

文本事件結構分類與 語意分析

台師大通識教育課程

文本分析與程式設計

授課：卓騰語言科技 _ PeterWolf

什麼是事件？(簡單版 vs. 詳情版)

當 (i), (ii), (iii) 同時為 TRUE 的時候, 發生了一個 "to close (關閉)" 的事件。

(i) 有個事件 e , 是屬於 $CLOSE()$ 事件集合的成員

(ii) 觸發事件的, 叫 AGENT

(iii) 被事件影響的, 叫 THEME

a. *to close*: $\lambda y \lambda x \lambda e [CLOSE(e) \& AGENT(e, x) \& THEME(e, y)]$

b. *to close*: $\lambda y \lambda x \lambda e [AGENT(e, x) \& THEME(e, y) \& \exists e' [CAUSE(e, e') \& THEME(e', y) \& \exists s [BECOME(e', s) \& CLOSED(s) \& THEME(s, y)]]]$

上面這個「詳細版」讀做 ...

有個 x 觸發了一個事件 e , 且有個 y 被事件 e 所影響。而這個 e 將導致 y 改變狀態 s , 成為 e' 。當這個狀態 s 是由 e' 的改變而來, 而且狀態 s 是「關閉的狀態」, 且 y 被於這個「關閉狀態」所影響時, 發生了一個 "to close(關閉)" 的事件

人的智慧以「事件」為單位運作，會是什麼樣子？

某甲對某乙說：「Kitty 喵喵叫！」
某乙要怎麼確定這句話「有意義」？

把「意義」定義為「其言為真(True)」，
否則就是「無意義(False)」。



$\lambda x. \text{meow}(x)$
[[Kitty meows]] = $\lambda x[\text{meow}(x) \ \& \ \text{Kitty}(x)]$

程式語言裡有對應的寫法嗎？
有，也叫 lambda abstraction
(或「無名函式」)



```
def knowledge(entry, collection):
    knowledge = {"KITTY": ["KITTY", "Kitty", "kitty"]}
    if entry in knowledge[collection]:
        return True
    else:
        return False

def meow(entry, collection):
    knowledge = {"MEOW": ["KITTY", "Kitty", "kitty"]}
    if entry in knowledge[collection]:
        return True
    else:
        return False

checkLIST = ["kitty", "snopy", "pikachu"]
list(filter(lambda x: knowledge(x, "KITTY") and meow(x, "MEOW"), checkLIST))
['kitty']
```

語意學 (semantics) 和程式語言 (Python, haskell, java...) 的 **Lambda abstraction** 本來就是同一件事哦！



lambda abstraction |



🕒 lambda abstraction python

移除

🕒 lambda abstraction semantics

移除

🔍 lambda abstraction haskell

🔍 lambda abstraction

🔍 lambda abstraction in ppl

🔍 lambda abstraction java

🔍 lambda abstraction algebras

🔍 lambda abstraction logic

🔍 lambda abstraction operator

🔍 lambda abstraction eta conversion

簡單版還是很複雜呀！沒關係，我們有 **Articut lv3**

- 他關上大門
 - event: 關上
- 他把大門關上
 - event: 關上
- 大門被他關上了
 - event 關上

THEME
+ 大門

ps.「存現句」不是事件。「存現」是指描述「存在、發生、發現」...等意義的句子。
e.g.,「房間裡有一張桌子」、「他發現一支筆」、「村子裡發現一隻山羊在跳舞」

試試 Articut lv3



設定： lv3 語意分析 ▾ A 拼音： 注音 ▾

WikiData： × 關閉 自定義詞典 ▾

他關上大門，把窗戶也關上了。

動作事件 (Event)

他 關上 大門 ，把 窗戶 也 關上了 。

機讀結果

input : [[0,5],[6,13]]

entity : [[[3,5,"大門"]],[[1,3,"窗戶"]]]

event : [['關上',"大門"],['關上了',"窗戶"]]

```
1  #!/usr/bin/env python3
2  #-*- coding:utf-8 -*-
3
4  from ArticutAPI import ArticutAPI
5
6
7  def main(inputSTR, nlptool):
8      resultDict = articut.parse(inputSTR, level="lv3")
9      return resultDict
10
11
12
13  if __name__ == "__main__":
14
15      inputSTR = "他關上大門，把窗戶也關上了。"
16      articut = ArticutAPI.Articut()
17      resultLIST = main(inputSTR, articut)
18      eventLIST = resultLIST["event"]
19      print(eventLIST)
```

