

**REAL-TIME TIMELINE  
SUMMARISATION FOR  
HIGH-IMPACT EVENTS  
IN TWITTER**

**YIWEI ZHOU, ALEXANDRA CRISTEA, UNIVERSITY OF WARWICK, UK  
NATTIYA KANHABUA, AALBORG UNIVERSITY, DENMARK**

# USER-GENERATED TIMELINE OF "EBOLA OUTBREAK"

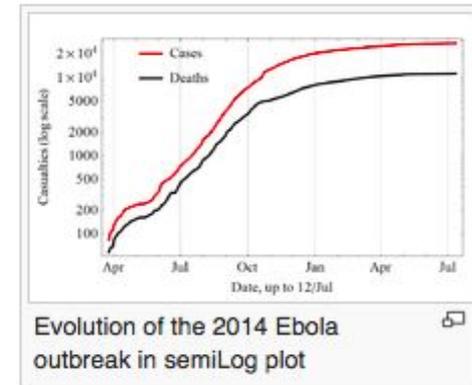
## March 2014 [ edit ]

 **Guinea** On March 18 Guinean health officials announce the outbreak of a mysterious hemorrhagic fever "which strikes like lightning." 35 cases are reported, and at least 23 people died.<sup>[11]</sup> On March 22, Guinea confirms the fever as Ebola, with 59 possible fatalities.<sup>[12]</sup>

 **Liberia** On March 24, two suspected cases in Liberia are announced by the Liberian Ministries of Information, Culture, Tourism, and Health. The government had also stated that Ebola had 'crossed over into Liberia,' but did not confirm the information.<sup>[13]</sup>

 **Médecins Sans Frontières** On March 24, [Médecins Sans Frontières / Doctors Without Borders](#) (MSF) establish an isolation center in Guéckédou Prefecture, the first in West Africa to be built expressly for the purpose of halting the epidemic.<sup>[14]</sup>

 **United Nations** On March 25, the [World Health Organization](#) (WHO) released its first report concerning Ebola in West Africa, wherein it is stated that thirteen Guinean cases from four districts of the country were confirmed by the [Institut Pasteur](#) and the Centre International de Recherche en Infectiologie in Lyon.<sup>[15]</sup>



**OBJECTIVE: AUTOMATIC GENERATION OF TIMELINE SUMMARIES IN REAL-TIME FOR HIGH-IMPACT EVENTS**

A large, textured red circle with a white dot in the center. The circle has a watercolor-like appearance with some darker and lighter shades of red. The background is a light, textured grey.

1.

# **METHODOLOGY**

## DEFINITION

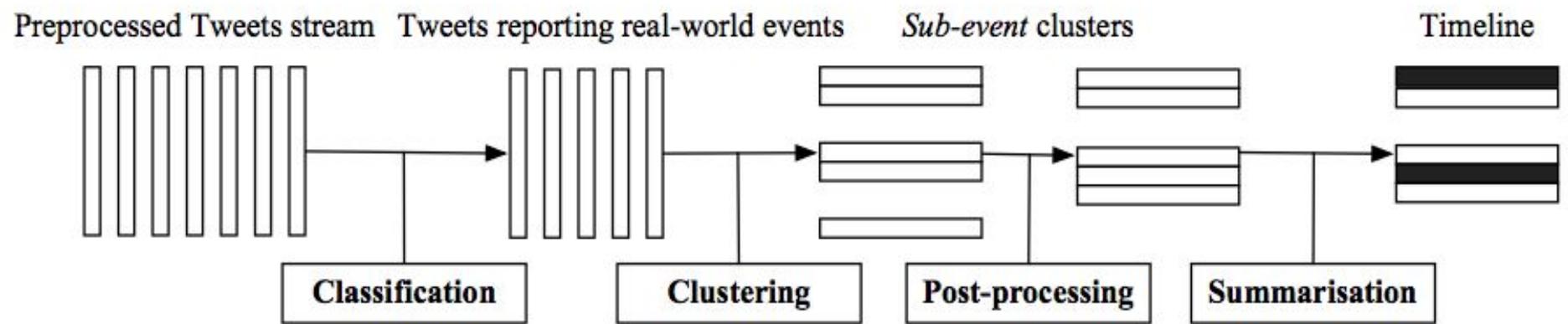
**Major event:** **high impact** events with **worldwide influence**: "Gulf of Mexico oil spill", "Ebola outbreak", "Zika virus outbreak".

