出國報告（出國類別：國際會議）

參加 IAE 協會在倫敦舉辦的「資料探勘與知識工程」國際會議出國報告

服務機關：行政院主計處電子處理資料中心

姓名職稱：陳設計師嘉華

派赴國家：英國

出國期間：100/7/4~100/7/10

報告日期：100/9/23

目　　錄

壹、目的

由 IAE (International Association of Engineers)協會在英國倫敦舉辦的國際工程大會(World Congress on Engineering；簡稱 WCE) 2011 年會，自 7 月 6 至 8 日爲期 3 天，是由不同領域的主題會議所合辦而成，專題發表的題材豐富且內容廣泛，此次參加的「資料探勘與知識工程」(ICDMKE)國際會議，係屬該年會 15 個主題會議之一。

希冀藉由參加本次會議，蒐集國際對「網路探勘」、「資料探勘」、「資料倉儲」、「知識系統」、「WEB 服務」、「憑證驗證快速更新、「雲端服務安全」...等等諸多議題的技術經驗及研究成果，用以學習暨引用於現行公文及檔管的業務當中，以期能對未來新版公文資訊系統之上線及推行能有所助益。

貳、過程

此次參加的 ICDMKE 國際會議，因屬 WCE 2011 年會 15 個主題會議之一。而 WCE 年會每年固定由 IAE 協會(International Association of Engineers)定期舉辦，今年會議於 2011 年 7 月 6 日至 8 日在倫敦帝國理工大學南肯辛頓校區(South Kensington campus, Imperial College London)舉行，是國際上重要的大型綜合研討會之一。以下爲會議的網址與舉辦地點「帝國理工大學南肯辛頓校區」。

http://www.iaeng.org/WCE2011/

WCE2011 年會之 15 個不同領域的主題會議名稱羅列如下：

- The 2011 International Conference of Applied and Engineering Mathematics (ICAEM'11)
- The 2011 International Conference of Computational Intelligence and Intelligent Systems (ICCIIS'11)
- The 2011 International Conference of Computational Statistics and Data

Engineering (ICCSDE'11)

- The 2011 International Conference of Computer Science and Engineering (ICCSE'11)
- The 2011 International Conference of Data Mining and Knowledge Engineering (ICDMKE'11)
- The 2011 International Conference of Electrical and Electronics Engineering (ICEEE'11)
- The 2011 International Conference of Financial Engineering (ICFE'11)
- The 2011 International Conference of Information Engineering (ICIE'11)
- The 2011 International Conference of Information Security and Internet Engineering (ICISIE'11)
- The 2011 International Conference of Mechanical Engineering (ICME'11)
- The 2011 International Conference of Manufacturing Engineering and Engineering Management (ICMEEM'11)
- The 2011 International Conference of Parallel and Distributed Computing (ICPDC'11)
- The 2011 International Conference of Systems Biology and Bioengineering (ICSBB'11)
- The 2011 International Conference of Signal and Image Engineering (ICSIE'11)
- The 2011 International Conference of Wireless Networks (ICWIN'11)

15 個主題會議中，主要領域包含有：應用和工程數學(ICAEM'11)、計算智能與智能系統(ICCIIS'11)、統計計算和數據工程(ICCSDE'11)、計算機科學與工程(ICCSE'11)、資料探勘和知識工程(ICDMKE'11)、電氣和電子工程(ICEEE'11)、金融工程(ICFE'11)、資訊工程(ICIE'11)、資訊安全和網路工程(ICISIE'11)、機械工程(ICME'11)、製造工程及工程管理(ICMEEM'11)、並行和分散式計算(ICPDC'11)、系統生物學與生物工程(ICSBB'11)、信號與影像工程(ICSIE'11)及無線網路(ICWIN'11)等。

根據 WCE2011 年會統計，今年計有來自 50 餘個國家、高達 1328 篇投稿論文，其中有 756 篇的論文順利被審查通過，錄取率約爲 56.93%，而每一領域大約有四十

餘篇的論文被發表，所發表的文章最後彙整成冊出版(ISBN：978-988-18210-6-5)。

摘錄「資料探勘與知識工程」(ICDMKE'11)、「統計計算和數據工程」(ICCSDE'11)及「資訊安全和網路工程」(ICISIE'11)會議，所發表的場次及主題如下：

數據挖掘和知識工程 ICDMKE I [第一場次]	
主持人：博士納思 GANAPATHY	
1	**A User-Interests Approach to Music Recommendation**
	Mr. Chen-Chang Wu
2	**Representing, Storing and Mining Moving Objects Data**
	Mr. Jorge Luis Huere Pena
3	**The Robust Classification for Large Data (Case: Classification of Jakarta Vegetation area by Using Remote Sensing Data)**
	Dr. DYAH ERNY HERWINDIATI, and Ms. Desi Arisandi
4	**Matching and Merging of Ontologies Using Conceptual Graphs**
	Prof. RAVI LOURDUSAMY
5	**A Study of Intrusion Detection in Data Mining**
	Prof. EKAMBARAM KESAVULU REDDY
6	**StreamSVC: A New Approach To Cluster Large And High-Dimensional Data Streams**
	Mr. Hasan Saberi
7	**Gaining Competitive Advantage on the Basis of Data Warehousing and Data Architecture**
	Dr. Virivada Venakata Raghava Raman
8	**An Intelligent Mechanism for GIS Contract Automation**
	Mr. Muhammad Shaheen
ICDMKE II [第二場次]	
主持人：博士 DYAH ERNY HERWINDIATI	
1	**Recommendation System Using Information Needs Radar Model**
	Mr. Cho-Wei Shih
2	**Towards Ontology Development for Teaching Programming Language**

	Dr. GOPINATH GANAPATHY
3	**Knowledge Management via Facebook: Building a Framework for Knowledge Management on a Social Network by Aligning Business, IT and Knowledge Management**
	Mr. Satidchoke Phosaard
4	**Extraction of Method Signatures from Ontology Towards Reusability for the Given System Requirement Specification**
	Prof. S. SAGAYA RAJ
5	**Achieving High Recall and Precision with HTLM Documents: An Innovation Approach in Information Retrieval**
	Mr. AMMAR AL DALLAL
6	**A Fuzzy Clustering Approach to Filter Spam E-Mail**
	Miss Nehaya Mohammad
7	**Hierarchical Sequence Clustering Algorithm for Data Mining**
	Prof. V UMADEVI CHEZHIAN
8	**Applying Weighted KNN to Word Sense Disambiguation**
	Dr. S.M. Fakhrahmad
9	**Simultaneous Spectrophotometric and Chemometric Determination of Oleic, Linoleic, and Linolenic Fatty Acids in Vegetable Oils**
	Mr. Gerard Dumancas
10	**Chinese Researcher Profile Annotation Based on Conditional Random Fields with Semantic Rules**
	Dr. Jungang Xu
資訊安全和網路工程 ICISIE I [第一場次]	
主持人：博士 Ping An Wang	
1	**An Agent Subscription-based Model for CRLs Validation**
	Mr. Sekpon Juntapremjitt
2	**An Interactive Firewall Simulator for Information Assurance Education**
	Prof. Huiming Yu
3	**Models for Recommender Systems in Web Usage Mining Based on User Ratings**
	Prof. ARUNESH K.

4	**Dynamic Bandwidth Shaping Algorithm for Internet Traffic Sharing Environments** Mr. Phongphan Danphitsanuphan
5	**Broker Architecture for Quality of Service** Prof. Dinesh Saini
6	**New Approach in Creating of Block Ciphers Based on Wavelet Decomposition of Splines** Dr. Alla Levina
ICISIE II [第二場次]	
主持人：教授 ARUNESH K	
1	**Uncertainties of Online Phishing Risks and Consumer Decision Making in B2C E-Commerce** Dr. Ping An Wang
2	**Security and Vulnerability Issues in University Networks** Prof. Dinesh Saini
3	**The SOAP Protocol Used for Building and Testing Web Services** Dr. Pirnau Mironela
4	**Securing Cloud System via Internal Control Management** Mr. Ashwin Alfred Pinto
5	**A Proposed Framework to Prevent Financial Fraud through ATM Card Cloning** Miss Divya Singh
6	**A Novel Encryption System using Layered Cellular Automata** Miss Bangaru Bhavya Balanagu
7	**Self-adaptability in Secure Embedded Systems: an Energy-Performance Trade-off** Mr. Nicolae Botezatu
統計計算和數據工程 ICCSDE I [第一場次]	
主持人：博士 Ilya Gluhovsky	
1	**Initial Values in Estimation Procedures for State Space Models (SSMs)** Dr. Raed Alzghool

2	**On Statistic for a New Test of Discordancy in Circular Data** Dr. ABDUL GHAPOR HUSSIN
3	**Detection of Glioma (Tumor) Growth by Advanced Diameter Technique Using MRI Data** Mrs. Karpagam S
	……
ICCSDE II [第二場次]	
主持人：博士 ***Raed Alzghool***	
1	**Multinomial Least Angle Regression with Application to Web Personal ization** Dr. Ilya Gluhovsky
2	**Clickstream Data Warehousing for Web Crawlers Profiling** Prof. Orlando Belo
	……

參、會議內容摘錄

主　　題：

一. 建造儲存網站點選資料流(Clickstream Data)紀錄之資料倉儲，以分析及察覺出「網路爬蟲(Web Crawlers)」行爲

二. SOAP 協定(Simple Object Access Protocol)用於建造和測試 Web 服務(Web Service)

三. 經由內部控制管理以建立安全的雲端系統

四. 基於網站使用評價探勘(Web Usage Mining)的推薦系統(Recommender System；RS)模型

五. 一個模糊的叢集處理來過濾垃圾郵件電子郵件

六. 使用 Facebook 來做知識管理─利用社群網路建立知識管理架構，來調整企業、IT 的知識管理

七. 使用代理訂閱爲基礎的模型，以更新與驗證憑證廢止清冊(CRL)

一. 建造儲存網站點選資料流(Clickstream Data)紀錄之資料倉儲，以分析及察覺出「網路爬蟲(Web Crawlers)」行爲

搜尋引擎是網際網路興起後最常被使用的工具之一，其主要技術包含前端的全文檢索與後端的網頁蒐集兩類， Google 與 Yahoo 等網站搜尋引擎後端的強大的網頁蒐集程式，可以將全世界各處的網頁通通捉取回去儲存以便提供搜尋之用，這個程式就稱爲「爬蟲(Crawler)」，也有人索性將其稱爲「蜘蛛(Spider)」,因爲它就好像在網路上爬來爬去的蜘蛛一樣,到處扒網頁資料回家放。

也因爲網路爬蟲 (Web Crawler)致力目標於「把別人家的資料庫都變成自家的資料庫」,在網路世界中早已算是個貪婪又古老的技術，再加上技術門檻不算高，只要能送出 HTTP Request 再加上正規表示法(Regular Expression)將網頁原始碼中的資訊解析出來，就算是具備基本的爬蟲功能，所以養出一隻在網路上

