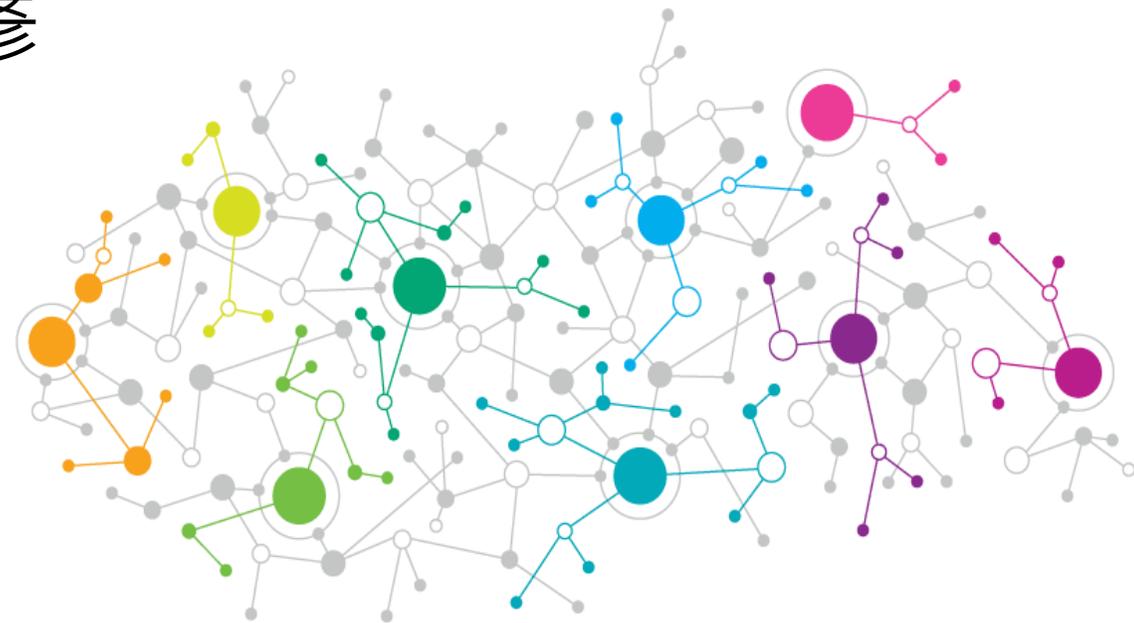


裁判書的深度解析 - 從裁判書到知識圖譜

藍景彥



Outline

- Introduction
- 1st Legaltech Hackathon
- 2nd Legaltech Hackathon
- Graph DB & Knowledge Graph
- Demo

Introduction

分享簡介

- 介紹兩屆法律黑客松，我們在解析裁判書上所運用的技術以及最後落地的成品
- 進一步介紹Graph DB以及知識圖譜，透過概念的介紹以及我們的發想，提供運用的場域以及實際上的Demo實作



1st Legaltech Hackathone

目前資料

- 各級法院民刑事裁判書
- 法律常用詞彙
- 法條

- 架構使用ElasticSearch，可以
提供便捷的查詢提取資料

green	open	臺灣新竹地方法院	sJqedkP6Rdufwir1K6SanQ	1	0	256510	0	1.8gb	1.8gb
yellow	open	臺灣高等法院高雄分院	r6fie0TsT0mwdckKxswTrQ	1	1	86676	0	1.9gb	1.9gb
yellow	open	news	tUSel3aRRwuYR8pTxpYAJg	1	1	0	0	283b	283b
green	open	臺灣苗栗地方法院	SNZoD0ywTLGlc3_HwXnFpQ	1	0	156921	0	980.9mb	980.9mb
yellow	open	臺灣橋頭地方法院	viNmU4hCRL-0kc3n53Tj4g	1	1	93334	0	539.7mb	539.7mb
green	open	臺灣基隆地方法院	awB0tgHsT0i7zxE9WzHuIQ	1	0	249000	0	1.4gb	1.4gb
green	open	臺灣彰化地方法院	_4IqFrAfQaCGcuYQDU3LFQ	1	0	435947	0	2.3gb	2.3gb
yellow	open	法律常用詞彙	gGM10IG9Rg6ZFq1DSTPWLw	1	1	2206	0	2.3mb	2.3mb
yellow	open	臺灣澎湖地方法院	suXZS8qnSUGJ017Z2ZASwg	1	1	19997	0	104.2mb	104.2mb
yellow	open	福建連江地方法院	dgdT3vsURGGwudCia8tv9g	1	1	1411	0	15.1mb	15.1mb
yellow	open	福建金門地方法院	R7_P3TFvSp2uAtKAoxi50A	1	1	22130	0	131.9mb	131.9mb
yellow	open	福建高等法院金門分院	LyiULJqpSKy21LoS0wCilQ	1	1	1752	0	60mb	60mb
green	open	臺灣臺北地方法院	mBmXVLqySfa_zw9sJ-bCWg	1	0	773487	0	5.8gb	5.8gb
green	open	臺灣臺中地方法院	T9GjavaWSwus332DCYmVEA	1	0	898861	0	6.2gb	6.2gb
green	open	臺灣嘉義地方法院	QdlLbphxTnq0FfbrBudjhw	1	0	214604	0	1.4gb	1.4gb
yellow	open	臺灣高雄少年及家事法院	U47g-C2BSd6K1Zn8AJURZA	1	1	18856	0	144.7mb	144.7mb
yellow	open	臺灣高等法院臺中分院	5vAIkzNqSrehR_zG4Mhs3A	1	1	101439	0	2.9gb	2.9gb
green	open	臺灣雲林地方法院	3agVXhA_S4Wzskh8yVgecw	1	0	185765	0	1.1gb	1.1gb
green	open	臺灣桃園地方法院	ng-awY73Rua3ZYyai4clzg	1	0	768954	0	4.7gb	4.7gb
yellow	open	臺灣高等法院	DMThIm-JREqVXtrJtgW9Dw	1	1	215138	0	5.6gb	5.6gb
green	open	臺灣高雄地方法院	1Qm38QjVRDeqvsspP7reUQ	1	0	1196709	0	5.7gb	5.7gb
green	open	臺灣屏東地方法院	zyJbubgxTPeLRt5iEYE44g	1	0	286027	0	1.6gb	1.6gb
yellow	open	最高法院	OGSsVpBDT02p-s5DvV7E5Q	1	1	137661	0	1.3gb	1.3gb
green	open	臺灣臺東地方法院	MX720J8pQUid1Q8GBvf8qg	1	0	119396	0	591.1mb	591.1mb
green	open	臺灣宜蘭地方法院	1jh3oZP1SzaHiw0wKDoYrg	1	0	164410	0	873.1mb	873.1mb
green	open	臺灣南投地方法院	AGPvX1VmSuuZ5Y2maU_HaQ	1	0	128068	0	1022.4mb	1022.4mb
green	open	臺灣高等法院臺南分院	SUPq01PZRtKWHUH_KpIHwg	1	0	61903	0	1.5gb	1.5gb
green	open	臺灣高等法院花蓮分院	Pdd8UFh1Tj2m5KRzHeuPA	1	0	15066	0	419.8mb	419.8mb
green	open	臺灣新北地方法院	3KBBRl10TGuuUJAcdduEWw	1	0	897178	0	5.8gb	5.8gb
yellow	open	智慧財產法院	rUy9SfihTiaBwbwryyHsiw	1	1	7491	0	292mb	292mb
green	open	臺灣士林地方法院	gqV6BsdmQQqe0kqHXWj15w	1	0	538881	0	2.7gb	2.7gb
green	open	臺灣臺南地方法院	RnWNq404TISGCCk60Wzq0g	1	0	628288	0	3.3gb	3.3gb
green	open	臺灣花蓮地方法院	0qTRoK9SSZ2iZmgWHhmP6A	1	0	114631	0	821.9mb	821.9mb
yellow	open	司法院 - 刑事補償	fPF6c2hZQyy-7U0j-ayZfA	1	1	2073	0	13.1mb	13.1mb

服務發想

- 裁判書包含民事、刑事、行政等訴訟判決，同時也包含裁定
 - 以往的裁判書，通常都是律師、法官或從事、就讀法律相關人員才會使用
 - 法律相關人士，基本上使用作為輔助資訊，因此常常會需要進行查詢以及擷取部分重點閱讀
 - 除法律相關人士，一般民眾通常難以有效閱讀判決書
-
- 加強裁判書的搜尋以及擷取解析 - 提供法律相關人士更精準的查詢及使用
 - 使一般民眾也能夠有效利用裁判書 - 一般常民的使用情境



