

# Modeling Multiple-Event Situations across News Articles

Earl J. Wagner

Northwestern University  
Evanston IL USA

ewagner@u.northwestern.edu

Larry Birnbaum

Northwestern University  
Evanston IL USA

birnbaum@cs.northwestern.edu

Kenneth D. Forbus

Northwestern University  
Evanston IL USA

forbus@northwestern.edu

## ABSTRACT

Readers interested in the context of an event covered in the news such as the dismissal of a lawsuit can benefit from easily finding out about the overall news situation, the legal trial, of which the event is a part. Guided by abstract models of news situation types such as legal trials, corporate acquisitions, and kidnappings, *Brussell* is a system that presents situation instances it creates by reading multiple articles about the specific events that comprise them. We discuss how these situation models are structured and how they drive the creation of particular instances.

## Categories and Subject Descriptors

I.2.4 Knowledge Representation Formalisms and Methods  
– frames and scripts.

## General Terms

Algorithms, Design.

## INTRODUCTION

News articles typically discuss a new event in detail and mention relevant information about the people and organizations involved in the event. Still, in reading about an event that just occurred, a reader may want to know about the earlier events that led up to it. In focusing on details that are new or have changed, however, articles often leave out contextual information of this sort.

Consider an individual reading a news article about a lawsuit filed by Oracle against SAP that notes: “The 16-month-old lawsuit focuses on TomorrowNow, a software maintenance specialist that SAP bought in 2005 to counter Oracle’s \$11.1 billion acquisition of PeopleSoft”. Suppose the reader were not familiar with this acquisition. He would likely be interested in finding out more about it, including how it started and how it ended. More generally, he could benefit from a presentation that builds upon his *expectations* for how an acquisition proceeds.

What do we mean by having expectations about the acquisition? In talking about a specific acquisition news situation, we refer to the events that make up “*the acquisition*” of the individual. Particular news situations like the acquisition of PeopleSoft by Oracle, or the acquisition of TomorrowNow by SAP, involve events that are linked through stereotypical, causal relationships: first is a *negotiation* that may or may not be covered in the news followed by a *public offer*. The prospective *acquiree* may *reject* the offer and the *acquirer* may *increase* the offer or *withdraw* it. If the acquisition proceeds and the acquiree chooses to *accept* the

offer then the acquisition concludes with a *close* event and possibly subsequent *layoffs*. These abstract events likely to occur in an acquisition make up the acquisition *situation type*, the limited sequence of causally-related events constituting it that are covered in multiple news articles.

To address the need identified above, we have developed *Brussell*, a system that extracts and organizes information about news situations across multiple articles as *situation instances* and presents timeline-like “storyline” views for these situation instances in the context of any web page mentioning the situation’s events. Consider again the case of a reader seeking to learn more about the acquisition situation, and suppose the person were using Brussell which works in conjunction with the web browser software. Rather than interacting with the article at the textual level by selecting keywords to search with, he can interact directly with a representation of the situation through a phrase that mentions it, e.g., “Oracle’s \$11.1 billion acquisition of PeopleSoft.”

## EXTRACTING NEWS SITUATIONS

To support this interaction, Brussell consists of a Firefox browser plugin and server software, which may both run on the same computer. The back-end component requires manually created situation model types (inspired by scripts [4]) and currently supports kidnappings, legal trials and corporate acquisitions, each of which has multiple possible outcomes and on the order of 8-12 possible events. The system runs daily to retrieve news articles from several English-language news websites via RSS feeds. After retrieving new articles, it selects the new articles with keywords associated with the situation types it supports and reads through the articles to create and extend situation instances. These instances may span a single mention up to several hundred if they are well publicized.

## STRUCTURE OF SITUATION MODELS

Situation models guide the creation of situation instances in several ways. They specify:

- What information to extract and aggregate within situation instances
- How to select documents and sentences to analyze
- How to extract information from sentences
- How to select among conflicting information that is extracted
- How to distinguish between different situation instances of the same type.

Situation models specify the participant roles and their types. Recall the acquisition of PeopleSoft by Oracle. The

corresponding situation instance features Oracle as the *acquirer* participant and PeopleSoft as the *acquiree* participant. The acquisition situation model includes these as well as other participants such as a *seller*. It specifies that the acquiree is an organization while the acquirer and seller may be organizations or persons.

