

## 運用論壇文章作情緒文字分析系統

林嘉淦\*, 康富城

National Taipei University, Department of Electrical Engineering, Taiwan

E-mail: jiachuan@mail.ntpu.edu.tw\*, leo77705@gmail.com

### 摘要:

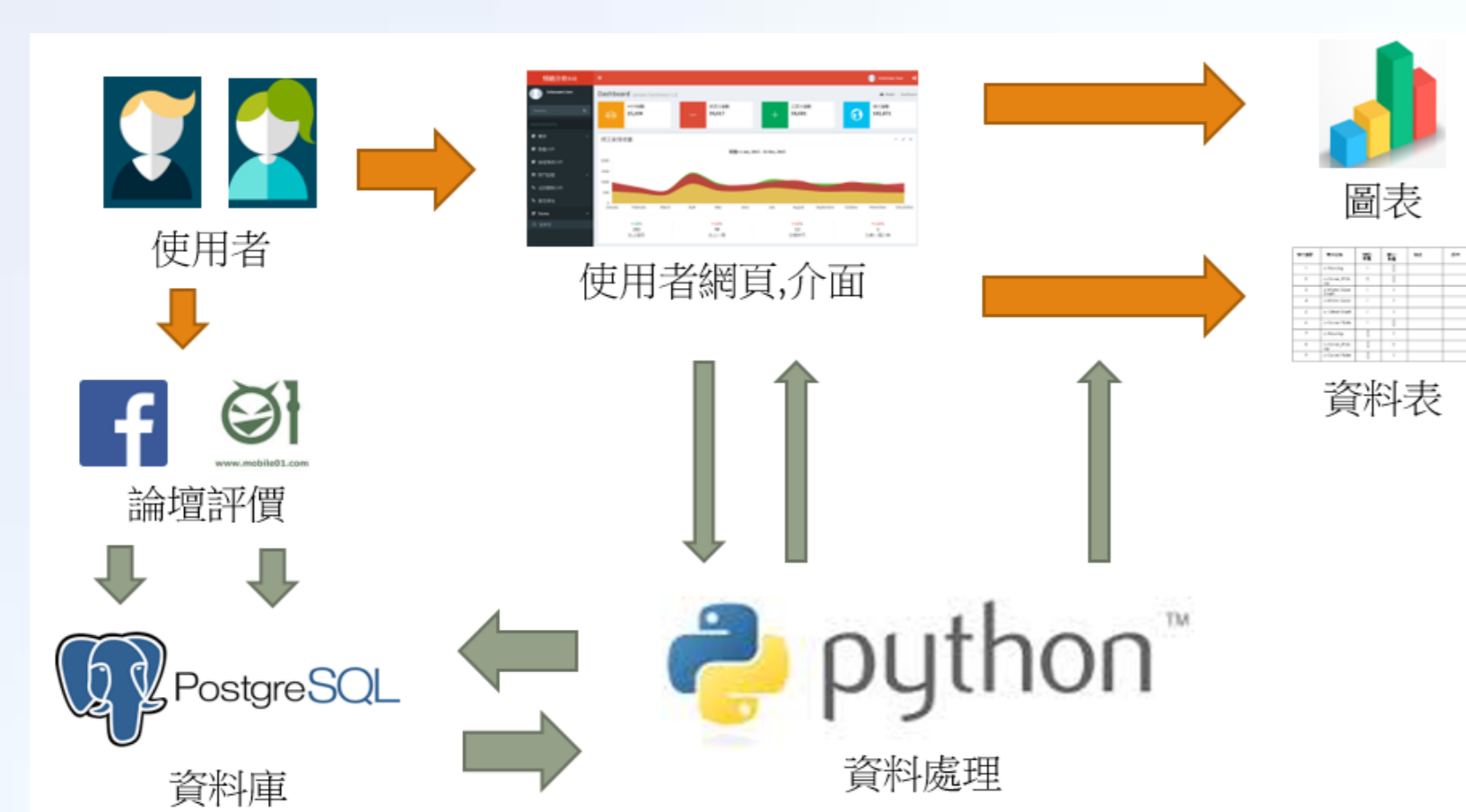
在較以前的時代裡，人們購買商品的評價，大多透過鄰居，人傳人傳達到整個社區，企業難以透過有效且便利的方式取得消費者對於該產品的評價，企業因此失去改善產品的機會。近年來，由於網路科技發達，使得消費者習慣更容易從各種不同管道抒發自己對於產品的看法，如撰寫在線上網站，或是透過填寫商品的回饋單，甚至是透過客服信箱進行客訴等等。於現今通訊發達，網路無所不在的環境中，消費者對於商品的資訊可以說是隨手可得。其中，在消費者的眾多評論中，有大多數的消費者會從中留下各種情緒句子，像是喜歡、討厭等形容詞。在上述的情形中，每一個句子都有隱含正面、中立、負面評價的情緒字眼。而企業為了獲取更多的商機，則想辦法收集分析這些消費者的評價後，做出商品的改良與推廣。基於我去年跟同學一起發表的”建構情緒分析系統-以車電產業為例”後，今年將整個系統變更環境做出改良，呈現更有效率、更美觀的介面。結合文字探勘與輿情分析，對分析的結果做出更精準的判斷，另外建構一個實質網站，方便使用者簡單閱讀判斷。並且利用定期更新資料的方式隨時提供當下情資。

