

---



# 研究方法期末報告

從使用者自建內容挖掘網民對議題的立場：  
以兩岸服貿為例

# 目錄

- 01 背景動機與研究問題
- 02 研究流程與方法
- 03 彙總與觀察
- 04 結論與未來展望



---

# 背景與動機

作者 sunbysea (忠言逆耳良藥苦口)  
標題 [新聞] 50萬臺灣人反服貿抵制中共 全球49個城市  
時間 Mon Mar 31 21:41:54 2014

臺灣重量級出版業者郝明義、作家平路等十多名各行各業的專家輪番在集會上發言，陳述跟中共的服貿協議傷害臺灣本土自身行業及危害整體國家安全，要求馬英九政府尊重民意，退回跟中共的服貿協議。

---

# 研究問題



電池壽命、外觀  
大小、重量

---

# 研究問題

- ✓ 對於社會和政治議題的立場分析
- ✓ 協助使用者將複雜、非結構化的文本 轉換為 結構化資料

# 研究流程

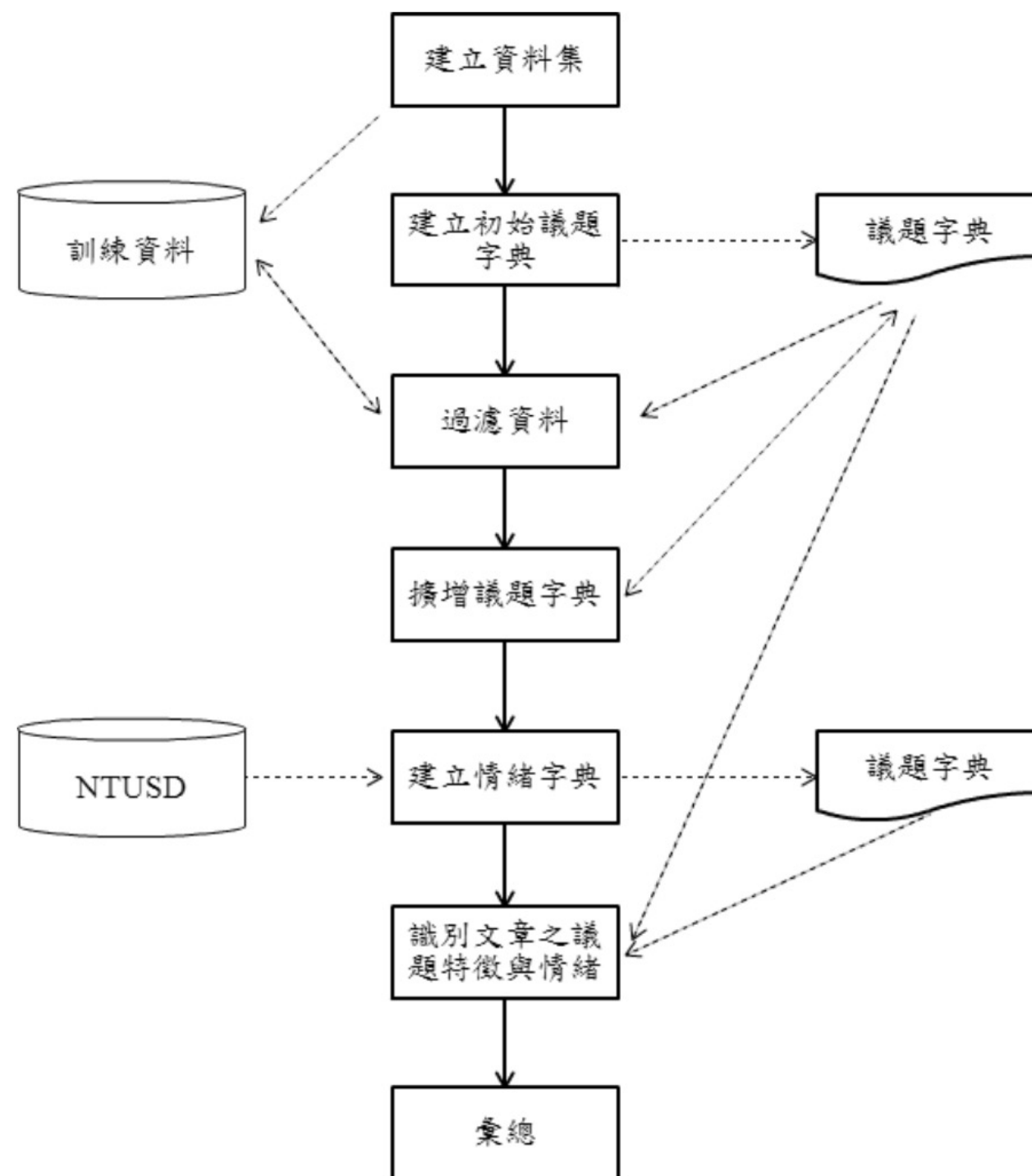


圖 3-1、研究方法流程圖

---

## 建立資料集

- ✓ 資料來源：PTT 八卦版、政黑版、政治版、國民黨版、民進黨版  
與「服務貿易協定」相關文章

---

## 建立資料集

### ✓ 文章收錄結構：

文章 = {文章編號, 標題, 作者, 內文, 時間, 推文數量, 回應文章對象}



# 建立資料集

作者 indium111 (原子核的奧秘) 看板 Gossiping  
標題 Re: [問卦] 服貿對醫療業勞工的影響  
時間 Wed Mar 19 21:44:57 2014

\* 引述《Merlion (台台)》之銘言：  
：對台灣醫院怎麼影響我先不想管  
：就算全變中資醫院 只是老闆換成中國人 資金變成中資  
：但醫生護士還是台灣人  
：然後不管是中國老闆還是台灣老闆都得依法遵照台灣的醫師訓練制度去走  
：但究竟對醫生護士藥師收入影響是好是壞???

台灣的醫療市場比大陸小，加上健保的框架下  
要有突破性的發展也很難，因為總額制度下，賺太多就等著被核刪  
不過台灣醫療人員的平均水準比大陸高  
如果我是陸資老闆的話，拿下醫院經營權  
就把台灣的醫療人員調到大陸的醫院，提升大陸醫院的水平來擊敗對手

至於台灣醫院的空缺，把大陸不好的醫療人員調到台灣來就解決了  
