

# Computable News: Milestone 1

## Automated Entity Tag Clouds

James Curran (and the Computable News team)

æ-lab

School of Information Technologies  
University of Sydney

23rd March, 2011



## ✓ Computable News team has been hired

- Postdoctoral fellows:
  - **Ben Hachey**
  - **Matthew Honnibal**
  - **David Vadas**
- Project manager:
  - **Tim Yates**
- PhD students:
  - **Joel Nothman**
  - **Will Radford**
  - **Tim O'Keefe**
- additional annotators from  $\alpha$ -lab and outside

# M1 Theme: Data collection, annotation and evaluation

- Demo: interactive tag clouds of entities using existing technology
- Collection and analysis of Fairfax Digital raw text
- Annotation schema design and testing
- Annotation tool development
- evaluation of existing schema/systems/technology  
e.g. OpenCalais, Extractiv, and Evri
- technology selection, e.g. choice of DBMS

# M1 Deliverables (high level)

- ✓✓ Demo: interactive entity tag clouds from SMH data
  - ✓ Schemas that target what we want to get out of news articles
- ✓✓ 1000 SMH articles annotated with entities
- ✓✓ 500 SMH articles annotated with sentiment
  - ✗ 500 SMH articles annotated with events
- ✓✓ A powerful yet simple-to-use annotation tool
  - ✓ infrastructure that supports these tools and future development

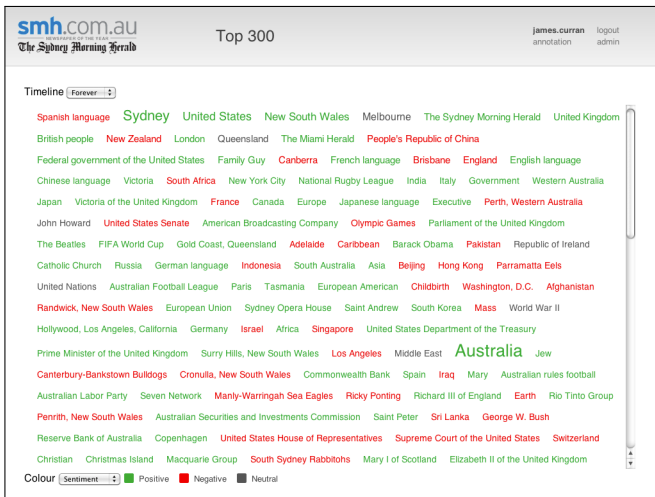
# M1 Demo

- ✓ Interactive entity tag clouds from Sydney Morning Herald data
  - Web interface showing entities and how often they occur
  - Entities link to the articles they appear in and other related entities
  - Visualisation of entity types and sentiment
- A thousand Sydney Morning Herald articles annotated with entities and sentiment
  - A powerful yet simple-to-use annotation tool
  - Schemas that target what we want to get out of news articles
- Back-end infrastructure that supports these tools and future development for the project

