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

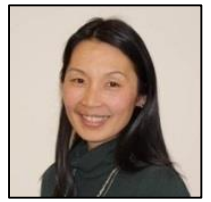
PROTECTING AND EMPOWERING CANADIANS TO IMPROVE THEIR HEALTH

InSIGHT

Internet-based Surveillance Informing Global Health Threats

Co-project Leads:

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

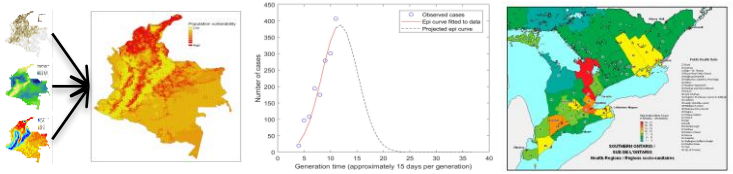
2. Data integration

Social media and search engine data

Google Trends

Human mobility and other data

4. Visualisation of outputs

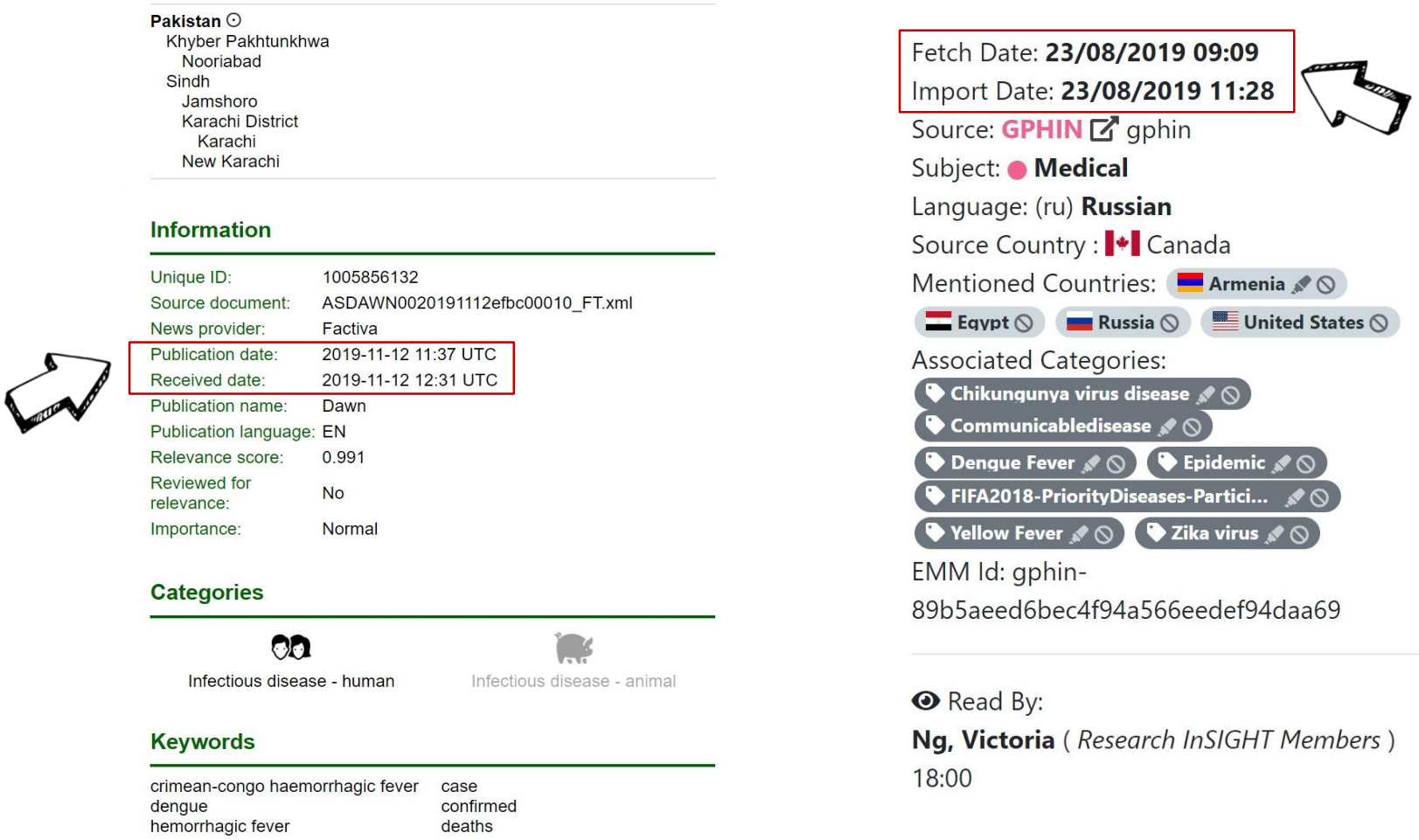


Automated to produce real-time outputs

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

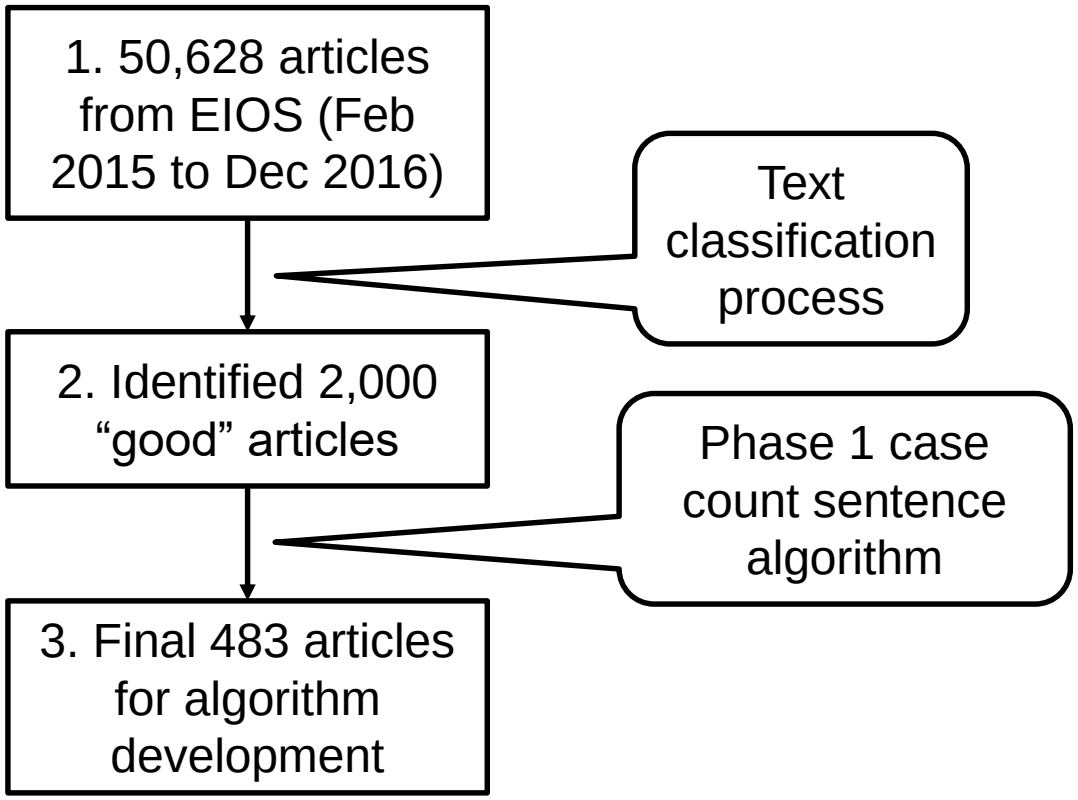
Refine and develop ML algorithms

- Algorithms in GPHIN and EIOS systems:



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)

