專利分析

資策會 中壢 大數據巨量分析班16期

第二組

CONTENT

01 組員介紹

02 主題介紹

03 系統架構

04 資料處理

05 機器學習

06 視覺化

07 展望

組員介紹

講者:王　昱

王予欣
資料處理
視覺化

王嘉宏
系統管理
資料管理及處理

顧艾菱
資料處理
機器學習

王昱
組長
系統管理

曾昱璇
網路爬蟲
視覺化

楊雅婷
前端架構建置管理
視覺化

專利分析

- 專利文件具有商業價值及戰略意義
- WIPO收錄超過9000萬筆專利文件，需要巨量資料技術進行分析
- 專利分類由專利官審批，長期為人詬病

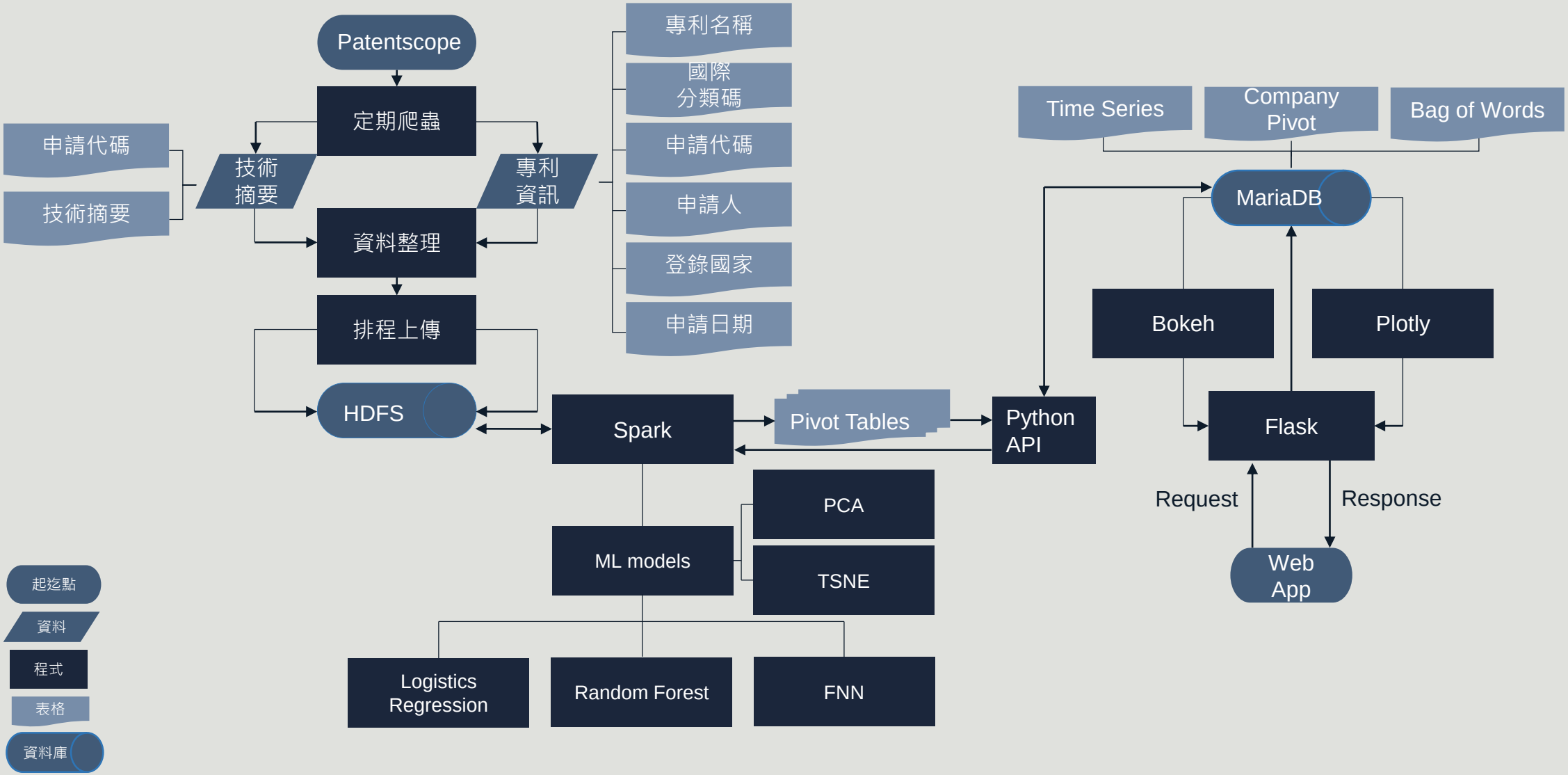
講者:王　昱

專案目標

- 建立專利資料庫
- 繪製圖形互動式專利地圖
  - 尋找專利技術破口
  - 協助企業研發策略布局
- 以AI文字探勘技術分類IPC
- 建立易於操作及可視化之儀表板
  - 可了解產業趨勢
  - 分析公司競合關係
- 建立智慧專利分類落點分析

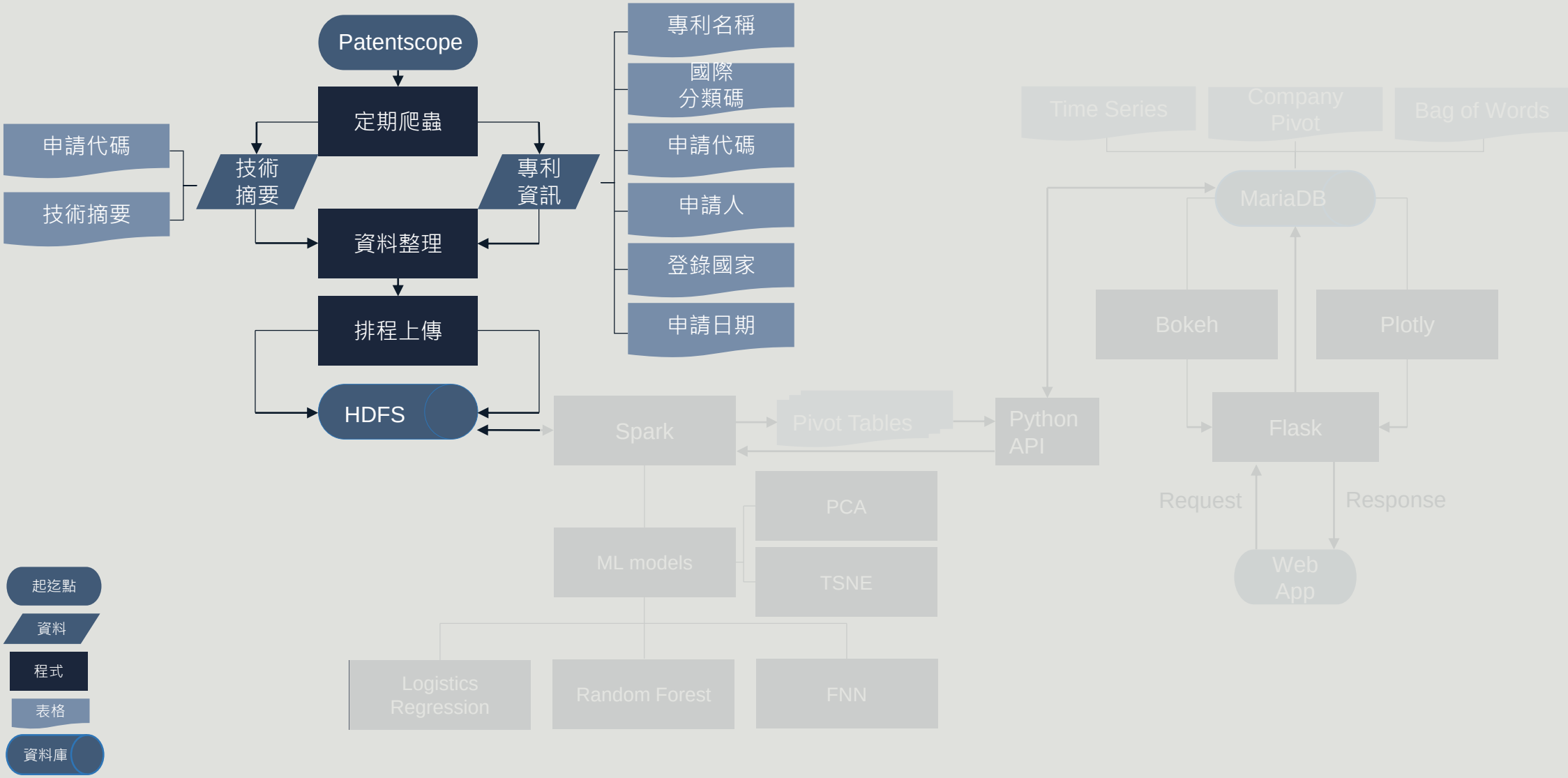
專案流程

講者:王　昱


Patentscope
定期爬蟲
申請代碼
技術摘要
技術摘要
專利資訊
專利名稱
國際分類碼
申請代碼
申請人
登錄國家
申請日期
資料整理
排程上傳
HDFS
Spark
Pivot Tables
Python API
Time Series
Company Pivot
Bag of Words
MariaDB
Bokeh
Plotly
Flask
Request
Response
Web App
PCA
TSNE
ML models
Logistics Regression
Random Forest
FNN
起迄點
資料
程式
表格
資料庫

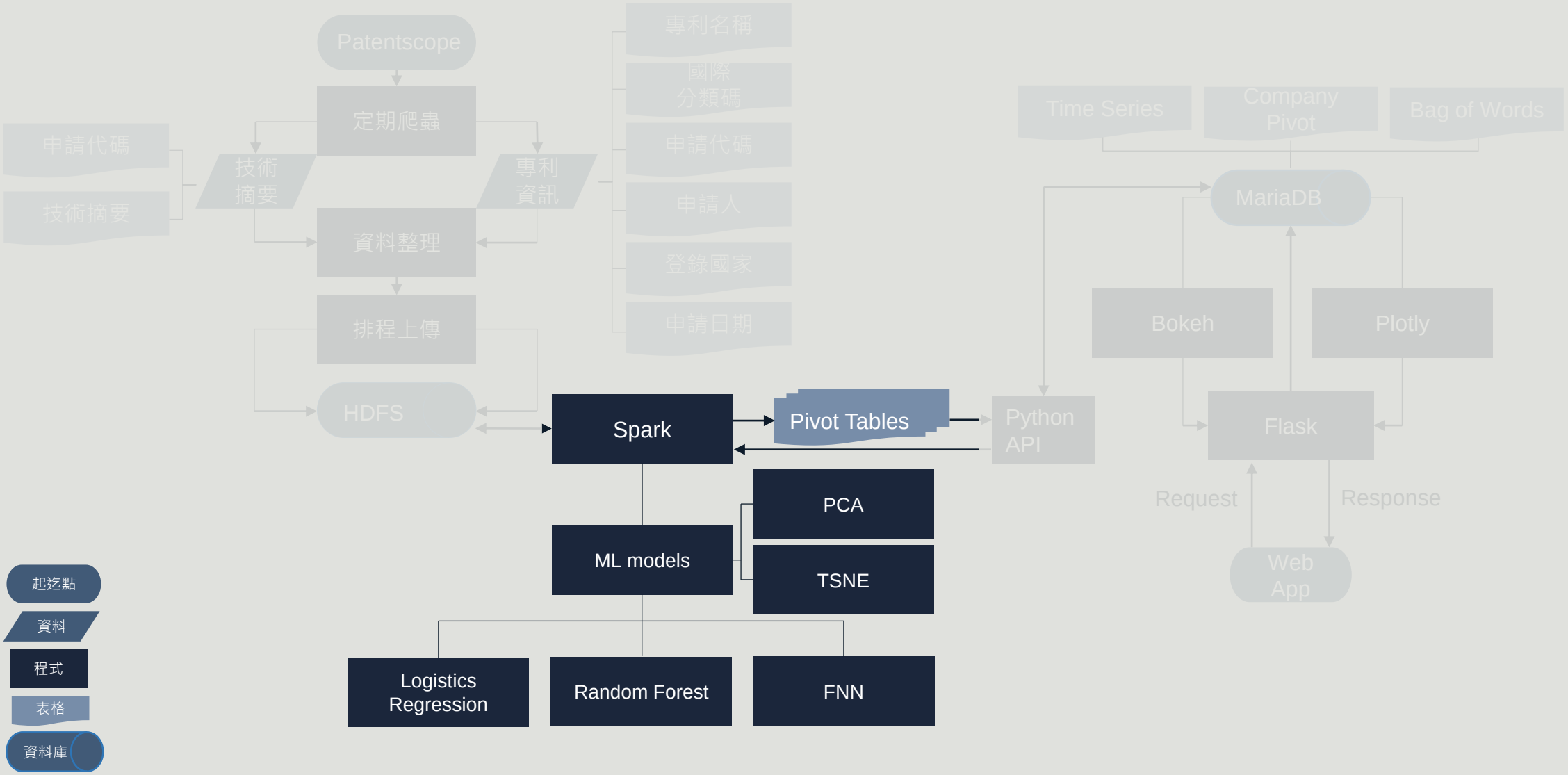
專案流程

講者:王　昱


Patentscope
定期爬蟲
申請代碼
技術摘要
技術摘要
專利資訊
專利名稱
國際分類碼
申請代碼
申請人
登錄國家
申請日期
資料整理
排程上傳
HDFS
Spark
Pivot Tables
Python API
Time Series
Company Pivot
Bag of Words
MariaDB
Bokeh
Plotly
Flask
Request
Response
Web App
PCA
ML models
TSNE
Logistics Regression
Random Forest
FNN
起迄點
資料
程式
表格
資料庫

