage. After submitting the review of all of the passages in the thread, the reviewers are presented with the documents that comprise any passages that were associated with the reviewed thread. Figure 2 shows the Document Review interface, which illustrates the sensitivity judgements for the passages that were deemed as sensitive (or non-sensitive) during the Thread Review stage (c.f. Fig. 1). In particular, the system enables the reviewer to refer to these passage reviews while reviewing the entire document to make more informed sensitivity judgements.

Architecture: The system is implemented in Python. Our system's review interfaces, shown in Figs. 1 and 2, are implemented using Django [3]. Our system deploys a layered architecture that comprises: (1) Data Layer, which stores the document collection, the identified threads and the reviewers' judgements, (2) Service Layer, which integrates the HINT [9] method to identify hierarchical information threads in the collection, and (3) Application Layer, which comprises the thread and document review interfaces to present the threads and documents to the reviewers, and captures their sensitivity judgements.

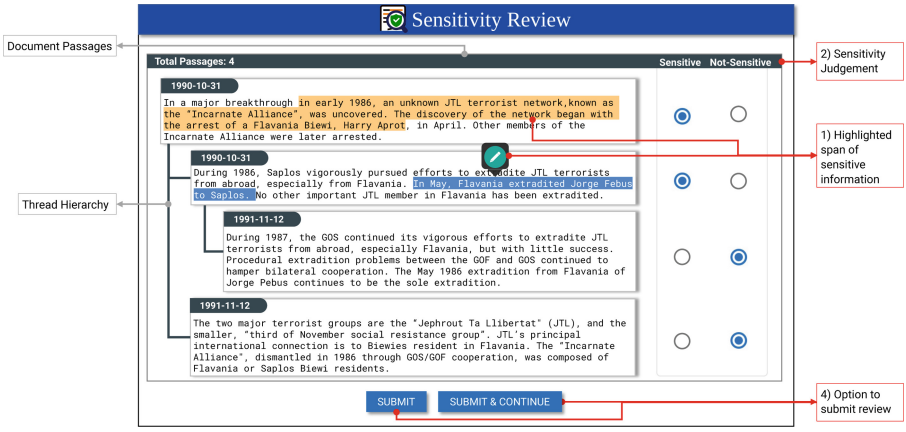

Fig. 1. Interface for collectively reviewing hierarchical information threads.

Fig. 2. Interface for reviewing a document after reviewing its corresponding threads.

3 Evaluation

We now present a user study, which evaluates the effectiveness of collectively reviewing related documents in automatically identified hierarchical information threads compared to the sequential (i.e., document-by-document) review in semantic category clusters [7]. We use the GovSensitivity collection [6] with real sensitivities to sample 25 documents (20% sensitive) to present to our study participants. We recruited 36 participants using the Prolific [11] crowdsourcing platform, and allocated the participants to review the documents using either threads or semantic clusters (i.e., 18 participants each). We measure the performance of participants using the BAC and NPS [2] metrics (same as used by Narvala et al. [7]). In particular, we use BAC to evaluate the accuracy of the participants' document reviews as well as the correctness of the reviews of passages that appear in the threads (to compute passage BAC, we use the participant's sensitivity annotations; c.f. Fig. 2). We use the NPS metric to evaluate the participants' reviewing speed.

Table 1 presents the study results. From Table 1 we observe that the participants who reviewed the threads before reviewing documents (i.e. the Thread condition) achieved significantly (independent samples t-Test; $p < 0.05$) higher Document BAC, Passage BAC, and NPS compared to the participants who reviewed documents in semantic clusters (i.e., the Cluster condition). These results demonstrate that collectively reviewing coherent threads can significantly improve the speed and accuracy of sensitivity reviewers compared to the document-by-document review. Moreover, by enabling the reviewers to analyse multiple related documents about a particular event, hierarchical information threads can assist the reviewers in accurately identifying specific portions (e.g. passages) of sensitivity within a document. Overall, our user study illustrates the promising benefits of our proposed system, where the incorporation of information threads into the sensitivity review process can improve both the effectiveness and efficiency of human sensitivity reviews.

