



Public Health
Agency of Canada

Agence de la santé
publique du Canada

Canada

Developing information extraction algorithms for open-source data from event-based surveillance systems

2019 EIOS Global Technical Meeting
Day 2: November 13, 2019

PROTECTING AND EMPOWERING CANADIANS
TO IMPROVE THEIR HEALTH





Erin E. Rees



Victoria Ng

1. AI and IT enhancements

Refine and develop ML algorithms



Digitize data sources



Real-time data capture



3. Advanced analytics



2. Data integration

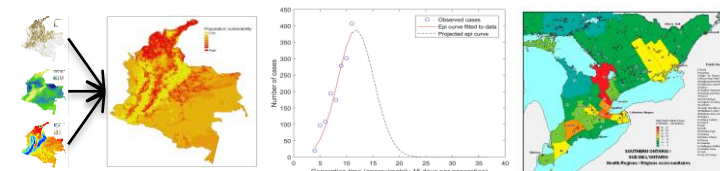
Social media and search engine data



Human mobility and other data



4. Visualisation of outputs



Automated to produce real-time outputs



GOAL: Less time screening for relevancy and more time for threat assessment

Refine and develop ML algorithms


- Algorithms in GPHIN and EIOS systems:


Pakistan 🌐
Khyber Pakhtunkhwa
Nooriabad
Sindh
Jamshoro
Karachi District
Karachi
New Karachi

Information

Unique ID:	1005856132
Source document:	ASDAWN0020191112efbc00010_FT.xml
News provider:	Factiva
Publication date:	2019-11-12 11:37 UTC
Received date:	2019-11-12 12:31 UTC
Publication name:	Dawn
Publication language:	EN
Relevance score:	0.991
Reviewed for relevance:	No
Importance:	Normal

Categories



 Infectious disease - human

 Infectious disease - animal

Keywords


crimean-congo haemorrhagic fever	case
dengue	confirmed
hemorrhagic fever	deaths









Fetch Date: **23/08/2019 09:09**
Import Date: **23/08/2019 11:28**

Source: **GPHIN**  gphin 















Subject: **Medical**

Language: (ru) **Russian**


Source Country :  Canada

Mentioned Countries:  Armenia 
 Egypt   Russia   United States 

Associated Categories:

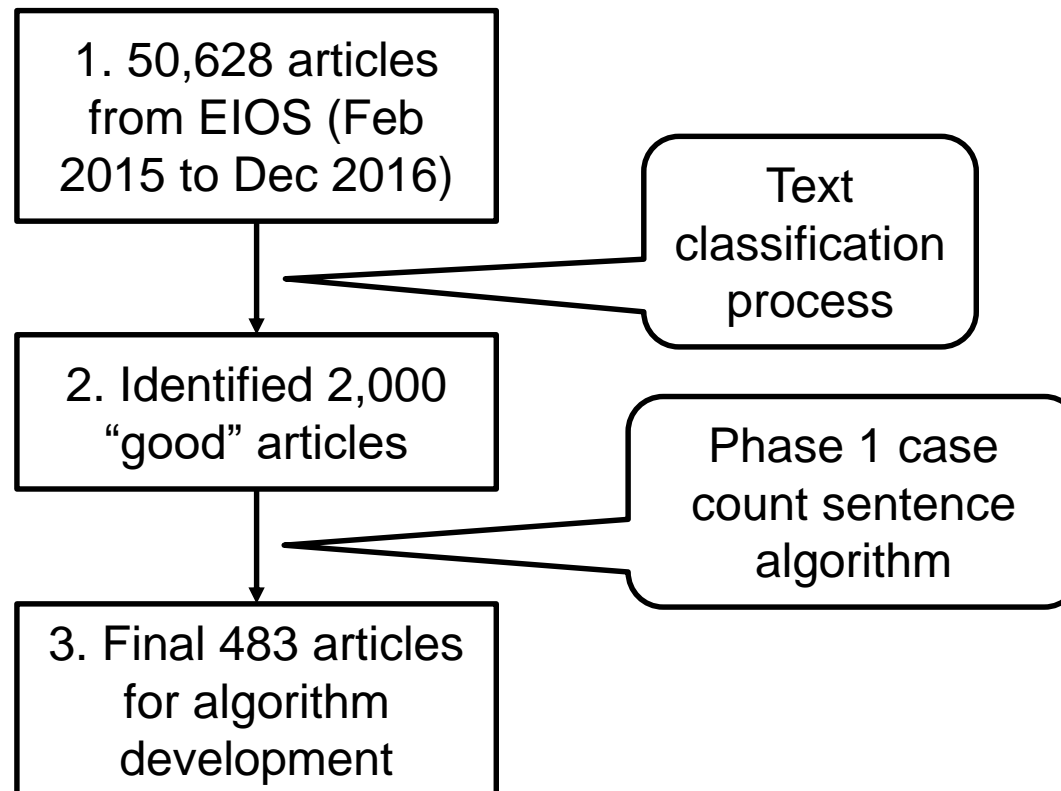
-  Chikungunya virus disease 
-  Communicable disease 
-  Dengue Fever   Epidemic 
-  FIFA2018-Priority Diseases-Partici... 
-  Yellow Fever   Zika virus 