Debug I/O Python Shell Debug Console Watch Modules Bookmarks Messages OS < >

[pid 24344] lv3.py Debug I/O (stdin, stdout, stderr) appears below

No module named 'graphene'
Articut-graphQL requires 'graphene' module.
Please use pip/conda install graphene-python to install the module and reload ArticutAPI.
[['關上', '大門'], ['關上了', '窗戶']]

知道文本的「事件」，可以幹嘛？可以算數學呀！



實體物件 (Entity)

存現句

事件句

疑問句
(詢問事件結果)

桌 上有3條 魚，貓咪 偷了1條 魚，還剩幾條？

動作事件 (Event)

桌上有3條魚，貓咪偷了1條魚，還剩幾條？

Calculation

魚

貓

條

-1

subtraction()
處理任何會導致 Qty 減少的計算
回傳「減少的數」
e.g., 吃掉兩個蛋糕
蛋糕 = -2

addition()
處理任何會導致 Qty 增加的計算
回傳「增加的數」
e.g., 買了九個蛋糕
蛋糕 = +9

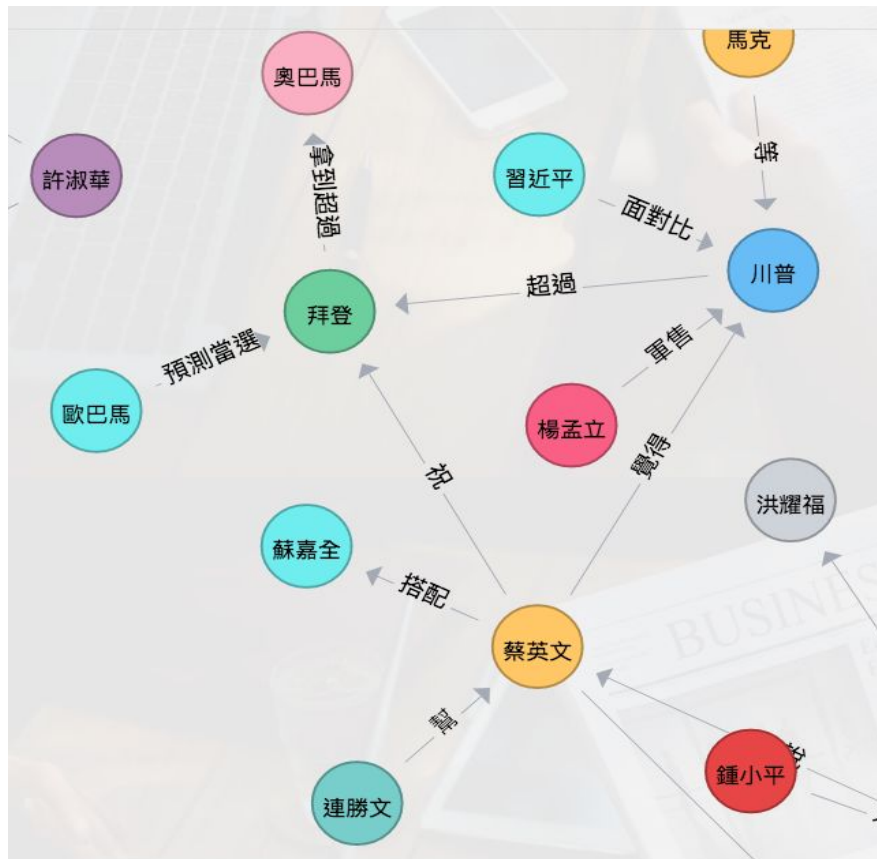
referential()
處理任何以參照法求 Qty 的計算
回傳「變化的量」
e.g., 比妹妹多七個蛋糕
蛋糕 = 妹妹 蛋糕 +7

else()

圖四、事件池 (Event Pool)

還可以拿來做輿情分析

網路觀察家: <https://api.droidtown.co/observer/>



肉搜青紅燈: <https://api.droidtown.co/doxing/>

```
articut.getPersonLIST()  
articut.parse(inputSTR, level="lv3")["event"]  
[[ ],[ ],[PERSON, PERSON],[ ]]  
[[ ],[ ],[EVENT],[ ]]
```

每日 10:00 (GMT+8) 更新, 今天增加 拜登對川普、川普對拜登、黃大洲對柯文哲、柯韓對李彥秀、柯文哲對黃大洲...

