DCU

National
Taiwan
University

Detecting Incidents with Limited Linguistic Knowledge for Low Resource Languages

EnvComp at ISGC 2017

Chao-Hong Liu, Qun Liu and Ching-Hsien Lee

ADAPT Centre, DCU, Ireland; NTU, Taiwan

Background

- ADAPT - The Centre for Digital Content Technology, Ireland
  - http://www.adaptcentre.ie/
  - Formerly known as CNGL
    (Centre for Next Generation Localisation)
  - Crisis MT (EU), kick-off in April, 2017
- Industrial Technology Research Institute (ITRI), Hsinchu, Taiwan
- Central Geological Survey, MOEA, Taiwan
  - 1999 Jiji earthquake

Related Projects

- NIST LoReHLT 2016
  - Low Resource human language technologies
  - https://www.nist.gov/itl/iad/mig/lorehlt16-evaluations
- Crisis Machine Translation (Crisis MT)
- Exploitation of Social Media for Emergency Relief and Preparedness (SMERP)
  - http://computing.dcu.ie/~dganguly/smerp2017/index.html
  - microblogs posted during the earthquake in Italy in August 2016
  - Shared Tasks on text retrieval and text summarization
- NAPES (EU FP7)
  - Next Generation Analytical Platforms for Environmental Sensing
  - http://www.napes.eu/
  - Workshop 2: March 27-28, 2017 at Croke Park, Dublin
    - INSIGHT Centre
    - NCSR (National Centre for Sensor Research)

Use of Language Processing

In Sensing the Environments

- "Social media in Ebola outbreak," Hossain et al.
  - doi: 10.1017/S095026881600039X
- "On the use of the Internet to collect earthquake information," edited by Bossu and Earle
  - http://www.annalsofgeophysics.eu/index.php/annals/article/view/5364/5494
- Citizen science, "crowd" sensing
- etc.

An Open Participant

From: Chao-Hong Liu
Subject: Coordination on ADAPT Shared Tasks Participation

BRIEF3:
New (and Interesting) NIST evaluations for Low Resourced languages to register on May 16 . . .

9. NIST announced new **Low resource human language technologies (LoReHLT)**evaluation series

The LoReHLT 2016 Evaluations include three sessions:
1) Machine Translation (MT)
2) Named Entity Recognition (NER)
3) Situation Frame (SF)

Situation Frame is now being defined and can be briefed as to detect (the underlined values):

Household supply (Type) 99% for sure (TypeConfidence)
in Dublin (EntityType=Location, Begin=2015, End=Ongoing)
with Urgency (Status).

Team EarthSensory

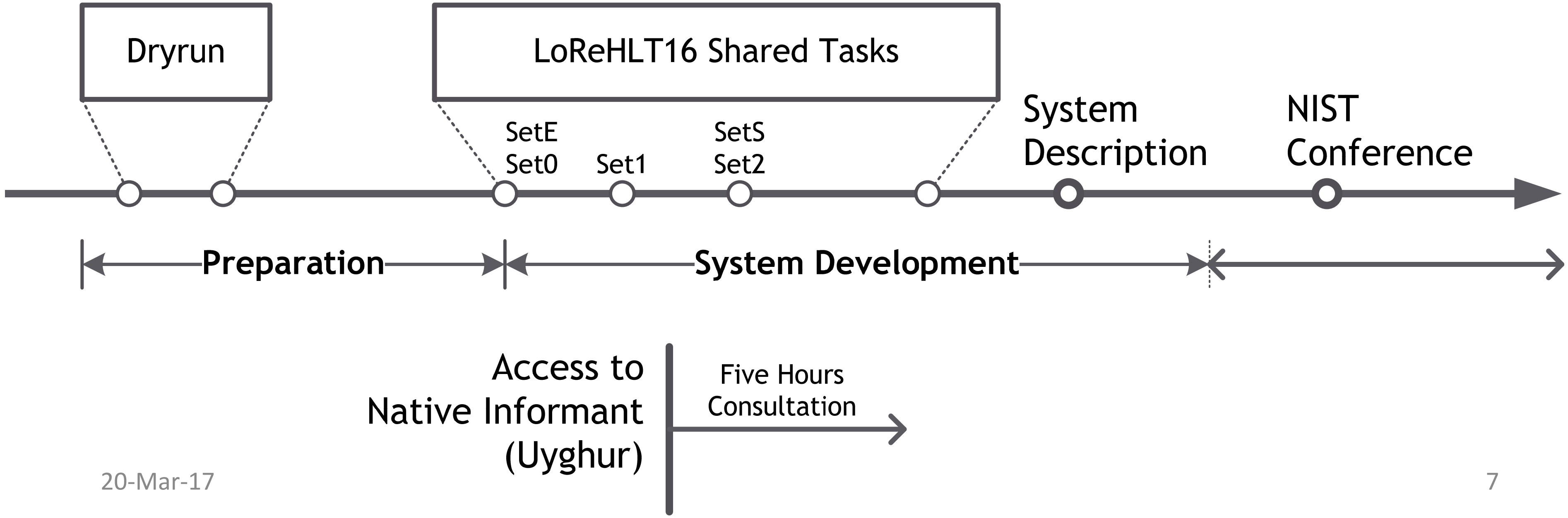
- Chao-Hong Liu and Ching-Hsien Lee
- Professor: Qun Liu

- The DCU-NTU incident detection systems for LoReHLT16 SF task are designed as keyword spotting systems.
  - ChineseUyghurStarDict.txt (ZH to UY)
  - Google translate (**EN to ZH** to UY)
    - SetS: English Scenario Model

Task: Situation Frame (SF)

Situation Frame is now being defined and can be briefed as to detect (the underlined values):

Household supply (Type) 99% for sure (TypeConfidence)
in Dublin (EntityType=Location, Begin=2015, End=Ongoing)
with Urgency (Status).



