

# 具智能研討會雲端平台基於 Hadoop

呂欣峰

國立虎尾科技大學

資訊工程所

19963120@gm.nfu.edu.tw

許乙清

國立虎尾科技大學

資訊工程所

hsuic@nfu.edu.tw

## 摘要

社群活動的資訊,如會議資訊,公益活動,發佈在網際網路已是一種趨勢且這類的資訊量很大,主要限制是使用者需在大量資訊中手動過濾出興趣的社群活動資訊,且使用者取得這類活動資訊是局部訊息、非主動性與即時性的,這主要是因為目前的服務是缺乏智能,且這類資訊量龐大無法快速比對與過濾出使用者感興趣的活動資訊。而雲端運算是近年來在網際網路上新興的資訊服務與發展模式,提供分散式電腦主機架構,能分散並行計算與備份資料是其特點,本研究建置 Web-based 的會議系統平台,使用語意網技術來解決平台缺乏智能的問題並使用 Hadoop 雲端運算來改善效能。

**關鍵詞：**雲端運算、語意網、知識本體、Hadoop、MapReduce

## Abstract

Information of community activities, such as conference information, public service activities, etc., published on the Internet has been a trend and a large amount. The main restriction is that users manually filter interested community activities information. Additionally, users obtain these information are part of message, non-initiative and real-time. Existing services lack of intelligence to quickly filter out the interested information for user needs. Cloud computing is an emerging business model that provides resources on-demand to cloud consumers over the Internet. It provides distributed computing architecture and data backup. This work developed a Web-based conference system platform that adopted Semantic Web technology to solve the intelligence and used Hadoop to improve the

efficiency issues.

**Keywords:** Cloud Computing, Semantic Web, Ontology, Hadoop, MapReduce

## 1. 介紹

現代社會的資訊量成長速度極快,因雲端平台的蓬勃發展,使用雲端運算技術來處理海量資料是目前的趨勢,日後不管是個人資料或商業資料幾乎都將儲存在雲裡,所以藉助雲端運算對這些資料進行分析和處理來獲得更多有價值的資訊。如圖 1 所示,本研究提出以雲端運算平台 Hadoop 為基礎之多層式雲端架構來涵蓋雲端運算的三種服務模式,以整合會議資訊、知識本體與 Web 2.0 應用網站。底層為叢集式架構使用大量實體電腦組成,以提供運算資源與儲存空間為基礎設施即是服務 (Infrastructure as a Service, IaaS),並以該多層式的雲端架構為基礎發展具智能研討會雲端平台,該平台在雲端運算架構中是扮演平台即是服務 (Platform as a Service, PaaS) 的架構,於該智能研討會雲端平台上再建置研討會雲端網站,而該研討會雲端網站在雲端運算架構中是一軟體即是服務 (Software as a Service, SaaS) 的系統。

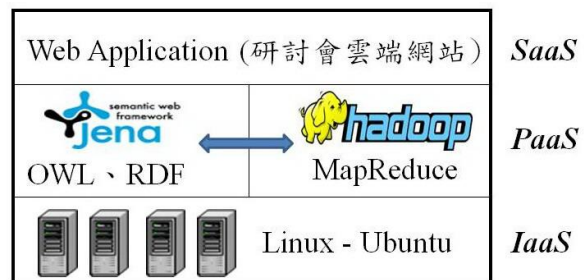


圖 1 多層式雲端架構

研討會雲端網站提供主辦方登錄研討會、演講活動等相關資訊,提供大量會議資訊給使用者,使用者透過網站可以搜尋與參加有興趣的會議主題。網站提供使用者註冊勾選個人感興趣的主題與類別,並結合知識本體技術網站將大量會議與使用者資訊,傳送給智能研

討會雲端平台透過 Hadoop 的 MapReduce 技術分散至不同的電腦進行並行的語意推論，以過濾出使用者有興趣的會議資訊。研討會雲端網站也結合現有的 Google 雲端服務包括 Google Calendar、Google Map、Google Mail 等等，此外，也將結合 QR Code 技術來做為物聯網中識別使用者身分與儲存會議資訊。

具智能研討會雲端平台特色對主辦方而言擁有可宣傳該會議的線上平台，不僅增加會議的曝光率，也可推廣會議地點附近文化、觀光產業達到雙贏；對與會者而言網站上大量會議資訊與結合語意網技術的推薦功能，並根據與會者喜好來過濾適合的會議，幫助與會者節省搜尋的時間，其他項目如透過 Google Map 規劃會議地點的交通，與會者的 Google Calendar 則會由系統寫入參與的會議紀錄，並會以 Mail 或手機簡訊來通知與會者會議時間即將到來。目前平台雖只支援學校與學術圈的會議活動，但基於圖 1 多層式雲端架構的建置與技術，可擴充在各種不同產業領域所舉辦相似性質活動，為具智能之混搭雲端服務，如科技研討會、商品發表會、藝文演講、旅遊展覽會、競技比賽等等。