「有一天我...，對方說要告我...」 常民法律知識需求情境



情境一

已知案件類別
不會檢索系統

情境二

僅有情境自述
不知案件類別

常民四不行

不熟關鍵字

污辱、侮辱、誹謗
酒駕、酒後駕車

不懂判決書結構

資料探勘老半天才
看懂文白交雜結構

不明判決要件

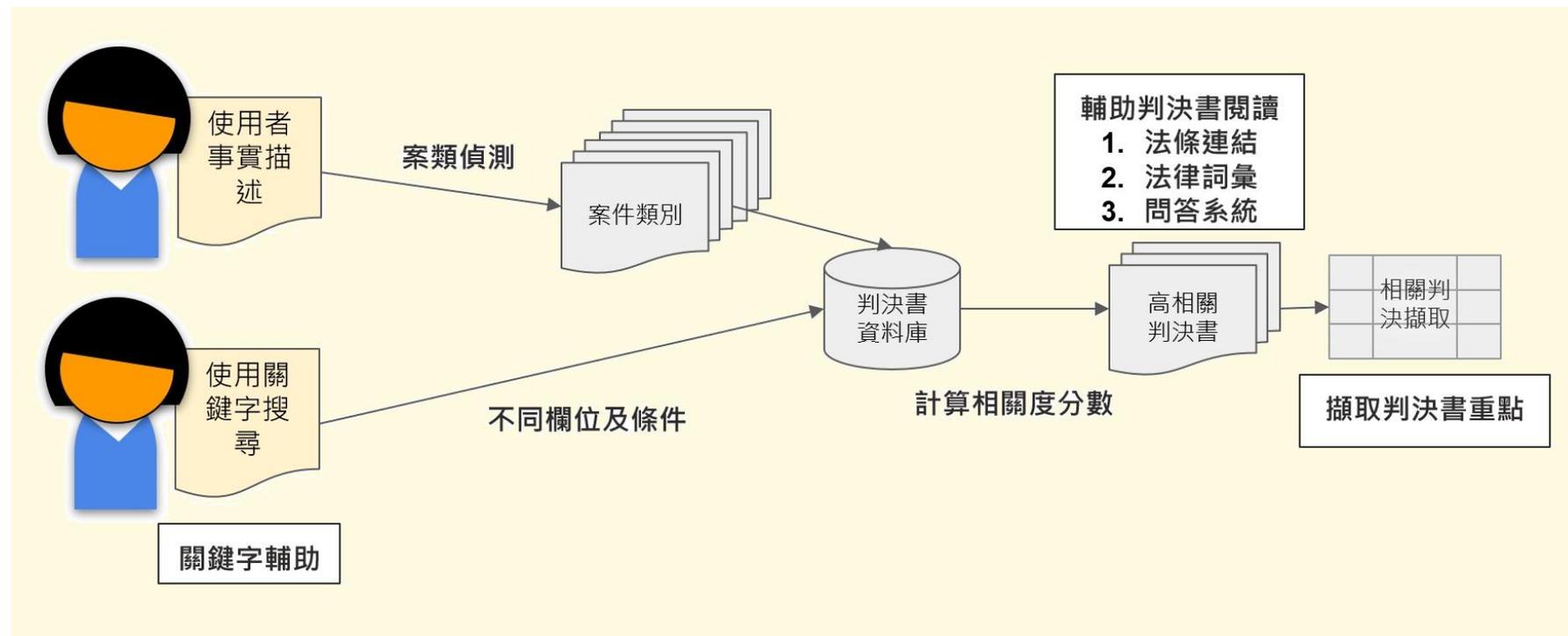
我們也只有法律系的
隊友熟，很痛苦

不知檢索系統邏輯

看來得有點先備知
識才知道怎麼查

服務架構

- 分成兩個類型的使用者
 - 一般常民
 - 法律相關人士





JudgeTalk

智慧型裁判書對話機器人

仿法律諮詢問答機器人

自然語言口語白話檢索

輔助視覺化引導

不熟關鍵字

污辱、侮辱、誹謗
酒駕、酒後駕車

不懂判決書結構

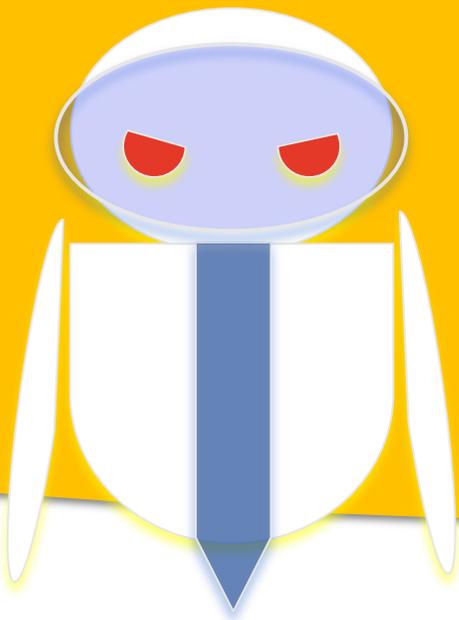
資料探勘老半天才
看懂文白交雜結構

不明判決要件

我們也只有法律系的
隊友熟，很痛苦

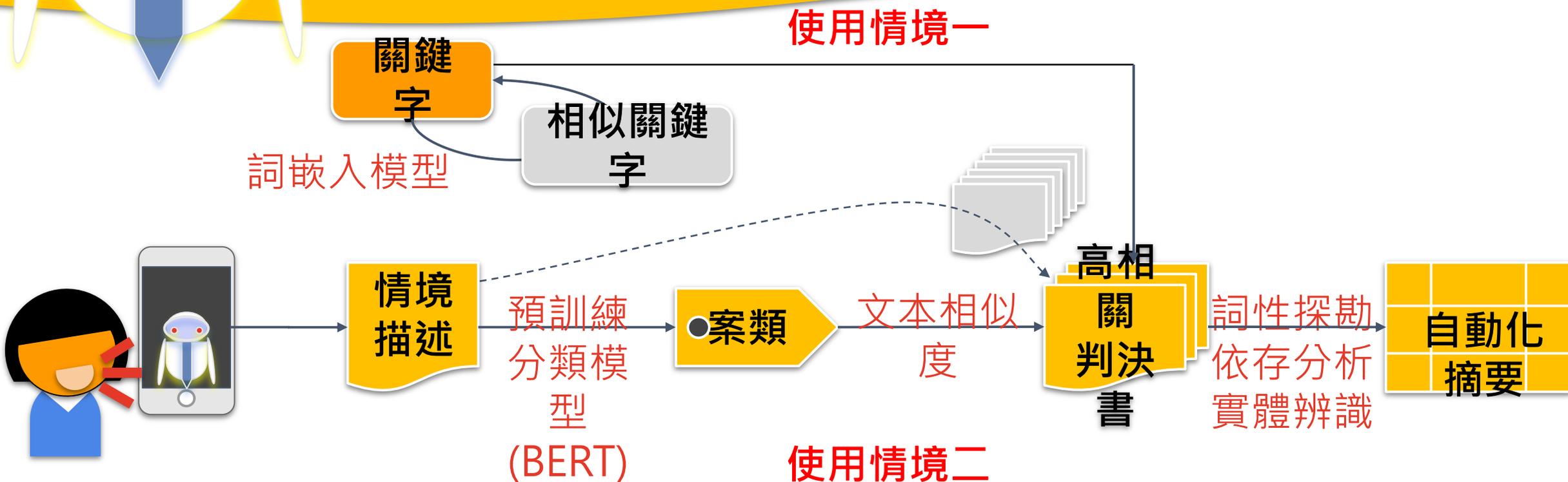
不知檢索系統邏輯

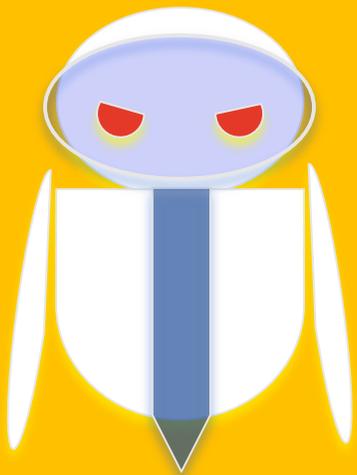
看來得有點先備知
識才知道怎麼查



JudgeTalk 系統架構

智慧型裁判書對話機器人





JudgeTalk 查詢模式一

關鍵字推薦 → 判決書閱覽

資料庫列表

裁判書查詢

事實釐清小助手

隨機裁判書標記

污辱

按日期排序

按相關度排序

相似關鍵字助手

污辱 侮辱 羞辱 污穢 汙穢 辱罵 貶抑 辱

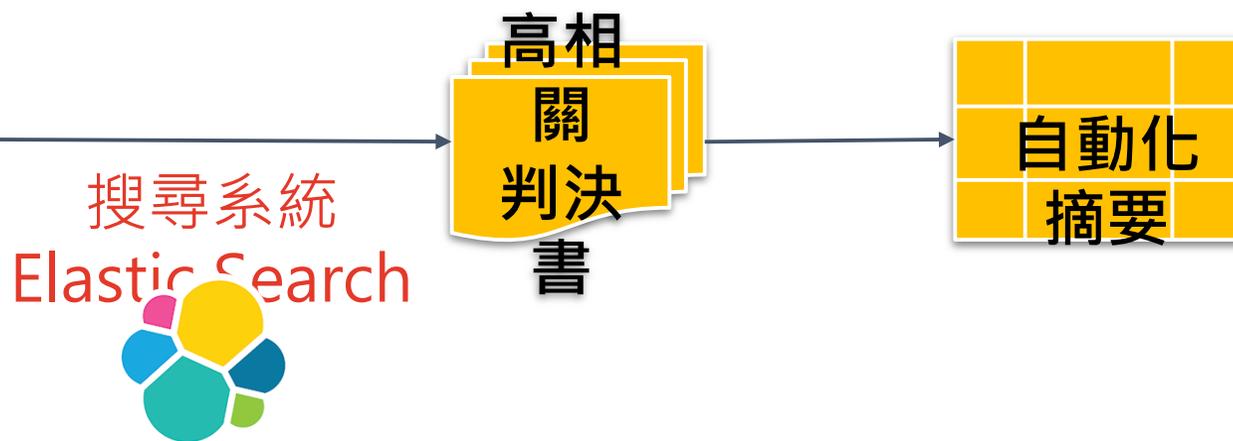
資料庫列表 裁判書查詢 事實釐清小助手 隨機裁判書標記

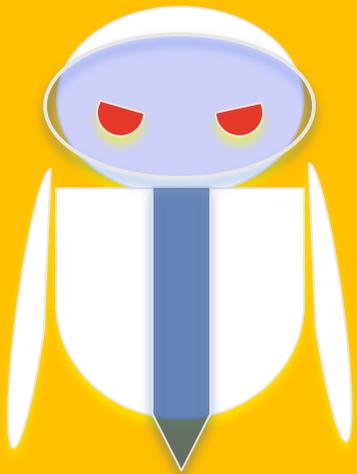
請輸入搜尋文字

按日期排序

按相關度排序

相似關鍵字助手





JudgeTalk 語音白話文搜尋 案類預測 + 搜尋最相關判決

法律諮詢小幫手

1:1
昨天臨檢的時候我因為不爽罵了警察

1:1
根據我的判斷，你的這個情況，屬於：妨害公務

1:1
這邊提供給您相關的裁判書參考，請稍等

1:1
104,易,178 妨害公務
102,簡,4147 妨害公務
102,簡上,631 妨害公務
103,簡上,513 妨害公務
99,簡上,954 妨害公務

- 裁判書全文
- 相關法條
- 難字解釋彈出視窗

相關法條

- 中華民國刑法 140 +
- 中華民國刑法 140 1 +
- 刑事訴訟法 159-2 +
- 中華民國刑法 309 +
- 刑事訴訟法 159-1 +
- 刑事訴訟法 159-4 +
- 中華民國刑法 309 1 +
- 中華民國刑法 41 1 +
- 中華民國刑法 55 +
- 刑事訴訟法 159-5 1 +
- 刑事訴訟法 299 1 +
- 中華民國刑法施行法 1-1 2 +

裁判字號：104,易,178

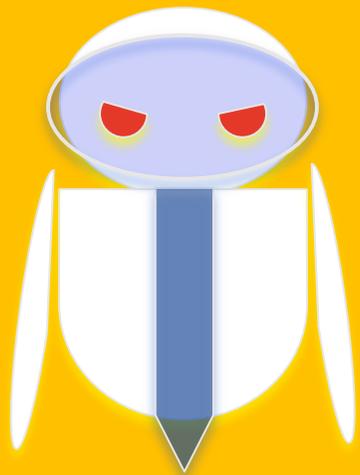
裁判日期：2015-04-27T00:00:00+08:00
裁判類型：刑事
裁判案由：妨害公務
裁判內文：
臺灣臺南地方法院刑事判決 104年度易字第178號
公 訴 人 臺灣臺南地方法院檢察處檢察官
被 告 尤錦福
上列被告因妨害公務案，經公訴（103年度偵續字第348號），本院判決：
主 文
尤錦福犯侮辱公務員罪，處有期徒刑肆月，如易科罰金，以新臺幣壹仟元折算壹日。
事 實
一、尤錦福係臺南市○○區○○路000號臺南市政府警察局第六分局鯤鯤派出所警員（現已調任為臺南市政府警察局第六分局喜樹派出所警員），其於民國103年7月5日上午8時起至10時止、10時至12時止，分別執行值班及勤區查察勤務，並預定下午10時起至12時止，再回派出所執行值班勤務。

CCLA

資料庫列表 裁判書查詢 事實釐清小助手
隨機裁判書標記

目前僅提供 公共危險,毒品危害防制條例,竊盜,詐欺,過失傷害,偽造文書,傷害,賭博,違反毒品危害防制條例,妨害自由,侵占,違反槍砲彈藥刀械管制條例,妨害風化,妨害公務,妨害名譽 這些案例的釐清喔！

法律諮詢小幫手
您好，請問有任何法律問題需要詢問的嗎？



JudgeTalk 自動化摘要 (I)

關係人偵測 + 視覺導覽

自然語言處理方法：利用詞性與依存句法剖析提高關係人抽取準確率



資料庫列表 裁判書查詢 事實釐清小助手 隨機裁判書標記

回到上一頁 裁判書探勘結果

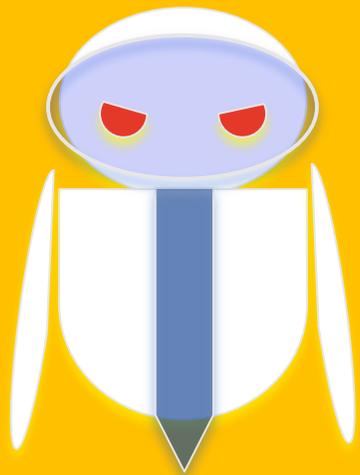
相關法條

- 刑事訴訟法 451 1 +
- 中華民國刑法施行法 1-1 1 +
- 刑事訴訟法 454 2 +
- 交通安全規則 102 1 送出
- 中華民國刑法 41 1 +
- 中華民國刑法施行法 1-1 2 +
- 刑事訴訟法 449 3 +
- 中華民國刑法 284 +

