

## 以鏈結資料發展電影資訊系統之研究

### A Study on the Development of Linked Data-based Movie Information System

**鄭智遠 Chih-Yuan Cheng**

國立臺灣師範大學圖書資訊學研究所研究生

Master Student, Graduate Institute of Library and Information Studies,

National Taiwan Normal University

E-mail: boyman1025@gmail.com

**柯皓仁 Hao-Ren Ke**

國立臺灣師範大學圖書資訊學研究所教授兼圖書館館長

Professor & University Librarian, Graduate Institute of Library and

Information Studies, National Taiwan Normal University

E-mail: clavenke@ntnu.edu.tw

#### 【摘要 Abstract】

電影是人們普遍的休閒活動之一，電影資訊也是人們經常透過網際網路查找的資訊之一，然而網際網路上龐大的電影資訊卻也有著資訊分散、重複性和不一致的問題。鏈結資料是實踐語意網的最佳實務，其基於資料的機制有助於不同系統間的資料交換與整合。本研究運用內容管理系統 Drupal 建構一套融入鏈結資料精神，以 IMDb 的電影、導演和演員資料為基礎，整合 DBpedia、LinkedMDB、DBtune，以及大英圖書館等鏈結資料來源的電影資訊系統，使電影資訊內容更加豐富且提升鏈結資料的閱讀性。最後並藉由系統使用性量表來評估該系統的使用性，了解使用者對本系統的接受度，評估結果顯示本系統的使用性是可接受的。

Movie watching is a common leisure activity, and thus people seek for information about movies from the Internet very often. However, for such a large data, there are problems derived like information scattering, duplication and inconsistency.

Linked Data is a best practice for Semantic Web, and the data-driven technique helps the exchange and integration of data among systems. This study creates a linked data-based movie information system by the content management system Drupal. This system exploits the information on movies, directors, and actors from IMDb, and integrates the information from several linked-data producers, including DBpedia, LinkedMDB, DBtune, and British Library. In this manner, the system provides rich movie information and enhance readability of linked data. Finally, this study evaluates this linked data-based movie information system by the System Usability Scale (SUS) to understand user usability and acceptance of the system, and the SUS evaluations show that this system's usability is acceptable.

### 【關鍵詞 Keywords】

語意網、鏈結資料、RDF、SPARQL、使用性量表

Semantic Web; Linked Data; RDF; SPARQL; System Usability Scale

## 一、前言

網際網路和全球資訊網發展迄今已成為人們獲取資訊的重要管道，網際網路和全球資訊網上各種系統的整合、資料的交換則成為一項重要的議題，亦因而促成了語意網（Semantic Web）的發展。Tim Berners-Lee（Berners-Lee, 2000）指出，所謂的語意網即是建構一個基礎建設，在這基礎建設下賦予資訊明確的定義，使得電腦能夠理解富含語意的文件和資料，從而促成電腦與人們的協同工作。要達到語意網的第一步，便是以電腦能夠理解的方式來描述任何具體或抽象、實體或虛擬的事物（thing），以及事物間的關聯性，而鏈結資料（Linked Data）便是實現語意網的最佳實務（Linked Data, n.d.; 柯皓仁、陳亞寧，2013）。

電影是門綜合藝術，也是人們普遍的休閒活動之一。無論是觀賞電影前後，人們經常在網際網路上查找電影資訊。現今提供電影資訊的網站眾多，然而該等網站（如IMDb、開眼電影網等）通常僅提供單一資訊，如想知道電影其他周邊相關的資訊（如其原著小說、原聲帶等），使用者通常必須額外利用搜尋引擎輸入關鍵字來獲取。此

外，各電影資訊網站大多都有獨立運作的資料庫、呈現各自的電影相關資訊，讓使用者自行查找其相關的資料，對於如此龐大資料來說，利用鏈結資料的概念描述，協助電影相關資料被發現、取得和使用具有令人期待的潛力。故本研究擬透過現存的電影資料，以鏈結資料技術實作一個電影資訊系統，促成不同電影資訊相關系統間的資料交換與整合。

## 二、文獻探討

本節就鏈結資料、RDF 與 SPARQL，以及系統使用性量表之相關文獻進行探討。

### （一）鏈結資料

鏈結資料（Linked Data）是來自於語意網概念下的資料分享的方式，它所提出的原則促進網際網路上來自不同地方的資源在同一框架下能被存取，這些資料或後設資料能夠由電腦辨識，再進一步的分享與再利用（Berners-Lee, 2006）。鏈結資料進一步解釋指得是將結構化的、電腦可讀的資料，透過發布技術並建立資料之間相互連接的一種實際方式，並且使用者擁有更多的資料操作權限。鏈結資料依循四個原則（柯皓仁、陳亞寧，2013；Berners-Lee, 2006; Heath & Bizer, 2011），進一步說明如下。

1. 以統一資源標識符（Uniform Resource Identifier, URI）為任何具體或抽象、實體或虛擬的事物命名。
2. 使用 HTTP URI 來查找特定事物：以 HTTP 做為客戶端和伺服器端之間查詢及傳送 URI 的機制，使人或電腦可以查詢特定 URI 所代表事物的相關資訊。
3. 伺服器端使用 RDF 與 SPARQL 等標準，提供特定事物的相關資訊：當伺服器端接獲客戶端運用 HTTP URI 或 SPARQL 對於特定事物的請求時，伺服器端會以 RDF 標準格式將其相關資訊傳回給客戶端。
4. 對於特定 URI 所代表事物的相關資訊中，應包含其他相關事物的 URI，使得事物間得以串連，提供使用者多樣化資訊。

根據最新資訊，鏈結開放資料雲（Linked Open Data Cloud）在 2014 年 4 月已有涵蓋政府、出版、生命科學、使用者產生內容、跨領域、媒體、地理、社會網絡等領域共 1,014 個鏈結資料集（<http://lod-cloud.net/>）。

鏈結資料的相關應用已逐漸普及，其中 DBpedia Mobile (Becker & Bizer, 2008) 乃是一個行動環境中鏈結資料的應用，DBpedia 萃取自維基百科，將其轉為鏈結資料型態，並結合 GeoNames、CIA Factbook、EuroStat 和 US Census 等資料集提供豐富的資訊。DBpedia Mobile 充分使用 DBpedia 的地理資料，藉由 GPS 的技術自動探索使用者所在位置附近的相關內容，並藉由鏈結資料瀏覽器提供視覺化的地圖介面。鏈結資料亦曾運用於複雜的生態資料管理，將森林火災、植物標本、昆蟲採集、森林動態樣區資料和臺灣物種名錄等五個資料庫透過鏈結資料串連，提供生態資料一個新的分享方法 (林朝欽、麥館碩、夏禹九，2011)。至於在圖書館界，則有虛擬國際權威檔 (Virtual International Authority File, VIAF)、美國國會圖書館鏈結資料服務 (LC Linked Data Service)、大英圖書館的自由化資訊服務 (British Library Free Data Services)、OCLC 的 WorldCat、歐盟的 Europeana 等鏈結資料服務 (柯皓仁、陳亞寧，2013)。

## (二) RDF 與 SPARQL 概述

鏈結資料所遵循的格式標準為資源描述框架 (Resource Description Framework, RDF) (W3C, 2015)。RDF 乃是一種以圖論 (Graph) 模型 (由相連的節點與邊所組成) 為基礎的資料結構呈現方式，將資料編碼組成「三元組」(Triples) 句型結構，由主語 (Subject) (類似於主詞)、述語 (Predicate) (類似於動詞) 和賓語 (Object) (類似於受詞) 所組成，一個三元組又可稱之為一個「陳述句」(Statement)，用以描述事物的屬性和事物間的關聯性。通常 RDF 在鏈結資料的架構中，以被序列化 (Serialize) 成其他格式的資料 (如 RDF/XML、RDFa、RDF/N3 和 RDF/N-Triples 等)。

W3C 針對 RDF 格式設計了 OWL-QL、RQL、RDQL 以及 SPARQL 等查詢語言，其中 SPARQL 是目前最廣為使用的 RDF 查詢語言。SPARQL (全名為 SPARQL Protocol and RDF Query Language) 以類似 SQL 的語法提供使用者查詢 RDF 資料 (W3C, 2013)。RDF 儲存系統大多提供有 SPARQL 查詢端點 (endpoint)，藉此讓使用者以 SPARQL 語言對 RDF 儲存系統進行資料查詢。

## (三) 系統使用性量表

使用性 (usability) 的評估方法有很多種，每種都有其適用之目的、限制。使用性評估的主要目的在於找出系統之問題，並衡量系統使用的情況，以了解使用者真正操作系統的經驗與感受。本研究採用 Brooke (1986) 的系統使用性量表 (System

Usability Scale, SUS) 作為系統評估的方式, 該量表針對網站使用時較常遇到的主觀感受, 提出了 10 個問題計分的五等第量表, 計算後的分數結果稱之為「SUS 分數」, 此分數能評估一個網站使用性的好壞。SUS 分數代表使用者對系統使用性的綜合評量, 可作為不同網站或系統之間使用性的比較, 但每題題目所表示的分數並不具有意義, 不適合被單獨比較或作為其他用途。「SUS 分數」計算的方式是將題目分為兩組, 一組為奇數組 1、3、5、7、9 題, 另一組為偶數組 2、4、6、8、10 題, 將奇數組的各題原始分數減去 1, 可得到各題對應的分數, 再用 5 減去偶數題各題的原始分數, 便可得到各題對應的分數, 最後計算總分, 將奇數組和偶數組兩組對應的分數相加, 相加結果乘以 2.5 後即為 SUS 分數。