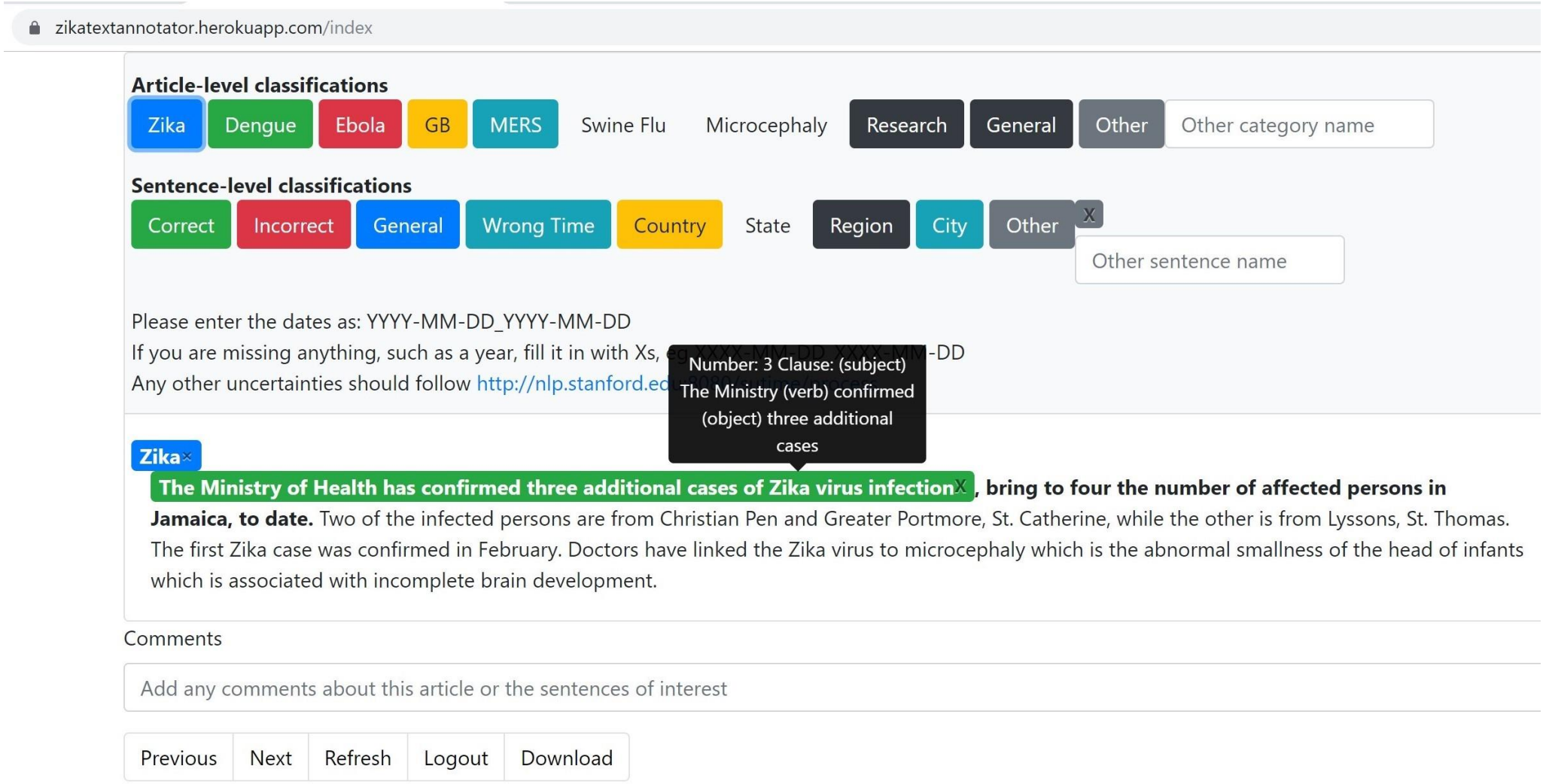
Algorithm development and evaluation

Current status

Dr. Adrian Verster

Health Canada / Santé Canada

The Zik-meister

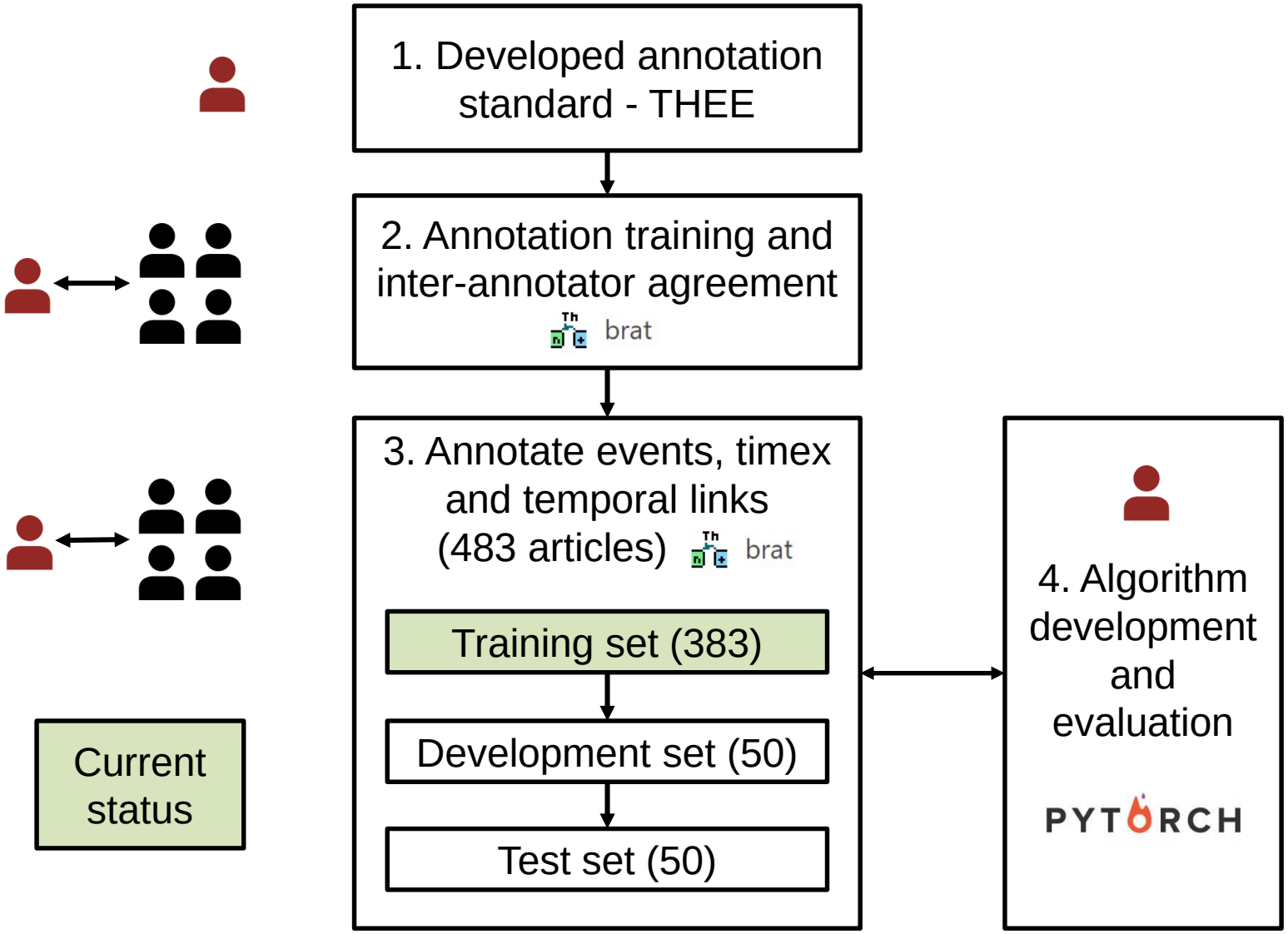


Preliminary results

- 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

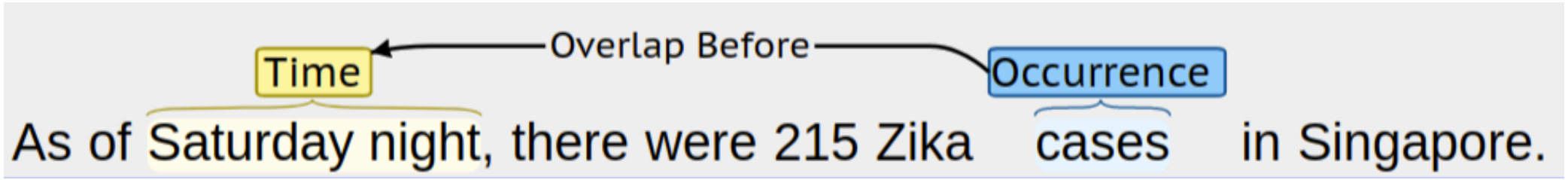


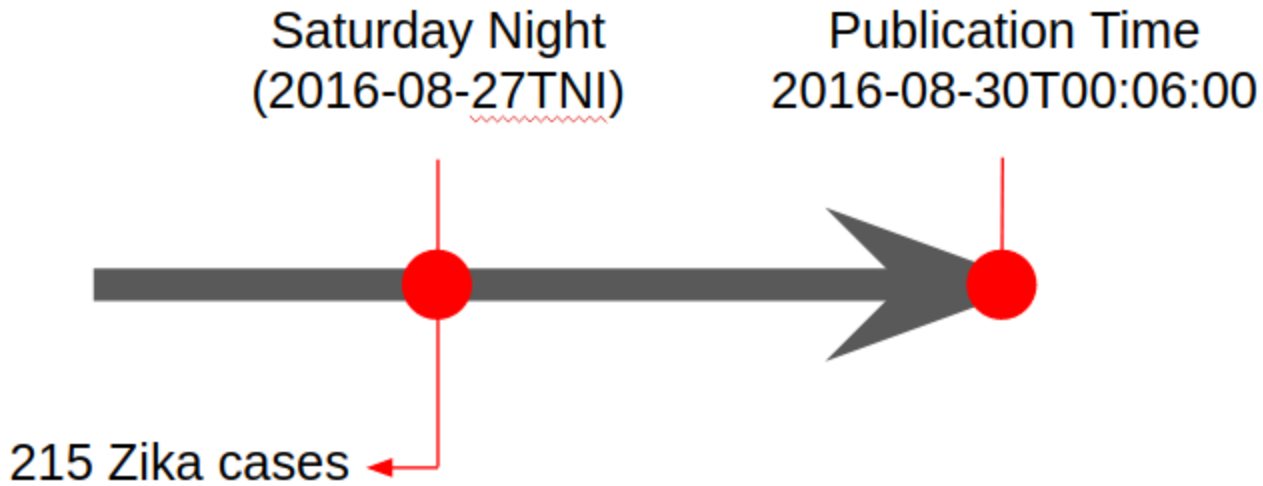
Computer Science
UNIVERSITY OF TORONTO

Jingcheng (Frank) Niu