主題人物

川普

人物交換

同類人物

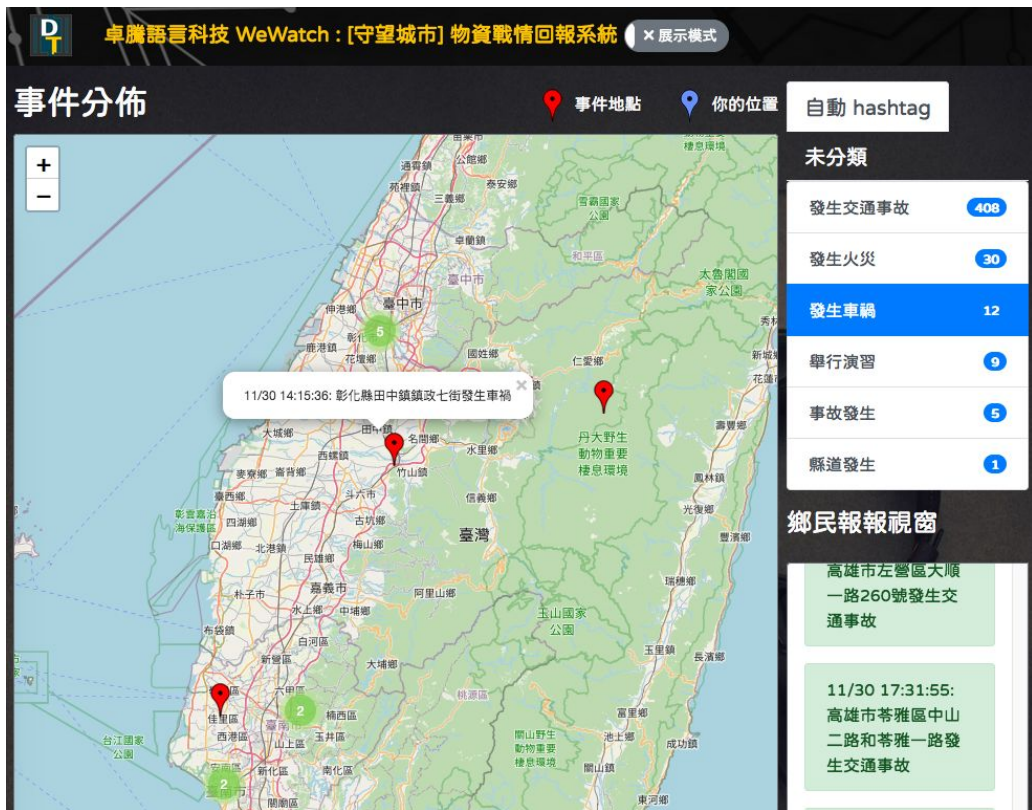
拜登

川
普

拜
登

+	2020-11-18	6
+	2020-11-19	9
+	2020-11-20	13
+	2020-11-21	11
+	2020-11-22	6
+	2020-11-23	6
+	2020-11-24	11
+	2020-11-25	5
+	2020-11-26	2
+	2020-11-27	7

反過來用，可以讓電腦理解「世界發生了什麼事」



文藝 Articut

設定: lv2 詞組斷詞 景點資料庫: 關閉 Wiki

自定義詞典 免費字數: 1742

台中市北屯區中清路發生車禍

機讀結果: 台中市/北屯區...(more)

台中市 location 北屯區 location 中清路 routetw 發生 verb 車禍 nouny

地點



存現



自動學習知識表徵 (知識圖譜、本體知識、知識網、**Ontology...**)

