

UNIVERSITY OF BERGEN

Media Content Analysis in News Angler and MediaFutures

Andreas L Opdahl

I2S Research Group and MediaFutures WP3

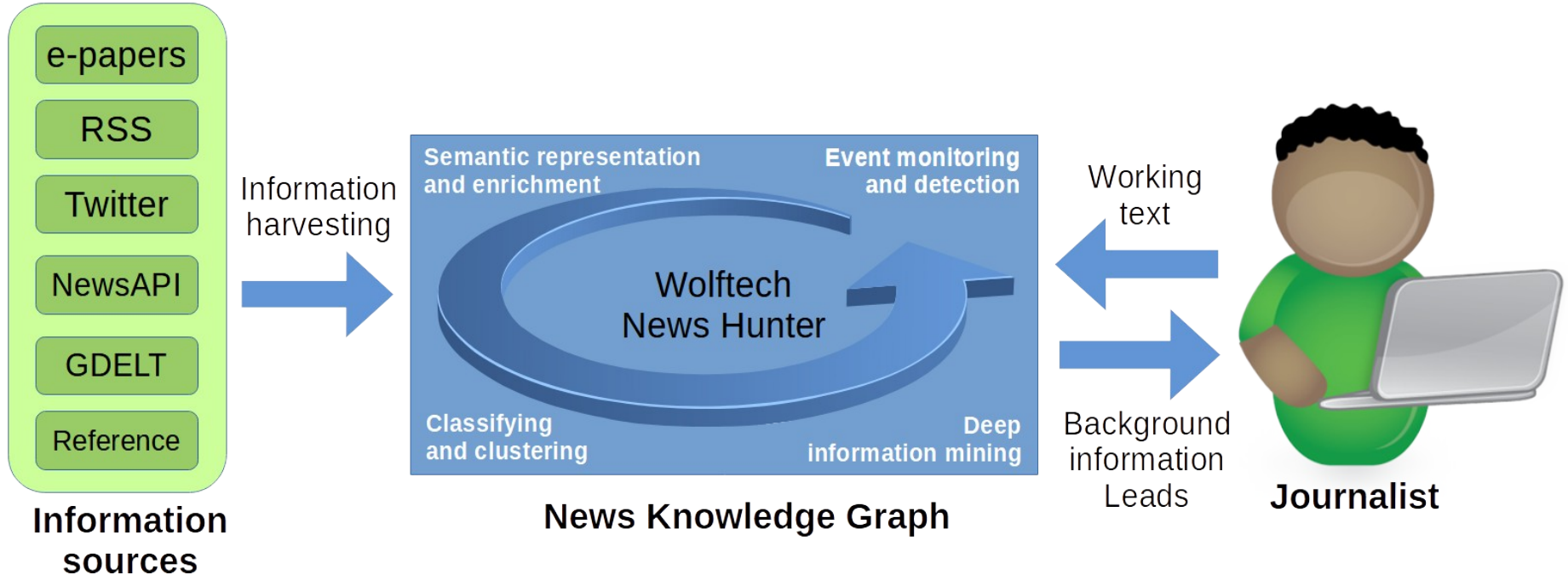
UNIVERSITY OF BERGEN



Joint work with: ...a lot of people...

The News Hunter Platform

- Harvesting news-related information from social media and other sources; analysing, organising, enriching and presenting news-related information to journalists.



Usage scenarios

- Automatic tagging and categorisation
- Background facts on demand
- Related stories on demand
- Suggesting informants
- Who is this guy?
- Who is this guy's friends?
- What's happening?
- What's more?
- What's my angle?
- Fill in my angle



News Editor

Write and publish the latest news

Brexit court defeat for UK government

He said: "The country voted to leave the European Union in a referendum approved by Act of Parliament. And the government is determined to respect the result of the referendum."
Government lawyers had argued that prerogative powers were a legitimate way to give effect "to the will of the people".

Berven, Christensen, Moldeklev, Opdahl & Villanger (2020).
A knowledge-graph platform for newsrooms



Helly Hansen
Journalist

Live inspect™

Refresh

Entities 3

Concepts 8

Concepts

brexit court defeat
european union
prerogative powers
country voted
give effect
uk government
government lawyers
referendum approved

Usage scenarios

- Automatic tagging and categorisation
 - Background facts on demand
 - Related stories on demand
 - Suggesting informants
 - Who is this guy?
 - Who is this guy's friends?
 - What's happening?
 - What's more?
 - What's my angle?
 - Fill in my angle
- Breaking news
 - 1) News Hunter finds a new solid event in the news graph.
 - 2) News Hunter retrieves angles that fit the new solid event.
 - 3) News Hunter retrieves a small knowledge graph that contains
 - the named entities, concepts/topics/categories, and relations in the new event
 - related facts that fit the selected angle
 - 4) News Hunter may suggests the new solid and angled event to the journalist.
 - 5) If so, the front end presents the solid event, the recommended angles, and the graph of facts to the journalist.