### 關鍵字:

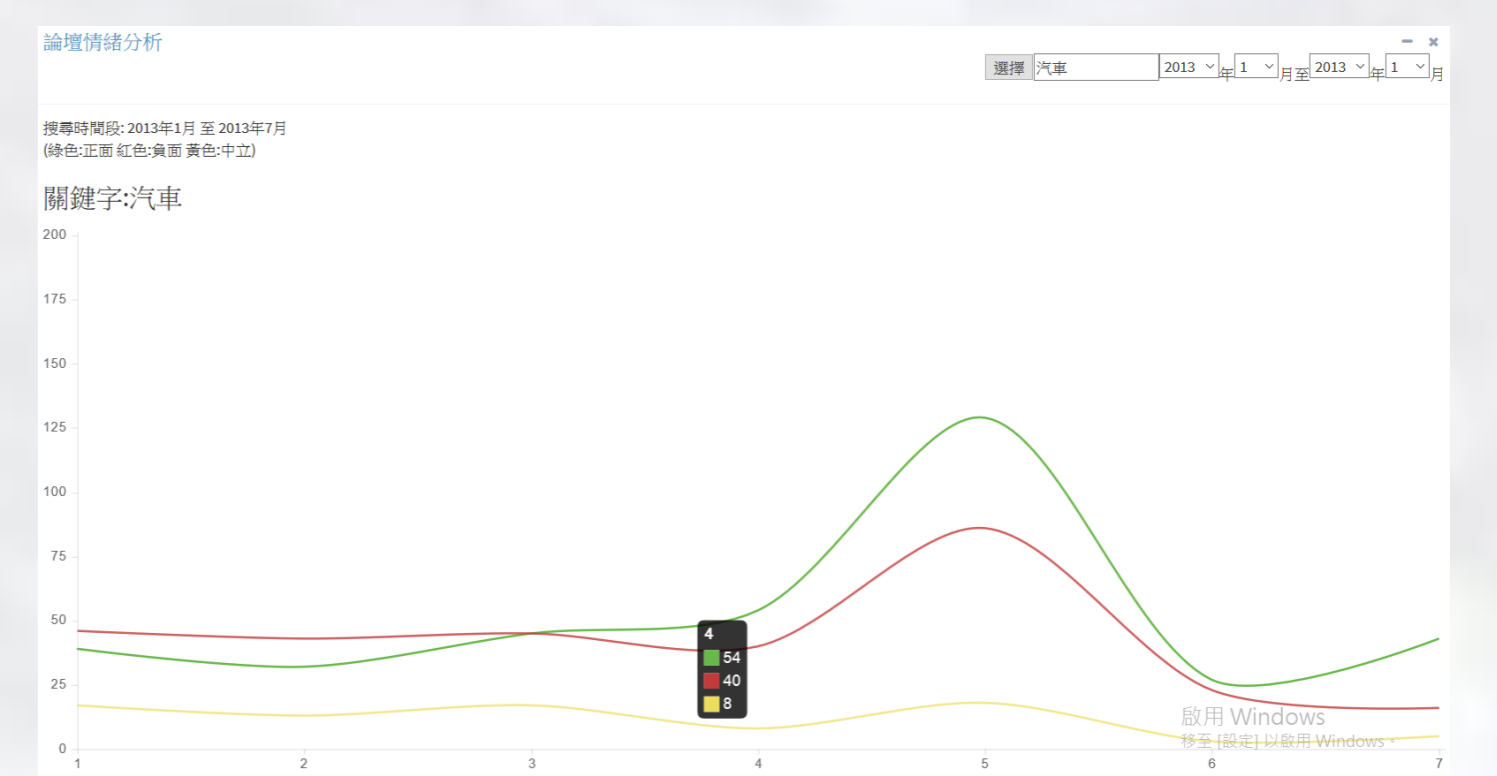
評價、情緒字眼、文字探勘、輿情分析、網路輿情

### 本文:

消費者將自身使用過商品的經驗寫在臉書、推特、論壇等。透過搜尋引擎搜尋關鍵字，並且找出網頁連結將文章、評論等一切重要資訊儲存在資料倉儲。透過程式幫忙將句子切開並與取出情緒字詞和時間綁定，把這些句子拆成字或詞的來比對。其中有用到python套件word2vec與jieba去做分詞與相關度判斷。並且將他們把時間綁定，再來拿這些資料做統計和分析，最後儲存在資料庫當中，製作一個網頁介面，給予使用者相關的報告和回饋來反映目前情況。