裁判字號：104,交簡,4326

裁判日期：2016-07-29T00:00:00+08:00
裁判類型：刑事
裁判案由：過失傷害
裁判內文：
臺灣新北地方法院刑事簡易判決 104年度交簡字第4326號
聲請人 臺灣新北地方法院檢察署檢察官
被告 蘇郁婷
選任辯護人 楊凱吉律師（法律扶助律師）
上列被告因過失傷害案件，經檢察官聲請以簡易判決處刑（104年度調偵字第1953號），本院判決如下：
主文
蘇郁婷犯業務過失傷害罪，處有期徒刑參月，如易科罰金，以新臺幣壹仟元折算壹日。
事實及理由

法律諮詢小幫手 14:48
您好，請問有任何法律問題需要詢問的嗎？
送出



JudgeTalk 自動化摘要 (II)

裁判書結構拆解 + 視覺導覽

裁判書深訪 (律師建議功能)

- 判決事實抽取
- 判決理由抽取
- 段落切割與搜尋
- 詞性標註與選取

裁判字號：104,交簡,4326

裁判日期：2016-07-29T00:00:00+08:00

裁判類型：刑事

裁判案由：過失傷害

裁判內文：

臺灣新北地方法院刑事簡易判決 104年度交簡字第4326號

聲 請 人 臺灣新北地方法院檢察署檢察官

被 告 蘇郁婷

搜尋詞彙 辯護人 楊凱吉律師 (法律扶助律師)

送出 被告因過失傷害案件，經檢察官聲請以簡易判決處刑 (104年度調偵字第1953號)，本院判決如下：

主 文

蘇郁婷犯業務過失傷害罪，處有期徒刑參月，如易科罰金，以新臺幣壹仟元折算壹日。

事實及理由

一、本件犯罪事實及證據，除下列事項應補充外，其餘均引用如附件檢察官聲請簡易判決處刑書之記載：

(一) 犯罪事實欄一、第10行所載「等傷害」前應補充「左側脛

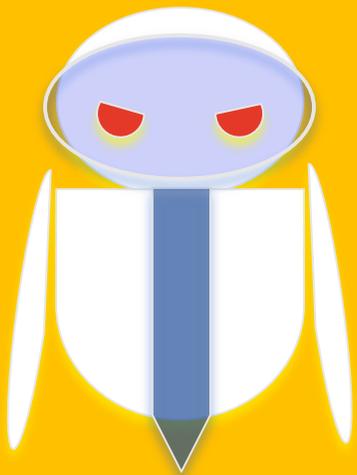
法律諮詢小幫手

14:21



您好，請問有任何法律問題需要詢問的嗎？

送出



JudgeTalk 白話文問答與互動 裁判事實問答模型

資料庫列表

裁判書查詢

事實釐清小助手

隨機裁判書標記

相關法條

中華民國刑法 284 2 +

中華民國刑法 41 1 +

刑事訴訟法 449 3 +

中華民國刑法施行法 1-1

2 +

刑事訴訟法 451 1 +

送出 道路交通安全規則 102 1

1 +

刑事訴訟法 454 2 +

刑事訴訟法 449 1 +

道路交通安全規則 90

+

道路交通安全規則 94 3

+

裁判字號：104,交簡,4326

裁判日期：2016-07-29T00:00:00+08:00

裁判類型：刑事

裁判案由：過失傷害

裁判內文：

臺灣新北地方法院刑事簡易判決 104年度交簡字第4326號

聲請人 臺灣新北地方法院檢察署檢察官

被告 蘇郁婷

選任辯護人 楊凱吉律師（法律扶助律師）

上列被告因過失傷害案件，經檢察官聲請以簡易判決處刑（104年度調偵字第1953號），本院判決如下：

主文

蘇郁婷犯業務過失傷害罪，處有期徒刑參月，如易科罰金，以新臺幣壹仟元折算壹日。

事實及理由

一、本件犯罪事實及證據，除下列事項應補充外，其餘均引用如附件檢察官聲請簡易判決處刑書之記載：

（一）犯罪事實欄一、第10行所載「等傷害」前應補充「左側脛骨平台粉碎性骨折合併軟組織受損、左髕關節軟組織挫傷

回到上一頁

裁判書探勘結果

法律諮詢小幫手

14:15

您好，請問有任何法律問題需要詢問的嗎？

他用了什麼毒品

16:25

第二級毒品

第二級毒品是什麼

16:25

甲基安非他命殘渣

檢察官是誰

14:24

許恭仁

法官是誰

14:24

劉瓊雯

送出

請問誰是被告

1:17

黃月惠

是誰犯了罪呢？

1:17

黃月惠

這個案子裡檢察官起訴了誰呢？

1:18

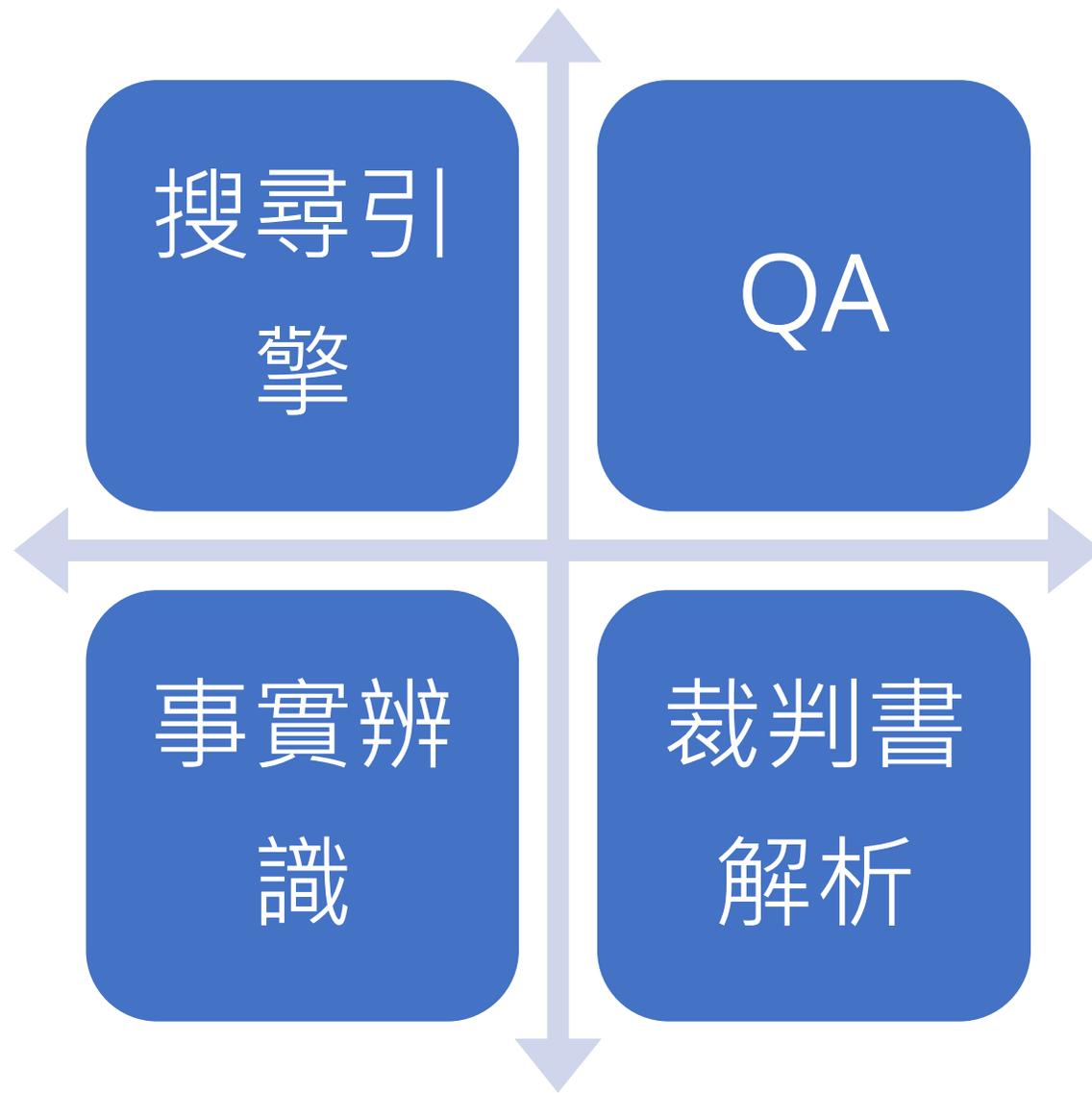
黃月惠

是誰被判刑？

1:20

黃月惠

處理流程



搜尋引擎

- 建構Elastic Search架構
- 建立Index存放裁判書
- 設定elasticsearch(Cluster、搜尋權重.....)
- 建立Cache協助查詢加速

green	open	臺灣新竹地方法院	sJqedkP6Rdufir1K6SanQ	1	0	256510	0	1.8gb	1.8gb
yellow	open	臺灣高等法院高雄分院	r6fie0TsT0mwdckKxswTrQ	1	1	86676	0	1.9gb	1.9gb
yellow	open	news	tuSel3aRRwuYR8pTxxpYAJg	1	1	0	0	283b	283b
green	open	臺灣苗栗地方法院	SNZoD0ywtLGlc3_HwXnFpQ	1	0	156921	0	980.9mb	980.9mb
yellow	open	臺灣板橋地方法院	viNmU4hCRL-0kc3n53Tj4g	1	1	93334	0	539.7mb	539.7mb
green	open	臺灣基隆地方法院	awB0tgHsT0i7zxE9WzHuIQ	1	0	249000	0	1.4gb	1.4gb
green	open	臺灣彰化地方法院	_4IqFrAfQaCgcuYQDU3LFQ	1	0	435947	0	2.3gb	2.3gb
yellow	open	法律常用詞彙	gM10IG9Rg6ZFq1DSTPWLw	1	1	2206	0	2.3mb	2.3mb
yellow	open	臺灣澎湖地方法院	suXZS8qnSUGJ017Z2ZASwg	1	1	19997	0	104.2mb	104.2mb
yellow	open	福建連江地方法院	dgdT3vsURGwudCia8tv9g	1	1	1411	0	15.1mb	15.1mb
yellow	open	福建金門地方法院	R7_P3TFvSp2uAtKAoxi50A	1	1	22130	0	131.9mb	131.9mb
yellow	open	福建高等法院金門分院	LyiuLJqp5Ky21LoS0wCiIQ	1	1	1752	0	60mb	60mb
green	open	臺灣臺北地方法院	mBmXVLqySfa_zw9s3-bCWg	1	0	773487	0	5.8gb	5.8gb
green	open	臺灣臺中地方法院	T9GjavaWswus33ZDCYmVEA	1	0	898861	0	6.2gb	6.2gb
green	open	臺灣嘉義地方法院	Qd1LbphxTnq0FfbrBudjhw	1	0	214604	0	1.4gb	1.4gb
yellow	open	臺灣高雄少年及家事法院	U47g-C2B5d6K1zn8AJURZA	1	1	18856	0	144.7mb	144.7mb
yellow	open	臺灣高等法院臺中分院	5vAikzNqSrehR_z6Mhs3A	1	1	101439	0	2.9gb	2.9gb
green	open	臺灣雲林地方法院	3agVXhA_S4Wzskh8yVgecw	1	0	185765	0	1.1gb	1.1gb
green	open	臺灣桃園地方法院	ng-awY73Rua32Yyai4clzj	1	0	768954	0	4.7gb	4.7gb
yellow	open	臺灣高等法院	DMThIm-JREqVXtrJtgW9Dw	1	1	215138	0	5.6gb	5.6gb
green	open	臺灣高雄地方法院	1Qm38QjVRDeqvsspP7rneUQ	1	0	1196709	0	5.7gb	5.7gb
green	open	臺灣屏東地方法院	zyJbubgxTpeLrt5iEYE44g	1	0	286027	0	1.6gb	1.6gb
yellow	open	最高法院	OGSsVpB0T02p-s5DvV7E5Q	1	1	137661	0	1.3gb	1.3gb
green	open	臺灣臺東地方法院	MX720J8pQUidIQ8GbvF8qg	1	0	119396	0	591.1mb	591.1mb
green	open	臺灣宜蘭地方法院	1jh3oZP1SzaHiwOwKDoYrg	1	0	164410	0	873.1mb	873.1mb
green	open	臺灣南投地方法院	AGPvX1VmSuuZ52maU_HaQ	1	0	128068	0	1022.4mb	1022.4mb
green	open	臺灣高等法院臺南分院	SUPq01PZrtKWHUH_KpIHwg	1	0	61903	0	1.5gb	1.5gb
green	open	臺灣高等法院花蓮分院	Pdd8UFh1Tj2mM5KRzHeuPA	1	0	15066	0	419.8mb	419.8mb
green	open	臺灣新北地方法院	3KBBR1LOTGuuUJAcdduEkw	1	0	897178	0	5.8gb	5.8gb
yellow	open	智慧財產法院	rUy9SfihTiaBwbwryyHsiw	1	1	7491	0	292mb	292mb
green	open	臺灣士林地方法院	gqV6BsdmQQqe0kqHXWj15w	1	0	538881	0	2.7gb	2.7gb
green	open	臺灣臺南地方法院	RnWnq404TISGCK60WzqOg	1	0	628288	0	3.3gb	3.3gb
green	open	臺灣花蓮地方法院	0qTRoK9SSZ2iZmgWlhmP6A	1	0	114631	0	821.9mb	821.9mb
yellow	open	司法院—刑事補償	fPF6c2H2Qyy-7U0j-ayZfa	1	1	2073	0	13.1mb	13.1mb

```
{
  "court": "最高法院",
  "date": "1996-03-21T00:00:0000",
  "no": "85,台上,611",
  "sys": "民事",
  "reason": "禁止地上權登記",
  "judgement": "最高法院民事判決...判決如主文...",
  "type": "判決",
  "historyHash": "bc09e3da6e9e78b4908b29beac0667fb",
  "mainText": "原判決廢棄，發回台灣高等法院高雄分院。",
  "opinion": "原審維持第一審所為上訴人敗訴判決，駁回其上訴，無非以：上訴人起訴...據上論結，本件上訴為有理由。依民事訴訟法第四百七十七條第一項、第四百七十八條第一項，判決如主文。",
  "argument": "被上訴人則以：伊聲請地...據為上訴人不利之判斷，亦有判決不備理由之違法。上訴論旨，執以指摘原判決不當，求予廢棄，非無理由。",
  "relatedIssues": [
    { "lawName": "土地登記規則", "issueRef": "28 1 7" },
    { "lawName": "土地登記規則", "issueRef": "51 1 3" },
    { "lawName": "民法", "issueRef": "767" }
  ],
  "party": [
    {
      "group": ["plaintiff"],
      "title": "上訴人",
      "value": "甲○○"
    },
    {
      "group": ["plaintiff", "agentAdLitem", "lawyer"],
      "title": "訴訟代理人",
      "value": "王程風律師"
    },
    {
      "group": ["defendant"],
      "title": "被上訴人",
      "value": "乙○○"
    }
  ]
}
```