到處蒐集的爬蟲並不需花費太久的時間，而且網路爬蟲再爬行網站的行爲與一般使用者瀏覽網頁時的方式是十分類似的，這對網站管理者不啻是個噩夢，可能面臨如網站商業資料外流、伺服器負載過重…等無法掌控的情形。

本篇主題提出一種研究方法，藉由使用網路伺服器日誌檔案，建構一記錄有「網站點選資料流數據」的資料倉儲系統，並探勘及分析這些大量的點選資料流，據以產生可辨識出瀏覽網站者身份的特徵資訊，使網站管理者易於區分是一般網站使用者或者網路爬蟲(Web Crawler)來瀏覽網站的。這種研究方法除了可用來強化網站的資安保護機制，如預防任何未經授權的內容存取、防止可能會造成網站效能下降的超載使用等，另外的助益則是能從網站用戶所點選資料的瀏覽行爲中，挖掘出更有價值的資訊，如針對電子商務網站，可透過使用所挖掘出的有價值資訊，改進網站的設計、提高客戶服務的品質、開發新的客戶及優化目標市場的行銷，藉以實現商業智慧並幫助提升業績。

因爲在分析點選資料流數據的過程中，對於分析結果會採取適當的行動，以致此研究方法的過程可以被看作是一個不停自我學習及回饋的循環週期，此過程包括三個主要元件，即：(1)點選資料流處理(ETL)元件：處理網站伺服器日誌內容時，提供準確的差異化分析以察覺偵測出網路爬蟲; (2)機器學習元件：研究網站用戶行爲，建立可辨識網路爬蟲的模式規則，以及(3)封鎖遏制元件：一旦網站伺服器的效能或安全性受到影響時，對監測出確爲網站爬蟲的瀏覽加以限制其繼續的到訪。

此研究方法並使用「WEKA」機器學習演算法軟體及「NATURA 網站」6個月 Web 伺服器日誌數據，進行實驗性質的資料探勘及分析驗證。其中實驗對象 NATURA 是一個有關自然語言處理(NLP)的研究專案，成立的網站(http://natura.di.uminho.pt)雖不具商業性質，但因其多樣性的內容仍吸引來自四面八方的不同的網路用戶造訪，網站伺服器日誌涵蓋有大量及多樣化的點選資料流數據可供研究。

二. SOAP 協定(Simple Object Access Protocol)用於建造和測試 Web 服務 (Web Service)

源自於 2001 年就開始被微軟炒作的分散式元件概念─Web 服務─它是一種應用軟體元件，使用標準 W3C 通訊協定及開放式標準的資料格式(例如：SOAP、XML 及 HTTP 等)，透過非同步訊息(Asynchronous Messages)或遠端程序的呼叫(Remote Procedure Call；RPC)，來提供客戶端服務。

換言之，Web 服務便是以 XML 爲資料格式，將軟體服務封裝成一堆可以在遠端呼叫的函數，並將服務註冊及公布在網路上(UDDI)，讓客戶端可藉由網路來搜尋及使用 Web 服務，透過標準的網路通訊協定(HTTP)及 XML 標準格式的底層通訊協定(SOAP 與 WSDL)來進行資料傳送、處理溝通與辨認程式碼等動作。可以看到，WSDL、SOAP 與 UDDI 皆是用 XML 方法來描述，接下來將詳細說明，這一些基礎的內容：

- SOAP (Simple Object Access Protocol)

是一種標準化的 W3C 通訊協定，它的出現是爲了簡化網頁伺服器（Web Server）由 XML 數據庫提取資料時，不需再花時間去格式化頁面，並且可以讓不同應用程式之間透過 HTTP 通訊協定，相互以 XML 格式交換資料，使其與程式語言、作業平台和硬體環境無關。左圖顯示 SOAP 訊息的結構，每一個 SOAP 訊息都是被一個稱爲 SOAP Envelope(信封)的根元素所包覆，包覆的內容中可放置多組 SOAP HEAD(表頭)及 SOAP BODY(內容

SOAP message structure

本體)，訊息本身就是一份 XML 文件。

- XML（eXtensible Markup Language）由 W3C 組織所制定及推薦、是用來描述資料的一種標記語言，使用者可以自己去定義標籤(Tag)的意義，是一套資料儲存工具，可以用來建立包含結構化格式資料的文件。

```
<?xml version="1.0" encoding="UTF -8"?>
<!-- Document that organizes the information upon the
person-computer -->
<Person sex='F'>
<Surname>Nicolae</ Surname >
<Name>John </name>or
<Datan>11.03.2010.</Datan>
<!—The address of the person is a composed element-->
<Address>
<Street number="1" block of flats="1">Republicii </Street>
<Entrance>A</Entrance>
<Apartment>1</Apartment>
<City>Bucharest </Bucharest>
<District>. </District>
<Country>Romania</Country>
</Address>
<!—A person may have more than a single contact
information-->
<Contact Info >
<Contact type="telephone">..... </Contact>
<Contact type="email">..... </Contact>
</Contact Info></Person>
```

XML document

- WSDL(WEB SERVICES DESCRIPTION LANGUAGE)

  爲用來描述 Web Services 細節的 XML 格式語言，主要的用途是讓 Web 服務能以一種標準方法來描述自己擁有哪些能力，以便和各種用戶端應用程式相互整合，附檔名爲.WSDL。亦即在用戶端與伺服器端準備以 SOAP 對話前，一定要先知道「有哪些服務可使用？」、「服務需要哪些參數？」，顧名思義，WSDL 就是用來處理這些問題。具有 WSDL 文件的 Web 服務，可以讓客戶端很容易撰寫程式去呼叫它。客戶端撰寫的呼叫程式必須包括：呼叫該 Web 服務的 URI、對應的函數名稱及輸出入參數。

- UDDI（Universal Description, Discovery and Integration）

  指的是一種有關於 Web Services 的目錄註冊服務，其架構亦是以 XML 爲基礎，UDDI 主要的目的爲讓 Web 服務提供者，透過 UDDI 告知其他人提供者有提供 Web 服務，因此 UDDI 的功能類似電話簿，目的是要快速告知服務使用者，有哪些的 Web 服務可以使用。

提供客戶端使用的 Web 服務，本身應滿足一些基本的要求，包括(1)易於被

重覆使用及擴展功能。(2)提供一種跨越平台、作業系統及程式語言的服務使用。(3)允許選用適當的傳輸協定，在網路中傳輸資料及訊息，最常被企業採用的爲不被防火牆所阻擋的 HTTP 協定。(4)多用 WSDL 語言描述 Web 服務，讓客戶端程式易於開發應用。(5)以 UDDI 目錄註冊 Web 服務，讓 Web 服務易於在網路中被搜尋及加以應用。

這篇主題除了探討 Web 服務的理論及應用外，並建議 Web 服務提供者在網路上發布 Web 服務前，可以先行在 Web 伺服器中，模擬大量使用者、在同一時間執行測試此 Web 服務，以避免和更正可能出現的錯誤，例如：使用 Google 的 Web 服務 API 執行模擬測試。因網路上有許多像 Google 這種現成的 Web 服務壓力測試工具可供測試，這些工具實際模擬 Web 服務被使用者端呼叫的程序，以 SOAP 請求，來發送 XML 格式的數據，並也取回同爲 XML 格式的測試結果。

但以 SOAP 爲核心的 Web 服務並非十全十美，若想提供一個 Web 服務，需安裝若干很龐大的軟體環境，如.NET Framework，以產生與伺服器溝通的 SOAP(底層需求很高)。而受限於 XML 格式長度的影響，SOAP 訊息的通訊、解譯，執行 Web 服務的方法、編譯，回傳再於客戶端解譯，大量的網路傳輸量所需耗費的時間，使得其傳輸量重、執行較爲緩慢的問題浮上檯面(Overhead 很大)。

另一種非標準的 Web 服務──被稱爲 REST(全名是 REpresentional State Transfer)──於焉誕生，它藉由指定 URL 來存取資源，存取方式包含了查詢、新增、修改以及刪除，正好對應到 HTTP Request 的 GET、POST、PUT 、DELETE 等類型的請求方法，是以資源(Resource)而非行動(Action)爲核心來提供 Web 服務。舉販售書籍或光碟的 Amazon.com 網站爲例，所提供的 Web 服務若採用 SOAP 的機制，需先開發完成 BookService 或 DVDService 這樣的服務，開發的內容則包含這 2 種服務的個別的、一連串的、繁複的可供遠端呼叫的函數及

WSDL 描述文件，才能讓客戶端使用 http://amazonlite.com/BookService?wsdl 或 http://amazonlite.com/DvdService?wsdl 這樣的 URL 來訂購，但若採用 REST 的機制，則僅需簡單的借用 HTTP Request 的 GET、POST、PUT、DELETE 等類型的方法，以 http://amazonlite.com/Books 或 http://amazonlite.com/Dvds 的 URL 客戶端就可進行訂購，因不需要再以 SOAP 這樣複雜的機制來開發一堆服務或函數程式，亦不需安裝一堆不必要的軟體環境，省略了開發端網站許多額外的處理，所以當 Amazon.com 同時提供了 REST 及 SOAP 的 2 套 Web 服務，發現使用 REST 來存取的流量比 SOAP 高出許多。

不管開發端網站或 Web 服務提供者，是選擇使用 Web 服務的哪一種類型架構，來提供客戶端使用和實作，例如 Google 選擇 SOAP 類型的服務，而 Yahoo 則是使用 REST 類型，每一架構都有其既有的優勢及缺點，藉由分析不同架構的特性去選擇應用，比如 SOAP 架構發展較久，又有來自 WS-I (Web Service Interoperability)組織的規範制訂，不論在基礎建設或安全性規範均較爲完備，且有許多大型應用程式會依賴它進行資料交換，是以它就十分適合應用在企業資訊系統間交換資料或做流程操作，可以 SOAP 與 XML 確保資料格式的正確與完整性，而 XML-Sig 規格則負責確保訊息的安全性。另一方面，REST API 就十分適合前端應用程式與服務之間的資料交換以及與使用者互動的部份，因爲它提供了相對簡便的方式來達到資料交換的效果，而且對資料的內容並無多加限制或規範，有效率的降低開發人員的學習門檻，對 Web 服務提供者以及第三方開發者而言可能是一種更好的選擇。

