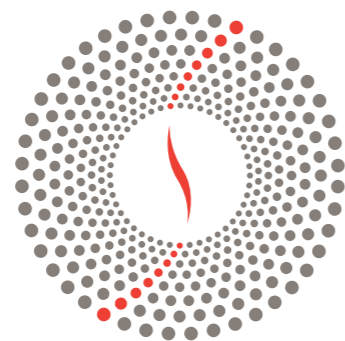


A Case Study on Automated Risk Assessment of Ships Using Newspaper-Based Event Extraction



Jesper Hoeksema
VU University Amsterdam



synerScope

connecting the dots

Willem Robert van Hage
SynerScope B.V.

The Problem

- The Dutch Coastguard monitors the Netherlands Exclusive Economic Zone
- 154.011 km², monitored by 51 full-time operators
- Typically contains around 1300-1400 vessels, with one entering or leaving approximately every 30 seconds

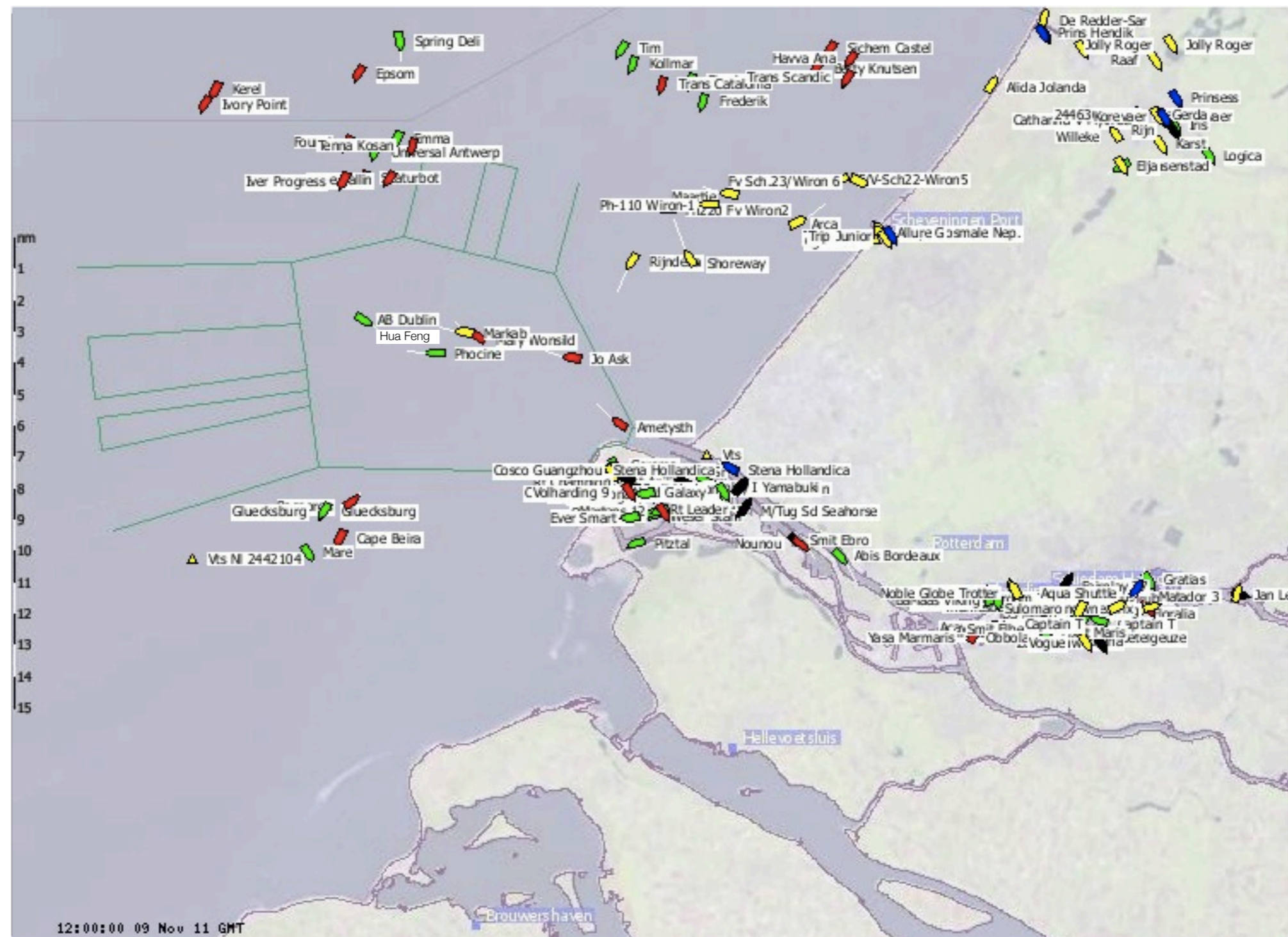
The Problem (continued)

- Currently, no outside sources are taken into account by the observation system when investigating vessels
- This is done manually by the operators, usually just by ‘googling’ on a second computer
- Number of ships is too big to process manually, so operators prioritize, often based on incomplete information



Example

The Hua Feng enters the harbor, and nobody cares



However...