Table 1. User study results. "⋆" denotes statistical significance (t-Test; $p < 0.05$).

Condition	Document BAC	Passage BAC	NPS (wpm)
Cluster (baseline)	0.653	0.625	209.42
Thread (our system)	**0.757**⋆	**0.709**⋆	**516.03**⋆

4 Conclusions

In this paper, we have presented a novel system to assist human sensitivity reviewers by enabling the collective review of coherent information about an event's evolution across multiple documents. Our system identifies hierarchical information threads to help the sensitivity reviewers in quickly and accurately identifying specific portions of sensitive information. Using a user study, we showed that presenting hierarchical information threads can significantly ($p < 0.05$) improve the reviewers' reviewing speed and accuracy (+146.41% NPS and +15.93% BAC) of the provided sensitivity judgements for the documents compared to traditional document-by-document review. Therefore, our system has the potential to aid government departments in complying with FOI laws by expediting the release of documents to the public while effectively protecting sensitive information.

References

1. Abualsaud, M., Ghelani, N., Zhang, H., Smucker, M.D., Cormack, G.V., Grossman, M.R.: A system for efficient high-recall retrieval. In: Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1317–1320 (2018). https://doi.org/10.1145/3209978.3210176
2. Damessie, T.T., Scholer, F., Culpepper, J.S.: The influence of topic difficulty, relevance level, and document ordering on relevance judging. In: Proceedings of the 21st Australasian Document Computing Symposium, pp. 41–48 (2016). https://doi.org/10.1145/3015022.3015033
3. Django Software Foundation (2021). https://djangoproject.com. Accessed 21 Oct 2023
4. McDonald, G., Macdonald, C., Ounis, I.: How the accuracy and confidence of sensitivity classification affects digital sensitivity review. ACM Trans. Inf. Syst. **39**(1), 1–34 (2020). https://doi.org/10.1145/3417334
5. Narvala, H., McDonald, G., Ounis, I.: Receptor: a platform for exploring latent relations in sensitive documents. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2161–2164 (2020). https://doi.org/10.1145/3397271.3401407
6. Narvala, H., McDonald, G., Ounis, I.: RelDiff: enriching knowledge graph relation representations for sensitivity classification. In: Findings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 3671–3681 (2021). https://doi.org/10.18653/v1/2021.findings-emnlp.311
7. Narvala, H., McDonald, G., Ounis, I.: The role of latent semantic categories and clustering in enhancing the efficiency of human sensitivity review. In: Proceedings

of the 2022 Conference on Human Information Interaction and Retrieval, pp. 56–66 (2022). https://doi.org/10.1145/3498366.3505824
8. Narvala, H., McDonald, G., Ounis, I.: Sensitivity review of large collections by identifying and prioritising coherent documents groups. In: Proceedings of the 31st ACM International Conference on Information and Knowledge Management, pp. 4931–4935 (2022). https://doi.org/10.1145/3511808.3557182
9. Narvala, H., McDonald, G., Ounis, I.: Effective hierarchical information threading using network community detection. In: Proceedings of 45th European Conference on Information Retrieval, pp. 701–716 (2023). https://doi.org/10.1007/978-3-031-28244-7_44
10. Narvala, H., McDonald, G., Ounis, I.: Identifying chronological and coherent information threads using 5W1H questions and temporal relationships. Inf. Process. Manag. **60**(3), 103274 (2023). https://doi.org/10.1016/j.ipm.2023.103274
11. Prolific (2014). https://www.prolific.com. Accessed 21 Dec 2023
12. UNESCO: Global report on the implementation of access to information laws (2022). https://unesdoc.unesco.org/ark:/48223/pf0000383160
13. Vo, N.P.A., Guillot, F., Privault, C.: DISCO: a system leveraging semantic search in document review. In: Proceedings of the 26th International Conference on Computational Linguistics, pp. 64–68 (2016). https://aclanthology.org/C16-2014