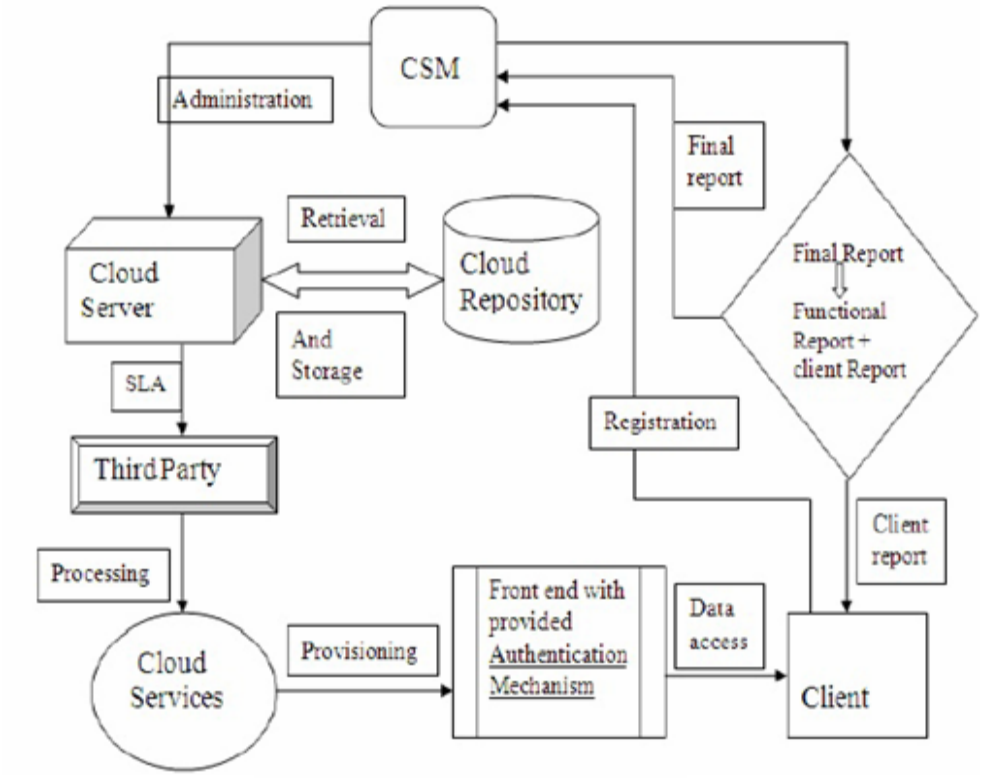
三. 經由內部控制管理以建立安全的雲端系統

雲端運算的興起，讓我們可以利用在遠方的雲端伺服器上存取自己的資料數據，這意味著使用者可以隨時將資料儲存在雲端伺服器的資料庫中，亦可以在需要時隨時取出使用。但這也衍生出使用者的一些疑問，比如儲存在雲端的數據是否有足夠的安全性保護、由誰來處理我們儲存的數據、被處理的程序又

是如何運作的，還有什麼樣的防護控制被應用以保全我們的數據，許許多多的疑問仍有待解答。

也因爲在雲端系統總是有一些百廢待興、不完整、不安全或未被建立的安全顧慮，比如每當有來自客戶端的造訪行爲時，數據存取應該適當的被監控，同時爲了確保敏感資料的機密性，應該有經由內部檢查的隱私權控管政策被採行等，是故在雲端系統中，數據資料的保護一直是一個不斷被討論的議題。雖然已經有大量的研究工作投入在這個領域，但仍有些爭議仍未被討論過，包括(1)客戶端應該有一個可以查看資料全程被存取的記錄，以確認資料被合法的使用。(2)每個業務需求都應被內部控管，並應不時檢視適用性聲明(Statement of Applicability；SOA)的控制內容。

根基於此，在本篇主題中提議一種被命名爲「Cloud Management Model」的模式(如下圖)，它的成員組成包括了(1)一種新被提出的實體——稱爲「雲端安全管理機制」(Cloud security manager；CSM)——它的責任爲管理雲端系統，每月定期產生監控報表，報表的內容除了記錄每一個客戶端的造訪行爲外，還有



Cloud Management Model

監控是否有違反規定及政策的事件發生，並發佈制定有哪些內部控制(2)雲端伺服器(3)第三方合作廠商(Third party)(4)雲端儲存裝置(Cloud repository) (5)客戶端(Client)。另外它的運作方式顯示如下圖，並解釋如後述。

在這模式中，CSM 被規劃指定爲權威機構，全權負責監測雲端系統的所有活動：「管理雲端伺服器；負責定義內部控制和懲罰；確保客戶端在雲端資訊庫所儲存的資料；在功能性報告及客戶端報告發布前先行審查並簽署；與已認可的第三方簽署一項服務水平協議(Service Level Agreements；SLA)，該協議明確規範有第三方應負的職責和法律責任」。並針對每個已登記的客戶端，在使用雲端服務時，需經由前端的認證機制驗證後，再由第三方提供協助以滿足客戶端的要求。最後提出內部控制矩陣(INTERNAL CONTROL MATRIX)，它是一個關鍵的功能，用以觀察監控整個雲端系統，規劃實施在不同的級別：(1)雲端的政策管理(Cloud governance)(2)雲端的應用(Cloud application) (3)第三方團體(Third party association)(4)實施數據保護法案(Implementing data protection act；DPA)。這個矩陣提供足夠且有效的內部控制，來協助並引導 CSM 針對整體的雲端系統，執行監視系統、客戶和第三方的管理工作，同時它也針對客戶端的個人數據資料，促進與實施了 DPA(數據保護法)。

本主題中所建議的「Cloud Management Model」模式及內部控制矩陣較現有的模型更爲有效，作爲 CSM 的作用是一個相當新的概念，其責任是有足夠的能力處理在雲端系統上的漏洞，而依據內部控制矩陣的監控，則提供了絕佳的優勢，因爲涵蓋了所有的控制管理，如：政策管理、第三方責任以及 DPA(數據保護法)。

四. 基於網站使用評價探勘(Web Usage Mining)的推薦系統(Recommender System；RS)模型

在網路世界中，網路使用者對推薦技術可推薦符合他們自身特別的需要及

偏好是相當感興趣的。記載有使用者偏好的自動預測系統，可以從很多使用者處，收集其評價；推薦系統(Recommender System；RS)就是基於網路瀏覽者的行爲和評價，進一步使用統計及知識挖掘的技術，來向網站的用戶提出推薦及建議。Amazon、google、Netflix 和 last.fmh 等電子商務網站，在它們的應用程式中都已嵌入所謂的推薦系統，因此這些網站使用者的滿意度和忠實度都比零售商來得高得許多。

網路使用者推薦引擎和技術在現今是個熱門的技術，以使用者導向的角度分類，推薦系統可概分爲 3 類(如下表)：(1)基於使用者過去個別的行爲，所提出的個人化推薦(Personalized recommendations)。(2)以背景、品味相似的使用者的過去行爲，所提出的同儕式推薦(social recommendations)。(3)依據使用者對網站有興趣的項目，所提出的項目建議(item recommendations)。

Approaches to Recommendation Systems

Usage behavior based Recommendations	
1.	Actual Items, pages (Personalized recommendations)
2.	Related items (item recommendations)
3.	Similar users tastes (social recommendations)

像 Amazon 網站就同時應用到前述的 3 種分類。另外一些商業的網站如 Pandora.com，因是一個提供音樂服務的網站，則專注於項目的分析，基於音樂檔案的架構分析來向使用者推薦歌曲。Strands 這個社交推薦引擎，則是基於相似使用者的社會性行爲分析，可經由電腦、手機或其他可上 internet 的裝置，即時的推薦給使用者，使他們得到他們想要的事物。

另外，在早期研究開發推薦系統時，研究者是嘗試從使用者的個人資料和過去的使用紀錄，以資料挖掘(Data Mining)技術分析使用者的行爲，亦即是以網站使用者的行爲來做導航推薦的，算是一種隱性評比，這被稱爲內容基礎

(Content-based approach)推薦系統。但它會面臨諸如新進的使用者，根本未曾有過使用記錄，如何提供適合的推薦呢？所以研究者嘗試將視野放大到所有的使用者用戶，藉由網站使用者對網頁或服務項目的評價，並納入參酌其他使用者相對應的評價，以得到預估的推薦效果，亦即是以網站使用者評價，來作爲推薦系統的基礎，算是一種顯性評比，這被稱爲協同過濾技術(Collaborative filtering；CF)。圖 1 及 圖 2 顯示推薦系統的行爲、評價和安全。

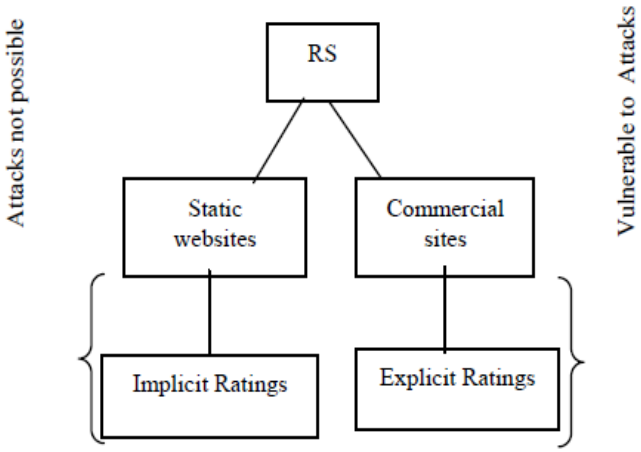


Fig 1. Rating for RS

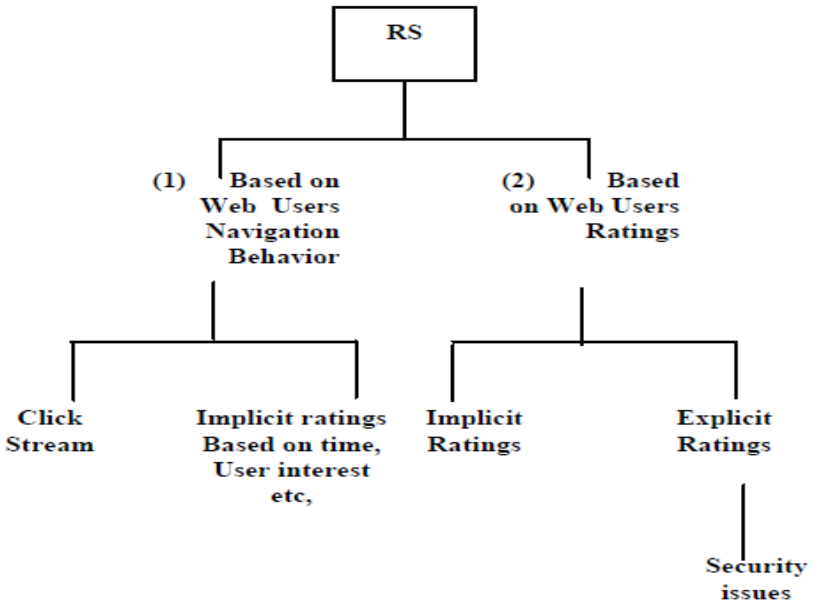


Fig. 2. RS Categories.

