Event Detection in Unstructured Text

(Using Common Lisp)

Jason Cornez, CTO RavenPack

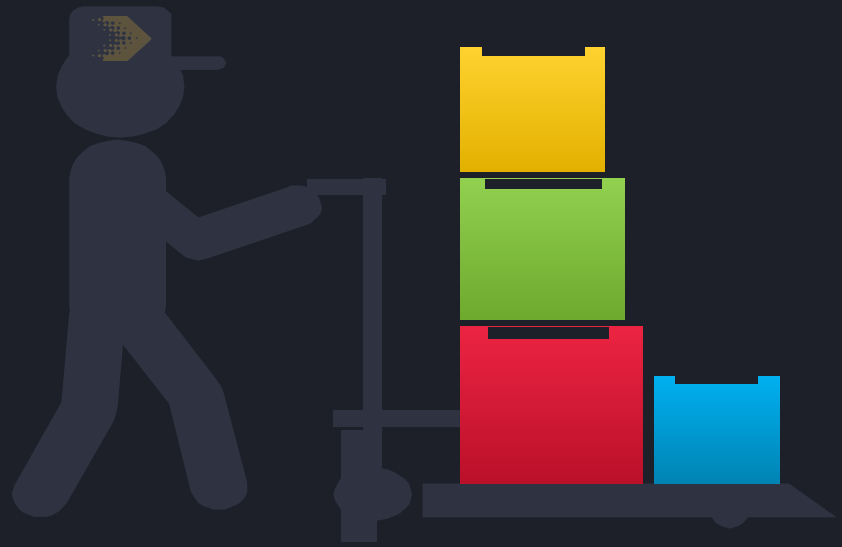
What does RavenPack Do?

RavenPack extracts Meaning from Unstructured BigData

www.ravenpack.com

Unstructured data is typically natural language text documents.

- 100.000+ Documents Daily
- Low Latency: about 250ms
- Archive of more than 300 million
- AWS Cloud, 24x7



RAVENPACK ANALYTICS

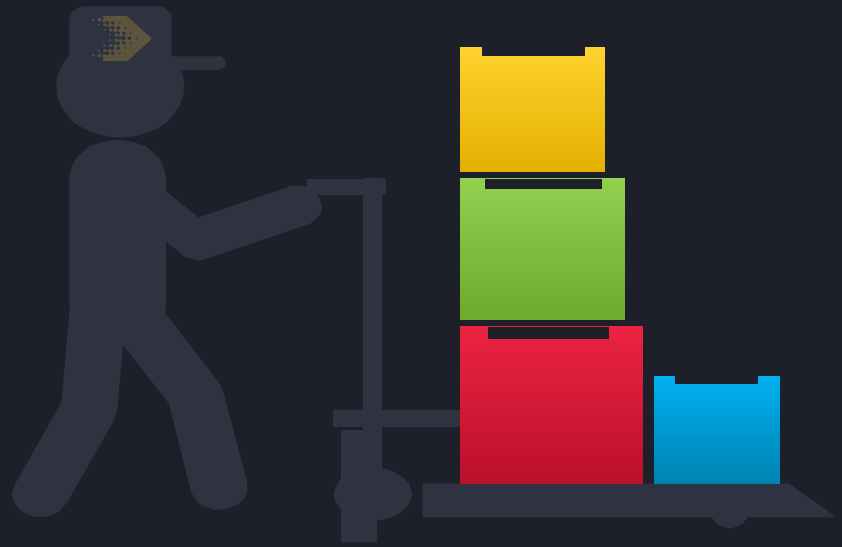
Actionable Insights from News and Social Media

www.ravenpack.com

- Entities and Events
- Relevance
- Sentiment
- Novelty

RavenPack also processes Private Content

- Email, Skype, Slack, Files
- Custom Entities
- We build Great Software
- We sell Data and Services

