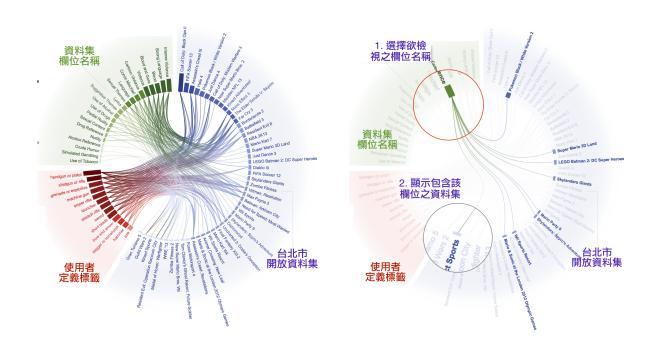
「大數據」或「開放資料」等是近年來熱門的資料議題之一,這顯示目前正進入一個能夠 大量取得資料的時代。在擁有大量且多元的資料之後,若要對所關心的主題進行分析時, 會面臨「如何挑選資料集?」的問題,並且從實際案例中發現[1-2],透過跨資料集分析 能夠找出有趣且關鍵的結果。

此外, 傳統標籤 (tag) 的檢索功能並不完善, 例如: 「地址」標籤未能列出所有相關的資料集。因此, 我們希望改善現有的功能, 包括

- 1. 標籤的生成由資料集內的欄位決定, 而非資料集名稱或手動輸入。
- 2. 標籤能集中具相同意義的欄位, 如「地理」可以包含「地址」與「經緯度」。
- 3. 利用欄位名稱及使用者定義標籤. 分析資料集間是否具有相關性。
- 4. 視覺化呈現資料屬性、標籤及資料集等三方的關係網絡。

我們提出一個跨資料集的檢索系統(如圖一),預計從臺北市政府資料開放平台中取出部分資料集進行展示。系統畫面分成欄位名稱(綠)、使用者定義標籤(紅)及開放資料集(藍)等三個部份。在系統初始畫面中會顯示關係網絡的全貌,使用者可以透過不同的角度進行檢索,來得知(1)資料集包含的欄位名稱及標籤,(2)欄位名稱連結到的資料集及標籤,與(3)使用者定義標籤概念包含的資料集及欄位名稱。



圖一、系統畫面的示意圖:左圖為系統初始畫面,右圖為檢索中的顯示畫面(圖片改編自 http://goo.gl/lePHSC)。

參考資料:

[1] Big Data: A Revolution That Will Transform How We Live, Work, and Think, http://www.books.com.tw/products/0010587258.

[2] The Moment of Clarity: Using the Human Sciences to Solve Your Tougest Business Problems, http://www.books.com.tw/products/0010640512.