Example



Universal TV @tvuniversal · 23 Feb 2017



WAR DEG DEG: **Madaxweyne Farmaajo oo Raiisal Wasaare u magacaabay Xasan Khayre Cali** fb.me/5O5S7oLVY

 Translate Tweet



1



2



3



Example



Universal TV @tvuniversal · 23 Feb 2017

WAR DEG DEG: **Madaxweyne Farmaajo oo Raiisal Wasaare u magacaabay Xasan Khayre Cali** [fb.me/505S7oLVY](https://www.facebook.com/505S7oLVY)

Translate Tweet



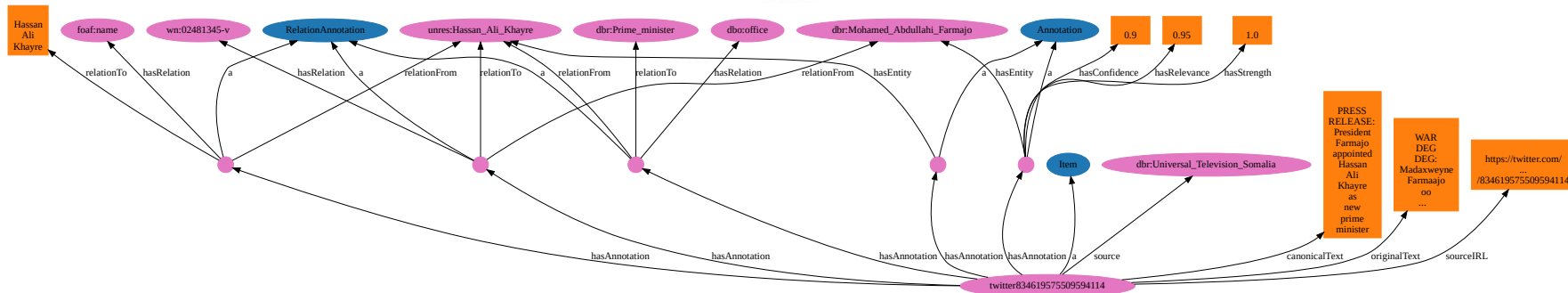
1



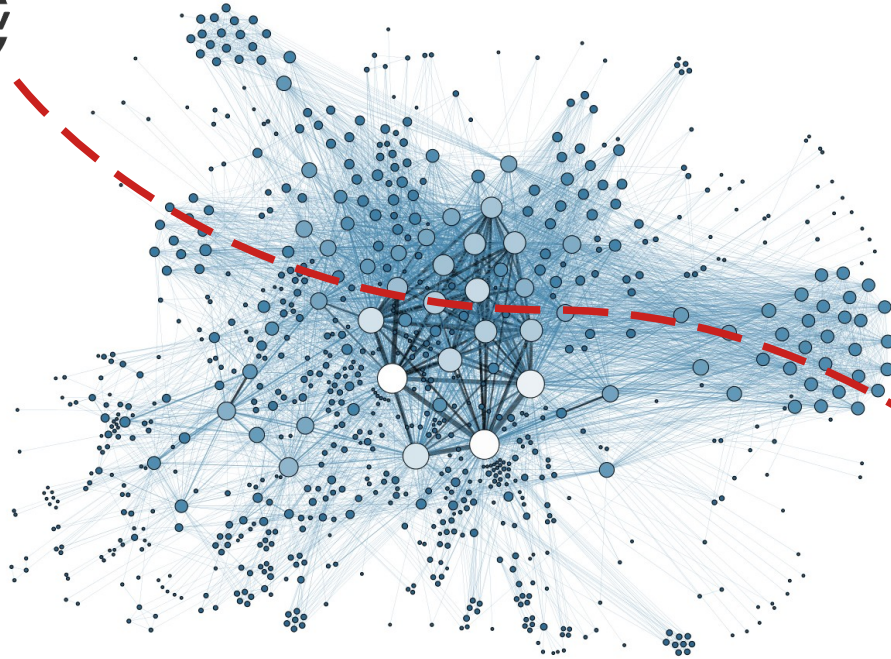
2



3

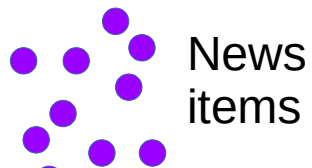


Opdahl & Tessem (2021): Ontologies for finding journalistic angles.
Software and Systems Modeling, 20(1), 71-87.