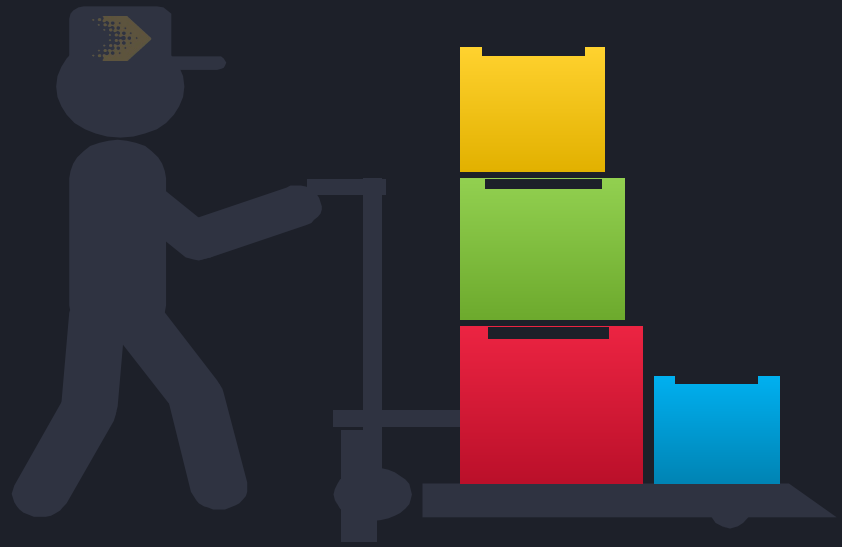


RavenPack

How do you accomplish this?

Modern Architecture

- Distributed
- Multi-threaded
- Easier Migration to the Cloud
- Horizontal Scaling

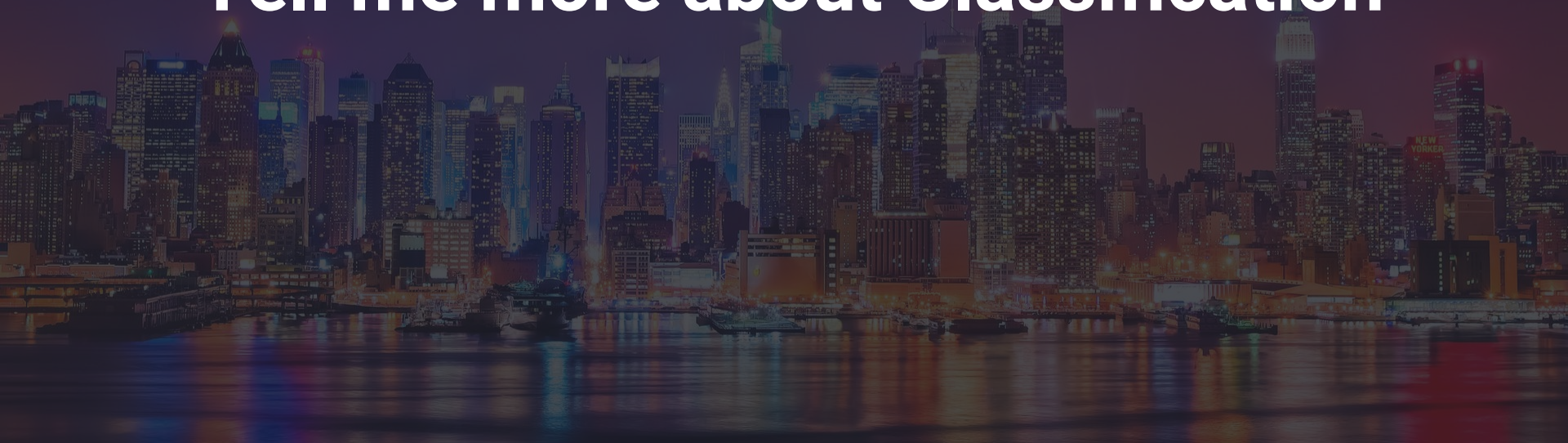


Separation of Concerns

- Collection - Java
- Classification - Lisp
- Analytics - Lisp
- Distribution - Lisp / Python

RavenPack

Tell me more about Classification



Streams-based Classification Framework

Presented at ELS 2013 in Madrid

www.ravenpack.com

- Entity Detection
- Attribute Matching
- Event Detection
- Many Others...

RavenPack

What is Event Detection?

Event Detection

RavenPack tracks thousands of event types.

