

## 鏈結開放資料為基礎之Web API服務--以觀光旅遊資訊為例

陳良駒 張有志 余丁榮

國防大學管理學院資訊管理學系

{ nctuhorse, Percy7, cecyu.tw }@gmail.com

### 摘要

鏈結開放資料的發展涵蓋語意技術及訊息結構化的特性，目的是建置一個相互連結且電腦能理解的資料網路，開放式的鏈結資料目前已應用於生命科學、地理、社群網絡、媒體、醫藥等各領域。此外，網際各項Web API的應用也蓬勃發展，透過混搭式的網頁應用程式來整合多樣的服務，以達到客製化的要求。本研究以台灣交通部觀光局網站中的各國來台旅客資料、世界銀行網站的各國人口資料、以及DBpedia的各項補充資料為不同來源的資料集，並以Google Fusion Tables為Web API的應用工具，將各項運用服務以視覺化的方式來呈現，提供關於來台旅客的相關知識和資訊，讓資訊傳遞更為鮮明及容易理解。

**關鍵詞：**語意網、鏈結開放資料、Google Fusion Tables、DBpedia、Mashup。

### Abstract

The development of Linked Data covers the characteristics of semantic technologies and structured knowledge and the purpose of Linked Data is to build an interlinking network that data can be interpreted. The Linked Open Data (LOD) has been used in a variety of knowledge domains including life sciences, geography, social networks, media, and medicine.

In addition, the applications of Web API are growing rapidly and they can integrate a variety of services with mash-up web design. In this study, we design a Web API application through Google Fusion Tables tool and embedded three different data sources: the tourist data from Taiwan Tourism Bureau Website, populations from World Bank, capital and other supplements information from DBpedia dataset into the application. The users can retrieve related information by visualized presentation and comprehend the information more conveniently.

**Keywords:** Semantic Web, Linked Open Data, Google Fusion Tables, DBpedia, Mashup

### 1. 前言

目前網路上的資料種類已呈現巨量的成長。為了滿足在網際網路裡快速連結到相關主題的資訊，網路的發展也進入到Web 3.0 世代[1-3]。主要目標在使機器能理解網路資源內容，做到自動化處理，進入語意網的時代，透過資料的結構化，再將

這些相關的結構性資料串連在一起，即能相當有效的提供查詢結果回饋給使用者。網際搜尋引擎(如Google、Yahoo、Bing等)雖提供快速的資料查詢功能及服務，然而，查詢的過程卻常常要耗費龐大的人力和時間；使用者不僅要有一定的使用經驗或技巧，而且要有相當的耐心，方可在既有的網路服務系統上找到所需要的內容。

為了解決上述的問題，資料格式的標準化及開放資料(Open Data)[4][5]是目前的主要課題。W3C也公布一些公開資料的標準格式，像CSV、JSON、RDF、OWL等等，目前國際上，諸多國家(如美國<sup>1</sup>、英國<sup>2</sup>及台灣<sup>3</sup>)已經實行政府開放資料政策，並積極鼓勵企業組織、人民團體、與公民個人來對政府開放資料作創新加值運用，以期增進公共價值。

鏈結資料(Linked Data)是一套將資料發布於網路的原則，運用此套原則發布資料將有助於資料集、詮釋資料(Metadata)、以及控制詞彙的彼此串聯。鏈結開放資料(Linked Open Data, LOD)即是透過開放資料與鏈結資料的結合效應，將有助於資料的再用性(Reuse)、降低冗餘(Redundancy)，並可透過資料互聯的加乘效果，展現資料的價值。LOD是實現語意網(Semantic Web)的具體機制。

LOD已經廣泛應用在許多領域，包含圖書館界在內，例如美國國會圖書館、大英圖書館、瑞典國家圖書館等都已採用LOD的概念發布資料。此外，也應用於生命科學、地理、社群網絡、媒體、醫藥等各領域。可見透過鏈結開放資料的重新組合可以創造出新的知識與對事物的理解，而這一切可能帶來在全新的領域上的運用。此外，網際各項Web API的應用也蓬勃發展，透過混搭式(Mashup)的網頁應用程式來整合多樣的服務，以達到客製化的要求。雖有部分學者[6][7]認為Web API與LOD的整合運用將會逐漸成長，然而相關的探討卻並不多見。

本研究以台灣交通部觀光局網站中的各國來台旅客資料、世界銀行網站的各國人口資料、及DBpedia為不同來源的開放資料集，並以Google Fusion Tables(GFT)為Web API的應用工具，將各項運用服務以視覺化的方式來呈現，讓資訊傳遞更為鮮明及容易理解。

