

Investigating the relationship between earthquakes and online news

Stephen Camilleri (1)
Joel Azzopardi (1)
Matthew R. Agius (2)

(1) Faculty of Information and Communication Technology, University of Malta,
(2) Faculty of Science, University of Malta
stephen.camilleri.03@um.edu.mt | joel.azzopardi@um.edu.mt | matthew.agius@um.edu.mt

System Inputs

1. Data Collection

- Continuous harvesting of data – minimum time gap between one cycle and another = every 30 minutes
- Earthquake readings from USGS are updated in database
- Differences in implementations to retrieve news articles (endpoints or direct from news agency site). Duplicate news articles are discarded



Multilingual news articles retrieved from 23 news agencies** worldwide



Earthquake readings retrieved from USGS in real-time

2. Content Processing

- Discarding unwanted text – e.g. author name, date and location at the beginning of the news content; concluding remarks; embedded Javascript
- Trim content (if need be)
- Identifying the language and translating of content (if need be)

3. Content Filtering

- Identifying which news articles mention the word 'quake'

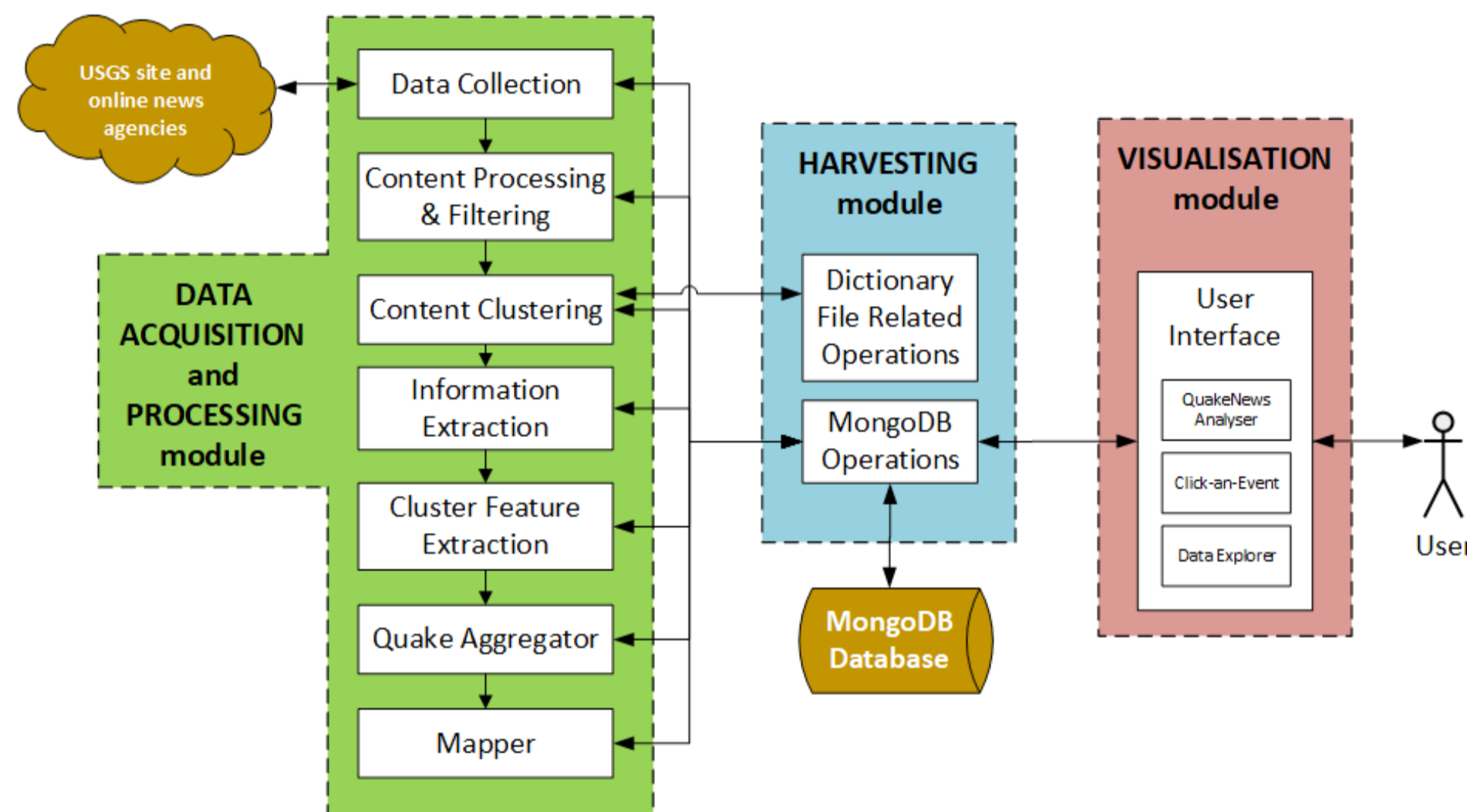
4. Content Clustering

- Sentence segmentation (split content into sentences) if news article mentions the word 'quake'
- Remove stopwords (common words) e.g. 'a', 'the' etc.
- Stemmatisation (reduce the inflectional forms from each word to root)
- Bag-of-words of each news article (count frequency of words in each article)
- Identify sentences with earthquake-related terms in each news article
- Term-by-term matrix of each news article
- Document similarity between news article and clusters by comparing bag-of-words