系統所推薦項目的描述，包括平均評價、其他使用者的評論、評論家的評比及個人化評價的預測。下圖的範例中，RS 系統對這些項目做評價並提供使用者。聰明的項目推薦需要明確的評價，因爲有太多的產品屬性，如價格，顏色，樣式，品牌…等，會假設對相同的顧客在不同時間做出不同等級的推薦，另外使用者網頁瀏覽的行爲也會改變項目的推薦。因此，當系統有足夠的評價資料時，如喜歡、不喜歡、看法或重要性，演算法將能預測使用者的喜好，並推薦合適的產品或服務。

Amazons' interface for recommendations

在網路採礦領域方面，有很多先進的、複雜的及顯著發展的技術，例如機率式潛在的語義分析(Probabilistic Latent Semantic Analysis；PLSA)、關連法則探勘(Association Rule Mining)、K 鄰演算法(K-Nearest Neighbor algorithms)、強健的協同過濾 (Robust Collaborative Filtering)、K-Means 分群演算法(K-Means clustering)及矩陣因素分解(Matrix Factorization)等，是近來常被研究人員應用於網站使用採礦(Web Usage Mining)的技術。

其中協同過濾技術因給予最貼近使用者需求的推薦技術，故時至今日，該技術已經廣泛地被網站推薦系統所採用。但傳統的協同過濾技術十分容易受到攻擊，許多不正確的評價，很容易就損害及扭曲了預測的品質，再加上大多數推薦技術是根基於模糊或不明確的網路瀏覽者的行爲和評價。是故，爲了減少不正確的評價，而影響了預測品質的這些問題，本篇主題提出 2 種獨特的推薦模型：階層測量推薦(Rank measure Recommendations；RANK-RECO)及測試測

量推薦(Testing measure Recommendations；TEST-RECO)，第一種方法 RANK-RECO 是建構在網路使用者的評價上，將相似愛好使用者的評價，轉換計算出多組基於階層關連性的建議，以提供高品質的推薦，給新的使用者。又受限於使用者的選擇，導致網站中某些項目的評價資料的缺乏，這造成在評價系統中只有稀少的實體被提供作爲預測，於是第二種方法 TEST-RECO 的推薦模型被提出，它預測出的有效的使用者喜好，是根據「相依 T 檢定(paried-t-test)」的統計工具去預測出來的，它的主要優勢是，推薦給新使用者的評價，不必受現存系統中的評價資料所限制及影響。並針對這 2 種模型，透過使用真實世界的應用資料—經由 Grouplens 網站下載有關 945 位使用者對 1682 部電影的大量評價的資料集合—來進行實驗，以評估它們的效力。實驗證實 RANK-RECO 及 TEST-RECO 的模型提供了更好的預測的準確性。

使用者對項目和網頁的評價，已應用網站視爲提供服務的標準慣例，而這些慣例可支援透過有價值的獨立的訊息，爲網路使用者產生建議，以供網路使用者和客戶來做採購的決策。因此，被提議的 RANK 和 TEST 模型，可支援網路使用者選擇他們所喜好的產品或網頁。本文提出的模型因尚在建造中，目前只有初步的結果，但希望它將來的運用能延伸到更大的範圍，以改善預測的準確性及靈敏度。

五. 一個模糊的叢集處理來過濾垃圾郵件電子郵件

垃圾郵件或者不需要的商業電子郵件，近年來已經成爲一個日益嚴重的問題，估計約有 70%的電子郵件流量是垃圾郵件(如下圖範例)。垃圾電子郵件，通常是由一大堆無法辨識的

Spam on the rise.
(Credit: MessageLabs)

Spam as a percentage of all email traffic

電子郵件帳號所發送或者是來自於商業廣告信函，經常發送給郵箱很多不需要的電子郵件訊息。因爲垃圾郵件常將郵件收件箱塞滿，不管它是已被收到郵箱內或是想預防它被郵箱收下，都要花相當多的時間和努力去刪除它。由於垃圾郵件的問題不斷的衍生，這已經成爲每個電子郵箱用戶都相當煩惱的事，估計平均每人每年要花 10 個工作日去應付及解決垃圾郵件。

即使目前有一個最佳的解決垃圾郵件問題的方法，數千名發送垃圾郵件的攻擊者也會尋找新方法戰勝它。正是因爲垃圾郵件發送者的技術不斷的精進，如何建造一個可持續學習的垃圾郵件過濾器，已成爲一個相當活躍的研究領域。大部分目前的研究都集中在使用數據挖掘的方法，來解決垃圾郵件的過濾。在這篇文章裡，使用模糊的叢集演算法(Fuzzy C-Means)來建造一個垃圾郵件過濾器。被提議使用的 Fuzzy 已被引用測試過—藉由在垃圾郵件手法收集來的、不同的實際用戶的電子郵件數據集合，並以資料採礦方式來研究如何過濾垃圾郵件，目的是評估當前狀態並且提議出一套解決方法。作者提出一個已實施的系統，它是根基於很強的區分垃圾郵件的功能和解決問題的一種分類技術。

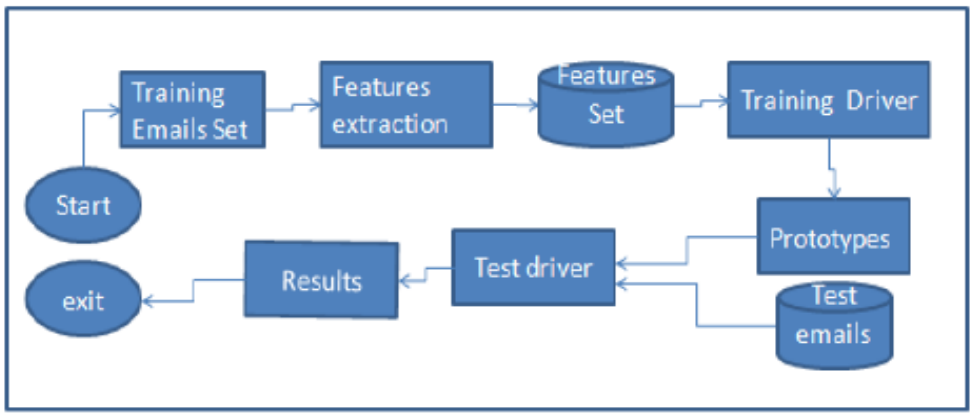
現存的郵件過濾器是利用白名單、黑名單、灰名單來過濾垃圾郵件。假如確認郵件來自已知的垃圾發送者，這封郵件會被標爲垃圾郵件。假如是來自使用者允許的位址，則會進入使用者的郵箱。用這種方法的困難是在使用者的負擔可能相當大。以規則基礎的過濾器，使用者必須要預先訂定垃圾郵件規則，但過濾器可能將垃圾郵件誤爲合法郵件，也有可能將合法郵件誤爲垃圾郵件，所以需要頻繁地更新清單。

以內容基礎的過濾器是基於能定出一套規則、範例、特徵的前提下，能描繪出垃圾郵件的分級及門檻。開始安裝這類過濾器是集中式的，但是跟垃圾郵件流量相比，其流量只佔 50%。另外機器學習技術是更多樣化及彈性的，如決策樹是依據先前收集的資料來判別垃圾郵件或合法郵件；貝氏網路則是目前更大眾化的反垃圾郵件的技術，但是它們都有相同的困難：在於如何衡量和依賴

許多特徵來做判斷。

本論文利用模糊叢集(fuzzy clustering)和文字採礦(text mining)這種種反垃圾郵件的技術，來過濾垃圾郵件。模糊叢集(fuzzy clustering)是可用及容易更新的技術，如果進來的每封電子郵件都可被當作資料庫的一部分，垃圾郵件特徵將被發現並且自動學習適應，如此即可判定未來的電子郵件是否爲合法，且不會有需重新計算所導致的較大成本花費。

模糊叢集模型(Fuzzy clustering Model)包括：特徵截取 (features extration)、訓練(training)、測試(testing)等 3 個主要階段，且在特徵截取階段之前，必須要找到可用的資料來做訓練及測試。下列是垃圾郵件過濾模式：



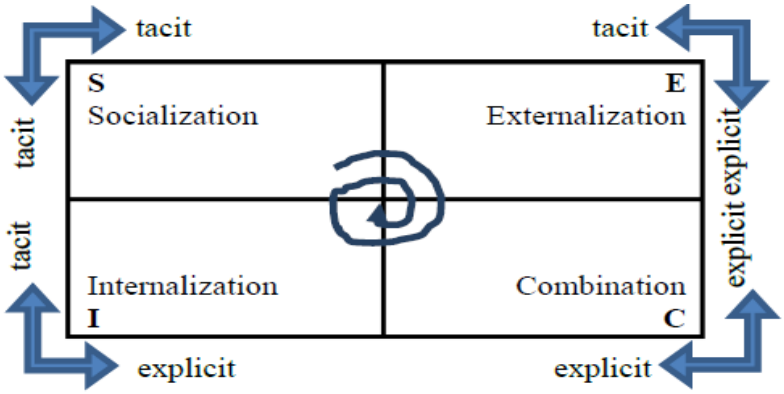
訂定特徵截取的前置作業中，主要面臨的困難是如何訂定特徵；本文提供初步的特徵清單，即利用公開的垃圾信件資料庫(SpamAssassin Spam Corpus)，它包含有 3,000 多封已被註記爲垃圾郵件或合法郵件的、來自於不同使用者的郵件資料，來放入模糊的叢集演算法中運算，透過擷取這些資料庫中的資料，蒐集更正確及可擴充的特徵資訊，以提供精確的垃圾郵件偵測。

垃圾郵件過濾的困難點是在於垃圾郵件一直被不斷的演變及更新，以致過濾器被要求必需因應這樣的趨勢，以具有高度預測垃圾郵件的準確性。本文所提出透過模糊叢集模型(Fuzzy clustering Model) ，經由實驗的結果，對垃圾郵件過濾的成效相當顯著，是可被當作一種垃圾郵件過濾器的工具，但目前尚在實驗階段中。

六. 使用 Facebook 來做知識管理—利用社群網路建立知識管理架構，來調整企業、IT 的知識管理

知識管理(KM)多年來已經被廣泛地認爲是能爲組織，創造價值的一個關鍵成功因素，因它能持續的改進一個組織處理知識的能力。KM 可分爲 4 個主要的階段變化：(1)知識創造(2)知識獲得和保存(3)知識分享(4)知識應用，這個概念是圍繞著 KM 形成一個循環，在個人(individual)、小組(team)、組織內(organizational)、組織間(inter-organizational)之間。

先說明本文所提出及採用的 SECI 模型，又稱爲「知識螺旋」，它描述了四種知識轉換的模式—「Socialization (社會化)」、「Externalization (外部明示) 」、「Combination (彙總組合)」和「Internalization (內部陞華)」(如下圖)，最初的原型是由野中鬱次郎(Ikujiro Nonaka)和竹內弘高(Hirotaka Takeuchi)於 1995 年在他們合作的《創新求勝(The Knowledge-Creating Company)》一書中提出，將企業