EMM Id: gphin-
89b5aead6bec4f94a566eedef94daa69

 Read By:
Ng, Victoria (*Research InSIGHT Members*)
18:00

Candidate disease: Zika virus

1. Case count extraction algorithm
2. Temporal information extraction algorithm



1. Case count extraction



Phase 1:
developed for
norovirus

100 randomly selected article:

- 56 sentences identified
- 34 of 56 correct
- 4 false positives

Phase 2:
developing for
Zika virus



Annotate case
count sentences
(483 articles)

Current status



Dr. Adrian Verster



Algorithm
development
and evaluation



Health
Canada

Santé
Canada

The Zik-meister



zikatextannotator.herokuapp.com/index

Article-level classifications

Zika

Dengue

Ebola

GB

MERS

Swine Flu

Microcephaly

Research

General

Other

Other category name

Sentence-level classifications

Correct

Incorrect

General

Wrong Time

Country

State

Region

City

Other

X

Other sentence name

Please enter the dates as: YYYY-MM-DD_YYYY-MM-DD

If you are missing anything, such as a year, fill it in with Xs, e.g. XXXX-MM-DD_XXXX-MM-DD

Any other uncertainties should follow <http://nlp.stanford.edu>

ZikaX

Number: 3 Clause: (subject)

The Ministry (verb) confirmed

(object) three additional cases

The Ministry of Health has confirmed three additional cases of Zika virus infectionX

, bring to four the number of affected persons in Jamaica, to date.

Two of the infected persons are from Christian Pen and Greater Portmore, St. Catherine, while the other is from Lyssons, St. Thomas.

The first Zika case was confirmed in February. Doctors have linked the Zika virus to microcephaly which is the abnormal smallness of the head of infants which is associated with incomplete brain development.