<sup>1</sup> <http://data.gov/>

<sup>2</sup> <http://data.gov.uk/>

<sup>3</sup> <http://data.gov.tw>

## 2. 文獻探討

### 2.1 鏈結開放資料

鏈結開放資料[8]即是鏈結資料技術與開放資料概念的組成。

根據開放知識基金會(The Open Knowledge Foundation: OKF)針對 Open data 所做的定義,說明「Open data」係指的是一種經過挑選與許可的資料,這些資料不受著作權、專利權,以及其他管理機制所限制,可以開放給社會公眾,任何人都可以自由出版使用,不論是要拿來出版或是做其他的運用都不加以限制。並要符合 11 項原則:(1)自由存取;(2)可再散布;(3)可再利用;(4)無技術限制;(5)署名;(6)完整性;(7)對任何個人或團體無差別對待;(8)對任何領域的應用無差別對待;(9)散布授權條款;(10)授權條款不得專屬於特定產品;(11)授權條款不得限制散布其他作品。

在傳統 Web 架構下使用者透過頁面鏈結(e.g. href)的方式進行資訊瀏覽及查詢工作,而 Linked Data 則強調每一個資料物件的識別及連結對應關係[9]。Linked Data 目的是建置一個機器能理解、能解讀的富含語意和具有結構化的資料網路。Linked Data 是一個正式的 W3C 項目,其讓資料開放並連接在一起,這樣人們就可以以其為基礎來建立有趣的應用。Tim Berners-Lee [10]於 2006 年提出 Linked Data 的四個基本原則,闡述如下:

- (1).以 URI 為任何事物的標記名稱。
- (2).使用 HTTP URI,讓任何人都可以找到這些標記名稱。
- (3).當某個使用者尋找某個 URI 時,以規範的標準(RDF、SPARQL)提供有用的資訊。
- (4).在提供使用者的資訊中,盡可能給予相關的 URI,以利使用者可以發現更多事物。

而 Passant[11]認為 Tim Berners-Lee 提出的這四個原則,從編程(Programmatic)的角度來看是非常有意義,因而進一步提出有關 LOD 資料集定義的概念式來表達:

Definition: Linked Open Data (LOD)

$LOD = \cup_i G_i$ ; 其中  $G = (R; L; I)$ ,

$R = \{r_1, r_2, \dots, r_n\}$  表示 URI 的資源(Resources)集合,  
 $L = \{l_1, l_2, \dots, l_n\}$  表示 URI 的鏈結狀態(Typed links)集合,

$I = \{i_1, i_2, \dots, i_n\}$  表示資源之間鏈結實例(Instances)的集合,亦即  $i_i = \langle l_j, r_a, r_b \rangle$ 。

資料來源:[11]

Linked Data 發展中會涉及到許多主要的標準和協定,包括 HTTP、URI、RDF、SPARQL 等[12]。簡單來說,Linked Data 就是使用 Http URI 標示任一物件或概念,讓使用者可以透過這些 URI 來查詢資料,並透過 RDF、SPARQL 等標準來提供資料,

在顯示的資料中會包含到其他 URI 的連結,使用者便可以依照這些 URI 找到更多資料。

自 2007 年底開始至 2011 年 9 月為止,LOD 已收錄 295 個資料集,310 億個 RDF 三元組(Triples),以及 5.04 億條 RDF 鏈結(Links) [13]。其中包含了很多著名的資料集,例如:DBpedia、DBLP Bibliography、GeoNames、FOAF、MusicBrainz 等,這些資料集涉及生命科學、地理、社群網絡、媒體、醫藥、出版社等不同領域。

LOD 的發展包括「鏈結型態(Link Types)」及「應用領域(Application Domain)」兩大面向。其中「鏈結型態」係以具語意特性的資料為對象,例如:人物(Person-related)、時間(Temporal-related)、空間(Spatial-related)等[14];而「應用領域」則以不同產業之特性為分類基準,例如:媒體(Media)、政府(Government)、出版物(Publication)、使用者自建內容(User-generated content)等[15]。相關資料雲圖如圖 1 所示。