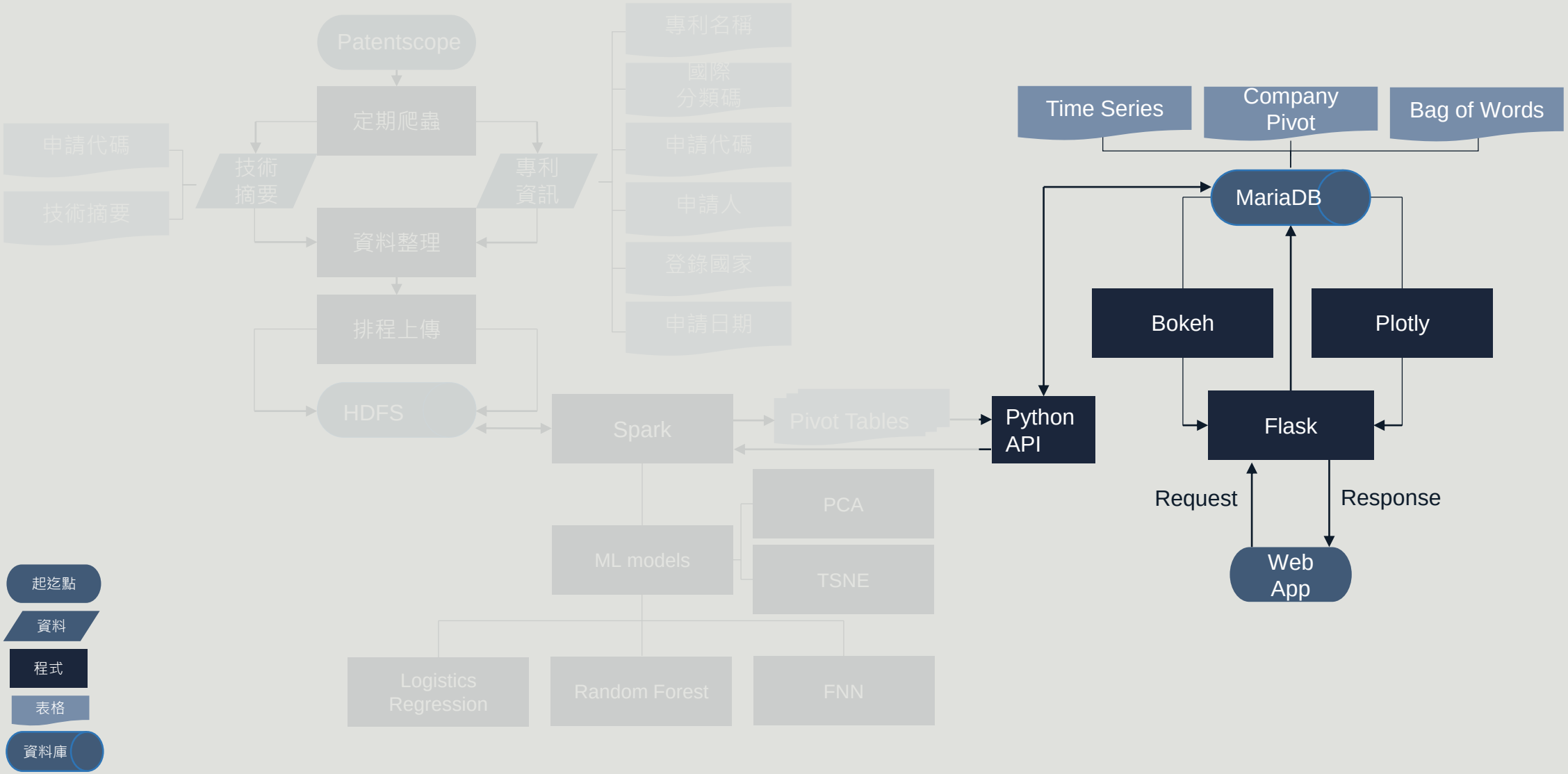
專案流程

講者:王　昱


Patentscope
定期爬蟲
申請代碼
技術摘要
技術摘要
專利資訊
專利名稱
國際分類碼
申請代碼
申請人
登錄國家
申請日期
資料整理
排程上傳
HDFS
Spark
Pivot Tables
Python API
Time Series
Company Pivot
Bag of Words
MariaDB
Bokeh
Plotly
Flask
Request
Response
Web App
PCA
ML models
TSNE
Logistics Regression
Random Forest
FNN
起迄點
資料
程式
表格
資料庫

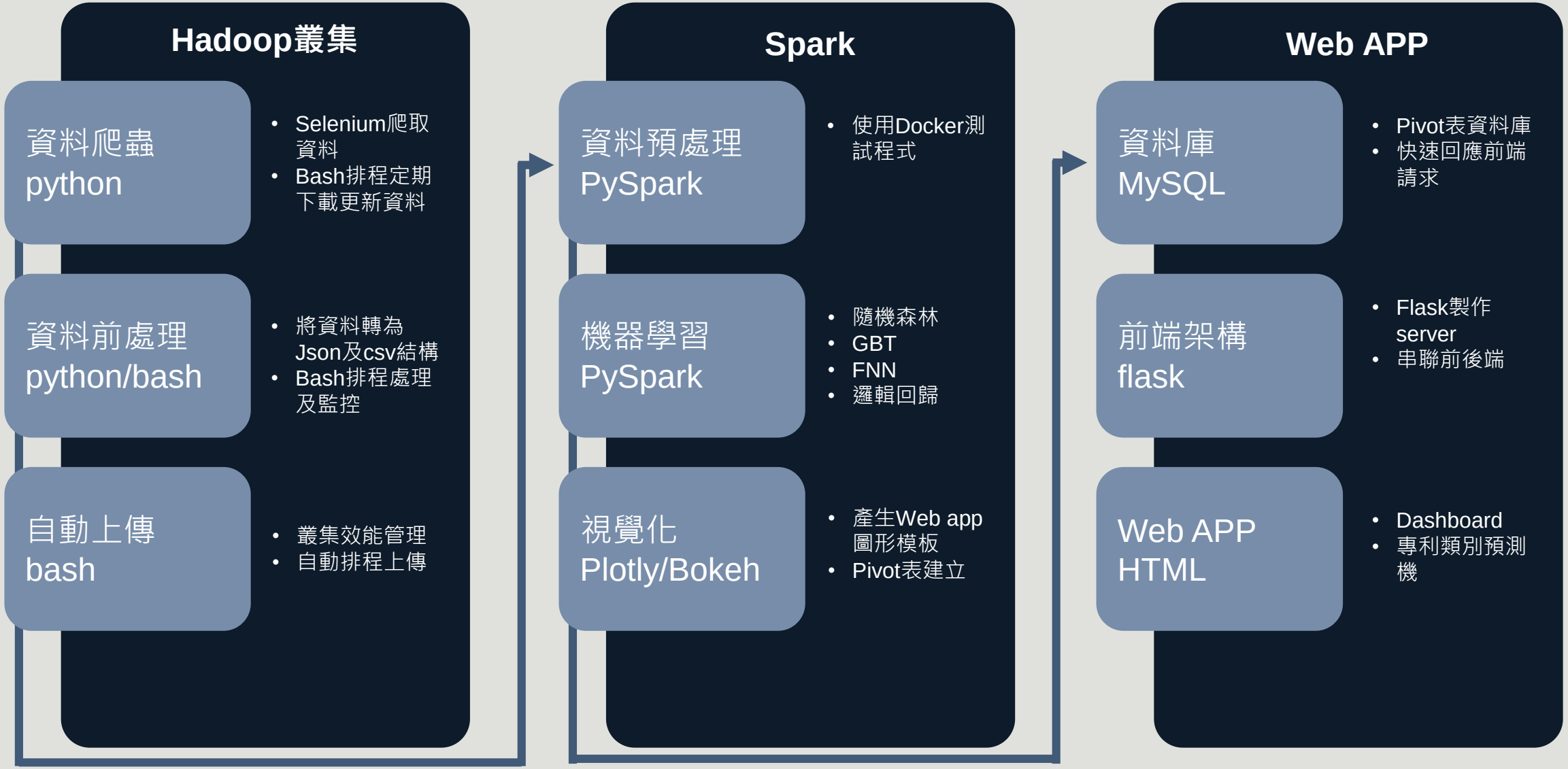
專案流程

講者:王　昱


Patentscope
定期爬蟲
申請代碼
技術摘要
技術摘要
專利資訊
專利名稱
國際分類碼
申請代碼
申請人
登錄國家
申請日期
資料整理
排程上傳
HDFS
Spark
Pivot Tables
Python API
Time Series
Company Pivot
Bag of Words
MariaDB
Bokeh
Plotly
Flask
Request
Response
Web App
ML models
PCA
TSNE
Logistics Regression
Random Forest
FNN
起迄點
資料
程式
表格
資料庫

系統架構

講者:王嘉宏


Hadoop叢集
資料爬蟲
python
Selenium爬取資料
Bash排程定期下載更新資料
資料前處理
python/bash
將資料轉為Json及csv結構
Bash排程處理及監控
自動上傳
bash
叢集效能管理
自動排程上傳
Spark
資料預處理
PySpark
使用Docker測試程式
機器學習
PySpark
隨機森林
GBT
FNN
邏輯回歸
視覺化
Plotly/Bokeh
產生Web app圖形模板
Pivot表建立
Web APP
資料庫
MySQL
Pivot表資料庫
快速回應前端請求
前端架構
flask
Flask製作server
串聯前後端
Web APP
HTML
Dashboard
專利類別預測機

資料簡介

專利公告日 →

國際類別分類碼 →

申請人資料 →

1. WO2021000007 - AN APPLICATOR

PCT Biblio. Data | Description | Claims | Drawings | National Phase | Notices | Documents

Submit observation | PermaLink | Machine translation

Publication Number
WO/2021/000007

Publication Date
07.01.2021

International Application No.
PCT/AU2020/050658

International Filing Date
26.06.2020

IPC
B05C 1/04 2006.01 | B05C 17/00 2006.01 | B43M 11/00 2006.01

Applicants
MCDONALD, Natlan [AU]/[AU]

Inventors
MCDONALD, Natlan

Agents
WYNNES PATENT AND TRADE MARK ATTORNEYS PTY LTD
Unit 4, 27 Godwin Street Bulimba Brisbane, Queensland 4171, AU

Priority Data
2019204795 03.07.2019 AU

Publication Language
English [EN]

Filing Language
English [EN]

Title
[EN] AN APPLICATOR
[FR] APPLICATEUR

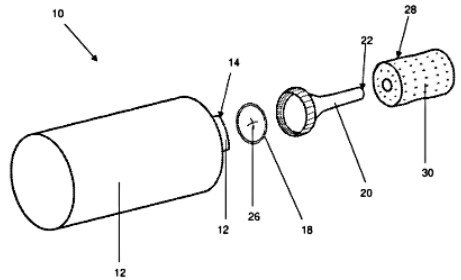


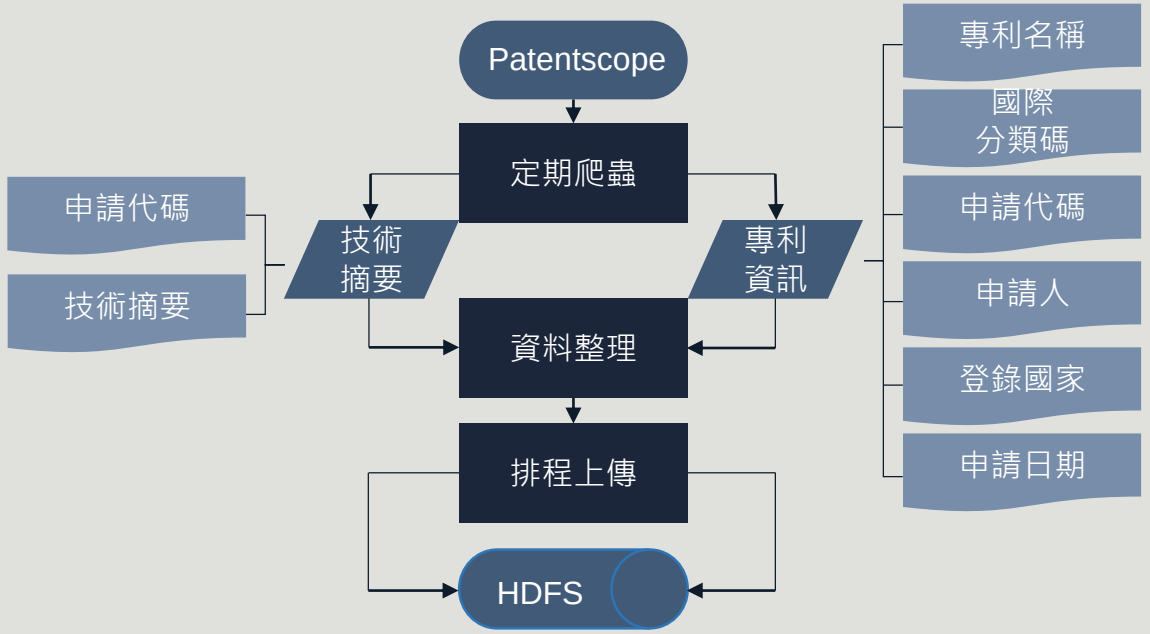
Figure 1

Abstract
[EN]
An applicator for applying a solvent cement, the applicator having a deformable container for containing the solvent cement, a nozzle, a valve located between the container and the nozzle, the valve changeable between an open condition where the solvent cement can move from the container to the nozzle and a closed condition where solvent cement is inhibited from moving from the container to the nozzle, and a disperser attached to the nozzle, the disperser adapted to disperse solvent cement that flows from the nozzle.
[FR]
La présente invention concerne un applicateur permettant d'appliquer une colle à solvant, l'applicateur ayant un récipient déformable pour contenir la colle à solvant, une buse, une vanne située entre le récipient et la buse, la vanne pouvant être modifiée entre un état ouvert dans lequel la colle à solvant peut se déplacer du récipient à la buse et un état fermé dans lequel la colle à solvant est empêchée de se déplacer du récipient à la buse, et un disperseur fixé à la buse, le disperseur étant conçu pour disperser la colle à solvant qui s'écoule à partir de la buse.