Dryrun

LoReHLT16 Shared Tasks

SetE
Set0

Set1

SetS
Set2

System Description

NIST Conference

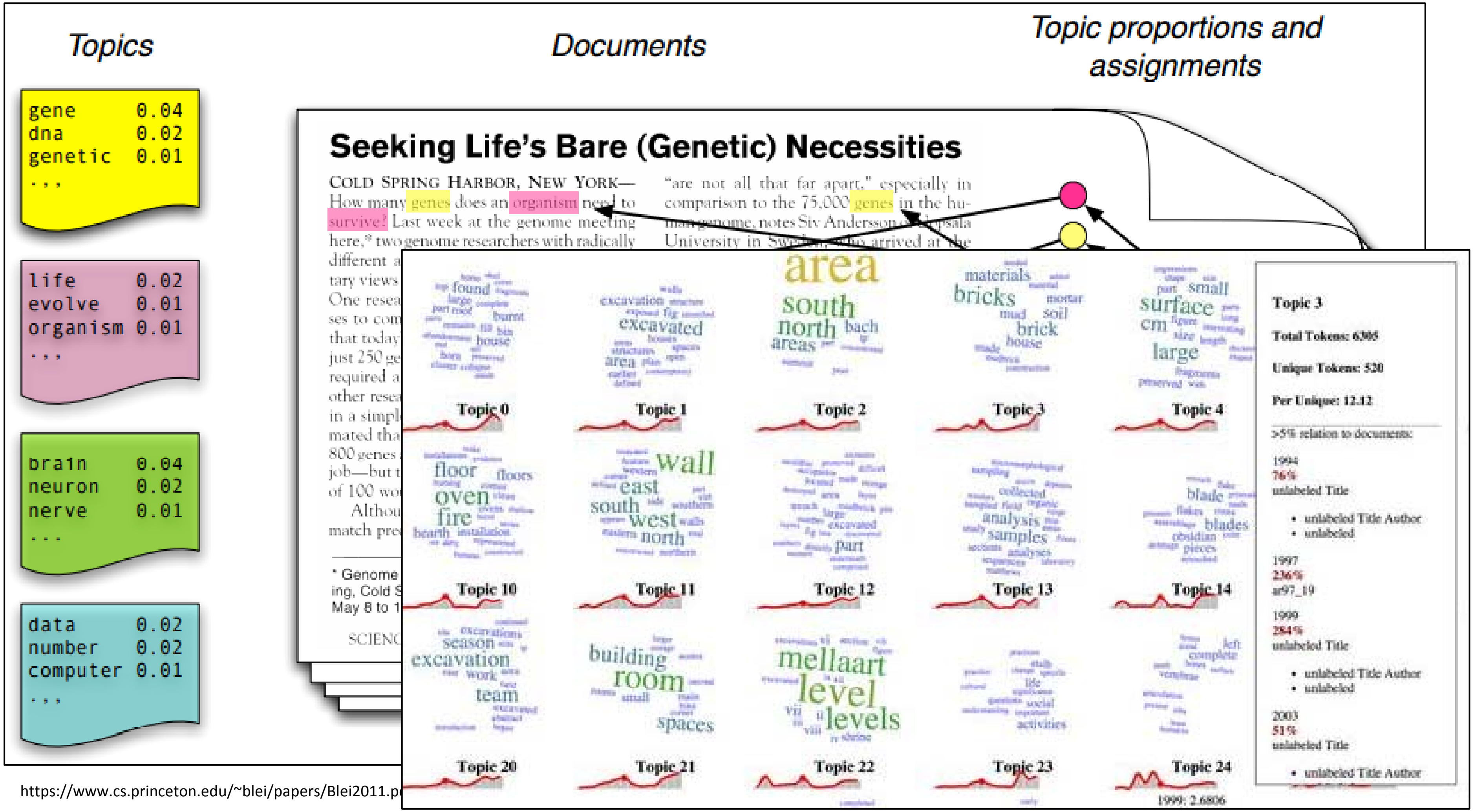
Preparation

System Development

Access to Native Informant (Uyghur)

Five Hours Consultation

Topic Modeling


Topics
Documents
Topic proportions and assignments
gene 0.04
dna 0.02
genetic 0.01
...
life 0.02
evolve 0.01
organism 0.01
...
brain 0.04
neuron 0.02
nerve 0.01
...
data 0.02
number 0.02
computer 0.01
...
Topic 0
Topic 1
Topic 2
Topic 3
Topic 4
Topic 10
Topic 11
Topic 12
Topic 13
Topic 14
Topic 20
Topic 21
Topic 22
Topic 23
Topic 24
1999: 2.6806

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK— How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different a… tary views … One resea… ses to com… that today… just 250 ge… required a… other resea… in a simpl… mated tha… 800 genes … job—but t… of 100 wo… Althou… match pre…

"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the

* Genome … ing, Cold S… May 8 to 1…

SCIENC…

Topic 3

Total Tokens: 6305

Unique Tokens: 520

Per Unique: 12.12

>5% relation to documents:

1994
76%
unlabeled Title
- unlabeled Title Author
- unlabeled

1997
236%
ar97_19

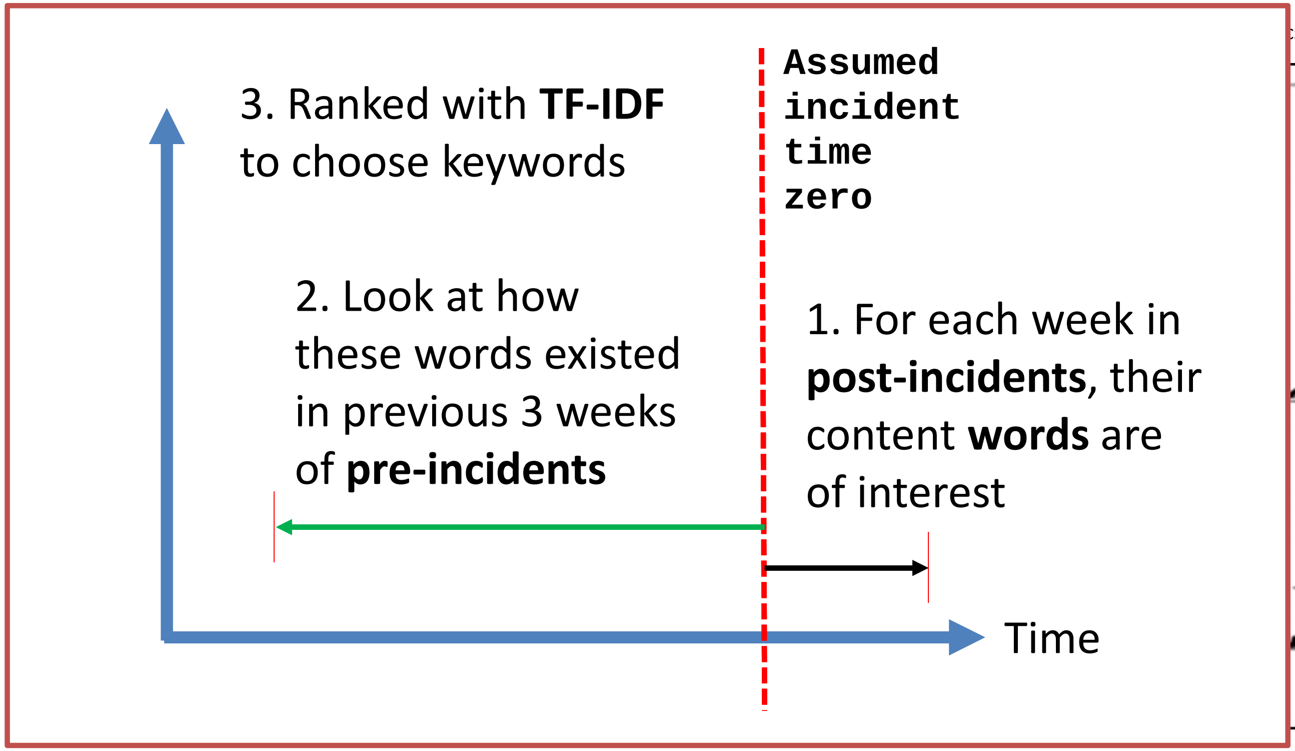
1999
284%
unlabeled Title
- unlabeled Title Author
- unlabeled

2003
51%
unlabeled Title
- unlabeled Title Author

https://www.cs.princeton.edu/~blei/papers/Blei2011.p…

https://twitter.com/elijah_meeks/status/458695317664194560

What we looked for?



3. Ranked with **TF-IDF** to choose keywords

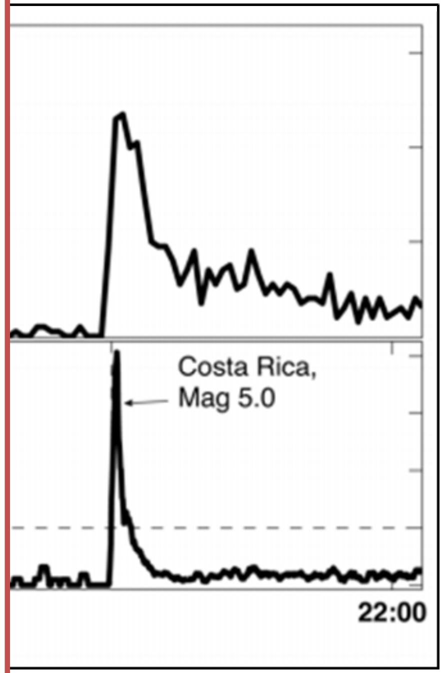
Assumed incident time zero

2. Look at how these words existed in previous 3 weeks of **pre-incidents**

1. For each week in **post-incidents**, their content **words** are of interest

Time

s.eu/index.php/annals/article/view/5364/5494



- We can’t find this “occurrence pattern” from pre/post incident posts/documents

Use of Native Informant (NI)

- Five hours consultation at the end of 2nd week
  - Only CP3 used NI's knowledge in the systems

- Asking NI for
  - Geographic indicator words (given en examples)
  - Keywords reviewing/editing (zh-uy, en-zh-uy)
  - Keyphrases translation (made-up en to uy)
  - Fundamental linguistic knowledge about Uyghur

NI: Geo-indicators

Geographic Indicators	Uyghur Translations
city	شەھەر
county	ناھيە
residential yard	قورو-جاي
Autonomous County (Oblast)	ئوبلاست
province	ئۆلكە
county	دۆلەت
(No corresponding English word)	تاغ

NI: Keywords/Keyphrases

NI edited	Keywords for Terrorism	Source
تۇيۇقسىز ھۇجۇم	تۇيۇقسىز ھۇجۇم	袭击 (Attack)

English Phrases	Uyghur Translations
medical resources are scarce	داۋالاش ئۈسكۈنلىرى يېتىشمەسلىك
to implement the curfew	نەزەربەنت قىلماق
water supply is insufficient	سۇ تەمىنلەش يېتىشمەسلىك

Results

condition	check point	eqvl. class	min error rate	max error rate	num subms.	score
constrained	CP 3	SFType	1.147	6.105	5	2.298
Un-constrained	CP 3	SFType	1.505	5.197	2	5.197

First Thoughts

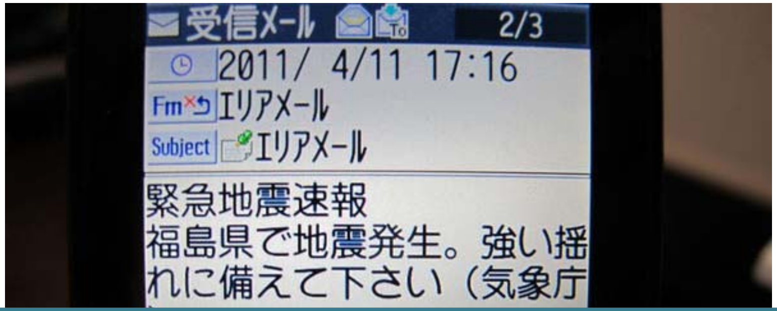
- Earthquakes
  - I saw these in situ
  - Prediction?
  - detection
  - **early warning**
    - sensors
  - **Logistics**
    - where
    - needs
    - emergency level - triage