Bangor、Kortum 與 Miller (2009) 對於計算出來的 SUS 分數予以分級, 提出了可接受度範圍 (Acceptability Ranges)、學校等級尺度 (School Grading Scale) 和感官形容評量 (Adjective Ratings) 等三種不同的評量標準, 並彙整出一份 SUS 分數對應量尺, 在該研究中指出 70 分是最低可接受的標準。Sauro (2013) 則提出 500 個產品的平均 SUS 分數為 68 分, 且能將 SUS 應用於 5 位測試者的小樣本。

### 三、研究方法

本研究採用研究與發展法 (Research & Development) 來開發以鏈結資料為基礎的電影資訊系統。本節說明該系統所採用的資料集、開發環境和平臺、系統流程與架構。

#### (一) 資料集

本研究所採用的資料集包含 IMDb、DBpedia、LinkedMDB、DBtune。

##### 1. 電影資料集 IMDb (<http://www.imdb.com/>)

本系統使用 IMDb (Internet Movie Database) 為電影名單來源, IMDb 是一個針對電影、電影演員、電視節目、電視藝人、電玩遊戲和電影製作小組的線上資料集, 包含過去、現在上映的全面性電影和電視資訊, 除了有豐富的作品資訊, 包括影片演員、導演、劇情、影評等基本資訊外, 更進一步提供相關的影音資料, 如預告片、宣傳影片等。近年來 IMDb 為因應越來越龐大的電影資訊內容以及網路應用的興起, 開始提供開放式資料集的下載, 進而使得許多

電影相關應用的出現。IMDb 資料集截至 2014 年 11 月提供約 313 萬的作品（包含電影、電視、電玩遊戲等作品）和 629 萬筆人物。

由於目前 IMDb 尚未提供鏈結資料格式或 SPARQL 查詢介面，因此在建構本系統前，需要對原始資料進行主題挑選和格式統一。本研究利用 IMDb 所釋出資料集當中的 movie list 檔案，清除非電影資料，例如電視劇、電視節目、電玩影音等，剩下名單皆為電影資料，以此做為系統之主要電影清單，再將電影名稱後方的上映年份和特殊標記符號清除，轉換結構化格式，以便後續將資料轉成鏈結資料格式進行連結配對。

## 2. DBpedia (<http://dbpedia.org/>)

本研究為了將電影資料延伸至本地資料庫的框架之外，並強調鏈結資料的重複使用性，因此採用 DBpedia 所開放的資料，以重複利用 Web 上已有的資料和已存在使用的知識本體 (ontology)。

DBpedia 基於鏈結資料的基本格式要求，將 Wikipedia 資料萃取成結構化的 RDF 語法表示，並提供 SPARQL 語法查詢介面。採用 DBpedia 作為電影補充資料的優點，一方面符合鏈結資料標準格式和重複使用已有的資源；另一方面可透過 SPARQL 指令的複寫將資料顯示在本研究建構的系統上，不須將 IMDb 其他資料完整匯入本地資料庫，節省空間和處理時間。DBpedia 使用不同的知識本體來表示不同的主題資源，同一主題內也會因資料欄位的不同，而有不同的屬性 (property)。本系統以電影資料為主，因此找出 DBpedia 中電影資料所使用的知識本體，主要來自己在 Web 上活躍使用的 rdf、rdfs、foaf 和 DBpedia 自身的 dbpedia 知識本體。表 1 詳細列出 DBpedia 電影常見的屬性和所代表的含意。

表 1

*DBpedia 電影常見詞彙*

屬性	URI	含意
rdfs:label	<a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a>	資源的名稱資訊
dbpedia-owl:abstract	<a href="http://dbpedia.org/ontology/abstract">http://dbpedia.org/ontology/abstract</a>	電影的摘要
dbpedia-owl:director	<a href="http://dbpedia.org/ontology/director">http://dbpedia.org/ontology/director</a>	電影的導演
dbpedia-owl:producer	<a href="http://dbpedia.org/ontology/producer">http://dbpedia.org/ontology/producer</a>	電影的製片
dbpedia-owl:cinematography	<a href="http://dbpedia.org/ontology/cinematography">http://dbpedia.org/ontology/cinematography</a>	電影的攝影師
dbpedia-owl:country	<a href="http://dbpedia.org/ontology/country">http://dbpedia.org/ontology/country</a>	電影的發行國家發行
dbpedia-owl:distributor	<a href="http://dbpedia.org/ontology/distributor">http://dbpedia.org/ontology/distributor</a>	電影的製片商
dbpedia-owl:editing	<a href="http://dbpedia.org/ontology/editing">http://dbpedia.org/ontology/editing</a>	電影的剪輯
dbpedia-owl:gross	<a href="http://dbpedia.org/ontology/gross">http://dbpedia.org/ontology/gross</a>	電影的營收
dbpedia-owl:musicComposer	<a href="http://dbpedia.org/ontology/musicComposer">http://dbpedia.org/ontology/musicComposer</a>	電影的配樂作曲
dbpedia-owl:runtime	<a href="http://dbpedia.org/ontology/runtime">http://dbpedia.org/ontology/runtime</a>	電影的片長
dbpedia-owl:starring	<a href="http://dbpedia.org/ontology/starring">http://dbpedia.org/ontology/starring</a>	電影的演員
dbpedia-owl:thumbnail	<a href="http://dbpedia.org/ontology/thumbnail">http://dbpedia.org/ontology/thumbnail</a>	電影對應的圖片
dbpedia-owl:wikiPageID	<a href="http://dbpedia.org/ontology/wikiPageID">http://dbpedia.org/ontology/wikiPageID</a>	電影對應 Wiki 文章的編號
dbpedia-owl:writer	<a href="http://dbpedia.org/ontology/writer">http://dbpedia.org/ontology/writer</a>	電影的編劇
rdf:type	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a>	資源所屬類型
rdfs:comment	<a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a>	資源簡短摘要(500 字以內)
foaf:name	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>	對應 foaf 的資源名稱
foaf:primaryTopic	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>	對應 foaf 的主題來源

資料來源：<http://mappings.dbpedia.org/server/ontology/classes/>

### 3. LinkedMDB (<http://www.linkedmdb.org/>)

LinkedMDB 的資料來源包含了 IMDb、FreeBase 和 DBpedia 的電影資訊，將其整合統一之後轉換成鏈結資料格式，也支持使用 SPARQL 查詢其資料，雖然 LinkedMDB 存取三個內容豐富的資料集，所提供的電影資訊不算完整，但提

供了其他 DBpedia 未包含的電影資訊內容，使得 LinkedMDB 部分資料在本系統能補足其他電影資訊。LinkedMDB 除了建立自己的電影知識本體外，也使用了 rdf、rdfs、Dublin Core (DC) 和 foaf (Friend of a Friend) 等知識本體的詞彙，表 2 詳細列出 LinkedMDB 常見的屬性和所代表的含意。

表 2  
LinkedMDB 常見詞彙

屬性	URI	含意
rdfs:label	<a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a>	資源的名稱資訊
dc:title	<a href="http://purl.org/dc/terms/date">http://purl.org/dc/terms/date</a>	對應 DC 的資源名稱
dc:date	<a href="http://purl.org/dc/terms/date">http://purl.org/dc/terms/date</a>	資源的日期資訊
movie:actor	<a href="http://data.linkedmdb.org/resource/movie/actor">http://data.linkedmdb.org/resource/movie/actor</a>	電影的演員
movie:cinematographer	<a href="http://data.linkedmdb.org/resource/movie/cinematographer">http://data.linkedmdb.org/resource/movie/cinematographer</a>	電影的攝影師
movie:director	<a href="http://data.linkedmdb.org/resource/movie/director">http://data.linkedmdb.org/resource/movie/director</a>	電影的導演
movie:editor	<a href="http://data.linkedmdb.org/resource/movie/editor">http://data.linkedmdb.org/resource/movie/editor</a>	電影的編劇
movie:film_cut	<a href="http://data.linkedmdb.org/resource/movie/film_cut">http://data.linkedmdb.org/resource/movie/film_cut</a>	電影的剪輯
movie:film_of_distributor	<a href="http://data.linkedmdb.org/resource/movie/film_of_distributor">http://data.linkedmdb.org/resource/movie/film_of_distributor</a>	電影的製片商
movie:filmid	<a href="http://data.linkedmdb.org/resource/movie/filmid">http://data.linkedmdb.org/resource/movie/filmid</a>	對應電影資源的編號
movie:initial_release_data	<a href="http://data.linkedmdb.org/resource/movie/initial_release_data">http://data.linkedmdb.org/resource/movie/initial_release_data</a>	電影的上映資訊
movie:music_contributor	<a href="http://data.linkedmdb.org/resource/movie/music_contributor">http://data.linkedmdb.org/resource/movie/music_contributor</a>	電影的配樂作曲
foaf:made	<a href="http://xmlns.com/foaf/0.1/made">http://xmlns.com/foaf/0.1/made</a>	對應 foaf 電影的導演
foaf:page	<a href="http://xmlns.com/foaf/0.1/page">http://xmlns.com/foaf/0.1/page</a>	對應 foaf 的外部連結

資料來源：LinkedMDB

#### 4. DBtune (<http://dbtune.org/>)

除了電影本身資訊外，電影原聲帶也是讓人印象深刻的，本研究以 MusicBrainz (<http://musicbrainz.org/>) 做為電影相關音樂的資料來源。MusicBrainz 被稱為「開放的音樂百科全書」，所有 MusicBrainz 資訊都是開放性的，可以很容易地獲得這些資料。DBtune 將 MusicBrainz 透過 D2R 工具建構 SPARQL 端點，提供鏈結資料查詢，藉由 URI，可以作為電影資訊所釋出之