本文將根據底下各章節來依序執行，首先第二章介紹本論文使用的相關研究技術與文獻探討，第三章將介紹研究方法，3.1 節中探討使用 OWL 來呈現知識本體，3.2 節中探討使用 MapReduce 所提出的方法。第四章介紹系統架構與流程，在 4.1 節為系統架構說明；4.2 節為系統流程說明。第五章展示系統程式與效能評估。最後在第六章提出結論。

## 2. 相關技術與研究

### 2.1 語意網

語意網(Semantic Web)[1,2,3]是網際網路發明人柏納斯李(Tim Berners-Lee)於 1998 年提出的概念，語意網是架構在現今網際網路之上來擴充其功能，能使得網際網路上的資訊能有更好的語意，這些豐富的語意可以描述網際網路上的任何資源，且這種語意不僅人類看得懂，軟體程式也能解讀語意網，進而作出具智慧能力的判斷。語意網是下一代網際網路發展的主流趨勢，目前很多人所討論的 Web3.0 [4]，就需採用語意網技術來達到智慧型網際網路的功能。

RDF 的全名是資源描述架構(Resource

Description Framework)[5]，也是一種資料模型，可用來描述網際網路上的任意資源，與描述資源之間的關聯性，所以 RDF 能提供簡單的語意能力來描述資源，且 RDF 這種資料模型本身也是採用 XML 語法來表示，所以每份 RDF 文件都必須是一份 Well-Formed XML 文件。但 RDF 並沒有提供一些基本的共同結構來幫助描述資源的類別(Class)與資源間關係的型態。

RDF Schema[6]是由一組字彙(Vocabulary)所組成，是一種字彙描述語言，用來制定資源的類別與屬性，這正是用來彌補 RDF 的不足，RDF Schema 擴展了 RDF 在語意描述資源上的能力，且 RDF Schema 本身也是架構在 RDF 之上，也須滿足 RDF 的資料模式與語法。

知識本體(Ontology)技術是語意網中主要使用的技術之一，可說是語意網的基礎組件[7]，在實體上知識本體是一個文件或檔案，對 Artificial Intelligence(AI)與 Web 領域來說每個知識本體都是針對某個特定領域而制定的，是用來分類網際網路上不同類別(Class)的資源與描述這些不同類別之間的關聯性，進而達到語意表達與知識的共享[8]。而用來發展知識本體的語言就稱為知識本體語言(Ontology language)，在過去幾年至今被發展出來的知識本體語言有 RDF Schema、SHOE、DAML、DAML+OIL[9]與 OWL[10]，其中 DAML+OIL 是較廣泛被使用的，DAML+OIL 是 OIL(Ontology Interchange Language)與 DAML(DARPA Agent Markup Language)所組合成的，而 OWL(Web Ontology Language)是 W3C 最新制定的知識本體語言。

目前語意網技術的推論引擎大都架構 Java 平台上，如 Jena [11]、Pellet[12]、RacerPro [13]，本研究是使用 Jena 來實踐系統推論功能的部份，Jena 是 HP LAB 所提供一語意網框架的 Java API。

### 2.2 雲端運算

依據美國國家標準與技術研究院(NIST)對於雲端運算(Cloud computing)的定義，雲端運算是一種模式，其依照需求能夠方便地存取網路上所提供的電腦資源(包括網路、伺服器、儲存空間、應用程式、服務)，可以快速地被供應，同時減少管理的工作，可降低成本並提昇效能[14]。表 1 為不同層級的雲端運算架構[15]。

表 1 層級式雲端架構

Client
Application
Platform
Infrastructure
Server

軟體即是服務(Software as a Service, SaaS) 提供雲端應用軟體在網際網路上，省去用戶在系統上安裝和執行該應用程式，基於網路市售軟體，從集中位置管理，使客戶通過網路遠程存取應用和管理，主要提供服務的有 Google、Oracle、IBM、Microsoft，其中 Google Apps 算是最廣泛的 SaaS 應用[16]。

平台即是服務 (Platform as a Service, PaaS) 提供使用雲端基礎設施的計算平台，它通常用來佈署客戶端的所有應用程式。因此客戶不必去購買和安裝所需的軟體和硬體，通過這項服務開發人員可以得到一個軟體生命週期所需的所有系統和環境的保持，無論是開發、測試、部署和託管 Web 應用程式，如 Google App Engine[17]、Microsoft Azure[18]。

基礎設施即是服務 (Infrastructure as a Service, IaaS) 提供了必要的基礎設施作為一種服務客戶端不需要購買所需的伺服器、資料中心或網路資源，這裡的關鍵優勢是客戶需要付出的只是他們使用的服務和時間。其結果是客戶可以以更低的成本實現更快的服務交付[19]。