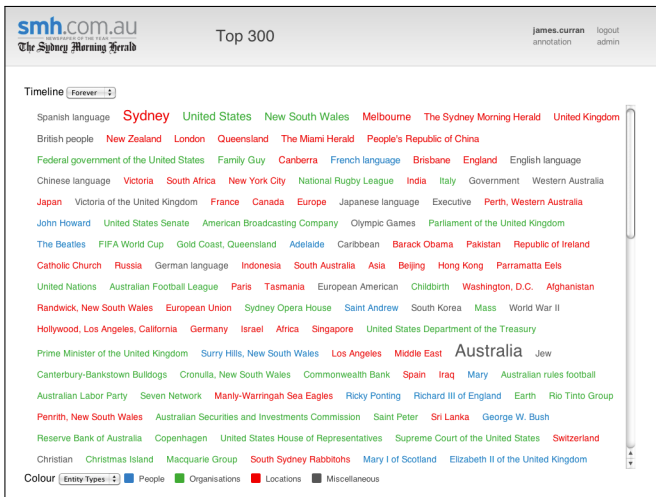
# Top 300 entity cloud



# Top 300 entity cloud with sentiment

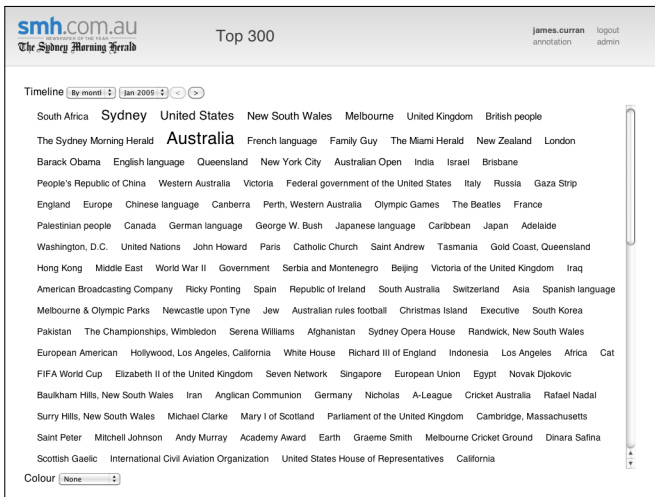


# Top 300 entity cloud with entity types





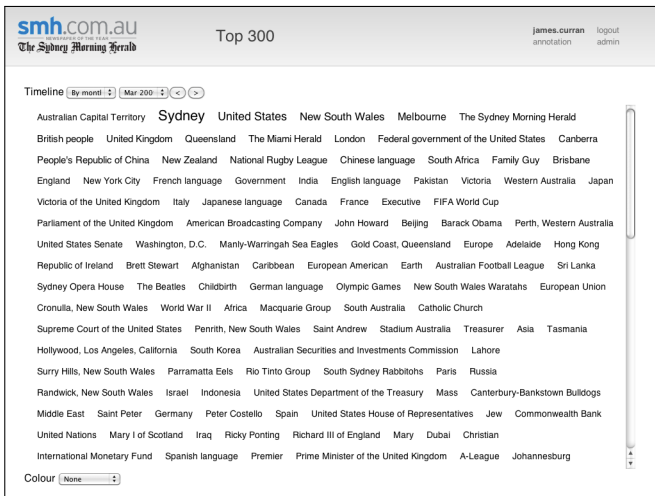
# Top 300 entity cloud by month



# Top 300 entity cloud by month




# Top 300 entity cloud by month




# Top 300 entity cloud by month



# Paul Keating entity cloud (together/strange/mutual)



MEMBERSHIP OF THE YEAR  
The Sydney Morning Herald



Paul Keating

james.curran  
annotation

logout  
admin

Timeline Forever

Related Entities Trends

Childbirth Peter Costello John Howard Barangaroo, New South Wales  
Gough Whitlam Liberal Party of Australia **Bob Hawke**  
Australian Labor Party Prime Minister of the United Kingdom Julia Gillard  
Malcolm Fraser Parliament of the United Kingdom Alexander Downer Brendan Nelson  
Treasurer Tony Abbott Israeli Labor Party Robert Menzies Kim Beazley John Hewson  
Labour economics Goat Island Bob Carr Rudd, Iowa Nathan Rees Government  
United States Department of the Treasury Liberal Party of Canada PM Parliament House, Canberra  
Julie Bishop Great Depression Member of Parliament Beijing Richard Rogers Canberra  
Bill Clinton Chris Johnson Lou Costello American Broadcasting Company United States Senate  
Glasgow Paul Kelly Christian Washington, D.C. Kerry Packer Reserve Bank of Australia  
International Monetary Fund The Daily Telegraph Executive


Colour None Measurement Together

Stories


- Give your pay packet a shave and help save jobs  
Friday, January 02, 2009
- BRAIN EXPLOSION  
Friday, January 02, 2009
- When a little knowledge is a cause for super concern  
Saturday, January 03, 2009
- Our global challenge: make civilisation an agent of peace, not conflict  
Monday, January 05, 2009
- Peacock's argument holds for the Kevin 747 express  
Tuesday, January 06, 2009

[See more](#)

# Paul Keating entity cloud (together/strange/mutual)



MEMBERSHIP OF THE YEAR  
The Sydney Morning Herald



Paul Keating

james.curran  
annotation

logout  
admin

Timeline Forever

Related Entities Trends

Treasurer John Howard Peter Costello Barangaroo, New South Wales

Gough Whitlam Richard Rogers Goat Island Childbirth Australian Labor Party

Liberal Party of Australia Malcolm Fraser Alexander Downer Chris Johnson

Prime Minister of the United Kingdom Lou Costello John Hewson Julia Gillard

Parliament of the United Kingdom Kim Beazley Brendan Nelson Robert Menzies

Institute, West Virginia **Bob Hawke** Israeli Labor Party John Grant Tony Abbott

Division of Higgins Australia Ben Chifley Government Paul Kelly Gareth Evans

Institute, Wisconsin Labour economics PM Bob Carr Rudd, Iowa Hansard Canberra

Julie Bishop John Curtin Kerry Packer Nathan Rees United States Department of the Treasury

Liberal Democrats Liberal Party of Canada Parliament House, Canberra Sydney

Member of Parliament Asia-Pacific Economic Cooperation


Colour None Measurement Strange

Stories


- Give your pay packet a shave and help save jobs  
Friday, January 02, 2009
- BRAIN EXPLOSION  
Friday, January 02, 2009
- When a little knowledge is a cause for super concern  
Saturday, January 03, 2009
- Our global challenge: make civilisation an agent of peace, not conflict  
Monday, January 05, 2009
- Peacock's argument holds for the Kevin 747 express  
Tuesday, January 06, 2009

See more

# Paul Keating entity cloud (together/strange/mutual)



MEMBERSHIP OF THE YEAR  
The Sydney Morning Herald



Paul Keating

james.curran  
annotation

logout  
admin

Timeline Forever

Related Entities Trends

**Bob Hawke** John Howard Peter Costello Australia

Barangaroo, New South Wales Childbirth Sydney Gough Whitlam Australian Labor Party

Liberal Party of Australia Prime Minister of the United Kingdom Parliament of the United Kingdom

Julia Gillard United States Malcolm Fraser Government Canberra Alexander Downer

Treasurer Goat Island Richard Rogers Tony Abbott Brendan Nelson The Sydney Morning Herald

London Israeli Labor Party John Hewson Kim Beazley Robert Menzies

United States Department of the Treasury Chris Johnson United Kingdom

American Broadcasting Company United States Senate Beijing Lou Costello

People's Republic of China Nathan Rees New South Wales British people Labour economics

Liberal Party of Canada Executive Bob Carr Rudd, Iowa PM Washington, D.C.