These can be corporate events like:

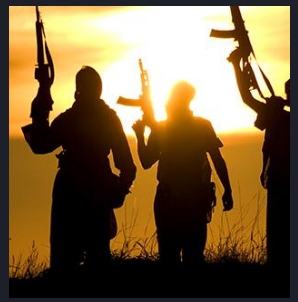
- bankruptcy
- layoffs
- product announcements
- analyst ratings
- earnings

Also global events like:

- currency exchange
- war
- terrorism
- crop yields
- floods

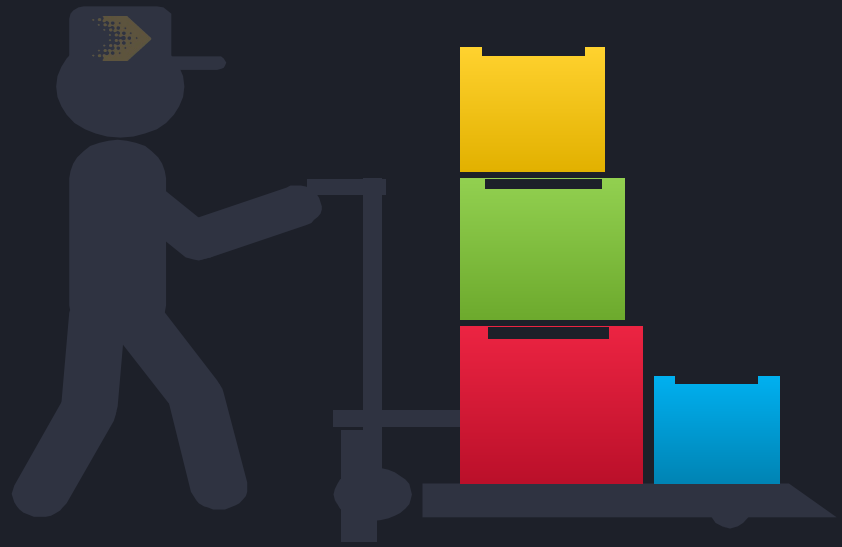
Future events could be related to:

- sports, entertainment, ...



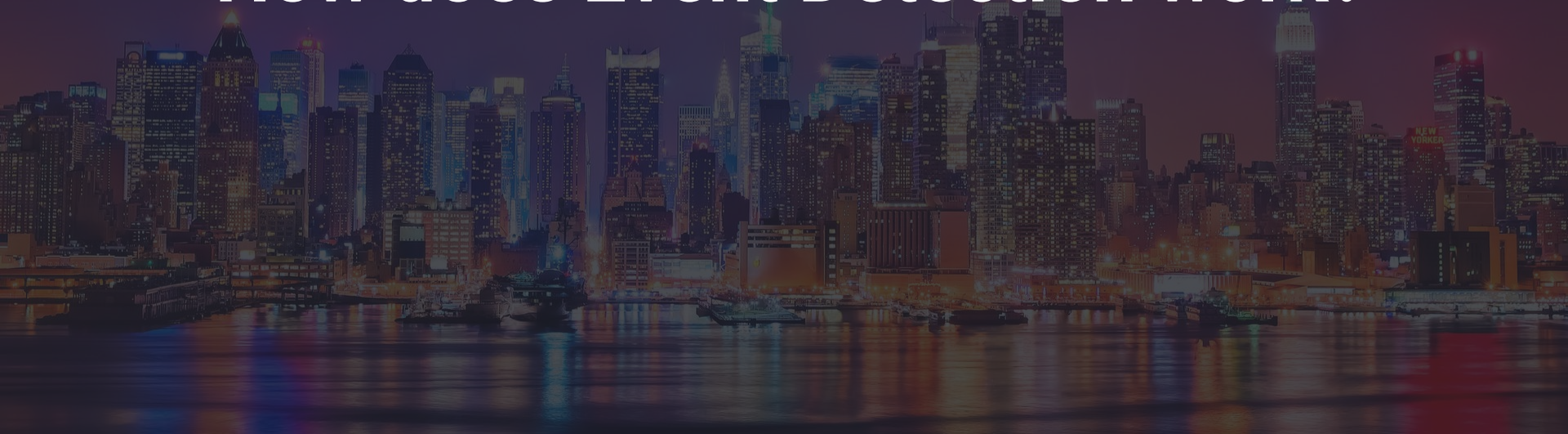
Event Detection

- Which Entities Participate
- What Role does each Play
- Dates, Magnitude, Sentiment, Trust
- Event Consolidation



RavenPack

How does Event Detection work?