### 2.3 Hadoop

Hadoop[20] 是一種雲端運算平台為 Apache 軟體基金會底下的開放原始碼計劃，主要運行於 Linux 上由 Java 寫成，提供大量資料的分散式運算環境專門用來架設叢集式電腦，其中 MapReduce 與 HDFS(Hadoop Distributed File System)是 Hadoop 的兩大核心，MapReduce 實現了分散式運算，而 HDFS 實現了分散式的檔案系統，並擁有高容錯率、資料本地化、簡化平行程式設計和擴充容易等優點，適合在超大資料量下執行，目前有許多大型企業都在使用，如 Yahoo、Google 等等，就算有龐大的資料量也可以利用簡單的擴充硬體設備來提升系統效能。

MapReduce[21]為 Google 在 2004 年所提出，為 Google 的核心運算模型，是一種並行的程式設計模型，它可以將工作分發到上千台電腦組成的叢集上進行運算處理。一個 MapReduce job 通常會先把輸入的資料集進行分割，並由 map task 以並行的方式處理資料，map 輸出的資料經過排序後，把結果輸入給 reduce 工作，作業的輸入和輸出都會被儲存在 HDFS 裡。

HDFS[22]主要採用了主從式架構模型，一個 HDFS 叢集是由一個 NameNode 與許多 DataNode 所組成的，NameNode 為主要伺服器管理檔案系統的命名空間與檔案的存取操作，叢集中的 DataNode 主要為管理儲存的檔案，檔案從內部來看被分為許多資料區塊存放在 DataNode，並由副本機制進行區塊的備份，HDFS 藉由分散式儲存技術來提升儲存空間與異地備援能力。

## 3. 研究方法

### 3.1 使用 OWL 呈現知識本體

知識本體(Ontology)主要用來對網路上任何資源進行知識描述和定義，例如同一個詞在不同的領域或不同的應用上，它的意義就不一定會一樣。知識本體主要在不同領域間用來呈現資源與資源的 Relation，其中包含許多特性不管是垂直父子類別關係或水平的邏輯關係，包括：Inverse、Symmetric、Transitive、Functional、equivalentClass 等等，可用來描述具特殊關係之 Property，例如：Inverse Property 可用來描述具相反關係之 Class，Symmetric 可用來描述具對稱關係之 Class，Transitive 可用來描述具遞移關係之 Class，讓不同的 Property 透過這些特性，將相關性能簡易地定義，而用來描述知識本體(Ontology)的語言為 Web Ontology Language(OWL)由 W3C 所制定。

圖 2 所示為具智能研討會雲端平台的知識本體，使用 UML 來呈現，領域制定在 Information Technology 下的會議資訊與演講活動。圖中 Information Technology 為父類別其 Internet、Computer Math、Artificial Intelligence、Information Security、Programming、Algorithm 為其子類別，Internet 類別下 Web Technology、Cloud Computing、Sensor Network、Internet of Things、Mobile

Computing 為其子類別，子類別會繼承父類別的屬性關係(Property)。

圖 2 中 IT 對於 Information Security 之間有 Require 的 Object Property，而 RequiredBy 的 inverseOf Property 為 Require。IT 對於 IT 之間有 SimilarTo 的 Transitive Property，底下接舉例說明 Property。

在 Cloud Computing 類別下有雲端策略研討會、漫步在雲端、雲端計算系列三個會議資訊物件，因在 IT 有定義了 SimilarTo 為 Transitive Property，所以雲端策略研討會相似

於漫步在雲端，漫步在雲端相似於雲端計算系列，則雲端策略研討會相似於雲端計算系列。

在 Information Security 類別下有雲端計算系列、建立安全的雲端運算環境、雲端計算系列兩個會議資訊物件，IT 對於 Information Security 之間有 Require 的 Object Property，而 RequiredBy 的 inverseOf Property 為 Require，所以雲端計算系列需要建立安全的雲端運算環境，則建立安全的雲端運算環境被雲端計算系列所需要。

