|  |  |
| --- | --- |
| 專題進度週報 | |
| 日期 | **2012/04/05** |
| 專題主題 | **新聞推薦系統** |
| 組員 | **朱奕安、戴均民、郭書佑** |
| 本週進度  **新聞來源蒐集**  在新聞來源蒐集的部分，我們決定採用RSS來幫助蒐集，各大新聞網站都有提供RSS的方式來取得新的資料，RSS是一個統一的格式，所以可以方便寫程式進行新聞蒐集。  RSS範例：[Google 重點新聞RSS](http://news.google.com.tw/news?pz=1&cf=all&ned=tw&hl=zh-TW&output=rss) (截圖如下)    但是通常RSS內通常只會提供一定數量的新文章(在上面的範例中，是10個最新的新聞)，如果沒有辦法一直抓取，就沒有辦法抓到所有的文章，於是，必須再利用「[Google閱讀器](http://www.google.com.tw/reader/view/)」進行輔助，然後再利用 Google API把新聞從Google閱讀器取出，Google API這部分仍須進行研究([說明](http://code.google.com/p/r2-release/wiki/GoogleReaderApi))。  目前我們這組已經在Google閱讀器中，放入了「Google新聞」、「YAHOO!奇摩新聞」、「UDN聯合新聞網」、「中國時報」等網站共156個RSS的來源。  **對新聞進行分類**  目前想到的方法，就是對標題先進行斷詞之後，在丟進搜尋引擎，以便確定其分類，但是仍然需要先訂出所有分類，才有辦法進行分類。  提供斷詞服務之網站：「[中研院中文斷詞系統](http://ckipsvr.iis.sinica.edu.tw/)」(測試截圖如下) | |
| 檢討與改進 | |