

# Event Information Retrieval from Text

Rashmi Sankepally

Department of Computer Science  
University of Maryland, College Park (USA)  
rashmi@umd.edu

## ABSTRACT

Events are an integral part of our day-to-day search needs. Users search for various kinds of events such as political events, organizational announcements, policy changes, personal events, criminal activity and so on. In linguistics, events are often thought of as discourse entities with associated complex structure and attributes.

Many professionals look for patterns that involve event occurrences. Journalists, financial analysts, intelligence analysts, attorneys conducting investigations, auditors examining corporate records are examples of users who may want to find such information and to arrange it in ways that can help to produce meaningful analyses.

The goal of this work is to develop effective information retrieval systems that can help users to satisfy event-related information needs. My particular interest is in events that are decomposable into subevents in ways that can be anticipated. I am interested in modeling decomposable events, and in automatically recognizing references to subevents, both to help with finding relevant documents and to help with presenting diverse results to the user.

I plan to pursue this broader goal in three stages, each of which involves creating a test collection. I have started by developing information retrieval test collections for news, building on existing collections of news stories from the Text Analysis Conference (TAC) [4] and the Topic Detection and Tracking Evaluations [1]. Next, I plan to build a new event ontology for email, and use that ontology as a basis for building an information retrieval collection from the Avocado Email collection. Finally, I plan to extend one or both of these test collections to support research on characterizing event impact, thus perhaps providing an additional basis for ranking.

In each stage of my work, new test collection(s) will enable new research. In my first stage, which is in progress, I have studied the effect of automatically detected subevents on ranking effectiveness. Using the Rich Entities, Relations and Events (ERE) ontology of event types and subtypes from the TAC Event track [5], and two existing automated event detection systems, I have developed a simple bag of words-and-events search system that uses the automatically detected event type information. I also built an information retrieval test collection from a Topic Detection and Tracking (TDT) collection, for which event-based topics exist. Evaluation results show promise when compared to baseline approaches. I plan to

further develop this line of work to identify ways to automatically decompose high level events into their components.

With the goal of extending the event retrieval work to an organizational setting, in my second line of research I plan to work with the Avocado email collection [3]. While newswire text and other Web pages are easily available for everyone's perusal, email content is interesting because it may contain organizational and personal events that differ from those found in news, and those events may be referred to in less fully contextualized ways. In this line of research, I want to build a retrieval system for events in this genre of personal and organizational content. The retrieval system will provide a diverse ranking covering the full range of subevents for which information is available. To do this, I will need a reusable test collection containing event-related topics, with the event nuggets within those documents annotated for relevance. This annotation at the event nugget level will support computing measures like  $\alpha$ -nDCG [2] in which the gain reflects in part the number of different event nuggets in a document.

The third stage of my work will be more exploratory, since there are many ways in which one might conceptualize event impact. My initial approach will be to explore alternative indicators for different types of impact. For example, the societal impact of news stories might be characterized by the number of readers, whereas the personal impact of an email might be characterized by the time before a reply is received. I am particularly interested in how sentiment analysis might be used to characterize event impact, particularly with reference to publicly available user-generated content.

## CCS CONCEPTS

• **Information systems** → **Information extraction**; *Ontologies*; *Novelty in information retrieval*; *Test collections*.

## KEYWORDS

event, event type, event detection, diversity, event impact

### ACM Reference Format:

Rashmi Sankepally. 2019. Event Information Retrieval from Text. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*, July 21–25, 2019, Paris, France. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3331184.3331415>

## REFERENCES

- [1] James Allan. 2002. Introduction to Topic Detection and Tracking. In *Topic Detection and Tracking*. Springer, 1–16.
- [2] Charles Clarke et al. 2008. Novelty and diversity in information retrieval evaluation. In *SIGIR*. 659–666.
- [3] Douglas Oard et al. 2015. Avocado Research Email Collection LDC2015T03. <https://catalog.ldc.upenn.edu/LDC2015T03>
- [4] Teruko Mitamura et al. 2017. Events Detection, Coreference and Sequencing: What's next? Overview of the TAC KBP 2017 Event Track. In *TAC*.
- [5] LDC. 2016. Rich ERE Annotation Guidelines Overview V4.2. [https://tac.nist.gov/2016/KBP/guidelines/summary\\_rich\\_ere\\_v4.2.pdf](https://tac.nist.gov/2016/KBP/guidelines/summary_rich_ere_v4.2.pdf)

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR '19, July 21–25, 2019, Paris, France

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6172-9/19/07.

<https://doi.org/10.1145/3331184.3331415>