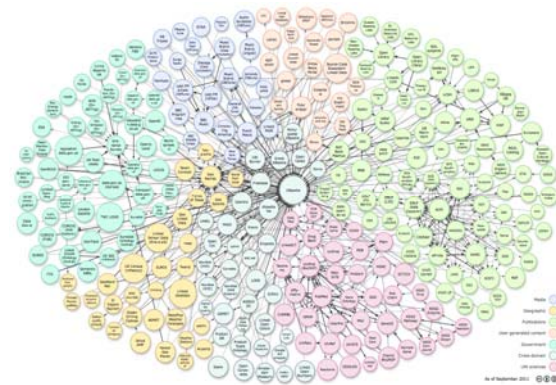


圖 1、LOD 資料雲圖 (資料來源:[15])

目前 Linked Data 主要的核心為 DBpedia。DBpedia [16]係指以結構化的形式萃取維基百科(Wikipedia)中的重要資訊,以強化維基百科的搜尋功能。目前 DBpedia 是世界上最大的多領域知識本體之一,也是 Linked Data 的一部分。由圖 1 可知,DBpedia 因主題內容涵蓋的廣泛度,與其他各類大型資料集均有通用性的連結,故 DBpedia 亦被稱為 Interlinking Hub。

目前英文版的 DBpedia 3.8 版[17]的資料集擁有超過 377 萬的事件(Things),包括 76 萬 4000 人、57 萬 3000 個地點、33 萬 3000 張唱片、7 萬 2 千部電影、1 萬 8000 種電腦遊戲、19 萬 2000 個組織、20 萬 2000 種物種和 5500 個疾病。DBpedia 的資料不僅被 BBC、Thomson Reuters、New York Times 所採用,也是 Google、Yahoo 等搜尋引擎檢索的對象。

### 2.2 網頁應用程式介面

網頁應用程式介面(Web API)是透過 Web 通訊協定及資料格式的開放式標準(例如 HTTP、XML 及 SOAP 等)去實作 Web Service 的一種軟體

服務[18][19]。

透過 Web API 可將不同來源的網路服務加以混搭，以滿足各種不同的需求 [20]。如 Google、Microsoft、Yahoo及Flickr等公司都有提供地圖、影音及新聞等網路服務，使用者可以很簡單的方式整合不同來源的網路服務，以創建新的網路應用程式或服務，如 Flickr 網路相簿和 Google Maps 結合的地圖服務。

目前網路上有眾多的 Web API，例如 Google、Yahoo 網路業者或各政府、公司行號等等都有提供各式各樣 Web API，根據 programmableweb 網站的統計(如圖 2)，截至 2013 年 9 月 11 日，全球計有 9940 個 API 及 7178 個 Mashup，且仍在快速成長中。



圖 2 programmableweb 網站的統計資料

一般來說，Web API 可提供 LOD 解決無法即時更新的資訊的困難，但這些 API 基本上僅提供 JSON 和 XML 格式的資料，並不是 RDF 格式，且 Web API 應用程式的資料來源往往都是單一的，所以 Web API 存在各有不同的存取機制、資料來源及資料格式的問題[21]，但是整合不同來源資料通常是網站開發的基本需求，透過近來興起的 Linked Data 原則及技術，鏈結多個不同資料來源提供更豐富的應用服務，似乎成為一個可行的解決方案[22]。

然而，目前已有許多研究開始探討 Linked Data 技術結合 Web API 開發應用服務系統的應用。例如：「Books@HPCLab」透過 Linked Data 鏈結不同網路書店資訊提供混搭式個人化服務的[23]；「DBpedia Mobile」結合 DBpedia 及 Google Maps，並鏈結 GeoNames, Revyu, EuroStat and Flickr 多個資料集，透過 GPS 定位，提供行動裝置所在位置附近的背景資訊[24]；「MORE」電影推薦系統，以 DBpedia 為中心，鏈結 Internet Movie Database (IMDB); LinkedMDB, 再參照使用者 Facebook 的個人資料來推薦電影[25]。各項研究成果均具體呈現出混搭式網路服務之效益。

本研究撰寫使用者網頁介面，結合 Google Fusion Tables API，透過使用者選擇過濾條件，連結多個不同來源的資料，包含以 SPARQL 語法查詢 DBpedia 資料集所獲取的資料，呈現更多元的相關

資訊，除了讓 API 與開放資料產生連結外，更能透過 Linked Open Data，提供更具豐富的資訊。

### 3. 研究架構及方法

#### 3.1 研究架構

本研究系統架構圖，分為資料源、資料儲存與轉換、資料呈現三大部分(如圖 3)。研究資料來源分別為：(1)台灣交通部觀光局；(2)世界銀行；(3)Wikipedia；(4)DBpedia。研究資料經彙整後存放於 MySQL 資料庫，再自動匯出 CSV 檔輸出到 Google Fusion Tables。撰寫使用者介面，依據需求進行簡易的條件查詢及選擇過濾條件，再依過濾條件讀取 Google Fusion Tables 資料，最後於網頁上透過統計地圖作視覺化呈現，提供相關資訊。