← 技術摘要

資料前處理

- 資料抓取後按內容與需求
  將資料分為兩批並進行前處理
- 將處理完的檔案上傳至叢集



資料前處理

- Python+Bash+Crontab，爬蟲、資料前處理、上傳及監控，**全自動化完成**

```
#
# m h  dom mon dow   command

#專利號碼爬蟲
0 0 * * 1 P_number.py
#專利爬蟲
0 12 * * 1 P_scraper.PY
#資料前處理
0 12 * * 2 Make_JSON.py
0 0 * * 3 abs.py
#資料上傳
0 6 * * 3 bash Update_ABS.sh
0 12 * * 4 bash Update_JSO.sh
```

專利資料庫 I

- 專利摘要資料表
  - 紀錄所有專利之技術簡介
  - 自然語言
  - 包含大量專業詞彙、化學式等
  - 資料庫蒐集超過400,000筆
  - 用於訓練智慧專利預測模型

```
root
 |-- Application_Number : string (nullable = true)
 |-- Abstract: string (nullable = true)
```

_c0	_c1
PCT/EP2001/009832	Liquid crystal mi...
PCT/KR2000/000814	The present inven...
PCT/IL2000/000667	The present inven...
PCT/DK2001/000750	The invention rel...
PCT/EP2001/012972	The invention rel...
PCT/GB1992/000681	Compounds having ...
PCT/EP2007/000257	The invention rel...
PCT/GB1992/001189	Compounds of gene...
PCT/RU2001/000190	The inventive met...
PCT/EP2004/007957	The invention rel...
PCT/IL2000/000459	The present inven...
PCT/KR2006/002440	A diagnostic kit ...
PCT/KR2001/001941	The present inven...
PCT/SE2002/000250	To facilitate eg ...
PCT/GR1993/000005	A technique for t...
PCT/DE2001/001579	The invention rel...
PCT/IL2001/000088	A method and circ...
PCT/KR2001/000965	A preferable embo...
PCT/GB2001/002487	New spisulosine d...
PCT/HU2001/000068	A system of appar...

only showing top 20 rows

專利資料庫 II

- 專利元資料表
  - 紀錄該專利之所有元資料，包含申請人、申請年分、行業別等資訊
  - 資料庫蒐集超過400,000筆
  - 用以實作統計描述，可萃取出具戰略價值之資訊
  - 統計描述結果可以儀表板方式呈現

```
root
 |-- Agents: string (nullable = true)
 |-- Applicants: array (nullable = true)
 |    |-- element: string (containsNull = true)
 |-- Application_Number: string (nullable = true)
 |-- Designated_States: struct (nullable = true)
 |    |-- African Intellectual Property Organization : array (nullable = true)
 |    |    |-- element: string (containsNull = true)
 |    |-- African Regional Intellectual Property Organization  : array (nullable = true)
 |    |    |-- element: string (containsNull = true)
 |    |-- Eurasian Patent Organization : array (nullable = true)
 |    |    |-- element: string (containsNull = true)
 |    |-- European Patent Office : array (nullable = true)
 |    |    |-- element: string (containsNull = true)
 |    |-- Global: array (nullable = true)
 |    |    |-- element: string (containsNull = true)
 |-- IPC: array (nullable = true)
 |    |-- element: string (containsNull = true)
 |-- Inventors: string (nullable = true)
 |-- PublicationNo_Name: string (nullable = true)
 |-- Publication_Date: string (nullable = true)
 |-- Publication_Number: string (nullable = true)
 |-- Title: array (nullable = true)
 |    |-- element: string (containsNull = true)
 |-- mono_ipc: string (nullable = true)
 |-- qua_ipc: string (nullable = true)
 |-- tri_ipc: string (nullable = true)
```

專利資料庫 II

- 專利元資料表
  - 紀錄該專利之所有元資料，包含申請人、申請年分、行業別等資訊
  - 資料庫蒐集超過400,000筆
  - 用以實作統計描述，可萃取出具戰略價值之資訊
  - 統計描述結果可以儀表板方式呈現

```
|-- Applicants: array (nullable = true)
|    |-- element: string (containsNull = true)

|-- Designated_States: struct (nullable = true)

|-- IPC: array (nullable = true)
|    |-- element: string (containsNull = true)

|-- Publication_Date: string (nullable = true)
```

專利資料庫 II

- 專利元資料表

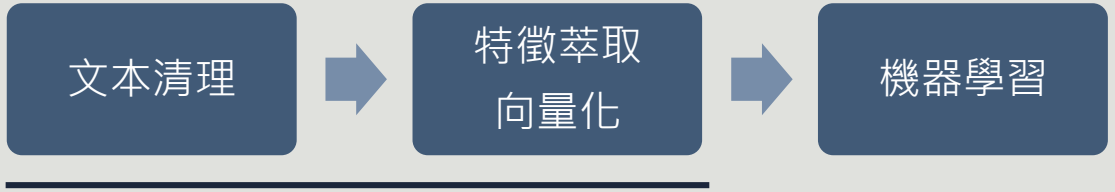
Agents	Applicants	Application_Number	Designated_States	Inventors	PublicationNo_Name	Publication_Date	Publication_Number	Title	ipc_simplified	mono_ipc
'ОСИПОВА, Наталья...	[' [CY]/[CY] (All...	PCT/IB2011/002779	[[BF, BJ, CF, ...	'ЯСНЕЦОВ, Владими...	1. WO2012028965 -...	08-03-2012	WO/2012/028965	[EN), COMBINATION...	[A]	A
nan	[' [BY]/[BY] (AM,...	PCT/BY2015/000005	[[BF, BJ, CF, ...	'ЖАВНЕРКО, Геннад...	1. WO2017070769 -...	04-05-2017	WO/2017/070769	[EN), COMPOSITE O...	[C, G]	C
nan	[' [BY]/[BY] (AM,...	PCT/BY2015/000005	[[BF, BJ, CF, ...	'ЖАВНЕРКО, Геннад...	1. WO2017070769 -...	04-05-2017	WO/2017/070769	[EN), COMPOSITE O...	[C, G]	G
"ООО 'МЕЖДУНАРОДН...	[' [RU]/[RU] (All...	PCT/EA2007/000006	[[BF, BJ, CF, ...	'ТАХАУТДИНОВ Шафа...	1. WO2008046426 -...	24-04-2008	WO/2008/046426	[EN), METHOD FOR ...	[B]	B
nan	[' [AZ]/[AZ]': 'Б...	PCT/AZ2017/000007	[[BF, BJ, CF, ...	'АЛЫЕВ, Абульфат ...	1. WO2019056075 -...	28-03-2019	WO/2019/056075	[EN), METHOD FOR ...	[C, G]	C
nan	[' [AZ]/[AZ]': 'Б...	PCT/AZ2017/000007	[[BF, BJ, CF, ...	'АЛЫЕВ, Абульфат ...	1. WO2019056075 -...	28-03-2019	WO/2019/056075	[EN), METHOD FOR ...	[C, G]	G
'BENEDETTO, Marco...	[' [IT]/[IT]': 'C...	PCT/IB2015/056276	[[BF, BJ, CF, ...	'MOGNA, Giovanni'...	1. WO2016027231 -...	25-02-2016	WO/2016/027231	[EN), METHOD FOR ...	[A]	A
nan	[' [BY]/[BY] (AL,...	PCT/BY2014/000007	[[BF, BJ, CF, ...	'ШИРИПОВ, Владими...	1. WO2016033674 -...	10-03-2016	WO/2016/033674	[EN), DEVICE FOR ...	[B]	B
nan	[' [US]/[US] (All...	PCT/IB2009/050270	[[BF, BJ, CF, ...	'ИВАЩЕНКО Андрей ...	1. WO2009093206 -...	30-07-2009	WO/2009/093206	[EN), 3-SULFONYL-...	[C, A]	C
nan	[' [US]/[US] (All...	PCT/IB2009/050270	[[BF, BJ, CF, ...	'ИВАЩЕНКО Андрей ...	1. WO2009093206 -...	30-07-2009	WO/2009/093206	[EN), 3-SULFONYL-...	[C, A]	A
nan	[' [DE]/[DE] (All...	PCT/EP2011/062178	[[BF, BJ, CF, ...	'FISCHER, Rüdiger...	1. WO2012010525 -...	26-01-2012	WO/2012/010525	[DE), VERWENDUNG ...	[A]	A
nan	[' [AZ]/[AZ] (All...	PCT/AZ2008/000002	[[BF, BJ, CF, ...	'ГАШИМОВ, Ариф Ма...	1. WO2009105840 -...	03-09-2009	WO/2009/105840	[EN), METHOD FOR ...	[H]	H
'HOFFMANN EITLE S...	[' [IT]/[IT]': 'A...	PCT/IB2018/058051	[[BF, BJ, CF, ...	'FERRARI, Alessio'	1. WO2019077521 -...	25-04-2019	WO/2019/077521	[EN), LIQUID COMP...	[A]	A
'БАЗАНОВ, Юрий Бо...	[' [RU]/[RU]': 'O...	PCT/IB2015/050897	[[BF, BJ, CF, ...	'ЕРИМБЕТОВ, Кенес...	1. WO2015118488 -...	13-08-2015	WO/2015/118488	[EN), AGENT FOR T...	[A]	A
'DR. LASSE WEINMA...	[' [DE]/[DE]': 'J...	PCT/EP2016/072524	[[BF, BJ, CF, ...	'HUDECEK, Michael...	1. WO2017050884 -...	30-03-2017	WO/2017/050884	[EN), A METHOD FO...	[C, A]	C
'DR. LASSE WEINMA...	[' [DE]/[DE]': 'J...	PCT/EP2016/072524	[[BF, BJ, CF, ...	'HUDECEK, Michael...	1. WO2017050884 -...	30-03-2017	WO/2017/050884	[EN), A METHOD FO...	[C, A]	A
'ФЕДОРОВ, Дмитрий...	[' [RU]/[RU] (All...	PCT/IB2012/052483	[[BF, BJ, CF, ...	'ШУРИГИН, Михаил ...	1. WO2012156938 -...	22-11-2012	WO/2012/156938	[EN), COMPOUNDS P...	[C, A]	C
'ФЕДОРОВ, Дмитрий...	[' [RU]/[RU] (All...	PCT/IB2012/052483	[[BF, BJ, CF, ...	'ШУРИГИН, Михаил ...	1. WO2012156938 -...	22-11-2012	WO/2012/156938	[EN), COMPOUNDS P...	[C, A]	A
'ASTRAZENECA INTE...	[' [SE]/[SE] (AE,...	PCT/GB2010/050822	[[BF, BJ, CF, ...	'BENNETT, Nichola...	1. WO2010133882 -...	25-11-2010	WO/2010/133882	[EN), DISACCHARIN...	[C, A]	C
'ASTRAZENECA INTE...	[' [SE]/[SE] (AE,...	PCT/GB2010/050822	[[BF, BJ, CF, ...	'BENNETT, Nichola...	1. WO2010133882 -...	25-11-2010	WO/2010/133882	[EN), DISACCHARIN...	[C, A]	A

資料預處理

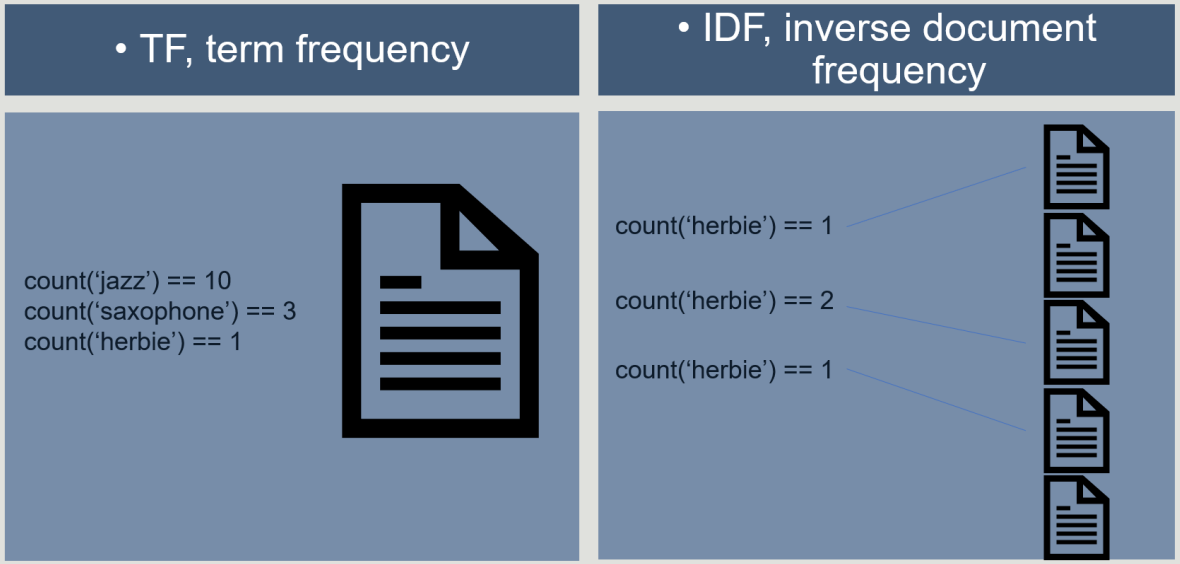
- 將專利摘要，轉換整**文件集**為機器學習模型可處理的**向量型態**。
- 使用TF-IDF技術
  - TF：詞頻，該單字在文本內出現次數。
  - IDF：逆向文件頻率，該單字在整份文件集的稀有程度。

範例：

1. ‘Python python Spark Spark’ ->
   (5,[0,4],[2.0,2.0]),
   (5,[0,4],[0.8109302162163288,0.0])
2. ‘Python SQL’ ->
   (5,[1,4], [1.0,1.0]),
   (5,[1,4],[0.4054651081081644,0.0])



資料預處理



資料預處理

- 專利摘要為技術之敘述，屬自然語言。
- 去除標段落及標點符號

The invention relates to a method for positioning an operating apparatus (28) in respect of a bending machine (3) of a production line (1) for bending sheet metal. A work surface (39) having specified boundaries is defined for the operating apparatus (28), within which work surface the operating apparatus (28) can be relocated. A target position is transmitted to an operating-control unit (35) via a first communication interface (33) and a current actual position of the operating apparatus (28) is determined. In addition, a target travel path (40) between the positions is determined and a generated travel command is transmitted to a travel drive (36). The operating apparatus (28) is automatically and autonomously relocated along the target travel path (40) and parked positioned at the target position.