電影原聲帶音樂曲目做連結，使得此服務具有獨一無二的優勢。除此之外，DBtune 也採用為音樂主題設計的知識本體，是一個很好的電影補充資料來源。

表 3 詳細列出 DBtune 用於電影原聲帶資源的屬性和所代表的含意。

表 3

DBtune 常見詞彙

屬性	URI	含意
rdfs:label	<a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a>	資源的名稱資訊
dc:title	<a href="http://purl.org/dc/terms/date">http://purl.org/dc/terms/date</a>	對應 Dublin Core 的資源名稱
mo:amazon_asin	<a href="http://purl.org/ontology/mo/amazon_asin">http://purl.org/ontology/mo/amazon_asin</a>	對應 Amazon 的標準識別號
mo:musicbrainz	<a href="http://purl.org/ontology/mo/musicbrainz">http://purl.org/ontology/mo/musicbrainz</a>	對應 MusicBrainz 的曲目
mo:image	<a href="http://purl.org/ontology/mo/image">http://purl.org/ontology/mo/image</a>	音樂資源的作品表達資訊

資料來源：DBtune

## （二）開發環境和平臺

本研究運用開放原始碼的內容管理系統（Content Management System, CMS）—Drupal (<https://www.drupal.org/>)—發展鏈結資料系統。Drupal 具有靈活的内容管理功能，提供使用者針對不同的内容類型（Content Type）、使用者檔案、内容回應、分類或其他實體（Entities）自訂適合的欄位。由於其開放原始碼的特性，凡具有 PHP 程式設計能力者都可以參與開發合作，以持續改進 Drupal 核心、模組、版型及分發安裝功能，且隨著 Drupal 升級，許多重要模組亦能隨之更新。Drupal 在一個具有高擴展性核心（Core）上搭載各式各樣的模組（Module）組織建構一個網站系統，在 Drupal 7 版本提供資料發布者更廣泛的操作性介面，資料能透過核心程式 RDF Mapping API 轉換成 RDF 格式公開，任何格式的欄位都能對應到一個 RDF 屬性。

在 Drupal 中，使用者可建立不同的内容類型，每種内容型態都可自行定義其欄位（Content Field），透過内容類型所建立的頁面稱為節點（Node）。Drupal 能讓内容類型、欄位和節點分別對應到 RDF 三元組句型結構中的主語、述語和賓語。定義好内容類型，系統會產生 rdfs 的詞彙表，建立節點之後，系統會產生一個 HTTP URI，透過此機制，Drupal 中的内容都能轉換成 RDF 的語意格式，並獲取 HTTP URI 發布在網際網路上。

Drupal 釋出 RDF 功能後，其開發團隊也加入其他領域的知識本體，如 FOAF、SKOS、SIOC、OWL 和 Dublin Core 等，將這些關於網站的基本描述詞彙放入基本頁面和文章內容中，相關內容類型之間的對應如圖 1。

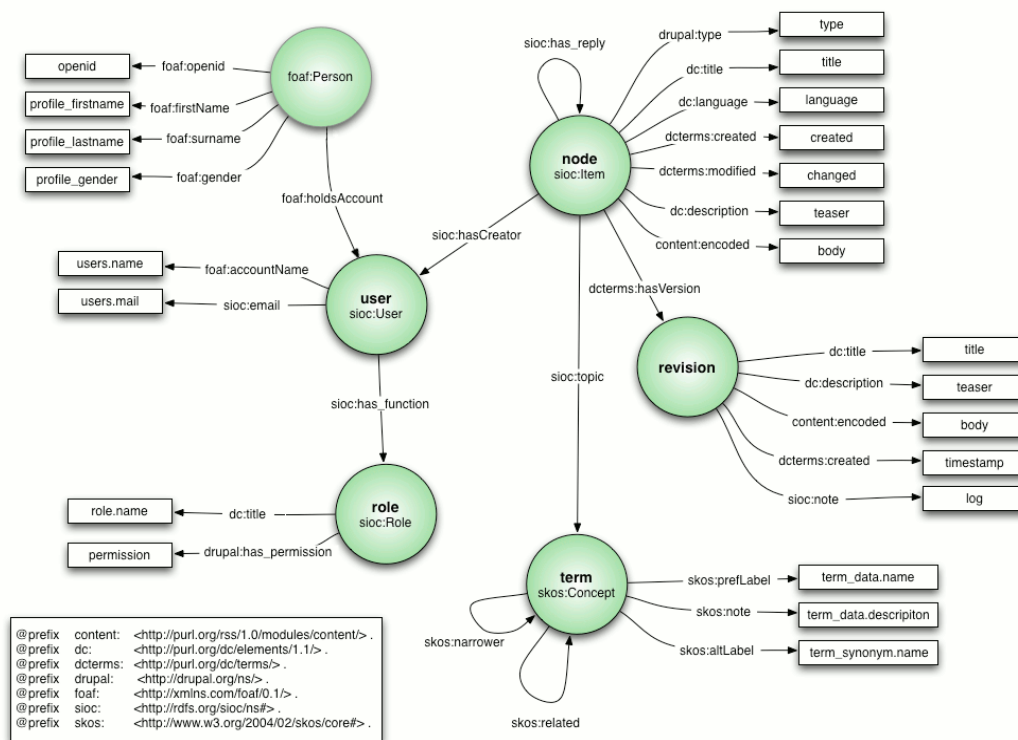


圖 1 Drupal 7 RDF Schema。取自  
<https://groups.drupal.org/files/drupalRDFschema4.png>

Drupal 提供了 RDF 的鏈結，使用者能在不同的資料集之間有目的的瀏覽。在 Drupal 原有的功能中，可把欄位值設定為網站內的另一個節點，也就是用欄位表示節點 A 和節點 B 之間的關係。當欄位被對應到一個 RDF 屬性，節點也同時轉換成 RDF 格式之後，節點 A 和節點 B 之間就能由 RDF 屬性來定義他們的關係。Drupal 每項功能，都以模組的型式開發而成的，只需透過啟用或關閉模組來決定功能是否可用，也可新增模組來增加新功能，自從加入語意網 RDF 的概念後，與 RDF 相關的模組也越來越多，本研究利用相關模組加以改寫，進而達成在 Drupal 上發展電影的鏈結資料。本研究運用的模組詳細內容如表 4。

表 4  
本系統採用的 *Drupal* 關鍵模組

模組名稱	功能
CCK	CCK (Content Construction Kit)，能在內容類型裡面添加新的欄位，滿足不同的需求。CCK 可以增加一些自訂的欄位，一般發文僅有內文、標題而已。使用 CCK 模組，能增加其他欄位以擴充內容類型所蒐集的資料型態。
RDFx	RDFx (RDF Extensions) 模組為 Drupal 提供較為全面的 RDF 支援，包括 RDF 資料的逐一建立和匯入其他的外部 RDF 資料，並支援 RDF 一系列序列化的格式，例如 RDF/XML、NTriples、Turtle。
SPARQL、 SPARQL View	SPARQL 和 SPARQL View 這兩個模組為 RDF 資料提供檢索技術，並且 SPARQL View 模組是依賴於 SPARQL 模組的。
RDF Proxy	RDF Proxy 模組用來從其它的 RDF 網站蒐集 RDF 資料。

### (三) 系統流程與架構

圖 2 為本系統鏈結資料的整體運作流程圖，首先向系統告知資料欄位所使用的 RDF 詞彙來源，再向系統說明外部鏈結資料集的 SPARQL 端點。當知道端點後，敘述資料集的資料內容，接著再從資料內容中定義系統的欄位，並作 RDF 屬性的對應，緊接著將資料欄位透過 SPARQL 查詢語法將資料取出，最後將取出的結果在本系統裡呈現。



圖 2 系統流程

圖 3 為本系統之系統架構圖，主要透過原始資料的處理與分析 (filtering)、從資料庫轉換為鏈結資料格式 (RDFizing) 和 RDF 屬性與欄位對應 (mapping) 而達成電影鏈結資料的整合，分述如下：

1. 本系統的核心資料主要來自 IMDb 資料庫，對資料進行分析，檢核與校正彙整並建立與導演、演員的關聯，將資料匯入本地資料庫，使其成為適用於本研究的資料格式，以利於與其他資料集進行比對時，能使資料在單一框架下被描述使用。
2. 透過 Drupal 內容管理系統的 RDFx 核心應用模組，將存在於本系統資料庫的資料正式轉換成鏈結資料的型式，並儲存於 RDF 倉儲 (ARC2)，等待與外部資料集的鏈結。
3. 查找和使用鏈結資料雲上與電影相關的開放資料 (如 DBpedia、LinkedMDB)，經過程式的編寫、詞彙與詞彙的對應和來源資料內容的辨識，串連載入至系統，最後將資料正式以鏈結資料的型式發布於網路，提供可連結和查找 SPARQL 端點，並呈現鏈結資料電影資訊系統介面供使用者探索。

