Natural Language Processing for Disease Surveillance

Mike Conway
Division of Biomedical Informatics
University of California, San Diego

mconway@ucsd.edu

Stockholm University Workshop - 21st June 2012

What have I been doing for the last four years?

- KR for phenotyping algorithms
- Building collaborative Knowledge Organisation System editing tools (WebProtege SKOS editor)
- Synonym detection (using statistical methods & semantic web technologies)
- **Natural Language Processing for Disease Surveillance**
  - 1. Outbreak detection using news text
  - 2. Developing resources to process chief complaints
  - 3. Using Twitter as a data source for public health surveillance

PART 1: MONITORING NEWS FOR DISEASE OUTBREAKS WITH BIOCASTER

How can NLP help in disease surveillance?

- Online news text is free (or at least relatively inexpensive)
- Online news text is accessible (without co-operation from governments)
- NLP allows us to process local languages
- NLP allows for <u>real-time</u> processing
- No consent issues for online text
- **NLP is not reliant on formal public health reporting mechanisms**

Biocaster system

- Based at the National Institute of Informatics, Tokyo
- Publicly accessible web-based interface
- Email alerting system
- Project began in 2006 (grant funded by JSPS & JST)
- Initially targeting Pacific Rim languages (currently, English, Japanese, Chinese & Vietnamese, but expanding to include Thai, Chinese, Korean, Malaysian & Indonesian)
- Uses a publicly available Knowledge Organisation System/Application Ontology (Biocaster Ontology)
- Collaborators:
  - Okayama University (Japan)
  - National Institute of Genetics (Japan)
  - Kasetsart University (Thailand)
  - National Institute of Infectious Diseases (Japan)
  - Vietnam National University
- PI: Dr Nigel Collier

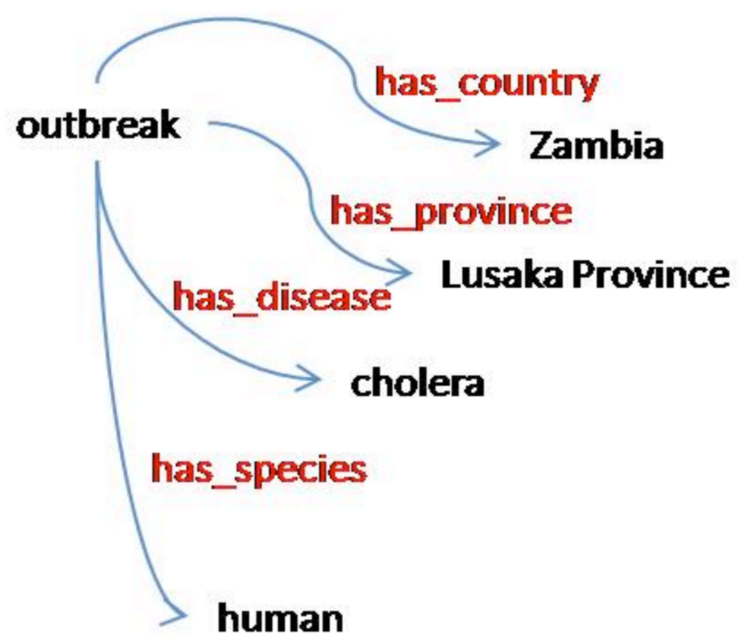
Simplified Example

<HTML> <head> <meta...><script...> </head><body>< p> Lusaka sufre la peor epidemia de cólera en más de diez años con 120 muertos</p><p> Pese a la esperanza de que la epidemia remitiera, las fuertes lluvias, que han ocasionado inundaciones en la capital zambia, podrían incluso empeorar la situación en las próximas semanas, dice MSF en su nota. </p></body><html>