Harvesting



News items

Lifting



Item graphs

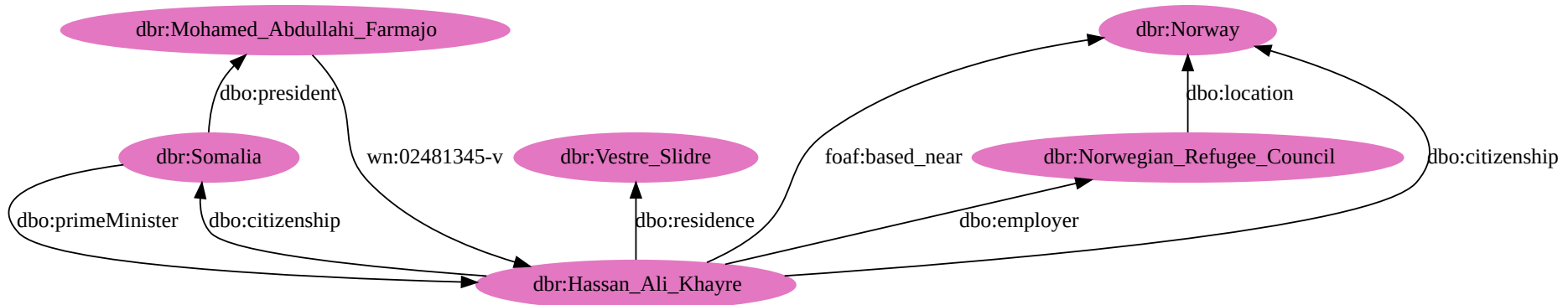


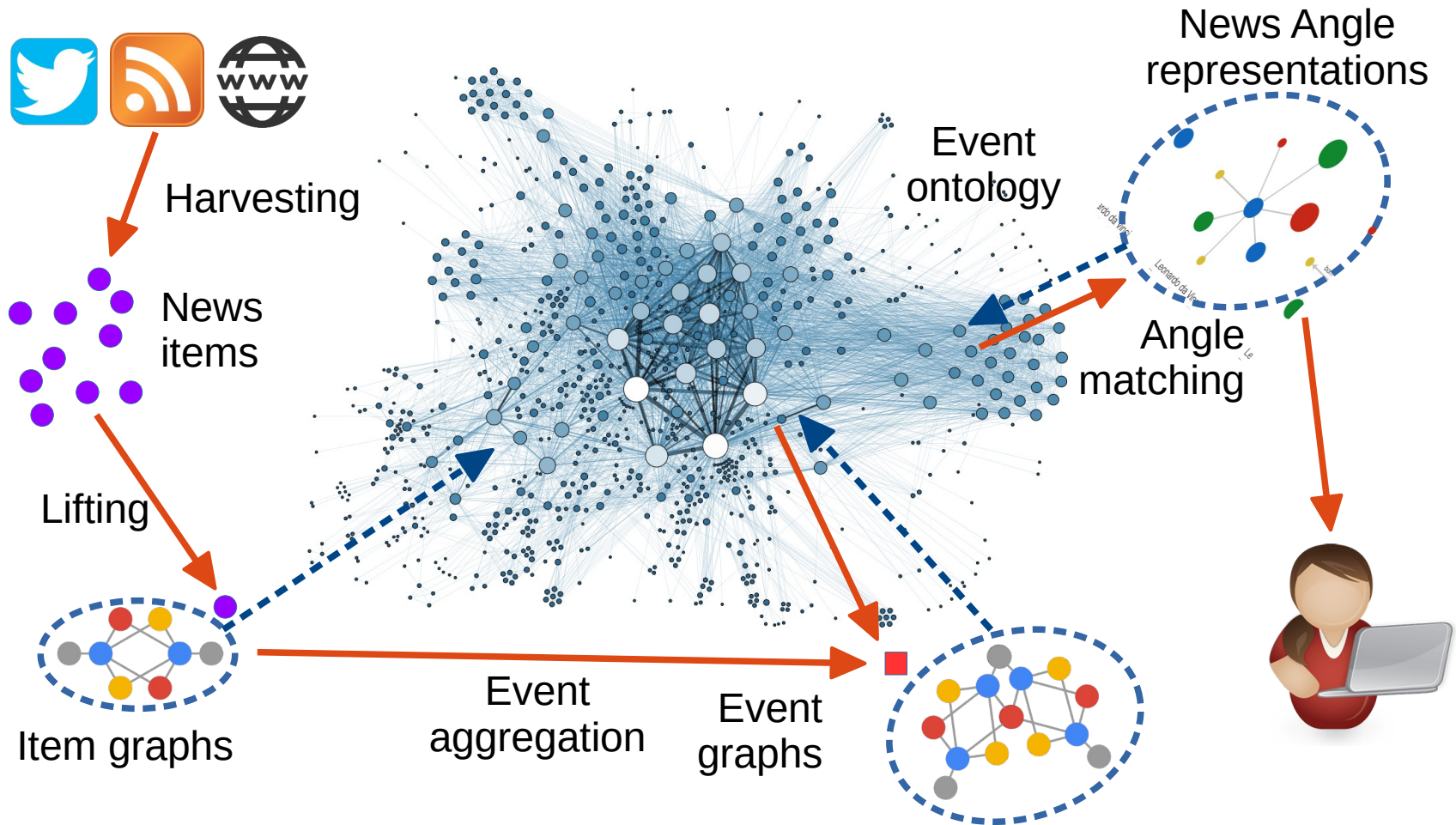
Knowledge Graphs

- A standard representation format for factual information
 - syntactic integration: uniform representation, standard technologies
 - semantic integration: standard identifiers and vocabularies
 - also enrichment, external linking, reasoning, ...

Knowledge Graphs

- A standard representation format for factual information
 - syntactic integration: uniform representation, standard technologies
 - semantic integration: standard identifiers and vocabularies