elasticsearch

QA

- 設定Question Answer不同目的
- 訓練Answer-Span模型(BERT)透過網路上資料集，以及使用者所標記之問答
- 解析文字模型(將使用者Input進行模型適用)
- 前端問答流程邏輯(Javascript、React)
- 前端呈現介面



事實辨識

- 裁判書前處理(透過文字探勘及Regex解析出事實)以及其相對應的罪名
- 訓練分類任務模型(BERT) , Input為審判事實 , Output為相對應罪名
- 模型預測API
- 前端介面呈現

法律諮詢小幫手

您好，請問有任何法律問題需要詢問的嗎？

請在下面欄位簡單描述您所面臨的狀況（字數請盡量不要超過500字）

昨天走在路上，看到有人拿槍

根據我的判斷，你的這個情況，屬於：違反槍砲彈藥刀械管制條例

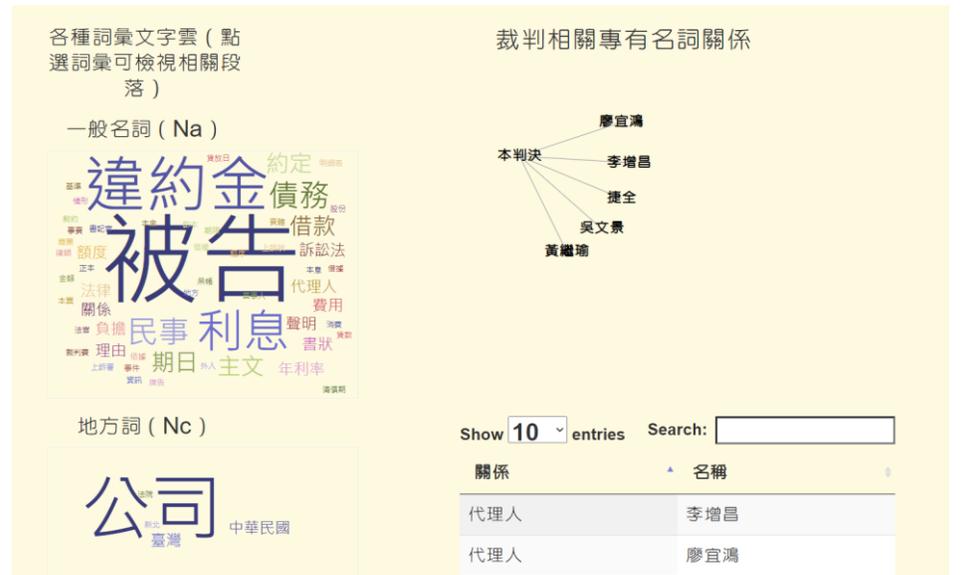
這邊提供給您相關的裁判書參考，請稍等

- 100, 訴緝, 30 違反槍砲彈藥刀械管制條例
- 100, 重訴, 16 違反槍砲彈藥刀械管制條例
- 97, 訴緝, 56 違反槍砲彈藥刀械管制條例
- 97, 訴, 807 違反槍砲彈藥刀械管制條例
- 102, 訴, 459 違反槍砲彈藥刀械管制條例

送出

審判書解析

- 文字處理相關模型API(詞性斷詞)
- Event Extraction(解析出裁判書中關係)
- 相關知識資料庫
 - 法律相關用語API(難解法律詞彙)
 - 法條API
- 前端呈現(Network、文字雲)使用D3.js



2nd Legaltech Hackathone

根基在上一屆的基礎

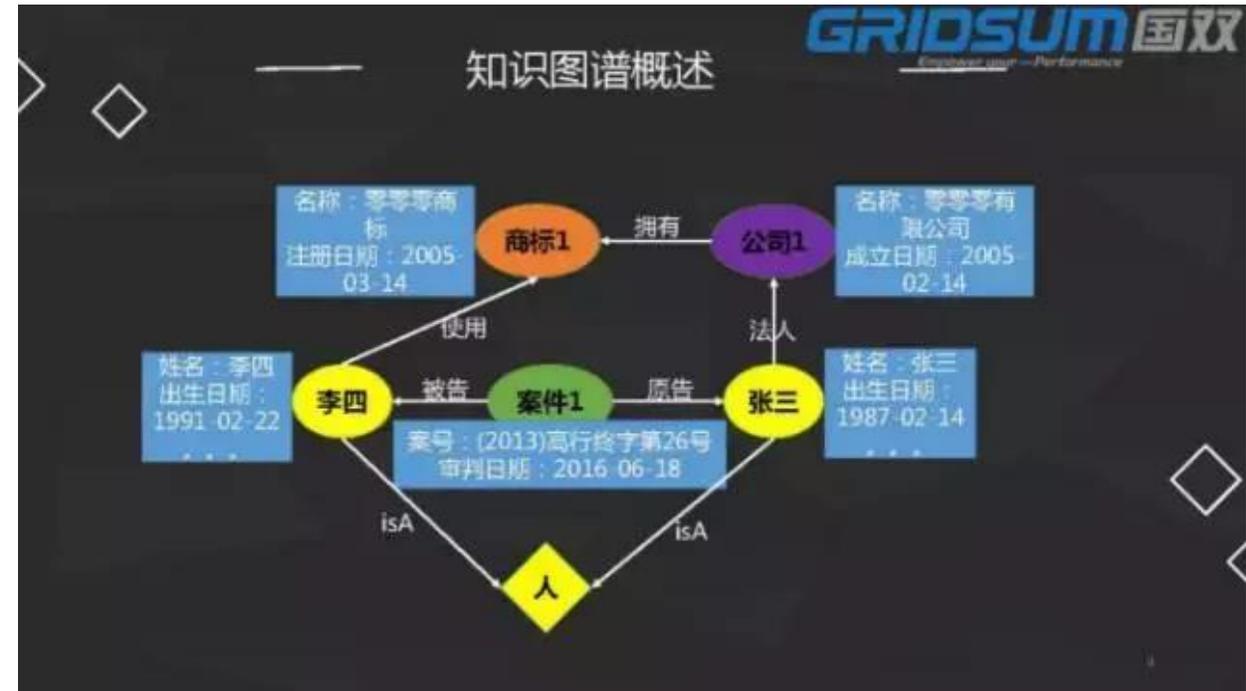
- 上次的應用主要包含幾大主軸：一、查詢 二、問答 三、分類（事實）
- 其中裁判書的解析上，僅透過法律常用辭彙的Match去做Tooltip；另外透過詞性解析出該裁判書的專有名詞人物，以及各類詞彙文字雲
- ▣ 根基在上一屆所建構的查詢及分類上，希望進一步能夠深入解析裁判書

裁判書深度解析

- 前面提到的解析，大部分都停留在詞彙層級，並沒有建構詞彙之間與知識的連結
- 希望能夠透過自動化方式解析每一篇裁判書的詞彙並進一步建構知識連結；同時透過使用者的反饋，建構更完善的知識連結
- 深度解析 - 將一則裁判書中的文字，萃取出具體的知識、事件；並進一步將其建構至Graph上，達到知識間的連結

裁判書中知識連結的意義

- 將裁判書中各詞類定義後，連結能夠使使用者更快理解該則裁判書
- 對於專業使用者，將過往知識進行連結，能夠有效減少其需要理解裁判書之時間



實體標記+詞性

書記官、被告、法官、家屬、
原告、檢察官

問答模型

抽取事實

資料庫

+ ...