First, I want to say sorry for this very poor quality, but it was my lense, which was decentered in 2008. To the photo: Ex "Gulf Jash" (today "Hua Feng") passing Cuxhaven bound for Hamburg on the 07.10.2008.



Ship data from: Clarkson Research Services Limited

Gulf Jash (Ex: Probo Koala) Dead/Demolished 47,980 DWT, Bulk/Oil Carrier, Built 1989

Owners/Managers are Gulf Navigation, Built at Hanjin H.I., Double Hull, Panama Flagged, RINA Classed, Ice Strengthened C Class, Length Overall of 182.80 m., Length Between Perpendiculars of 175.72 m., Draught of 13.09 m., Beam of 31.98 m., 53.35 Tonnes per Centimetre Immersion, Gross Tonnage of 31,255, MAN B. & W. Engine, Speed of 15.00 kts at 28.00 tonnes per day, Heavy Fuel Oil, Horsepower of 12,800B at ...

Owner/Manager Details

Gulf Navigation Group

Specialised Details

Cargo Capacities of 51,000 cu.m. and 321,000 Barrels, Segregated Ballast Tanks, Clean Ballast Tanks, 7 Tanks, Epoxy Tank Coating, Grain Capacity of 50,669 cu.m., 7 Holds, 7 Hatches, Strengthened for Heavy Cargo, 2 Gantry Crane(s) with a safe working load of 28 tonnes.

Additional information

IDENTIFICATION: Exnames are Probo Koala, Probo Baoning. Call Sign 3FOG9, IMO Number 8309816.
DIMENSIONS/TONNAGES: Moulded Depth of 18.85 m., Keel to mast air draft of 46.60 m., Tonnage of 26,146 Panama Canal Net, 26,837 Suez Canal Net, 14,257 International Net, 14,389 Light Displacement and 47,222 Dwt (long). ENGINE DETAILS: Engine Description 2 S.A. 4-cyl., Engine Model 4L80MC-E Mk1. CARGO HANDLING: 7 Centre Tanks with a capacity of 51,000 cu.m., Crude Oil Washing, Ore Cargo Capacity of 51,000 cu.m., Grab Fitted, Hatch Dimensions are 6 @ 13.50 x 20.20 m., 1 @ 13.50 x 12.20 m. SAFETY AND OTHER DETAILS: Last known special survey in July 2004, Clean Ballast Tanks, Inert Gas System.

Reported Sale and purchase information

Reported Sales available on [Shipping Intelligence Network](#)

Reported Fixture information

Reported Fixtures available on [Shipping Intelligence Network](#)

2006 Côte d'Ivoire toxic waste dump

From Wikipedia, the free encyclopedia

The **2006 Côte d'Ivoire toxic waste dump** was a [health crisis](#) in [Côte d'Ivoire](#) in which a ship registered in [Panama](#), the ***Probo Koala***, chartered by the Dutch-based oil and commodity shipping company [Trafigura Beheer BV](#), offloaded [toxic waste](#) at the Ivorian port of [Abidjan](#). The waste was then dumped by a local contractor at as many as 12 sites in and around the city of Abidjan in August 2006.

The gas caused by the release of these chemicals is blamed by the UN and the [government of Côte d'Ivoire](#) for the deaths of 17 and the injury of over 30,000 Ivorians with injuries that ranged from mild headaches to severe burns of skin and lungs. Almost 100,000 Ivorians sought medical attention for the effects of these chemicals.^[1]

The substance was claimed by Trafigura to have been "slops", or waste water from the washing of the *Probo Koala's* tanks. An inquiry in [the Netherlands](#) in late 2006 revealed the substance was more than 500 tonnes of a mixture of fuel, [caustic soda](#), and [hydrogen sulfide](#) for which Trafigura chose to not pay a €1,000 per cubic metre disposal charge at the [port of Amsterdam](#). The *Probo Koala* was turned away by several countries before offloading the toxic waste at the Port of Abidjan.^{[2][3]}

Trafigura denied any waste was transported from the Netherlands, saying that the substances contained only tiny amounts of hydrogen sulfide, and that the company did not know the substance was to be disposed of improperly. In early 2007, the company paid US\$198 million for cleanup to the Ivorian government without admitting wrongdoing, and the Ivorian government pledged not to prosecute the company.^[4] A series of protests and resignations of Ivorian government officials followed this deal.