 - also enrichment, external linking, reasoning, ...







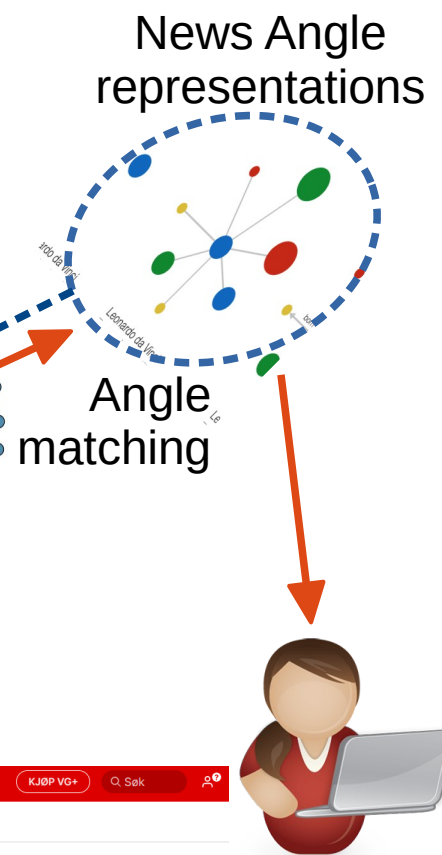
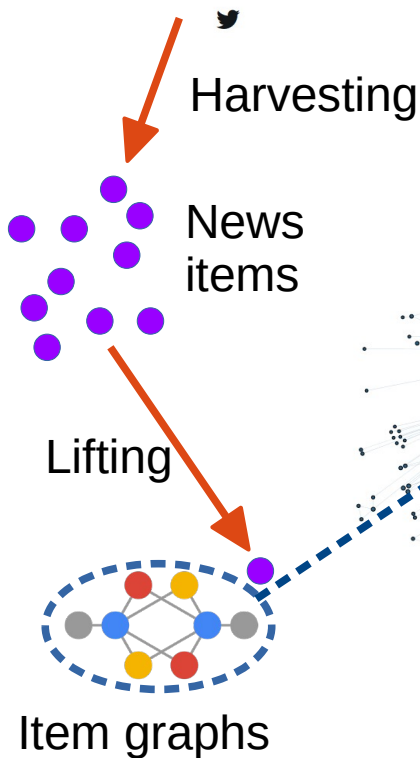
Universal TV @tvuniversal · 23 Feb 2017

WAR DEG DEG: Madaxweyne Farmaajo oo Raiisal Wasaare u magacaabay

Xasan Khayre Cali fb.me/5O5S7oLVY

Translate Tweet

1 2 3



VG VG LIVE VGTV VG+ SPORT TV-GUIDE PENT KJØP VG+ Q. Sek 👤

NYHETER | Alle nyheter Innenriks Utenriks Meninger Coronaviruset Stortingsvalget

Norske Hassan ny statsminister i Somalia

Norsk-somaliske Hassan Ali Khaire (49) er utnevnt til ny statsminister i Somalia.