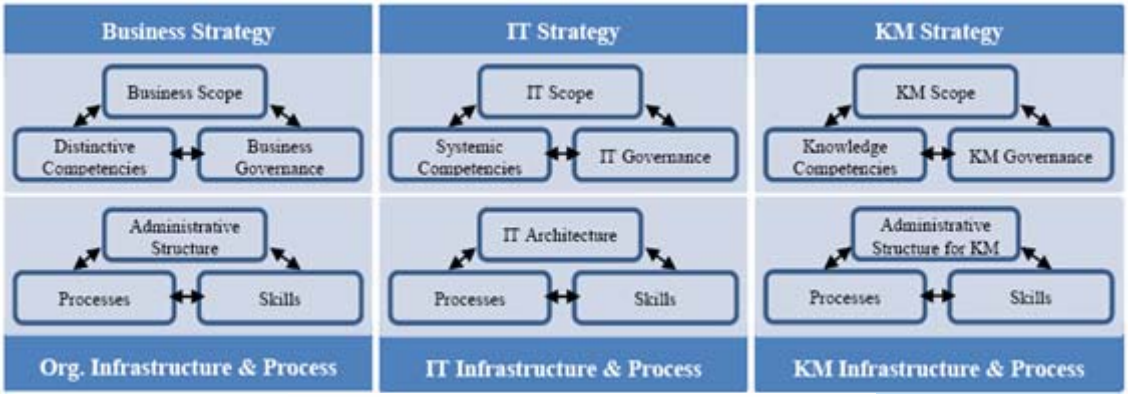


SECI Model

知識劃分爲「Tacit Knowledge(隱性知識)」和「Explicit Knowledge (顯性知識)」兩類；所謂隱性知識包括信仰、隱喻、直覺、思維模式和所謂的「訣竅」；而顯性知識則是可以用規範和系統化的方式進行傳播，又稱爲可本文化的知識。在企業創新活動的過程中隱性知識和顯性知識二者之間互相作用、互相轉化，知識轉化的過程實際上就是知識創造的過程。

TABLE I
BUSINESS-IT-KM STRATEGIC ALIGNMENT

Business Strategy	IT Strategy	KM Strategy
Business Scope • Objectives • People: Project Manager, Staffs, Exhibitors, Participants, etc. • Budgets, Venue, Plan and Schedule	**IT Scope** • IT Roles • IT Staffs • IT Budget • IT Plan and Schedule	**KM Scope** • KM Roles, KM Staffs • People: Project Manager, Staffs, Exhibitors, Participants, etc. • KM Budget, Plan and Schedule
Business Competencies • Event Organization Skills • Project Management, Communication • Promotion	**Systematic Competencies** • IT Support • Technical	**Knowledge Competencies** • Socialization, Externalization, Combination, Integration • Tacit & Explicit Knowledge
Business Governance • Operational-Level Governance Checklist	**IT Governance** • Operational-Level Governance Checklist	**KM Governance** • Operational-Level Governance Checklist



Strategic Alignment Model for Business, KM and IT.

本文有鑑於 KM 因缺乏與企業的策略相銜接，所引發企業內部對 KM 投資的爭論，提出一套包含有企業(Business)、資訊管理(IT)和知識管理(KM)的策略調整模型(顯示如下圖) ，這套模型已經被廣泛地證明及採用，可透過將 IT 策略性模型，調整延伸爲 KM 的策略性模型來應用，詳細模型內容則列於 Table I。KM 的架構可以被應在所有的企業流程方面，如用於持續性的流程改善、日常的工作或其他的企業架構，如客戶關係管理(CRM) 、企業創新….等等，都已經證明是成功利用 KM 的結果。

此外，新興資訊技術提供了人們溝通方式的改變，也影響了企業經營的模式，並且開啓探索新的商業模式的機會。以目前網路通訊中最火熱的、最大的社群網 Facebook，根據統計在 2010 年底會有超過 5 億用戶註冊使用。因此，在本文的研究過程中，調整模型將 Facebook 納入作爲一種 KM 工具的研究，亦即讓 Facebook 成爲 KM 的一種選擇，以證明此架構的可行性。這項研究列出了

3 個主要的目標：(1)建造一種企業組織在 KM 應用的架構。(2)提供一個關於如何透過提議的架構在 KM 裡使用 Facebook 的詳細案例。(3)初步的評估此架構如何在真實世界中作業。

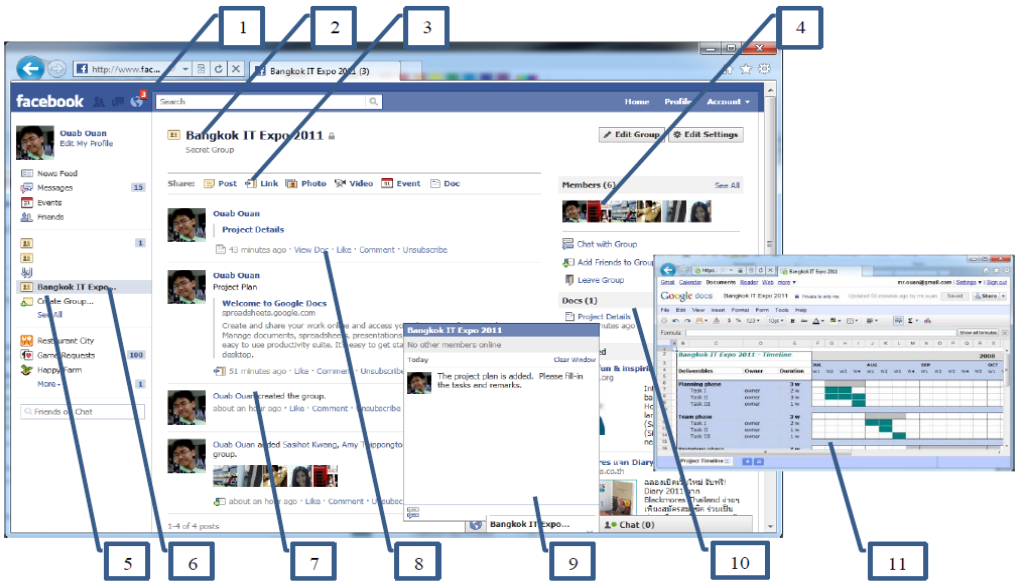
在下表 II 的 IT 策略中之 IT Architecture and Resources 列出那些工具是必需的，專案的 Facebook 群組被個別建立以供內部使用，KM 的導入不會影響企業的日常工作也不會影響公司組織，只有 KM 的員工要額外負擔 KM 的管理職責。在 IT processes 方面，任何 email 通訊方式將會被移至 Facebook，透過 Facebook 來交流。任何訊息需要溝通，寄件人可利用 Facebook 佈告欄分享的方式與整個 team 來交流溝通。也可透過 Facebook 傳遞媒體檔、文件、附加檔。這明確的說明如何在系統內獲得知識—問題解決和最佳實踐的方式，這些知識應該被收集和儲存在知識庫中。在 Skills 方面，作者透過下列的圖例來描繪出如何經由 facebook 群組及 Google doc 平台來做爲專案管理及溝通的工具。

TABLE II
BUSINESS-IT-KM PROCESS AND INFRASTRUCTURE ALIGNMENT

Business Strategy	IT Strategy	KM Strategy
Business Administrative Structure and Recourses • People: Project Manager, Staffs, Exhibitors, Participants, etc.	IT Architecture and Resources • Event Website • Event Facebook/Project Facebook • Google Docs Documents: Spreadsheets • IT Staffs	Administrative Structure and Resources for KM • Project Manager • Staffs • Knowledge Gathering Staffs
Business Processes • Project Planning • Promotion • Booth Sales • Venue Preparation • Event Activities • Evaluations	IT Processes • Using Facebook Group for Privacy and Immediate Notifications • Sending Facebook Message for Personal Questions, Posting on Facebook Wall Post for Group Questions • Project Tracking via Google Docs Spreadsheets • Evaluations	KM Processes • Information and Issues Related to Each Project Tasks – Explicit Knowledge Seamlessly Captured • Issues Compilation on Walls, Message, and Boxes – Implicit & Explicit Knowledge Captured • Evaluations • SECI
Skills • Project Management • Communication • Interpersonal • Leadership	Skills • Support • Technical • Internet • Application Software	Skills • SECI: Socialization, Externalization, Combination, Integration • Explicit/Tacit Knowledge of the Project • Willingness to Learn

Facebook 能讓使用者能建立實體的網頁(如個人、公司)，或者能建立讓人們想要連接上的一粉絲網頁。在這個研究案例中，首先，群組類型應該被使用，粉絲頁或正式頁供外部使用，專案管理則與 KM 整合使用。專案成員必須有個

別的 Facebook 帳號，專案經理或者技術人員再於 Facebook 上建立一群組，群



組的創造者應該建立成機密和私人的群組，並指派只有群組成員能瀏覽並且參加討論，以利內部的管理，當有任何活動在群組頁發生時，在 Facebook 左上方會顯示通知訊息，這非常方便，組員能透過訊息通知了解目前狀況 (如上圖之 No.1)，而在 Member 處可以出現群組成員(如上圖之 No.4)，並可在網頁上展示專案計劃和文件(如上圖之 No.8, No.10) 。如需專案詳細資料，需要額外的工具，可無縫結合到 Facebook 上，如透過 Google document 能附上一份文件和一張工作表格以利專案進度追蹤(如上圖之 No.11)，並允許組員不斷更新其內容。附加檔案能被群組成員容易的由右方的網頁取得(如上圖之 No.10)。

此外專案成員必須在專案上不斷更新進度，以利專案進度追蹤，成員每天可以總結工作並發佈到 Facebook 上的意見公佈，如果有相關的資訊或有用的知識，可透過媒介來分享(如上圖之 No.3) ，如果是私人訊息，成員能透過訊息功能來交流，訊息功能亦可以發送訊息到電子郵件，也能與客戶進行訊息交流並且能方便的整合到專案網頁中。簡單來說， Facebook 群組頁和 Google 文件功能，能整合電子郵件、訊息傳遞、網頁管理、內容管理、事件行事曆、辦公室應用….等功能，能符合企業內部作業需求。這個與 Facebook 整合的 KM 架構本身，可依據 KM 基本原則來評估進行。首先，在 KM 中重要的人、組織、文

化和過程等要素，必須被識別及正確處理，並被評估如何快速地整合到現有的組織架構中。再則，應用 Facebook 連續分享的風格可鼓勵群組組員追蹤問題的後續進展，Facebook 和此架構正提供了一個平台，以促使 KM 利用 SECI 模型，讓這個平台可激勵組織的不斷學習，並可達到企業內部知識的轉化—社會化、外部明示、彙總組合、內部陞華的知識創造過程。最後，Facebook 被證明可成功地成爲專案流程中的重要溝通管道。

而 Facebook 的環境是一種令人愉快的工作環境，它不僅促進組織成員專案的交流，也可供組織成員即時分享溝通，且已被驗證可成爲 KM 架構中的一部份，也適用於前述的策略調整模型(Business, IT and KM Strategic Alignment Model)。但目前研究只局限於一項專案管理，將來需要更大規模的投入，才能更進一步的研究如何支援整個模型架構。

七. 使用代理訂閱爲基礎的模型，以更新與驗證憑證廢止清冊(CRL)