Federal government of the United States John Grant The Miami Herald

Colour None Measurement Mutual

**Stories**

- Give your pay packet a shave and help save jobs  
Friday, January 02, 2009
- BRAIN EXPLOSION  
Friday, January 02, 2009
- When a little knowledge is a cause for super concern  
Saturday, January 03, 2009
- Our global challenge: make civilisation an agent of peace, not conflict  
Monday, January 05, 2009
- Peacock's argument holds for the Kevin 747 express  
Tuesday, January 06, 2009

[See more](#)

# Entity relatedness == co-occurrence statistics

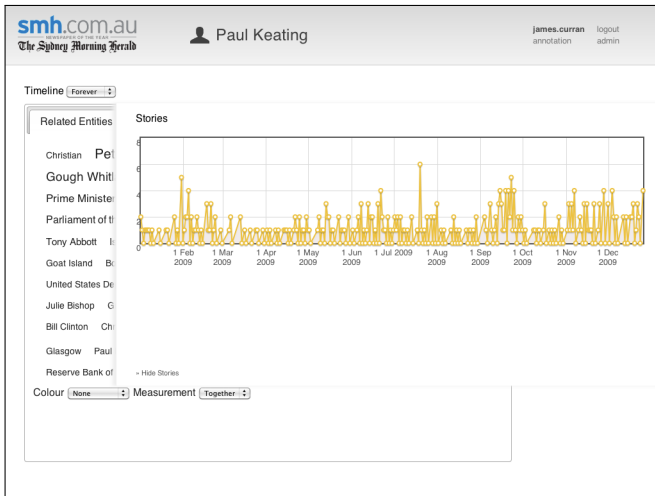
- Frequent – joint frequency co-occurrence count
  - $C(i, j)$
- Strange –  $\chi^2$ 
  - $$\frac{N(C(i, j)C(\bar{i}, \bar{j}) - C(\bar{i}, j)C(i, \bar{j}))^2}{C(i)C(j)(C(\bar{i}, j) + C(\bar{i}, \bar{j}))(C(i, \bar{j}) + C(\bar{i}, \bar{j}))}$$
- Together – joint normalised by individual frequencies
  - $$\frac{C(i, j)}{C(i) + C(j)}$$
- Mutual – Mutual Information
  - $$\frac{C(i, j)^3}{C(i)C(j)}$$



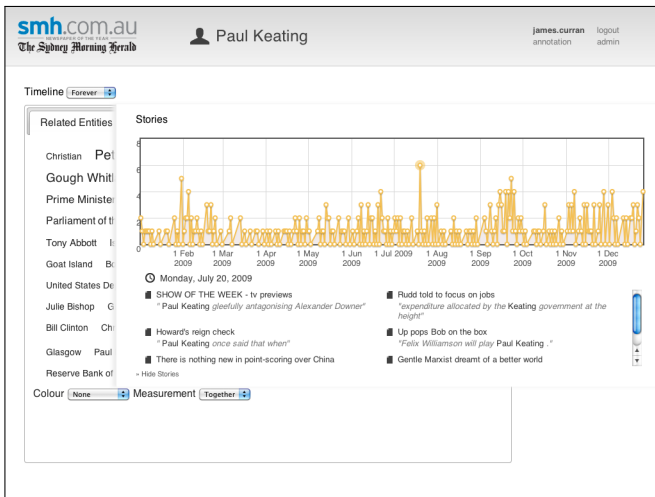
# M1 Demo: Feature checklist

- ✓ Tag clouds show associations between related entities
- ✓ Selectable time period, on month or year
  - Colours/opacity are controllable by:
    - ✓ entity type – person, location, organisation or miscellaneous
    - ✓ sentiment – positive, negative, neutral
    - ✗ recentness of mentions (better as visualisation)
  - Size/weight is controllable by:
    - ✓ overall importance of the entity
    - ✓ unusualness of the entity or relationship
- ✓ creating tag clouds for entities related to a particular entity


# M1 Bonus: Paul Keating story statistics



# M1 Bonus: Paul Keating story statistics



# M1 Bonus: Paul Keating annotated story


SMH090103NA68UD1JQST
james.curran  
annotation
logout  
admin

## When a little knowledge is a cause for super concern

*Business » None*  
Sat 03 Jan 2009

ALEXANDER POPE observed in An Essay On Criticism that "a little knowledge is a dangerous thing".

He was writing in 1711 but the adage aptly describes a conundrum faced by **Australian** superannuation funds today.

For all the headlines, our super funds, particularly the not-for-profit industry funds, have not been hit nearly as hard by the global financial crisis as other financial institutions.

The reason is liquidity.

Many of last year's failed firms have been victims of a liquidity crisis ; a good old-fashioned run on the bank, in which more people have tried to leave than there was cash set aside to meet withdrawals.

It is difficult to get high returns when a large chunk of your funds must sit idly in cash to pay out people who suddenly decide to leave.

Most of the best performers of the past few years, such as hedge funds, have avoided this problem by imposing lock-up periods on investors and freezing redemptions when too many tried to head for the exits.

Historically, industry super funds have never had this problem.

Since **Paul Keating** introduced the super guarantee in 1992, a portion of **Australia**'s wages have flowed faithfully into these funds every month, and because the accounts cannot be accessed until the age of 65, there has not been same fear of a run on the funds.

This has allowed industry super funds to fill their portfolios with unlisted, illiquid investments, such as infrastructure, property, private equity and hedge funds.