The invention relates to a method for positioning an operating apparatus in respect of a bending machine of a production line for bending sheet metal. A work surface having specified boundaries is defined for the operating apparatus within which work surface the operating apparatus can be relocated. A target position is transmitted to an operating-control unit via a first communication interface and a current actual position of the operating apparatus is determined. In addition, a target travel path between the positions is determined and a generated travel command is transmitted to a travel drive The operating apparatus is automatically and autonomously relocated along the target travel path and parked positioned at the target position

專利摘要 → 建立詞袋 → 文字向量化

資料預處理

- 去除常見、無代表性之單字
- 計算詞頻(Term Frequency)
- 建立詞袋
- 計算逆向文件頻率(Inverse Document Frequency)

vocabList	counts
invention	115492.0
one	102200.0
least	79562.0
said	75129.0
method	66887.0
first	65188.0
device	60588.0
comprising	60137.0
relates	54586.0
second	52090.0
means	51299.0
system	48751.0
provided	47785.0
comprises	45356.0
material	44986.0
surface	44534.0
present	43967.0
also	41774.0
wherein	39181.0
may	37113.0

only showing top 20 rows

資料預處理

- 計算逆向文件頻率(Inverse Document Frequency)
- 資料集欄位數量：197097欄
- 文件完全轉為向量，可進行後續NLP模型訓用

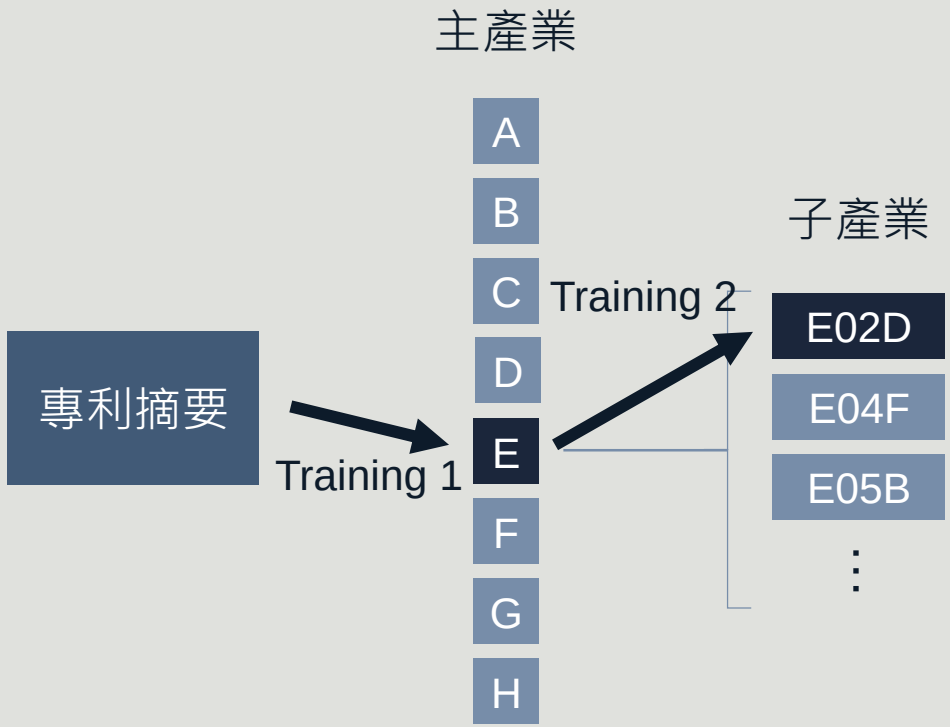
The invention relates to a method for positioning an operating apparatus in respect of a bending machine of a production line for bending sheet metal. A work surface having specified boundaries is defined for the operating apparatus within which work surface the operating apparatus can be relocated. A target position is transmitted to an operating-control unit via a first communication interface and a current actual position of the operating apparatus is determined. In addition, a target travel path between the positions is determined and a generated travel command is transmitted to a travel drive The operating apparatus is automatically and autonomously relocated along the target travel path and parked positioned at the target position

(197091,[1,2,10,12,32,34,42,56,65,79,107,176,274,327,409,457,1454,1506,1978,5511,9097,20862,111267],[1.461182438344008,1.737658437916606,2.2927314140520427,1.9731288448602595,3.0686465493899537,2.95946759801359,3.1997371715396246,3.397178705855117,10.001082332066941,3.9640480312097166,3.693323452225631,3.4594725571388643,3.8269897521016234,8.619112448393286,4.592682480112563,9.439976943617742,5.405258149925722,5.247502070531855,5.614978680907791,7.166959538317104,7.984255148400509,9.40564082933167,0.0])

専利摘要 → 建立詞袋 → 文字向量化

智慧分類模型建立

- 將向量化之**專利摘要**與之**國際分類代碼**(IPC)放入模型中訓練，並篩選表現最佳者
- 訓練目標類別
  - 大分類：8大類別
  - 子分類：654 子類別
- 使用以下三種演算法做訓練：
  - 羅吉斯回歸
  - 隨機森林
  - 前饋神經網絡FNN
- 困難：
  - 樣本具超高維度
  - 類別標籤數量龐大



模型訓練結果

模型精準度	羅吉斯回歸	隨機森林	前饋神經網絡
8大分類	0.435	0.275	0.392
654子分類	0.448	0.298	0.573

大分類

mono_ipc	rawFeatures	features	indexedLabel	rawPrediction	probability	prediction
A	(18778,[0,2,3,4,6...	(18778,[0,2,3,4,6...	0.0	[-104.76359778962...	[0.0,1.0,0.0,0.0,...	1.0
A	(18778,[6,30,104,...	(18778,[6,30,104,...	0.0	[-30.771753986580...	[9.68753173195840...	4.0
B	(18778,[0,3,4,6,7...	(18778,[0,3,4,6,7...	2.0	[-44.143478410535...	[8.36028938053069...	6.0
G	(18778,[0,4,5,13,...	(18778,[0,4,5,13,...	3.0	[-591.92806829425...	[0.0,1.0,0.0,1.41...	1.0
H	(18778,[0,2,3,4,5...	(18778,[0,2,3,4,5...	4.0	[-128.38596039992...	[0.0,1.0,6.126823...	1.0
A	(18778,[0,2,10,12...	(18778,[0,2,10,12...	0.0	[123.477958694314...	[1.0,1.3851310503...	0.0
A	(18778,[0,4,10,22...	(18778,[0,4,10,22...	0.0	[-64.171678946210...	[1.04878786001143...	2.0
A	(18778,[0,4,13,18...	(18778,[0,4,13,18...	0.0	[-103.12443197521...	[9.19483065458986...	1.0
A	(18778,[0,29,35,3...	(18778,[0,29,35,3...	0.0	[-64.812648897485...	[5.62490008980729...	1.0
B	(18778,[0,1,2,4,1...	(18778,[0,1,2,4,1...	2.0	[362.210024830556...	[0.50166976777357...	0.0
C	(18778,[0,1,4,19,...	(18778,[0,1,4,19,...	1.0	[260.589861064609...	[1.36651082401933...	1.0
C	(18778,[1,5,6,10,...	(18778,[1,5,6,10,...	1.0	[-237.93966907686...	[0.0,1.0,0.0,0.0,...	1.0
C	(18778,[1,21,27,4...	(18778,[1,21,27,4...	1.0	[442.683092368988...	[1.0,1.5864390454...	0.0
H	(18778,[2,5,11,13...	(18778,[2,5,11,13...	4.0	[-287.82276444534...	[0.0,0.0,0.0,0.0,...	4.0
A	(18778,[0,4,6,7,8...	(18778,[0,4,6,7,8...	0.0	[617.470316544188...	[1.0,0.0,4.026423...	0.0
A	(18778,[6,14,25,2...	(18778,[6,14,25,2...	0.0	[527.002157581215...	[1.0,0.0,5.405655...	0.0
A	(18778,[12,13,17,...	(18778,[12,13,17,...	0.0	[446.542771845157...	[1.0,0.0,6.030050...	0.0
B	(18778,[0,1,4,5,1...	(18778,[0,1,4,5,1...	2.0	[-75.970922642582...	[1.98575288057063...	1.0
B	(18778,[1,2,8,24,...	(18778,[1,2,8,24,...	2.0	[-437.05160358736...	[0.0,4.9343454593...	2.0
B	(18778,[33,67,74,...	(18778,[33,67,74,...	2.0	[-122.15498839713...	[1.57253193804042...	1.0

only showing top 20 rows

子分類

qua_ipc	rawFeatures	features	indexedLabel	rawPrediction	probability	prediction
G06F	(7639,[0,2,4,5,10...	(7639,[0,2,4,5,10...	42.0	[15.6640025456533...	[7.46096478311355...	10.0
E02D	(7639,[0,1,2,3,8,...	(7639,[0,1,2,3,8,...	4.0	[1.77767684816139...	[0.00230542461085...	10.0
E21B	(7639,[1,7,8,10,1...	(7639,[1,7,8,10,1...	0.0	[52.0229897849143...	[1.0,2.6766560717...	0.0
E21B	(7639,[3,7,12,15,...	(7639,[3,7,12,15,...	0.0	[39.7653993751308...	[0.99999999999395...	0.0
E21B	(7639,[6,9,10,11,...	(7639,[6,9,10,11,...	0.0	[52.0800113979501...	[0.99999999999999...	0.0
E05B	(7639,[6,13,14,32...	(7639,[6,13,14,32...	3.0	[-14.759699366231...	[4.91360304031870...	3.0
E21B	(7639,[0,4,10,11,...	(7639,[0,4,10,11,...	0.0	[32.7894998876437...	[0.99999999008463...	0.0
E01F	(7639,[0,1,2,4,5,...	(7639,[0,1,2,4,5,...	17.0	[15.5314053225411...	[0.75042430289094...	0.0
E04G	(7639,[2,4,7,9,10...	(7639,[2,4,7,9,10...	11.0	[12.4515389380917...	[0.00452199175885...	6.0
E06B	(7639,[0,1,4,9,10...	(7639,[0,1,4,9,10...	2.0	[-3.3033431012007...	[1.35350069937513...	2.0
E04F	(7639,[0,1,3,4,5,...	(7639,[0,1,3,4,5,...	5.0	[0.80486068135602...	[7.52630682918305...	5.0
E02D	(7639,[0,1,2,3,4,...	(7639,[0,1,2,3,4,...	4.0	[0.25900809221544...	[1.60346273539367...	12.0
E04F	(7639,[4,5,6,8,9,...	(7639,[4,5,6,8,9,...	5.0	[-2.6942295703053...	[3.80655763141517...	1.0
E05B	(7639,[5,8,11,14,...	(7639,[5,8,11,14,...	3.0	[11.9750531760996...	[1.34162422812507...	3.0
E05B	(7639,[0,6,14,30,...	(7639,[0,6,14,30,...	3.0	[-2.6728806228139...	[5.85350730477927...	3.0
E21B	(7639,[1,2,9,10,2...	(7639,[1,2,9,10,2...	0.0	[42.9613195052480...	[0.99999999999999...	0.0
E21B	(7639,[26,31,32,3...	(7639,[26,31,32,3...	0.0	[10.2162036509116...	[0.80265519253776...	0.0
E04B	(7639,[2,5,6,10,1...	(7639,[2,5,6,10,1...	1.0	[3.60328195088580...	[2.09515775121467...	1.0
E21B	(7639,[1,2,4,15,1...	(7639,[1,2,4,15,1...	0.0	[21.8988986377348...	[0.99999884488019...	0.0
E21B	(7639,[7,21,26,35...	(7639,[7,21,26,35...	0.0	[25.2543261396910...	[0.99978150042154...	0.0

only showing top 20 rows

資料解讀

- 化繁為簡
- 觀眾導向
- 可讀性
- 表明觀點
- 資料重複性
- 由果推因，回溯推理


```
{
  "PublicationNo_Name":"1. WO1982002074 - DEVICE FOR PRESSING GEOPHYSICAL DEVICES AGAINST THE WALL OF A BORE HOLE",
  "Title":[
    "EN), DEVICE FOR PRESSING GEOPHYSICAL DEVICES AGAINST THE WALL OF A BORE HOLE), ",
    "FR), \"DISPOSITIF DE COMPRESSION DAPPAREILS GEOPHYSIQUES SUR LA PAROI DUN TROU DE FORAGE\"),"
  ],
  "Publication_Number":"WO/1982/002074",
  "Publication_Date":"24-06-1982",
  "Application_Number":"PCT/SU1980/000201",
  "IPC":[
    "'E21B 23/01', '2006.01'",
    "'E21B 47/00', '2012.01'",
    "'E21B 49/10', '2006.01'",
    "'G01V 1/147', '2006.01'",
    "'G01V 11/00', '2006.01'",
    "'G01V 3/18', '2006.01'"
  ],
  "Applicants":[
    "' [SU]/[SU] (AllExceptUS)': 'VSESOYUZNY NAUCHNO-ISSLEDOVATELSKY INSTITUT RAZVED",
    "' [SU]/[SU] (UsOnly)': 'KUTUKOV, Viktor, Pavlovich'"
  ],
  "Inventors":"'PEVZNER, Aleksandr, Abramovich', 'PEVZNER, Lev, Abramovich', 'PRITSKER, Leonid, Semenovich', 'BUTUZOV, Yury, Alekseevich'",
  "Agents":"'THE USSR CHAMBER OF COMMERCE AND INDUSTRY'",
  "Designated_States":{
    "Global":[
      "AU",
      " DE",
      " GB",
      " US"
    ]
  }
}
```