A civil lawsuit in London was launched in 2008 by almost 30,000 Ivorians against Trafigura. In May 2009, Trafigura announced it would sue the [BBC](#) for [libel](#) after its *Newsnight* program alleged the company had knowingly sought to cover up its role in the incident. In September 2009 *The Guardian* obtained and published internal Trafigura emails showing that the traders responsible knew how dangerous the chemicals were. Shortly afterwards Trafigura offered an unnamed settlement figure to the [class action](#) suit against it.^[5]

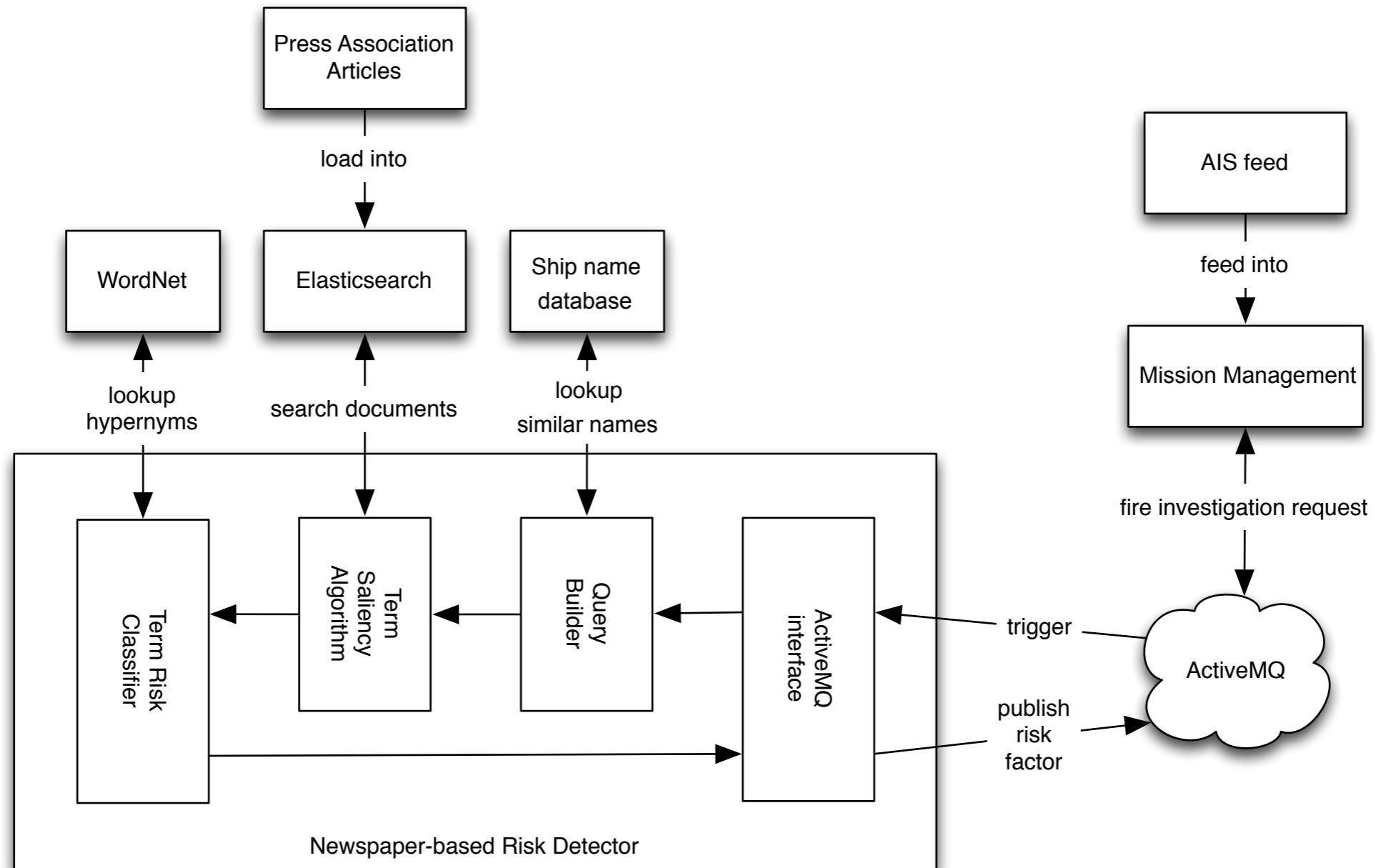


A truck full of garbage on the streets of Abidjan. Much of Trafigura's toxic waste was dumped in large open areas in the poor suburbs of the city