https://www.youtube.com/watch?v=8cqQJCmcLWQ

http://www.webpages.uidaho.edu/~simkat/geol345_files/345lecture9.htm

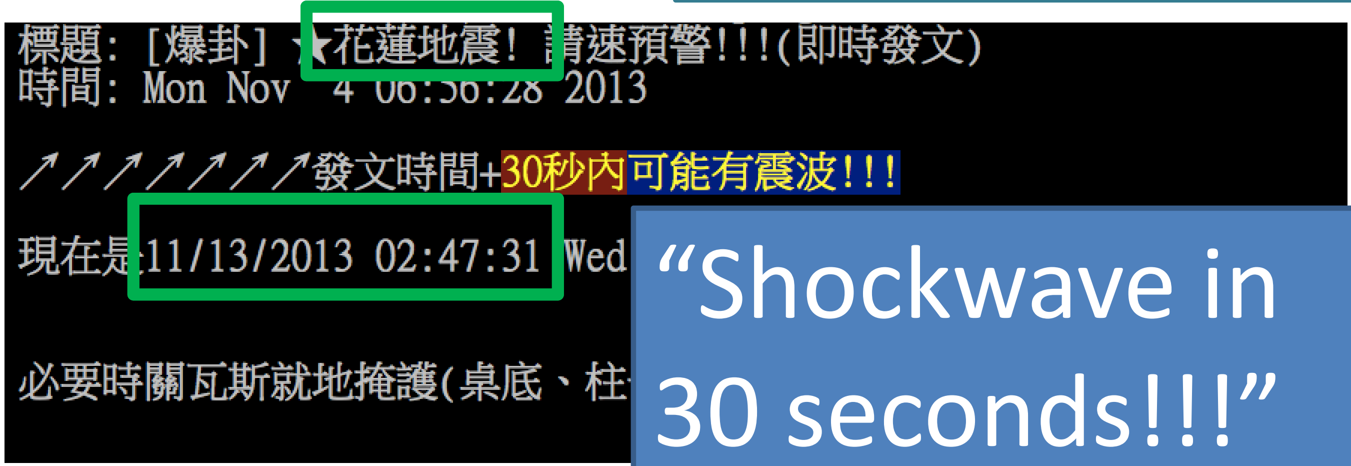
Earthquakes

- Early warning system?
  - Japan, 2011
  - Taiwan, 2013
    - Hey, we have the PTT villagers' earthquake early warning **"system"**



受信メール 2/3
2011/ 4/11 17:16
Fm エリアメール
Subject エリアメール
緊急地震速報
福島県で地震発生。強い揺
れに備えて下さい（気象庁

Japan has automatic system . . .
"Emergency Notification: Earthquake in Fukushima. Get prepared for strong shockwave."



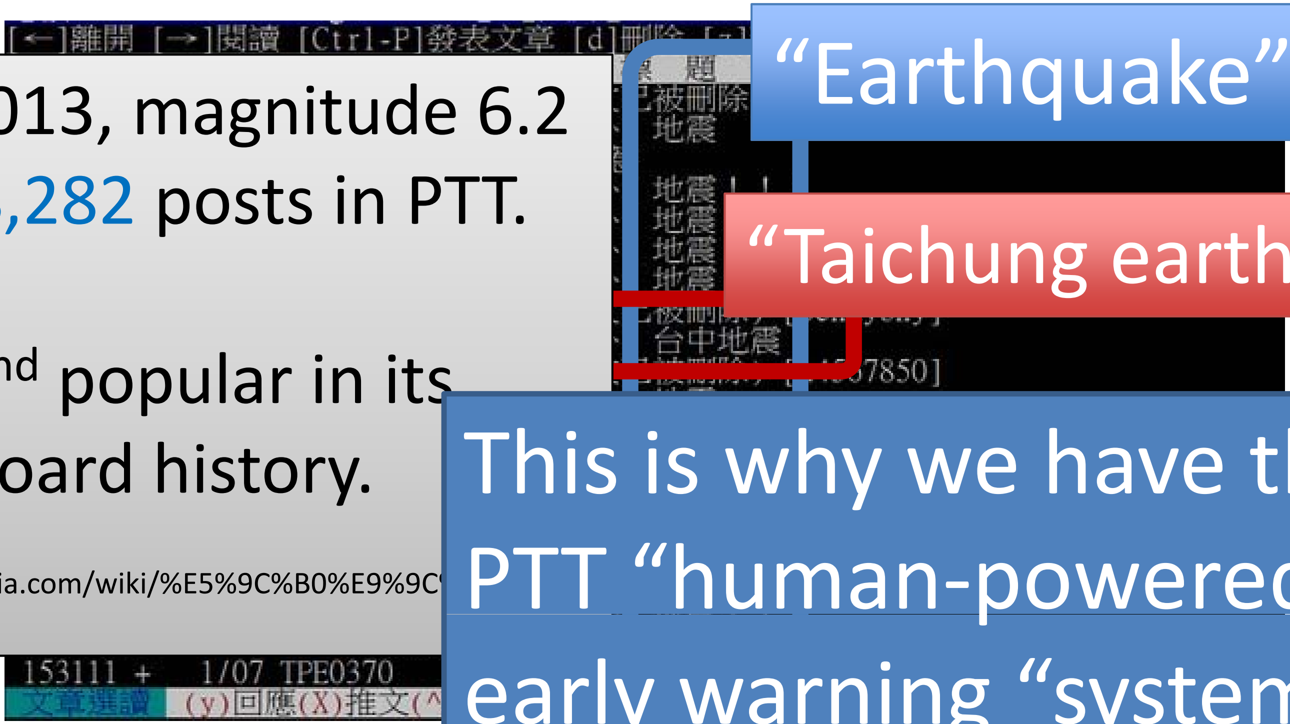
```
標題: [爆卦] ★花蓮地震! 請速預警!!!(即時發文)
時間: Mon Nov  4 06:56:28 2013

↗↗↗↗↗↗↗發文時間+30秒內可能有震波!!!

現在是:11/13/2013 02:47:31 Wed

必要時關瓦斯就地掩護(桌底、柱
```

"Shockwave in 30 seconds!!!"

http://techsea.cc/2013/11/13/ptt-earthquake/

Earthquakes

- "Is this right? http://www.appledaily.com.tw/realtimenews/article/new/20150107/537649/
  - . . . the first thing you sensed earthquakes is **NOT to run or hide, but TO POST IT on the internet? "**


[←]離開 [→]閱讀 [Ctrl-P]發表文章 [d]刪除
題
已被刪除
地震
地震！！
地震
地震
地震
台中地震
7850]
153111 + 1/07 TPE0370
文章選讀 (y)回應(X)推文(^

"Earthquake"

"Taichung earthquake"

Oct 31, 2013, magnitude 6.2 quake, 38,282 posts in PTT.