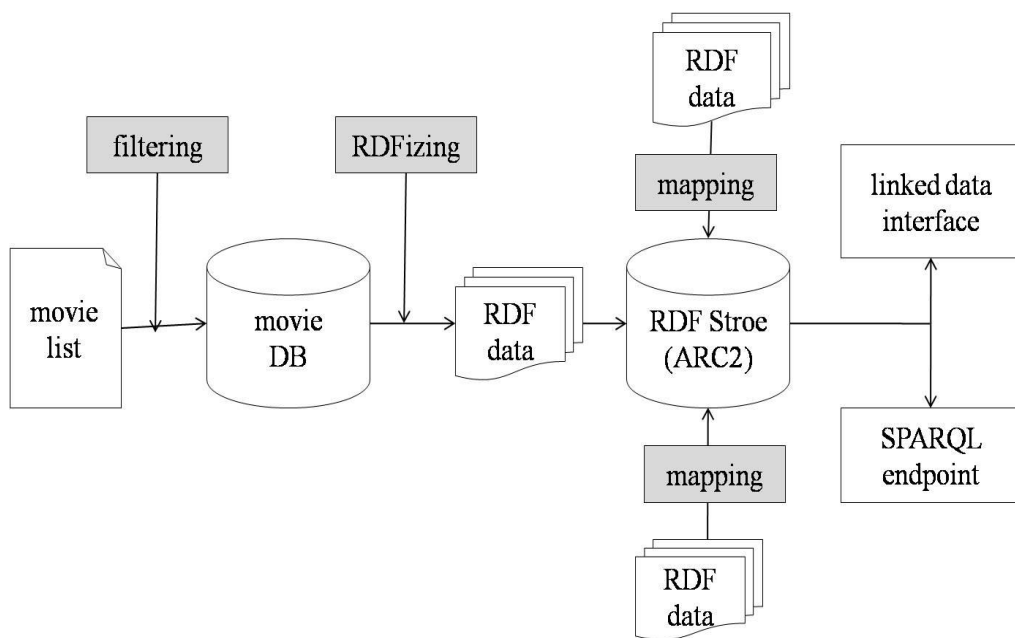


圖 3 系統架構圖

## 四、系統建構

本研究採用鏈結資料技術，將電影資訊網站提供的電影資料庫與鏈結資料雲上其他相關資料集相互連結與對應，以使本研究所建立的電影資料能盡到最大的資料分享及使用效益。本節將說明如何整合使用 Web 上的開放原始碼工具與開放的詞彙集，以將電影資料進行處理並且發布上網。首先說明本系統所規劃的內容與欄位，與所用到的模組；第二節闡述資料處理與轉換過程，包含對資料內容的檢核、依據資料內容所收集的知識本體或詞彙、資料庫中的內容轉換為 RDF 與對應，以及本系統的 URI 命名方式；第三節說明將資料載入 RDF 倉儲系統的過程，並將資料發布至 Web 上與其他資料集之間建立起連結。

### （一）內容建置

準備進行鏈結資料前，本研究將 Drupal 內容管理系統安裝至伺服器，建立電影（film）、導演（director）和演員（actor）等三種內容類型，並且規劃此三種內容類型之欄位，以便後續外部資料能與本系統鏈結，此三種內容類型的欄位說明見表 5。

表 5  
內容類型與欄位

內容類型	欄位	說明
film	field_f_title	電影名稱
	field_f_genre	電影分類
	field_f_director	電影導演
	field_f_starring	電影演員
	field_f_language	電影出版語言
	field_f_country	電影發行國家
	field_f_abstract	電影摘要
	field_f_homepage	電影官方網站
	field_f_runtime	電影播放長度
	field_f_release_date	電影上映日期
	field_f_pages	電影相關網站
	field_f_book	電影相關書籍名稱
	field_f_isbn	電影相關書籍 ISBN
	field_f_amazon	電影購買原聲帶網站
	field_f_musicbrainz	電影原聲帶相關資訊
	field_f_picture	電影相關圖片

表 5 (續)

內容類型	欄位	說明
director	field_d_title	導演名稱
	field_d_birthdate	導演出生日
	field_d_birthplace	導演出生地
	field_d_nationality	導演國籍
	field_d_residence	導演居住地
	field_d_abstract	導演簡介
	field_d_picture	導演照片
actor	field_a_title	演員名稱
	field_a_birthdate	演員出生日
	field_a_birthplace	演員出生地
	field_a_nationality	演員國籍
	field_a_residence	演員居住地
	field_a_abstract	演員簡介
	field_a_picture	演員照片

## (二) 資料處理與轉換

鏈結資料初始階段首先要對資料內容進行分析，找出適合本研究的主題以及相關性。在資料轉換前，需要進行檢核與校正，使機器得以操作資料轉換。鏈結資料的目標在於將資料作為一個可供眾人使用的資料庫，處理與分析原始資料後，下一步驟應為每筆資料加上獨一無二的 URI，使其成為網路上的資源。而後建立統一的資料格式，並選擇和重複使用 Web 上合適的詞彙來增加網路資料與資料之間的交互操作性。

### 1. 資料準備與分析

本研究使用 IMDb 的資料，內容來源有五，分別為 movie.list、directors.list、actors.list、country.list、language.list 和 genre.list。雖說各資料皆從 IMDb 獲取，但原始資料凌亂且無法直接匯入資料庫，必須刪除其他非電影資料，再與導演、演員等資料建立關聯；並且為了之後電影資料能順利與其他資料集鏈結，需將篩選後的電影資料轉換成相同格式。IMDb 的電影原始資料 (movie.list) 中包含有五種非電影的資料，分別是電視劇 (television series)、迷你電視劇 (television mini-series)、電視製作只在電視上放映的電影 (TV movie)、視訊影音 (video movie) 和電視遊戲 (video game)，本研究著重於電影資料，因此上述提及的資料皆篩選過濾。

**movie.list** 原始資料的規則，電視劇名稱前後會加上「" "」（雙引號）；迷你電視劇名稱除了前後加上雙引號之外，尚在電視劇名稱後加上「(mini)」；只在電視放映的電影加上「(TV)」，視訊影音加上「(V)」，電視遊戲則加上「(VG)」，本研究透過上述規則，過濾剩下的建立編號並匯入資料庫。**actors.list** 和 **director.list** 原始資料的規則與 **movie.list** 大致相同，導過演過電視劇的加上雙引號，以此類推，不同之處在於導過或演過非電影的導演或演員，可能也演過一般的電影，因此在過濾上述規則前，需先比對過濾後的 **movie.list** 和原始資料 **actors.list**、**director.list** 是否有同樣的電影資料，有相同電影資料則保留，完全無相同電影資料則過濾。

資料篩選過濾完成後，整理出適合本系統的電影資料概念與關係，以電影（**Movie**）資料為中心匯入至電影內容類型，包含有電影發行的國家（**Countries**）、電影出版的語言（**Language**）和電影所屬分類（**Genre**）等欄位，其餘分別匯入導演資料（**Directors**）至導演內容類型和演員（**Actors**）資料至演員內容類型，總計共有 58 萬筆電影資料、322 萬筆演員資料和 37 萬筆導演資料。

## 2. URI 設計

本研究採用 **Drupal** 作為將資料庫轉換為 **RDF** 的工具，核心是 **RDFx** 模組，在這個模組中定義使用 **W3C** 推薦的 **RDFa**。在 **RDFx** 模組下，**Drupal** 可以分析使用者的資料結構自動產生對應，關聯式資料庫轉 **RDF** 的預設對應規則是內容類型（**content type**）對應到 **RDF schema** 中的類別（**RDF class**），資料表中的欄位（**field**）作為該類別中的個體所具備的屬性（**RDF property**），資料表中的節點（**node**）視為該資源的來源（**RDF resource**）。根據鏈結資料的規則，類別、屬性與資源都會依照設定規則給定 **URI**，而文字類的資料或是非 **URI** 的資料可加上資料型態，例如字串、整數或浮點數等來幫助資源的辨識。本研究亦設計每部電影有自己的 **URI**，以便往後系統建立 **SPARQL** 端點，能找到有該電影資料的 **SPARQL** 節點。

本研究自行定義的 **URI**，乃是直接以網域名稱與目錄階層當作命名空間，對於 **URI** 中的本地名稱，則以人可閱讀理解之字詞命名取代 **Drupal** 預設的節點編號。由於本研究的來源資料中有大量的英文電影名稱和人名，因此針對以上命名方式，以「\_」（底線）取代「 」 （空格）作為本地名稱，例如電影名為 **Guardians of the Galaxy** 取代成 **Guardians\_of\_the\_Galaxy**。**URI** 命名通則如表 6。例如

Guardians of the Galaxy 這部電影，內容類型屬於 film，URI 則會以 [http://140.122.104.69:8080/movie/film/Guardians\\_of\\_the\\_Galaxy](http://140.122.104.69:8080/movie/film/Guardians_of_the_Galaxy) 呈現，此部電影導演是 [http://140.122.104.69:8080/movie/director/James\\_Gunn](http://140.122.104.69:8080/movie/director/James_Gunn)，而演員其中之一是 [http://140.122.104.69:8080/movie/actor/Chris\\_Pratt](http://140.122.104.69:8080/movie/actor/Chris_Pratt)。

表 6  
URI 命名通則

Node	URI
Movie	<a href="http://140.122.104.69:8080/movie/{contenttype}/{moviename}">http://140.122.104.69:8080/movie/{contenttype}/{moviename}</a>
Director	<a href="http://140.122.104.69:8080/movie/{contenttype}/{directorname}">http://140.122.104.69:8080/movie/{contenttype}/{directorname}</a>
Actor	<a href="http://140.122.104.69:8080/movie/{contenttype}/{actorname}">http://140.122.104.69:8080/movie/{contenttype}/{actorname}</a>

