Structured Narratives as a Framework for Journalism: A Work in Progress

David A. Caswell

Structured Stories
Los Angeles, California, USA
david@structuredstories.com

Abstract

This paper describes Structured Stories, a platform for producing and consuming journalism as structured narratives based on instantiations of event frames. The event frames are defined using FrameNet and are instantiated as structured events using references to nodes in various knowledge graphs. Structured narratives with recursive, fractal and network characteristics are then assembled from these structured events. The approach requires the direct reporting of journalistic events into structure by untrained reporters, and utilizes a simplified sequential user interface to achieve this. A prototype has been built and published, and is being applied to the reporting of local government journalism to explore editorial aspects of the approach.

1998 ACM Subject Classification 1.2.4 Knowledge Representation Formalisms and Methods, F.4.1 Frames and Scripts

Keywords and phrases journalism, frames, narrative structure, event representation

Digital Object Identifier 10.4230/OASIcs.CMN.2015.39

1 Introduction

Journalism has historically been produced, distributed and consumed within the context of loosely-defined supra-document products such as edited newspapers and magazines. These products provide not merely collections of discrete text articles, but also larger-scale informal narrative functions across articles, such as story continuity, consistency of voice, de-duplication, indicators of importance, variance in detail, loose organization of sub-narratives, etc. They are often perceived by their producers and by their consumers to be conveyers of coherent supra-document narratives [3].

More recently, due to technological disruption, the economic basis of these products has started to break down, or 'unbundle', and they are increasingly being replaced by digital streams of isolated text documents, often clustered and ranked using topic models and named entity references. This unbundling has had negative consequences for professional journalism producers, for whom the economic and competitive advantages of supra-document journalism products have been replaced by intense article-to-article competition. It has also had some negative consequences for journalism consumers, who have gained access to far greater quantities of text articles but who have simultaneously lost the large-scale organizing and narrative functions that supra-document journalism products provided.

Computational models of narrative may offer an alternative form of supra-document journalism product that could resolve some of the consequences of unbundling for producers and consumers of journalism, and that may be sustainable in the current economic and technological environment. Considerable work has been performed on this, most often focused on extracting structured storylines from vast corpora of text articles using supervised and semi-supervised natural language processing techniques that are trained on small sets of documents

6th Workshop on Computational Models of Narrative (CMN'15).
Editors: Mark A. Finlayson, Ben Miller, Antonio Lieto, and Remi Ronfard; pp. 39–44

OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

carefully annotated using various annotation schemes – an approach that is exemplified by the ongoing EU NewsReader project [12]. These automated story understanding systems must directly confront the complexity of natural language, albeit via machine learning, and remain dependent on sources of high-quality natural language text articles that are under severe and increasing economic threat.

Alternative approaches that provide mechanisms for the direct creation and maintenance of structured narratives as journalistic artifacts have not been widely explored in recent years, perhaps because the structures used by earlier direct-entry narrative modeling systems, such as the scripts of Ableson and Schank [1] and even the sketchy scripts of DeJong [5], have been formal, complex and therefore difficult to apply in a production journalism environment. The more recent availability of new networked knowledge management technologies does not appear to have been applied to new attempts at direct-entry narrative modeling beyond a few examples such the BBC storyline ontology [11] and Facebook's custom stories [9].

Structured Stories is an attempt to build and test a platform for supra-document journalism products using event and narrative data structures. The approach does not attempt a formal representation of events and narratives equivalent to that expressible in natural language, but instead provides a 'computational pidgin' for narrative somewhat similar to that proposed by Margaret Masterman and Martin Kay for machine translation in 1960 [10]. Events within Structured Stories are considered to be discrete things in the world, in the Davidson sense [4], and not linguistic artifacts originating in text. The arrangement of these events into narrative structures seeks to align with human narrative cognition concerning the relative importance of events and the encapsulation of detail within narratives.

The Structured Stories platform was designed and built during late 2013 and 2014, and has been implemented as a cloud-hosted and API-accessible database of event and narrative information. It is currently being populated with structured narratives in the local government domain, and is consumable in five languages.

2 Description of the Platform

The building blocks of Structured Stories are event frames, which are abstractions of discrete journalistic events and are defined as subsets of FrameNet semantic frames [2]. Event frames are light-weight and flexible and are gathered into a searchable library that can grow to many tens of thousands of frames. Each event frame contains a set of type-constrained event roles that are referenced to semantic roles within the parent semantic frame, and a set of natural language phrases that are centered on a verb lexical unit from the semantic frame and that express event-level context. Although rooted in the semantic formalism of FrameNet, these contextual phrases characterize event frames as editorial artifacts, and not as formal structures. As editorial artifacts they are therefore relatively simple and flexible, and are intended to be created, managed and used by journalists for journalistic purposes.