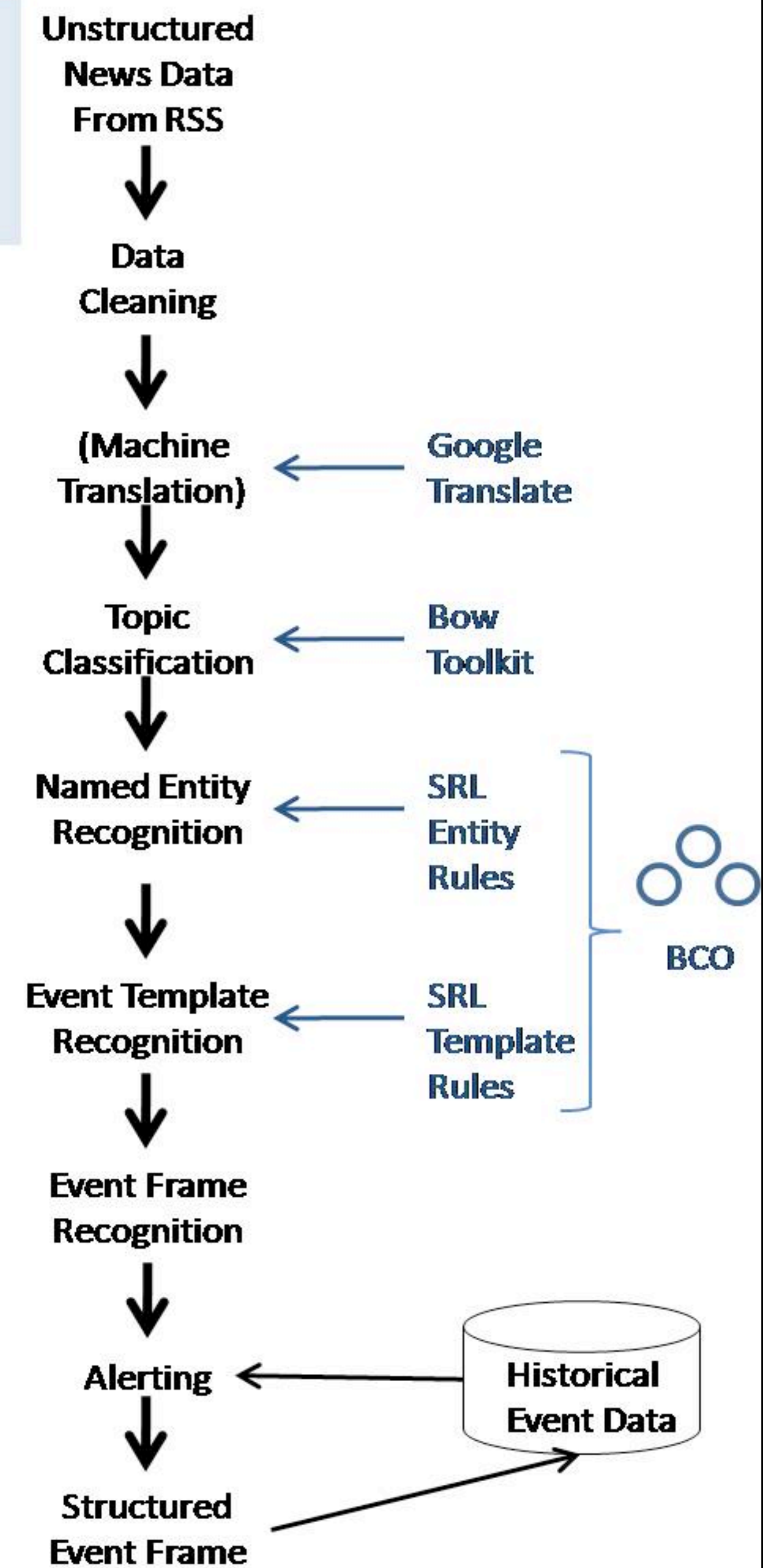
Lusaka suffered the worst cholera epidemic in more than ten years with 120 deaths. Despite the hope that the epidemic submit, heavy rains which have caused flooding in the Zambian capital, could even worsen the situation in the coming weeks, MSF said in his note.