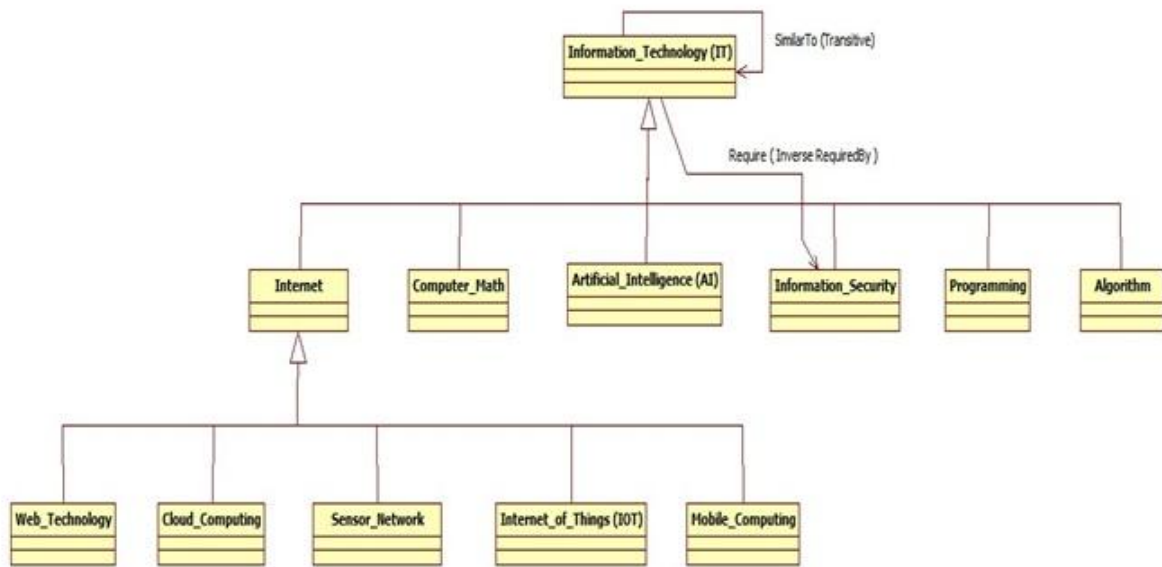


圖 2 知識本體

### 3.2 使用 MapReduce 結合 Jena

本研究提出使用語意網推論引擎 Jena 架構在 MapReduce 框架上，運用 MapReduce 並行處理的能力來減少大量資料進行推論時所耗費的時間。圖 3 所示為 MapReduce 流程，在一份 Fact RDF 檔案裡會記錄著所有物件與物件間的關係，這邊所指的物件即是會議資訊，當會議資訊的不斷增加資料量不斷變大，所以在輸入資料的部分由 MapReduce 框架去讀取 HDFS 內分割好的 Fact RDF 檔案資料集，來進行 Map 階段的處理，每個 Map 程序置入知識本體 OWL 檔案與 Fact RDF 檔案使用 Jena API 進行並行式的語意推論，來增加其效能。在中間值與 Reduce 階段在本例上是可有可無的階段，最後的推論結果可以選擇在 Map 程序結束後輸出，也可以經由 Reduce 進行一些資料的整合或處理後，輸出到 HDFS 上。

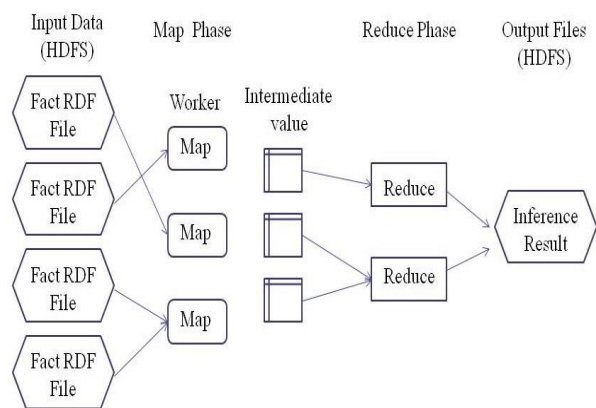


圖 3 MapReduce 流程



## 4. 系統架構與流程

圖 4 所示為混搭雲端服務平台之精神，由雲端運算、物聯網、語意網所整合之應用，強調智慧型物聯網應用於雲端運算。

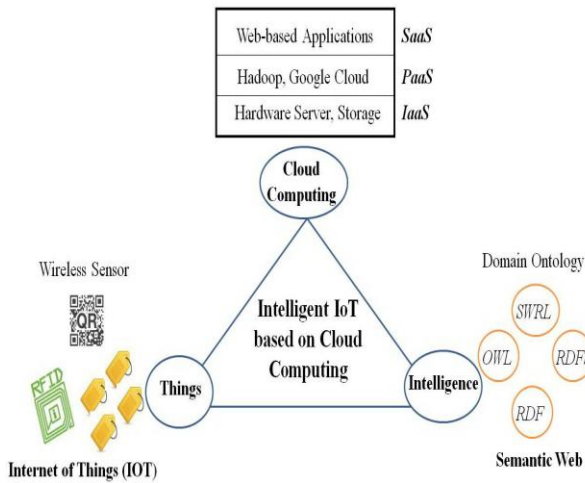


圖 4 混搭架構圖

### 4.1 系統架構

具智能研討會雲端平台網站為 Web-based 平台，表 2 為系統開發工具與技術，運行於 Apache Server 由 PHP 撰寫而成，後端為雲端運算平台 Hadoop 運行於 Linux Ubuntu 系統上，用來架設叢集式電腦，提供分散並行計算與備份資料的能力，另外提供 Web Service 主要有 RSS Parser 服務與會員身份確認(Identification)服務。主辦方除了透過網頁表單至平台建置會議資訊外，並提供 RSS Parser Web Service 來剖析主辦方會議網站的 RSS Feed 資訊來新建會議資訊，而身份確認的 Web Service 為主辦方送出與會者識別碼至平台上檢查是否為我們網站會員，以便提供後續的服務。而後端 Jena 推論引擎結合 Hadoop 來處理語意網推論部分。平台上也結合 Google Calendar 幫助與會者會議日程紀錄與通知、Google Map 路線規劃，平台也會產生二種 QR Code，其一記錄與會者資訊達到識別身份，其二記錄主辦方建置的相關會議資訊，供主辦方列印海報時可附上 QR Code。