Event Detection Classifier

Event Detection matches Templates against Annotated Text

Templates

"Regular" patterns
Authored by humans
About 100,000 templates
Built using In-House tool

Annotated Text

Entity Detections
Attributes

The same text is often annotated in multiple ways

Some Example Templates


```
$COMPANY  %DECLARE  %BANKRUPTCY-KEYWORD
$ORGANIZATION  %FORECAST  %UNEMPLOYMENT  %FALL
$PLACE  %CONSUMER-CONFIDENCE  %RISE
$COMPANY  *  NET  INCOME  %FALL
```

Hmm, something's not quite right...

Some Example Templates


```
( :$COMPANY :%DECLARE :%BANKRUPTCY-KEYWORD )
( :$ORGANIZATION :%FORECAST :%UNEMPLOYMENT :%FALL )
( :$PLACE :%CONSUMER-CONFIDENCE :%RISE )
( :$COMPANY :* :NET :INCOME :%FALL )
```

That’s better...

Some Example Annotated Text

www.ravenpack.com

THE	WEINSTEIN	CO	DECLARES	BANKRUPTCY	GENE	MADDAUS	/	VARIETY
$FACT				%BANKRUPTCY-KEYWORD			%AND	!MAGNITUDE-SENTIMENT
$COMPANY(*) / 1			%ANNOUNCE				$FORWARDSLASH	$SENTIMENT / -
%THE		%CORPORATE-FORM	%DECLARE					%POSITIVE-WORDS
		%COMPANY	%TESTIFY					
?	?	?	?	?	+	*	?	?

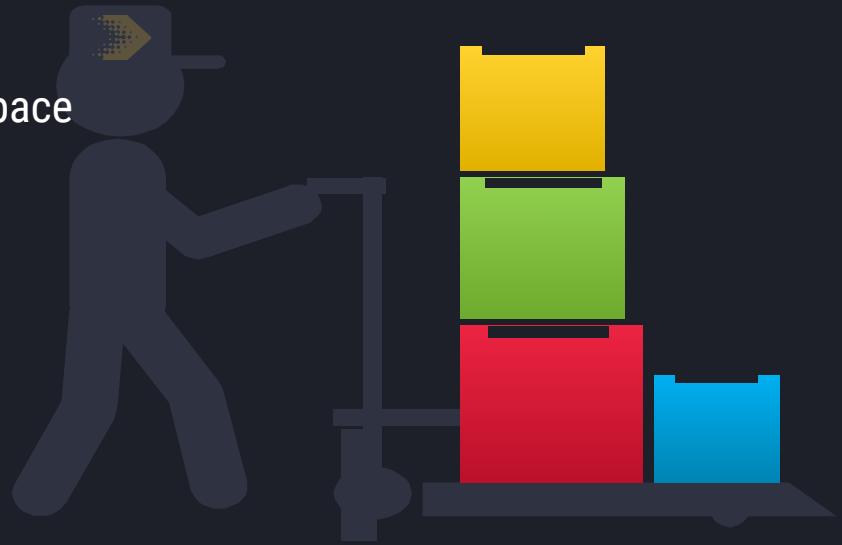
FOMC	EXPECT	UNEMPLOYMENT	TO	DECLINE	ONLY	SLOWLY
$ORGANIZATION(*) / 6	%FORECAST	%ECONOMIC-INDICATOR	!MAGNITUDE-SENTIMENT			!MAGNITUDE-SENTIMENT
$ORGANIZATION(*) / -	%EXPECT	%UNEMPLOYMENT	$SENTIMENT / -			$SENTIMENT / -
			%NEGATIVE-WORDS			$MAGNITUDE / -
			%FALL			%LOW-MAGNITUDE
			%REJECT			%NEGATIVE-WORDS
			%STOCK-LOSS-ACTION			
			%PREPOSITION	!MAGNITUDE-SENTIMENT		
			%PREPOSITION-ACTUAL-INDICATOR	$SENTIMENT / -		
			%PREPOSITION-PLACE	%FALL		
			%PREPOSITION-TIME	%NEGATIVE-WORDS		
				%REJECT		
				%STOCK-LOSS-ACTION		
?	?	?	?	?	?	?