Ranked 2nd popular in its Gossips board history.

http://zh.pttpedia.wikia.com/wiki/%E5%9C%B0%E9%9C
6%87

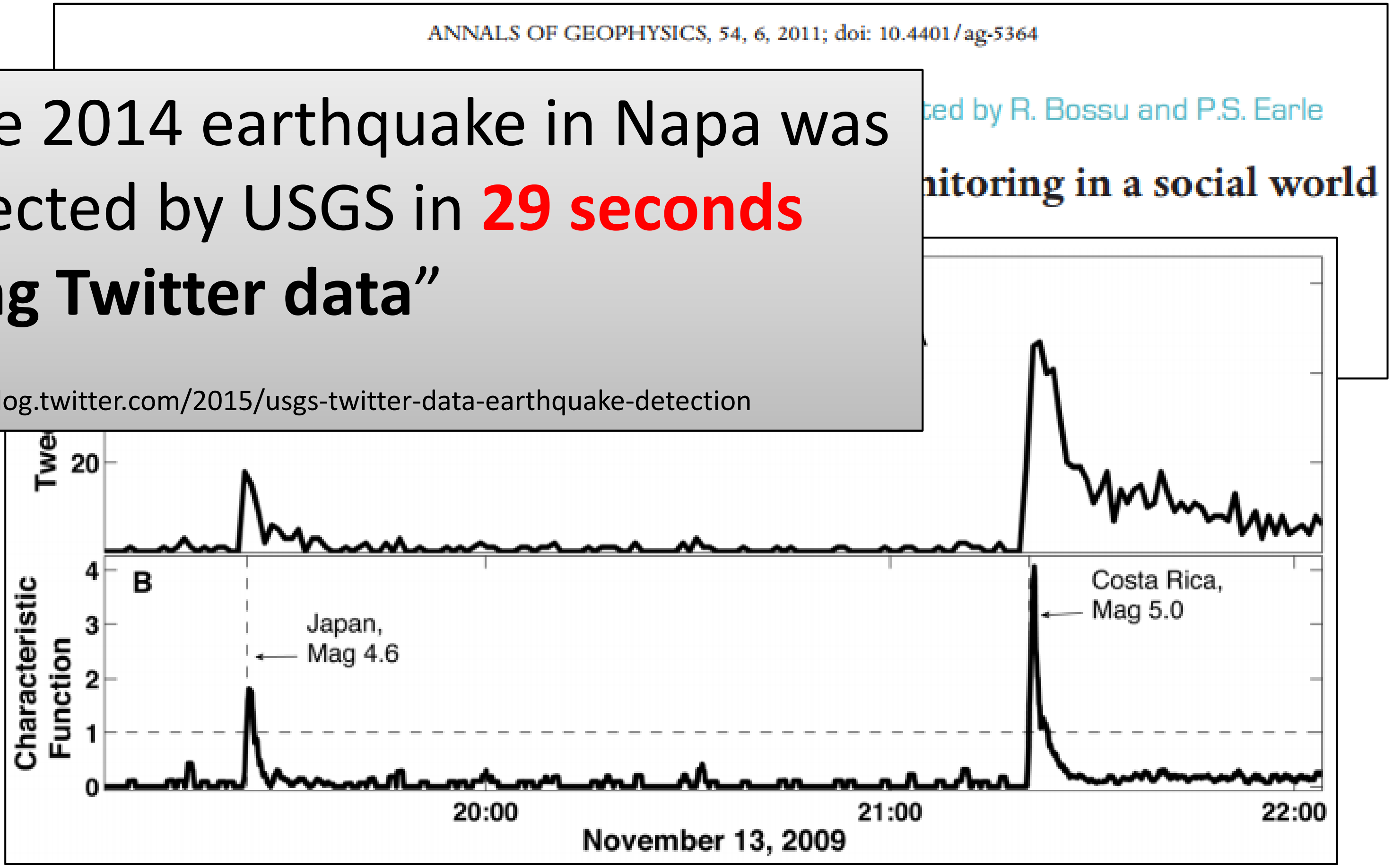
This is why we have the PTT "human-powered" early warning "system."

http://www.appledaily.com.tw/realtimenews/a

Earthquakes

- **On the use of the Internet to collect earthquake information** / ed. by Rémy Bossu and Paul S. Earle
  - U.S. Geological Survey

http://www.annalsofgeophysics.eu/index.php/annals/article/view/5364/5494



ANNALS OF GEOPHYSICS, 54, 6, 2011; doi: 10.4401/ag-5364

ted by R. Bossu and P.S. Earle

nitoring in a social world

“The 2014 earthquake in Napa was detected by USGS in **29 seconds** **using Twitter data**”

https://blog.twitter.com/2015/usgs-twitter-data-earthquake-detection

Earthquakes

- Beyond incident detection
  - logistics
    - News:

      “STOP sending aids . . .”