表 2 開發工具與技術

項目	開發環境與技術
作業系統	Linux Ubuntu 12.04
雲端運算平台	Hadoop 1.0.3

Web Server	Apache HTTP Server
程式語言	Java、PHP
Java Development kit	JDK 1.6
整合開發環境	Eclipse SDK 3.7.2
XML Parser	JDOM 1.1.1
資料庫	MySQL
語意網技術	Ontology、OWL、RDF
推論引擎	Jena API 2.6.4
物聯網	QR Code
API Tools	Google API Tools

由圖 5 所示為多層式系統架構圖，由下而上堆疊而成。實體機器層(IaaS)由許多機器組成叢集式的運算環境，提供運算能力與儲存空間。HDFS 儲存層則存放運算所需的資料集、知識庫等等。演算層(PaaS)使用 MapReduce 框架結合 Jena 推論引擎，來實現並行式語意推論系統加快資料處理的速度。應用層(SaaS)可提供各種不同領域的語意雲端應用系統，本研究領域是資訊技術先關的研討會、會議或演講活動。終端用戶層不管是透過瀏覽器或行動設備皆能存取語意雲端應用服務。

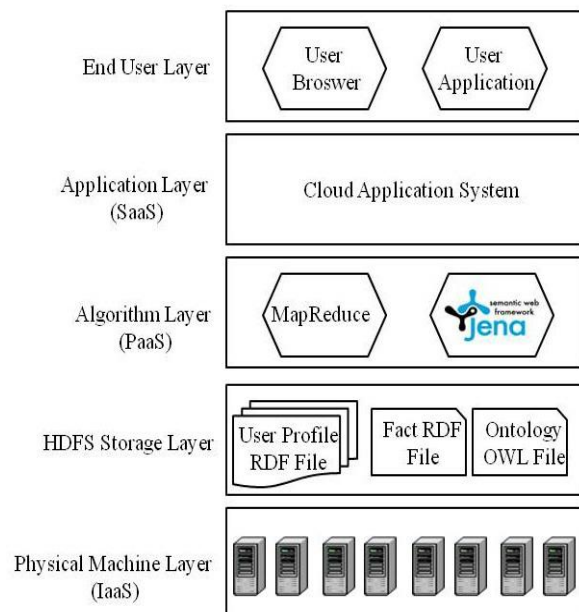


圖 5 多層式系統架構

## 4.2 系統流程

圖 6 為具智能研討會雲端網站流程圖，1.輸入資料的部分，由主辦方透過 Web Service 或網頁表單來建置會議資訊於網站上，頁面上顯示相關會議資訊供與會者瀏覽，並提供會議資訊 QRCode 給主辦方以結合物聯網應用。2. 網站使用了很多 Google 雲端服務與 API 的應用，如 QRCode 的產生，Google Calendar 幫助與會者會議日程紀錄與通知，Google Map 路線規劃，Google 通訊錄的會議推薦給好友服務..等等。3.當會議資訊被建立時，網站會根據知識庫知識本體的定義來更新 Fact RDF 檔案，建立會議資訊間的關聯性。4.使用者在網

站上註冊基本資料外也會勾選領域與喜好。5.使用者註冊資訊產生 User Profile RDF 檔案用來幫助網站過濾使用者有興趣的研討會資訊。6. 使用者註冊後可得到平台提供的智慧型推薦功能，由後端 Jena 推論引擎結合 Hadoop 使用 MapReduce 進行語意推論。9.語意推論結果使用 User Profile RDF 檔案來進行過濾，並返回符合使用者之結果。10.網站提供的服務除了語意推薦功能外，當與會者選擇參加某場會議，平台會將會議日程寫入與會者 Google Calendar，且產生 QR Code 記錄所參加的會議資訊與身份識別用，會議時間即將來臨也會寄發 Mail 或簡訊來通知與會者。