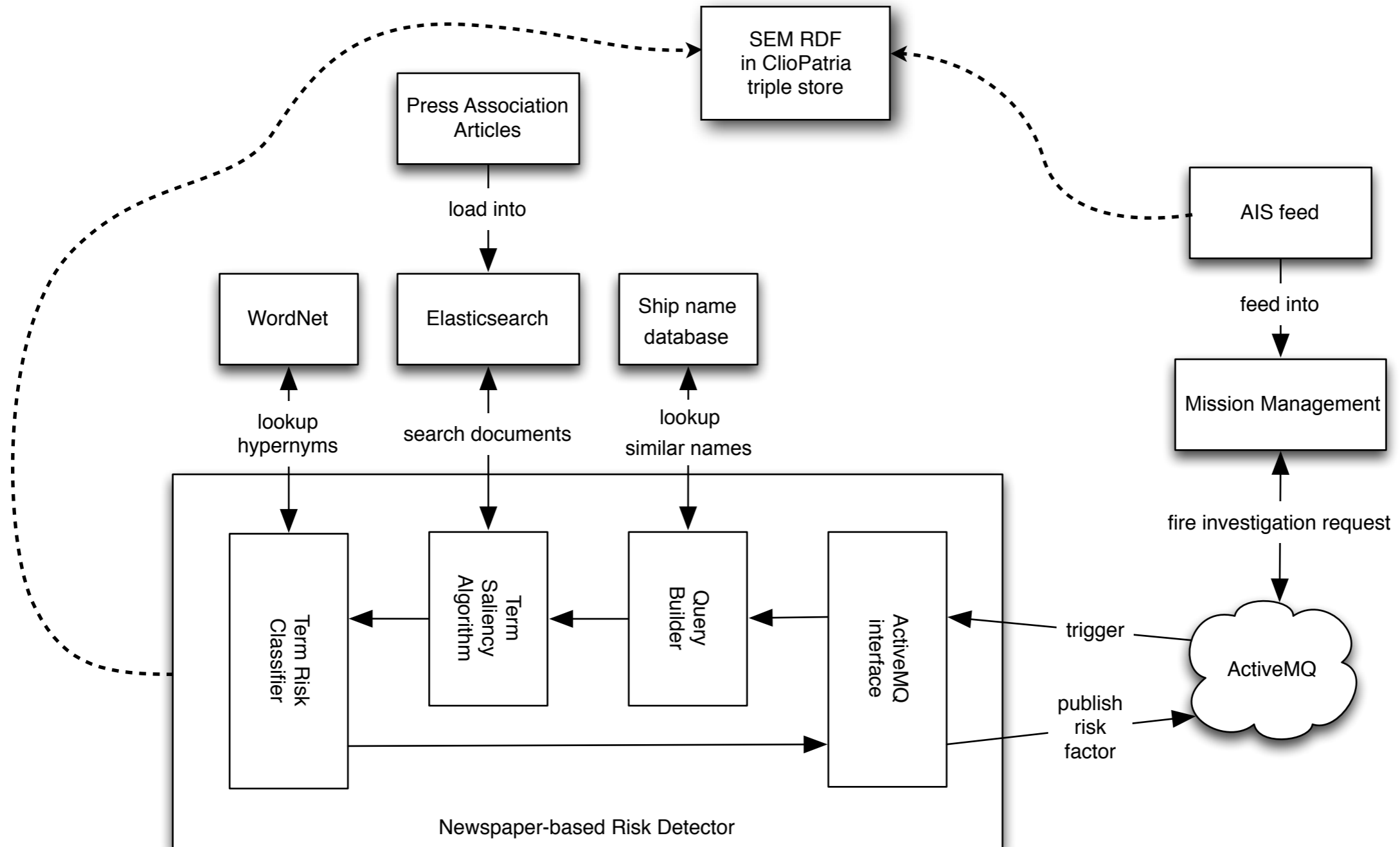
Approach

- Full prioritization based on ship's history
- Combination of relevance feedback, lexical databases and domain information
- Quickly determine the 'types' of the events in a ship's history, and only look for types we are interested in
- No automatic action is taken, operator is alerted