5. Feature Extraction

- Date Resolution (identifying and converting dates to a particular format)
- Tokenisation (splits a sentence into words)
- Part of speech tagging (identify nouns, verbs, adjectives etc.)
- Location Resolution verified by GeoNames
- Extract magnitude, number of casualties, injured and quantifiable structural damage

System Architecture



Observations

- **Most mentioned event**
Papua New Guinea 7.5-magnitude earthquake on 25th February 2018 – 50 articles over a span of 48 days
- **Most reported event**
Taiwan's 6.4-magnitude earthquake on 6th February 2018 – 47 articles over a span of 5 days

Contributions

- Quicker & more accurate analysis of the possible correlations between earthquake events and news articles speaking about these events
- Built news corpus (253,129 articles) that can be used to explore other areas of interest
- Provided practical solutions on how to exploit natural language processing tools to extract earthquake features
- Library for extracting temporal expressions

6. Cluster Feature Extraction
Identifying a common magnitude, date (or range of dates) and geographical location/s

7. Quake Aggregator
Aggregating daily earthquake readings into one event

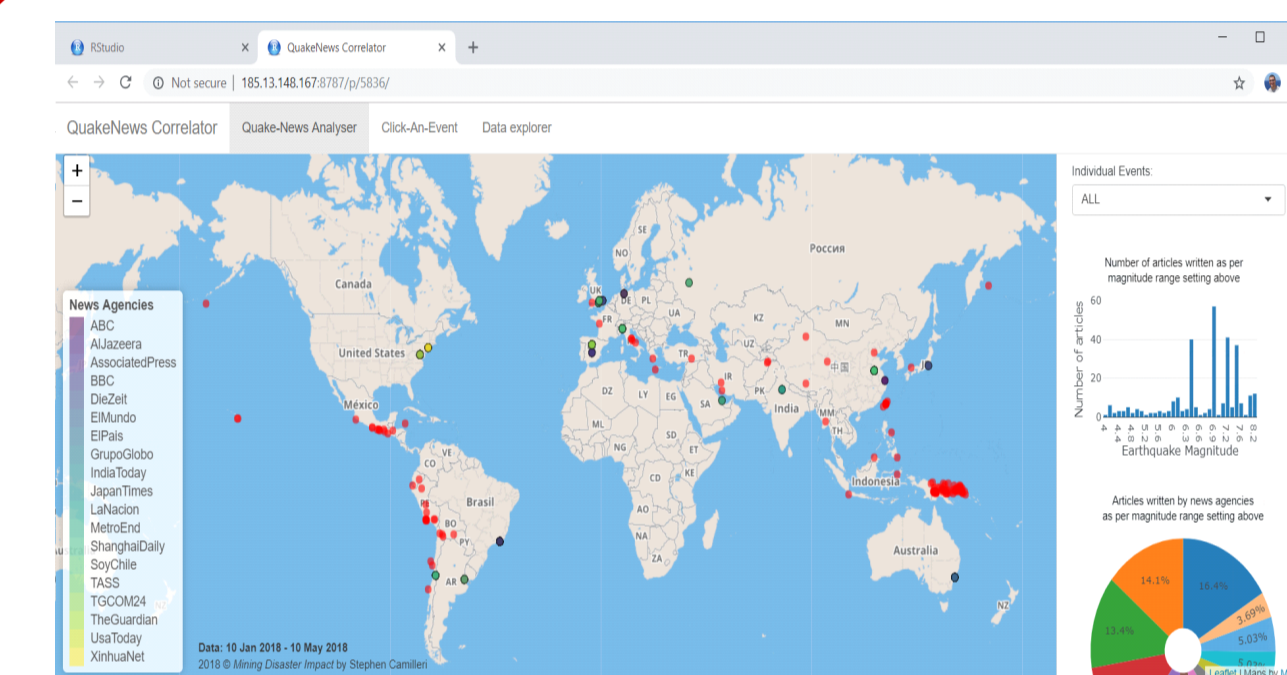
8. Mapper
Mapping clusters with earthquake events. An earthquake event can only be mapped against the most appropriate cluster (using location scores). A cluster can be mapped against one / multiple events.