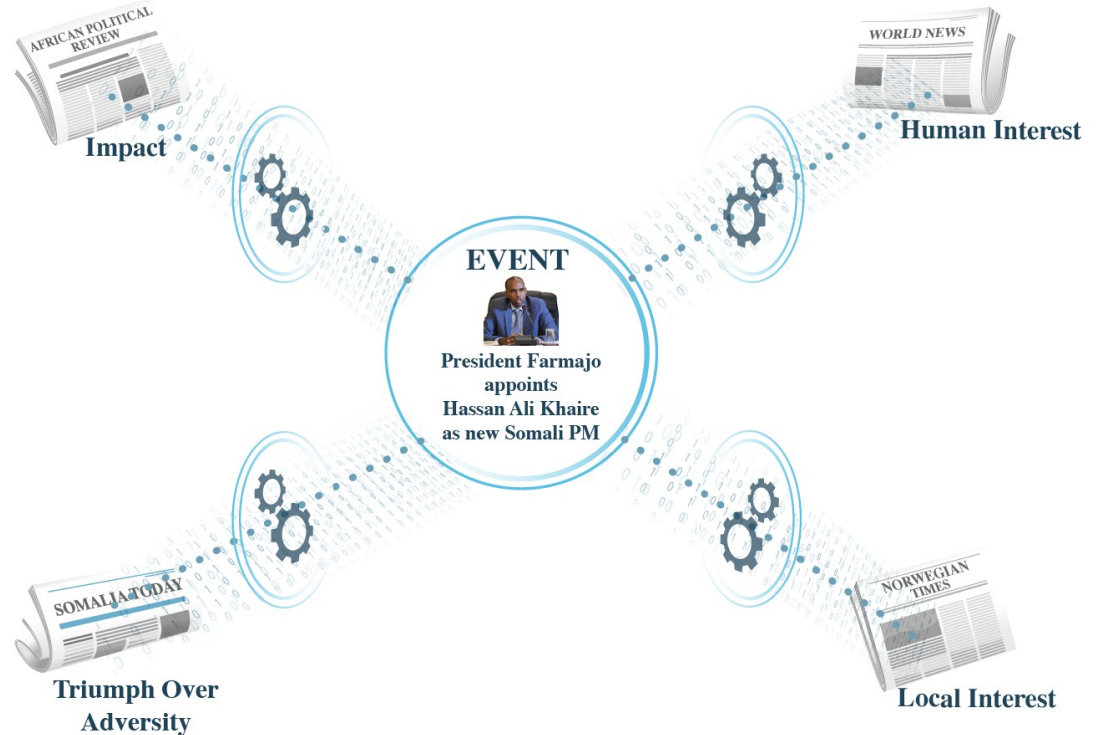
News Angles

“A predefined story ‘angle,’ [...] provides reporters a theme around which to build a story. [...] news values distill what people find interesting and important to know about. They include importance, interest, controversy, the unusual, timeliness, and proximity.”
(Shoemaker & Reese 1991/96)

News Angles

“A predefined story ‘angle,’ [...] provides reporters a theme around which to build a story. [...] news values distill what people find interesting and important to know about. They include importance, interest, controversy, the unusual, timeliness, and proximity.”
(*Shoemaker & Reese 1991/96*)

Can we formalise news angles and use them to detect newsworthy events in knowledge graphs?



Motta, Daga, Opdahl & Tessem (2021):
Analysis and design of computational news angles.
IEEE Access, 8, 120613-120626.