Architecture



Architecture



Querying

- 25000 maritime-related press releases from the Press Association
- Stored in (a modified) Elasticsearch cluster
- Retrieve documents by name, taking extra care to exclude similar names (using a list of known names)
 - For example, when looking for the ship “*Mary*”, explicitly exclude the “*Queen Mary*”
- Retrieve the term vectors for returned document, as well as for a sample of 100 random documents that were not returned

Rocchio

- Relevance feedback technique
- Reshape the query into one term vector that describes documents about the vessel with respect to the other documents

$$\vec{Q}_m = \left(a * \vec{Q}_o \right) + \left(b * \frac{1}{|D_r|} * \sum_{\vec{D}_j \in D_r} \vec{D}_j \right) - \left(c * \frac{1}{|D_{nr}|} * \sum_{\vec{D}_k \in D_{nr}} \vec{D}_k \right)$$

D_r – hijack, pirates, ship, captain, etc.

D_{nr} – ship, captain, sea, engine, etc.

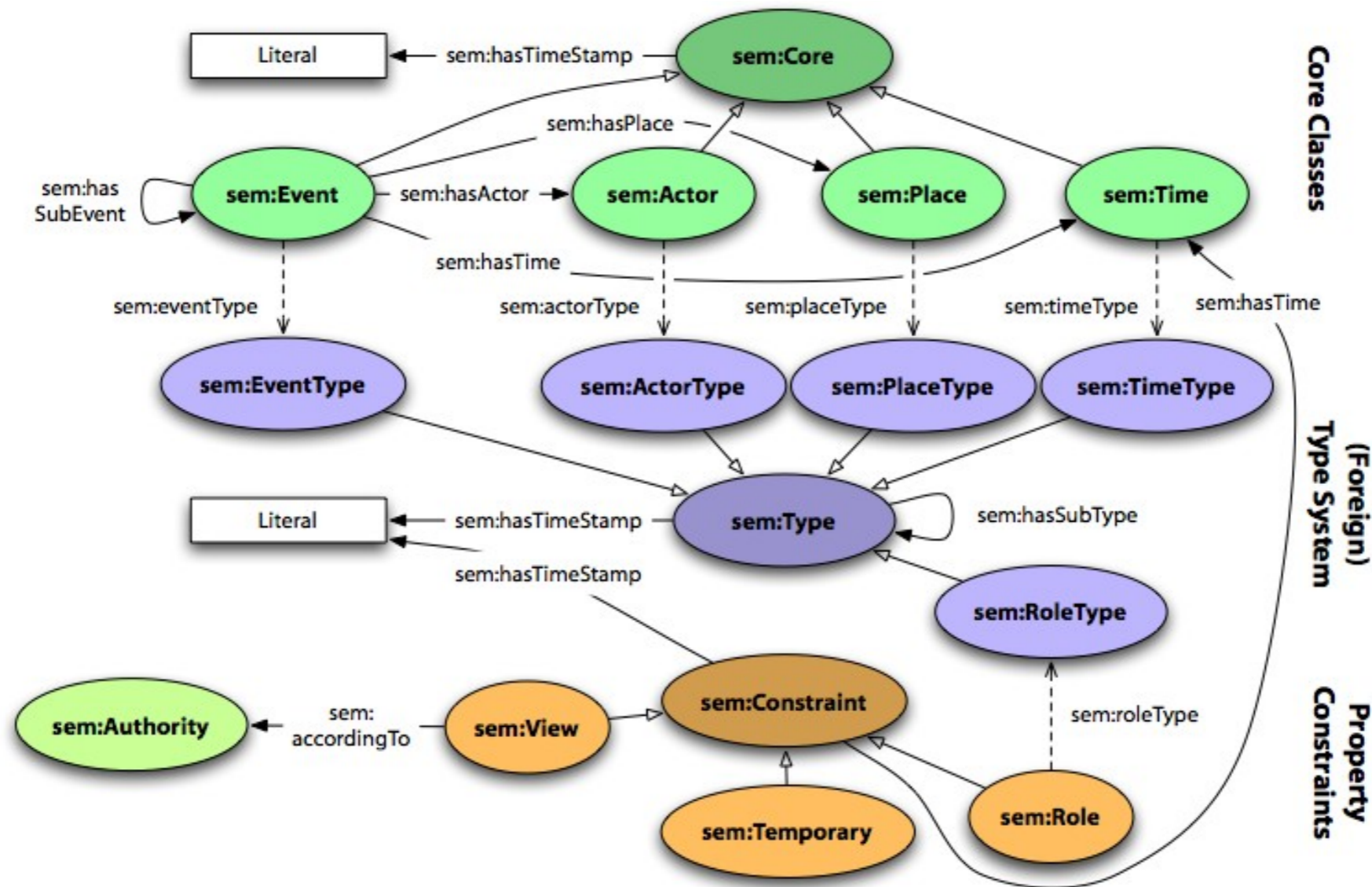
Q_m – **hijack, pirates**, ship, captain, etc.