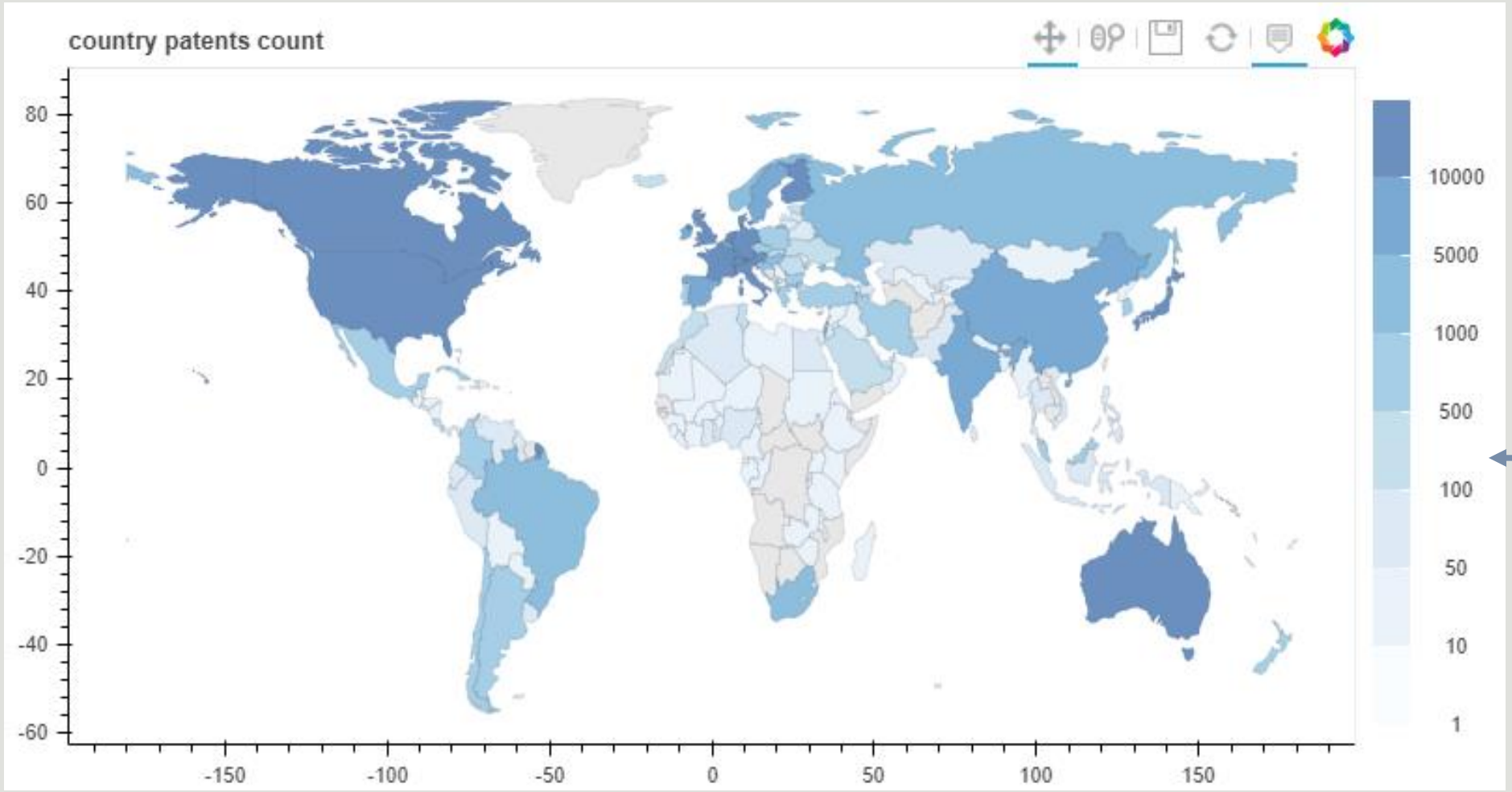

申請日期

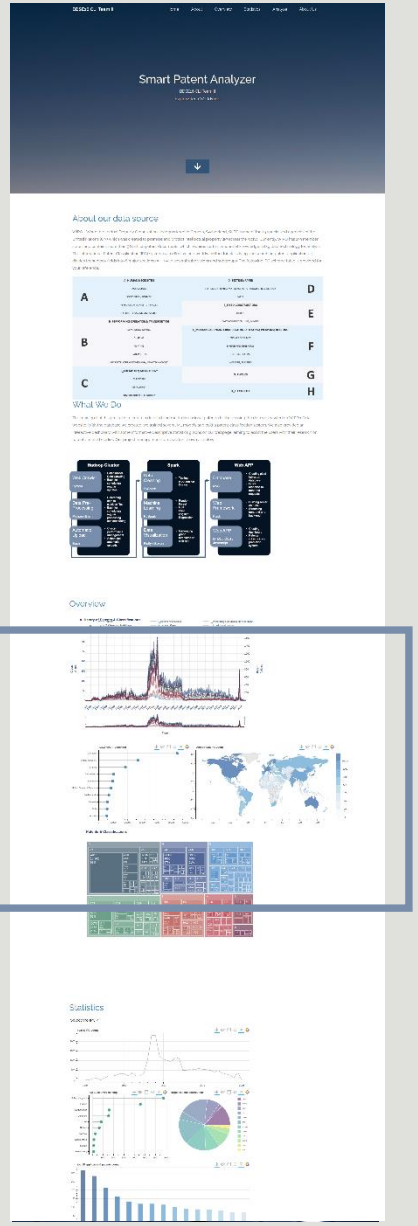
國際類別分類碼

申請人資料/國家

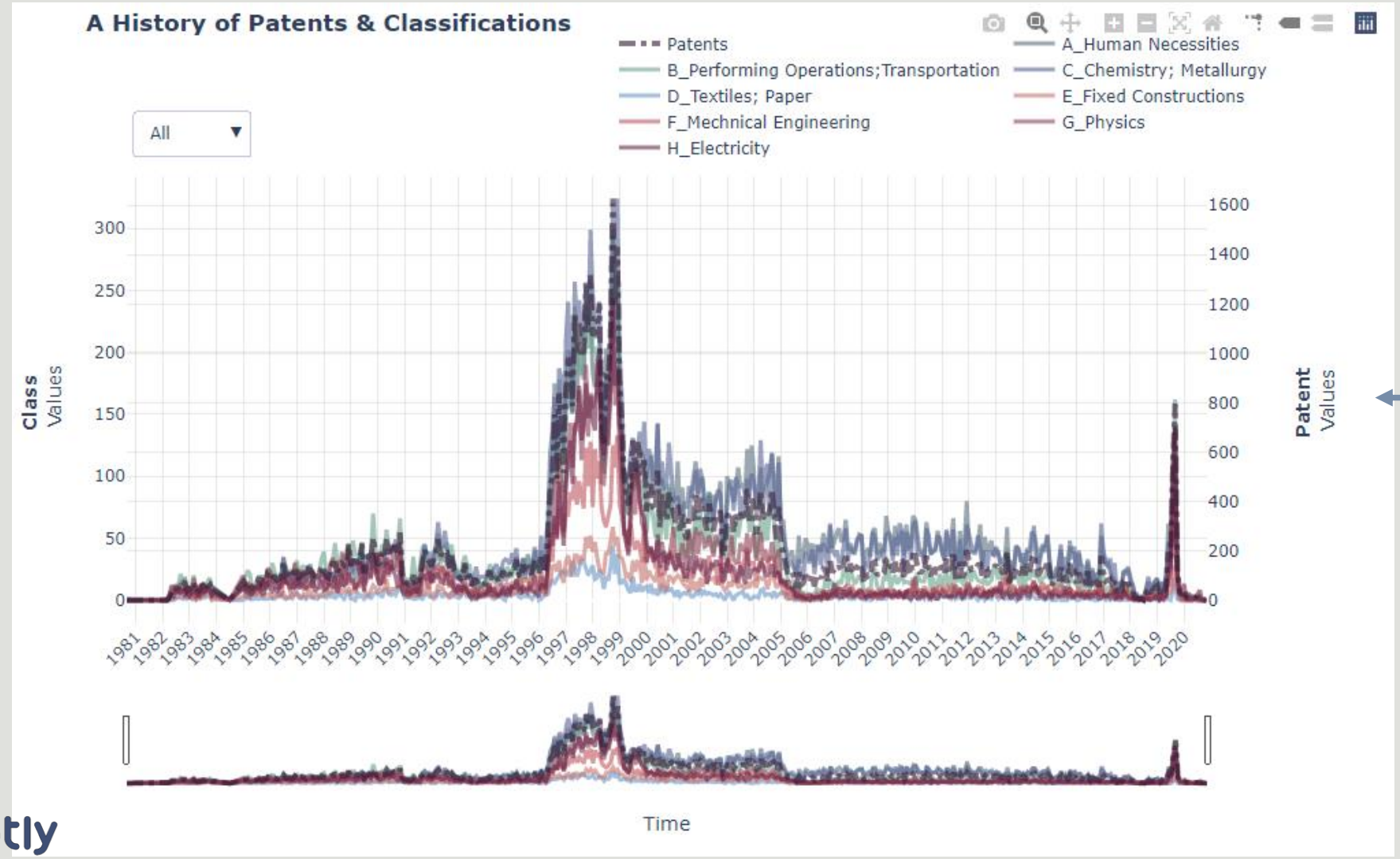
申請國家

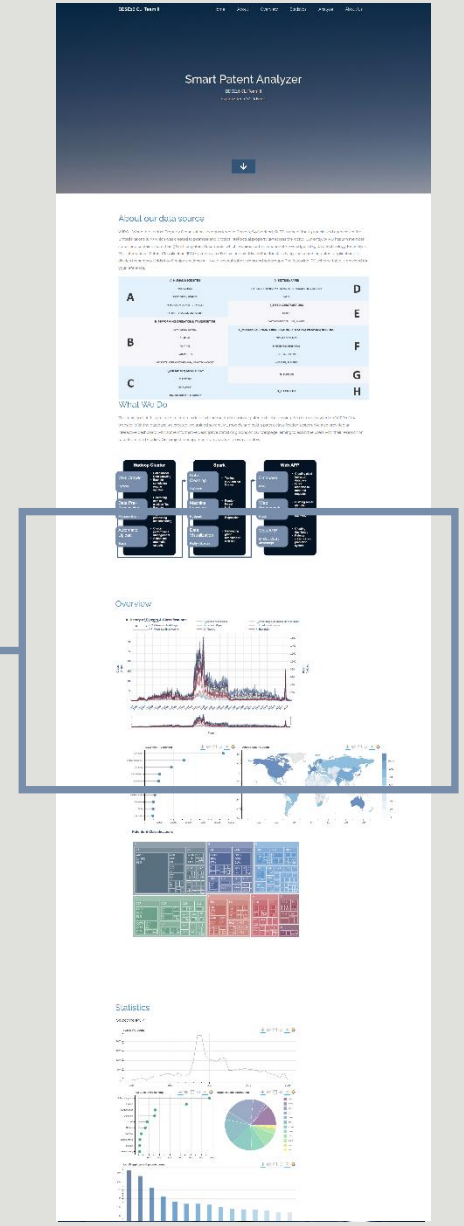
國家的專利申請數


country patents count
80
60
40
20
0
-20
-40
-60
-150
-100
-50
0
50
100
150
10000
5000
1000
500
100
50
10
1

