Knowledge-Embedded Narrative Construction from Open Source Intelligence

Priyanka Ranade

University of Maryland, Baltimore County
Department of Computer Science & Electrical Engineering
1000 Hilltop Circle, Baltimore, MD 21250
priyankaranade@umbc.edu

Abstract

Storytelling is an innate part of language-based communication. Today, current events are reported via Open Source Intelligence (OSINT) sources like news websites, blogs, and discussion forums. Scattered and fragmented sources such as these can be better understood when organized as chains of event plot points, or *narratives*, that have the ability to communicate end to end stories. Though search engines can retrieve aggregated event information, they lack the ability to *sequence* relevant events together to form narratives about different topics. I propose an AI system inspired by Gustav Freytag's narrative theory called the *Plot Element Pyramid* and use knowledge graphs to represent, chain, and reason over narratives from disparately sourced event details to better comprehend convoluted, noisy information about critical events during intelligence analysis.

Introduction

Storytelling, and the delivery of *narratives*, have intuitively enabled human users to communicate and understand one another. Today, massive amounts of Open Source Intelligence (OSINT) about *events*, such as breaking news occurrences, are published daily over the Internet via blogs, social networks, and news sources. Breaking events related to higher level *topics* are typically fetched by querying search engines and news aggregators. Events contain *plot points*, which are incidents that directly impact the development of that particular event. For example, during a pandemic event, the development of a vaccine is an example of a plot point. This volume of events needs to be parsed by interested users who individually piece together and order event plot points to form coherent explanations about associated themes, context, and temporal developments. The process of piecing event plot points into a chronological and contextually-aware sequence is known as *narrative construction*. This process is generally convoluted, as events are reported and instantiated by potentially millions of users at a time, introducing noise such as misinformation, redundant or outdated data, and conflicting reporting. Current AI systems generally lack construction abilities that address above challenges.

While search engines have the ability to retrieve and rank events from an abundance of online information, support for *contextually tracking evolving event plots* and *ordering large volumes of unstructured, noisy events* into a narrative is an open research problem. The challenges associated with traditional search engines and constraints faced by human users when manually sequencing events, have simultaneously inspired my dissertation research of representing thematic narrative properties using semantic knowledge graphs. We believe that the use of narrative theories developed by linguists, AI researchers, storytellers, etc [5], can guide an AI system to contextually order events. Additionally, narrative theories are proven to be *universal*, enabling these systems to generalize across domains [1]. Specifically, I hypothesize that leveraging knowledge graphs for narrative construction will assist in automatically revealing hidden relationships between information sources and modeling the evolution of events through story plot lines.

The bulk of my thesis work will explore the development of novel representations that are able to generalize across a diversity of narrative types. Additionally, I will explore sequencing techniques for ordering events based on plot-driven perspective framing.

Existing Challenges and Thesis Goals

Current work in narrative understanding explores tasks such as story completion, generation, and construction. This dissertation is focused on narrative *construction* from sources such as online news, blogs, and social media.

Traditionally, a typical fragmented narrative construction problem is set up such that, a set of articles d_1, d_2..., d_n, retrieved by a keyword search query about a topic t, contain several events e_1, e_2..., e_n. Each event is extracted, ranked, and ordered into a chronological sequence. Consider a real-world example where an *intelligence analyst* requires information about a critical topic t. The analyst will input a query about t in a search engine and piece together information from a multiple set of sources (news blogs, social media posts) to understand related events. The results are ranked based on contextual relevance and similarity, rather than published time. In this scenario, the analyst will construct together the most applicable *plot points* to form a narrative that answers the query. An informative narrative is one that communicates the progression of plot points in events with little

to no information gaps. Current narrative construction methods do not provide support for ordering based on contextual features such as thematic frames, and are limited to *fill in* tasks, given start and end events [6, 2]. Constructing narratives based on temporal features alone does not support the *dynamic* nature of events reported online. New information is constantly being published, changing the scope of plots that make up a narrative associated with a particular event. For example, plot updates in an existing narrative may not be completely captured if published in the remote future. This causes information-gaps in the plot chaining process, diminishing the situational awareness about the overall event.