↓

問題：只能抽取答案、特徵，不易做深度推論

↓

目標：從大數據方法做判決書的深度解析

↓

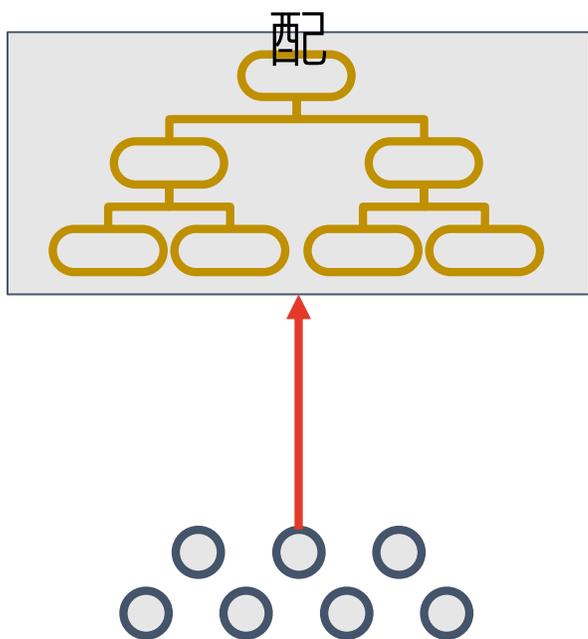
專業人員用的

判決書知識圖譜

建立圖譜的兩種途徑

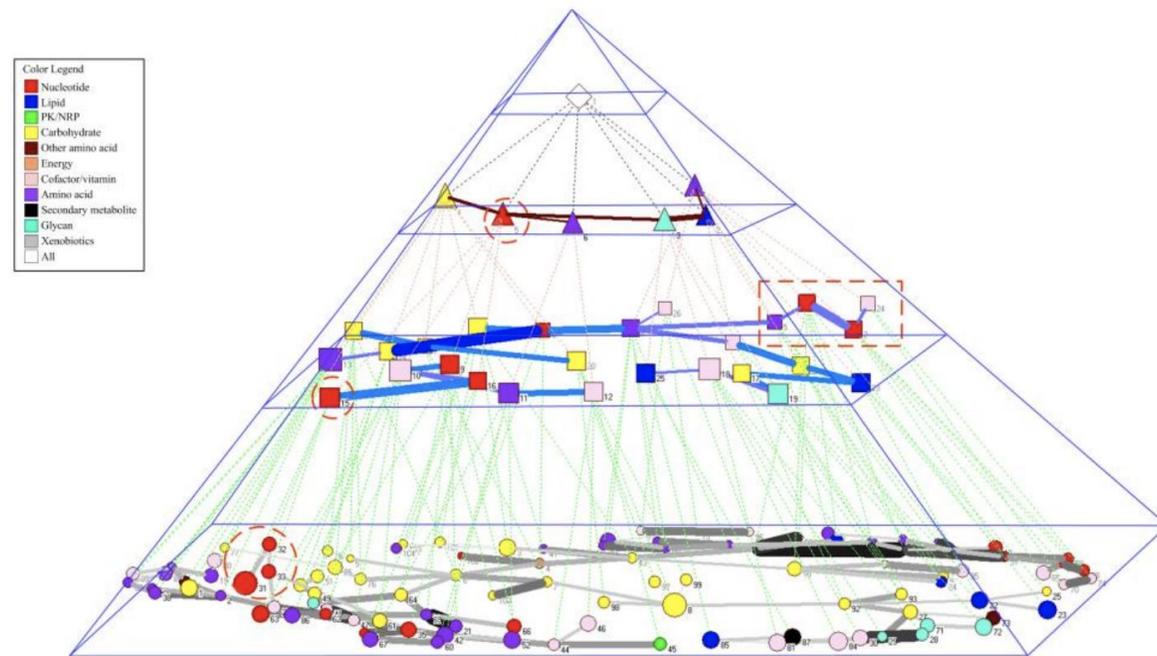
Top-down 專業優先

專家建好Schema → 萃取與匹



Bottom-up 大數據導向

句法剖析 → 建立實體關聯 → 形成網絡



網絡抽象化示意圖。取自鄭家胤 (2009)

Top-down

專業與匹配

專業內的溝通與協調
建立圖譜費心費力
訂多細？相對主觀
不同案類是否相同？

技術上的問題
如何一一匹配？

特徵	特徵描述	特徵數
Fact_tfidf (ftfidf)	標記為「事實」區段文字之 TF-IDF 值	10034
Behavior_tfidf (btfidf)	標記為「行爲」區段文字之 TF-IDF 值	2897
行爲關鍵字(bkey)	標記為「行爲」的文字中是否出現表五所列之關鍵字	23
被害人反應關鍵字(vkey)	標記為「被害人反應」的文字中是否出現表五所列之關鍵字	23
行爲人個數(actor)	1~2 人、3~5 人、6 人以上	3
行爲人特徵(afeature)	行爲人犯罪背景描述，包括無前科、有前科、有精神疾病、累犯	4
共犯個數(accomplice)	0 共犯、1~2 人、3~5 人、6 人以上	4
被害人個數(victim)	1~2 人、3~5 人、6 人以上	3
犯後態度(after)	未坦承犯行/否認犯行/未見悔意/態度不佳、坦承部分犯行、坦承犯行	3
財物(property)	分爲一千以下、一千到五千、五千到一萬、一萬到十萬、十萬到百萬、錢、百萬以上、手機電子用品、本票等文件、證件、交通工具、珠寶金飾、皮包、其他	14
犯案工具(tool)	分爲電話、錄影光碟、廣義刀械、槍、道具槍、言語行爲、交通工具、棒棍、電擊棒、縱火工具、信函、掩飾衣物、綑綁工具、衣物、藥物、金屬工具、其他	17
Local label 標記次數	各 local label 在判決書中被標記的次數，原始標記資料中若兩斷詞標同一種 label 算標一次 林琬真等 (2012)	16
Local label 標記順序	各 local label 在判決書中標記的順序	256

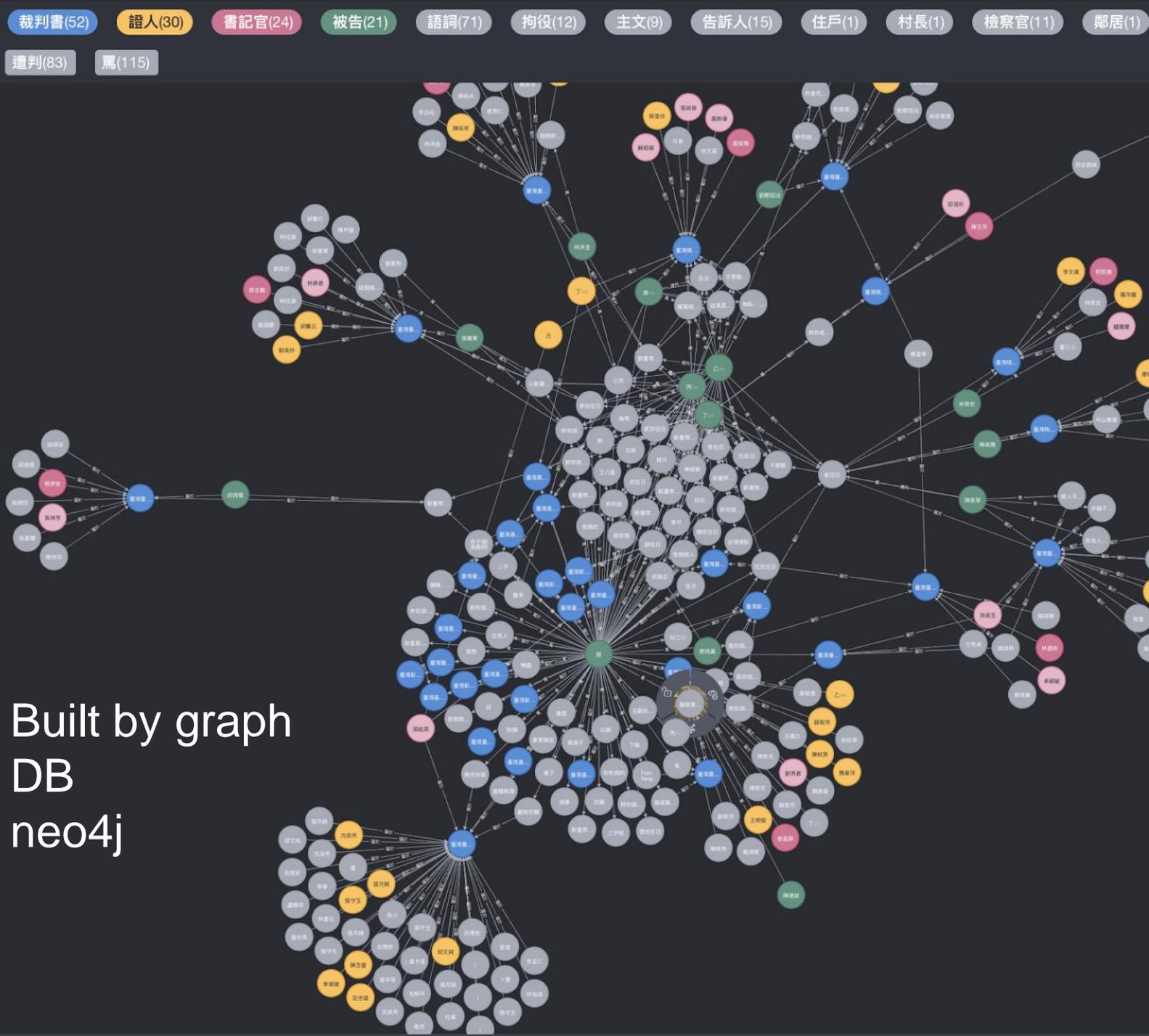
Bottom-up 自動化句法剖析

萃取出大量的特徵細節
(e.g., 欠債? 精神? 小三?)

如何連結特徵形成知識?

需要專家解析

Built by graph
DB
neo4j



人機合作 智慧型圖譜构建輔助系統

語言學知識

句法剖析、實體偵測

+排序、篩選

專家勾選要保留的關係

+設計關係階層

知識圖譜整合與比較

+跨案別、專家比較

Future

提供知識圖譜查詢

妨害名譽案類實體關係

實體標籤 名字 新增

Show 10 entries

Search:

實體標籤	次數	保留	上位標籤	下位標籤
罰金	6441	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
原告	2669	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
辯護人	2253	<input checked="" type="checkbox"/>	<input type="text"/>	正方辯護人
律師	1109	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
代理人	826	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
有期徒刑	674	<input checked="" type="checkbox"/>	刑度	<input type="text"/>
友人	446	<input checked="" type="checkbox"/>	人	摯友
配偶	426	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
委員	405	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
自訴人	395	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>

Showing 11 to 20 of 2,054 entries

Previous 1 2 3 4 5

機器輔助建立的 具上下位關係的專家知識圖譜

過失傷害 ▾

實體標籤 名字 新增 送出勾選

Show 10 ▾ entries Search:

實體標籤	次數 ▲	保留	上位標籤	下位標籤	submit
-	1	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	提交
6年級	1	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	提交
員	1	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	提交
師	1	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	提交
政	1	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	提交
輪功	1	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	提交
銷商	1	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	提交
一生	1	<input type="checkbox"/>	<input type="text"/>	<input type="text"/>	提交

2. 照次數排序

被記錄的實體上下位關係 (可供刪修)

實體標籤 ▲	上位標籤	下位標籤	remove
妻子	人	前妻	✖
拘役	判決	天數	✖
檢察官	法律從業人員	主任檢察官	✖
法官	法律從業人員	地方法院法官	✖
管理員	人	大樓管理員	✖
證人	人	關係人	✖
證人	人	關鍵證人	✖
起訴書	物		✖