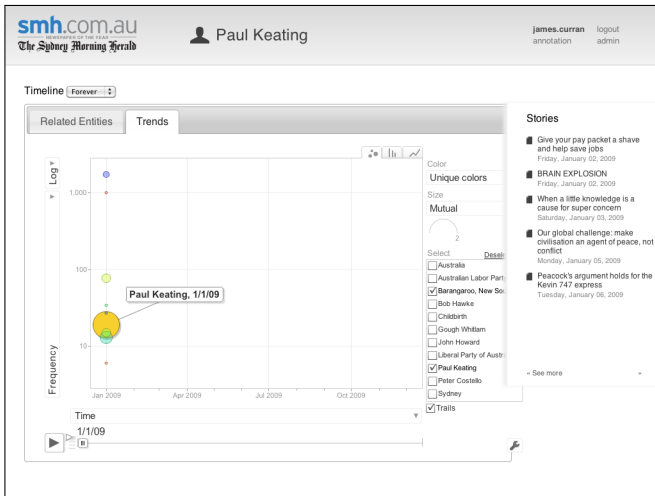
Not only do these assets tend to outperform shares and bonds but their returns are less closely linked with the vagaries of the sharemarket.

The best performing funds over the past five years, MTAA, Cbus, Hostplus, AustralianSuper and Westscheme, have up to half of their money in these investments.

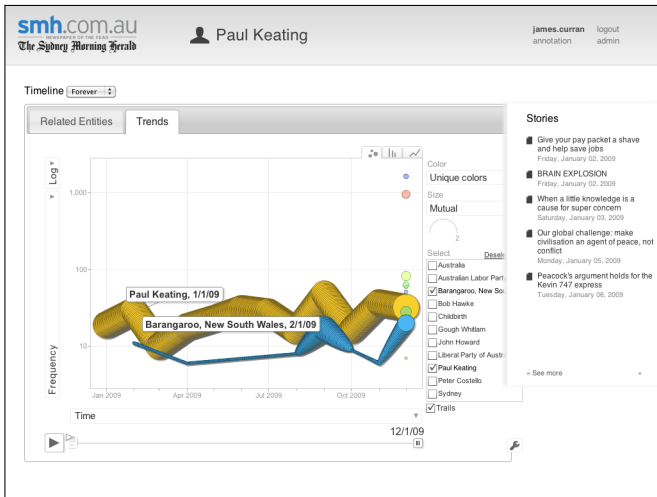
Those who are forced to be more liquid will argue that these assets have done better lately simply because they are not valued as often, and that when they eventually are, there will be massive write-downs for industry funds.

There may be some truth to this but industry funds have also outperformed in bull markets, where, by such logic, unlisted assets would have been priced lower than the volatile

# M1 Bonus: Paul Keating trending topics



# M1 Bonus: Paul Keating trending topics



## ✓ M1 Demo: Bonus features

- ✓ linking to/from articles mentioning entities (was in deep linking)
- ✓✓ mention frequency and daily stories sidebar
- ✓✓ Gap Minder entity trend visualisation

# Technologies

- CSS, HTML, Javascript (jQuery)
- Database: PostgreSQL – integrates easily with Django
  - explored using Tokyo Tyrant to store article text
- Django – Python framework for HTML templating and ORM
- MongoDB + Tokyo Tyrant for Wikipedia data storage
- TinyMCE used for rich text editor in annotation tool



# News data feeds and gold standard

- ✓ Back-end system set up to store Fairfax Digital data
- ✓ Feed set up and receiving breaking news articles
- ✗ Need to expand to all Fairfax Digital feeds
  
- ✓ stories for gold standard randomly chosen from:
  - the full year 2009
  - between 100 and 1500 words
  - excluding articles from Births, Deaths, Marriages and Tenders

# What is a named entity?

Failed entrepreneur and former CEO Eddy Groves will defend a criminal charge relating to the collapse of Australia's biggest childcare chain, ABC.

“I've pleaded not guilty today and I will vigorously defend the charge, and that's really all I can say right now,” Groves said.

Based on <http://www.news.com.au/business/breaking-news/abc-learning-founder-eddy-groves-pleads-not-guilty/story-e6frfkur-1225996040971>

# What is a named entity?

Failed entrepreneur and former CEO [Eddy Groves](#) will defend a criminal charge relating to the collapse of [Australia](#)'s biggest childcare chain, [ABC](#).

‘‘I’ve pleaded not guilty today and I will vigorously defend the charge, and that’s really all I can say right now,’’ [Groves](#) said.

- [Eddy Groves](#) → wiki/Eddy\_Groves
- [ABC](#) → wiki/ABC\_Learning
- [ABC](#) → wiki/Australian\_Broadcasting\_Corporation

# Entity recognition comparison

- A number of commercial systems perform entity recognition: OpenCalais, Extractiv, Evri (see wiki)
- some systems also perform entity linking
  - they are all closed systems (model/features cannot be changed)
- also academic systems for performing these tasks: C&C (Sydney/Cambridge), Stanford NE recogniser, LT tools (Edinburgh)
  - trained on very different text to real-world SMH text

## ✓ M1 Entities: deliverables

- ✓ entity annotation scheme defined
- ✓✓ entity annotation tool implemented
- ✓✓ 1,000 historical SMH articles annotated with entities
  - ✓ Comparison of existing entity recognition packages
  - ✓ Engine for highlighting entities and linking

# Entity type scheme (1)

- **Individual** – a person, animal, deity, or fictional character
  - Nathan Rees; Angus Houston; Phar Lap; Lady Gaga; Miles Davis
- **Location** – a geographical, geopolitical, or celestial feature
  - Uluru; Australia; Mars
- **Organisation** – a company, sports league or team, band, theatre company, etc
  - Sportingbet; Wallabies; Pink Floyd; Sydney Theatre Company
- **Product** – a toy, game, software, drug, weapon
  - Prozac, AK-47, Monopoly, Microsoft Windows, Nikon D300s
- **Work of Art** – a book, play, movie, song, album, etc
  - Miracles, Harry Potter and the Prisoner of Azkaban