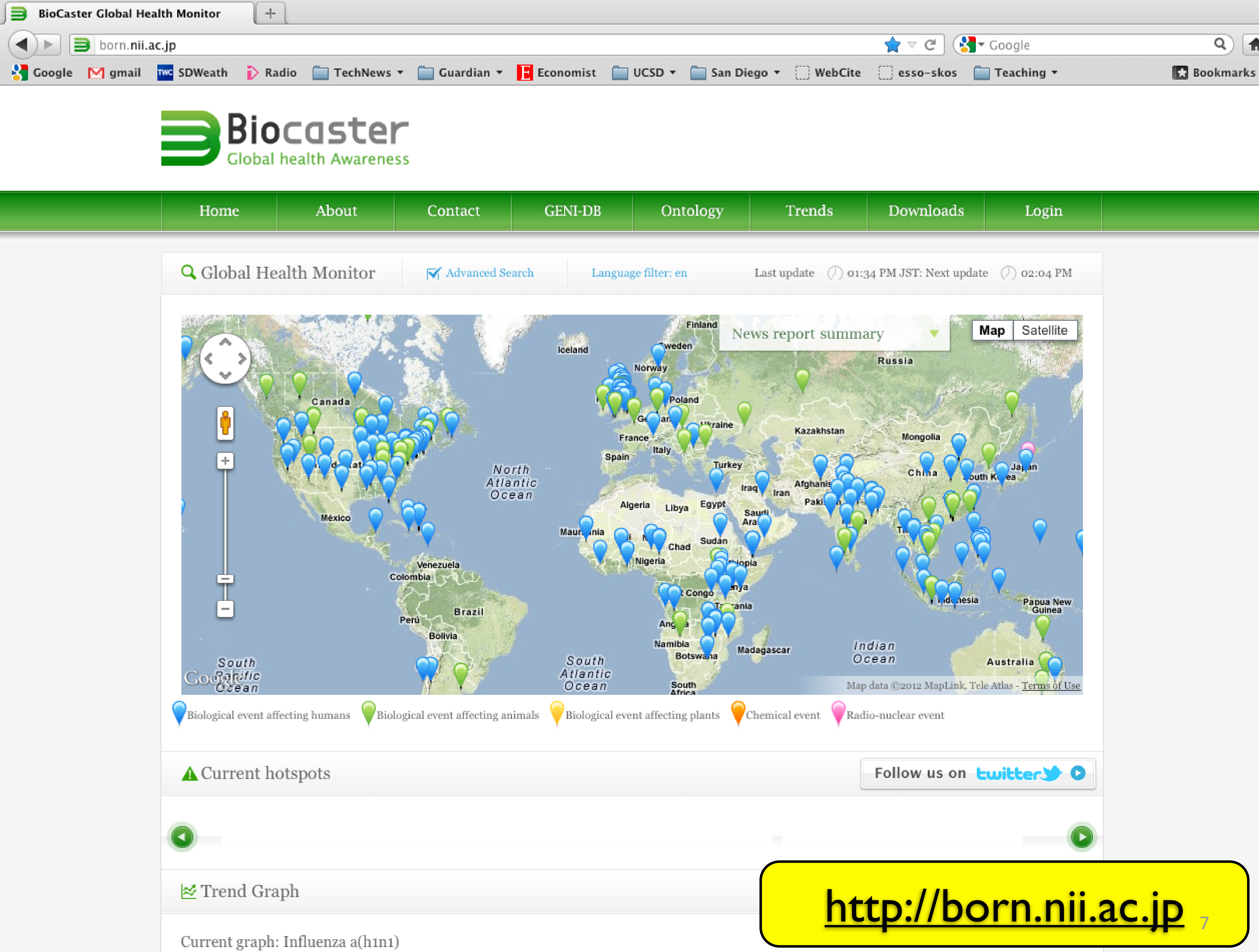
Topical relevancy = true

<LOCATION>Lusaka</LOCATION> suffered the worst <DISEASE>Cholera </DISEASE> epidemic in <TIME>more than ten years</TIME> with <PERSON>120 deaths</PERSON>. Despite the hope that the epidemic submit, heavy rains which have caused flooding in the <LOCATION>Zambian capital</LOCATION>, could even worsen the situation in the <TIME>coming weeks</TIME>, <ORGANIZATION>MSF </ORGANIZATION> said in his note.



