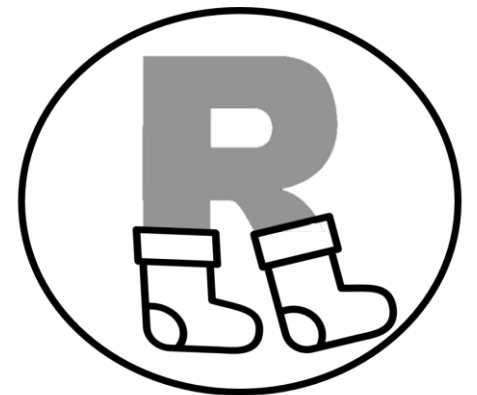# the GDELT Project

Pei Lee

the GDELT Project

- the Global Database of Events, Language, and Tone

- 全球事件、語言和音調資料庫

- Jigsaw（前 Google 智庫部門 Google Ideas）

- 世界上最大的社會事件開放資料庫：監控全世界 100 多種語言的新聞媒體

the GDELT Project

- an initiative to construct a catalog of human societal-scale behavior and beliefs across all countries of the world, connecting every person, organization, location, count, theme, news source, and event across the planet into a single massive network that **captures what's happening around the world, what its context is and who's involved, and** how the world is feeling about it, every single day.

the GDELT Project

事件發生 → 印刷、廣播、網頁 → Jigsaw →

角色？
關係？
地點？
事件類型？
影響該國穩定程度的指標
正負面評價

每 15 分鐘更新一次

the GDELT Project

- 資料存取方式：

	GDELT Analysis Service	Google BigQuery	GDELT API
費用	免費	收費	免費
篩選彈性	小	大	中
技術要求	中	高	低
資料齊全	中	高	低
限制	EVENT：前 20,000 筆資料 GKG：前 200,000 筆資料	$$$	近 12 週 網路新聞

the GDELT Project

- GDELT Analysis Service

1. GDELT Event Database：

   - EVENT TABLE：58個欄位
   - MENTIONS TABLE：15個欄位
   - 詳細說明：http://data.gdeltproject.org/documentation/GDELT-Event_Codebook-V2.0.pdf

2. GDELT Global Knowledge Graph：

   - 詳細說明：http://data.gdeltproject.org/documentation/GDELT-Global_Knowledge_Graph_Codebook-V2.1.pdf

the GDELT Project

- ## Google BigQuery

  - 例如：

    SELECT theme, COUNT(*) as count

    FROM (

    select REGEXP_REPLACE(SPLIT(V2Themes,';'), r',.*', '') theme

    from [gdelt-bq:gdeltv2.gkg]

    where DATE>20150302000000 and DATE < 20150304000000 and V2Persons like '%Tsipras%'

)

    group by theme

    ORDER BY 2 DESC

    LIMIT 300

the GDELT Project

- Google BigQuery

  - 結果（部分）：

Row	theme	count
1	GENERAL_GOVERNMENT	5502
2	TAX_FNCACT_MINISTER	5184
3	TAX_WORLDLANGUAGES_GREEK	4251
4	TAX_ETHNICITY_GREEK	4205
5	ECON_BANKRUPTCY	3514
6	LEADER	2743
7	TAX_FNCACT_PRIME_MINISTER	2286
8	ECON_WORLDCURRENCIES_EURO	1036
9	NEGOTIATIONS	1030

the GDELT Project

- Google BigQuery 收費方式

Action	Cost	Notes
Storage	$0.02 per GB, per month	First 10 GB is free each month
Long Term Storage	$0.01 per GB, per month	
Streaming Inserts	$0.05 per GB	
Queries	$5 per TB	First 1 TB per month is free

the GDELT Project

- GDELT API：

  - 使用說明：http://blog.gdeltproject.org/gdelt-doc-2-0-api-debuts/

  - 例如：
    ```
    http://api.gdeltproject.org/api/v2/doc/doc?
    query= theme:ACT_MAKESTATEMENT  "trump"  Domain:cnn.com &
    mode=ArtList &
    format=CSV &
    STARTDATETIME=20170418000000 &
    ENDDATETIME=20170711000000
    ```

the GDELT Project

- GDELT API :

  - QUERY :

    - "(KEYWORD) "
    - (a OR b)
    - -
    - Domain
    - ImageFaceTone
    - ImageNumFaces
    - ImageOCRMeta
    - ImageTag

the GDELT Project

- GDELT API :

  - QUERY :

    - ImageWebCount
    - ImageWebTag
    - SourceCountry
    - SourceLang
    - Theme
    - Tone
    - ToneAbs

the GDELT Project

- GDELT API :

  - MODE :

    - ArtList
    - ImageCollage
    - ImageCollageInfo : the number of times (up to 200) it has been seen
    - ImageCollageShare
    - TimelineVol
    - TimelineVolInfo : the top 10 most relevant articles

the GDELT Project

- GDELT API :

  - MODE :

    - TimelineTone :
      http://api.gdeltproject.org/api/v2/doc/doc?query=%22internet%20of%20things%22&mode=TimelineTone
    - TimelineLang :
      http://api.gdeltproject.org/api/v2/doc/doc?query=%22internet%20of%20things%22&mode=TimelineLang
    - TimelineSourceCountry :
      http://api.gdeltproject.org/api/v2/doc/doc?query=%22internet%20of%20things%22&mode=TimelineSourceCountry
    - ToneChart :
      http://api.gdeltproject.org/api/v2/doc/doc?query=%22internet%20of%20things%22&mode=ToneChart

the GDELT Project

- GDELT API :

  - MODE :

    - WordCloudEnglish
    - WordCloudNative : http://api.gdeltproject.org/api/v2/doc/doc?query=%22artificial%20intelligence%22&mode=WordCloudNative
    - WordCloudTheme : http://api.gdeltproject.org/api/v2/doc/doc?query=%22artificial%20intelligence%22&mode=WordCloudTheme
    - WordCloudImageTags : http://api.gdeltproject.org/api/v2/doc/doc?query=%22artificial%20intelligence%22&mode=WordCloudImageTags
    - WordCloudImageWebTags : http://api.gdeltproject.org/api/v2/doc/doc?query=%22artificial%20intelligence%22&mode=WordCloudImageWebTag

簡報結束