### 3. 詞彙選擇與對應

此階段的目的是按照本系統現有資料的欄位，將目前非 RDF 的資料（包括電影、導演、演員、發行國家、出版語言和電影分類）轉換成 RDF 格式，使資料有一致的標準描述標籤，本步驟利用 Drupal 的核心程式 RDF Mapping API，將儲存至資料庫的欄位對應到 RDF 詞彙。系統所採用對應詞彙皆選擇來自網路上最大且最多領域的 DBpedia Ontology，它不只提供一系列關於電影領域的知識本體，也因其處於鏈結資料雲的中心位置，擁有龐大豐富的外部連結，並且持續的更新和維護，適合做為本系統詞彙對應的來源。系統現有資料欄位與 RDF 詞彙對應如表 7。

表 7  
現有資料欄位與 RDF 詞彙對應

內容類型	欄位	RDF 詞彙
Film	field_f_title	rdfs:label
	field_f_director	dbpedia-owl:director
	field_f_starring	dbpedia-owl:starring
	field_f_country	dbpedia-owl:country
	field_f_language	dbpedia-owl:language
	field_f_genre	dbpedia-owl:genre
Director	field_d_title	rdfs:label
Actor	field_a_title	rdfs:label

Drupal 會以 RDFa 格式呈現所對應的欄位詞彙，圖 4 是 Guardians of the Galaxy 資料為例呈現的 RDFa 格式。



電影：

```
<div about=" http://140.122.104.69:8080/movie/film/Guardians_of_the_Galaxy">
<h1 property="rdfs:label">Guardians of the Galaxy</h1>
</div>
```

導演：

```
<div about=" http://140.122.104.69:8080/movie/director/James_Gunn">
<div property="dbpedia-owl:director">James Gunn</div>
</div>
```

演員：

```
<div about=" http://140.122.104.69:8080/movie/actor/Chris_Pratt">
<div property="dbpedia-owl:actor">Chris Pratt</div>
</div>
```

發行國家：

```
<div class="fields">
<div property="dbpedia-owl:country ">USA</div>
<div property="dbpedia-owl:country ">UK</div>
</div>
```

採用語言：

```
<div class="fields">
<div property=" dbpedia-owl:language">English</div>
</div>
```

電影分類：

```
<div class="fields">
<div property=" dbpedia-owl:genre">Drama</div>
<div property=" dbpedia-owl:genre">Mystery</div>
<div property=" dbpedia-owl:genre">Thriller</div>
</div>
```

圖 4 Guardians of the Galaxy 資料的 RDFa 格式呈現方式

### (三) 連結資料與發布

經過資料處理和轉換階段，系統的电影資料已經轉換成 RDF 格式。本研究查找了網路上的相關資料集，並且改寫 SPARQL View 模組，目的是透過 SPARQL 查詢

方式，依據系統所需要的鏈結的資料，向對方的 SPARQL 端點請求資料。另外是將所查詢到的資料，與本系統的电影進行比對，減少電影配對的錯誤發生。

在運用 SPARQL 查詢指令請求資料前，本系統先前已將 IMDb 資料進行處理分析和匯入資料庫，並且將欄位與 RDF 詞彙對應（見表），接著需將剩下未有資料並等待外部鏈結的欄位，像是摘要（field\_f\_abstract）、官方網站（field\_f\_homepage）等和 DBpedia、LinkedMDB、DBtune、大英圖書館等資料集所提供的 RDF 詞彙作對應，見表 8。

表 8

鏈結欄位與 RDF 詞彙對應

鏈結欄位	RDF 屬性	說明
field_f_abstract	dbpedia-owl:abstract	鏈結 DBpedia 所提供的電影摘要資料
field_f_homepage	dbpedia-owl:homepage	鏈結 DBpedia 所提供的電影官方網站資料
field_f_runtime	movie:runtime	鏈結 LinkedMDB 所提供的電影播放長度資料
field_f_release_date	movie:release_date	鏈結 LinkedMDB 所提供的電影上映日期資料
field_f_pages	foaf:page	鏈結 LinkedMDB 所提供的電影相關網站資料
field_f_book	dc:title	鏈結大英圖書館所提供的電影相關書籍資料
field_f_isbn	bibo:isbn10 bibo:isbn13	鏈結大英圖書館所提供的電影相關書籍 ISBN 資料
field_f_amazon	mo:amazon_asin	鏈結 DBtune 所提供的購買電影原聲帶的網站
field_f_musicbrainz	mo:musicbrainz	鏈結 DBtune 所提供的電影原聲帶相關資料
field_f_picture	dbpedia-owl:thumbnail	鏈結 DBpedia 所提供的電影相關圖片
field_d_birthdate	dbpedia-owl:birthDate	鏈結 DBpedia 所提供的導演出生日資料
field_d_birthplace	dbpedia-owl:birthPlace	鏈結 DBpedia 所提供的導演出生地資料
field_d_nationality	dbpedia-owl:nationality	鏈結 DBpedia 所提供的導演國籍資料
field_d_residence	dbpedia-owl:residence	鏈結 DBpedia 所提供的導演居住地資料
field_d_abstract	dbpedia-owl:abstract	鏈結 DBpedia 所提供的導演摘要資料
field_d_picture	dbpedia-owl:thumbnail	鏈結 DBpedia 所提供的導演圖片
field_a_birthdate	dbpedia-owl:birthDate	鏈結 DBpedia 所提供的演員出生日資料
field_a_birthplace	dbpedia-owl:birthPlace	鏈結 DBpedia 所提供的演員出生地資料
field_a_nationality	dbpedia-owl:nationality	鏈結 DBpedia 所提供的演員國籍資料
field_a_residence	dbpedia-owl:residence	鏈結 DBpedia 所提供的演員居住地資料
field_a_abstract	dbpedia-owl:abstract	鏈結 DBpedia 所提供的演員摘要資料
field_a_picture	dbpedia-owl:thumbnail	鏈結 DBpedia 所提供的演員圖片

SPARQL 指令於資料集進行查詢後，有資料時會回傳資料放入 SPARQL 語法所定義的查詢結果欄位內，再藉由改寫過後的 SPARQL View 模組，將查詢結果與系統欄位連結，將資料放入欄位中，再顯示於系統頁面上。

為了使 SPARQL 查詢能與本系統的電影資料相互對應，本研究將電影名稱（field\_f\_title）、導演名稱（field\_d\_title）和演員名稱（field\_a\_title）等欄位定義為一個{name}變數，透過此變數利用 SPARQL 查詢指令，向資料集請求查詢 rdfs:label 屬性內的資料與{name}變數一致的資料，如圖 5。

本系統鏈結的 SPARQL 查詢語法大致可分為電影、導演、演員、電影相關書籍、電影原聲帶，詳細說明如下。

### 1. 電影資料

圖 5 以 PREFIX 開頭的文字定義了詞彙的命名空間縮寫，例如 dbpedia-owl:Film 即為<http://dbpedia.org/ontology/Film>的縮寫，通常是一個可查找的 URI。語法中如果套用了 dbpedia-owl:Film 屬性，便接受了該屬性的所有特性。語法一（圖 5）和語法二（圖 6）分別告訴了 DBpedia 和 LinkedMDB 的 SPARQL 端點，需要使用到 rdf、dbpedia-owl、movie、foaf 的詞彙表，給定暫時定義的資料集變數和查詢的欄位，請求 rdfs:label 屬性的資料，如果這個 rdfs:label 屬性在 DBpedia 資料集內有提供 foaf:homepage、dbpedia-owl:abstract、dbpedia-owl:thumbnail 屬性的資料，以及在 LinkedMDB 資料集內有提供 movie:initial\_release\_date、movie:runtime 屬性的資料，就分別放入到 d\_homepage、d\_abstract、d\_image、mdb\_release\_date、mdb\_runtime 的欄位，並且根據 DBpedia 資料集限制，foaf:homepage、dbpedia-owl:abstract、dbpedia-owl:thumbnail 必須同時符合 rdf:type 屬性是 dbpedia-owl:Film 的資料才能回傳。本研究發現 DBpedia 資料集中有部分使用 rdfs:label 屬性也使用 foaf:name 屬性作為代替資料的名字描述，因此當語法一（圖 5）找不到符合查詢的資料就改以使用語法二（圖 6）查詢。

```
# DBpedia dataset!
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>

SELECT DISTINCT ?dbp ?d_homepage ?d_title ?d_abstract ?d_image
WHERE {
  ?dbp rdf:type dbpedia-owl:Film;
  rdfs:label "{name}";
  rdfs:label ?d_title.
  OPTIONAL { ?dbp foaf:homepage ?d_homepage }
  OPTIONAL { ?dbp dbpedia-owl:abstract ?d_abstract }
  OPTIONAL { ?dbp dbpedia-owl:thumbnail ?d_image }
}

# LinkedMDB dataset!
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX movie: <http://data.linkedmdb.org/resource/movie/>

SELECT DISTINCT ?linkedmdb ?mdb_title ? mdb_release_date ?mdb_runtime
WHERE {
  ?linkedmdb rdfs:label "{name}";
  rdfs:label mdb_title;
  OPTIONAL { ?linkedmdbmovie:initial_release_date ? mdb_release_date }
  OPTIONAL { ?linkedmdbmovie:runtime ? mdb_runtime }
}
```