反正兩邊都是講中文的，沒有語言溝通上的問題  
至於醫療執照的問題，反正立法院都可以30秒過服貿了  
要讓大陸的醫生來台灣有那麼難嗎?

--  
\* 發信站: 批踢踢實業坊(ptt.cc)  
◆ From: 220.134.57.127  
→ taso5566: 台灣的好醫師只負責權貴, 還輪不到平民百姓 03/19 21:46  
→ innominate: A. 醫院服務業: (4)除模式(3)所列內容外, 不予承諾。 03/19 21:46  
推 sa374724: 醫院現在就極其所能的壓榨護理人員了, 未來直接嗆我有大 03/19 21:46  
→ sa374724: 陸護士要不要做隨便你 03/19 21:46  
推 Merlion: 所以被調到大陸收入會不會三級跳 ㄎ~~~三級跳就爽喔! 03/19 21:47  
推 aqw445566: 經濟部表示: 讓青年人... 03/19 21:47

文章 = {P0002, RE:服貿對醫療業勞工的影響, indium111, 台灣的醫療市場..., 2014/03/19, 6, 標題為: 服貿對...(P0001) }

---

## 建立初始議題字典

- ✓ 目的：從資料集中過濾不相干的文章內容
- ✓ 議題字典 = { 同義字集(S), 等位詞集(E), 議題面向字集(A) }

---

## 建立初始議題字典

- ✓ 同義字集 (S)：欲分析議題的同義字眼（服貿、福懋…等等）
- ✓ 等位詞集 (E)：議題的相關字眼（自由貿易協定、跨太平洋夥伴關係…等等）
- ✓ 議題面向字集 (A)：此議題的面相（產業、經濟、國安…等等）

---

## 過濾資料

看板	起迄時間	文章總數	過濾剩餘數	是否採用
國民黨版	2003.11-2014.07	18077	20	否
民進黨版	2001.08-2014.07	18083	95	否
政治版	2001.12-2014.07	18278	266	否
政黑版	2006.04-2014.07	50492	7600	是
八卦版	2005.06-2014.07	95891	9504	是

---

## 擴增議題字典

- ✓ 依照名詞出現的頻率進行篩選
  - 利用斷詞，取出頻率超過10次以上的名詞單詞
  - 優點：可找出網民於論壇中較戲謔的特徵值字眼，進行字眼取代
  - 缺點：需要人工的判別