Risk classification

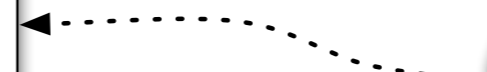
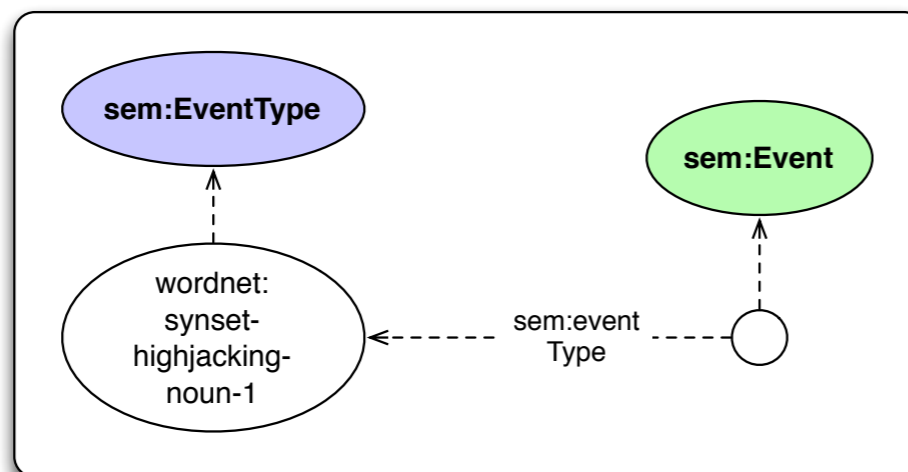
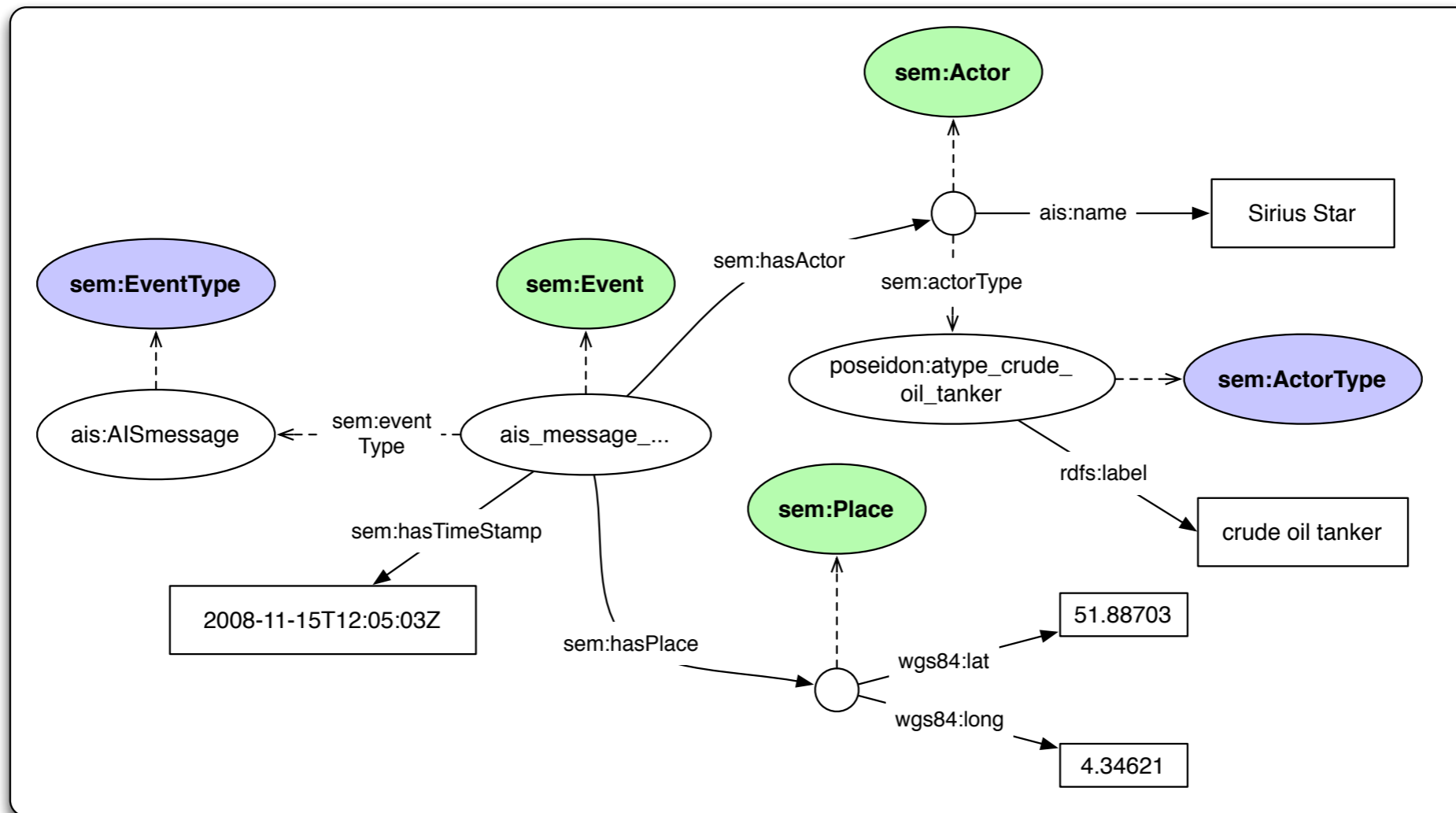
- Pre-defined set of concepts for each event type we want to detect
- Each term in the calculated vector is lemmatized and looked up in WordNet
- When matched with a pre-defined concept, a score is added to the event type's score
- This score is calculated by taking the frequency of the term in the term vector, divided by the ambiguity (number of synsets that contain the term)