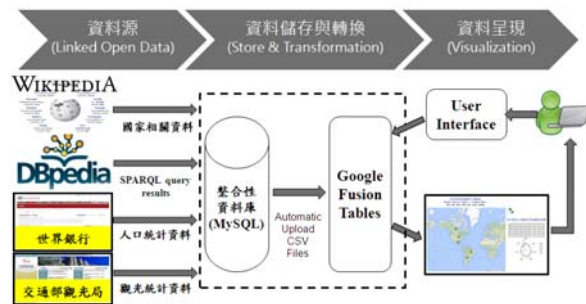


圖 3 系統架構圖

#### 3.2 Google Fusion Tables

過去要製作專業的統計地圖都需要仰賴專業的 GIS 軟體，不管是商業或免費的 GIS 軟體，除了操作上的門檻之外，如何取得繪圖所需的統計單元也是一個大困難。

現在透過 Google 推出新的服務稱為 Fusion Tables[26]，讓一般使用者可以很簡單地製作出專業地統計地圖，統計資料都可以利用統計地圖來做視覺化的呈現。Google 的 Fusion Tables 屬於一個實驗性的應用程序，讓資料可以儲存、共享、查詢，並可透過視覺化呈現。Fusion Tables提供 REST API 來管理資料表、資訊視窗樣板和樣式，並且透過查詢管理介面，可以執行資料的插入/更新/刪除等動作，也可以依照查詢條件進行資料的篩選，查詢結果可以以 CSV 或 JSON 格式輸出，或者將資料提供給 Google Maps API or Google Chart Tools 使用。

台灣交通部觀光局的所開放或統計每年來台旅遊觀光人數，均以 CSV 表格呈現，一般民眾難以從中獲取資訊，本研究想以世界地圖的樣貌做不同的資訊呈現，希望社會大眾可以從政府所開放的資料中，獲取更多的資訊或知識，礙於製作統計地圖的 GIS 軟體在使用或獲取上有一定的門檻，所以本研究選用目前屬免費且易於取得的 Google Fusion Tables，以便於結合 Google Maps API 將資料做視覺化的呈現。

### 3.3 SPARQL

SPARQL[27]簡單協定與資源描述框架查詢語言(Simple Protocol and RDF Query Language)，主要是用來查詢 RDF 的語言，語法跟 SQL 相似，SPARQL 主要有四種不同類型的查詢語法：

- 1.SELECT 標準查詢。以標準 SPARQL XML 結果格式呈現查詢結果。
- 2.ASK 查詢結果是 yes/no，沒有具體內容。
- 3.DESCRIBE 查詢結果以三元組來描述特定資源。
- 4.CONSTRUCT 根據查詢結果生成 RDF。

本研究即透過 SPARQL 語法萃取出 DBpedia 中與國家相關的補充資訊。

### 4. 研究成果

本研究以觀光旅遊資訊為範例，整合自不同來源資料(1)交通部觀光局：2008-2012 年各年度來台旅遊人數統計；(2)世界銀行：截至 2012 年世界各國最新人口數統計資料；(3)DBpedia：世界各國首都及 URI 連結；(4)Wikipedia：世界各國國家簡介及旗幟，應用服務系統以網頁的形式呈現，設計的功能及服務透過 URI 發布，提供免費使用，藉以達到有效的資訊散佈，以及強化系統的使用率，以下將概述系統各功能的畫面操作。

- (1).將各資料來源所獲取的資料彙整到 MySQL 資料庫後，經由 JavaScript 所撰寫程式自動轉換出 CSV 檔如圖 4，再藉由系統工具匯入 Google 的 Fusion Tables(如圖 5)。
- (2).透由 Google Fusion Tables API 所提供之功能，結合 Google MAPs 製作統計地圖如圖 6，將資料做視覺化呈現。
- (3).使用 JavaScript 撰寫使用介面，該介面提供數種過濾條件給使用者查詢。使用者可依據需求設定查詢條件，系統將提供各年度來台人數的統計數據及雷達分析圖表(如圖 7)。
- (4).使用者若選取地圖標記，將於 Google Fusion Tables 的資訊視窗(Infowindow)中提供進階的資訊如圖 8，包含國家介紹、人口、首都、來台統計數據等資料及 DBpedia、Wikipedia 等，達到外部連結資料的目的。