**Sub-events:** key incidents showing the status of the ongoing major event; earn **compatible attention**, but can **only last for a few days, or even shorter**.

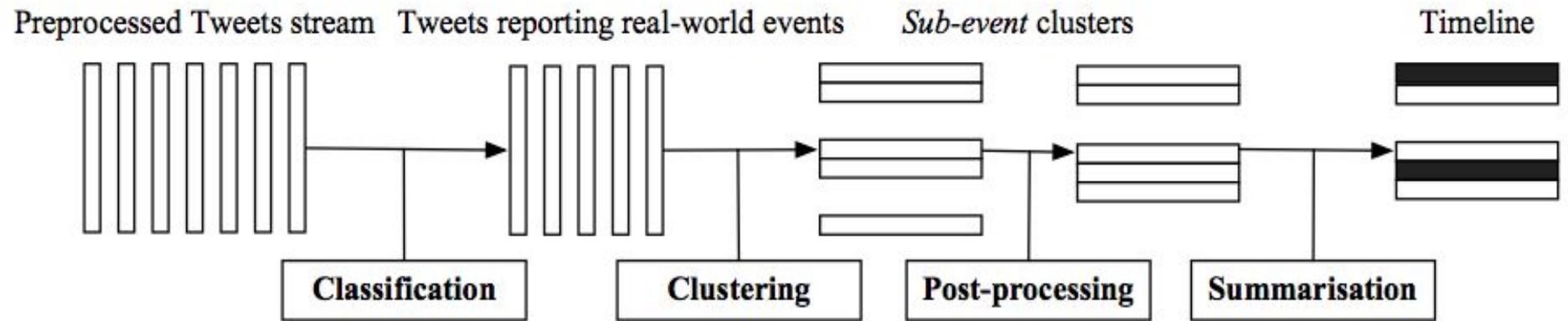
E.g., "On March 24, two suspected cases in Liberia are announced by the Liberian Ministries of Information, Culture, Tourism, and Health. The government had also stated that Ebola had 'crossed over into Liberia,' but did not confirm the information." is a sub-event of the major event "Ebola outbreak".

**Timeline:** the **temporal development** of a major event; i.e., all the sub-events relevant to the major event, listed in chronological order.