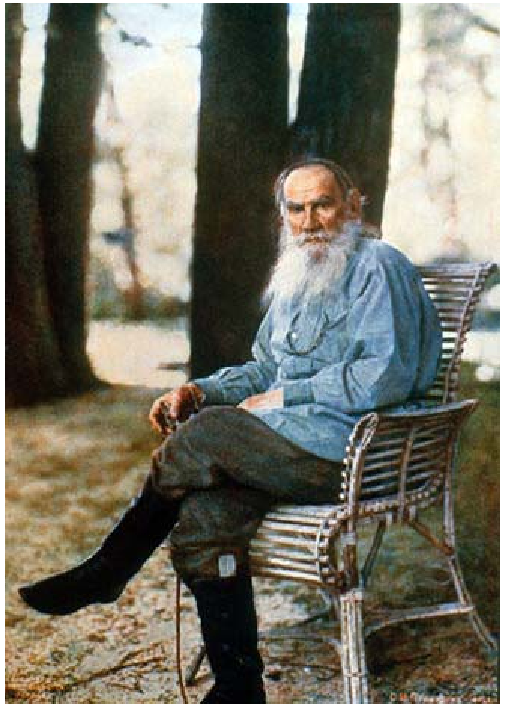
      Too much for some, and

      NONE for other regions.
    - Social networks (SN):
      - Taichung **needs** water
      - Nantou **needs** 1,000 tents
      - Foods **NOT** needed in . . . the next week

http://atc.archives.gov.tw/921/c_01_01.swf

What to model?

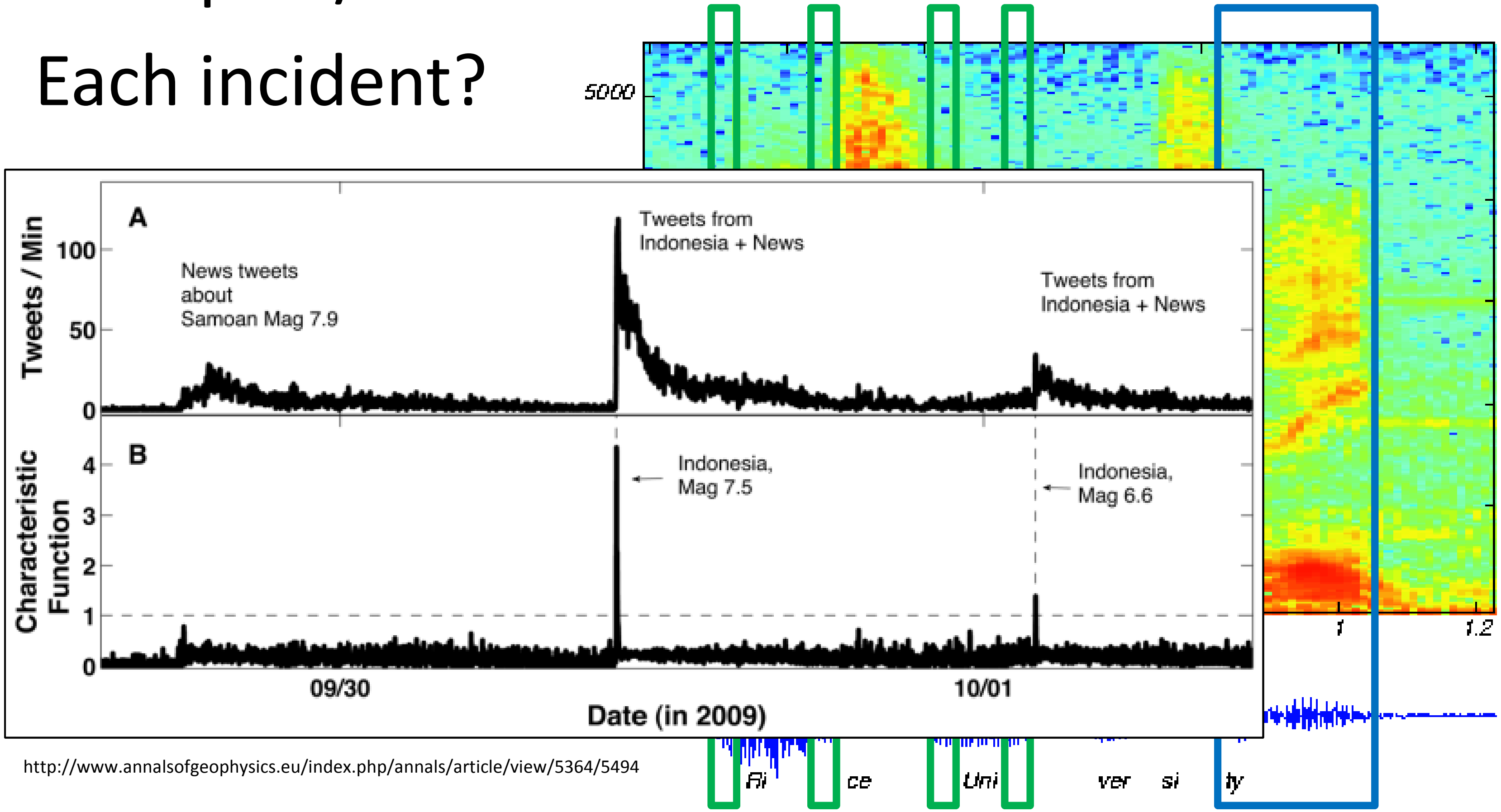
- Leo Tolstoy
  - “All happy families are like one another; each unhappy family is unhappy in its own way.”
- Happy families?
  - Posting patterns?
- Unhappy families
  - Situation frames . . .
  - Annotations



https://en.wikipedia.org/wiki/Leo_Tolstoy

What to model/**annotate**?

- Each post/document?
- Each incident?