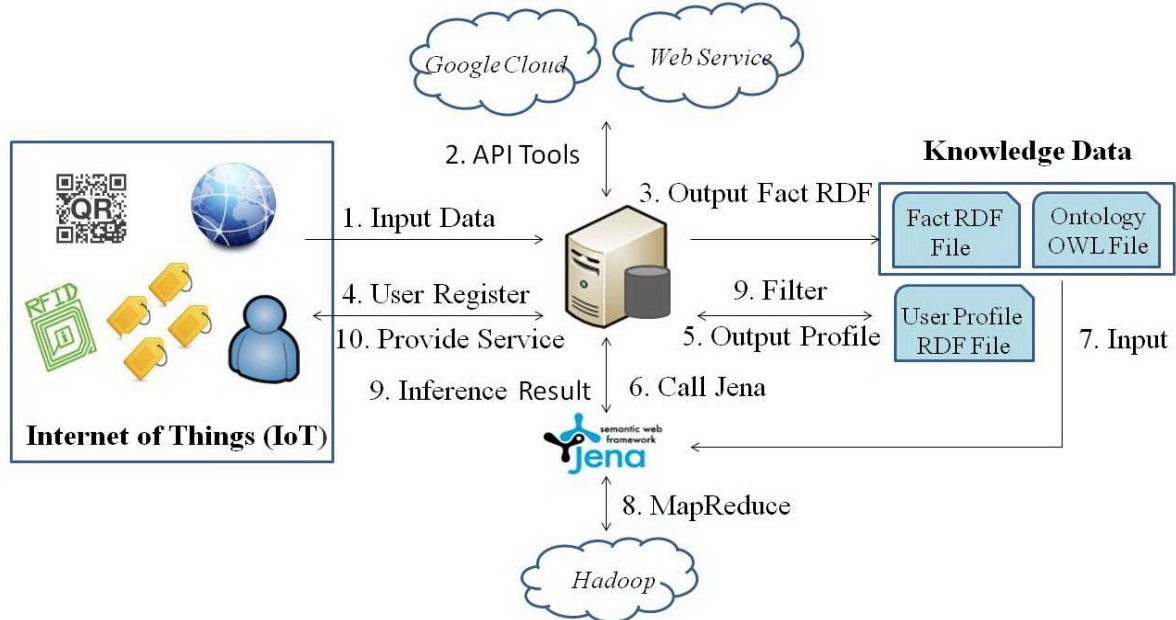


圖 6 具智能研討會雲端平台流程

## 5. 系統展示與效能評估

### 5.1 系統展示

如圖 7 所示為網站首頁，提供基本的註冊功能、RSS 訂閱服務、會議資訊的瀏覽。圖 8 為註冊頁面，使用者可在網站註冊申請帳號，並填寫 Gmail 帳號以便提供後續 QRCode 的寄送與行事曆的追蹤。勾選感興趣的領域與喜好由平台產生 User Profile RDF 檔案，檔案內記錄了使用者會員資訊與喜好的類別，用來協助網站過濾出使用者感興趣的會議資訊。



圖 7 網站首頁

◎歡迎註冊

帳號： (請輸入 Gmail 帳號)

密碼： (請輸入 Gmail 密碼，系統將會加密處理)

信箱： (請輸入 Gmail)

姓名：

電話：

身分證字號： (必須)

◎請勾選個人領域與喜好：

模糊邏輯  雲端運算  語意網  資訊安全  網路網路  行動運算  人工智慧

計算機數學  程式設計  演算法  物聯網  感測網路  Web技術

圖 8 註冊頁面

圖 9 為登入頁面，提供註冊使用者更多的服務，使用者可登錄會議，使用智能會議推薦功能，使用 Google 行事曆追蹤與提醒已參加的會議等等。

◎登入會員

具智能研討會雲端平台

◎會議表單填寫

名稱	主辦方	日期	地址
未來的學習 - 雲端教學創新服務	李進賢所長	2012-10-04	財團法人資訊工業策進會數位教育研究所
可視化設計能力 程式設計 Mentor 平台	程學院	2012-09-28	402 台中市南區國光路250號
演算法如何塑造世界	Kevin Slavin	2011-07-01	
1012 雲端計算系列演講	大同大學資訊管理系	2012-09-12	台北市中山區104中山北路三段四十四號
物聯網學術研討會	王心童博士	2013-05-17	716 台南市南門外東區大馬路 949 號
ANET2012 台灣網路研討會	銘傳大學	2012-10-23	桃園縣龜山鄉德明路 5 號
臺南市社大聯盟 - 第六屆台灣資訊社會研究學會年會暨學術研討會(TAIS 2009)	曾淑芬	2009-11-14	新竹市大馬路1001號 交通大學
2010 年第 10 屆資訊管理學術暨實務研討會	林宜璇	2009-05-12	桃園縣龜山鄉大崗村行人路56號 中央警察大學榮光樓
2013 年第 7 屆資訊管理學術暨實務研討會	陳德輝、陳志輝	2003-05-09	桃園縣龜山鄉大崗村行人路56號 中央警察大學榮光樓100會議室
ANET2011 台灣網路研討會	國立宜蘭大學	2011-10-24	832 雲林縣虎尾鎮文化路64號 蔣經國大學