圖 5 電影資料查詢語法一

```
# DBpedia dataset!

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>


SELECTDISTINCT ?film ?d_homepage ?d_title ?d_abstract ?d_image
WHERE {
  ?film rdf:type dbpedia-owl:Film;
  foaf:name "{name}";
  foaf:name ?d_title.
  OPTIONAL { ?film foaf:homepage ?d_homepage}
  OPTIONAL { ?film dbpedia-owl:abstract ?d_abstract}
  OPTIONAL { ?film dbpedia-owl:thumbnail ?d_image}
}


# LinkedMDB dataset!

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX movie: <http://data.linkedmdb.org/resource/movie/>


SELECT DISTINCT ?linkedmdb ?mdb_title ?mdb_release_date ?mdb_runtime
WHERE {
  ?linkedmdb rdfs:label "{name}";
  rdfs:labelmdb_title;
  OPTIONAL { ?linkedmdbmovie:initial_release_date ?mdb_release_date}
  OPTIONAL { ?linkedmdbmovie:runtime ?mdb_runtime}
}
```

圖 6 電影資料查詢語法二

## 2. 導演和演員資料

導演與演員的資料來源亦來自於 DBpedia 資料集（如圖 7），SPARQL 查詢模式與電影資料查詢語法大致相同，除了請求的欄位資料不同，還增加了 dbpedia-owl:nationality、dbpedia-owl:birthDate、dbpedia-owl:birthPlace、dbpedia-owl:residence 屬性，分別是導演與演員的國籍、出生日、出生地和居住

地。為了減少導演與演員之間可能人名相同，導演的 `rdf:type` 屬性要符合 `yago:Director` 值，而演員需符合 `yago:Actor` 值，才能將資料回傳。

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
SELECTDISTINCT
?dbp ?p_title ?p_abstract ?p_image
?p_nationality ?p_date ?p_place ?p_residence
WHERE {
?dbp rdf:type yago:Director;# Director content
?dbp rdf:type yago:Actor;# Actor content
    rdfs:label "{name}"; rdfs:label ?p_title.
OPTIONAL { ?dbp dbpedia-owl:abstract ?p_abstract}
OPTIONAL { ?dbp dbpedia-owl:thumbnail ?p_image}
OPTIONAL { ?dbp dbpedia-owl: nationality ?p_nationality}
OPTIONAL { ?dbp dbpedia-owl:birthDate ?p_date}
OPTIONAL { ?dbp dbpedia-owl:birthPlace ?p_place}
OPTIONAL { ?dbp dbpedia-owl: residence ?p_image}
}

```

圖 7 導演、演員資料查詢語法

### 3. 電影相關書籍資料

電影相關書籍的資料查詢語法如圖 8，資料集是來自大英圖書館所提供的鏈結資料，此資料集所使用的詞彙表，除了使用常用於網站的 `rdf`、`Dublin Core`、`SKOS` 的詞彙，也利用書日本體知識 (`Bibliographic Ontology`) 和 `ISBD` 詞彙表來表達書本資料的 `ISBN` 和頁數，本研究亦匯入上述兩者的 `namespace`，作為屬性的對應。此資料集與其他資料集不同之處在於名稱採用 `dc:title` 屬性，而將 `rdfs:label` 作為書籍的作者屬性，並使用 `bibo:isbn10` 和 `bibo:isbn13` 代表此書 `ISBN` 的屬性，`isbd:P1053` 為頁碼、`skos:notation` 為此書的版本。假如查詢的資料是空值，在本系統即不會顯示。

```

PREFIX dc: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX isbd: <http://iflastandards.info/ns/isbd/elements/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT DISTINCT
?book ?b_title ?b_author ?b_isbn10 ?b_isbn13 ?b_pages ?b_edition
WHERE {
?book dc:title "{name}"; dc:title ?b_title.
OPTIONAL { ?book rdfs:label ?b_author }
OPTIONAL { ?book bibo:isbn10 ?b_isbn10 }
OPTIONAL { ?book bibo:isbn13 ?b_isbn13 }
OPTIONAL { ?book isbd:P1053 ?b_pages }
OPTIONAL { ?book skos:notation ?b_edition }
}

```

圖 8 電影相關書籍資料查詢語法

#### 4. 電影原聲帶資料

DBtune 使用較多已存在的音樂本體知識 (Music Ontology) 來描述 MusicBrainz 的資料集，但也有自行建立詞彙表，以 vocab 為開頭的屬性。本系統的電影原聲帶資料，鏈結 mo:musicbrainz 屬性來顯示相關的音樂，亦鏈結 mo:amazon\_asin 屬性來抓取亞馬遜標準識別號碼 (Amazon Standard Identification Number) 成為能提供亞馬遜的電影原聲帶購買連結，並使用 vocab:albummeta\_coverarturl 屬性提供電影原聲帶的圖片。電影原聲帶的資料查詢語法如圖 9。

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX vocab: <http://dbtune.org/musicbrainz/resource/vocab/>

SELECTDISTINCT ?music ?m_title ?m_amazon ?m_musicbrainz ?m_image
WHERE {
?music rdfs:label "<name>"; rdfs:label ?m_title.
OPTIONAL { ? music mo:musicbrainz ?m_musicbrainz }
OPTIONAL { ? music mo:amazon_asin ?m_amazon }
OPTIONAL { ? music vocab:albummeta_coverarturl ?m_image }
}

```

圖 9 電影原聲帶資料查詢語法

總結本節鏈結了 DBpedia 的電影摘要和圖片，導演和演員的簡介、出生日、出生地、國籍、居住地和照片；LinkedMDB 的電影播放長度、電影上映日期；大英圖書館的電影相關書籍、作者、ISBN、頁數和版本；DBtune 的電影原聲帶、亞馬遜購買網站和專輯圖片。本系統所鏈結的資料集以及資料集所提供的 SPARQL 端點如表 9。

表 9  
資料集與 SPARQL 端點

資料集	SPARQL 端點
DBpedia	<a href="http://dbpedia.org/sparql">http://dbpedia.org/sparql</a>
LinkedMDB	<a href="http://data.linkedmdb.org/sparql">http://data.linkedmdb.org/sparql</a>
大英圖書館	<a href="http://bnb.data.bl.uk/sparql">http://bnb.data.bl.uk/sparql</a>
DBtune	<a href="http://dbtune.org/musicbrainz/sparql">http://dbtune.org/musicbrainz/sparql</a>

最後步驟為公開本系統的 SPARQL 端點，讓使用者也能在網路上查找本系統的資料，成為正式的鏈結資料集。Drupal 透過安裝 SPARQL 模組，開啟 SPARQL 端點的功能，藉由內建的 RDF index 功能，建立 RDF 索引，能使在 RDF 倉儲內的每筆資料經由屬性對應，並透過查詢介面，在系統的 SPARQL 端點進行查詢。本系統所提供的 SPARQL endpoint 介面如圖 10，網址為 <http://140.122.104.69:8080/movie/sparql>。

圖 10 SPARQL 端點介面



## 五、系統展示與評估

本節首先展示本系統呈現的畫面與功能(<http://140.122.104.69:8080/movie/>)，再藉由測量工具 SUS 進行系統的使用性評估。

### (一) 系統展示

使用者進入系統後，藉由主選單選擇查找的內容（電影、導演、演員），經由文字方塊輸入關鍵字，系統會將此關鍵字至匯入的 IMDb 資料做搜尋，將查詢結果呈現於畫面中。使用者也可不透過搜尋方式，讓系統預設所產生的隨機 100 筆資料進行瀏覽，如圖 11。

系統的鏈結資料呈現方式整體來說以 IMDb 資料為核心，呈現該部電影 Genre（分類）、Director（導演）、Language（語言）、Country（國家）和 Starring（演員）的內容，同時分別連接至 DBpedia 鏈結資料獲取 RDF 為 dbpedia-owl:abstract 的電影簡介和 dbpedia-owl:thumbnail 的圖片；至 LinkedMDB 鏈結資料獲取 RDF 為 movie:release\_date 的電影上映日期和 movie:runtime 的播放長度；至 DBtune 鏈結資料獲取 RDF 為 mo:amazon\_asin、mo:musicbrainz 的電影相關音樂和原聲帶照片；至大英圖書館鏈結資料獲取 RDF 為 dc:title、bibo:isbn10、bibo:isbn13 的電影相關書籍資料。

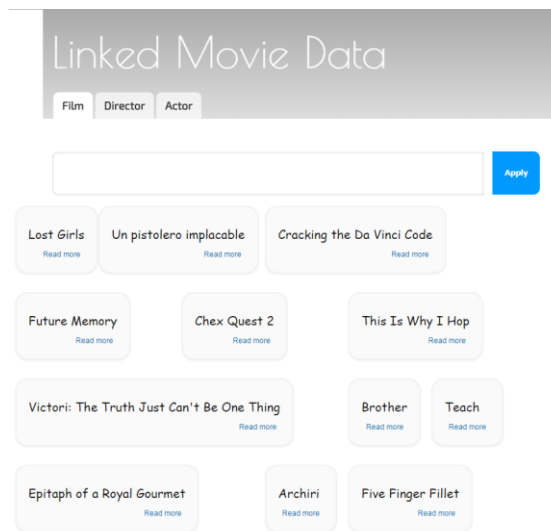



圖 11 系統首頁

圖 12 為查找「Forrest Gump」電影的例子，系統一進入電影內容頁面，便會即時發出 SPARQL 查詢至 DBpedia、LinkedMDB、DBtune 和大英圖書館的 SPARQL 端點請求此部電影的資料，分別將相對應的資料呈現於系統上。



Release Date: 1994


**Pages**

- <http://www.freebase.com/view/guid/9202a8c04000641f800000000005320c>

**Books**

[Forrest Gump / John Escott](#)  
ISBN10: 0140816127  
pages: 43p.

[Forrest Gump / Winston Groom](#)  
ISBN10: 0552996092  
pages: 238p.