Alert = true





http://born.nii.ac.jp

PART 2: CHIEF COMPLAINT BASED SYNDROMIC SURVEILLANCE

Syndromic Surveillance

Definition: "Syndromic surveillance is surveillance using health-related data that precede diagnosis and signal a sufficient probability of a case or outbreak to warrant further public health response"

US Centers for Disease Control

Data sources for syndromic surveillance

- Over-the-counter pharmacy sales
- School absenteeism
- Calls to NHS Direct (in UK)
- **Emergency Room reports (textual)**

Shifting from news text to clinical text

Chief Complaint: Acute and persistent **fever**, **chills**, and **cough**.

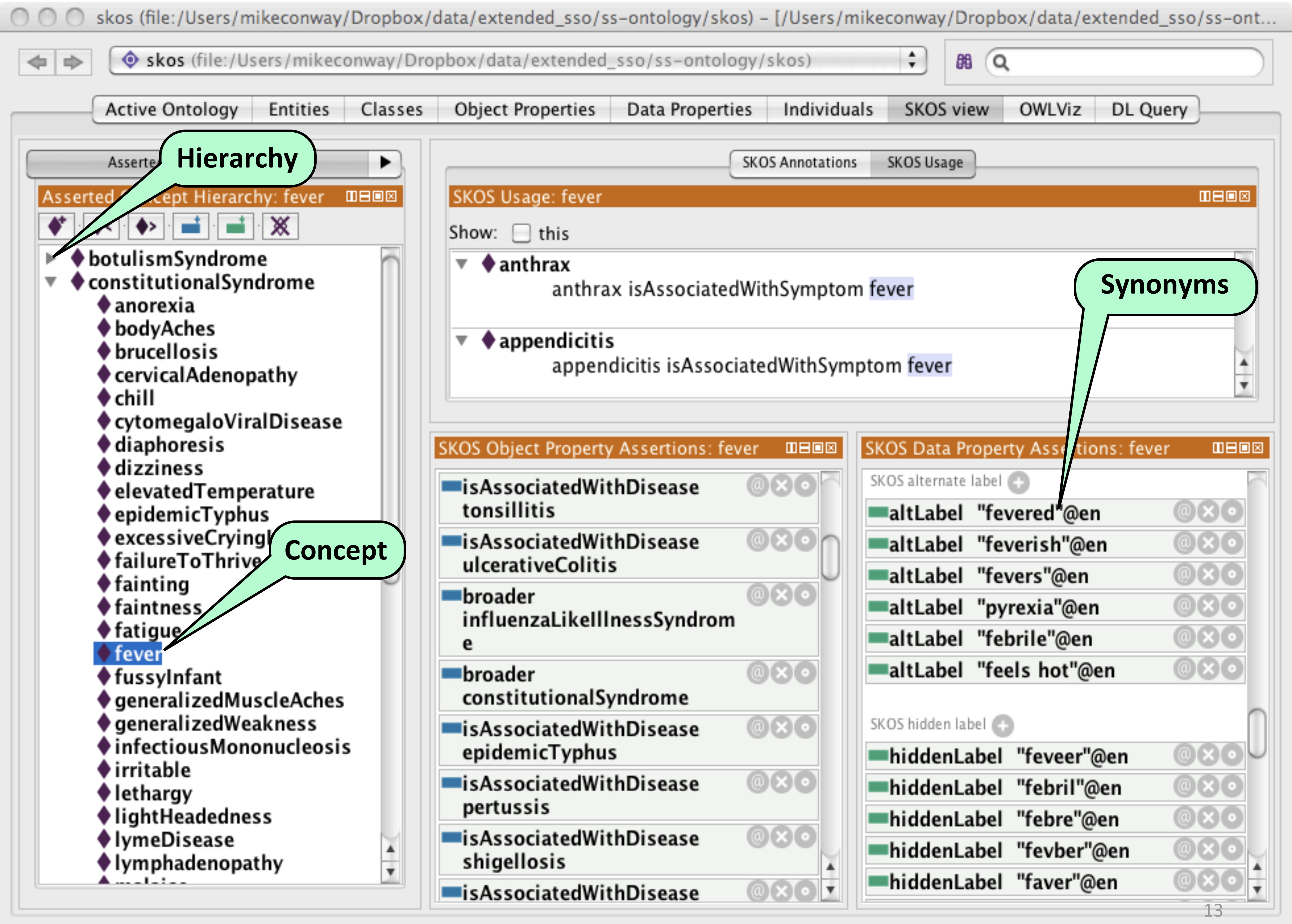
ubiquitous

easily accessible

This is a **AGE[in 50s]-year-old male with a prior history over the past year of an undiagnosed lung problem. He has been seen by pulmonary and infectious disease among others with significant testing including repeated chest CT and a PET scan. No history of cancer. He may have some bronchiectasis or mucus plugging primarily the lingular area. Over the past week, he has had intermittent **fever** and **chills**, although he is only running a 99-degree temperature here. He has had somewhat of a dry at least nonpurulent productive cough. Some **rhinorrhea**, **sore throat**, **headache**, **muscle aches,** and **flu-like symptoms**.

ILI syndrome definition

- Influenza-like illness is a diagnosis of possible influenza with a common set of symptoms
- Most people would agree that **fever**, **chills**, **dry cough** & **body aches** should be part of the syndrome definition
- Disagreement concerning other symptoms: **hoarseness**, **photophobia**, **conjunctivitis**, **wheezing**, etc.


skos (file:/Users/mikeconway/Dropbox/data/extended_sso/ss-ontology/skos) - [/Users/mikeconway/Dropbox/data/extended_sso/ss-ont...
skos (file:/Users/mikeconway/Dropbox/data/extended_sso/ss-ontology/skos)
Active Ontology
Entities
Classes
Object Properties
Data Properties
Individuals
SKOS view