As we saw in the example, situation models represent the kinds of events that may occur as the situation proceeds such as an offer and the acceptance or withdrawal of the offer. Each situation event type specifies text patterns for recognizing mentions of the event in establishing that the event occurred such as “ACQUIRER offered to (acquire OR purchase OR buy) ACQUIREE” for offer event or “ACQUIRER withdraws offer to (acquire OR purchase OR buy) ACQUIREE” for the withdraw event. Brussell uses keywords specific to situation models to select documents to read, and reads only sentences with terms appearing in event patterns.

In recognizing an event mention, the system must link it to other, related events. If there is a mention of an acquisition closing, to which acquisition should this event be assigned? Brussell distinguishes situation instances both by their type and the identities of *focal* participants. The focal participant in an acquisition is the acquiree and any acquisition-related events found involving the same acquiree are aggregated within the same acquisition instance. This limits extraction to event mentions explicitly specifying participants in a such a way that their roles can be inferred.

After extracting situation events from news, the system finalizes the sequence of events in the situation instance by applying two kinds of semantic constraints specified in the models. First, to reflect the ordering of actual events, ordering constraints are imposed on events requiring that, for example, an *offer* is made before the offer is *accepted* or *withdrawn*. Second, constraints also hold preventing mutually-exclusive events from both being recognized. The occurrence of a *withdraw* event will prevent a *close* event from occurring.

By abstracting knowledge specific to situations types, Brussell can be extended with additional models including the kidnapping situation model in Table 1.

## BACKGROUND

No other research systems extract information about stereotyped sequences of causally-related events from multiple articles. Related work in single-document extraction includes the original work in using scripts to model human memory and for story-understanding, and their application in later systems such as FRUMP, for reading articles directly from newswires [2]. Brussell updates this approach by filling in situation templates of multiple events by reading and integrating information from many articles about situation events rather than just one. It does so by extracting “focal” participant event arguments specified in situation models to unify events and link events known to be causally related [5]. More recent work has described unsupervised learning of narrative event chains, similar to Brussell's situation models [1].

Situation Models	
Corporate Acquisition	Kidnapping
<b>Document Keywords</b>	<b>Document Keywords</b>
acquis*, acquir*	kidnap*, abduct*, hostage*, captur*
<b>Participants and Types</b>	<b>Participants and Types</b>
ACQUIRER: Organization or Person	KIDNAPPER: Organization or Group of People or Person
ACQUIREE: Organization ( <i>focal</i> )	KIDNAP-VICTIM: Group of People or Person ( <i>focal</i> )
SELLER: Organization or Person	RESCUER: Organization or Group of People or Person
<b>Events</b>	<b>Events</b>
NEGOTIATIONS	ABDUCTION
OFFER	VIDEOTAPE
ANNOUNCE	THREAT
INCREASE-OFFER	DEMAND-RANSOM
DECREASE-OFFER	NEGOTIATION
WITHDRAW	APPEAL-FOR-RELEASE
REJECT	PAY-RANSOM
CLOSE	RELEASE
LAYOFFS	RESCUE
	ESCAPE
	KILLING

Table 1: Situation Models for Acquisition and Kidnapping

Recent work in information extraction has focused on aggregating information extracted about entities and events across multiple documents. The Automatic Content Extraction Evaluation (ACE08) tested cross-document extraction of entities and relations among entities such as "employer [of]" and "inventor [of]" [3]. Brussell extends this work not only by extracting entire instances of situations spanning multiple events, but also by using these instances to perform cross-document inference propagating extracted event arguments such as dates, locations and participants.

## REFERENCES

- [1] Chambers, N. and Jurafsky, D. “Unsupervised Learning of Narrative Event Chains.” In *Proceedings of ACL/HLT 2008*, (2008).
- [2] DeJong, G. F. 1982. “An Overview of the FRUMP System.” In Lehnert, W. G. and Ringle, M. H. (Eds.), *Strategies for Natural Language Processing*. Hillsdale, Erlbaum, pp. 149--176.
- [3] NIST. “Automatic Content Extraction 2008 Evaluation Plan (ACE08)” <http://www.itl.nist.gov/iad/mig/tests/ace/2008/doc/ace08-evalplan.v1.2d.pdf>
- [4] Schank, R. C. and Abelson, R. P. *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, (1977)
- [5] Wagner, E. J., Liu, J., Birnbaum, L., Forbus, K. D. “Rich Interfaces for Reading News on the Web.” In *Proceedings of the 2009 International Conference on Intelligent User Interfaces (IUI 2009)* (2009).