知識表徵

Knowledge Representation

陳李綱

林猷舜

莊志洋

2000年12月

教育大辭書



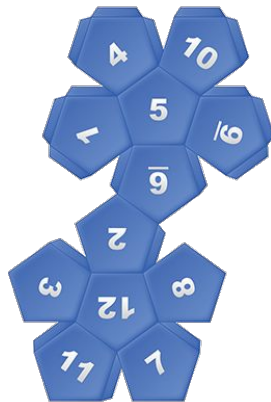
名詞解釋： 知識表徵指代表知識的各種方式或形式。從認知心理學中訊息處理論的觀點言，人類的知識系統中有各種不同形式的表徵方式，如命題、條件語句、基模及心像等方式。「命題」代表知識的基本單位，每一個命題代表每一個概念。概念與概念的聯絡與溝通以命題網絡來表示，可將一些事實性資料聯結而形成敘述性知識。而程序性知識則以「條件語句」的方式表示，以「若…，則…」的方式表徵一些應用性方法或程序等方面的知識。「基模」則是代表敘述性知識與程序性知識關係的有組織的參考架構。「心像」是一種空間和連續資料的表徵知識。由此可見，人類知識的獲得與學習如同電腦儲存訊息的歷程一樣，是以一種有系統的、有組織的方式，將不同種類的資料與訊息以「命題」、「心像」、「基模」與「條件語句」等方式進入記憶系統中，累積形成敘述性知識與程序性知識的認知架構。在電腦使用方面，知識表徵又譯為「知識表示」，目前以知識表示為以一種語言或邏輯規則將知識表示出來，使電腦能夠辨認人類的知識且能用在自動推理上。這些描述法或邏輯規則有特定的語法(syntax)和語意(semantics)。語法是形成句子的規則，而語意是解譯句子和用來演繹其他句子的規則。將人類知識依據這些語法表示成句子，電腦再依所用語言之語意來解讀句子，即為知識表示，一階邏輯(first-order logic)便是一種較為人知的知識表示法。

知識圖譜

1. 本質上是一種語意網路。是「用以表達意義」的一種網路結構。
2. 網路結構可視為一個多面體，每一面是一個「節點 (Point)」，而面和面之間則以「邊 (Edge)」相連結
3. 在知識圖譜里，每個節點表示現實世界中存在的「實體 (Entity)」，每條邊為實體與實體之間的「關係 (Relation)」

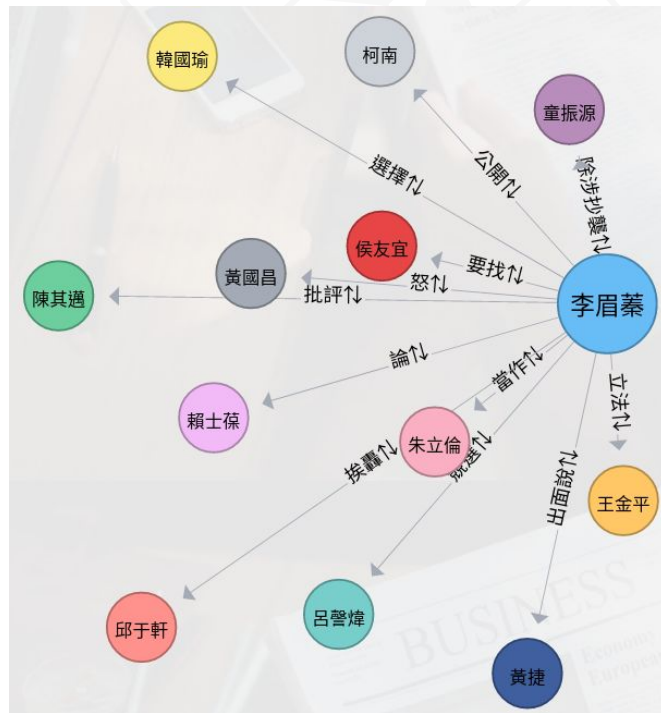


多面體



展開的多面體

圖譜化多面體



各種知識網

廣義知識網本體架構2.0版線上瀏覽 (本介面呈現約二分之一的常用詞彙供參考查詢)

The screenshot shows the XLORE 2.0 online browser interface. On the left, there is a sidebar with buttons for 'Collapse', '說明', '查詢點', '查定義式', and '進階定義式查詢'. The main area displays a hierarchical tree structure of concepts. The root is 'TopNode', which branches into 'entity|事物 [事物]', 'event|事件', 'object|物體 [對象, 客體]', 'thing|萬物 [東西, 萬物, 萬有, 東東]', 'physical|物質', 'abstract|抽象物', 'part|部分 [部分, 份, 部份, 角, 成分, 成份, 片段, 部件, 片斷]', 'group|群體', 'foundation|基礎 [基本, 基礎, 根本, 基, 宗, 主體, 根基, 基石, 母體, 本體, 基底]', 'obstacle|障礙 [障礙, 障礙物]', 'Miscellaneous(thing|萬物)', 'time|時間', 'space|空間 [空間, 空隙, 間隙]', 'relation|關聯 [關係]', and 'Miscellaneous(object|物體)'. The 'relation' node is highlighted in blue, and its sub-node 'FeatureConcept' is also highlighted.