Example: Local interest angle

(def-relation event-with-local-angle (?remote-event ?place ?local-event)

"event-with-local-angle is an event that takes place in a place (?remote-place), but concerns an actor that has been involved in an event concerning a place of local interest (?place)"

:sufficient (and (event ?remote-event)

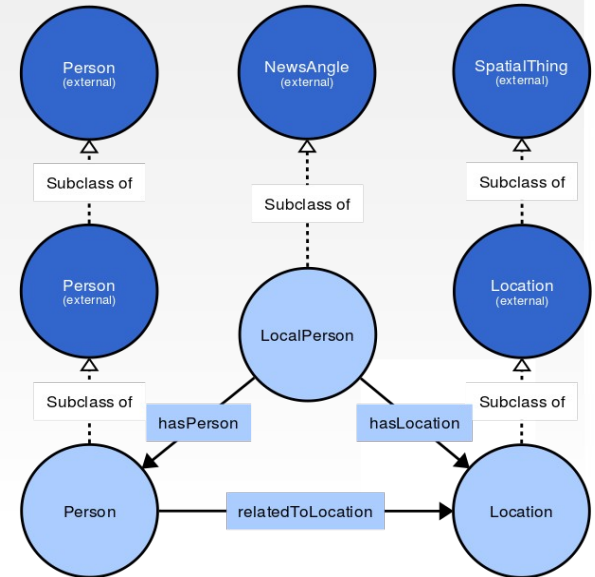
(has-place ?remote-event ?remote-place)

(not (= ?remote-place ?place))

(has-actor ?remote-event ?actor)

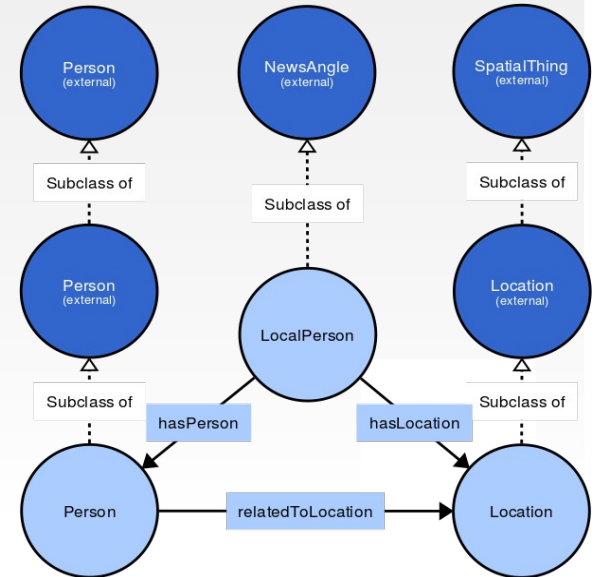
(has-actor ?local-event ?actor)

(has-place ?local-event ?place)))



Example: Local interest angle

```
CONSTRUCT {  
  [ rdf:type na:LocalPerson;  
    na:hasPerson ?person;  
    na:hasLocation ?location  
  ].  
  ?person na:relatedToLocation ?location .  
} WHERE {  
  ?person rdf:type na:Person .  
  ?location rdf:type na:Location .  
  ?person na:relatedToLocation / na:basedNear? ?location .  
}
```