Listing 1 Event frame – simplified structure for a 2-role event frame

```
Event frame ID
FrameNet frame ID
Role1 ( Event Frame Role, FrameNet Role, allowed type )
Role 2 ( Event Frame Role, FrameNet Role, allowed type )
Phrase 1 ( Journalistic Phrase, Verb Lexical Unit  )
```

Discrete journalistic events are represented within the platform as structured events. Each structured event is defined by an event frame, and each of the event roles from the defining event frame is assigned a typed reference to a Uniform Resource Identifier (URI) – typically

an entry in a knowledge graph. These URIs are constrained by type and the platform recognizes seven top-level types: characters, entities, locations, information artifacts, other events, narratives and constants. The knowledge graphs used include Freebase, WikiData and Facebook, and the event type and narrative type are referenced to structured events and structured narratives within the Structured Stories database. Structured events are also associated with various discourse elements, including natural language bullet points, summaries describing the event, images illustrating the event, etc., and events are also linked by cause and effect relationships.

Listing 2 Structured Event – simplified structure

```
Event ID
Event frame ID
Time ( reference time, temporal granularity, temporal duration )
Role references:
   Characters ( event frame roles, knowledge graph IDs )
   Entities/concepts (event frame roles, knowledge graph IDs)
   Locations (event frame roles, knowledge graph IDs)
   Information artifacts (event frame roles, local references)
   Reference Events (event frame roles, event IDs)
   Referenced Stories (event frame roles, story IDs)
   Constants (event frame roles, local references)
Discourse elements (text summary, image, audio, video, etc.)
Causal relationships (causing event IDs, cause types)
```

The platform represents narrative structures as ordered collections of references to structured events, with each reference carrying information about the function of the event within the structured narrative. The relative importance of the event within the structured narrative is represented, and the encapsulation of detail about the event is captured using references to other structured narratives. This fractal-like [6] and recursive structuring enables single structured narratives of many tens of thousands of discrete events to be represented coherently and explored with a few clicks. The narrative structure also enables linkages between structured narratives using common events, common characters, common locations and several other factors, enabling very large-scale narrative networks to be assembled and navigated.

Listing 3 Structured Narrative – simplified structure

```
Story ID
Story events
        ( Event ID, Importance Value, Subnarrative Story ID )
        ( Event ID, Importance Value, Subnarrative Story ID )
        ( Event ID, Importance Value, Subnarrative Story ID )
                                ...
```

These event and narrative structures enable an array of features that facilitate the consumption of journalism. The presentation of narratives can be extensively controlled, enabling the use of different kinds of discourse elements to provide different media experiences of the narrative. The use of structured narratives appears to substantially improve the consumption efficiency of narratives compared with consumption from documents by providing explicit control of detail, access to sub-narratives and navigation of the narrative network. Source documents and related documents are linked from individual structured events and are therefore easily findable within the narrative structure. Text discourse elements can be translated at the event level using machine translation or single-sentence human translation

– a much easier task than the translation of large multi-sentence narratives within text documents. The basis of structured narratives as a permanent and cumulative data store enables the publishing of journalism as a 'pull' (user decides) model rather than as a 'push' (publisher decides) model. Individual events are published as URIs and might therefore be used in mashups or in fact-checking applications, and explicit querying of the event and story database using knowledge graph references, semantic frame references and other structural elements is straightforward. Even reasoning on structured narratives may be possible.

The technical implementation of the prototype platform is centered on a RESTful API, powered by a Node.JS server application. The databases are hosted on Amazon AWS EC2 and S3, and combine Redis, a file system and a graph database. The front-end application is based on the AngularJS application framework.

3 Discussion

Significant uncertainty exists regarding the ease with which untrained users can create and edit structured events and structured narratives within the platform, and also regarding their motivation to do so. Exploring this uncertainty is one of the primary goals for the project and has driven the design of several features within the platform.

The approach seeks to provide sufficient expressive power in its representation of events and narratives to be useful for journalism, but simultaneously seeks to be simple enough to enable easy use by untrained users – typically professional and citizen journalists. This 'goldilocks' goal has been addressed through the light-weight and flexible nature of the event frames, and through a sequential user interface technique that has been shown to enable the entry of individual events by an untrained reporter within 20 seconds.

