### 植基於個人本體論的 新聞推薦系統

資工四 戴均民 朱奕安 郭書佑 指導教授 廖宜恩

# 研究內容

- **)** 1.新聞系統
  - 2.推薦功能
  - 3.系統定位

#### 新聞系統





#### 推薦功能--Yahoo



後昨天首度造訪紐約,據美聯社報導,比賽收視率創下MSG本季轉播新高。 麥迪遜花園廣場電視台(MSG Network )本季轉播的前18場比賽中,平均收視率為

会想域佔國廣場電視日(MISO NetWOR)本字轉編的則10域比賽中,平均收視平為 3.41,昨天火箭與尼克之戰,收視率顯升至4.77,顯示紐約球迷,還是非常懷念這個擁有 高學歷、黃色皮膚的籃球高手。

本季轉披休士頓火箭隊戰袍的林書豪,昨天前往紐約麥迪遜花園廣場踢館,面對前東家, 他出手不留情,全場15投9中,拿下全隊次高的22分,另送出8助攻,率隊以109比96扳倒 尼克,打破了尼克主場不敗金身。

昨天比賽開始前,現場進行球員介紹,兩隊球員依序唱名,當叫到林書豪時,球迷仍給予 他熱烈掌聲,當第4節星聲,他被換下場時,仍有不少死忠粉絲起立鼓掌。

上季在尼克隊打出身價的林書家,季末火箭隊開出3年2500萬美元的合約,希望網羅這位「哈佛小子」,尼克原有機會留下他,只要球團願意跟進出價,林書家就得繼續留在紐約,但尼克並沒這麼做。

林書家確定加盟火箭後,一度造成MSG股價暴跌,兩天內市值蒸發4836萬美元(約新台幣14億元),這些例子在在都顯示,豪小子在紐約客心目中的地位。1011219





### 推薦功能--Google





### 推薦功能

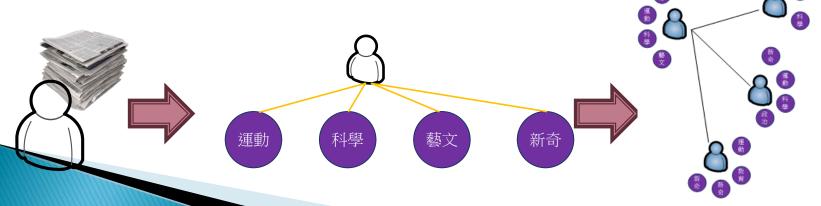




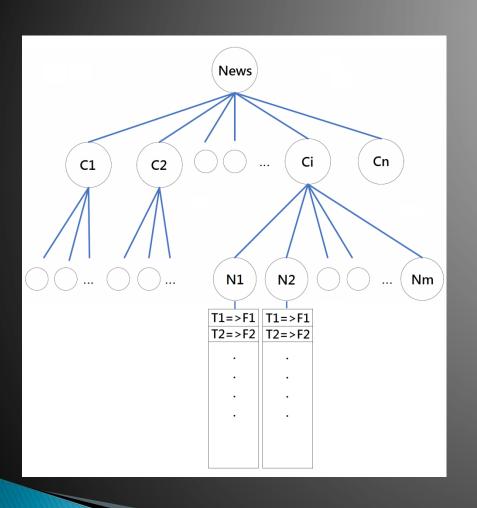
### 系統定位

- 挖掘使用者興趣特徵
  - 。個人本體論
- 根據興趣推薦新聞給使用者
  - 。 新分類新聞推薦
  - 。協同過濾推薦
- ▶ 隨時隨地取得新聞





### 個人本體論



- ▶ C表示分類(Class)
- ▶ N表示(News)
  - 。一個節點為詞(Term)與 頻率(Frequency)的列 表
- 最大的樹為所有新聞的總和
- 》視使用者為一個樹, 有屬於自己的樹圖。