Comments

Add any comments about this article or the sentences of interest

Previous

Next

Refresh

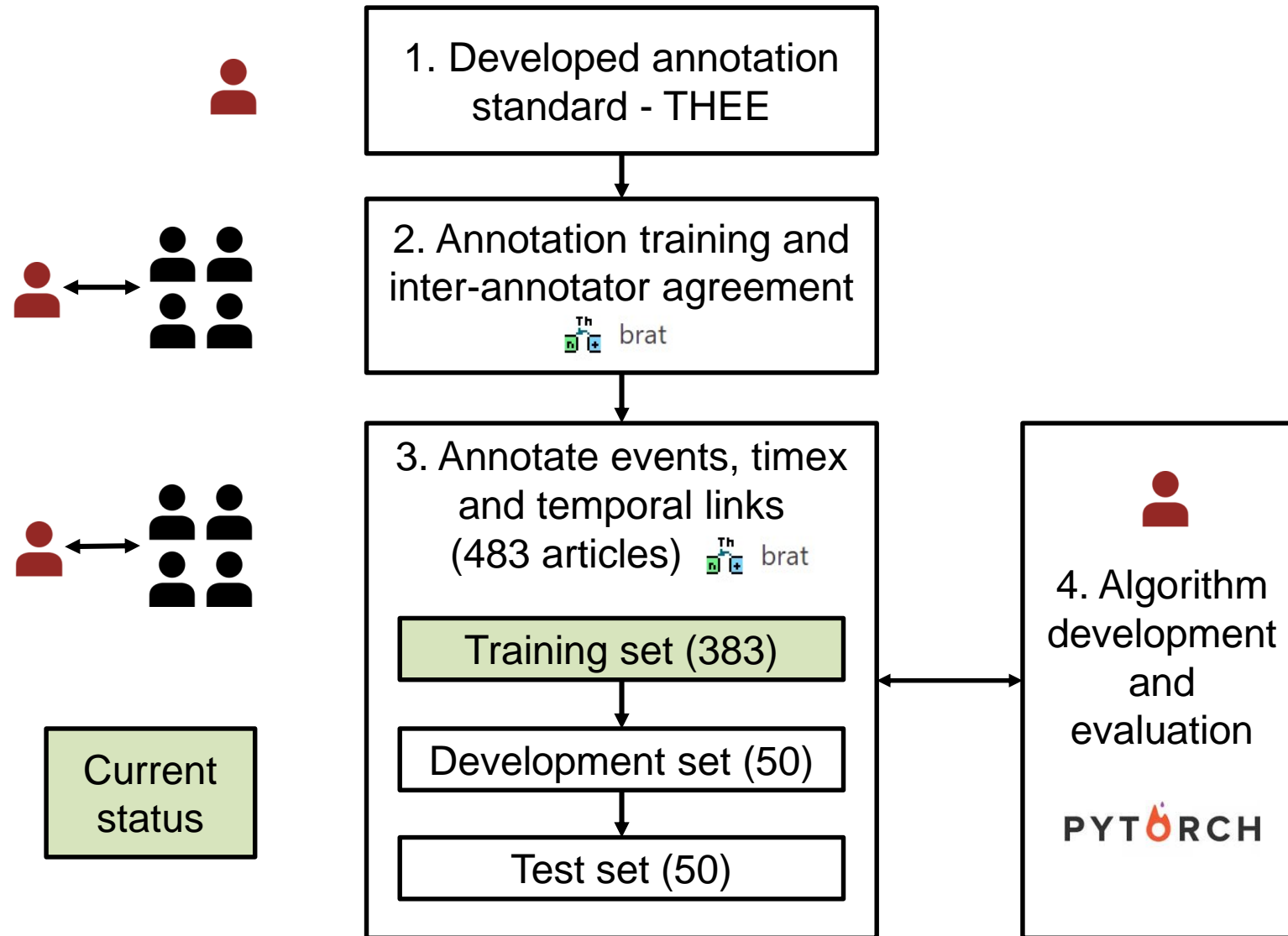
Logout

Download

Preliminary results

- Examples of true positive sentences:
 - “*Five new locally acquired cases* were announced Wednesday in Miami-Dade County”
 - “Officials in Connecticut on Thursday announced *two new cases* of the Zika virus”
 - “The announcement came as Australia confirmed its *first two cases* of the virus”
- Examples of true negative sentences:
 - “Approximately *one in five people infected* with the virus show symptoms”
 - “*Of California’s 77 cases*, the greatest numbers resulted from travel to El Salvador ..”
 - “There *have been at least four confirmed cases* of Louisiana residents who were diagnosed with the illness after traveling abroad”