RavenPack

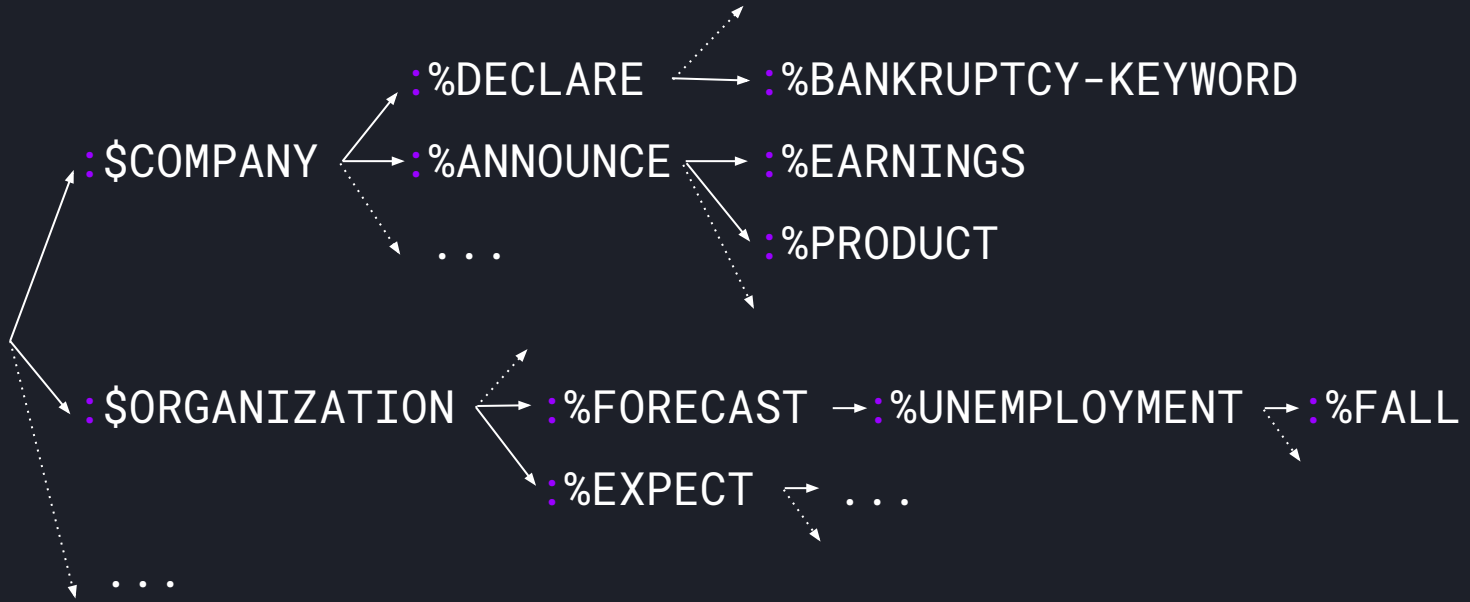
What about the Matching itself?

Matching Templates with Annotated Text

- We have something like a Rete algorithm
  But ours was “invented” independently
- Templates are stored in a Trie
  Additions have low impact on speed and space
- Multiple Templates can match given Text
  Scoring to choose a Winner
- Wildcards are the most expensive part