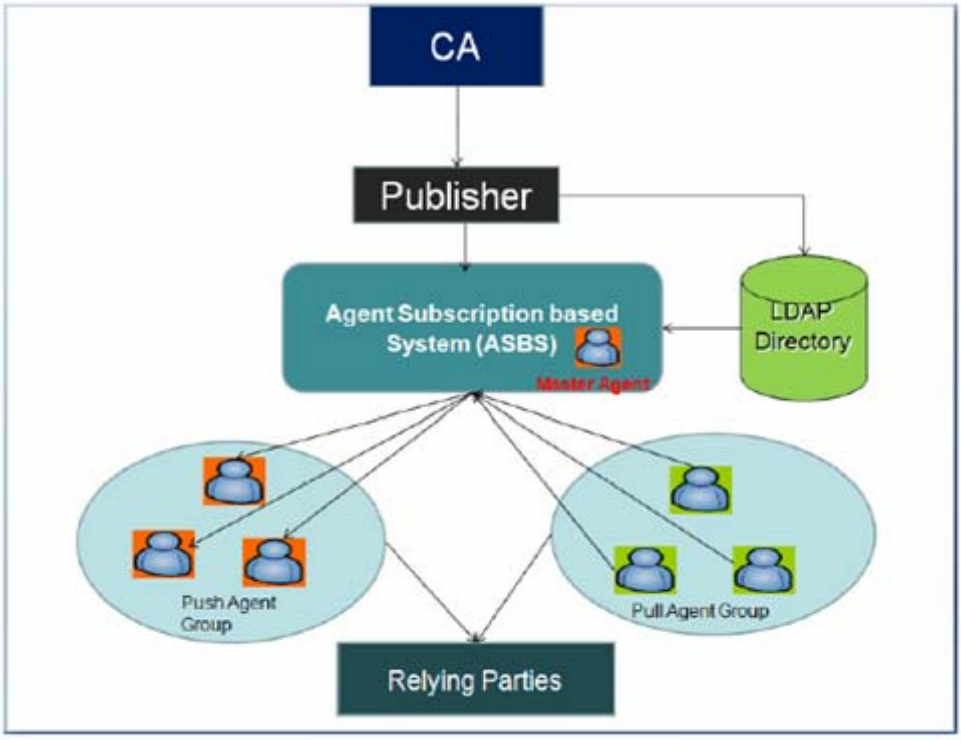
一個可靠性的「公開金鑰基礎建設(Public-key infrastructure；PKI)」系統，有賴於具公信力第三方身分的憑證管理中心(Certification Authorities；CA)，利用憑證管理資訊系統，提供及管理憑證的申請、簽發、廢止及產生稽核記錄等服務。已被簽發的憑證狀態可被區分爲有效或者是無效，而所有經 CA 無效的廢止的憑證資料，皆被公佈於 CA 的目錄伺服器中，以供外界隨時查詢。所以當憑證被用來執行交易時，任何系統均應針對該憑證之有效性來進行核對，以確保交易之安全。

憑證廢止機制是一個非常重要的應用程序，最常用的憑證廢止方式就是使用由 CA 發佈的「憑證廢止清冊(Certificate Revocation List；CRL)」，已標準化爲 X.509 標準，並支援 PKI 系統；該 CRL 包含全部無效憑證的列表，定期由 CA 更新及公佈在目錄伺服器中，而通常 CRL 檔案容量龐大，還會增加網路傳輸的負擔，且需要很長的下載時間。另一種憑證廢止方法「差異憑證廢止清冊(Delta CRL)」則是一種特殊的 CRL，包含一個自上次完整 CRL 之後廢止的憑證列表，

Delta CRL 的檔案大小，遠小於完整的 CRL，因此它可以降低冗長 CRL 的下載時間及減少佔用的頻寬。還有一種方式爲符合 IETF RFC 2560 的「線上憑證狀態通訊協定(Online Certificate Status Protocol；OCSP)」，它的訊息溝通是透過 HTTP 通訊協定的 request 及 response 的方法，來線上即時呼叫及回應憑證狀態的查詢，使用者可立即知道憑證的有效性，相較於 CRL 會定期發佈包含全部已經廢止的憑證資訊，OCSP 則只須處理用戶端查詢單一憑證狀態資訊的要求，顯然使用 OCSP 的方式應該要比使用 CRL 更有效率，但因爲所有的 OCSP 在查詢時，每一次當憑藉狀態有變動，仍需要由 CA 簽名，所以它仍面臨憑證的時效性是無法擴展的問題。

大部分解決 CRL 驗證計劃的技術，都集中在處理 CRL 該如何更新的方法，如減少 CRL 大小的 Delta CRL，及另一個將儲存廢止憑證的空間，劃分爲多個分區的模型，當某個憑證的狀態被更動，僅需更新該憑證所在的分區。儘管這些技術都可以解決 CRL 最佳化，以使 CRL 的更新過程執行更爲快速，但主動提供最新的 CRL 給第三方信賴憑證者(Relying Parties)的技術，仍付之闕如。針對這部分的技術，最近終於有 Datniel Kouril 提出的 CRL Push Delivery 方法，它提出信賴憑證者不需單獨再向個別的 CA 徵詢及索取 CRL，而是使用集中的 CRL 分發服務，這個服務先從一個或多個 CA 收集 CRL 數據並主動分配給客戶端；該模型提供了一個類似線上即時更新和發布 CRL 的方式，惟信賴憑證者肩負有——檢查憑證是否有效或是否有新的廢止資料已被發出——的責任。但由於 CRL Push Delivery 方法根基於主動通知的基礎架構，龐大 CRL 更新通知的數據流，也形成了極高的通信成本。

在本文中使用代理訂閱和簽名廣播的技術，提供一種憑證廢止管理系統的建議模式，以大幅減少 CRL 的大小及透過執行即時的 CRL 更新，來努力克服傳統 CRL 和 OCSP 機制的侷限。下圖說明了所提出的框架，主要的組成內容包括：



Proposed ASBS Framework

(1)LDAP 目錄是一個用來存放憑證和由 CA 發行和出版 CRL 的資料庫。(2)發行商(Publisher)是一個系統元件，用來開發彙集 CA 系統調用的所有憑證和 CRL，並將已發佈的 CRL 提供給 LDAP 目錄及代理訂閱的 ASBS 系統使用。(3)代理訂閱爲基礎的系統(ASBS)則是一個用來與 Publisher 及 LDAP 目錄一起運作的元件，包含有 3 種代理機制：Master Agent、Pull Agent 及 Push Agent，並由 Master Agent 做爲主要負責，諸如監控發布 CRL 給信賴憑證者的信息和建構由 CA 授權服務簽名的 CRL，廣播給其他代理商。並負責與 Pull Agent 及 Push Agent 溝通。另外 Pull Agent 則負責檢查 CRL 的變更，並在一定的期間向 ASBS 取回 Delta CRL 以執行更新；而如果 CRL 有任何的更新，Push Agent 則是接收由 ASBS 主動通知的已即時更新的 CRL，因此，Push Agent 總能得到最新的 CRL，以提供信賴憑證者使用。此架構透過使用 Delta CRL 更新的方式，僅將差異變化的 CRL 調用給代理商和信賴憑證者使用，亦即由 CA 發布的 CRL 僅儲存在 LDAP 目錄和 ASBS 中；再由 Master Agent 計算和傳播 Delta CRL 到訂閱的代理商和信賴憑證者處。它可以讓應用程序僅連接到 ASBS，而不是直接連接到 LDAP

目錄，這是因爲需要 Agent 系統執行 Dealta CRL 計算和動態的調整更新等工作。

藉由使用在 Linux 平台上運行的 Open CA 系統及 Java 程式開發出的發行商(Publisher)與 ASBS 系統，並以 MySQL 作爲一個 LDAP 目錄的資料庫，來實驗上述建議的架構，以評估它的功能和性能，且將它與微軟 CA 作一比較。在測試案例中，由 Open CA 及微軟 CA 系統共核發出 50 張憑證，且 CRL 每隔 1 小時被更新及設置它的生命週期爲 3 天。先測試 Open CA 的基本功能，包括憑證核發、發佈 CRL 及憑證管理系統中諸如暫停和廢止的功能，並檢查及確認 CRL 的內容和憑證狀態的更新無誤，這證實了所建議的架構與 Open CA 可以支援正常的憑證管理功能。而另外在 Open CA 及微軟 CA 對照性的功能測試中，則使用電子支付的應用程序進行測試，並觀察 2 組 CA，在 CRL 的檢查更新所花費的時間。結果表明(如下表)，Open CA 的系統在 CRL 的檢查更新上，花費了較短的時間。這是因爲，Open CA 是藉由 Delta CRL 更新的方式，僅將差異變化的 CRL 調用給代理商和信賴憑證者使用，並透過由 ASBS 主動的通知，來執行即時的 CRL 更新。

COMPARING CRL UPDATE BETWEEN OPEN CA WITH ASBS AND MICROSOFT CA

	CRL Update Period (1-5)				
	1	2	3	4	5
Update time of Open CA with ASBS (seconds)	0.22	0.24	0.28	0.32	0.36
Update time of Microsoft CA (seconds)	0.62	0.65	0.69	0.74	0.78

本文提出的架構側重於支援最小 SIZE 的即時 CRL 更新，並提出了一個稱爲出版商(Publisher)的機制來彈性及靈活的支援憑證與 CRL 的發布；Delta CRL 的更新模型應用則提供了更好的 CRL 更新效能；而 3 種代理機制的概念更被用來作爲一個訂閱交付 CRL 的動態管理。這使得該架構與任何 PKI 的應用程序運作時，均可產生明顯的改善和具備可擴充的特性。在未來，有兩個方面的建議需

要持續的進行；首先，針對大量發行的 CRL，計劃引進更先進的技術和數據壓縮演算法，以提供 ASBS 系統更快速的更新效能和網路流量的控制。另一個方面，就是要在實際爲多應用程式和多用戶的真實環境下，執行完整的測試，以穫得更進一步的回饋與持續不斷的改善。

肆、心得與建議

一. 網路探勘(Web Mining)的廣泛應用

將資料探勘(Data Mining)的技術應用到全球資訊網上的「網路探勘(Web Mining)」應用，分爲 3 類，包含「網站結構探勘(Web Structure Mining)」、「網站內容探勘(Web Content Mining)」及「網站使用度探勘(Web Usage Mining)」。其中被廣爲使用的「網站使用度探勘」，重視找出使用者在全球資訊網上的瀏覽及存取行爲，以網頁的日誌檔(web log file)爲資料來源，進行探勘──或可建構在資料倉儲系統上運作。

前述的主題摘錄中，正透過 Web Mining「網頁的日誌檔」，不但可辨識出在網站內任意爬行瀏覽的網路爬蟲(Web Crawler)，用以強化網站的資安保護機制，增添資訊安全的防護；另外的助益則是可挖掘出網站中有價值的資訊，據以推薦用戶所偏好的服務項目或增加友善性的網站設計；亦可透過 Web Mining「網站使用者的評價資訊」，用以豐富網站內容的提供或客製化用戶專屬的網頁，這些都大幅提昇了客戶服務的品質和優化了目標市場的行銷，並形塑網站的附加價值與成本效益。應用於本處定期發布的統計資訊網，如：經濟成長率、家庭收支調查、失業率等，或可了解大眾最熱衷的統計數據是什麼，或可預做媒體報導的回應，或甚而可回饋至整體調查的調整，以發布最具前瞻性的、趨勢性的統計報告。