2. Temporal information extraction

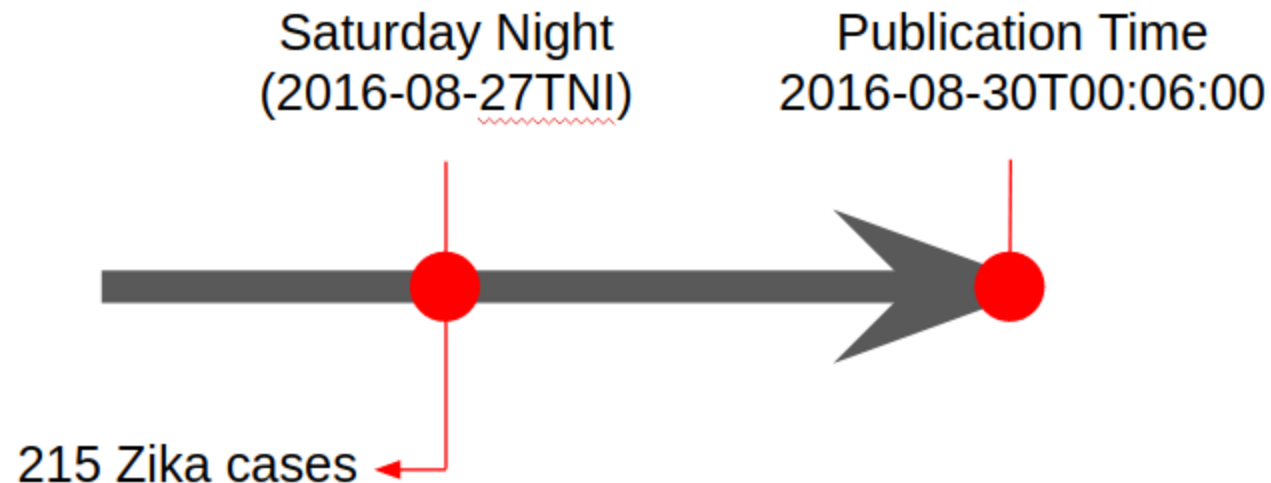
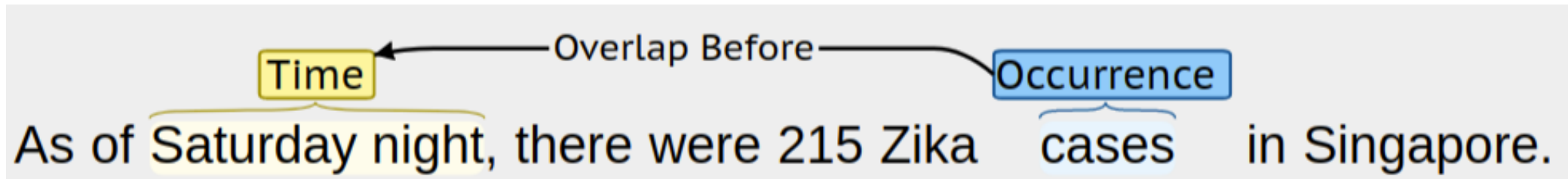