Example: Location angles

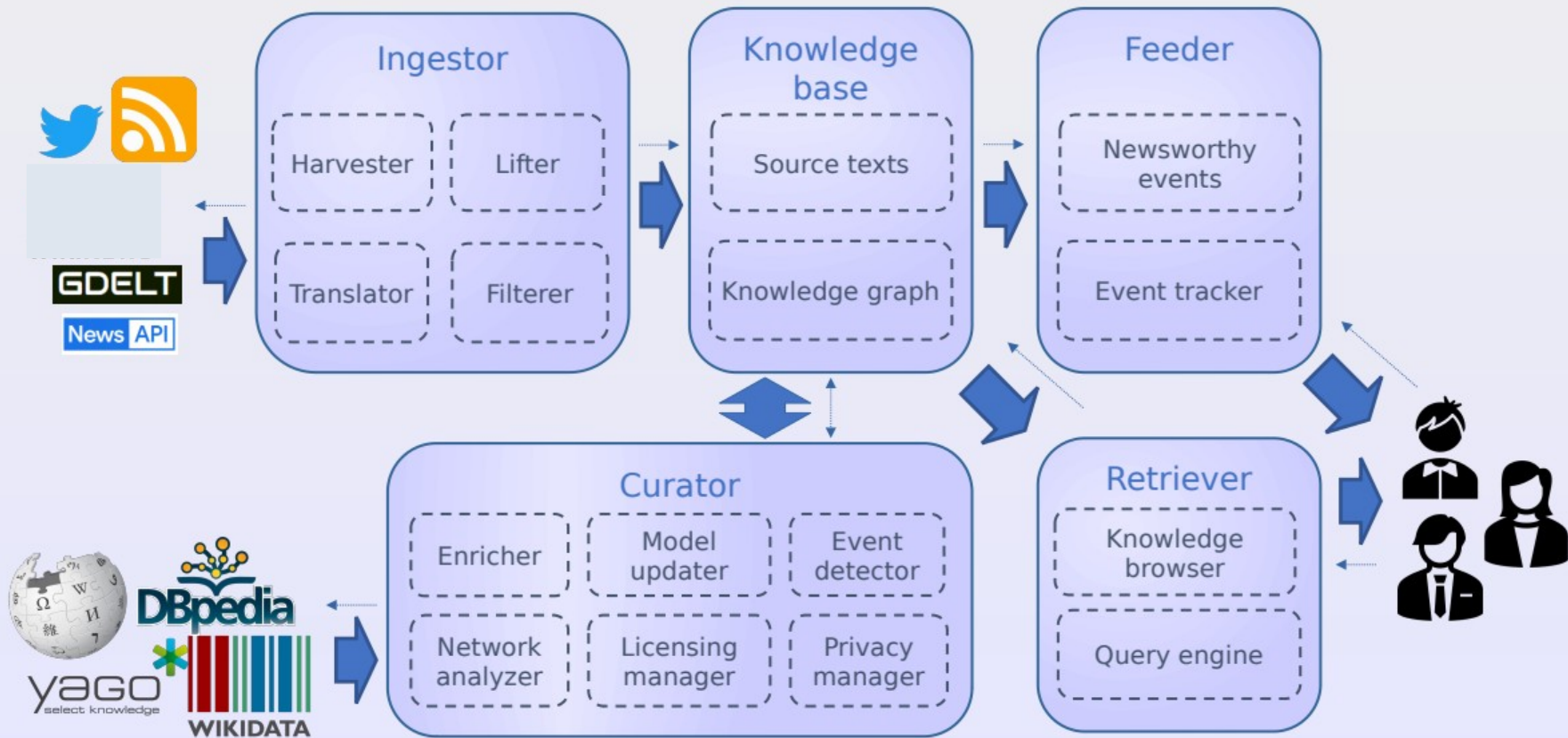
- Location graph service
 - select *Location of Interest (LoI)*
 - find Entities of Interest (EoIs) related to the LoI
 - people, companies, tourist locations, roads, and bridges...
 - create Location KG
 - distill as look-up table {Eoi: KG chain back to LoI}
- Angle detector service
 - find relations to LoI in incoming events
 - events represented as small graphs
 - perform table look-up
 - return chain(s) from Eoi(s) back to LoI
- Also *People of Interest?*





The News Hunter architecture

Harvesting news-related information from social media and other sources; analysing, organising, enriching and presenting news-related information to journalists. Implemented state-of-the-art big data and distributed technologies.



Architecture services

- **Twitter harvester:**
real-time-streams from news providers
- **RSS harvester:**
feeds from news providers
- **NewsAPI harvester:**
real-time news feeds
- **GDELT harvester:**
events, mentions and articles
- **GDELT lifter**
- **URL harvester:**
download full texts
- **Neuralcoref:**
co-reference resolver
- **DBpedia Spotlight:**
named entity linking (NEL)
- **SpaCy NEL:**
named entity linking
- **Kolitsas NEL:**
named entity linking
- **Early relation extractors:**
OpenNRE, SPN4RE
- **Automatic uploading:**
RDF in Kafka → Blazegraph

- Organised as a *micro-service architecture*





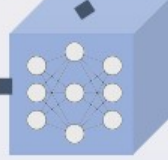
The News Hunter infrastructure



Service nodes

Web scraping, API, user interfaces, semantic lifting processes

- Light-to-medium processing
- Python, REST API, ...



Computation-intensive nodes

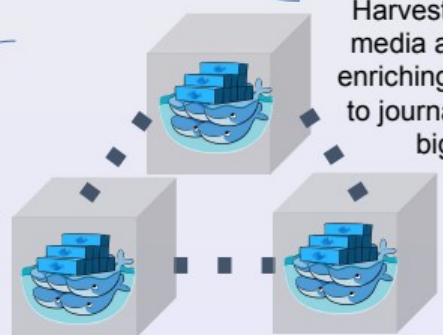
Complex AI services and training processes.

- CPU, RAM, GPU intensive
- Python, spaCy, ...