---

## 擴增議題字典

- ✓ 利用特徵值的句型關係（名詞 + 連接詞 + 名詞）
- 詞類標記後，進行迭代計算模式來添加額外的字眼

群眾 (Na)	現在 (Nd)	要 (D)	學習 (VC)	負起 (VC)	公民 (Na)	的 (DE)	義務 (Na)	， (COMMACATEGORY)	
學者 (Na)	和 (Caa)	媒體 (Na)	必須 (D)	扮演 (VC)	解釋 (Na)	和 (Caa)	傳播 (VC)	的 (DE)	角色 (Na)

---

## 建立情緒字典

- ✓ 使用NTU 情緒字典 (NTUSD)
- ✓ 考量該字詞於政治領域中情感極性的變換
- ✓ 負向詞過濾原則

---

## 建立情緒字典

### ✓ 負向詞過濾原則

- 負面詞：「噓」標籤 -> 負向詞
- 正面詞：若 NTUSD 的正向詞出現在某篇推文中 -> 維持正向
- 排除：若 NTUSD 一般性情緒字眼都未出現 -> 從情緒字典排除



---

## 識別文章之議題特徵與情緒

### ✓ 句子剖析 (建構依賴關係)

- 利用史丹佛剖析器 (Stanford parser) 剖析句子句法

#### Typed dependencies

```
nsubj (抹黑-6, 媒体-1)
prep (抹黑-6, 为了-2)
assmod (利益-5, 自己-3)
assm (自己-3, 的-4)
pobj (为了-2, 利益-5)
root (ROOT-0, 抹黑-6)
assmod (民众-9, 抗议-7)
assm (抗议-7, 的-8)
dobj (抹黑-6, 民众-9)
```

# 識別文章之議題特徵與情緒

## ✓ 依賴關係選取

- 採用四種依賴關係挑選潛在的特徵值

依賴關係代碼	句型結構	範例	範例
nsubj	主詞 + 任意字串 + 動詞 or 形容詞	媒體為了自己的利益， <b>抹黑</b> 抗議的民眾	nsubj (抹黑, 媒體)
dobj	動詞 + 任意字串 + 受詞	媒體為了自己的利益， <b>抹黑</b> 抗議的 <b>民眾</b>	dobj (抹黑, 民眾)
amod	形容詞 + 名詞	而是我一個 <b>無自由</b> 言論及受壓迫的黨體系中	amod (言論, 無自由)
nn	名詞作為形容詞 + 名詞	兩岸實質上仍存在敵對關係與 <b>安全</b> 風險	nn (風險, 安全)

# 識別文章之議題特徵與情緒

## ✓ 情緒指定\_變化性用語規則

<p>特徵值 \ 變化性用語</p>	<p>高、變高、新高、提高、攀高、高漲、上漲、飆升、飆漲、狂飆、攀升、調升、上升、多、遞增、增加、暴漲</p>	<p>少、遞減、減少、減低、下降、降低、低、倒退</p>
<p>房價、房租、房市、租金、稅租、失業率、物價、關稅、成本、社會成本</p>	<p>情緒極性由正轉負</p>	<p>情緒極性由負轉正</p>
<p>工資、薪水、薪資、生活品質、就業機會、工作機會</p>	<p>維持正向</p>	<p>維持負向</p>

---

## 識別文章之議題特徵與情緒

### ✓ 情緒指定\_否定詞反轉

例句：一種不以武力侵佔，改以法律吞食的后殖民主義

不 (相關用語)	其他
不可、不行、不能、不得、不準、絕不、決不、不許、不必、不用、不需、不須、不會、不再、毫不、從不	無法、無須、無需、無庸、並非、從位、毫無、免、犯不著、絕非、別、勿、弗、沒、沒有、但、卻

---

## 彙總與觀察

- ✓ 觀察一：PPT討論版中以**批評（負向）**的言論居多
- ✓ 觀察二：網民於論壇中間接面向的討論度是較高的，且其中對於**程序面向**的情緒發表強烈的負面情緒

---

## 結論與未來展望

- ✓ 有步驟性的分析來自網路論壇的使用者自建內容
- ✓ 觀察結果說明**批評負面**的言論居多
- ✓ 未來可以朝向處理政府對外的資訊流，了解政府與網民兩方的立場是否一致

---

Thanks!