Simple Event Model

- For representation, each detected event type is assumed to correspond to one distinct SEM event



Simple Event Model



Evaluation

- **E1** - Given the name of a ship, are relevant documents returned?
- **E2** - Given a set of documents (from E1), does the system classify the correct event types?
- **E3** - Given the name of a ship (with known risk behavior), does the system classify this ship correctly and completely?
- Behavior for non-remarkable vessels was tested qualitatively, looking for anomalies when classifying 76000 known ship names

Results

Vessel	E1			E2					E3				
	D_F	D_R	P	T_{TP}	T_{FP}	T_{FN}	P	R	T_{TP}	T_{FP}	T_{FN}	P	R
Exxon Valdez	34	8	0.24	3	0	0	1.00	1.00	3	0	0	1.00	1.00
Probo Koala	0	0	1.00	0	0	0	1.00	1.00	0	0	1	1.00	0.00
Costa Concordia	0	0	1.00	0	0	0	1.00	1.00	0	0	2	1.00	0.00
Estonia	136	80	0.59	0	0	1	1.00	0.00	0	0	1	1.00	0.00
Herald of Free Enterprise	95	52	0.55	1	1	0	0.50	1.00	1	1	0	0.50	1.00
Sirius Star	46	44	0.96	2	0	0	1.00	1.00	2	0	0	1.00	1.00
Vindo	26	26	1.00	2	0	0	1.00	1.00	2	0	0	1.00	1.00
Edinburgh Castle	4	0	0.00	0	1	0	0.00	1.00	0	1	1	0.00	0.00
Zeldenrust	1	1	1.00	2	1	0	0.67	1.00	2	1	0	0.67	1.00
Scandinavian Star	9	9	1.00	1	3	0	0.25	1.00	1	3	0	0.25	1.00
Lady Azza	0	0	1.00	0	0	0	1.00	1.00	0	0	2	1.00	0.00
Ronin	2	2	1.00	2	1	0	0.67	1.00	2	1	0	0.67	1.00
Union Pluto	1	1	1.00	2	1	0	0.67	1.00	2	1	0	0.67	1.00
Achille Lauro	72	48	0.67	0	0	3	0.00	0.00	0	0	2	1.00	0.00
Viking Victor	23	22	0.96	0	1	1	0.00	0.00	0	1	1	0.00	0.00
Astree	4	4	1.00	1	2	0	0.33	1.00	1	2	0	0.33	1.00
Total	453	297	0.66	16	11	5	0.59	0.76	16	11	10	0.59	0.62