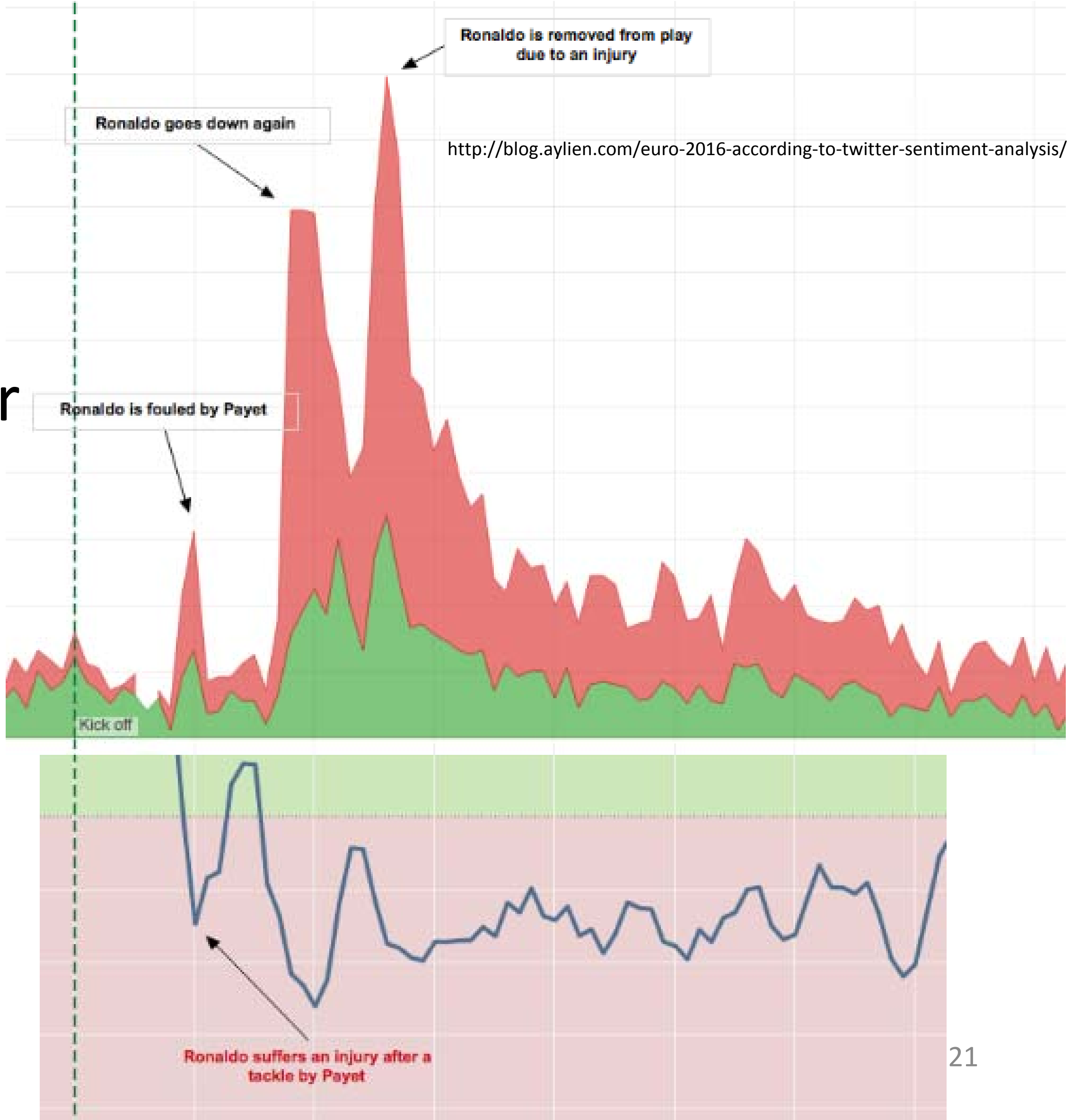


http://www.annalsofgeophysics.eu/index.php/annals/article/view/5364/5494

http://sound.stackexchange.com/questions/22536/what-type-of-information-can-i-get-from-the-waveform-and-the-spectrum-of-audio

What to model?

- Phenomena specific to earthquakes?
  - Euro 2016 soccer tournament
  - sentiment of all of the **English tweets**, which amounted to about **17 million in total**



http://blog.aylien.com/euro-2016-according-to-twitter-sentiment-analysis/

Discussions

- Text itself is not the only source
  - date/time
  - who posted it? (who = an account in social media)
  - where does s/he live? (most co-occurred place)
  - how posts/documents (and, how these whos) are related/connected?

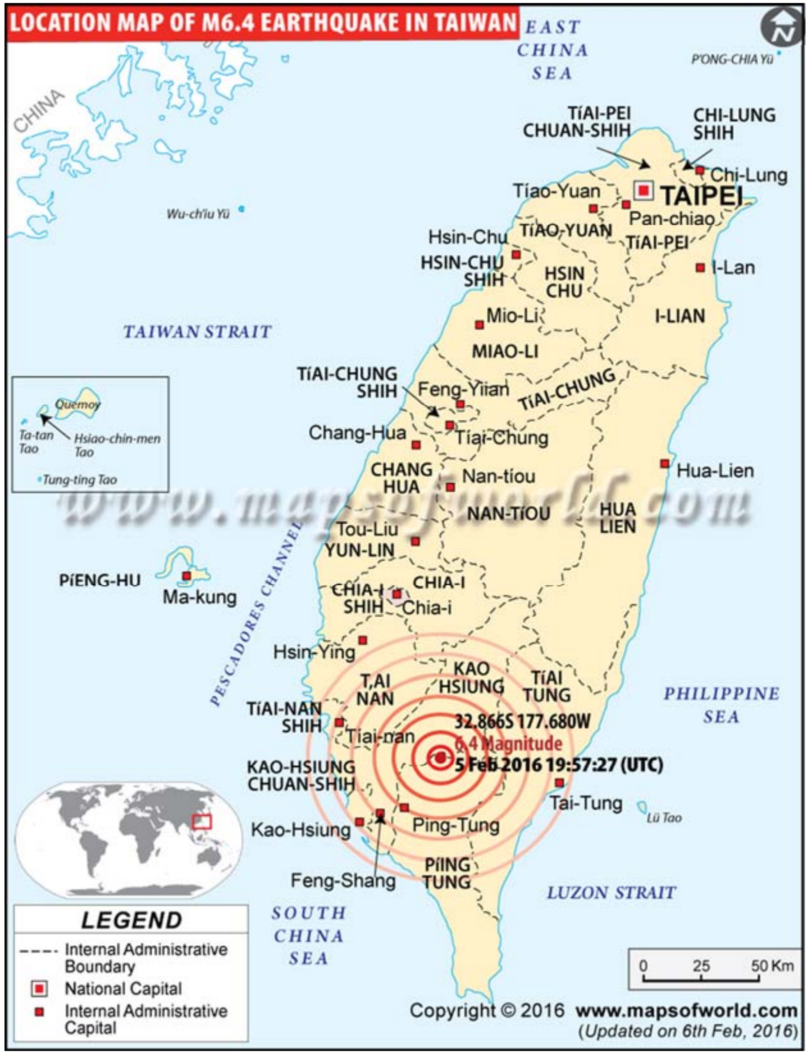
- How about, detecting
  - Situation of **unknown types**? No annotations . . .
  - Alien invasion?