二. 快速更新憑證廢止清冊的機制

本處刻正開發的新版公文檔管系統，規劃使用自然人憑證，執行電子線上簽

核；而線上簽核最終歸檔的公文，亦附加了各層級簽核者的電子簽章及簽核時戳。那這自然人憑證又是如何應用在這系統中呢？當同仁要在公文系統中使用憑證前，需先執行憑證登記─即將有效憑證與本處永久性帳號作連結，且每次在驗證憑證爲有效之後，使用者才可線上簽核。

自然人憑證是否有效，則是透過在公文伺服器上的排程，每日凌晨自內政部網站下載「憑證廢止清冊(CRL)」，來更新資料庫中憑證的狀態。其中內政部自然人憑證管理中心，每天會簽發、公布出一個完整的憑證廢止清冊(complete CRL)以及差異憑證廢止清冊(delta CRL)，並可下載兩者，清冊內容包括自然人憑證主體名稱序號、撤銷日期；同時提供有免費的線上憑證狀態查詢(OCSP)服務，以供應用系統線上查詢即時的憑證狀態之用。

而前述的主題摘錄中，所提出代理訂閱爲基礎的系統(ASBS)，捨去了可能會造成網路流量過大的 OCSP 即時線上查詢的方式，它更新憑證的狀態是以 Pull Agent 及 Push Agent 代理機制來運作，而公文系統更新 CRL 的方式正是類似其中的 Pull Agent 模式，但 Pull Agent 更有效率的是使用較小 SIZE 的 delta CRL 來更新。

因爲一個完整的自然人憑證 CRL 包含了全國所有的憑證廢止資訊，目前下載的 complete CRL 檔案大小約爲 11.7MB，遠較 delta CRL 約爲 31.9KB 來得大，且以自然人憑證截至 100 年 9 月的發卡量高達 235 萬 3 千餘張來看，該憑證的應用將會愈來愈普及化，尤其在每年 5 月的網路報稅的大幅成長的推動趨勢下，未來 complete CRL 檔案將會愈來愈龐大，屆時恐會影響到 CRL 下載及更新時的速度及效能。有鑑於此，爲避免憑證狀態的更新耗費了太多的系統效能，需先預作因應，調整公文系統 CRL 下載的方式；未來，在 ASBS 的應用趨於發展成熟後，或可引用其聰明的 Push Agent 代理機制，反向接收由 ASBS 主動通知的已即時更新的 CRL，將能得到最新的 CRL，以提供信賴憑證者使用。

三. Web 服務(Web Service)的跨平台使用

Web 服務具有不受平台、作業系統及程式語言限制，且易於被重覆使用及擴展功能等優點。相較之下，企業的專屬資訊系統因屬孤島系統，不但在規劃系統時企業成本花費較大，企業內各前後開發使用的系統間，亦無法相容並共享資源，更無法爲因應市場變化，而可彈性且快速的改變系統的業務流程；是以在各大網站平台中(如 Amazon 網站)，不管是以 SOAP 或是 REST API 爲基準的 Web Service，都早已被廣泛應用在它們的資訊服務中了。

而在企業中，以 Web Service 做爲現行各資訊系統的媒介，更是一種常被使用的解決方式；各資訊系統提供各種 Web Service 以與不同資訊系統達到資訊共享與資源共用的好處；透過呼叫 Web Service 及得到 XML 的查詢結果，可解決異質系統中無法快速整合服務的問題，而呼叫者再也不用受到被呼叫系統內部功能程式碼或資料表結構的修正所影響，而需牽一髮動全身的配合修改它呼叫的程序，除非提供呼叫的 URL 被改變了，充分享受到 Web Service 不受平台、作業系統及程式語言限制的好處。

明年將上線之公文系統，便是以.NET 元件來開發 Web 服務，提供本處行政知識網管理系統 AKM 網站呼叫顯示「逾期公文的待辦件數」、「存查續辦公文的待辦件數」及「表單申請的待辦件數」等查詢結果；另外，亦規劃整合 AKM 的差勤資訊，以省略公文系統中同仁需重覆設定請假代理人的作業，目前整合的方式是以開放 AKM 資料庫中的差勤資料表查詢權限，但基於保護各系統資料庫的安全性、及因差勤資料表若異動所導致的查詢服務中斷等考量，未來希望可改採行以呼叫 AKM 所提供的「差勤記錄查詢」Web Service 的方式來整合。

四. 聰明自學的垃圾郵件過濾器

市面上目前可買到的垃圾郵件過濾器，必須可判斷每封電子郵件是否爲垃圾郵件，或是爲合法的郵件(俗稱爲 ham)，它主要倚賴被手動建置的匹配規則模式，來篩選郵件，要建立這樣的規則，恐需要相當專業的知識與技術才能完成，並不是一個簡單的任務。

而且因爲一個錯誤的郵件分類，例如實際應判定爲垃圾郵件但卻被誤判爲合法郵件的情況，使得垃圾郵件仍持續被收信至用戶的郵箱中，這會讓人相當的困擾，而另一個反向的錯誤分類，例如實際應判定爲合法郵件但卻被誤判爲垃圾郵件的狀態，則可能讓用戶遺漏掉該看而未看的重要郵件，它所可能造成的商機、生產力或時間的損失，更加無法估算。再者，垃圾郵件攻擊者更積極以相似性的郵件，來規避垃圾郵件過濾器的追查，也因其技術日新月異，因此任何反垃圾郵件的技術都必須要能迅速反應。

如何分辨垃圾郵件和合法電子郵件，並且能被訓練爲可自動化更新的過濾垃圾郵件的方法，這對垃圾郵件過濾器而言，是非常重要的關鍵；前述的主題摘錄中，所設計提出的垃圾郵件過濾器，它正具備了一種好的垃圾郵件過濾技術所應具有的 3 種特性：(1)它將準確地分類垃圾郵件和合法郵件。(2)它可輕易的適應垃圾郵件不斷改變的技術。(3)它容易自動化更新，以應對新的手法。藉由透過這些特性，讓郵件用戶未來或可不再受到大量垃圾郵件的攻擊與騷擾。

五. 當 Facebook 與知識管理(KM)相遇

知識管理(KM)一直廣爲企業界所重視，正如前述主題摘要中所闡述的，知識轉化的 4 種基本模式，包含有 Socialization、Externalization、Combination 和 Internalization。其中所謂的「Socialization」，指的是隱性知識轉化爲隱性知識的方式，它是一個透過共用經歷所建立的隱性知識的過程，而獲取隱性知識的關鍵是經由觀察、模仿和實踐。「Externalization」，指隱性知識轉化爲顯性知識的方式，它是一個將隱性知識使用顯性化的概念和語言清晰表達的過程，其轉化手法有隱喻、類比、概念和模型等，是知識創造過程中至關重要的環節。「Combination」，指的是顯性知識和顯性知識的組合，它是一個通過各種媒體產生的語言，將各種顯性概念組合化和系統化的過程。「Internalization」，則是顯性知識轉化爲隱性知識的方式，它是一個將顯性知識形象化和具體化的過程，通過「彙總組合」產生新的顯性知識被組織內部員工吸收、消化，並陞華

成員工自己的隱性知識。

由此可知 KM 的重要性，不管是將資深員工個人的 know how 轉化爲組織業務的 know how，或者是將組織業務的 know how 快速轉化爲新進員工個人的 know how，這對企業知識的薪火傳承與組織的健全至關重大。尤其在政府機關未來可能面臨中堅階層的大批退休潮，建議有系統並有區隔的適度降低人治色彩，並將這群人寶貴的知識留傳下來，俾使接替者可以適度的接手，以促進組織的活化，減少因爲退休潮所造成的知識斷層影響。

而組織在導入 KM 時，關鍵的成功因素，包括有：塑造知識分享文化、正確的推動策略、推動知識社群經營及專業資訊管理的平台，而使用 Facebook 來做爲 KM 的管理工具平台，是一種相當新穎的應用方式。依據資策會 MIC 新近的調查顯示，83.8%的網路使用者曾經使用過網路社群的活動，吸引他們使用的原因分別爲「同好交流 (69.3%)」、「獲得有興趣的資訊(57.5%)」與「喜歡分享想法(36.3%)」等，Facebook 正因滿足網路人際互動、資訊傳遞與共享的需求，因此有近 70%使用社群網路的使用者，表示幾乎每天都會造訪 Facebook。正因 Facebook 與 KM 同樣均爲基於社群活動的本質在運作，故可運用 Facebook 做爲 KM 知識社群經營及專業資訊管理的平台，這除了可減少組織花費於建置 KM 平台的費用之外，它友善且活潑的介面，亦可吸引組織的使用者前來使用及分享，但需掛慮的是這種類似雲端服務的應用，如何確保組織 KM 資訊的安全性，與下一階段要討論的議題正相吻合。

六. 新興的雲端安全議題

大家都喜歡用雲端服務來做很多事，不管是企業或個人，如作者即運用 Google Mail 這雲服務來傳遞這份報告最新版本的文件，而可以不用使用 USB 儲存媒體來攜帶資料。又如爸爸媽媽最愛上傳小 baby 的相片至網路相簿分享，再如中小企業可透過 Google 許多服務，將業務資料儲存至雲端資料庫，以節省它們建置資訊環境的成本。但對於許多把關鍵業務應用轉移至雲端環境的企業

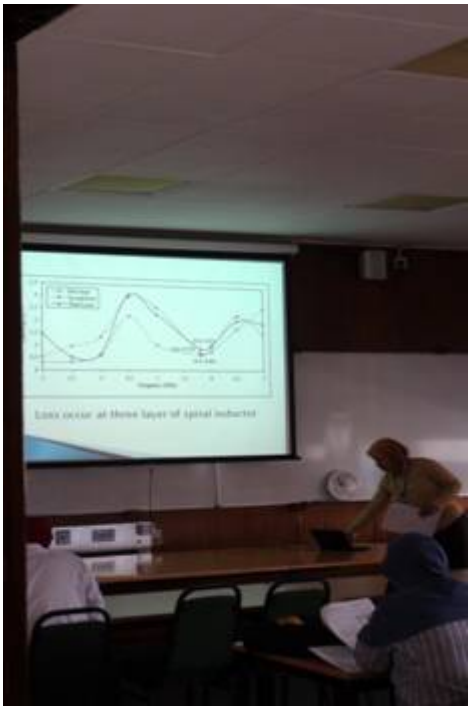
而言，資訊安全往往是首要的考量，儘管雲端服務能提供許多的優勢，但資訊主管們對資安的防範仍相當的謹慎，除非有方法能妥善來管理資訊安全，並可切實遵守資安法規，使雲端環境的安全性可達到實體資料安全的水準。