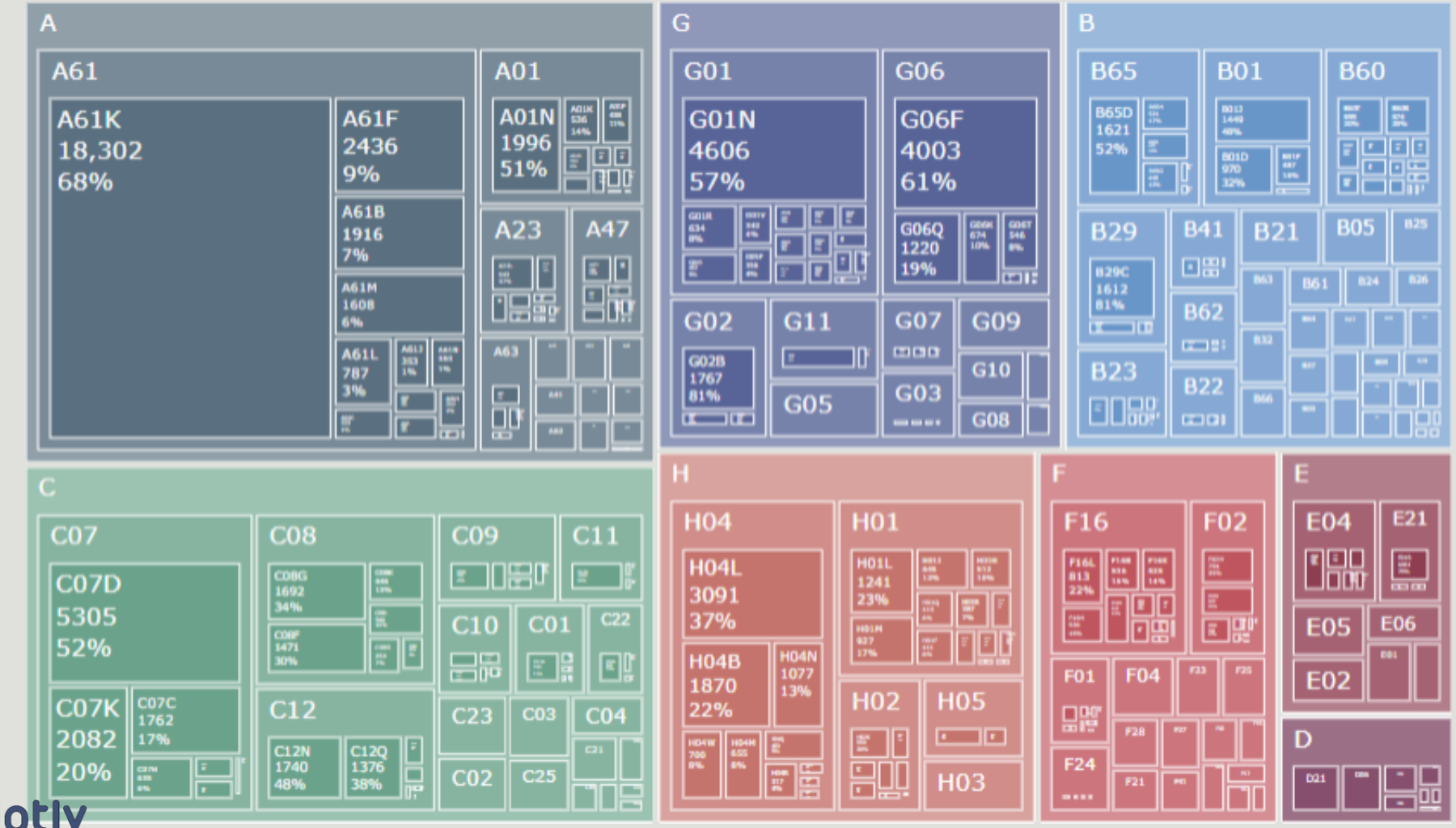

Smart Patent Analyzer
About our data source
What We Do
Overview
Statistics

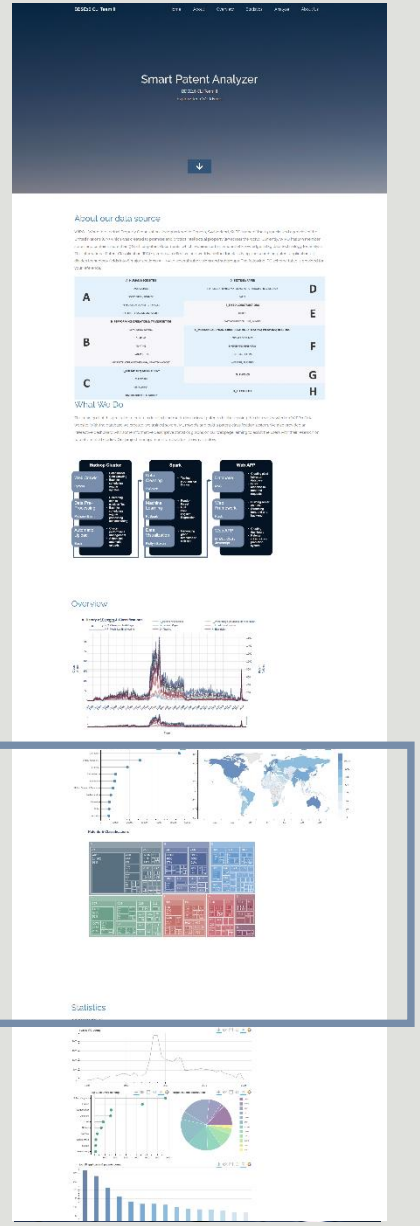
歷屆專利與分類變化


A History of Patents & Classifications
Patents
A_Human Necessities
B_Performing Operations;Transportation
C_Chemistry; Metallurgy
D_Textiles; Paper
E_Fixed Constructions
F_Mechnical Engineering
G_Physics
H_Electricity
All
Class Values
Patent Values
Time
plotly


Smart Patent Analyzer

專利與分類的階層關聯



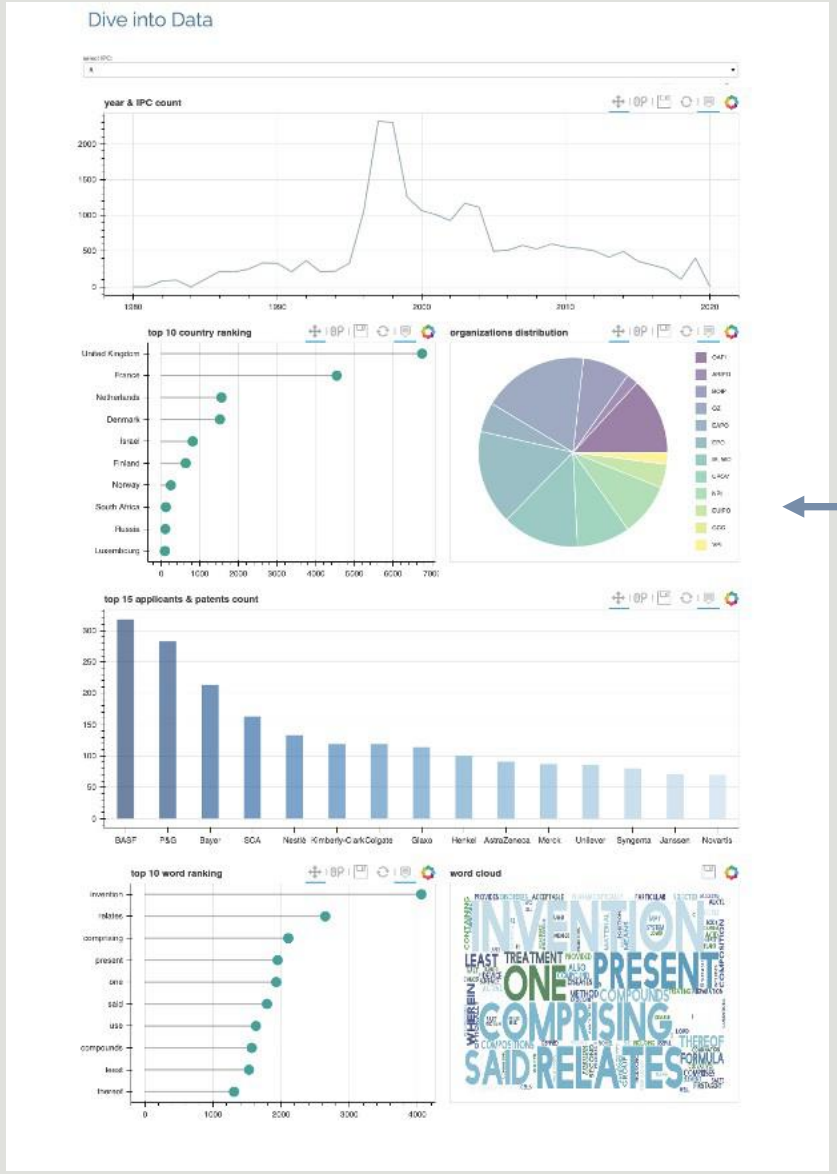


plotly

互動式圖表呈現

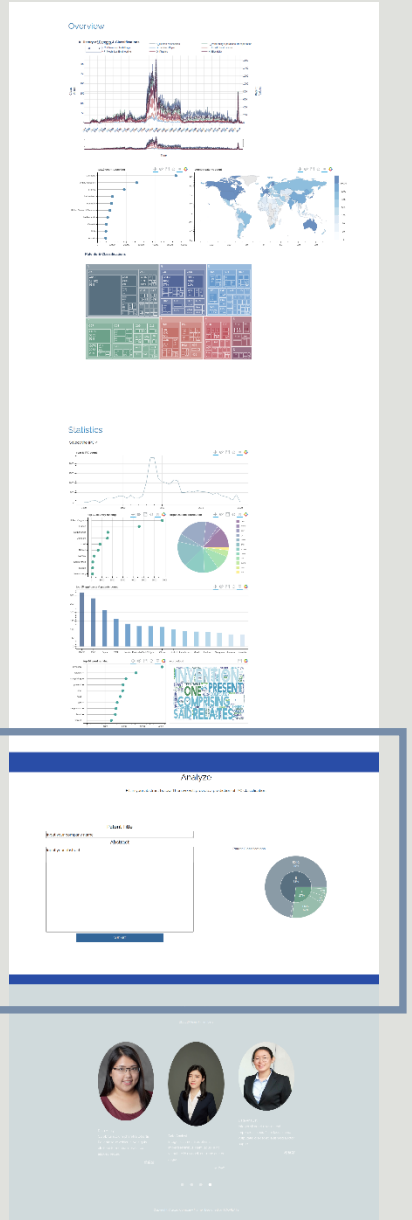
- 主要目的：
  - 協助企業研發策略布局

- 可篩選特定專利類別查看：
  - 年份專利數量增長趨勢
  - 國家/組織專利數量分佈
  - 各類別公司專利數量排名
  - 文字雲，可與專利分類落點分析對應



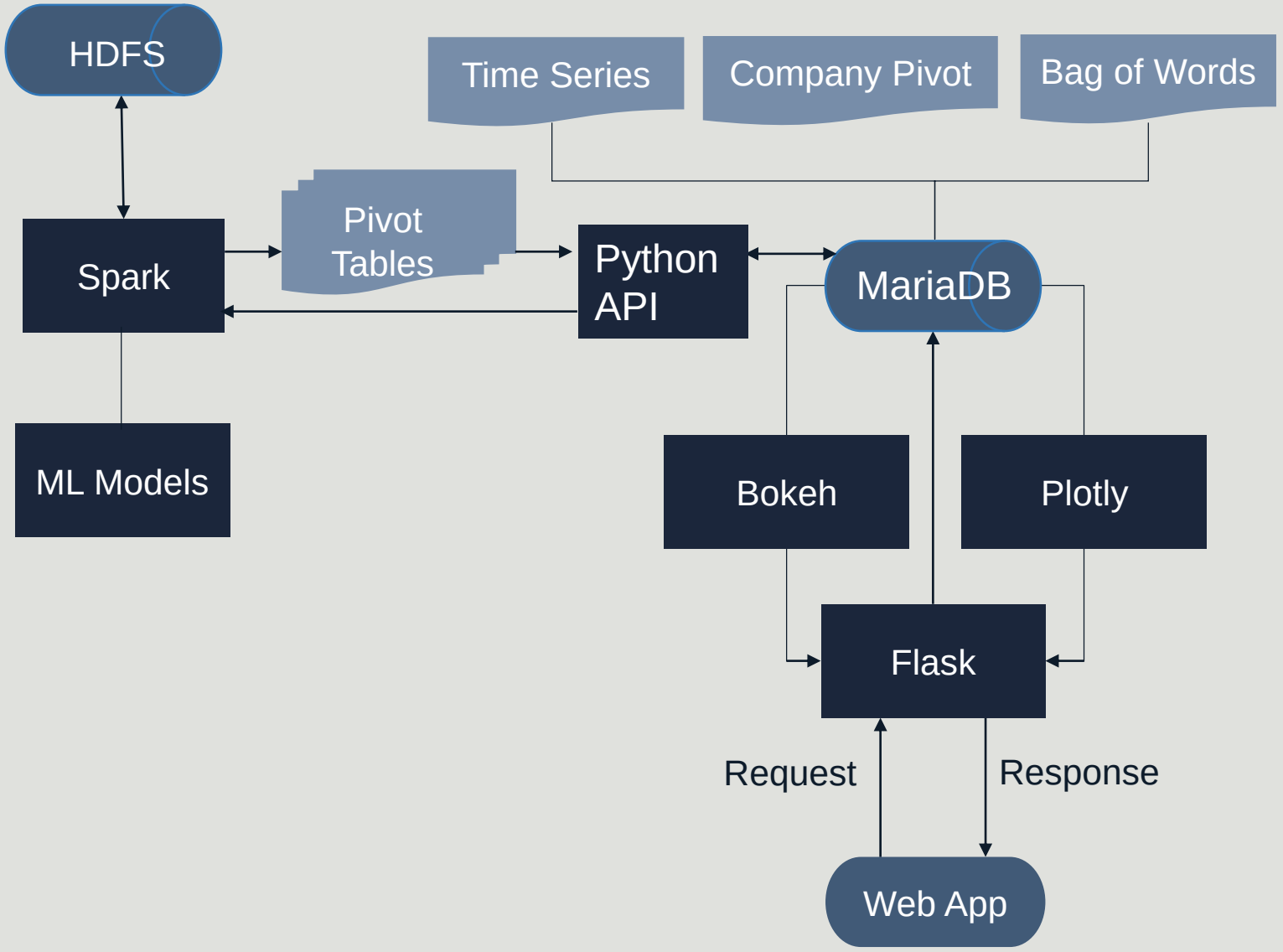
專利智慧分類落點分析


Analyze
Fill in your abstract below. Then we will provide a prediction of IPC classification.
Patent Title
Input your company name
Abstract
Input your abstract
SUBMIT
Patent Classifications
E04G
99%
E
73%
F
27%
F02C
67%