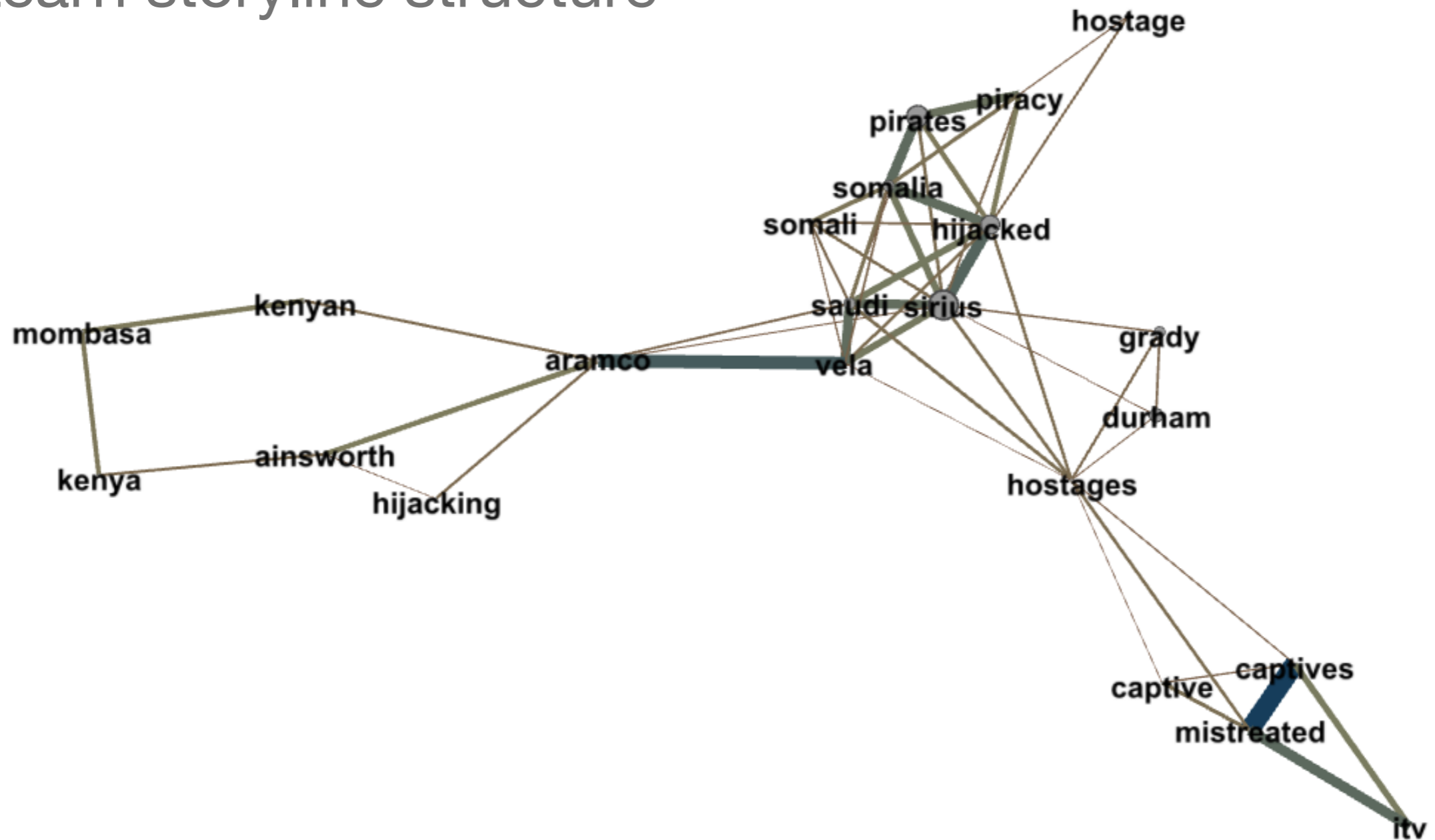
- 3064 of the 76696 vessels triggered an alert on at least one category,
- Would require an operator to confirm 5 vessels each hour, instead of 2 each minute (when manually assessing)

Future work

- Separate the distinct events, and
Cluster multiple mentions of the same event
- Improve identity resolution
(e.g. different ships with the same name)

Future work (continued)

- Learn storyline structure



Recap

- Event-type extractor for assisting maritime security operators
- Relevance feedback, lexical databases, domain information
- Raises 'red flags' fairly accurate given that articles about the ship are actually in our database
- Number of false negative is low enough not to overload an operator

Questions?