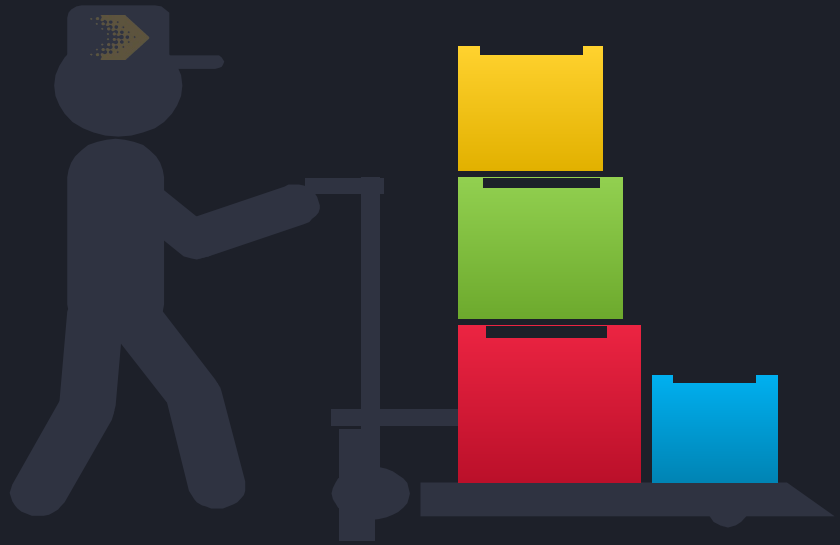


A Template Trie



Populating Event Types

- An Event Type is composed of multiple roles
- Each role allows particular types of data
- The matching engine populates roles with entities, attributes, or data
- Conditions further define a match



Event Type Examples

www.ravenpack.com

legal-issues

Template maps to type when:

– POS –
– POS –
– POS –
– POS –
– POS –

always

1.	legal-issues-plaintiff	nil	COMPANY, NATIONALITY, ORGANIZATION, PERSON
2.	legal-issues-defendant	nil	COMPANY, NATIONALITY, ORGANIZATION, PERSON
3.	legal-issues-period	nil	PERIOD(TIMESPAN)
4.	legal-issues-date	nil	PERIOD(Date)
5.	legal-issues-position-plaintiff	nil	POSITION
6.	legal-issues-position-defendant	nil	POSITION
7.	legal-issues-authority	nil	ORGANIZATION(GOVERNMENT)
8.	legal-issues-location	nil	PLACE
9.	legal-issues-sentiment-modifier	nil	MAGNITUDE
10.	legal-issues-sentiment	nil	SENTIMENT
11.	legal-issues-opinion-plaintiff	nil	COMPANY, NATIONALITY, ORGANIZATION, PERSON
12.	legal-issues-opinion-defendant	nil	COMPANY, NATIONALITY, ORGANIZATION, PERSON
13.	legal-issues-opinion-authority	nil	ORGANIZATION(GOVERNMENT)
14.	legal-issues-opinion-location	nil	PLACE
15.	legal-issues-rater	nil	COMPANY, NATIONALITY, ORGANIZATION, PERSON

earnings-up

Template maps to type when:

2 > 3
– POS –
– POS –
– POS –
– POS –

(AND (> $6 $8))

1.	earnings-up	nil	COMPANY
2.	earnings-up-actual	nil	CURRENCY, NUMBER
3.	earnings-up-previous	nil	CURRENCY, NUMBER
4.	earnings-up-percentage	nil	NUMBER, PERCENTAGE
5.	earnings-up-period	nil	PERIOD(TIMESPAN)
6.	earnings-up-date	nil	PERIOD(Date)
7.	earnings-up-method	nil	PERIOD(METHOD)
8.	earnings-up-previous-period	nil	PERIOD(TIMESPAN)
9.	earnings-up-type	nil	EARNINGS-TYPE
10.	earnings-up-product	nil	PRODUCT
11.	earnings-up-sentiment-modifier	nil	MAGNITUDE
12.	earnings-up-sentiment	nil	SENTIMENT
13.	earnings-up-opinion	nil	COMPANY
14.	earnings-up-opinion-product	nil	PRODUCT
15.	earnings-up-rater	nil	COMPANY, NATIONALITY, ORGANIZATION, PERSON

RavenPack

Are there any Limitations?

Event Detection Limitations

Opportunities for Improvement

Many Templates

Authored by humans and they are “fragile”. Even with 100,000 templates, there are many texts where we fail to match events.