# PIPELINE

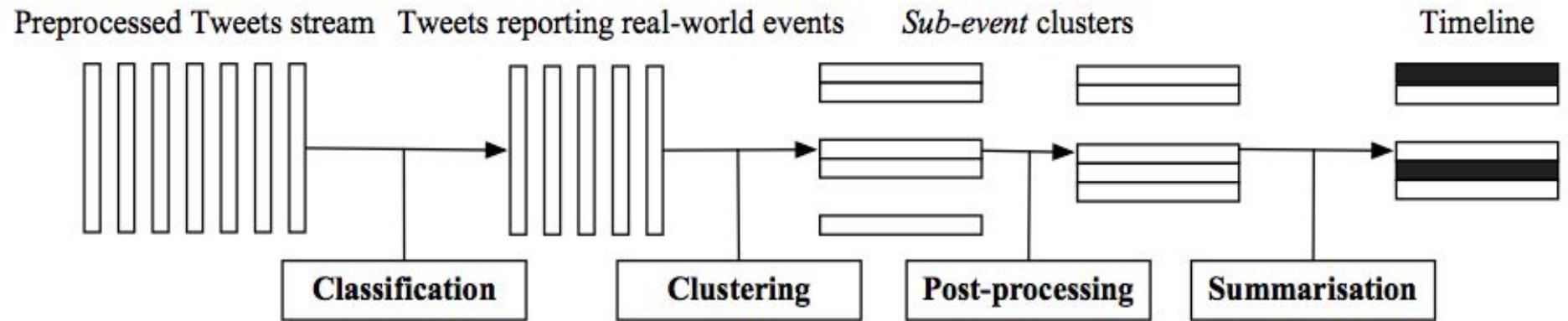


# PIPELINE



**Preprocessing:** entity-based filtering + POS tagging

# PIPELINE

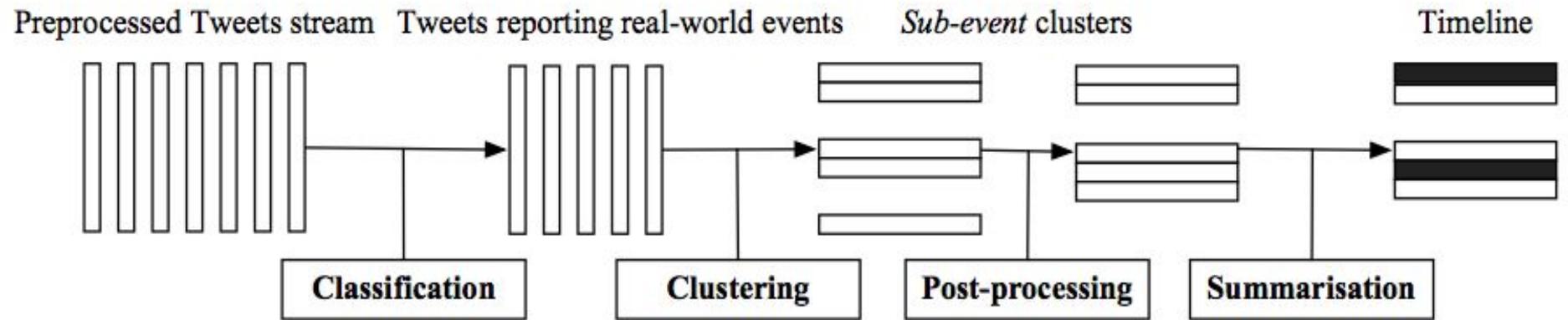


Extraction of tweets reporting real-world events

## EXTACTION OF TWEETS REPORTING REAL-WORLD EVENTS

- ▶ Filter out tweets expressing personal feelings, or pointless “babblers”.
- ▶ Employ the differences in **expression patterns**.
- ▶ Only **event-independent** features are included:
  - Twitter syntax features: hashtags, at-mentions;
  - other users' reactions: retweets, favourites;
  - informality: emoticons, abbreviations;
  - strong feelings: interjections, exclamation and question marks;
  - fact-related information: numbers, URLs, locations.

# PIPELINE



**Sub-event detection in tweets stream**

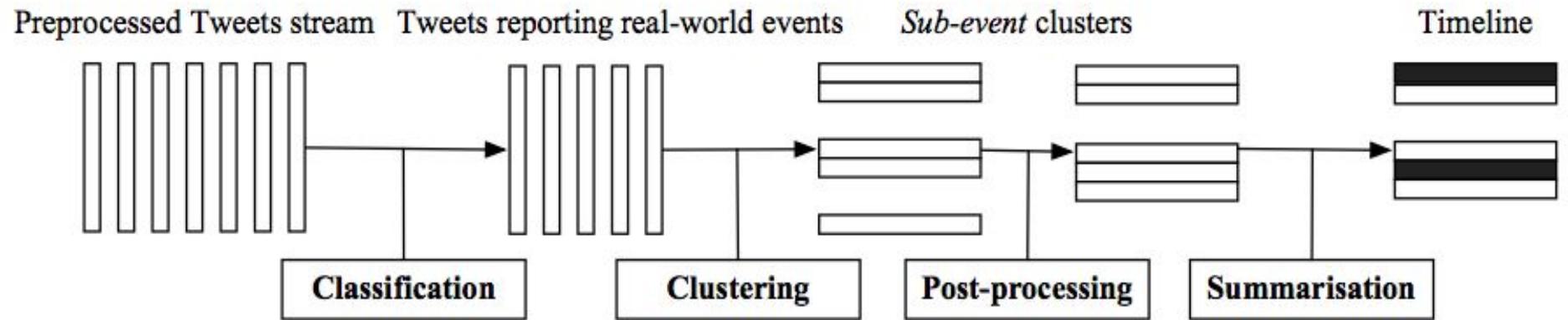
## SUB-EVENT DETECTION IN TWEETS STREAM

- ▶ Representation based on **key terms**: noun phrases, verbs in original forms, hashtags, URLs, numbers and at-mentions.
- ▶ **Prioritised URL-based clustering strategy**: cluster tweets contain common **useful** URL(s).
- ▶ **Threshold-based clustering strategy**:

$$\text{GetSimilarity}(c_1, c_2) = J_{JW}(K_1, K_2) = \frac{K'_1 \cap K'_2}{K'_1 \cup K'_2}$$

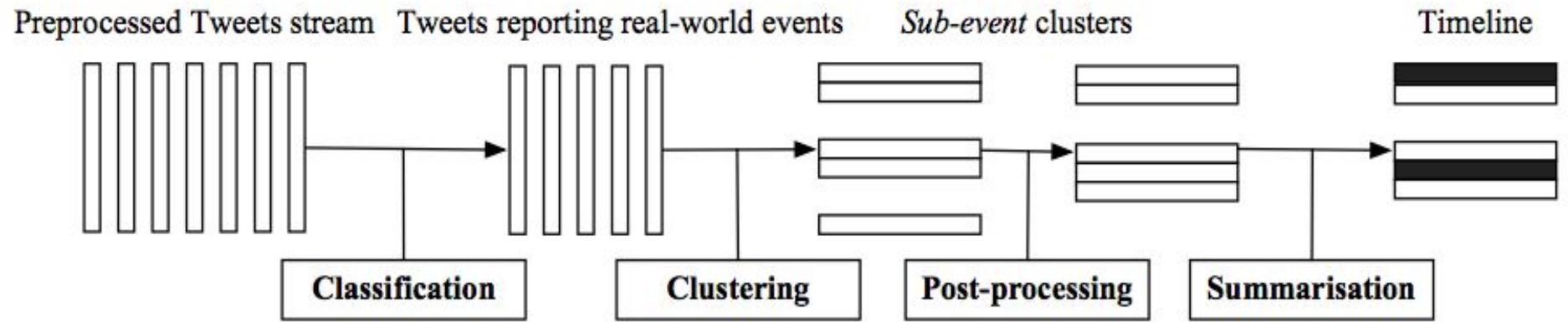
- Jaccard similarity metric based on **Jaro-Winkler matching**;
- $K'$ : sets of key terms sharing the same meaning.

# PIPELINE



Post-processing of detected sub-events

# PIPELINE



**Timeline Summarisation**

## POST-PROCESSING + TIMELINE SUMMARISATION

- ▶ A more rigid, computationally-intensive **hierarchical clustering** step, processed **offline**, to deal with the miss-outs of the online clustering.
- ▶ Consider **temporal feature** of sub-events: one sub-event will be seen as **terminated**, if there is no update for  $M$  days.
- ▶ Select the **most representative** tweet in one sub-event cluster:
  - **textual feature**: contains the most information
  - **temporal feature**: contains the newest updates (posted timestep+extracted from tweet)

2

**EVALUATION ON EBOLA  
TWEETS DATASET  
(90,823 TWEETS, PROCESSED  
IN POSTED TIME ORDER)**

# REAL-WORLD EVENTS REPORTING TWEETS EXTRACTION

	<i>OTHER</i>	<i>REAL-WORLD</i>	<i>AVERAGE</i>
Precision	<b>0.828</b>	<b>0.761</b>	<b>0.795</b>
Recall	<b>0.730</b>	<b>0.850</b>	<b>0.790</b>
F1 score	<b>0.799</b>	<b>0.805</b>	<b>0.792</b>