In order to address information gaps and contextual ordering constraints, my thesis specifically explores novel knowledge representation techniques grounded in narrative theory for constructing evolving narratives across a number of disparate OSINT sources. I will create an extensible ontology inspired by Gustav Freytag's narrative theory called the *Plot Element Pyramid* [1], which is a five-component model that outlines thematic and temporal stages in generic storytelling. The goal of this work is to embed traditional narrative plot properties to disparate online sources to contextually order events into *plot chains*. By grounding the construction process in narrative theory, we achieve two goals: (1) *Contextual ordering* can be assigned to event plot points: This narrative-based ordering is based on plot properties in an event, with respect to time, setting, and theme. For example, following Freytag's pyramid, an expository event will always precede a rising action event. Likewise, a rising action event will always precede a climax event and follow an expository event. (2): *Cross-domain interoperability* is supported: Universal plot elements have the ability to generalize across different domains [1]. This goal is particularly important when considering OSINT sources, which contain a variety of event types.

Knowledge Representation and Narrative Construction

My thesis includes the following research questions:

RQ1: How can narrative-inspired knowledge graphs effectively model narrative plots and enable contextual downstream plot ordering of events? - We propose the Narrative Event Ontology (NEO), which is an OWL ontology that represents Freytag's plot elements as classes. NEO will also extend global knowledge represented in existing event ontologies.

RQ2: Can we recognize narrative elements in unstructured text? - It is difficult to identify narrative features in natural language. Freytag's plot points have universal definitions that can be generalized to *any domain*, providing structure to the classification task. We will explore event and plot framing techniques for one domain through an extensive human evaluation study and use publicly quantified, factual sequences of plots in different events as ground truth. We will evaluate strategies to transfer the classified plot relationships to broader events and domains.

RQ3: Is there a way to chain events through ontological reasoning grounded in narrative theory? - The bulk of my thesis work will explore novel algorithms for sequencing event plots based on the class properties present in NEO. In particular, I seek to develop a set-theory based hierarchical knowledge graph question answering algorithm to group event plots by thematic context, time, and published source. I argue that a constructed event plot chain must fulfill three criteria: *(1) Semantic Relevance* - The events on the plot chain must be relevant to the input event query; *(2) Multivariance* - The events in each plot chain must be tightly coherent with each other and adapt to new or conflicting information; and *(3) Low Data Redundancy* - Many media outlets report the same information, causing unnecessary higher level processing on repeated data. The system should have the ability to prune information that is already represented in a knowledge graph.

Objective	Timeline
RQ1	May 2022- Sept 2022
RQ2	Oct 2022- March 2023
RQ3	April 2023-January 2024
Thesis Writing	February 2024-April 2024

Table 1: Research Timeline.

Preliminary Work and Research Timeline

My preliminary work has explored methods to fine-tune large language models to adapt to the cybersecurity domain [4, 3], and will be relevant to RQ2. In addition, our survey paper "Computational Understanding of Narratives" was recently accepted at IEEE Access. My current work addresses RQ1 and will be submitted to a high impact Neurosymbolic AI conference. My research timeline is shown in Table 1.

References

[1] Boyd, R. L.; Blackburn, K. G.; and Pennebaker, J. W. 2020. The narrative arc: Revealing core narrative structures through text analysis. *Science advances*, 6(32): eaba2196.

[2] Mostafazadeh, N.; Roth, M.; Louis, A.; Chambers, N.; and Allen, J. 2017. Lsdsem 2017 shared task: The story cloze test. In *2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*, 46–51.

[3] Ranade, P.; Piplai, A.; Joshi, A.; and Finin, T. 2021. CyBERT: Contextualized Embeddings for the Cybersecurity Domain. In *International Conference on Big Data (Big Data)*, 3334–3342. IEEE.

[4] Ranade, P.; Piplai, A.; Mittal, S.; Joshi, A.; and Finin, T. 2021. Generating fake cyber threat intelligence using transformer-based models. In *International Joint Conference on Neural Networks (IJCNN)*, 1–9. IEEE.

[5] Riedl, M. O.; and Young, R. M. 2010. Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, 39: 217–268.

[6] Zhu, X.; and Oates, T. 2012. Finding story chains in newswire articles. In *13th International Conference on Information Reuse & Integration (IRI)*, 93–100. IEEE.