圖一、操作程序之流程圖



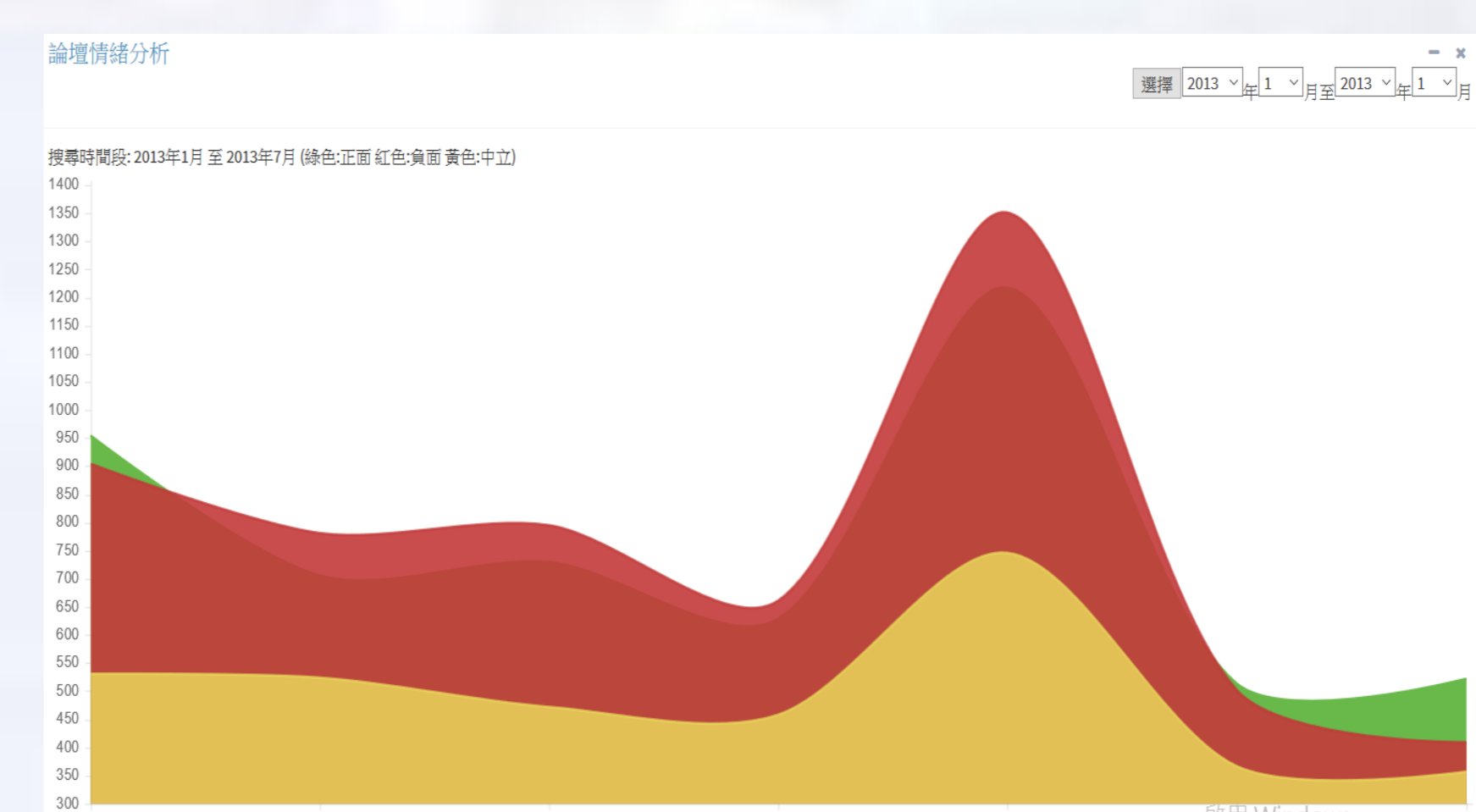
圖四

2015年各月份客訴量統整(整年每月份客訴量曲線圖)(圖四)

圖五

2015年關於“汽車”一詞分析(論壇情緒分析折線圖)(圖五)

X軸為月份，Y軸為數量。左圖為分析在2015一整年某家廠商蒐集的客訴量分析。右圖則是分析在論壇裡有關“汽車”一詞在2015年1月~7月有關的正面、負面與中立文章量。



2015年論壇月份情緒分析曲線圖(圖六)

| 日期                        | 標題                 | 類別 | 發表人       | 發文數 | 回覆數 | 文章數   | 正面情緒字數 | 負面情緒字數 | 中立情緒字數 |
|---------------------------|--------------------|----|-----------|-----|-----|---|--------|--------|--------|
| Aug. 1, 2015, 5:21 p.m.   | 新Alaris行車電腦... (4) | 資訊 | khoo      | 400 | 2   | http://www.robotforum.com/.../Alaris-446668201  | 400    | 0      | 0      |
| Aug. 3, 2015, 4:17 a.m.   | 致富車... (2)         | 資訊 | john      | 400 | 0   | http://www.robotforum.com/.../致富車-446668202     | 400    | 0      | 0      |
| Aug. 12, 2015, 2:11 p.m.  | 行車電腦... (4)        | 資訊 | khoo      | 410 | 0   | http://www.robotforum.com/.../行車電腦-446668203    | 410    | 0      | 0      |
| Feb. 11, 2015, 5:39 p.m.  | 台灣... (3)          | 資訊 | khoo      | 114 | 0   | http://www.robotforum.com/.../台灣-446668204      | 114    | 0      | 0      |
| Aug. 30, 2015, 10:54 a.m. | 新Alaris... (2)     | 資訊 | john      | 358 | 0   | http://www.robotforum.com/.../新Alaris-446668205 | 358    | 0      | 0      |
| Feb. 9, 2015, 5:49 a.m.   | Line... (3)        | 資訊 | periphere | 380 | 0   | http://www.robotforum.com/.../Line-446668206    | 380    | 0      | 0      |
| Oct. 26, 2015, 2:28 a.m.  | 行車... (2)          | 資訊 | john      | 328 | 0   | http://www.robotforum.com/.../行車-446668207      | 328    | 0      | 0      |
| Nov. 24, 2015, 8:28 a.m.  | Alaris... (2)      | 資訊 | khoo      | 288 | 0   | http://www.robotforum.com/.../Alaris-446668208  | 288    | 0      | 0      |
| Nov. 21, 2015, 5:59 p.m.  | Line... (2)        | 資訊 | periphere | 288 | 0   | http://www.robotforum.com/.../Line-446668209    | 288    | 0      | 0      |
| Aug. 31, 2015, 1:29 a.m.  | 行車... (2)          | 資訊 | khoo      | 238 | 0   | http://www.robotforum.com/.../行車-446668210      | 238    | 0      | 0      |