## Entity type scheme (2)

- **Facility** – a building, bridge, highway, airport, etc
  - Santa Cruz cemetery; Randwick Racecourse; Highway 61
- **Artefact** – a law, legal case, award, degree, currency, market index, vessel, tv/radio station, newspaper, magazine, etc
  - Corporations Act; ARIA Album of the Year; Australian Dollar; ASX 200; FBi Radio
- **Event** – a sporting event, holiday/observance, convention, war, award event, natural disaster, programme/campaign, etc
  - 2010-2011 Ashes; Earth Hour; World War II; Oscar Awards; Hurricane Katrina; Building the Education Revolution
- **Miscellaneous** – the O and R in NORP and languages
  - Catholic, Sunni, Italian

## M1 Entities: Bonus

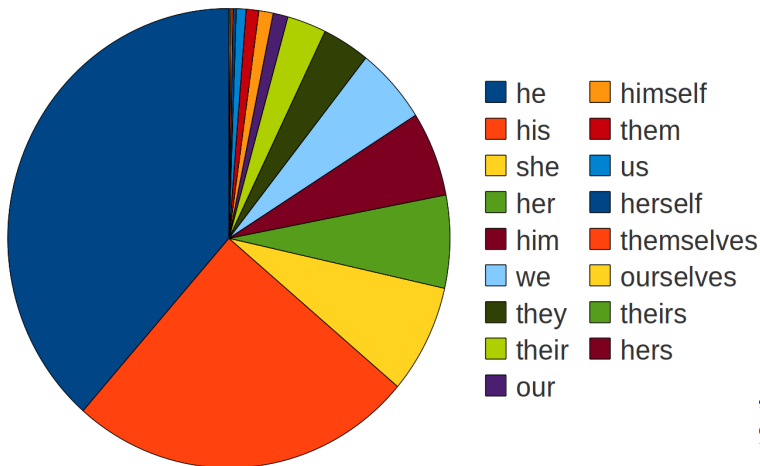
- ✓ NORPs mapped to the underlying nation, e.g. American → USA
- ✓✓ All entities linked against our knowledge base
- ✓✓ Entities are corefenced within an article
  - ✓ Including pronoun coreference



# Annotation analysis

- Over 500,000 words (over 320 man hours of work)
- 9,991 unique entities, about 40% of them non-Wikipedia
- 44,279 mentions, about 27% of them non-Wikipedia
- Top 5 Entities in Wikipedia:
  - Australia
  - United States
  - Kevin Rudd
  - Sydney
  - China
- Top 5 non-Wikipedia Entities - mostly obituaries:
  - Paddy Ward
  - Val Street
  - Max Roberts
  - Richard Aldridge
  - Isoken Allen Orobor

# Bonus round: 9,854 pronouns annotated



# What is an event?

Failed entrepreneur and former CEO [Eddy Groves](#) will defend a criminal charge relating to the collapse of [Australia](#)'s biggest childcare chain, [ABC](#).

“I’ve pleaded not guilty today and I will vigorously defend the charge, and that’s really all I can say right now,” [Groves](#) said.

- [Eddy Groves](#) *is involved in a* JUSTICE-EVENT
- [Eddy Groves](#) *is a CEO of* [ABC](#)

# Existing event systems

- Aim to identify predetermined types of mentioned events, extracting who, when, where, etc.
- ACE 2005 corpus
  - Marks event “trigger” terms, participants, coreference
  - 34 types focussing on war, politics, business
  - Low annotator agreement (ambiguous); biased topic
  - Extracting all its information requires “semantic completeness”
- TimeML
  - Generic: most verbs and many nouns are marked as events
  - A more theoretical approach, interested in temporal relations
- OpenCalais (and others from late 1990s, early 2000s)
  - A small list of pre-configured event types
  - What, who, when identified with trigger words and patterns

# Event detection is hard to define and do

- Previous efforts have not achieved high annotation consistency or system accuracy, because:
  - ① **Events don't have clear boundaries.** If a company lays off 100 workers, is that one EMPLOYMENT event, or 100? Is a house burning down in a bushfire part of the same event?
  - ② **Events can be mentioned indirectly.** An article might quote a spokesperson saying We have increased profits by reducing our salary expenditures.
  - ③ **It's unclear what counts as an event.** Are price movements events? Does an event need a definite beginning and end? What about long-term events, like wars?