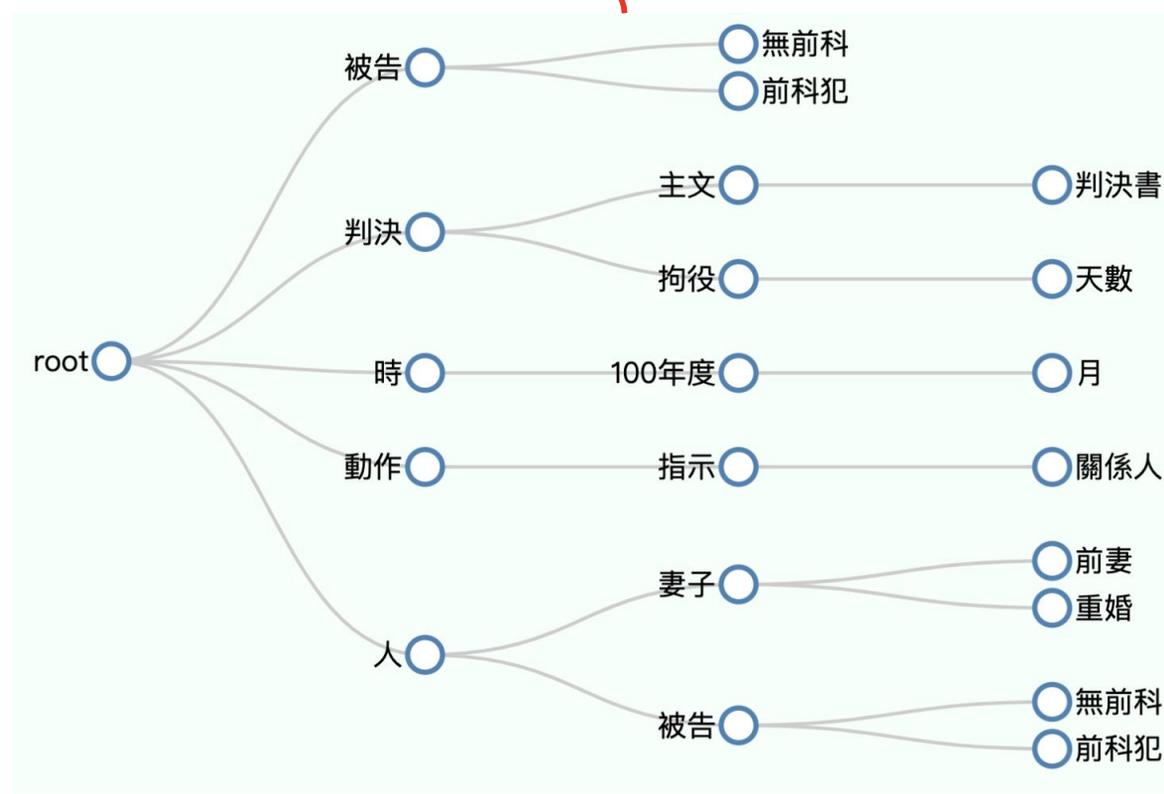
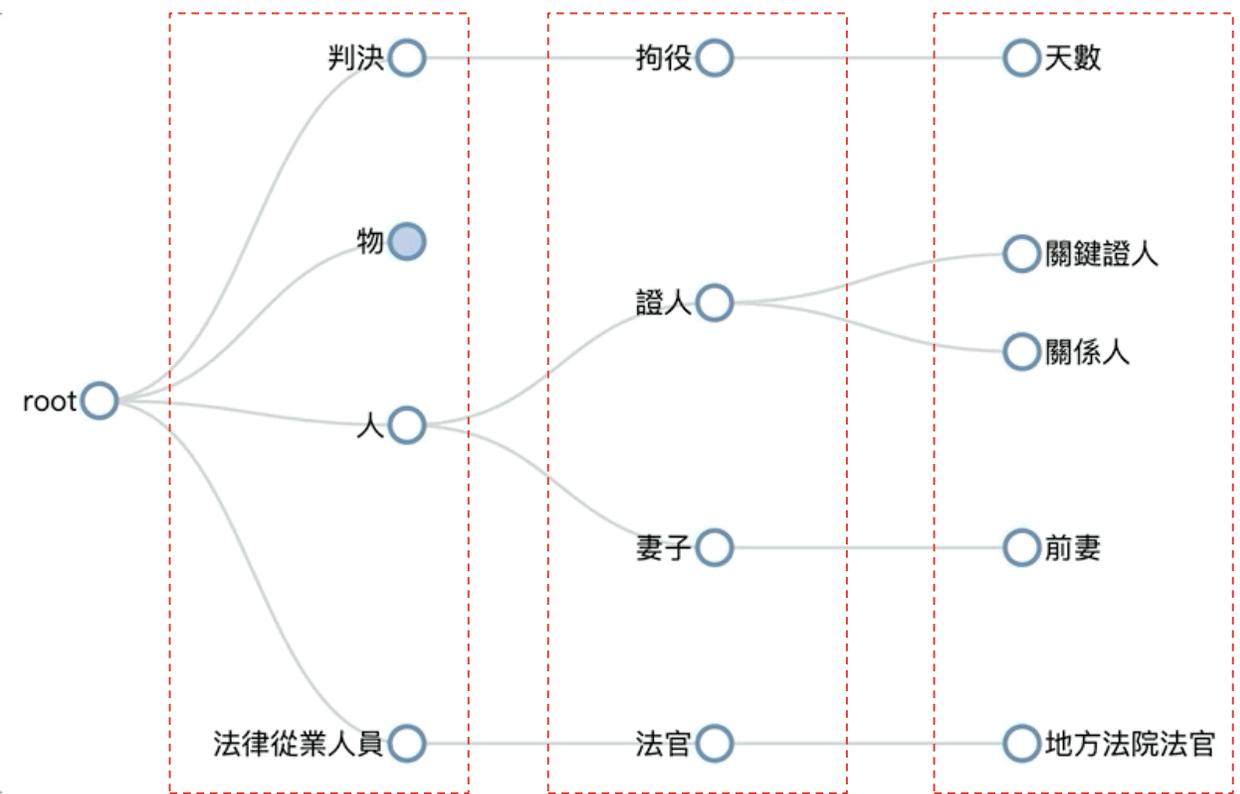
根據紀錄建立圖譜網絡以運算

上位標籤

實體標籤

下位標籤

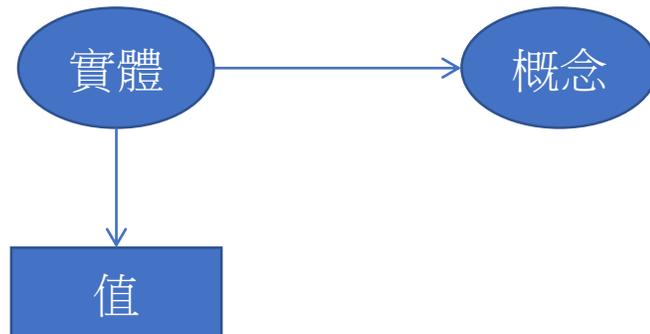
用Graph-embedding做跨圖運算



Graph DB&Knowledge
Graph

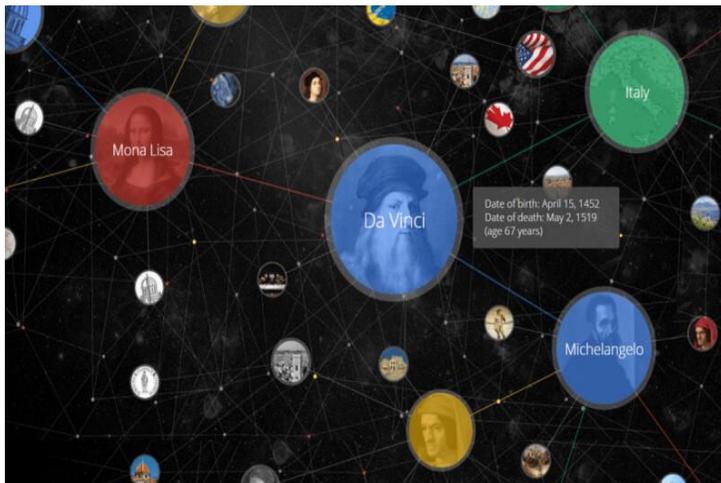
前情提要-背景

- 希望使電腦演算法展現“智能”，而智能的先決條件即是具備常識以及知識
 - 語義網絡
 - First Order Logic
 - 知識工程



前情提要-背景

- 傳統的語義網絡與知識圖譜的差異？
- 在網路的發展以及電腦運算能力加強的今天，傳統的知識工程得到了新的機會
- 建構一個巨量、可更新、質量高的大知識圖譜是一個有可能的事情(Google - 2012)

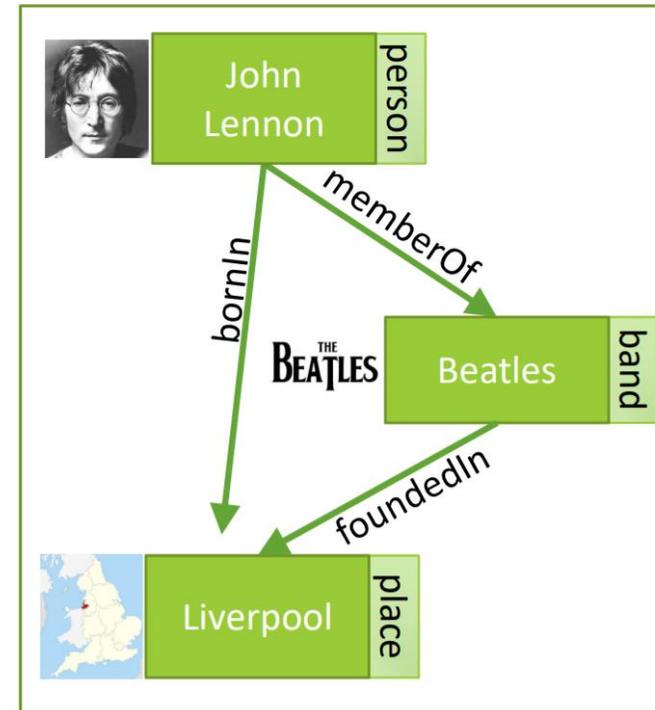
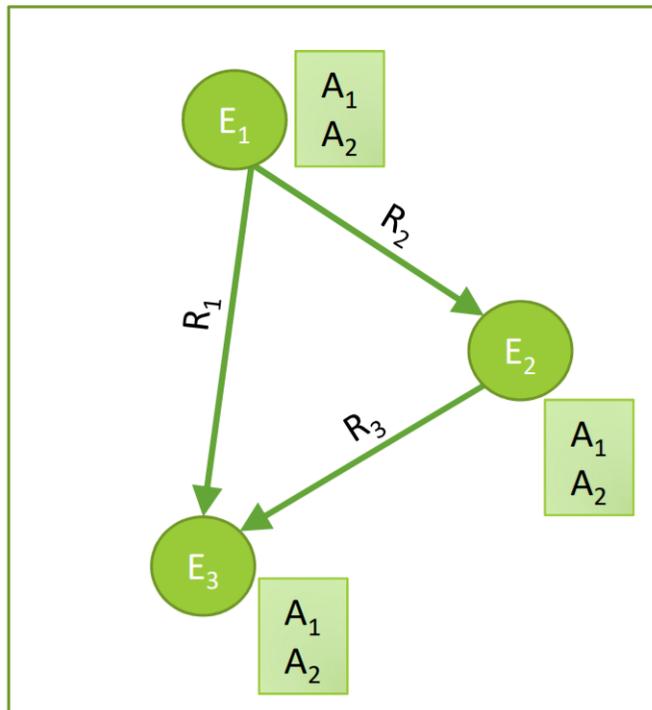


Google search results for "barry schwartz technologist". The search bar shows the query and the Google logo. Below the search bar, there are several search results. The first result is a Wikipedia entry for "Barry Schwartz (technologist)". The second result is "Barry Schwartz's Search Engine Roundtable Articles". The third result is "Barry Schwartz - Search Engine Land Author". The fourth result is "Barry Schwartz, CEO of RustyBrick". The fifth result is "Barry Schwartz - Google!". On the right side of the page, there is a profile card for "Barry Schwartz" with a photo and a bio. A red arrow points from the search results to the profile card.