歡迎 - xmlab-403 用戶

國立資訊科技大學資訊工程系  
Copyright © 2012 資訊管理學術暨實務研討會  
請於 2012 年 10 月 1 日以前完成報名

圖 9 登入頁面

如圖 10 所示，會議主辦方可在網站登錄會議資訊或演講活動，也提供會議主辦方以 RSS Feed 的方式輸入到網站來登錄會議資訊，由系統後端 RSS Parser Web Service 進行剖析，如圖 11 所示。

◎會議表單填寫

會議代碼：

會議名稱：

會議主辦：

會議位置：

會議日期：

會議網址：

會議類別：

模糊邏輯  雲端運算  語意網  資訊安全  網路網路  行動運算  人工智慧

計算機數學  程式設計  演算法  物聯網  感測網路  Web技術

圖 10 登錄會議

網站提供會議方網站RSS登錄會議，請遵循規範。

Input RSS URL:

◎範例下載

```
<?xml version="1.0" encoding="UTF-8"?>
- <rss version="2.0">
  - <channel>
    <!--主要剖析器下的會議資訊儲存在資料庫-->
    <title>第13屆模糊研討會</title>
    <!--研討會網站的主題-->
    <link>http://nfu.edu.csie/fuzzy/</link>
    <!--研討會網站的連結網址-->
    <description>632雲林縣虎尾鎮文化路64號管理大樓1F</description>
    <!--研討會網站的描述-->
    <pubDate>2010-12-13 20:25:1 GMT</pubDate>
    <!--研討會網站更新日期-->
    <language>zh-tw</language>
    <!--zh-TW 繁體中文語系-->
    <category>fuzzy</category>
    <!--加議本體定義之類別-->
    + <image>
    + <item>
    + <item>
  </channel>
</rss>
```

圖 11 RSS Parser Web Service

登錄會議成功後，網站會顯示相關會議資訊並產生 QRCode，QRCode 包含該會議的相關資訊以便使用於主辦方後續海報印刷等服務如圖 12 所示。

◎會議登錄成功

會議代碼：78964

會議名稱：未來的學習 - 雲端教學創新服務

會議主辦：李進賢所長

會議網址：

會議位置：財團法人資訊工業策進會數位教育研究所

會議日期：2012-10-04

會議類別：cloud computing



QRCode 已寄出!

圖 12 登錄會議成功

使用者可以在網站瀏覽相關感興趣的會議資訊後，連結到主辦方會議網站報名參與該會議或演講活動。主辦方在會議報名時程結束後，可提供給我們參與該會議的使用者識別碼，如圖 13 所示。透過系統後端驗證識別碼的 Web Service 來寫入資料庫記錄所有使用者所參與的會議，並產生 QRcode 寄發給與會者，如圖 14 所示。

◎登錄資料

Call Web Service (WSDL URL)

Input:

◎範例下載

```
<?xml version="1.0" encoding="UTF-8"?>
<!--IDCTYPE participants SYSTEM "http://140.130.34.47/participants.dtd"-->
<!--old為研討會會議識別碼-->
- <mid為與會者身份識別碼-->
- <conference id="63577">
  - <p>
    - <mid>P123596575</mid>
  </p>
</conference>
```

圖 13 身份驗證 Web Service



寫入資料庫完成!!  
寄發QRCode，資料量大時請稍候!!  
xm1ab403@gmail.com



編號: 63577  
名稱: 第13屆模糊研討會  
主辦方: 國立虎尾科大電資學院資訊工程所  
地址: 632 雲林縣虎尾鎮文化路64號管理大樓1F  
日期: 2012-11-14  
單位: 國立虎尾科大電資學院資訊工程所  
<http://140.130.34.50/conference/organizer/verification.php?num=P123190256>

kingispeak@gmail.com



編號: 63577  
名稱: 第13屆模糊研討會  
主辦方: 國立虎尾科大電資學院資訊工程所  
地址: 632 雲林縣虎尾鎮文化路64號管理大樓1F  
日期: 2012-11-14  
單位: 國立虎尾科大電資學院資訊工程所  
<http://140.130.34.50/conference/organizer/verification.php?num=P123190245>

圖 14 與會者 QRCode

網站結合 Google 行事曆 API 功能，將定期在後端資料庫系統中撈取使用者所參與的會議或演講活動，寫入 Google 行事曆，由行事曆以手機簡訊或信箱通知與會者會議時間，如圖 15 所示。



圖 15 Google 行事曆

與會者持系統寄發之 QRcode，到會場時透過掃描 QRCode 可導向到主辦方簽到頁面執行報到功能，如圖 16 所示。

簽到

請輸入身分證：R123456789

姓名：王傳一  
身分證：R123456789  
簽到確認!!