Uncaptured Context

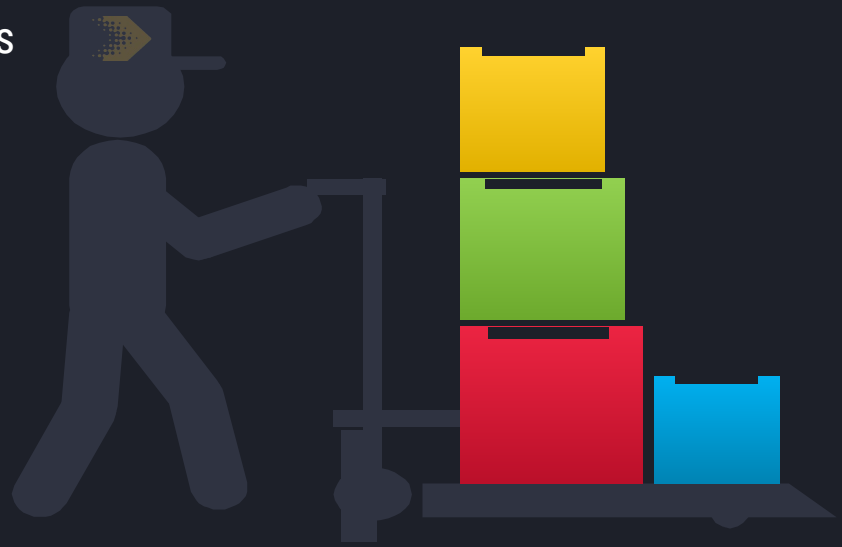
There is often text in wildcards, before or after the match that could contribute context, which we’d like to capture - without the need to write more templates.

RavenPack

How are you addressing these?

Event Detection Initiatives

- Augmenting matches to produce Enriched Events
- A matching Template might not fill all Roles
- Inspect Nearby Annotated Text
- Populate Dates, Magnitudes, Sentiment



Augmented Match Example 1

www.ravenpack.com

"MEXICO: Consumer Confidence Index Rises 1% In September On Monthly Basis"

(:$PLACE :%CONSUMER-CONFIDENCE :%RISE :$PERCENTAGE :%PREPOSITION-TIME :$PERIOD)

$PLACE	%CONSUMER-CONFIDENCE	%RISE	$PERCENTAGE	%PREPOSITION-TIME	$PERIOD(*)	*$PERIOD*
MEXICO	CONSUMER CONFIDENCE INDEX	RISES	1	IN	SEPTEMBER	*ON MONTHLY BASIS*

1.	consumer-confidence-up	42125631020B6113DF09246655383EE9 - Mexico (ESS: 0.53)
2.	consumer-confidence-up-actual	
3.	consumer-confidence-up-previous	
4.	consumer-confidence-up-percentage	1
5.	consumer-confidence-up-period	TIMESPAN: 2017-09-01 00:00:00 -> 2017-10-01 00:00:00 (Period: -----)
6.	consumer-confidence-up-rater	
7.	consumer-confidence-up-date	
8.	consumer-confidence-up-method	***METH: MOM***
9.	consumer-confidence-up-sentiment-modifier	
10.	consumer-confidence-up-sentiment	
11.	consumer-confidence-up-opinion	
12.	consumer-confidence-up-previous-period	

Augmented Match Example 2

"In first quarter 2018, the Company repurchased 50.6 million shares of its common stock"

(:$COMPANY :%BUYBACK :$NUMBER :%SHARES)

%PREPOSITION-TIME $PERIOD	$COMPANY(*)	%BUYBACK	$NUMBER	%SHARES
IN FIRST QUARTER 2018	THE COMPANY	REPURCHASED	50,600,000	SHARES

1. buybacks
2. buybacks-actual
3. buybacks-previous
4. buybacks-percentage
5. buybacks-period
6. buybacks-date
7. buybacks-sentiment-modifier
8. buybacks-sentiment
9. buybacks-opinion
10. buybacks-rater

E8846EF4BDF75A87A2EEC359DB4D6588 - Wells Fargo & Co. (ESS: 0.59) (Rep

TIMESPAN: 2018-01-01 00:00:00 -> 2018-04-01 00:00:00 (Period: 2018-Q1)

RavenPack

What’s Next?

Themes and Beyond

The Future of Event Detection

Themes

A Theme comprises the essential ingredients required for a match of a particular Event Type.

Machine Learning

Just have a human say, “this sentence is an example of that event type”. And have the computer do the rest.

Theme Considerations

(:$COMPANY :%BANKRUPTCY)

Company enters bankruptcy.

Company exits bankruptcy.

Company should consider bankruptcy.

Company avoids bankruptcy.

Company denies bankruptcy rumors.

RavenPack

Who Built This?

Event Detection Classifier Team

- Andrew Lawson
- Nick Levine
- André Thieme
- Maybe You?


RavenPack
Thank you!
Any Additional Questions?


RavenPack
Yes, We're in Marbella.
Yes, We're Hiring!
Jason Cornez
jcornez@ravenpack.com