# Problems

- Level of granularity
- Diverse language and implicit events
- Generic vs specific
- Tense, aspect, likelihood and rhetorical forms
- Incompleteness of the scheme, i.e. we expect there to be events that we won't annotate
- Low inter-annotator agreement

# Event types for sentence level scheme

Category	Examples
Lifecycle	Births, deaths, injuries, marriages
Organisation lifecycle	Mergers, major expansions, bankruptcy
Employment	Hiring, firing, contract extension
Conflict	Attacks, law suits, disputes
Justice	Investigation/arrest/trial/sentencing
Governance	Elections, laws, parliamentary proceedings
Sports match	horse races, league matches
Finance	Price movements, share issues, restructuring
Real estate	Property developments, construction, zoning
New release	Product launch, reports
Transaction	Buying/selling, bids, payments

# M1 Events: Deliverables

- ✓ Event annotation tool
- ✓ Event annotation scheme
- ✗ 500 SMH articles annotated with events (220 completed)
  - We annotated:
    - 13 broad event types at sentence level (with word-level anchors)
    - coreference and sub-event relations within the document
    - whether the event was “realised”



## A new plans for events: document-level

- Key event(s) of the document
  - Excluding back-story which introduces ambiguities
  - What event caused this article to be written?
  - Not just any activity: only newsworthy events
- Fine-grained, hierarchical event schema
  - Models specific and general event types
- More annotation detail: did/will it possibly/definitely happen?
- We will extract who/where/when from the document
- Better targets what users will want

# What is sentiment?

**Failed** entrepreneur and former CEO Eddy Groves will defend a **criminal charge** relating to the **collapse** of Australia's biggest childcare chain, ABC.

“I’ve pleaded **not guilty** today and I will vigorously defend the charge, and that’s really all I can say right now,” Groves said.

- **Failed** *applies negative sentiment towards* Eddy Groves
- **criminal charge** *applies negative sentiment towards* Eddy Groves
- **collapse** *applies negative sentiment towards* ABC
- **not guilty** *applies positive sentiment towards* Eddy Groves

## Existing sentiment systems

- Some text analytics packages are able to detect entity level sentiment such as:
  - Lexalytics (commercial) - Though sentiment analysis is a headline feature, their methods are very simple
  - OpenAmplify (commercial) - Designed for social media analysis, rather than news analysis
  - Appinions (commercial) - Only does entity level sentiment analysis about pre-defined topics
  - TextMap (academic) - Licensable research software, will only show aggregate sentiment around entities
- Word level sentiment dictionaries are often used to determine entity level sentiment

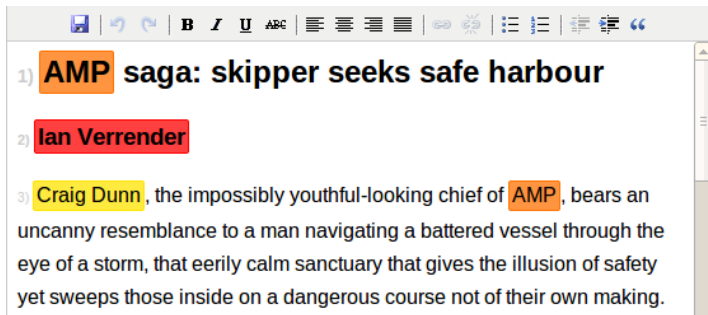
# Problems

- News articles are by nature quite objective, and sentiment tends to be mostly in the form of quotes
- Low inter-annotator agreement means that even humans struggle with the task

# M1 Sentiment: Deliverables

- annotate how the document encourages the reader to feel about each entity
- ✓ annotation scheme:
  - four options, no complicated granularities
  - Positive, negative, neutral and mixed
- ✓ Annotation tool extended to handle entity-level sentiment
- ✓✓ 500 SMH articles annotated with sentiment (we did 1000!)

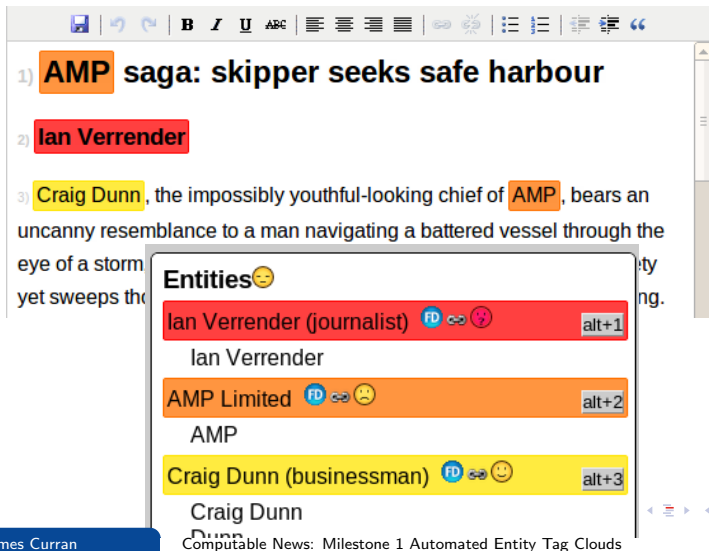
# Annotation process



The screenshot shows a web editor interface with a toolbar at the top containing icons for file operations, undo/redo, bold, italic, underline, text color, background color, bulleted list, numbered list, link, unlink, table, and quote. Below the toolbar, a text area contains the following content:

- 1) **AMP** saga: skipper seeks safe harbour
- 2) **Ian Verrender**
- 3) **Craig Dunn**, the impossibly youthful-looking chief of **AMP**, bears an uncanny resemblance to a man navigating a battered vessel through the eye of a storm, that eerily calm sanctuary that gives the illusion of safety yet sweeps those inside on a dangerous course not of their own making.