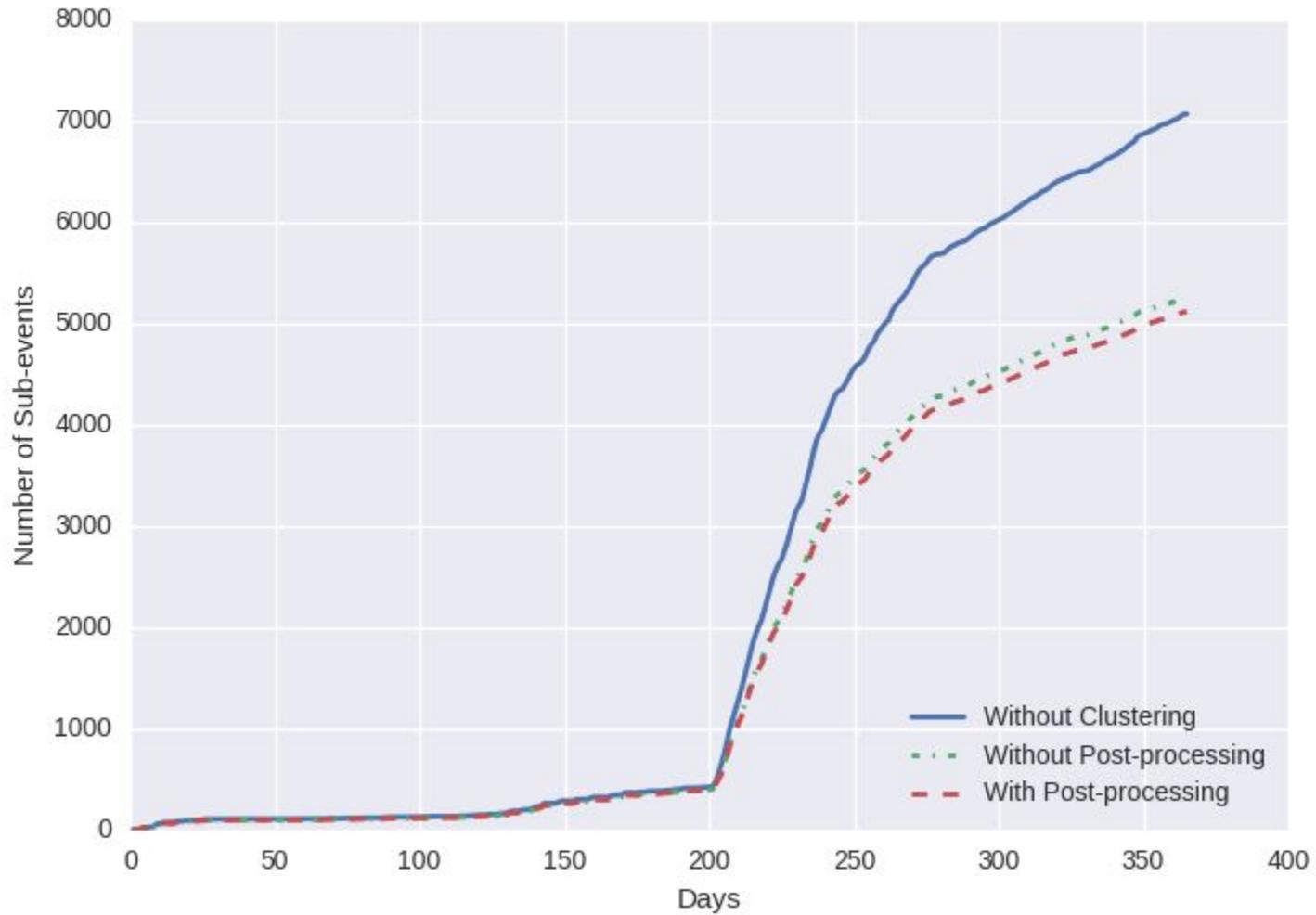
# SUB-EVENT DETECTION

CT	Cosine-TFIDF		Jaccard		Without Post-processing		With Post-processing	
	CP	CR	CP	CR	CP	CR	CP	CR
0.3	84.0%	94.0%	67.0%	65.7%	80.0%	60.9%	77.9%	59.2%
0.4	94.4%	97.1%	80.0%	74.6%	85.9%	70.4%	83.9%	68.3%
0.5	97.8%	98.6%	85.0%	78.1%	90.8%	74.4%	90.0%	72.4%
0.6	100.0%	99.2%	88.8%	81.5%	93.0%	76.4%	92.0%	75.6%
0.7	100.0%	99.3%	92.9%	85.1%	94.9%	77.9%	93.5%	77.6%
0.8	100.0%	99.4%	93.9%	90.1%	95.4%	79.9%	95.4%	79.8%
0.9	100.0%	99.5%	100.0%	99.2%	95.9%	80.8%	95.6%	80.8%