### 演算法—新分類新聞推薦

類別 排行	1	2	3	4	5
權重 =wi	5	4	3	2	1

- ▶ 1.對使用者樹圖
  - 。賦予類別權重
  - 賦予類別下詞權重

詞頻排行	1	2	3	4	5
權重再修正 =Wj	Wi*5	Wi*4	Wi*3	Wi*2	Wi*1

- ▶ 2.對新聞文本樹圖
  - 。 擷取詞與頻{Term,Frequency}
  - 。判斷字詞是否出現在使用者樹圖之下
  - 。若有則該新聞權重上升Frequency\*Wj
- 3.得到該新聞對使用者的重要性

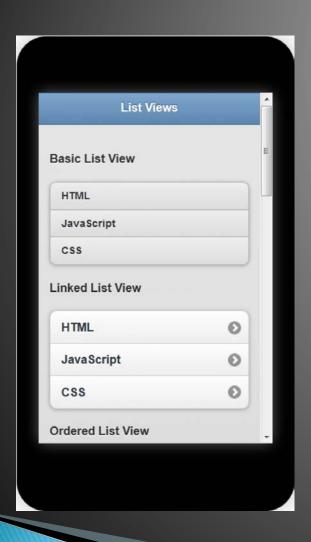
### 演算法—協同式過濾

▶ Cosine Similarity餘弦相似性

similarity = 
$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

- ▶ Ai、Bi表示AB分類下的字詞權重
- ▶ 此方法會計算出介於0~1的值
- 對每一個使用者都以此方式計算同好者

### 開發環境



- jQuery Mobile
  - 簡潔的使用環境
  - 好的可攜性

# 新聞推薦系統Demo

瀏覽功能推薦功能未來展望

### 瀏覽功能

<b>∢</b> Back	UDN焦	點要聞	▲ 登出
功能選單	今日推薦	個人資料	設定
Filter ite	ms		
<b>6死因伏法</b> 2012-12-21 21	<b>陳金火、廣復</b> :26:46	<b>感強劫財劫色</b>	分屍
法務部批准 2012-12-21 21	<b>6死囚今晚</b> 位 :26:46	犬法	0
詹宏志:小 2012-12-21 19	、確幸、微加: :54:24	幸福瀰漫臺灣	•
幸福感指標 2012-12-21 03	票不是只有 <b>GI</b> :49:57	DP	0
民眾自求多 2012-12-21 03	·福…「其實) ::18:25	台灣沒那麼糟	ı Đ
<b>你幸福嗎?</b> 2012-12-21 03	' 台灣居亞洲 :18:25	四小龍之首	0

<b>∢</b> Back	新聞ID	▲ 登出	
功能選單	今日推薦	個人資料	設定

#### 學者:馬雅沒末日說是指新紀元

2012-12-18 03:07:45

馬雅末日說最近在全球與起熱潮,曾在墨西哥鑽研馬雅文明的淡江大 學美洲所所長陳小雀,昨天以「破解謎樣馬雅」為題演講時指出,馬 雅文化不可能有末日之說,今年十二月廿一日其實只是新紀元的開 始。「中美洲國家都沒這麼熱中,其他地區卻瘋狂討論世界末日 說,真的很奇怪耶。」陳小雀在墨西哥國立自治大學攻請博士時,曾 深入研究馬雅文化。她說若要探討二o一二末日說,得先了解馬雅曆 法。馬雅人同時使用太陽曆、祭祀曆(卓爾金曆)及長紀曆三種曆 法;太陽曆以廿天為一個月,十八個月便是三百六十天;但善於觀星 她說,馬雅人此一特別的計算方式,是因為採廿進位法。從太陽曆可 看出馬賴人的天文及數學能力很職,因第十九月只有五天,被視為不 祥禁忌日,期間凡事須小心。 祭祀曆也以廿天為一個月,但只有十 三個月,太陽曆每走五十二年,會和祭祀曆重疊,馬雅人相信每五十 二年就是一個新的循環。 長紀曆計算日子的單位分成「金」(kin, 一天)、「鳥依納」(uinal, 廿天)、「敦」(tun, 三百六十 天)、「卡敦」(katun,七千二百天)及「巴京敦」(baktun,十 四萬四千天),其中又以十三巴克敦為分界,五千二百年為一個循 環。 自西元前三一一四年八月十三日開始計算,十三個巴克敦就是 今年十二月廿一日,因此被當成末日預言,但在馬雅文化只是代表新 紀元開始,沒有末日之說。 ※延伸閱讀》 · 影音/「天文學家談末 日, 系列影片・FB粉絲團/中央大學「末日重生」歳末慶典・朱學 悔評論孫維新談末日說 · 孫維新回應朱學恆險書批評