# Annotation process



1) **AMP** saga: skipper seeks safe harbour

2) **Ian Verrender**

3) **Craig Dunn**, the impossibly youthful-looking chief of **AMP**, bears an uncanny resemblance to a man navigating a battered vessel through the eye of a storm. yet sweeps the

**Entities** 😊

- Ian Verrender (journalist)** FD ∞ 😊 alt+1  
Ian Verrender
- AMP Limited** FD ∞ 😊 alt+2  
AMP
- Craig Dunn (businessman)** FD ∞ 😊 alt+3  
Craig Dunn

Computable News: Milestone 1 Automated Entity Tag Clouds

# Inter-annotator agreement (Cohen's Kappa)

- Entities
  - Annotating entity types: 0.80-0.88
  - Linking entities: 0.85-0.89
    - Choosing from set of all entities, so overstates agreement
  - Entity type agreement over aligned links: 0.93-0.97
- Events
  - Over sentences: 0.48, 0.00, 0.00, 0.54, 0.29, -0.01, 0.71, 0.77, 0.00, 0.12, 0.50, 0.40
  - Over doc: 0.48, 0.00, 0.00, 0.33, 0.47, -0.11, 0.44, 1.00, 0.00, 0.47, 0.63, 0.57
- Sentiment: 0.40 0.51 0.44
  - Unexpectedly low, but all of the disagreement comes from where to put the threshold



# Part I

## Extra slides

# M1 Sentiment engine

- Each adjective is given a sentiment score taken from SentiWordNet (v1)
- From this we calculate:
  - Story sentiment by summing the sentiment of adjectives appearing in the story
  - Mention sentiment by summing the sentiment of adjectives that appear in the sentence containing the mention
  - Entity sentiment by summing the sentiment of all mentions of the entity, where two or more mentions occur in the same sentence only one is used
- Note that this is quite similar to the techniques used by Lexalytics

# Computable News

- Joint project between CMCRC, Fairfax Digital and The University of Sydney.
- \$600K project led by James Curran:
  - 3 Postdocs: Ben Hachey, Matt Honnibal and David Vadas
  - 3 PhD students: Joel Nothman, Tim O'Keefe and Will Radford
  - CMCRC Project Manager: Tim Yeates
- Started November 2010 and runs for between 1 and 3 years.
- The first end-of-milestone meeting on February 17th.

# Entity Resource Pages at the New York Times

TIMES TOPICS &gt; PEOPLE &gt; M &gt; MADOFF, BERNARD L.

Bernard L. Madoff



Ruby Washington/The New York Times

Updated: Dec. 18, 2010

## Overview

On Wall Street, his name was legendary. Now it is infamous.

With money Bernard L. Madoff had earned as a lifeguard on the beaches of Long Island, he built a trading powerhouse that had prospered for more than four decades. At age 70, he had become an influential spokesman for the traders who are the hidden gears of the marketplace.

But on Dec. 11, 2008, Mr. Madoff was arrested at his Manhattan home by federal agents and charged in a 20 year Ponzi scheme that was the largest fraud in Wall Street history.

[http://topics.nytimes.com/top/reference/timestopics/people/m/bernard\\_l\\_madoff](http://topics.nytimes.com/top/reference/timestopics/people/m/bernard_l_madoff)

## Multimedia

Code	Unit	Hours	Prerequisite
10000	Introduction to Computer Science	3	None
10001	Computer Systems	3	10000
10002	Discrete Mathematics	3	10000
10003	Algorithms and Data Structures	3	10001
10004	Operating Systems	3	10001
10005	Database Systems	3	10001
10006	Computer Networks	3	10001
10007	Computer Graphics	3	10001
10008	Computer Security	3	10001
10009	Computer Architecture	3	10001
10010	Computer Systems	3	10001
10011	Computer Systems	3	10001
10012	Computer Systems	3	10001
10013	Computer Systems	3	10001
10014	Computer Systems	3	10001
10015	Computer Systems	3	10001
10016	Computer Systems	3	10001
10017	Computer Systems	3	10001
10018	Computer Systems	3	10001
10019	Computer Systems	3	10001
10020	Computer Systems	3	10001
10021	Computer Systems	3	10001
10022	Computer Systems	3	10001
10023	Computer Systems	3	10001
10024	Computer Systems	3	10001
10025	Computer Systems	3	10001
10026	Computer Systems	3	10001
10027	Computer Systems	3	10001
10028	Computer Systems	3	10001
10029	Computer Systems	3	10001
10030	Computer Systems	3	10001
10031	Computer Systems	3	10001
10032	Computer Systems	3	10001
10033	Computer Systems	3	10001
10034	Computer Systems	3	10001
10035	Computer Systems	3	10001
10036	Computer Systems	3	10001
10037	Computer Systems	3	10001
10038	Computer Systems	3	10001
10039	Computer Systems	3	10001
10040	Computer Systems	3	10001
10041	Computer Systems	3	10001
10042	Computer Systems	3	10001
10043	Computer Systems	3	10001
10044	Computer Systems	3	10001
10045	Computer Systems	3	10001
10046	Computer Systems	3	10001
10047	Computer Systems	3	10001
10048	Computer Systems	3	10001
10049	Computer Systems	3	10001
10050	Computer Systems	3	10001
10051	Computer Systems	3	10001
10052	Computer Systems	3	10001
10053	Computer Systems	3	10001
10054	Computer Systems	3	10001
10055	Computer Systems	3	10001
10056	Computer Systems	3	10001
10057	Computer Systems	3	10001
10058	Computer Systems	3	10001
10059	Computer Systems	3	10001
10060	Computer Systems	3	10001
10061	Computer Systems	3	10001
10062	Computer Systems	3	10001
10063	Computer Systems	3	10001
10064	Computer Systems	3	10001
10065	Computer Systems	3	10001
10066	Computer Systems	3	10001
10067	Computer Systems	3	10001
10068	Computer Systems	3	10001
10069	Computer Systems	3	10001
10070	Computer Systems	3	10001
10071	Computer Systems	3	10001
10072	Computer Systems	3	10001
10073	Computer Systems	3	10001
10074	Computer Systems	3	10001
10075	Computer Systems	3	10001
10076	Computer Systems	3	10001
10077	Computer Systems	3	10001
10078	Computer Systems	3	10001
10079	Computer Systems	3	10001
10080	Computer Systems	3	10001
10081	Computer Systems	3	10001
10082	Computer Systems	3	10001
10083	Computer Systems	3	10001
10084	Computer Systems	3	10001
10085	Computer Systems	3	10001
10086	Computer Systems	3	10001
10087	Computer Systems	3	10001
10088	Computer Systems	3	10001
10089	Computer Systems	3	10001
10090	Computer Systems	3	10001
10091	Computer Systems	3	10001