**CP**: clustering precision, the percentage of **positive clusters** in all the generated clusters that contain more than one tweet;

**CR**: compression ratio, number of detected sub-events divided by the total number of tweets.

# SUB-EVENT DETECTION



## SUB-EVENT SUMMARISATION

Date	Location	Timeline
2014.03.23	Guinea	No serious med infrastructure in the area for response. "Guinea confirms Ebola as source of epidemic <a href="http://aje.me/1gU7kpU">http://aje.me/1gU7kpU</a> via @AjEnglish"
2014.03.24	Senegal, Liberia, Guinea, Sierra Leone	#Senegal & #Liberia mobilise medics to ward off #Ebola spreading in #Guinea. #SierraLeone much closer to epicenter doing/saying nothing.
2014.03.24	Sierra Leone	#EbolaFever has hit eastern Sierra Leone. Fast action needed please madam #MinisterofHealthandSanitation. This is very serious. God help us.
2014.03.26	Guinea	#Guinea says it has contained #Ebola outbreak in its southeast, but death toll rises and people are scared Reuters <a href="http://in.reuters.com/article/2014/03/26/guinea-ebola-idINL5N0MN50D20140326...">http://in.reuters.com/article/2014/03/26/guinea-ebola-idINL5N0MN50D20140326...</a>
2014.03.26	Guinea	@WHO does not recommend any travel, trade restrictions to #Guinea & neighbouring countries in respect to this #Ebola outbreak #AskEbola
2014.03.26	Guinea, Liberia	#Ebola virus kills 90% of those it strikes - 63 people have died so far in #Guinea in latest outbreak, 5 in #Liberia <a href="http://www.bloomberg.com/news/2014-03-25/ebola-victims-face-90-death-risk-drugs-start-to-emerge.html...">http://www.bloomberg.com/news/2014-03-25/ebola-victims-face-90-death-risk-drugs-start-to-emerge.html...</a>
2015.03.11	UK, Sierra Leone	And now a UK Military Health Worker battling #Ebola in #sierraleone <a href="http://edition.cnn.com/2015/03/11/europe/uk-military-ebola/index.html">http://edition.cnn.com/2015/03/11/europe/uk-military-ebola/index.html</a>
2015.03.13	Liberia	WHO Confirms No Ebola Case in Liberia in Two Weeks - <a href="http://AllAfrica.com">http://AllAfrica.com</a> <a href="http://goo.gl/fb/aDB55B">http://goo.gl/fb/aDB55B</a> #LIBERIA
2015.03.20	Liberia	#Liberia confirms first #Ebola case in weeks, just as the authorities were beginning the countdown to an Ebola-free nation.
2015.03.20	Sierra Leone	Good news! No confirmed cases of Ebola recorded by the Sierra Leone government in their 20 March daily report. <a href="http://reliefweb.int/report/sierra-leone/ebola-outbreak-updates-march-20-2015...">http://reliefweb.int/report/sierra-leone/ebola-outbreak-updates-march-20-2015...</a>

201 ITEMS IN WIKIPEDIA TIMELINE, 126 CAN FIND THEIR  
CORRESPONDING ITEMS;  
WHO ISSUED FIRST WARNING ON 25TH MARCH, OUR  
FIRST DETECTED EVENT WAS ON 23RD MARCH.

# CONCLUSION

A GENERIC APPROACH TO GENERATE  
REAL-TIME TIMELINE FOR HIGH-IMPACT  
EVENTS;

A PROMISING SUPPLEMENT OR  
REPLACEMENT OF USER-GENERATED  
TIMELINE;

EARLY ALARMS FOR DISASTER SURVEILLANCE.

**THANKS!**



**any  
questions  
?**