前情提要-基本介紹

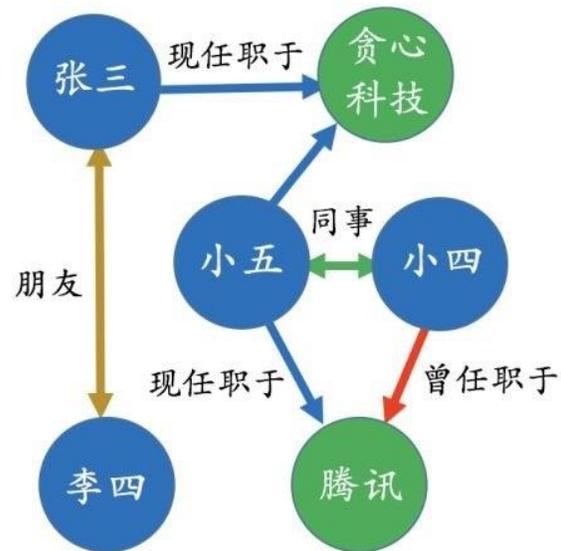
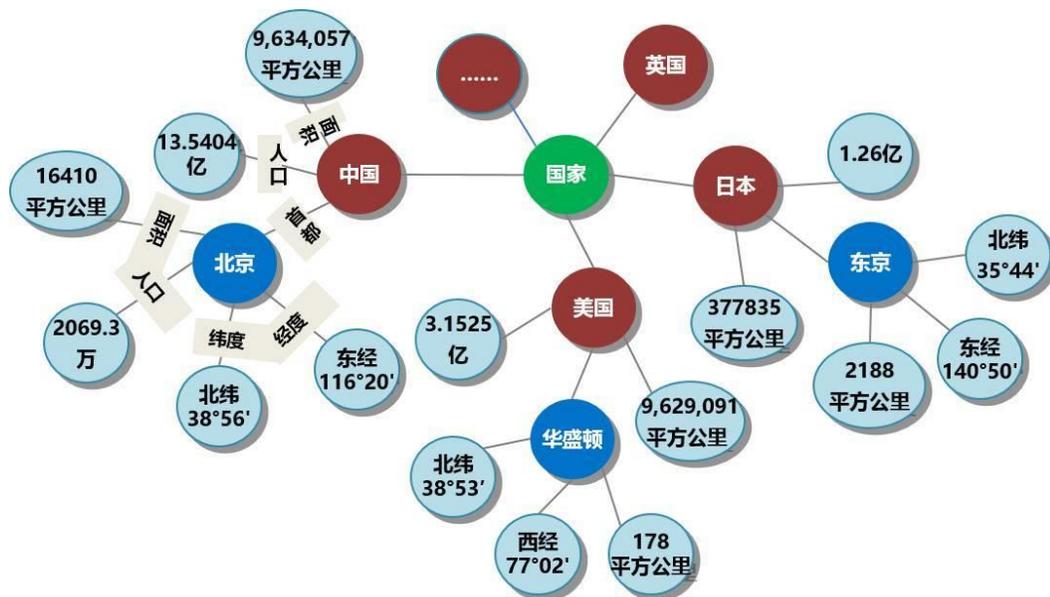
- What is Knowledge Graph?

節點(Node)、節點的屬性(Attribute)以及連線(Edge)組成



前情提要-基本介紹

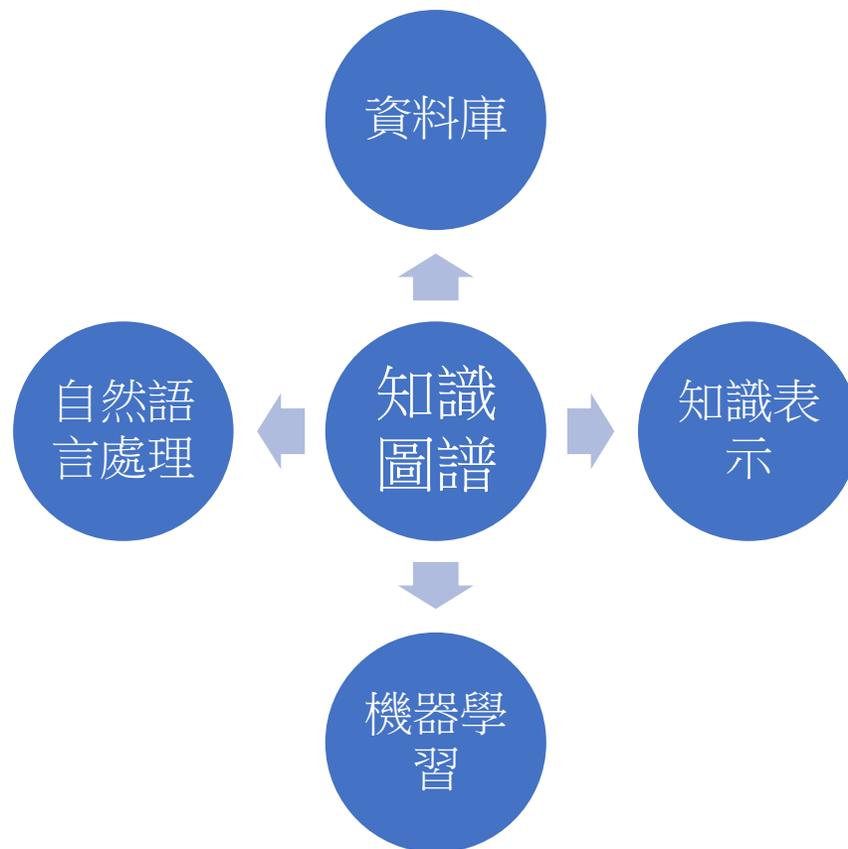
- Example Of Knowledge Graph



案例：社交网络

前情提要-基本介紹

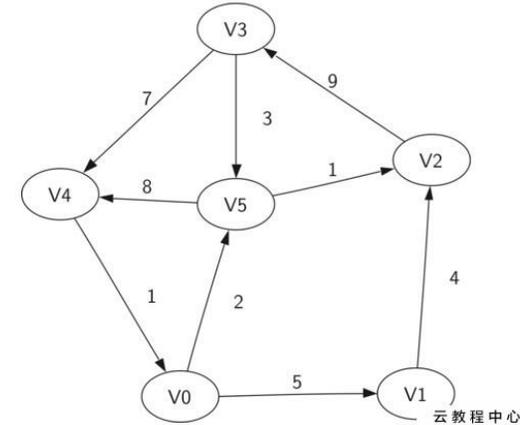
- Knowledge Graph應用到的領域技術



前情提要-知識表示

- 知識的表示意指如何將知識使用能夠讓電腦運算或存取的方式
 - 圖論
 - 描述點與點之間連線
 - 三元組
 - 描述物件的資源類型 (RDF)
 - 馬可夫鏈
 - 在點與點之間狀態轉換的機率
 - 謂詞邏輯
 - 物件推論的過程

$$E = \left\{ (v0, v1, 5), (v1, v2, 4), (v2, v3, 9), (v3, v4, 7), (v4, v0, 1), (v0, v5, 2), (v5, v4, 8), (v3, v5, 3), (v5, v2, 1) \right\}$$



Predicate

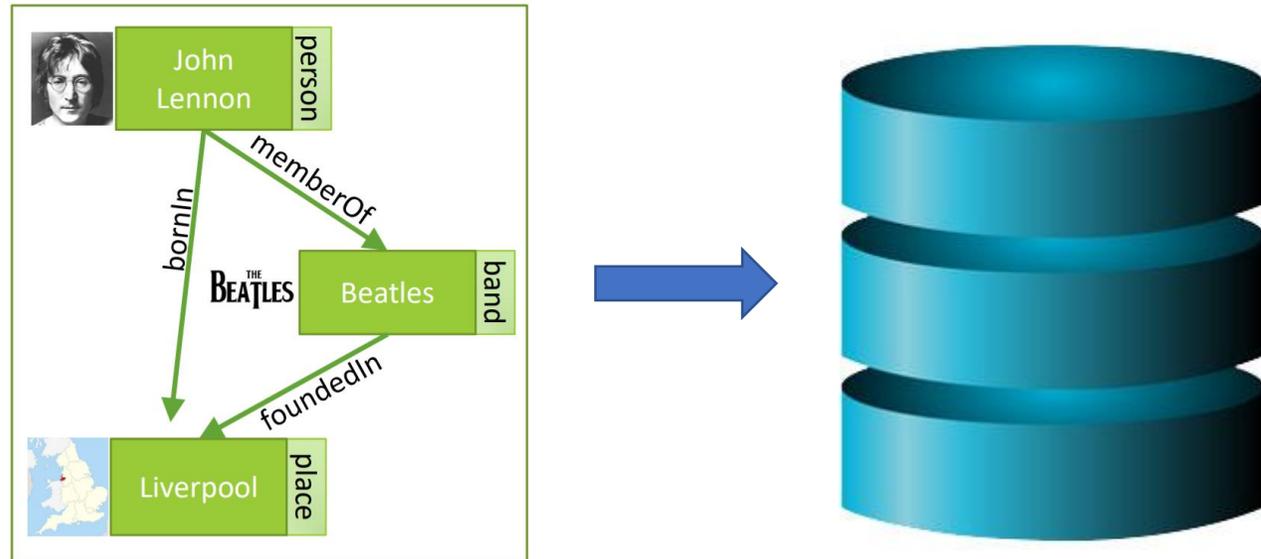
Arguments

LeftOf (x, y)

Atomic Sentence

前情提要-資料庫

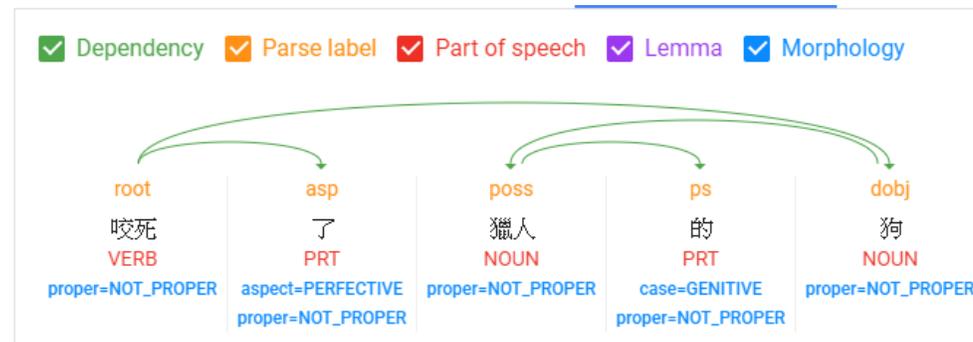
- 將知識透過一定方式表達後，需要存取的地方
 - SQL
 - NOSQL
 - GRAPH DB



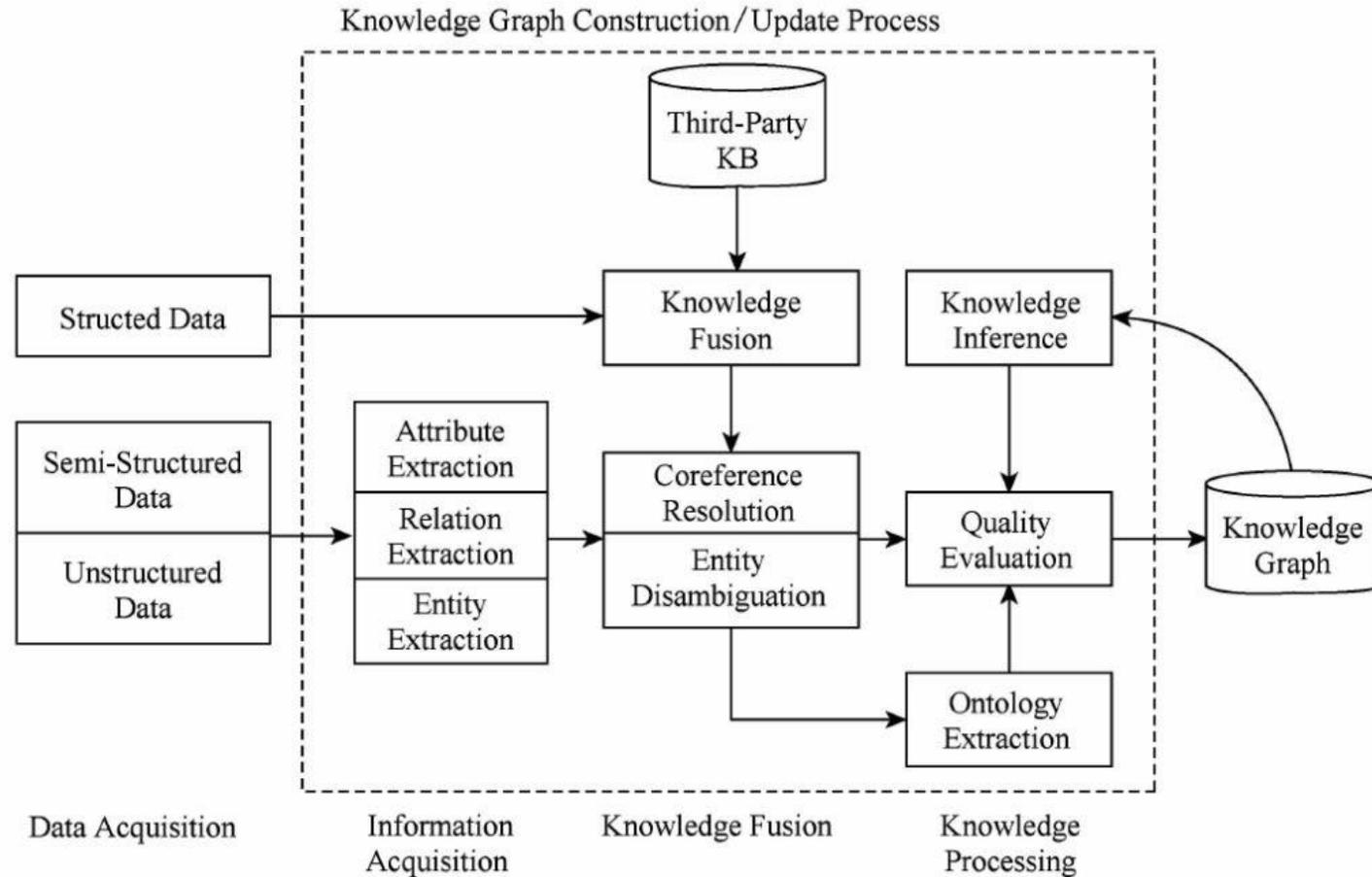
前情提要-自然語言處理

- 自然語言處理的應用可以應用在前期知識表示的抽取(實體及關係之抽取)以及後續在應用層次上(問答及查詢等等)之任務上。
 - 斷詞
 - 詞性
 - 命名實體辨識
 - 句法樹