9. Visualiser
Retrieves the earthquake events and news agency sites, which published articles about those events and displays them in the specific dashboards + makes them available for searching

Aim

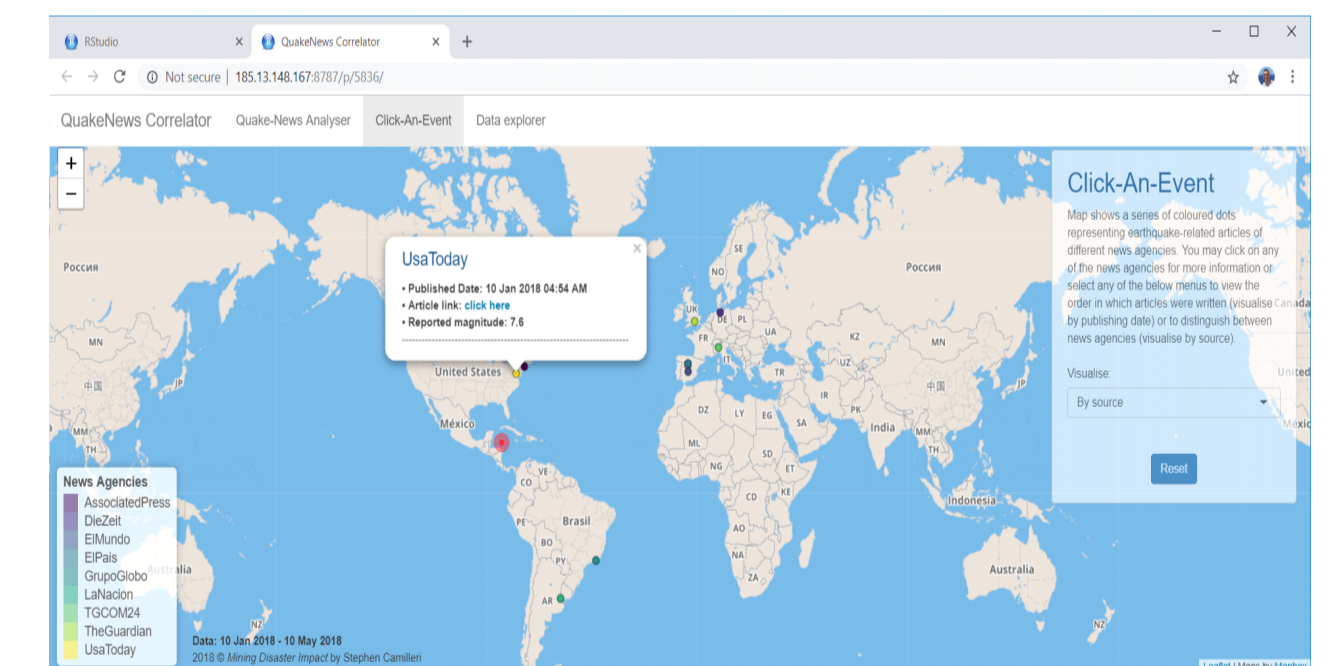
To automatically map in near real-time news articles with earthquake events and to visualise the results via dashboards

System Outputs



QuakeNews Analyser Dashboard provides at a glance view of how news agencies react to earthquakes of different magnitudes

Click-An-Event Dashboard provides analysis of which news agencies reacted to a specific earthquake across time



Data Explorer gives the user the possibility to search for articles (with filtering) and retrieves content of article and other information

Date	Headline	Source	View
10 Jan 2018	7.2 magnitude earthquake in the Caribbean Sea; tsunami alert begins	TGCOM24 International	View
10 Jan 2018	Strong earthquake of 7.6 Richter registered in Honduras	SoyChile International	View
10 Jan 2018	Caribbean hit by magnitude 7.8 earthquake, tsunami warning in place	ABC local	View
10 Jan 2018	Caribbean hit by magnitude 7.8 earthquake, tsunami warning in place	ABC International	View
10 Jan 2018	Strong earthquake hits Honduras and causes tsunami alert	GrupoGlobo International	View
10 Jan 2018	Tsunami warning in the Caribbean Sea after a powerful earthquake of magnitude 7.8	ElMundo	View

** The news agencies are: ABC (Australia - local and international); Al Jazeera (Qatar); Associated Press (USA); BBC (UK); China Daily (China - local and international); Die Zeit (Germany); El Mundo (Spain); El Pais (Spain - local and international); Grupo Globo (Brazil - local and international); India Today (India - local and international); Japan Times (Japan); La Nacion (Argentina - local and international); Metro End (UK); News24 (South Africa - local and international); Reuters (UK); Shanghai Daily (China - local and international); Soy Chile (Chile - local and international); Sudan Tribune (Sudan); TASS (Russia); TGCOM24 (Italy - local and international); The Guardian (UK); USA Today (USA); Xinhua Net (China).