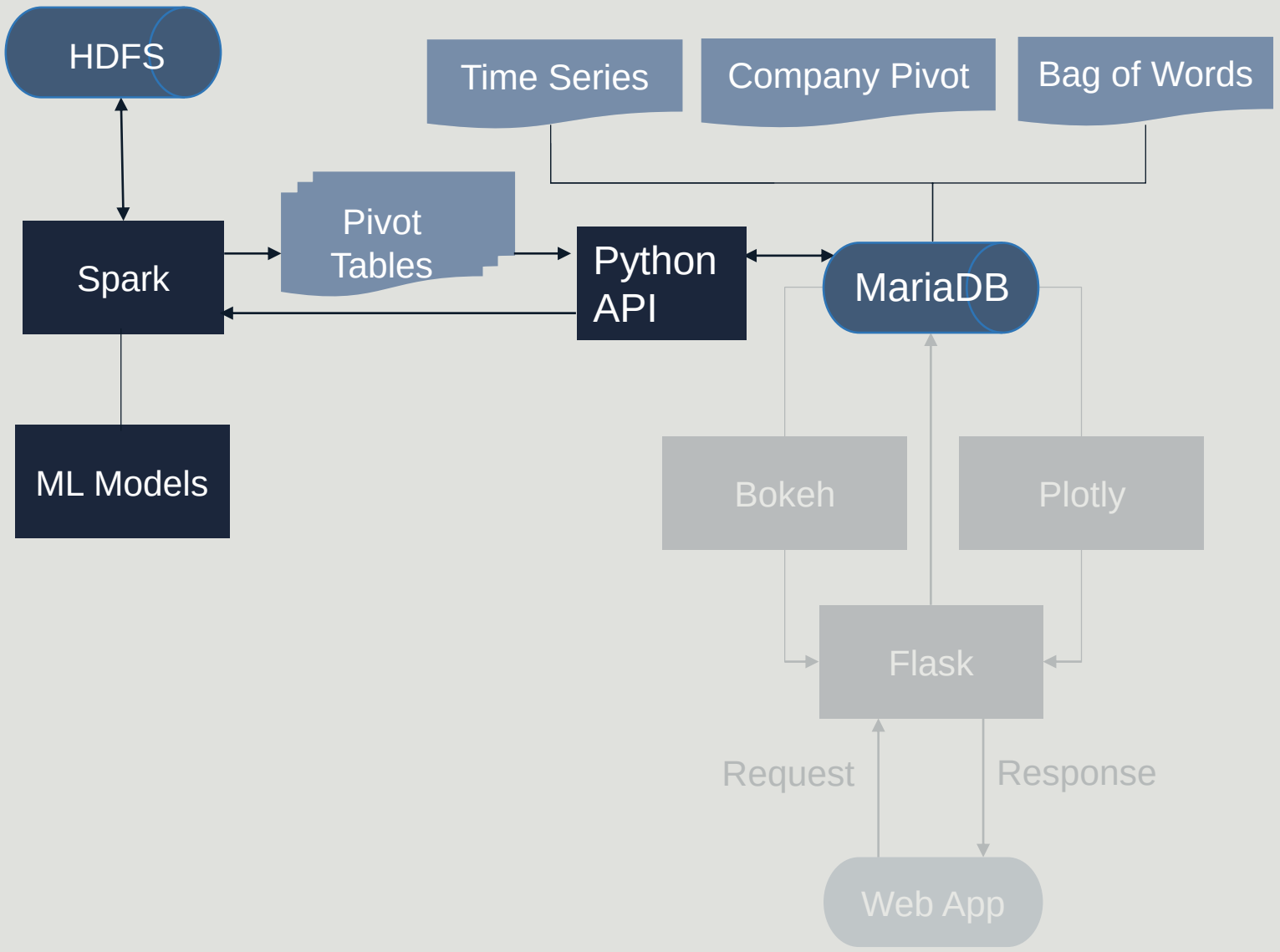
Web開發

- 目的：動態網頁、自動化更新
  - 串聯前端架構與後端資料庫



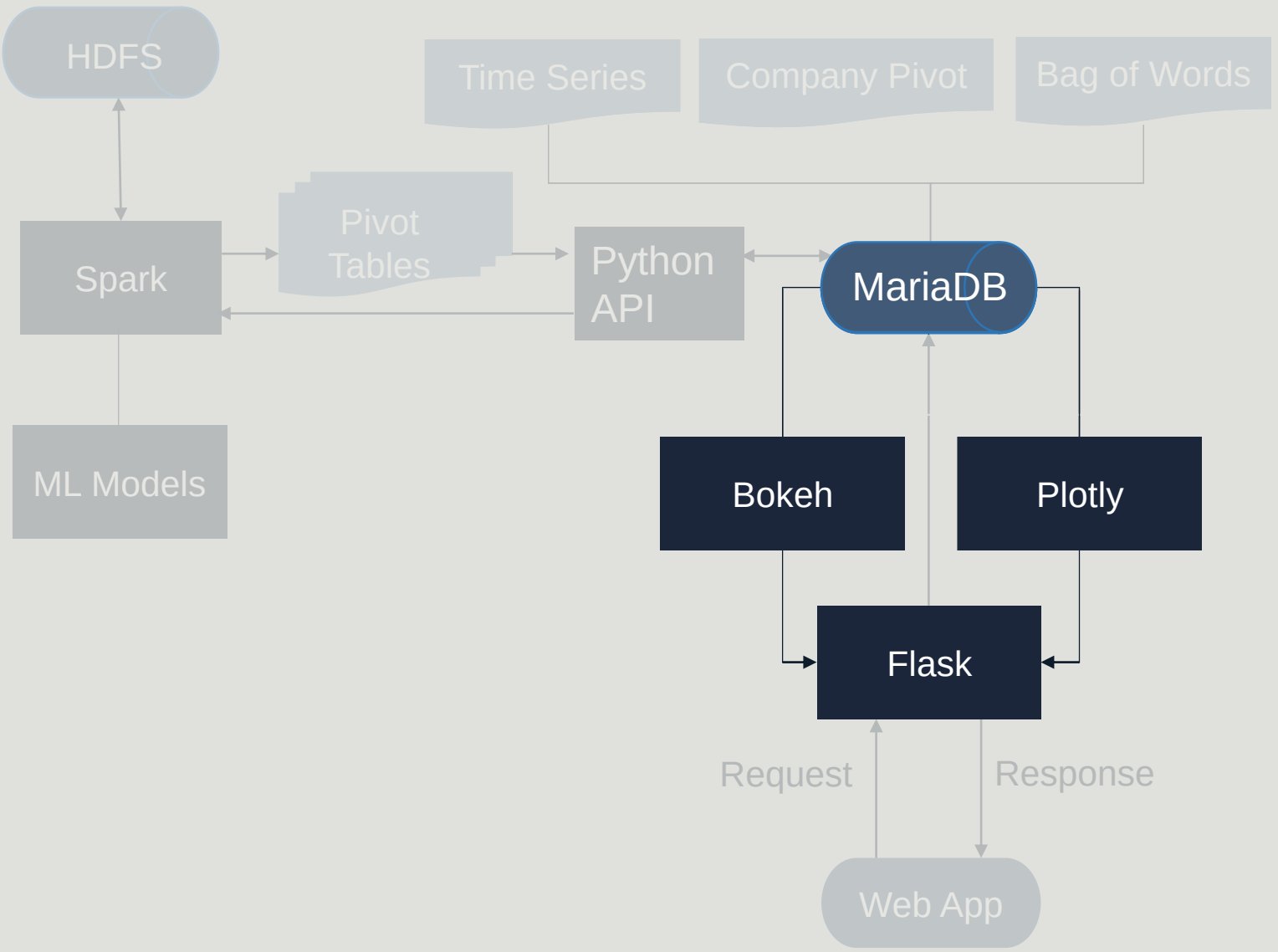
Web開發

- 目的：動態網頁、自動化更新
  - 串聯前端架構與後端資料庫
- 後端資料庫
  - Spark、MySQL



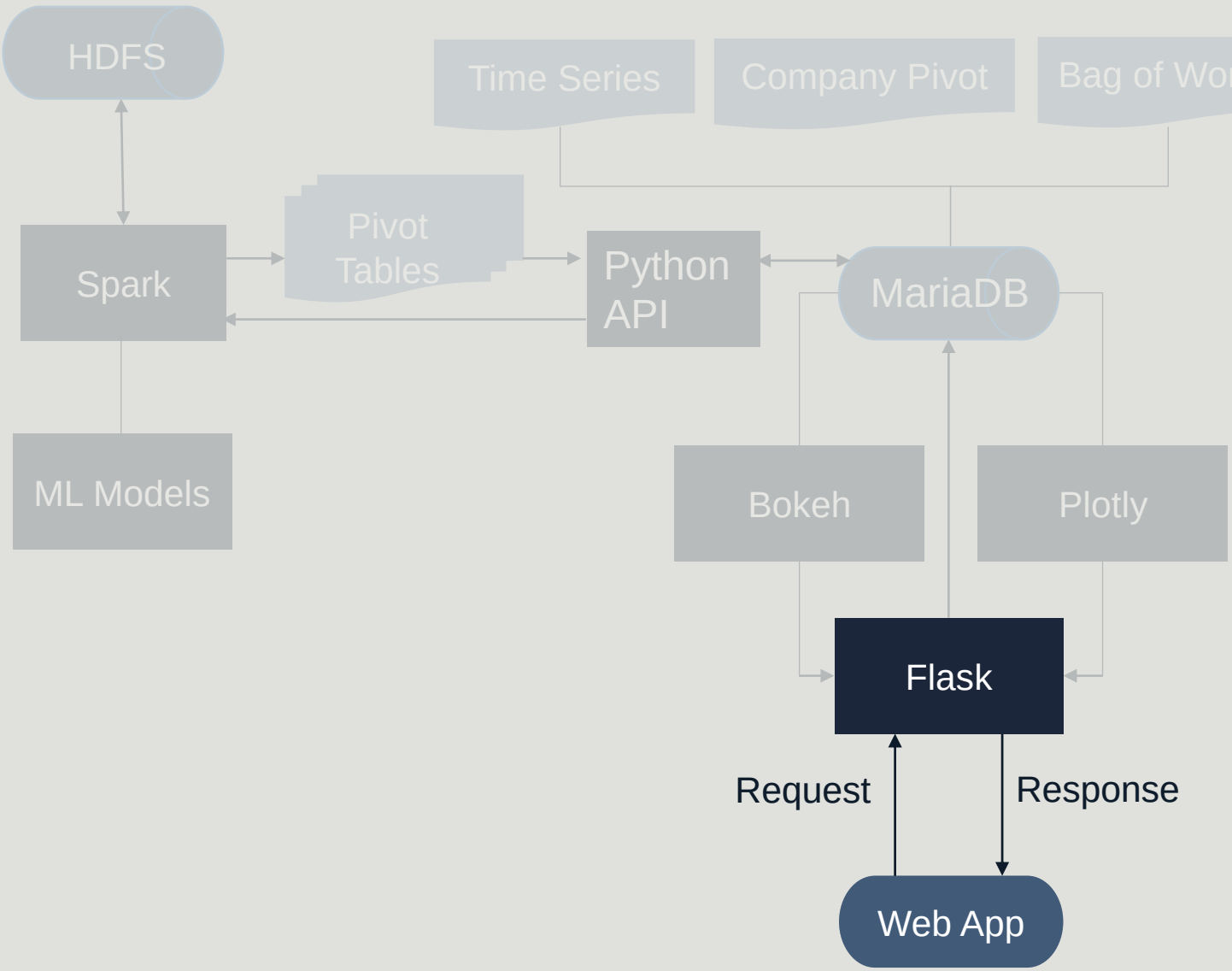
Web開發

- 目的：動態網頁、自動化更新
  - 串聯前端架構與後端資料庫
- 資料準備
  - Spark、MySQL
- 網頁後端
  - Bokeh、Plotly、 Flask