本研究係以透過鏈結開放資料技術連結不同來源的開放資料，呈現更豐富的各年度世界各國來台人數相關資訊，以圖 8 為例，在網頁右側顯示各區域(亞洲、美洲、歐洲、大洋洲及非洲)或全球(世界各國及各大洲)的各年度來台人數統計表及趨勢分析圖，左側的 Google Maps 根據選項進行圖資更新，並可點選地圖上國家標誌，開啟資訊視窗，如圖 8 中點選紐西蘭，呈現有關該國家更豐富的資訊，包括來自 Wikipedia 的國旗旗幟及國家介紹、世界銀行的人口數資料、DBpedia 的首都及 URI、交通部觀光局的各年度來台人數統計表及趨勢分

析圖等。經由各項統計資料可得知，以觀光及求學為目的之紐西蘭來台人數呈現大幅成長。

序	國家名稱	人口	人均 GDP	人均 GDP 成長率	人均 GDP 單位	國家代碼	國家名稱	國家代碼	國家名稱	國家代碼
1	中華人民共和國	1,370,531,000	1,284,720	1,284,720	1,284,720	china	中華人民共和國	china	中華人民共和國	china
2	印度	1,198,681,000	1,000,151	1,000,151	1,000,151	india	印度	india	印度	india
3	美國	307,298,000	49,241	49,241	49,241	usa	美國	usa	美國	usa

圖 4 MySQL 彙整資料集

Country	Population	GDP	GDP per capita	GDP growth rate	Capital	Flag
中華人民共和國	1,370,531,000	1,284,720,000,000	937,866	7.6%	北京	
印度	1,198,681,000	1,000,151,000,000	834,376	7.6%	新德里	
美國	307,298,000	15,000,000,000,000	48,821	2.3%	華盛頓	

圖 5 匯入 GFT 統計資料

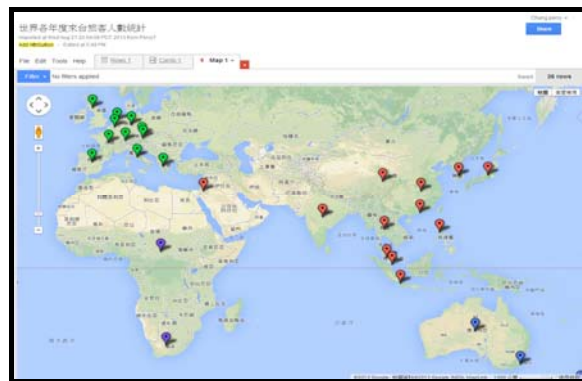


圖 6 由 GFT 製作統計地圖



圖 7 系統使用者介面



圖 8 GFT 資訊視窗

## 5. 結論

Linked data 雖然可以串聯各資料集，但要查詢開放資料需要了解 SPARQL 的基本語法及使用方法，Web API 雖可提供即時更新資料服務，但是卻無法整合不同來源資料集，而政府雖然開放資料，但都是一堆冷冰冰的原始數據，並無法讓社會大眾能輕易地從中獲取資訊或知識，本研究嘗試透過結合各項技術或工具的長處，將開放資料作連結及重新詮釋，並以視覺化的方式作呈現。

本論文係透過鏈結開放資料及 Google Fusion Tables 的開放性 API，鏈結 DBpedia、世界銀行及台灣交通部觀光局等開放資料，將資料以不同的面貌呈現，主要貢獻(1)透過 Linked Data 技術鏈結不同的資料來源，突破 Web API 單一資料來源的限制；(2) SPARQL 查詢 DBpedia 資料，已彙整至 MySQL 再輸出至 Google Fusion Tables，解決使用者學習 SPARQL 查詢語法的困擾及查詢限制；(3) 交通部觀光局各年度世界各國來台人數統計資料，透過趨勢分析圖，使用者可輕易分析或掌握背後資訊，解決面對枯燥且冷冰冰數據的解析困境。(4)於 Google Maps 上作資訊的視覺化呈現，讓使用者瞭解各國地理位置並簡單與資料融合在一起，解決地理空間概念或知識的不足。本研究除傳遞更豐富的知識外，更是 Linked Data 與 Web API 結合的最佳實踐。