The approach seeks to deliberately manage the risk of combinatorial explosion in the number of event frames in multiple ways. There is a deep design assumption that the distribution of the use of event frames for journalism will follow a scale-free power law [7], and therefore that the combination of a library of 'head' event frames, a fast method for creating new 'tail' event frames, and a fast search mechanism for finding event frames will enable frame numbers to be manageable. The risks of combinatorial explosion in editorial tasks, such as event frame de-duplication, are higher but are partly reduced by the use of FrameNet as a semantic foundation.

The near-term challenge of motivating participation by reporters during experimentation will be initially addressed by employing a small number of reporters to add structured events and assemble structured narratives in small domains with strong journalistic needs – specifically local government journalism in selected cities. In the medium term motivation will likely depend on the prospect of a sustainable economic rebundling of journalism as structured narrative products and on civic motivation by citizen journalists. In the long term motivating participation by reporters would depend on the efficacy of structured narratives as a mechanism for accumulating journalism and for distributing that journalism via novel products. There are also many additional significant uncertainties regarding the utility of the approach to consumers of journalism, upon which the motivation for participation by producers will ultimately depend.

4 Next Steps

The prototype of the Structured Stories platform is currently being populated with structured events and structured narratives relating to local government news stories in Los Angeles.

The next step for the project will focus on evaluating the feasibility of event and narrative entry and maintenance by untrained reporters, and on defining and evaluating editorial processes to facilitate the management of journalistic quality within structured narratives. This evaluation will occur concurrently with a major reporting project focused on local government in New York City, which will be undertaken during the summer of 2015. If reporting and editing prove feasible then a deep evaluation of the consumption side of the approach, using the captured structured narratives and an iOS app, will be attempted.

Regardless of the results of this testing the Structured Stories project should generate a dataset of hand-curated journalistic news events referenced to FrameNet frames and semantic roles, populated by knowledge graph references and linked to text articles that describe those news events. This dataset may be useful as a training set for supervised machine learning projects. Conversely, there are opportunities to use machine learning techniques such a relation extraction and frame parsing to facilitate capture of structured events into the platform. The Structured Stories approach to modeling narrative structure is therefore an alternative to, and also a complement to, the supervised machine learning approach.

Several extensions to the Structured Stories platform are anticipated, and include the addition of sources of event semantics beyond FrameNet (including VerbNet, PropBank and possibly the NewsReader Events and Situations Ontology), the inclusion of additional discourse elements at the structured event level (including audio, video and comics), and the possible extension of discourse elements to individual roles within the structured events. Improvements to the event reporting workflow, possibly including semi-automation of the workflow using the EVITA system [8] and various TF-IDF document clustering techniques such as the Associated Press Overview system will be explored following the assessment of reporting and editing using the prototype platform.

The Structured Stories prototype is publicly available at `http://www.structuredstories.org`.

References

1. Robert Abelson and Roger Schank. *Scripts, Plans, Goals and Understanding.* Psychology Press, USA, 1977.
2. Collin Baker. Framenet, present and future. In *First International Conference on Global Interoperability for Language Resources*, 2008.
3. Allan Bell. *The Language of News Media.* Wiley-Blackwell, New Jersey, USA, 1991.
4. Donald Davidson. The Logical Form of Action Sentences. In Nicholas Rescher, editor, *The Logic of Decision and Action.* University of Pittsburg Press, 1967.
5. Gerald DeJong. Prediction and Substantiation: A New Approach to Natural Language Processing. *Cognitive Science*, 3:251–273, 1979.
6. Charlie Hargood et al. Exploring the Poetics of Strange and Fractal Hypertexts. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media.* 23rd ACM Conference on Hypertext and Social Media, 2012.
7. Lev Muchnik et al. Origins of power law degree distribution in the heterogeneity of human activity in social networks. *Scientific Reports*, 3, 2013.
8. Roser Saurí et al. EVITA: A Robust Event Recognizer for QA Systems. In *Proceedings of HLT/EMNLP 2005*, pages 700–707, 2005.
9. Facebook Inc. Open Graph Stories (`https://developers.facebook.com/docs/opengraph`). *Facebook Developer Website*, 2015.
10. Margaret Masterman and Martin Kay. Mechanical Pidgin Translation (`http://mt-archive.info/CLRU-1960-Masterman.pdf`). *Cambridge Language Research Unit*, 1960.

11 Robin Pembrook. Storyline Data Model: Sharing the Ontology for BBC News (`http://www.bbc.co.uk/blogs/internet/entries/98d9d3f2-6932-371b-931c-7a13fbf1bacf`). *BBC Internet Blog*, 2013.

12 P. Vossen, G. Rigau, L. Serafini, P. Stouten, F. Irving, and W. R. V. Hage. Newsreader: recording history from daily news streams. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC'14)*, Reykjavik, Iceland, 2014.