

## 1. Computable News

News is all about **events**, with their *who*, *what*, *when* and *where* at the centre of each story. New events build upon old events.

The *Computable News* project seeks to identify the relationships between events in the news, and to provide readers with ways to access more stories of interest.

## 2. Event extraction

- *Information extraction* is about pulling structured data out of natural language
- Events are complex linguistically:
  - Synonymy: his passing, Bin Laden died, Operation Neptune Spear, fatal gunshot, kicked the bucket, the deceased's estate
  - Polysemy: could be political death
- Events are complex ontologically:
  - Sub-events: Rugby World Cup 2011, the grand final, the winning try
  - Co-events: win vs lose; disappear vs kidnap; poison vs murder vs die
  - Diversity: rose 5c, GFC, begin, plan, die
  - Factuality: certain, probable, possible
- This complexity makes extracting characteristics of generic events difficult
- We instead define events by how they are reported in news

## 3. Introducing: Event Linking

News providers like Fairfax have begun to link background information to archival articles (see figure 1). With them, users can read further without leaving the site.

Yet under 2% of smh.com.au articles have inline links to past articles. Coverage is low, inconsistent, and requires arduous manual effort to determine appropriate links.

**Automated event linking** intends to produce these links more systematically, with high coverage and low manual effort.

### An event linker:

1. finds references to past events in a given document  $D$ ; and
2. links each to its first report in a given news archive  $A$ , where possible.

This is like Named Entity Linking, but treats the news archive as a "Wikipedia" of events.

## 4. Annotating a news corpus

We are currently producing a collection of documents with manual event link annotations (overlaid in figure 1).

We ask annotators to:

1. identify mentions of past/ongoing events that are newsworthy
2. classify them into:
  - Basic** e.g. won the race
  - Complex** e.g. 2000 Summer Olympics
  - Trend/Change** e.g. grew 5%
  - Multiple** e.g. many races
3. link each basic event to the SMH archive, or mark it *Reported here* or *Not found*
4. link each complex event to Wikipedia or mark it *Not found*

**Figure 1:** Our annotations on an smh.com.au article with manual event links

## Sydney man carjacked<sup>1</sup> at knifepoint

Glenda Kwek  
September 21, 2011

There has been another carjacking<sup>1</sup> in Sydney, two weeks after two people were stabbed<sup>2</sup> in their cars in separate incidents.

A 32-year-old driver was walking to his station wagon on Hickson Road, Millers Point, after feeding his parking meter about 4.30pm yesterday when a man armed with a knife grabbed<sup>1</sup> him and told him to hand over his car keys and mobile phone, police said.

...

Police warned<sup>3</sup> Sydney drivers to keep their car doors locked after two stabbings<sup>2</sup> this month.

On September 4, a 40-year-old man was stabbed<sup>4</sup> when three men tried to steal his car on Rawson Street, Auburn, about 1.20am.

The next day, a 25-year-old woman was stabbed<sup>5</sup> in her lower back as she got into her car on Liverpool Road, Ashfield, about 8pm.

Police appealed<sup>1</sup> for anyone with information to phone Crime Stoppers on 1800 333 000.

**Our annotation:**

- New events
- Background events

**1:** Reported here

**2:** Reported across multiple articles

**3:** Reported in *Sydney drivers told: lock your doors*

**4:** Reported in *Driver stabbed after Sydney carjacking*

**5:** Reported in *Car attack: driver stabbed in the back*

**Figure 2:** A small event link graph derived from smh.com.au articles



## 5. Automated linking approaches

A precise event linker would match *who*, *what*, *when*, *where* between event references in a source and target document.

We propose two broad linking approaches:

### System 1 (c.f. entity linking workflow)

1. find each past event mention in  $D$ ; for it:
2. search  $A$  for candidate articles to link to
3. rank the candidates and select one/none

### System 2

1. collect all articles in  $A$  relevant to  $D$
2. identify new information in each article
3. match these to event mentions in  $D$

## 6. Discoveries so far

- Sports articles report events very differently to general news: most SMH sports articles are commentary, assuming readers already know the match outcome.
- When finding articles to link, filtering by date range is extremely helpful. Often, however, date information is not explicit in the article, and needs to be acquired from external knowledge sources like the web.

## 7. The bottom line

- Event linking is an approach to extracting structured event representations from text, centred on how they are reported in a news archive.
- An automated approach could add links for readers to trace the background to a story, as the news told it.
- This substitutes for the current low-coverage, manual approach to linking background stories in some news sites.
- Automated event linking can transform a news archive into a graph of event references (see figure 2).

## 8. Industry

The *Computable News* project is a cooperative project between The Capital Markets CRC, Fairfax Media and The University of Sydney. The team includes one senior lecturer, three Postdoctoral researchers and three PhD students.

The diverse combination of partners ensures that the project maintains clear research and commercial goals.