**Soundtrack**

- [track](#)
- [track](#)
- [track](#)
- [track](#)

### Forrest Gump


**Genre:** Drama Romance  
**Director:** Robert Zemeckis  
**Language:** English  
**Country:** USA  
**Starring:**

Joe Abby	Rob Adams	Todd Adamson	Joe Alaskey
Markus Alexander	Sam Anderson	Neil Armstrong	Pete Auster
Peter Bannon	Kenneth Bevington	Geoffrey Blake	Jim Boeke
Charles Boswell	Jack Bowden	Grady Bowman	Arthur Bremer
Stephen Bridgewater	David Brisbin	Greg Brown	Michael Burgess
Clint Calvert	Dick Cavett	Troy Christian	Dick Clark
John Connally	Richard D'Alessandro	Jim Damron	Ed Davis
Kevin Davis	Steve DeRelian	<b>Paulie DiCocco</b>	Peter Dobson
Ryan Duncan	James Ent	Don Fischer	Michael Flannery

Forrest Gump is a 1994 American epic romantic dramedy film based on the 1986 novel of the same name by Winston Groom. The film was directed by Robert Zemeckis and starred Tom Hanks, Robin Wright, Gary Sinise and Sally Field. The story depicts several decades in the life of Forrest Gump, a naive and slow-witted yet athletically prodigious native of Alabama who witnesses, and in some cases influences, some of the defining events of the latter half of the 20th century in the United States; more specifically, the period between Forrest's birth in 1944 and 1982. The film differs substantially from Winston Groom's novel on which it is based, including Gump's personality and several events that were depicted. Filming took place in late 1993, mainly in Georgia, North Carolina, and South Carolina. Extensive visual effects were used to incorporate the protagonist into archived footage and to develop other scenes. A comprehensive soundtrack was featured in the film, using music intended to pinpoint specific time periods portrayed on screen. Its commercial release made it a top-selling soundtrack, selling over 12 million copies worldwide. Released in the United States on July 6, 1994, Forrest Gump received critical acclaim and became a commercial success as the top grossing film in North America released that year, being the first major success for Paramount Pictures since the studio's sale to Viacom earlier in the year. The film earned over \$677 million worldwide during its theatrical run. The film won the Academy Awards for Best Picture, Best Director for Robert Zemeckis, Best Actor for Tom Hanks, Best Adapted Screenplay, Best Visual Effects and Best Film Editing. It also garnered multiple other awards and nominations, including Golden Globe Awards, People's Choice Awards and Young Artist Awards, among others. Since the film's release, varying interpretations have been made of the film's protagonist and its political symbolism. In 1996, a themed restaurant, Bubba Gump Shrimp Company, opened based on the film, and has since expanded to multiple locations worldwide. The scene of Gump running across the country is often referred to when real-life people attempt the feat. In 2011, the Library of Congress selected Forrest Gump for preservation in the United States National Film Registry as being "culturally, historically, or aesthetically significant".

圖 12 電影內容頁面

使用者查找導演內容或演員內容方式與電影內容相同，利用文字方塊輸入關鍵字搜尋，系統亦會即時送出 SPARQL 語法至 DBpedia 的 SPARQL 端點請求導演或演員的出生日、出生地、居住地、國籍、簡介和照片，系統並同時提供該導演或演員所導過或演過的電影列表，演員的顯示結果如圖 13。



## Tom Hanks

**Birth Date:** 1956-07-09+02:00  
**Birth Place:** [http://dbpedia.org/resource/Concord,\\_California](http://dbpedia.org/resource/Concord,_California)  
**Birth Date:** Los Angeles

Thomas „Tom“ Jeffrey Hanks (\* 9. Juli 1956 in Concord, Kalifornien) ist ein US-amerikanischer Schauspieler, Regisseur, Filmproduzent sowie Synchronsprecher vieler amerikanischer Film- und Fernsehproduktionen. Er gehört zu den profiliertesten Charakterdarstellern Hollywoods und wurde zweimal in Folge mit dem Oscar als bester Hauptdarsteller ausgezeichnet – für seine Hauptrollen in den Filmen Philadelphia (1993) und Forrest Gump (1994).

Thomas Jeffrey "Tom" Hanks (born July 9, 1956) is an American actor, producer, writer, and director. Hanks is best known for his roles in Big, A League of Their Own, Sleepless in Seattle, Forrest Gump, Apollo 13, Saving Private Ryan, You've Got Mail, The Green Mile, Cast Away, The Da Vinci Code, Captain Phillips, and Saving Mr. Banks, as well as the animated films The Polar Express and the Toy Story series. Hanks has earned and been nominated for numerous awards during his career, including winning a Golden Globe for Best Actor and an Academy Award for Best Actor for his role in Philadelphia and a Golden Globe, an Academy Award, a Screen Actors Guild Award, and a People's Choice Award for Best Actor for his role in Forrest Gump, and earning the Stanley Kubrick Britannia Award for Excellence in Film from the BAFTAs in 2004. Hanks is also known for his collaboration with film director Steven Spielberg on Saving Private Ryan, Catch Me If You Can, and The Terminal, as well as the 2001 mini-series Band of Brothers, which launched Hanks also as a successful director, producer and writer. In 2010, Spielberg and Hanks were executive producers on the HBO mini-series The Pacific (a companion piece to Band of Brothers). As of 2012, Hanks' films have grossed over \$4.2 billion at the United States box office alone, and over \$8.5 billion worldwide making him one of the highest-grossing actors in film history.

Thomas Jeffrey Hanks (Concord, 9 luglio 1956) è un attore, regista, sceneggiatore, e produttore cinematografico statunitense. La sua carriera ha avuto inizio negli anni ottanta quando, interpretando la serie Henry e Kip (Bosom Buddies), cominciò a esser conosciuto; da allora ha recitato in numerosi film ottenendo un grande successo di critica e di pubblico. Nel corso della sua carriera ha ricevuto 5 nomination agli Oscar, 7 ai Golden Globe, 4 ai BAFTA, 6 agli Screen Actors Guild Awards e 8 agli Emmy. È l'attuale Primo Vicepresidente dell'Academy, la società che ogni anno assegna il premio Oscar.

Thomas Jeffrey Hanks (Concord, 9 de julio de 1956) es un actor y productor estadounidense más conocido como Tom Hanks. Ganador del Óscar y del Globo de Oro, Hanks es uno de los actores (junto a Luise Rainer, Spencer Tracy y Katharine Hepburn) que han ganado dos Óscar consecutivos como Mejor actuación principal y el único intérprete masculino que lo haya hecho en la segunda mitad del siglo XX. Por su protagonismo en Ángeles y demonios (2009) se convirtió en el segundo actor mejor pagado tras ingresar cincuenta millones de dólares. La taquilla lograda por sus películas en Estados Unidos supera los 4.500 millones de dólares.

توماس جيفري "توم" هانكس (ولد في 9 يوليو 1956) هو ممثل، ومخرج، وممثل صوت، وكاتب، ومنتج أمريكي حائز على جائزة الأوسكار مرتين وجائزة إيمي وأربعة جوائز الغولدن غلوب.

**Movie**

- [A Hero's Journey: The Making of Beowulf](#)
- [A Hologram for the King](#)
- [A League of Their Own](#)
- [A Timeless Call](#)
- [AFI Life Achievement Award: A Tribute to George Lucas](#)

圖 13 演員內容頁面

## (二) 系統使用性評估

本研究使用 SUS 量表評估網站的使用性，填寫量表前，先向使用者說明本系統的建置目的與主要功能，再請受試者自由操作本系統。使用者能夠查找任何電影、導

演或演員資料，隨後即開始填寫問卷。參與問卷調查者共有 23 名，其背景分別為 3 位圖資系、17 位資管系與 3 位電信系的學生。表 10 分別是 23 位使用者填寫量表的數據與換算出來的 SUS 分數。

根據使用者使用系統後所填寫的使用性量表，由表 10 可知使用者對於本系統使用性程度，使用者 C、使用者 E 與使用者 F 給予的使用性程度最高為 80 分，而使用者 V 的使用性程度最低為 55 分，整體而言使用性程度為 72 分，顯示本系統達到系統使用性量表的標準 (Bangor, Kortum, & Miller, 2009; Sauro, 2013)。

表 10  
各使用者填寫 SUS 結果

題號	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	總分
使用者 A	4	3	4	2	4	3	4	2	4	2	70.0
使用者 B	3	1	3	4	4	1	3	2	4	1	70.0
使用者 C	3	1	4	1	4	2	4	1	4	2	80.0
使用者 D	3	2	5	2	3	2	5	2	4	1	77.5
使用者 E	4	1	5	2	3	3	5	2	5	2	80.0
使用者 F	3	2	4	1	4	2	5	2	4	1	80.0
使用者 G	4	1	3	1	2	2	5	1	3	3	72.5
使用者 H	4	1	5	2	4	1	3	3	3	1	77.5
使用者 I	2	3	4	1	4	2	5	3	4	1	72.5
使用者 J	3	3	3	1	3	1	4	1	3	3	67.5
使用者 K	4	3	5	2	2	3	5	2	3	3	65.0
使用者 L	4	1	4	2	3	2	3	2	3	1	72.5
使用者 M	4	1	3	2	4	3	4	1	4	2	75.0
使用者 N	3	1	3	1	3	3	3	3	4	1	67.5
使用者 O	4	1	4	1	4	2	4	1	3	3	77.5
使用者 P	4	1	4	2	2	1	4	3	4	1	75.0
使用者 Q	2	1	4	1	4	2	4	1	4	2	77.5
使用者 R	4	2	4	1	2	1	5	2	4	2	77.5
使用者 S	4	2	4	2	4	1	3	1	4	2	77.5
使用者 T	3	3	5	1	3	2	4	1	4	2	75.0
使用者 U	3	2	4	3	4	3	4	2	4	3	65.0
使用者 V	3	3	4	3	3	3	4	3	3	3	55.0
使用者 W	4	2	4	3	4	3	4	3	4	3	65.0
平均	3.43	1.78	4	1.78	3.35	2.09	4.09	1.91	3.74	1.96	72.0