Jingcheng
(Frank) Niu

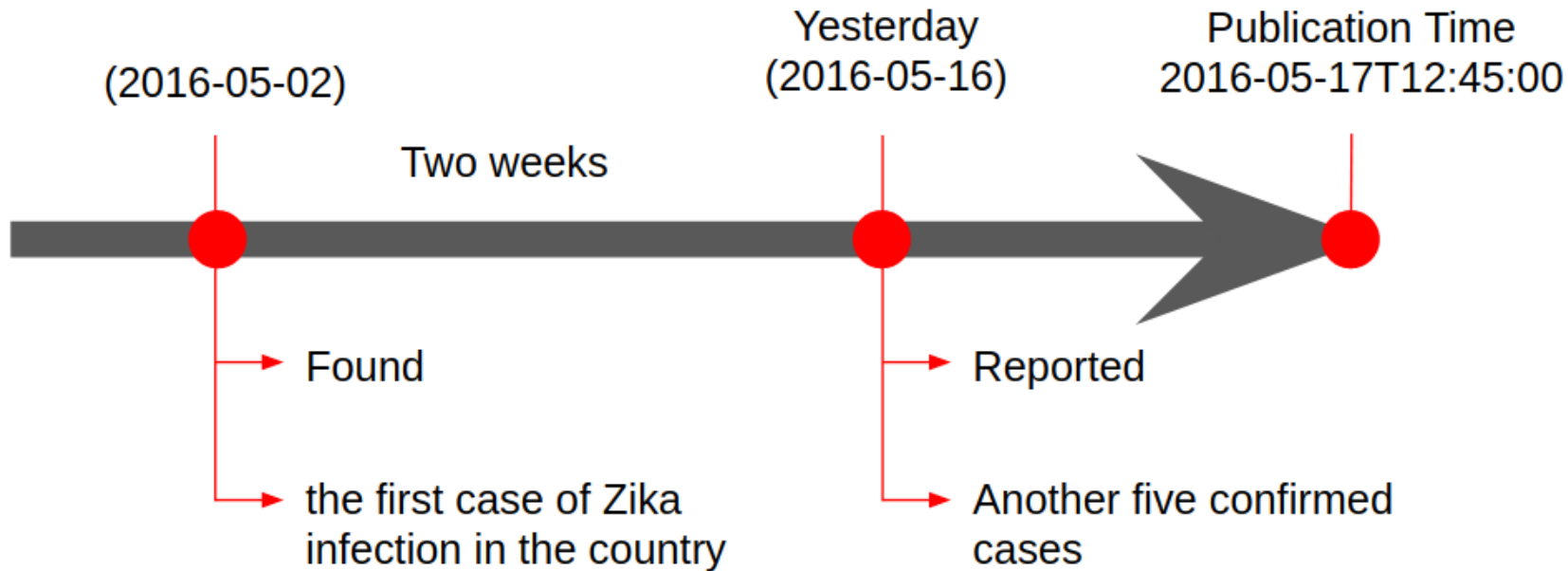
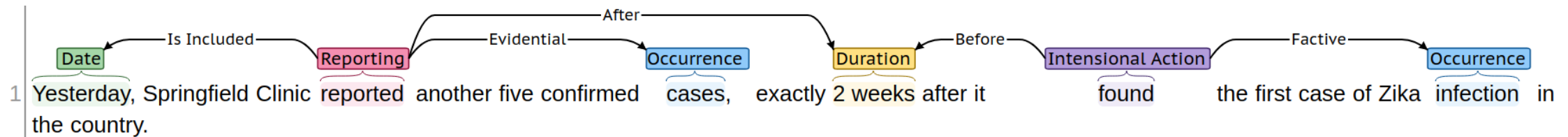


Dr. Gerald Penn

THEE annotation examples



THEE annotation examples



Preliminary results against TimeBank corpus

- TIMEX

- 'classification_f1': {'p': 90.95, 'r': 84.74, 'f1': 87.73}
- 'identification_f1': {'p': 90.95, 'r': 89.79, 'f1': 90.36}

Current state of the art:
p: 0.90, r: 0.85, f1:0.86

Source: Llorens et al (2013)

- EVENT

- 'classification_f1': {'p': 85.96, 'r': 55.43, 'f1': 67.4}
- 'identification_f1': {'p': 85.96, 'r': 61.69, 'f1': 71.83}

Current state of the art:
p: 0.81, r: 0.86, f1:0.83

Source: Llorens et al (2013)

$$\text{Precision} = \frac{\text{True Positive}}{\text{True positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True positive} + \text{False Negative}}$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Review

- Event
 - Reporting
 - Aspectual
 - IAction
 - IState
 - Occurrence
- TimeX
 - Date
 - Time
 - Duration
 - Set
- Temporal links
 - Before
 - After
 - Simultaneous
 - Is_Include
 - During
 - Identity
- Aspectual links
 - Initiation
 - Reinitiation
 - Termination
 - Culmination
 - Continuation
- SLINK:
 - Modal
 - Factive
 - Counter-Factive
 - Evidential
 - Negatively Evidential
 - Conditional

Questions and acknowledgements