查詢

請輸入身分證：R123456789

姓名：王傳一  
身分證：R123456789  
簽到：YES

圖 16 會場簽到與查詢頁面

## 5.2 效能評估

實驗使用了 1 個約 225K 大小的 Fact RDF 檔案，並分割為 20 個約 12K 大小的 Fact RDF 檔案來進行效能測試。表 3 為平台測試環境，表 4 與表 5 為效能測試結果。

表 3 測試環境

項目	開發環境與技術
作業系統	Linux Ubuntu 12.04
JDK	Java 1.7.0 openJDK
SSH	OpenSSH 5.9
雲端運算平台	Hadoop 1.0.3
整合開發環境	Eclipse SDK 3.7.2
推論引擎	Jena 2.6.4

表 4 與表 5 為執行 10 次效能平均時間，表 4 可以看出在一般 PC 未使用 MapReduce 框架進行語意推論運算時，當 Fact RDF 檔案因物件量的增長變大時，會影響到其效能。我們分割該 RDF Fact 檔案使用 MapReduce 執行分散式並行運算，所得到的執行時間是非常快速的，如表 5 所示。

表 4 未使用 Hadoop 效能測試

資料量 / 模式	10K~15K Fact RDF 檔案 1 個	200K~225K Fact RDF 檔案 1 個
一般 PC	29.70(s)	678.93(s)

表 5 使用 Hadoop 效能測試

資料量 / 模式	200K~225K Fact RDF 檔案 1 個	10K~15K Fact RDF 檔案 20 個
Hadoop (1-Node)	764.27(s)	120.33(s)

## 6. 結論

本研究提出多層式的雲端運算架構，強調整合語意網技術於雲端運算平台 Hadoop 上，運用 MapReduce 並行計算特點，提供執行大量語意推論時所需的運算能力，並實作雲端服務平台來驗證，且因知識本體領域的不同雲端服務平台擁有其高擴展性。

未來將整合 Eucalyptus 提供虛擬主機平



台功能以利商業服務與費用計算。在異業整合方面也將與會議地點的交通業及住宿業合作，提供較低折扣來吸引人潮，推廣地方產業，提供例如地方小吃推薦、特產推薦、景點推薦...等，而平台領域則可擴大到不只提供會議資訊，也可提供相似性質的活動資訊。

## 參考文獻

- [1] Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284(5), 34-43.
- [2] Shadbolt, N., Berners-Lee, T. & Hall, W. (2006). The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3), 96-101.
- [3] I-Ching Hsu, SXRS: An XLink-based Recommender System using Semantic Web technologies, *Expert Systems with Applications*. 36, 2009, 3795-3804.
- [4] J. Hendler (2008), Web 3.0: Chicken Farms on the Semantic Web, *Computer*, 41(1), 106 - 108.
- [5] Klyne, G. & Carroll, J. J. W3C Recommendation (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax. available at: <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/> [accessed 26 November 2008]
- [6] Brickley, D. & Guha, R. V. W3C Recommendation (2004). RDF Vocabulary Description Language 1.0: RDF Schema. available at: <http://www.w3.org/TR/rdf-schema> [accessed 26 November 2008]
- [7] Studer, R., Hotho, A., Stumme, G., et al. (2003). Semantic Web - State of the Art and Future Directions. *Künstliche Intelligenz*, 17(1), 5-9.
- [8] Maedche, A., Motik, B., Stojanovic, L., et al. (2003). Ontologies for Enterprise Knowledge Management. *IEEE Intelligent Systems*, 18(2), 26-33.
- [9] Horrocks, I., Patel-Schneider, P. F. & Harmelen, F. v. (2002). Reviewing the Design of DAML+OIL: An Ontology Language for the Semantic Web. *Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence*. AAAI Press (pp. 792-797), Edmonton, Alberta, Canada
- [10] Smith, M. K., Welty, C. & McGuinness, D. L. W3C Recommendation (2004). OWL Web Ontology Language Guide. available at: <http://www.w3.org/TR/owl-guide/> [accessed 26 November 2008]
- [11] Jena, <http://jena.sourceforge.net/>
- [12] Pellet, <http://clarkparsia.com/pellet/>
- [13] RacerPro, <http://www.racer-systems.com/>
- [14] National Institute of Standards and Technology - Computer Security Resource Center - [www.csrc.nist.gov](http://www.csrc.nist.gov)
- [15] <http://zh.wikipedia.org/wiki/雲端運算>
- [16] Google Apps, <http://www.google.com/enterprise/apps/business/>
- [17] Google App Engine, <https://developers.google.com/appengine/>
- [18] Microsoft Azure, <http://msdn.microsoft.com/zh-tw/windowsazure/centrum-windows-azure.aspx>
- [19] Bhaskar Prasad Rimal, Eunmi Choi, "A taxonomy and survey of cloud computing systems", 2009 Fifth International Joint Conference on INC, IMS and IDC, published by IEEE Computer Society.
- [20] Apache Hadoop, <http://hadoop.apache.org/>
- [21] Jeffrey Dean, Sanjay and Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*. Page107-113, Volume 51, Issue 1, 2008.1.
- [22] HDFS, [http://hadoop.apache.org/docs/hdfs/current/hdfs\\_design.html](http://hadoop.apache.org/docs/hdfs/current/hdfs_design.html)