Management nodes

Service orchestration and monitoring

- Lighter processing
- Docker Swarm

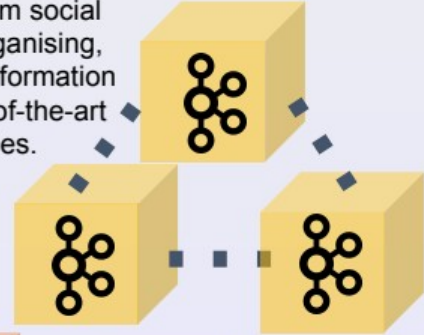


Harvesting news-related information from social media and other sources; analysing, organising, enriching and presenting news-related information to journalists. Implemented using state-of-the-art big data and distributed technologies.

Message queue nodes

Message exchange, queueing (TBD)

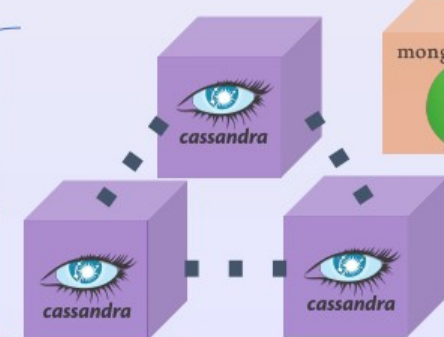
- Lighter processing
- Kafka



Raw data nodes

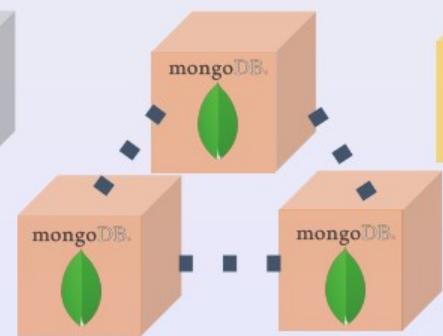
Distributed storage for raw data files (textual, multimedia)

- Disk intensive
- Cassandra, ...



Configuration nodes

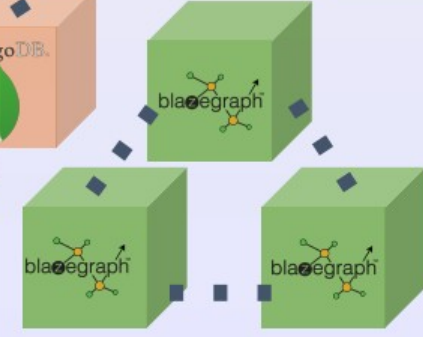
- Lighter processing
- MongoDB, files



Knowledge graph nodes

News semantic representation storage.

- Disk, CPU and RAM intensive
- Blazegraph

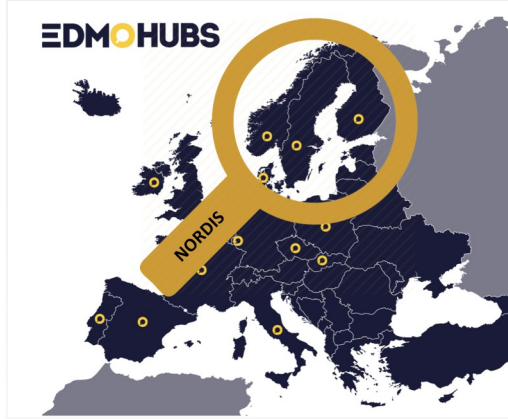


Infrastructure nodes

- Cloud-based, big-data ready
 - OpenStack, TerraForm, Ansible
- Deployed in **Docker Containers**:
 - running as Docker Swarm
- **Kafka**:
 - pub/sub message queue
 - stream database
- **Zookeeper**:
 - metadata for Kafka
- **MongoDB**:
 - configuration and metadata
 - IP addresses & ports
 - user names & passwords
- **Cassandra**:
 - original texts
 - JSON-LD metadata
- **Blazegraph**:
 - Knowledge Graph
 - scale-out
- **Python 3.8-3.9**:
 - specialised services
- **FastAPI**:
 - Python library for APIs



Context



**Media
Futures ●**

- **Augmented Journalism Network** – *Nordic seminar series.*
The Nordic research councils in the Humanities and Social Sciences (NOS-HS)
- **NORDIS** – *Nordic observatory for digital media and information disorders*
- **MediaFutures** – *Centre for Research-Driven Innovation of Responsible Media Technology*



Conclusion

- Technical side:
 - proof-of-concept *prototype*
 - state-of-the-art *infrastructure*
- Theoretical side:
 - *architecture* for journalistic knowledge platforms
 - understanding of *computational news angles*
- Further work:
 - implement scaling detection of angles
 - richer & more precise lifting
 - integration with Wolftech
 - other aspects: UI, visualization, terms-of-use, privacy, ...





Questions?