## 誌謝

本論文感謝國科會專題計畫經費贊助(計畫編號：NSC 102-2410-H-606 -007)。

## 參考文獻

- [1] Berners-Lee T, "The World Wide Web: A very short personal history," World Wide Web Consortium[Online] Available: <http://www.w3.org/People/Berners-Lee/ShortHistory.html> .19 98.
- [2] Hassanzadeh H, Keyvanpour M, "A Machine Learning Based Analytical Framework for Semantic Annotation Requirements," arXiv preprint arXiv:11044950. 2011.
- [3] Aghaei S, Nematbakhsh MA, Farsani HK, "Evolution of the world wide web: From WEB 1.0 TO WEB 4.0," International

- Journal Web & Semantic Technology, V. 3, No. 1. 2012, pp. 1-10.
- [4] Open Definition Version 1.1, URL <http://opendefinition.org/okd/>
- [5] The Open Knowledge Foundation, URL <http://okfn.org/opendata/>
- [6] Héder M, Solt I, "DBpedia Mashups," Semantic Mashups, Springer, 2013, pp. 119-43.
- [7] Taheriyani M, Knoblock CA, Szekely P, Ambite JL, "Rapidly integrating services into the Linked Data cloud," The Semantic Web—ISWC 2012, Springer, 2012, pp. 559-74.
- [8] Yu L, "A Developer's Guide to the Semantic Web." Springer; 2011, pp 409-466
- [9] Fionda V, Gutierrez C, Pirró G, "Semantic navigation on the web of data: specification of routes, web fragments and actions," Book Semantic navigation on the web of data: specification of routes, web fragments and actions, Editor, ed.^eds., ACM, City, 2012, pp. 281-90.
- [10] Berners-Lee T, "Linked data-design issues (2006)," URL <http://www.w3.org/DesignIssues/LinkedData.html>. 2011.
- [11] Passant A, "Dbrec—music recommendations using DBpedia," The Semantic Web—ISWC 2010, Springer, 2010, pp. 209-24.
- [12] Hausenblas M., "Linked Data Specifications, Linked Data Research Centre", URL <http://linkeddata-specs.info/>. 2010.
- [13] W3C SweoIG TaskForces CommunityProjects LinkingOpenData, URL <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.
- [14] Hausenblas M, "Linked data applications," First Community Draft, DERI. 2009.
- [15] Cyganiak R, Jentzsch A, "Linking open data cloud diagram," LOD Community (<http://lod-cloud.net/>). 2011.
- [16] Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z, "Dbpedia: A nucleus for a web of open data," The semantic web, Springer, 2007, pp. 722-35.
- [17] DBpedia, <http://wiki.dbpedia.org/About>
- [18] Alonso G, Casati F, Kuno H, Machiraju V, "Web Services: Concepts, Architectures and Applications." Springer Publishing Company, Incorporated; 2010.
- [19] W3C Web Services Activity, URL <http://www.w3.org/2002/ws/Activity>
- [20] Benslimane D, Dustdar S, Sheth A, "Services mashups: The new generation of web applications," Internet Computing, IEEE, V. 12, No. 5. 2008, pp. 13-5.
- [21] Bizer C, "The emerging web of linked data," Intelligent Systems, IEEE, V. 24, No. 5. 2009, pp. 87-92.
- [22] Bianchini D, "A classification of web API selection solutions over the linked web," Book A classification of web API selection solutions over the linked web, Editor, ed.^eds., ACM, City, 2012, pp. 4.
- [23] Solomou GD, Kalou AK, Koutsomitropoulos DA, Papatheodorou TS, "A Mashup Personalization Service based on Semantic Web Rules and Linked Data," Book A Mashup Personalization Service based on Semantic Web Rules and Linked Data, Editor, ed.^eds., IEEE, City, 2011, pp. 89-96.
- [24] Becker C, Bizer C, "DBpedia Mobile: A Location-Enabled Linked Data Browser," LDOW, V. 369. 2008.
- [25] Mirizzi R, Di Noia T, Ragone A, Ostuni VC, Di Sciascio E, "Movie Recommendation with DBpedia," Book Movie Recommendation with DBpedia, Editor, ed.^eds., Citeseer, City, 2012, pp. 101-12.
- [26] Gonzalez H, Halevy AY, Jensen CS, Langen A, Madhavan J, Shapley R, et al., "Google fusion tables: web-centered data management and collaboration," Book Google fusion tables: web-centered data management and collaboration, Editor, ed.^eds., ACM, City, 2010, pp. 1061-6.
- [27] Prud'hommeaux E, Seaborne A, "SPARQL query language for RDF. W3C Recommendation (2008)," URL <http://www.w3.org/TR/rdf-sparql-query/>. 2008.