#### 杳看原新間網址

### 瀏覽功能

<b>∢</b> Back	依日期	看新聞	▲ 登出
功能選單	今日推薦	個人資料	設定
2012-12-2	1		495
2012-12-2	0		490
2012-12-1	9		493
2012-12-1	8		486
2012-12-1	7		459
2012-12-1	6		449
2012-12-1	5		829
2012-12-1	4		1132
2012-12-1	3		1024
2012-12-1	2		1143

<b>∢</b> Back	依分類	看新聞	<b>A</b> 2	胜
功能選單	今日推薦	個人資料	詉	Ĕ
UDN熱門	推薦-最新報	導	13796	0
UDN財經	股市-股市要	間	1257	0
UDN熱門	推薦-經濟日	報發燒新聞	1112	0
UDN焦點	要聞-社會		1009	0
UDN焦點	要聞-兩岸		686	0
UDN焦點	要聞-國際		621	0
UDN熱門	推薦-聯合晚	報發燒新聞	589	0
UDN焦點	要聞		573	0
UDN健康	休閒-健康醫	藥	547	0
UDN熱門	推薦-發燒新	間	500	0

### 推薦功能—同好者推薦



<b>∢</b> Back	相似度推薦展示		▲ 登出
功能選單	今日推薦	個人資料	設定
© Filter item	S		
中市委外活動 2012-10-29 05:5	動年花 <b>13億「要</b> 分 0:08	〉(僕幹嘛)	•
最新調查:E 2012-10-29 21:5	<b>已婚婦女期待生</b> 身 2:07	身孩降至10%	•
山腳國小發展 2012-10-29 22:29	展客語教育 屢獲化 5:40	<b>圭績</b>	•
捷運宅失寵 2012-10-29 23:1	南方之星四成流 9:17	泛標	•
華通第3季獲 2012-10-30 12:3	利李增106.48% 5:50	每股純益0.21元	•
嘴唇被要求版 2012-11-04 04:16	豆應 賴雅妍吻戲好 6:10	序害羞	•
仁寶 本季獲 2012-11-02 04:12			0

### 未來展望—使用者分群

	修改	修改資料	
功能選單	今日推薦	個人資料	設定
電子郵件			
ivoivoitoito02@y	ahoo.com.tw		
目前密碼			
必須填寫密碼才可	可修改資料		
新密碼			
若不修改密碼此材	<b>闌請</b> 留空		
新密碼確認			
若不修改密碼此相	<b>闌請</b> 留空		
姓名			
朱奕安			
生日			
1990-12-12			
性別		男女	
回功能	選單	修改	



### 未來展望一同義字合併

a=林書豪 A=Jeremy Lin bcde為其他

- 文本A={a,a,A,b,c,d}
- ▶ 文本B={a,A,A,b,c,e}
- 餘弦相似性
  - $\circ = 6/8$
  - $\circ = 0.75$

若將a與A合併為aA

- 文本A={aA,aA,aA,b,c,d}
- ▶ 文本B={aA,aA,aA,b,c,e}
- ▶ 餘弦相似性
  - $\circ = 11/12$
  - $\circ = 0.917$

合併前相似度=0.75

合併後相似度=0.917

### Any Questions!

# Thank you for your attention!