OWLViz
DL Query
Hierarchy
Synonyms
Concept
SKOS Annotations
SKOS Usage
SKOS Usage: fever
Show: this
anthrax
anthrax isAssociatedWithSymptom fever
appendicitis
appendicitis isAssociatedWithSymptom fever
botulismSyndrome
constitutionalSyndrome
anorexia
bodyAches
brucellosis
cervicalAdenopathy
chill
cytomegaloViralDisease
diaphoresis
dizziness
elevatedTemperature
epidemicTyphus
failureToThrive
fainting
faintness
fatigue
fever
fussyInfant
generalizedMuscleAches
generalizedWeakness
infectiousMononucleosis
irritable
lethargy
lightHeadedness
lymeDisease
lymphadenopathy
SKOS Object Property Assertions: fever
isAssociatedWithDisease tonsillitis
isAssociatedWithDisease ulcerativeColitis
broader influenzaLikeIllnessSyndrome
broader constitutionalSyndrome
isAssociatedWithDisease epidemicTyphus
isAssociatedWithDisease pertussis
isAssociatedWithDisease shigellosis
isAssociatedWithDisease
SKOS Data Property Assertions: fever
SKOS alternate label
altLabel "fevered"@en
altLabel "feverish"@en
altLabel "fevers"@en
altLabel "pyrexia"@en
altLabel "febrile"@en
altLabel "feels hot"@en
SKOS hidden label
hiddenLabel "feveer"@en
hiddenLabel "febril"@en
hiddenLabel "febre"@en
hiddenLabel "fevber"@en
hiddenLabel "faver"@en

PART 3: USING TWITTER FOR DISEASE SURVEILLANCE

Three Strands for Twitter Surveillance

– Syndromic surveillance
– Mental health surveillance
– Tobacco surveillance


50g e
GOLDEN VIRGINIA
HAND ROLLING TOBACCO
Smoking seriously harms
you and others around you


LUCKY
STRIKE
IT'S TOASTED


50g e
DRUM
ORIGINAL
Smoking seriously harms
you and others around you


JACK LINK'S
ORIGINAL
JERKY
CHEW
JACK LINK'S


Marlboro
Marlboro
LIGHTS


SMOOTH MINT
SKOAL
SNUS
15
SPIT-FREE
TOBACCO POUCHES
WARNING: Smokeless
tobacco is
addictive.

Tasks, Challenges & Applications

- Identify underage smoking tweets
- Monitor “astroturfing”
- Monitor the impact of anti-smoking campaigns
- Distinguish between tobacco and weed related tweets

“light that shit. smoke that shit. pass that shit.”