1. 歐巴馬 Wikipedia Article Saliency: 0.54	PERSON	2. 總統 Saliency: 0.25	PERSON
3. 美國 Wikipedia Article Saliency: 0.21	LOCATION		



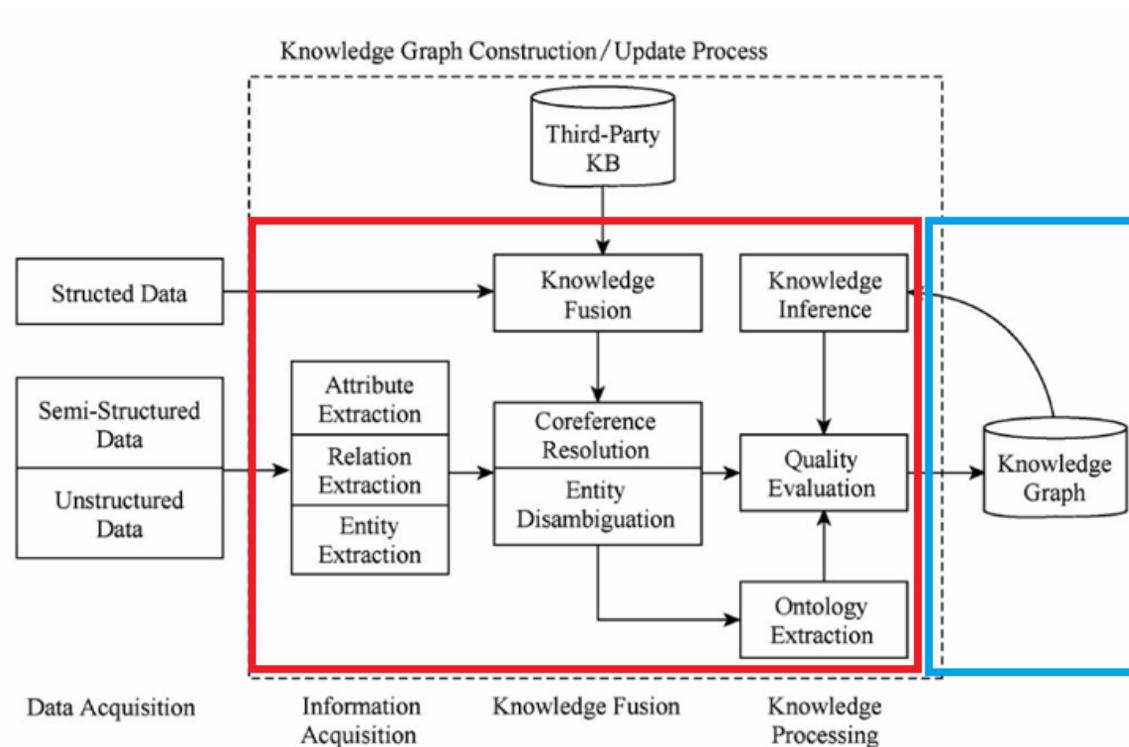
前情提要-建購知識圖譜



知識處理及圖資料庫 流程及建構

前言

- 前面複習了知識圖譜的一些概念，這邊我們會進一步去討論，知識處理及儲存的這個部分

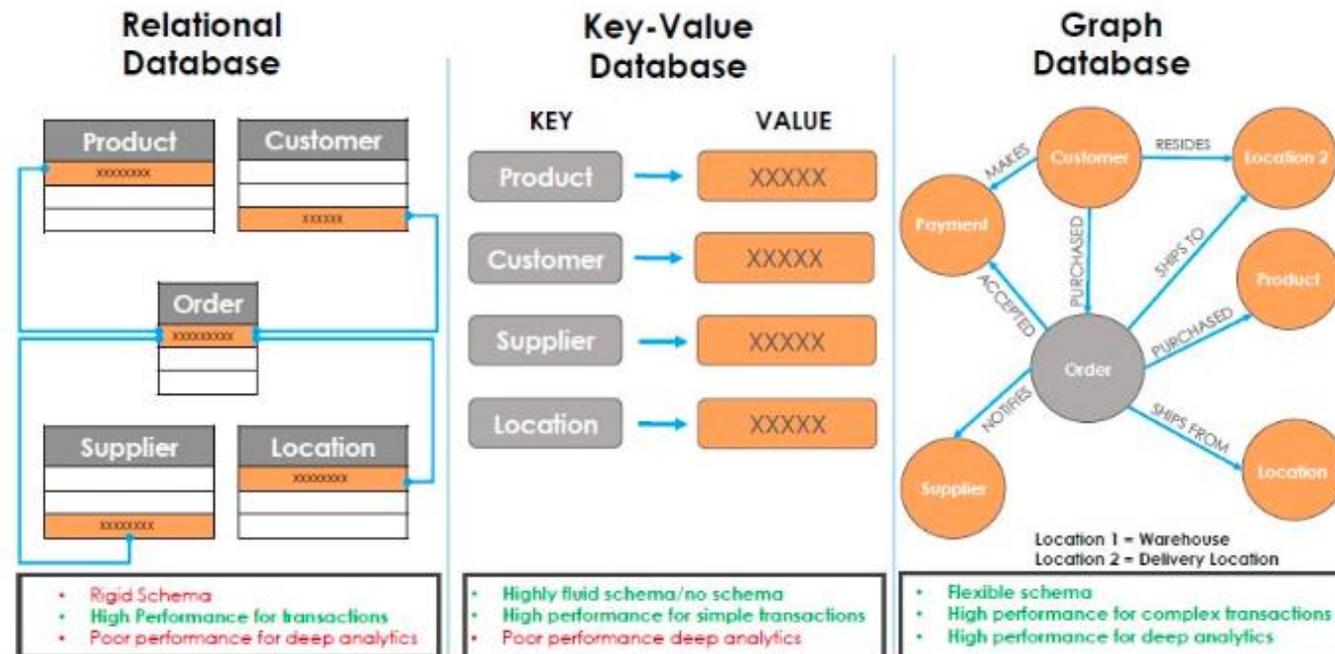


建構資料庫

- 知識圖譜會運用到很多層面的技術或知識，但儲存的基礎是資料庫，資料庫可以作為知識圖譜的基礎去存儲知識
- 資料庫可以幫我們實現知識圖譜中的兩個特性：
 - 大量儲存
 - 快速查找

資料庫種類

- 資料庫的概念出現以來，出現了許多種類的資料庫，而在知識圖譜中，GraphDB則是較為多人使用的類型

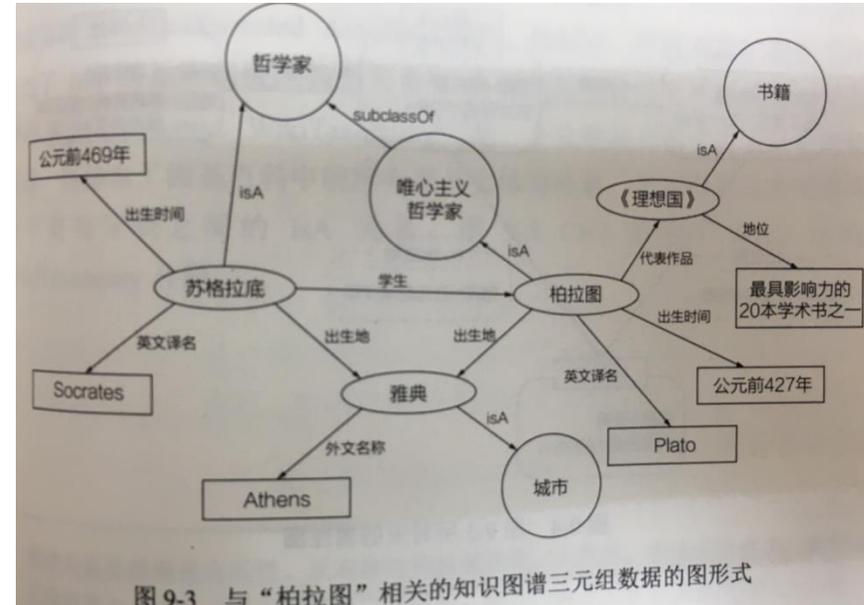


處理過的知識如何存取？

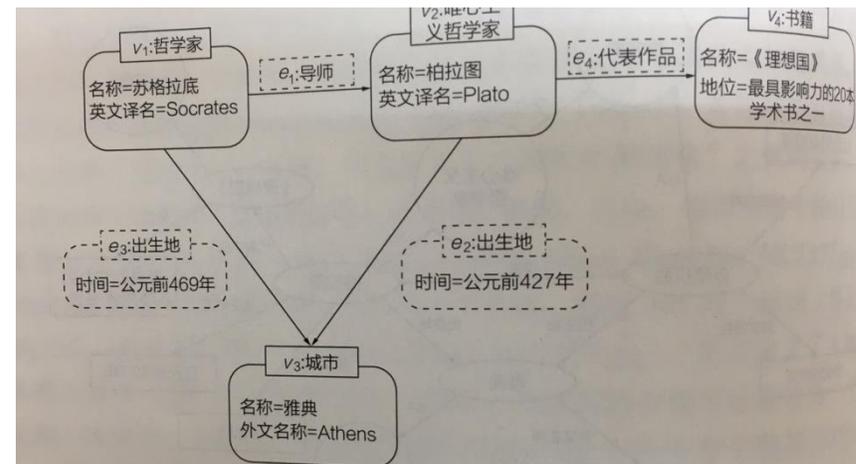
- 當我們試圖將知識存進資料庫時，往往會遇到如何處理所謂的“知識”的情況
- 前面提到我們在Knowledge Process過程中會進行實體辨識、關係抽取等等，但在這個過程後，我們該如何存取這些萃取過後的知識？

知識儲存的格式

- 三元組
 - RDF
 - RDF有向圖
- 屬性圖
 - 透過標籤標示類型
 - 加入屬性(property)



Source	Relation	Target
蘇格拉底	IsA	哲學家



Why Graph DB

- 以SQL來說，三元組的儲存格式，可以很輕易的存取在關聯資料表中
- 但一個非常巨大的三元組表，在查找時是十分耗時的(查詢某一種關係需要遍歷整個表) - 額外創取屬性表去降低連接操作次數
- 對於屬性圖來說，NOSQL的存取方式以及節點、邊屬性皆分散儲存，則是目前許多人採用的graphDB儲存方式 (Neo4j)

Neo4j儲存的方式

- Neo4j 運用類似SQL的語法 - Cypher
- 將處理過後的資訊，透過屬性圖的方式儲存至其DataBase中
- 查找同樣可以透過Cypher去進行大量且多連結的查詢



Demo 實作

Repo

- https://github.com/aron3312/GraphDB_sample