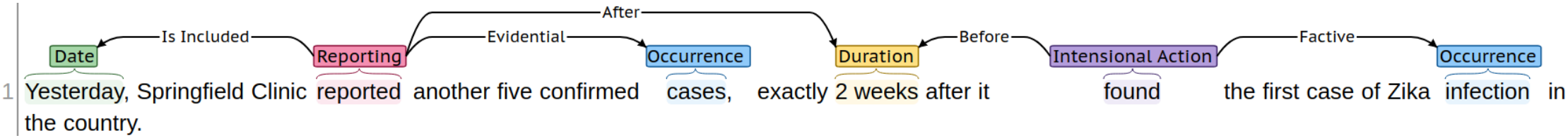
Dr. Gerald Penn

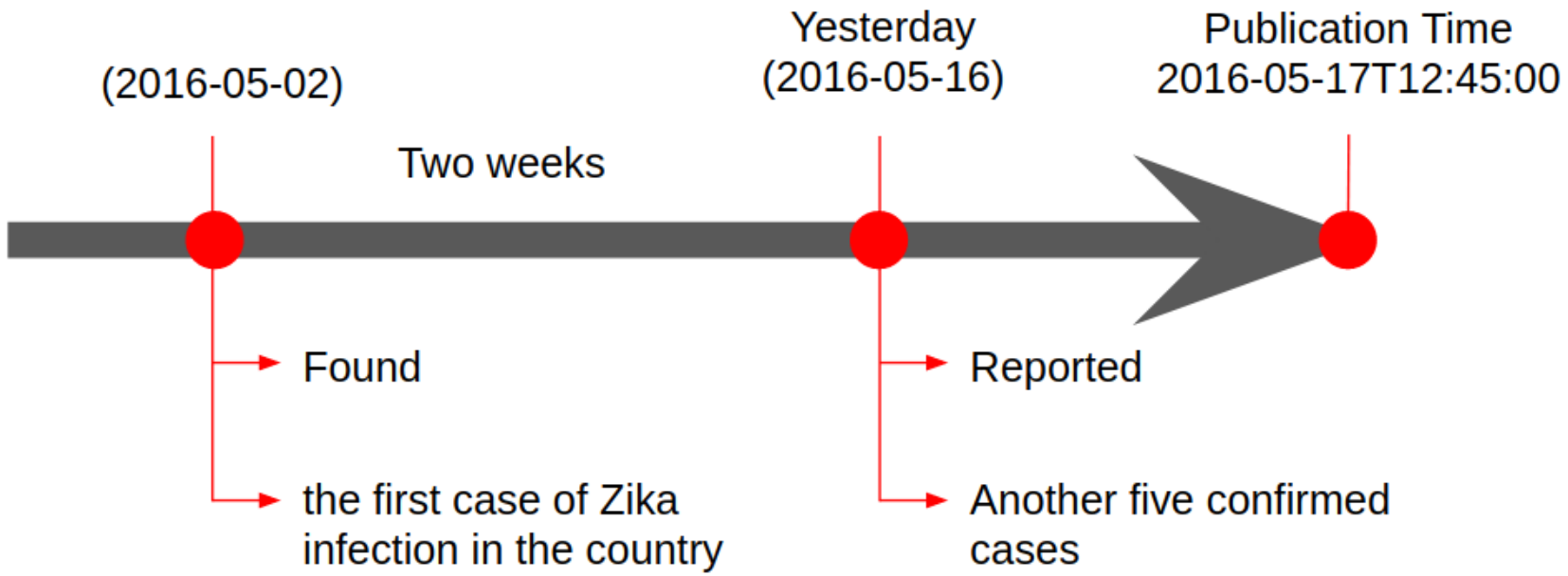
THEE annotation examples





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)

$$\text{Precision} = \frac{True\ Positive}{True\ positive + False\ Positive} \qquad \text{Recall} = \frac{True\ Positive}{True\ positive + False\ Negative} \qquad \text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

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

Government of Canada / Gouvernement du Canada
Canadian Safety and Security Program
Programme canadien pour la sûreté et la sécurité
CSSP is a federal program led by Defence Research and Development Canada's Centre for Security Science, in partnership with Public Safety Canada.
Le PCSS est un programme fédéral dirigé par le Centre des sciences pour la sécurité de Recherche et développement pour la défense Canada, en partenariat avec Sécurité publique Canada.
Canada

Computer Science
UNIVERSITY OF TORONTO

Health Canada
Santé Canada

EIOS
epidemic intelligence
from open sources

UNIVERSITY of GUELPH
CHANGING LIVES
IMPROVING LIFE

McGill
Department of
Epidemiology, Biostatistics
and Occupational Health

Natural Resources Canada
Ressources naturelles Canada
Canada

UNIVERSITÉ LAVAL

UQÀM
Université du Québec à Montréal