“i can rock polka-dots because i'm smokin' hot”

“anybody that thinks the carter iv is better than watch the throne should stop smoking crack as soon as possible”

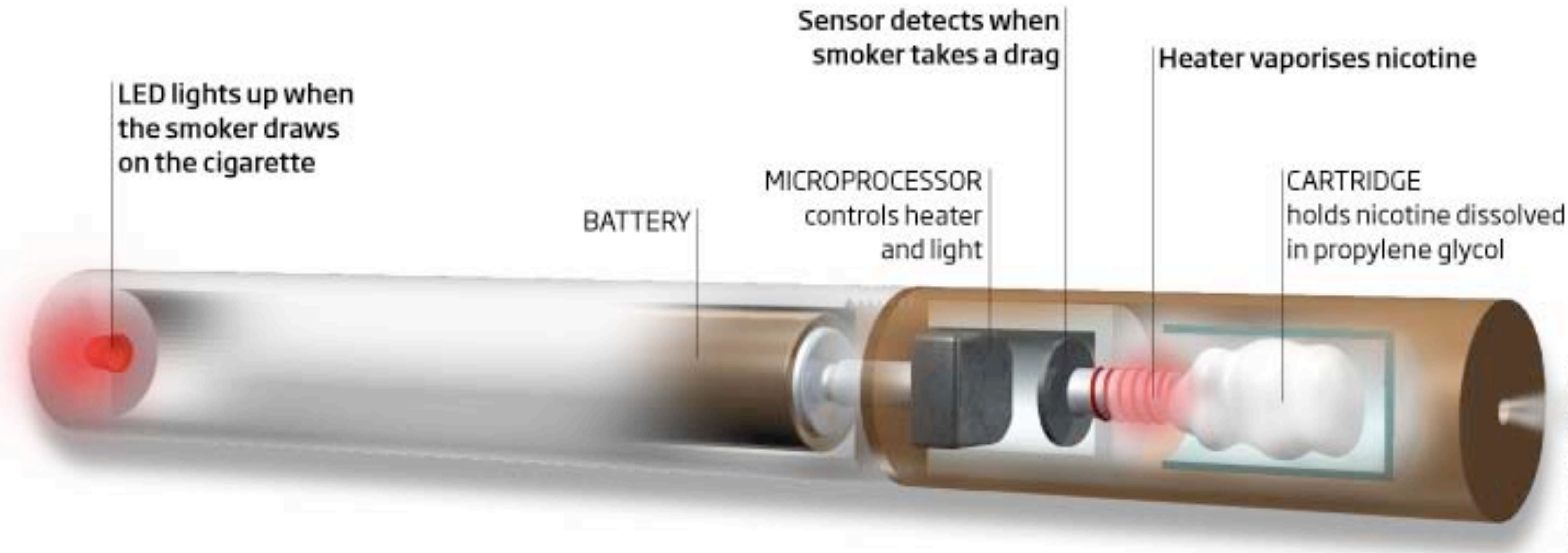
“omg i made the best burgers yday! southern fried chicken, melted cheese and crispy smoked bacons. yum!!! xxx”

Electronic cigarettes

- Patented in China in 2004
- Produces an inhaled vapour that looks and feels like tobacco smoke
- A heated element is used to vapourize nicotine containing liquid into a breathable mist
- Different “flavours” available (mint, strawberry, etc.)
- Relatively high start-up cost, but low running costs
- Legality varies by country - **in the US they are unregulated** (FDA lost case in 2010)
- No one has good data on electronic cigarettes
- How useful is Twitter?

Smoke without fire

Suck on an e-cigarette and it produces a cloud of nicotine-carrying vapour with none of the toxic by-products of burning tobacco



Initial work

- >100 brands
- > 60% of electronic cigarette tweets are advertising
- Currently performing content analysis of tweet types

"drunk investment of the night = electronic cigarette"

"new indianapolis smoking ban kicks in today: the new ordinance also forbids use of electronic cigarette devices"

"i hate electronic cigarettes this shit so point less its nothing but water an nicotine"

"i just got one of those electronic cigarettes. it's exactly like smoking a real cigarette, only shit"

THANK YOU