<http://ehownet.iis.sinica.edu.tw/ehownet.php>

貓[Cats]

CN | BD | EN

• 上位关系

都市动物 猫属 宠物

• 下位关系

猫健康 虚构猫 猫引诱剂 猫粮 猫品种 猫题材作品 猫形类 各国猫

• 相关概念

饲养动物当作宠物 阿姆斯特丹博物馆 狗 宠物用品 兽医学 水族饲养 宠物食品 思想实验

• 实例

猫咪论坛 猫肉 蒙德寺 猫柜 猫爪手术 幸运土猫 三色猫 奶油猫讨论 招财猫 恐猫症 >>

• 来源

<https://zh.wikipedia.org/wiki/Category:%E8%B2%93>

https://xlore.org/search_result?search=%E7%8C%AB



Main page
Community portal
Project chat
Create a new item
Create a new Lexeme
Recent changes
Random item
Query Service
Nearby
Help
Donate
Tools

What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Concept URI

Item

house cat (Q146)

domesticated feline

cat | domestic cat | housecat | Felis silvestris catus | cats

~ In more languages

Configure

Language	Label	Description	Also known as
English	house cat	domesticated feline	cat domestic cat housecat Felis silvestris catus cats
Chinese	家猫	小型猫科动物	貓 貓咪
Traditional Chinese	家猫	小型猫科动物・野猫的亞種	貓 貓咪
Taroko	No label defined	No description defined	

All entered languages

<https://www.wikidata.org/wiki/Q146>

利用「event」讓 AI 學習關於「貓會做什麼」的知識

貓成為世界上最為廣泛的寵物之一
貓咪三餐都吃主食罐
貓通常指家貓
貓和虎本就有著高相似度
貓二字古體差異甚大
貓是最少的
貓的發情期是冬末至夏初
貓三個月大時便有地盤觀念
貓若身體不舒服
貓亦能濃縮尿液



```
'event': [],  
  ('吃', '主食罐'),  
  ('指', '家貓'),  
  ('有著', '高相似度'),  
  [],  
  [],  
  [],  
  [],  
  [],  
  ('濃縮', '尿液 ')]
```

藉由 Articulate 的 "event" (事件) 分析, 再加上以「貓」為主詞的設定, 我們很快就能分出「貓能做的事」以及「對貓的描述」兩種文本類別。



Quiz:

課堂中已經演示了文本處理中，利用「事件」進行文本分析的許多用法。

1. 請回想你在中學時遇到的數學應用問題，如果能妥善地定義每一種描述計算步驟的事件句型，是否有可能打造一個「泛用數學問題求解器」？
 - a. 若可能，為什麼？若不可能，為什麼？
2. 試說明前幾前課堂中使用的 `.getVerbStemLIST()` 和今天的 "event" 有何異同
 - a. `.getVerbStemLIST()` 依賴什麼而定？
 - b. event 依賴什麼而定？
 - c. 哪一種的正確性較高？哪一種的資訊密度較高？

Assignment: 小組作業, 每組繳一份至你們的「組名目錄」即可

1. 從課程 github repo 中把課程中提供的 week12 的目錄 git pull 下來。
2. 把 week12.py 改名為 **week12_分組隊名.py**
3. 在 **week12_分組隊名.py** 中, 設計你的程式, 利用 Articut 完成以下指定規格:
 - a. 把 "text.txt" 內容用來產生關於「倉鼠」的知識。
 - b. 把 "text.txt" 的內容取出動詞, 比較一下和 a 有何不同
4. 請在 wikipedia 中搜尋「皇帝企鵝」, 並利用 Articut 的 lv3 event 功能建立關於「皇帝企鵝」的知識。
5. 將步驟 3 和 4 做出的結果儲存為 week12_分組隊名.json, 分別存入 "倉鼠" 和 "皇帝企鵝" 兩個條目中。e.g., {"倉鼠": [("吃", "瓜子"), ("使用", "運動輪")], "皇帝企鵝": [("吃", "魚")...]}