圖七

2015年1月分關於Altis文章分析(文字雲+堆疊圖)(圖七)

圖八

2015年1月分關於U6文章分析(文字雲+堆疊圖)(圖八)

由左圖可以看到資料表中，可以看到依據近幾年比較熱門的討論文章，他們的討論聲量與情緒偏向判斷，給予其URL方便使用者可以去查看其文章討論細節；而右圖則是對留言者與作者做出分析，依照他們最多發文留言數作排序，分析其個人發表的正面、中立以及負面的文章，來做為訪問或參考的依據，幫助企業判斷。

### 結論:

這是我的實體網站連結:

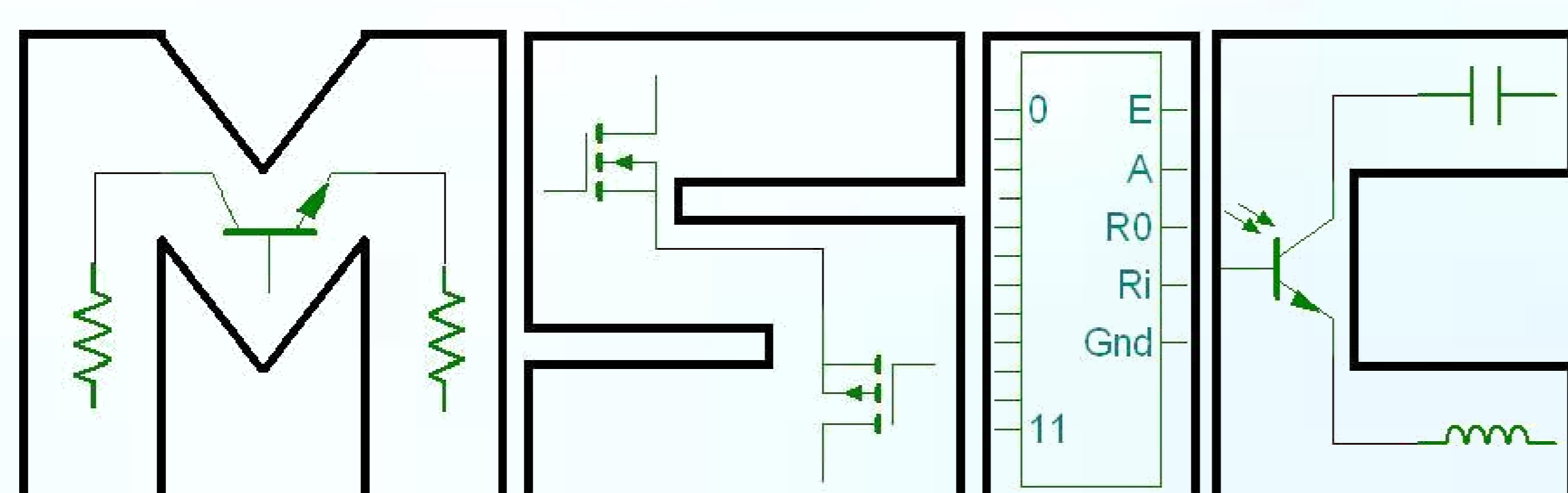
<http://caranalysis.cloudapp.net/>

從以上研究成果中，我比去年改進了系統上速度的優化，使用者可以選擇的功能更多元，前端網頁的設計與資料處理，以及學到了python與R不同的優點與缺點。

可以發現到這些方法的重點都是要嘗試從大量的文字中，辨識出正面與負面等情緒，好讓這些結果產生出商業價值，這也是口碑評價對於現在公司來說非常重要的一環，公司該在消費者評價這一部份的管理層面上多出一份力。

### 參考文獻:

- [1] Bing Liu 著，2012，“Sentiment Analysis and Opinion Mining”。
- [2] T Mikolov 等人著，2013，Distributed Representations of Words and Phrases and their Compositionality
- [3]吳柏翰學長的GITHUB:<https://github.com/rippleblue>
- [4]我本人的GITHUB:<https://github.com/kangfizz>



TEL:(02)8674-6502 FAX:(02)2673-6500

Address:151,University Rd San Shia, New Taipei City,237 Taiwan

新北市三峽區大學路151號人文大樓