於題號 Q1、Q3、Q5、Q7、Q9 為正向回饋，根據表 11 可看出，針對這 5 題勾選同意和非常同意加起來的比率達到 70%以上的有「Q3 我認為這個電影資訊系統容易使用」、「Q7 我認為大多數的人都能快速了解學會使用這個電影資訊系統」二項，由此可知使用者普遍覺得本系統易用性高且容易學習。

在另一方面，題號 Q2、Q4、Q6、Q8、Q10 為負向回饋，根據表 11 可看出，針對這 5 題勾選不同意和非常不同意加起來的比率達到 70%以上的有「Q2 我覺得這個電影資訊系統過於複雜」、「Q4 我想我會需要技術上的協助，才能使用這個電影資訊系統」、「Q8 我覺得這個電影資訊系統使用起來非常困難」與「Q10 我需要先學習很多知識，才能開始使用這個電影資訊系統」等四題，顯示使用者普遍覺得使用本系統不需要藉由他人的協助、也無須具備額外的專業知識即可使用此電影資訊系統。負向回饋題「Q6 我認為這個電影資訊系統缺乏一致性」不同意與非常不同意未達 70%，顯示在電影資訊系統功能與內容的一致性上仍有成長空間，經由晤談得知，使用者認為本系統較新或較舊的電影，其內容有部分資料不夠完善。此外，針對負向回饋題「Q4 我想我會需要技術上的協助，才能使用這個電影資訊系統」，仍然有 4%的使用者填寫同意，透過晤談得知，使用者對於本系統的隨機推薦電影頁面，無法明確地進行電影判斷和瀏覽，且使用者認為系統的執行速度影響到瀏覽時是否需要技術上的協助，期待改善系統效率與速度。

表 11  
SUS 人數與百分比結果

題號		非常不同意	不同意	沒有意見	同意	非常同意
Q1 我會願意經常使用這個電影資訊系統	人數	0	2	9	12	0
	百分比	0%	9%	39%	52%	0%
Q2 我覺得這個電影資訊系統過於複雜	人數	11	6	6	0	0
	百分比	48%	26%	26%	0%	0%
Q3 我認為這個電影資訊系統容易使用	人數	0	0	5	13	5
	百分比	0%	0%	22%	57%	22%
Q4 我想我會需要技術上的協助，才能使用這個電影資訊系統	人數	10	9	3	1	0
	百分比	43%	39%	13%	4%	0%
Q5 我覺得這個電影資訊系統將各種功能整合的很好	人數	0	4	7	12	0
	百分比	0%	17%	30%	52%	0%
Q6 我認為這個電影資訊系統缺乏一致性	人數	6	9	8	0	0
	百分比	26%	39%	35%	0%	0%



表 11 (續)

	題號		非常 不同意	不同意	沒有 意見	同意	非常 同意
Q7	我認為大多數的人都能快速了解學會使用這個電影資訊系統	人數	0	0	5	11	7
		百分比	0%	0%	22%	48%	30%
Q8	我覺得這個電影資訊系統使用起來非常困難	人數	8	9	6	0	0
		百分比	35%	39%	26%	0%	0%
Q9	我很有自信能夠使用這個電影資訊系統	人數	0	0	7	15	1
		百分比	0%	0%	30%	65%	4%
Q10	我需要先學習很多知識，才能開始使用這個電影資訊系統	人數	8	8	7	0	0
		百分比	35%	35%	30%	0%	0%

## 六、結論

本研究藉由實作鏈結電影資料的方法，示範鏈結資料於電影資料上的應用，提供電影的查詢與瀏覽。鏈結資料的運作必須依賴資料的開放程度和需要具備與外部資源作交換的能力，本研究採用 Tim Berners-Lee 所提出的鏈結資料原則，使用知識本體和 RDF 描述資料，完成了將電影資料轉換為鏈結資料格式，發布了資料至 SPARQL 端點，並且查找了目前線上活躍使用的資料集合，透過 SPARQL 鏈結了外部的資料。茲將本研究的貢獻陳述如下。

### (一) 建立鏈結電影資料平臺

隨著鏈結資料技術應用的日益擴展，相關的工具與技術逐漸推出，作為內容管理系統的 Drupal，除了可進行編輯、發布資訊等，也結合了鏈結資料技術，發布相關應用模組，惟因模組開發人員並未持續關注，未更改成符合其相依性模組的相容性，且未進行版本上的更新，以至於功能尚未成熟，一般使用者不易使用。本研究經反覆開發與測試，歸納出本系統於 Drupal 上建立鏈結資料的規則及流程，並修改其相關模組，建立出一套以鏈結資料為基礎的電影資訊系統。

除了著重於電影資料的部分，本研究所開發之電影資訊系統也提供了瀏覽導演和演員的相關資料，能將有關於導演執導的電影和演員參與演出的電影，顯示於平臺上，使用者只需瀏覽導演或演員的鏈結資料，即可迅速了解內容或連至其他電影的鏈結資料。本研究亦建立 SPARQL 端點供使用者查詢來自 IMDb 的部分電影資料。

## （二）使電影內容更為豐富

傳統電影資訊網站的內容，需要利用網路機器人、網路蜘蛛或爬蟲程式將所需要的資料全部儲存於伺服器上，系統開發者依據系統的功能去規劃資料庫，再根據資料呈現方式抓取資料庫。上述的方式，除了需要有足夠的資料庫儲存空間外（例如電影的照片儲存空間），資料需靠單方來編輯修改。在本研究中，透過系統中的 SPARQL View 模組和 SPARQL 查詢語法的設計，開發者可以快速地向鏈結資料集請求電影相關的資料，再將資料呈現於系統上。最大的優點除了能減少系統空間外，對於資料的更新，都能藉由鏈結資料的方式同步更新，當系統鏈結了其他資料集，當資料集的內容改變，本系統的內容也會改變。

## （三）提升鏈結資料的閱讀性

過去在網際網路世界，網站內容著重於「資料如何呈現」，較少琢磨在「資料的本身」，自從語意網概念的興起後，開始重視資料本身所蘊含的資訊力量，隨即發現資料散亂於各處且各格式不一致，或是資料重複性的問題。鏈結資料技術的出現，成為一個有效的，能讓資料格式統一，並且能使機器了解此資料意義的方式。大多數使用鏈結資料的系統，對於使用者，不論以 SPARQL 查詢所回傳的結果或是網頁內容，皆以用 URI 的方式呈現資料的內容，就像是一串網址一樣，必須透過查看 RDF 詞彙在此資料集對此內容的定義為何，才能理解此 URI 資訊。本系統利用 Drupal 的網頁內容建置架構，加上 CSS 排版，將鏈結回來的資料放入指定的欄位，並給予欄位名稱，亦將圖片的 URI 直接以 HTML 圖片格式呈現。使用者可省去對應 RDF 詞彙所代表的意義，增加鏈結資料的可讀性。

## 致謝

本研究為行政院科技部專題研究計畫「以 FRBR、Linked Data、Crowdsourcing 設計圖書館館藏查詢系統之研究」（計畫編號：102-2410-H-003-121-MY3）之部分成果。

## 參考文獻

- 林朝欽、麥館碩、夏禹九 (2011)。生態資料管理與分享的新嘗試：語意連結。《國家公園學報》，21(2)，38-48。
- 柯皓仁、陳亞寧 (2013)。鏈結資料在圖書館的應用。在全國學術電子資訊資源共享聯盟，*海量資料：學術研究新境界研討會*，臺北市。Retrieved from <http://concert.stpi.narl.org.tw/Images/upload/2013/11/HaoRenKe-Report.pdf>.
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114-123.
- Becker, C., & Bizer, C. (2008, April). DBpedia mobile: A location-enabled linked data browser. Paper presented at the *1st Workshop about Linked Data on the Web (LDOW2008)*, Beijing, China.
- Berners-Lee, T. (2000). *Weaving the Web: The original design and ultimate destiny of the World Wide Web*. Harper Business, 1st edition.
- Berners-Lee, T. (2006). *Design issues of Linked Data*. Available at: <http://www.w3.org/DesignIssues/LinkedData.html>
- Brooke, J. (1986). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. *Usability Evaluation in Industry*. London: Taylor and Francis. Retrieved from <http://www.usabilitynet.org/trump/documents/Suschapt.doc>
- Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space (1st edition). *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1-136. Morgan & Claypool. Available at: <http://linkeddatabook.com/editions/1.0/>
- Linked Data. (n.d.). *What is the relationship between Linked Data and the Semantic Web? Frequently Asked Questions (FAQs)*. Retrieved from <http://linkeddata.org/faq>
- Sauro, J. (2013). *10 things to know about the System Usability Scale (SUS)*. Retrieved from <http://www.measuringu.com/blog/10-things-SUS.php>
- W3C (2013). *SPARQL Query Language for RDF*. Retrieved from <http://www.w3.org/TR/rdf-sparql-query/>
- W3C (2015). *RDF Current Status*. Retrieved from [http://www.w3.org/standards/techs/rdf#w3c\\_all](http://www.w3.org/standards/techs/rdf#w3c_all)