LoReLEI Language Packs

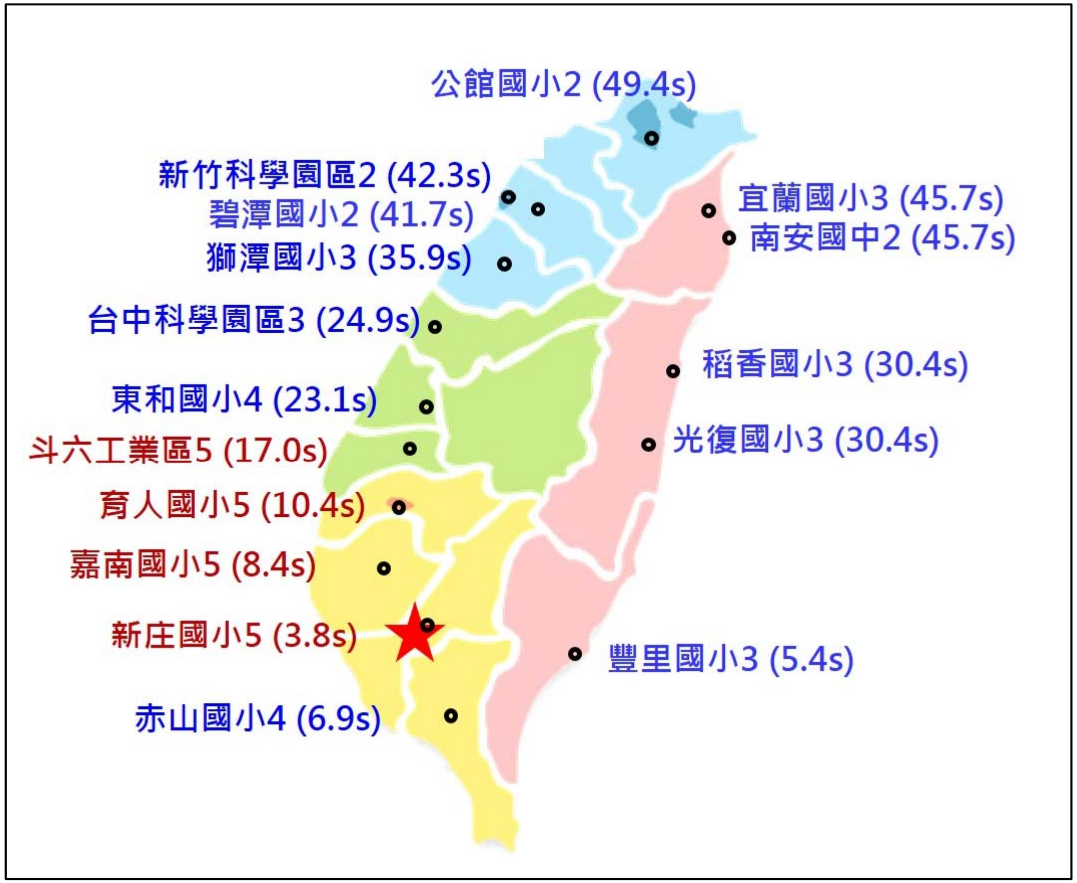
- Haitian Creole
- Turkish
- Uzbek
- Tamil
- Hausa
- Kurmanji Kurdish
- Zulu
- Mandarin
- Lao
- Vietnamese

Thank you!

- Comments?


LOCATION MAP OF M6.4 EARTHQUAKE IN TAIWAN
EAST CHINA SEA
P'ONG-CHIA Yü
CHINA
TíAI-PEI CHUAN-SHIH
CHI-LUNG SHIH
Chi-Lung
Tíao-Yuan
TAIPEI
Pan-chiao
Wu-ch'iu Yü
Hsin-Chu
TíAO-YUAN
TíAI-PEI
HSIN-CHU SHIH
HSIN CHU
I-Lan
Mio-Li
I-LIAN
TAIWAN STRAIT
MIAO-LI
TíAI-CHUNG SHIH
Feng-Yiian
TíAI-CHUNG
Quemoy
Ta-tan Tao
Hsiao-chin-men Tao
Tung-ting Tao
Tiai-Chung
Chang-Hua
CHANG HUA
Nan-tiou
Hua-Lien
www.mapsofworld.com
NAN-TíOU
HUA LIEN
Tou-Liu
YUN-LIN
PESCADORES CHANNEL
PíENG-HU
Ma-kung
CHIA-I SHIH
CHIA-I
Chia-i
Hsin-Ying
KAO HSIUNG
TíAI TUNG
T,AI NAN
PHILIPPINE SEA
TíAI-NAN SHIH
32.866S 177.680W
Tiai-nan
6.4 Magnitude
5 Feb 2016 19:57:27 (UTC)
KAO-HSIUNG CHUAN-SHIH
Tai-Tung
Lü Tao
Kao-Hsiung
Ping-Tung
PíING TUNG
Feng-Shang
LUZON STRAIT
LEGEND
Internal Administrative Boundary
National Capital
Internal Administrative Capital
SOUTH CHINA SEA
0 25 50 Km
Copyright © 2016 www.mapsofworld.com
(Updated on 6th Feb, 2016)

http://www.mapsofworld.com/thematic-maps/earthquake/taiwan.html


公館國小2 (49.4s)
新竹科學園區2 (42.3s)
碧潭國小2 (41.7s)
獅潭國小3 (35.9s)
宜蘭國小3 (45.7s)
南安國中2 (45.7s)
台中科學園區3 (24.9s)
稻香國小3 (30.4s)
東和國小4 (23.1s)
光復國小3 (30.4s)
斗六工業區5 (17.0s)
育人國小5 (10.4s)
嘉南國小5 (8.4s)
新庄國小5 (3.8s)
豐里國小3 (5.4s)
赤山國小4 (6.9s)

http://www.narlabs.org.tw/tw/news/news.php?news_id=1376