 Interactive

## Madoff's Clients

A list of individuals, banks, charitable foundations and investment firms exposed to losses in Bernard L. Madoff's investment funds.

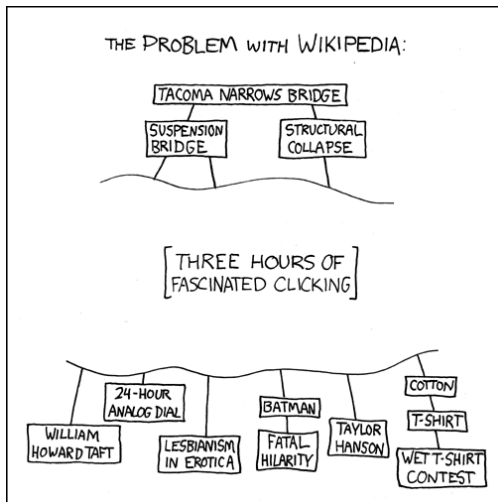


Interactive Feature

### A Timeline of the Madoff Fraud

Key dates in the Ponzi scheme and case of Bernard L. Madoff

# A good problem to have



<http://xkcd.com/214/>

# Quotation and Sentiment Extraction

Failed entrepreneur and former CEO [Eddy Groves](#) will defend a criminal charge relating to the collapse of [Australia](#)'s biggest childcare chain, [ABC](#).

“I’ve pleaded not guilty today and I will vigorously defend the charge, and that’s really all I can say right now,” [Groves](#) said.

- [Eddy Groves](#) said “I’ve pleaded not guilty today...”
- [Eddy Groves](#) → ☹️

# User-defined Alerts

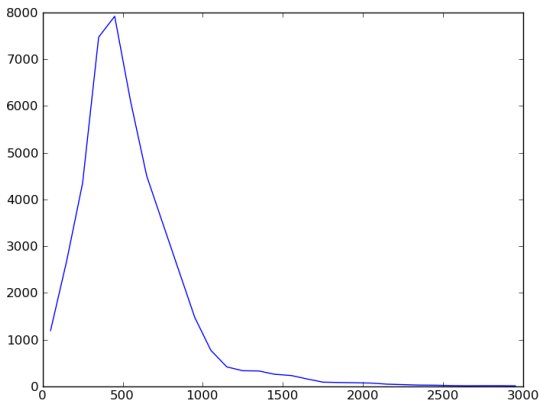
- Stories about ORGANISATION.
- What PERSON has said about ORGANISATION.
- New building projects in LOCATION.
- ...

## Postgres vs Tokyo Tyrant

TASK	POSTGRES SQL	TOKYO TYRANT	
Load 500,000 stories	171.303/248.292	71.256/79.392	13
Retrieve all 500,000 stories	11.281/10.410	75.317/81.909	14
Retrieve 1,000 random stories	0.275/0.218	0.071/0.082	
Concurrent retrieval	6.645/6.027	3.402/3.404	



	SECTION	AVERAGE WORD COUNT
Word counts by section	Births	48
	Marriages	54
	Tenders	80
	Deaths	102
	SS	200
	Domain East	324
	News	381
	Property	459
	Business	560
	Drive	623
	Sport	656
	Good Weekend	1260



# Entity type scheme v0.9 – Core types

- Animal
- Artefact
- Award
- Event
- Facility
- Organisation
- Person
- Product
- Vessel
- Work of Art
  
- Miscellaneous
- Adjectival (nationality, religion, political party, etc.)

# Entity type scheme v0.9 – Facets for core types

- Art/Architecture
- Business
- Education
- Entertainment
- Environment/Weather
- Family/Genealogy (clan, tribe, etc.)
- Food/Drink
- Government/Civil
- Housing/Real Estate
- Law
- Media
- Medicine/Health
- Military
- Political
- Religious
- Science/Technology
- Sport
- Transportation

# Entity type scheme v0.9 – Location and address types

- Country
- Region
- City/Suburb
- Natural Feature
- Celestial Body
  
- Post/Street
- Phone/Fax
- Email
- Url/Path

# Entity type scheme v0.9 – Number and date types

- Monetary Amount
- Percent
- Quantity
- Score
  
- Date
- Time

## Entity type scheme v0.9 – Terminological types

- Currency (e.g., AUD)
- Natural Disaster (e.g., Cyclone Larry)
- Gene/Protein (e.g., Tinman)
- Language (e.g., Chinese)
- Market Index (e.g., ASX 200)
- Medical Condition (e.g., Cholera)
- Person Role (e.g., Minister of Foreign Affairs)
- Species (e.g., Drosophila Melanogaster)
- Substance (e.g., Penicilin)
- Treatment (e.g., Chemotherapy)