News Article Retrieval in Context for Event-centric Narrative Creation

Nikos Voskarides[1*] **Edgar Meij**[2] **Sabrina Sauer**[3] **Maarten de Rijke**[4]
[1] Amazon, Barcelona, Spain
[2] Bloomberg, London, United Kingdom
[3] University of Groningen, Groningen, The Netherlands
[4] University of Amsterdam, Amsterdam, The Netherlands
nvvoskar@amazon.com, emeij@bloomberg.net
s.c.sauer@rug.nl, m.derijke@uva.nl

Abstract

Writers such as journalists often use automatic tools to find relevant content to include in their narratives. In this paper, we focus on supporting writers in the news domain to develop event-centric narratives. Given an incomplete narrative that specifies a main event and a context, we aim to retrieve news articles that discuss relevant events that would enable the continuation of the narrative. We formally define this task and propose a retrieval dataset construction procedure that relies on existing news articles to simulate incomplete narratives and relevant articles. Experiments on two datasets derived from this procedure show that state-of-the-art lexical and semantic rankers are not sufficient for this task. We show that combining those with a ranker that ranks articles by reverse chronological order outperforms those rankers alone. We also perform analysis of the results that sheds light on the characteristics of this task.[1]

1 Introduction

Professional writers such as journalists generate narratives centered around specific events or topics. As shown in recent studies, such writers envision automatic systems that suggest material relevant to the narrative they are creating (Diakopoulos, 2019). This material may provide background information or connections that can help writers generate new angles on the narrative and thus help engage the reader (Kirkpatrick, 2015).

Writers in the news domain often develop narratives around a single main event, and refer to other, related events that can serve different functions in relation to the narrative (van Dijk, 1988). These include explaining the cause or the context of the main event or providing supporting information (Choubey et al., 2020). Recent work has focused on automatically profiling news article content (i.e., paragraphs or sentences) in relation to their discourse function (Yarlott et al., 2018).

In this paper, instead of profiling existing narratives, we consider a scenario where a writer has generated an incomplete narrative about a specific event up to a certain point, and aims to explore other news articles that discuss relevant events to include in their narrative. A news article that discusses a different event from the past is relevant to the writer's incomplete narrative if it relates to the narrative's main event and to the *narrative's context*. Relevance to the narrative's main event is topical in nature but, importantly, relevance to the narrative's context is not only topical: to be relevant to the narrative's context, a news article should enable the continuation of the narrative by expanding the narrative discourse (Caswell and Dörr, 2018).

We model the problem of finding a relevant news article given an incomplete narrative as a retrieval task where the query is an incomplete narrative and the unit of retrieval is a news article. We automatically generate retrieval datasets for this task by harvesting links from existing narratives manually created by journalists. Using the generated datasets, we analyze the characteristics of this task and study the performance of different rankers on this task. We find that state-of-the-art lexical and semantic rankers are not sufficient for this task and that combining those with a ranker that ranks articles by their reverse chronological order outperforms those rankers alone.

Our main contributions are: (i) we propose the task of news article retrieval in context for event–centric narrative creation; (ii) we propose an automatic retrieval dataset construction procedure for this task; and (iii) we empirically evaluate the performance of different rankers on this task and perform an in-depth analysis of the results to better understand the characteristics of this task.

[*] Research conducted when the first author was at the University of Amsterdam.

[1] This is an extended abstract of a paper published at ACM ICTIR 2021: https://dl.acm.org/doi/10.1145/3471158.3472247.

Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022), pages 72 - 73

References

David Caswell and Konstantin Dörr. 2018. Automated Journalism 2.0: Event-driven narratives. *Journalism Practice*, 12(4).

Prafulla Kumar Choubey, Aaron Lee, Ruihong Huang, and Lu Wang. 2020. Discourse as a Function of Event: Profiling Discourse Structure in News Articles around the Main Event. In *ACL*. ACL.

Nicholas Diakopoulos. 2019. *Automating the News: How Algorithms Are Rewriting the Media*. Harvard University Press.

Keith Kirkpatrick. 2015. Putting the Data Science into Journalism. *Commun. ACM*, 58(5).

Teun A. van Dijk. 1988. *News as Discourse*. University of Groningen.

W. Victor Yarlott, Cristina Cornelio, Tian Gao, and Mark Finlayson. 2018. Identifying the Discourse Function of News Article Paragraphs. In *Workshop on Events and Stories in the News 2018*. ACL.