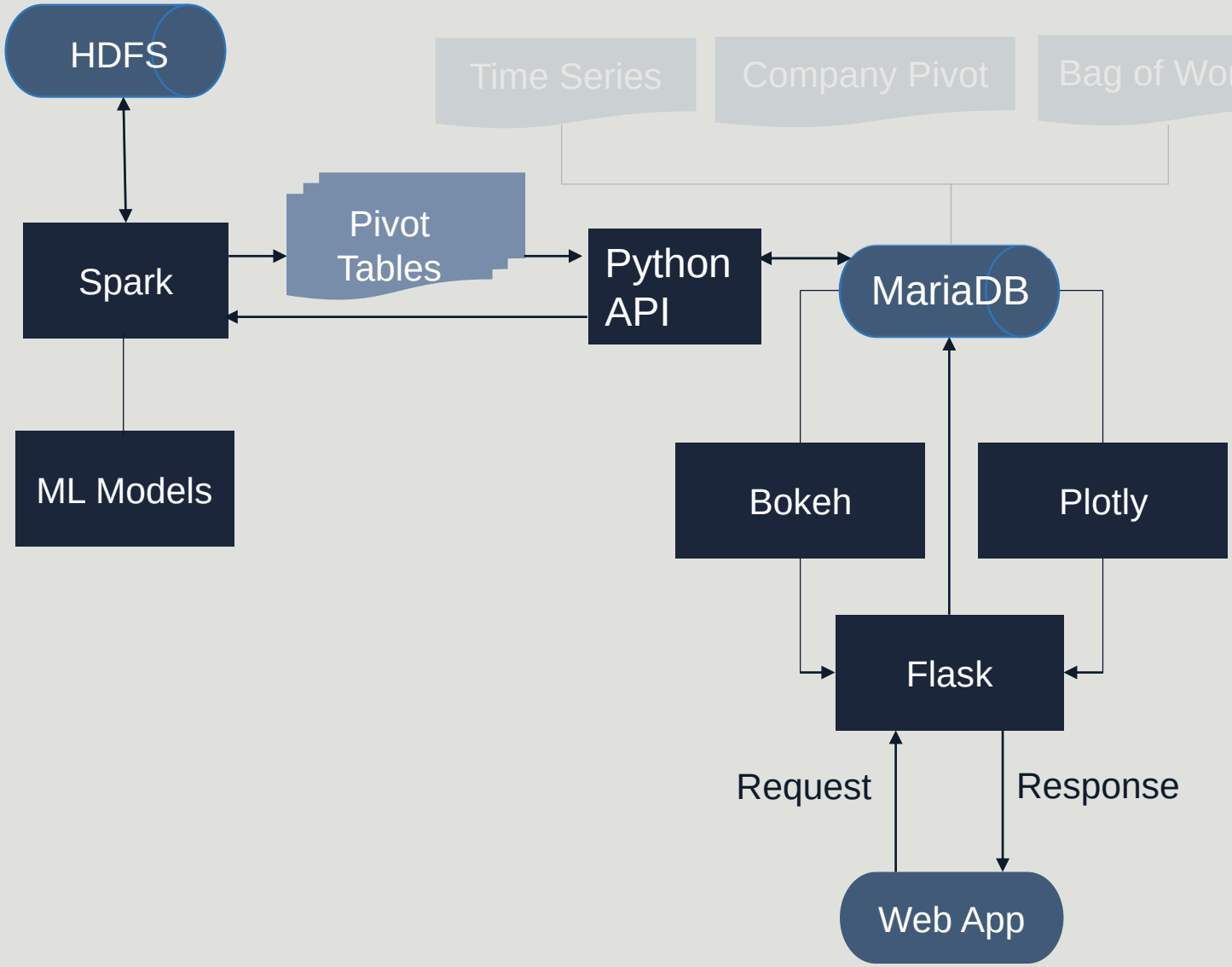
Web開發

- 目的：動態網頁、自動化更新
  - 串聯前端架構與後端資料庫
- 資料準備
  - Spark、MySQL
- 網頁後端
  - Bokeh、Plotly、Flask
- 網頁前端
  - HTML、CSS、BootStrap、JavaScript

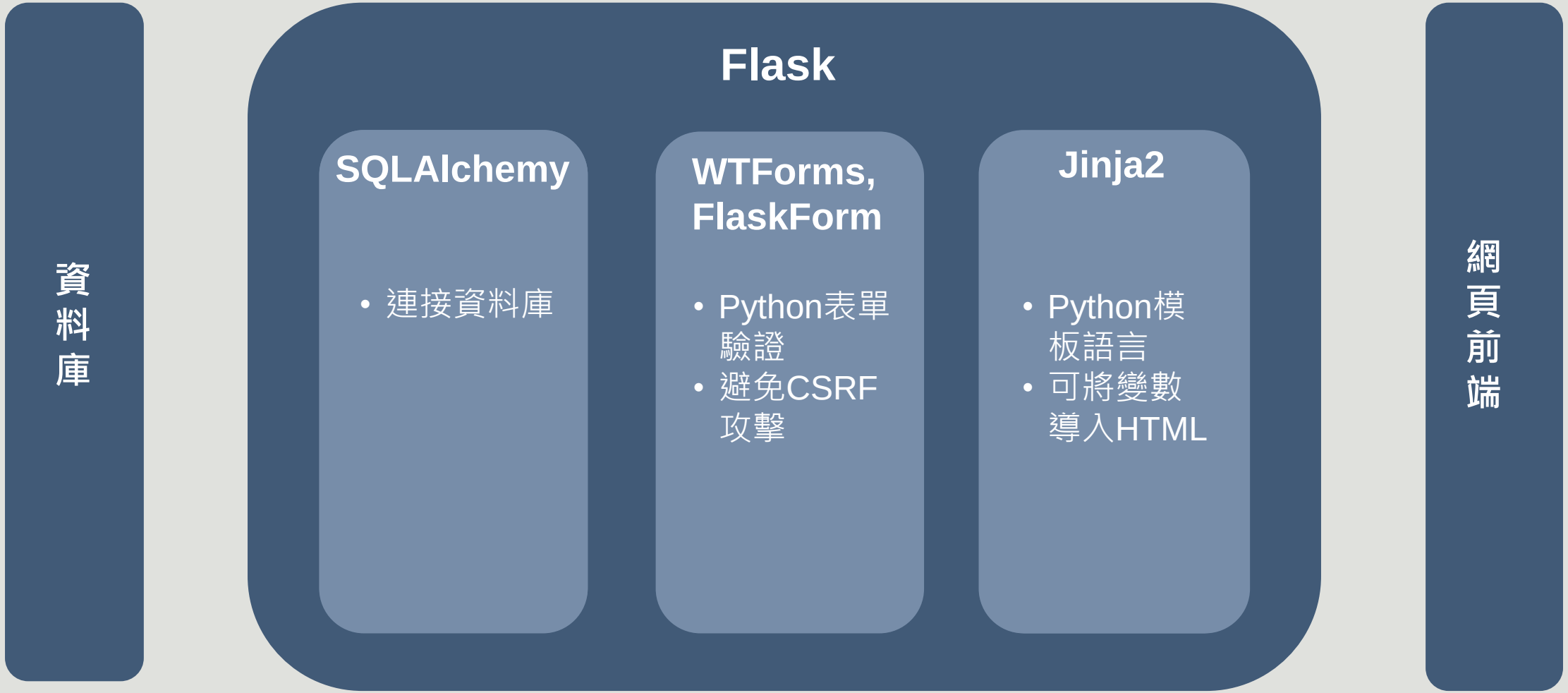


Web開發

- 目的：動態網頁、自動化更新
  - 串聯前端架構與後端資料庫
- 資料準備
  - Spark、MySQL
- 網頁後端
  - Bokeh、Plotly、 Flask
- 網頁前端
  - HTML、CSS、BootStrap、JavaScript



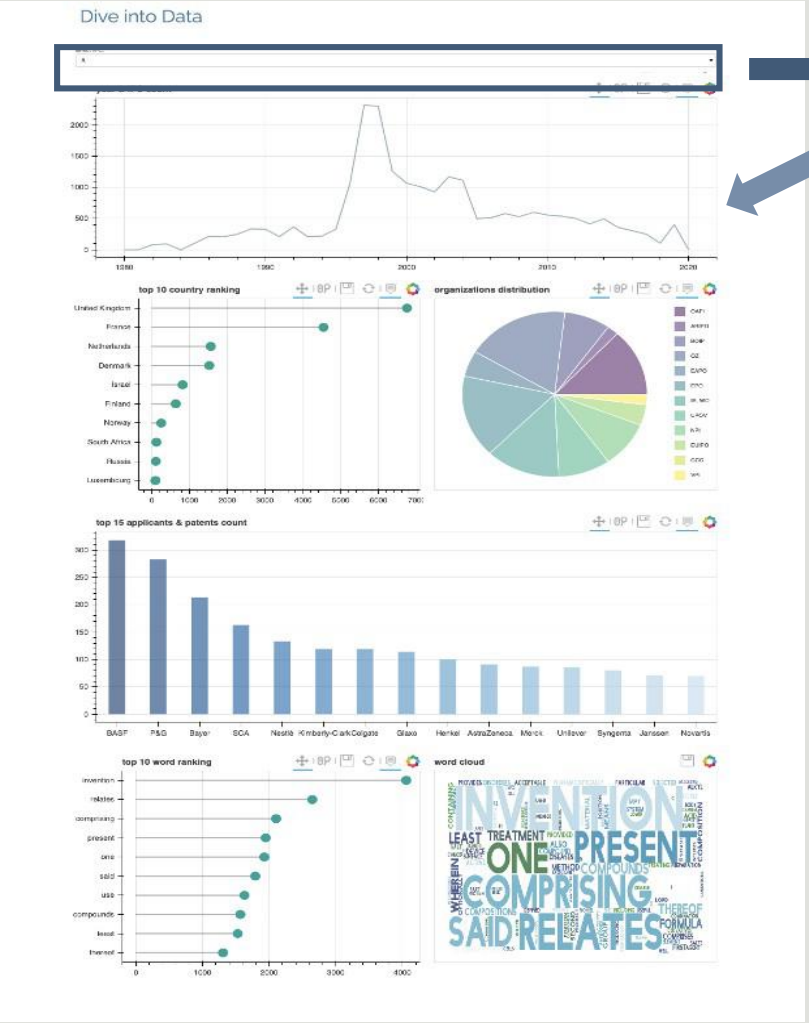
Flask


資料庫
Flask
SQLAlchemy
• 連接資料庫
WTForms, FlaskForm
• Python表單驗證
• 避免CSRF攻擊
Jinja2
• Python模板語言
• 可將變數導入HTML
網頁前端

資料視覺化

講者:楊雅婷

儀表板結果呈現



Web APP

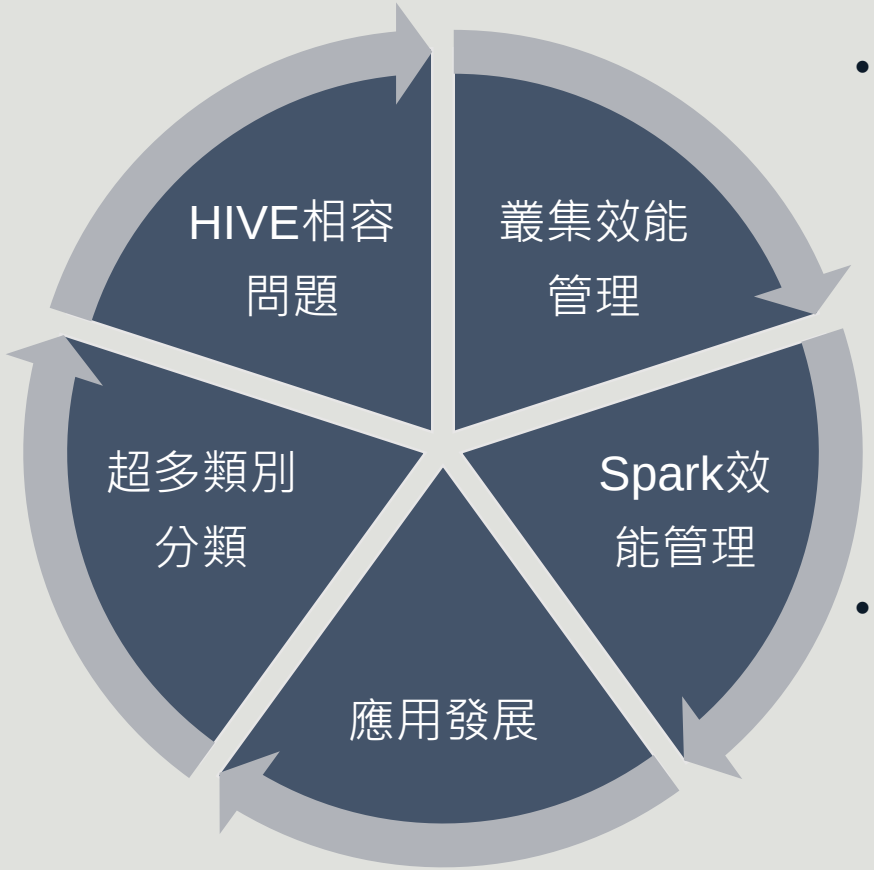
網頁前端
HTML/CSS/
JS/BootStrap

前端架構
Flask

資料庫
MySQL

未來展望

- 將資料輸出為樞紐分析表，並存入MySQL資料庫中作中轉站



- 批次處理上傳資料，並建立監控機制
- 改用許多二元分類模型組合為一套訓練模型集合，分別訓練各個不同之類別。結果亦表明此法可提升準確率
- 定期壓縮資料成parquet及bzip2檔，並調校spark session

- 改善前端網頁應用功能
- 優化機器學習之模型表現
- 連結更多與產業、財經方面之分析，完善並最大化專利分析所帶來之效用

16th 第二組

智慧戰略
專利分析

Thank You

王予欣 / 王昱 / 王嘉宏 / 曾昱璇 / 顧芠菱 / 楊雅婷