而且也肇因於雲端環境的安全性，無從讓使用者加以考證，新興的雲端安全議題也開始爲大家所大量關注與討論，雲端安全的議題演變至今，已產生 2 個主要的面向被討論，一個便是一般使用者在使用雲端服務，所應注意的安全議題，而另外一個議題則是如何反向思考從雲端，來提供更好的安全性產品或服務給雲端的使用者，許多資訊安全服務公司如 Symantec 等，也陸續將它們的產品或服務導入了雲端服務當中。

前述的主題摘要，便提出了一種新的「雲端安全管理機制」(Cloud security manager；CSM)，做爲管理雲端服務的安全應用。透過這種模式，CSM 全權負責監測雲端系統的所有活動，針對已登記的客戶端，需經由前端的認證機制驗證後，才可使用雲端服務，並由第三方協助滿足客戶端的要求，另外使用內部控制矩陣，來協助並引導 CSM 針對整體的雲端系統，執行監視系統、客戶和第三方的管理工作，同時它也針對客戶端的個人數據資料，促進與實施了數據的保護法。可見未來在雲端安全的議題與應用仍會是相當火紅的話題。

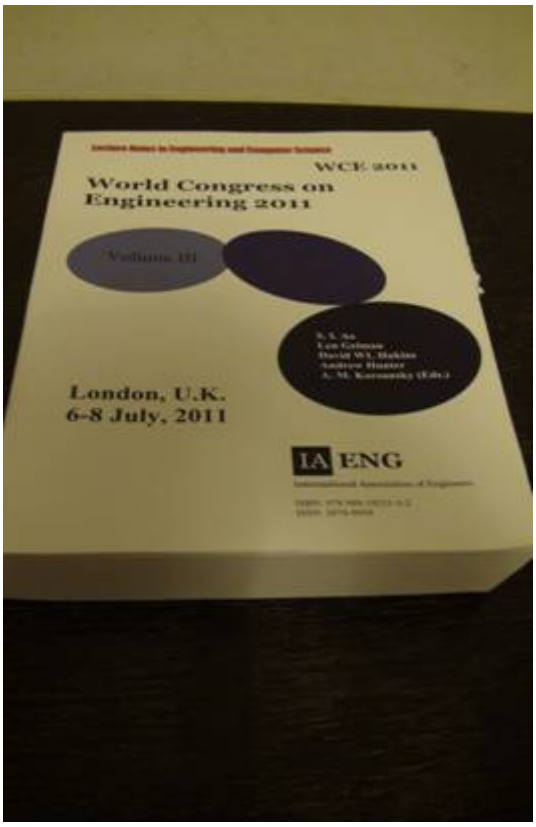
伍、參考資料

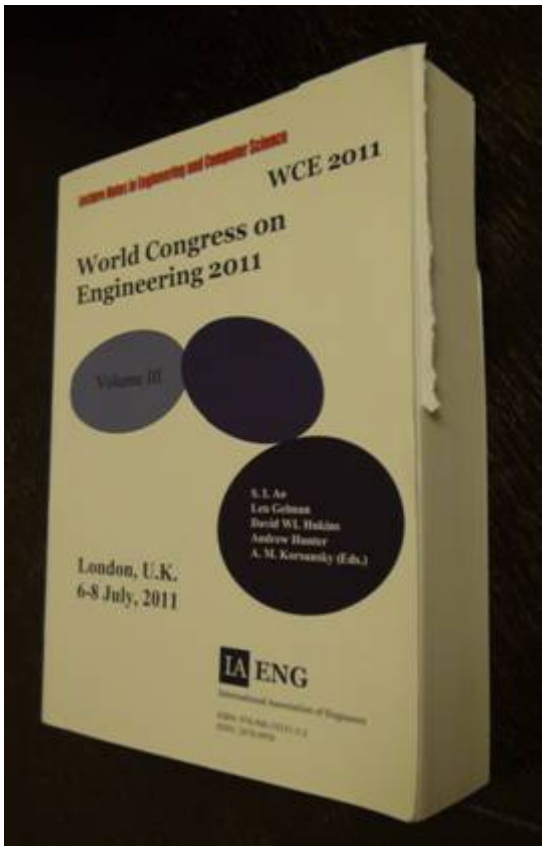

WORLD CONGRESS ON
ENGINEERING 2011
WCE 2011
LONDON, U.K., 6-8 JULY, 2011
IAENG



會場花絮

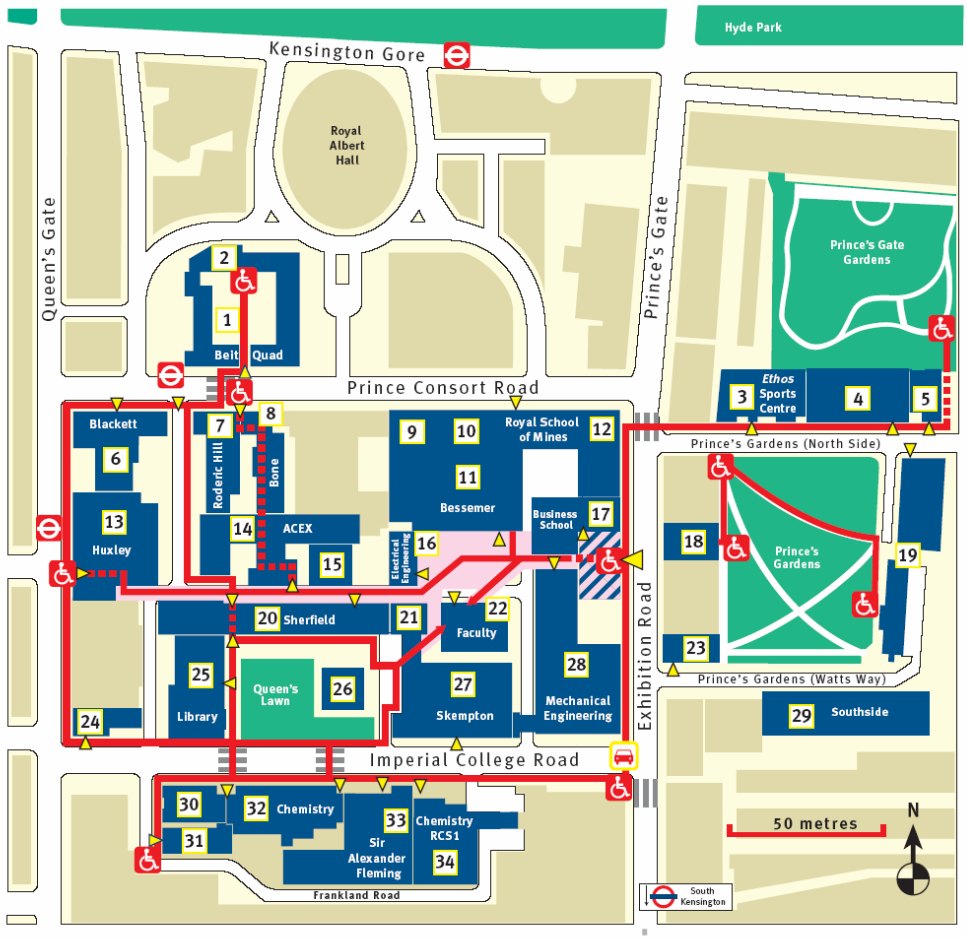

Lecture Notes in Engineering and Computer Science
WCE 2011
World Congress on Engineering 2011
Volume III
S. I. Ao
Len Gelman
David WL Hukins
Andrew Hunter
A. M. Korsunsky (Eds.)
London, U.K.
6-8 July, 2011
IA ENG


WCE 2011
World Congress on Engineering 2011
London, U.K.
6-8 July, 2011
IA ENG


WCE 2011
World Congress on Engineering 2011
London, U.K.
6-8 July, 2011
IA ENG

Imperial College London

South Kensington Campus



Buildings where wheelchair access is not possible at this time

1	Beit Quadrangle
2	Imperial College Union
3	*Ethos* Sports Centre
4	Prince's Gdns, North Side Garden Hall
5	Weeks Hall
6	Blackett Laboratory
7	Roderic Hill Building
8	Bone Building
9	Royal School of Mines
10	Aston Webb
11	Bessemer Building
12	Goldsmiths Building
13	Huxley Building
14	ACE Extension
15	William Penney Laboratory
16	Electrical Engineering
17	Business School
18	53 Prince's Gate
19	Eastside (under construction)
20	Sherfield Building Student Hub Conference Office
21	Grantham Institute for Climate Change
22	Faculty Building
23	58 Prince's Gate
24	170 Queen's Gate
25	Imperial College and Science Museum Libraries
26	Queen's Tower
27	Skempton Building
28	Mechanical Engineering Building
29	Southside
30	Wolfson Building
31	Flowers Building
32	Chemistry Building
33	Sir Alexander Fleming Building
34	Chemistry RCS1

英國「倫敦帝國學院」南肯辛頓校園地點(地圖)

Date: 22 June 2011

To Whom It May Concern:

Official Invitation Letter for WCE 2011

Participant Name: Miss Chia-hua Chen

Registration Number: WCE2011_1307078365

The World Congress on Engineering 2011 (WCE 2011) will take place in London, U.K., 6-8 July, 2011. We are happy to tell that the above participant has registered for WCE 2011.

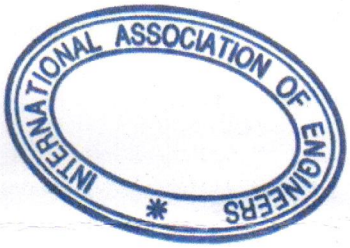
It would be our pleasure that our participants can enjoy the conference by learning the latest research development and exchanging ideas with others. Our conference serves as a good platform of networking the engineering community.

If you have any question, you are very welcome to contact us at any time.

Best regards,

may tang

May Tang
WCE 2011 Organizing Committee
IAENG Assistant Secretary
Email: wce@iaeng.org
http://www.iaeng.org/WCE2011
http://www.iaeng.org



IAENG INTERNATIONAL ASSOCIATION OF ENGINEERS

World Congress on Engineering 2011

London, U.K., 6 – 8 July, 2011

Certificate of Attendance

Miss Chia-hua Chen

IAENG INTERNATIONAL ASSOCIATION OF ENGINEERS

Official Payment Receipt

Date: 6 July 2011

Issued By Olive Choi
IAENG Assistant Secretary

The World Congress on Engineering 2011

London, U.K., 6-8 July, 2011

The registration number is WCE2011_1307078365

Your Registration Information

Participant:

Title*	Miss		
First Name*	chia-hua	Email	anne0105@dgbas.gov.tw
Last Name	chen	Mailing Addr	12F., No.216, Sec. 2, Minsheng Rd., Banqiao Dist., New Taipei City 220, Taiwan (R.O.C.)
Affiliation	none	ZIP/Postal Code	22041
Telephone	886-933884112	City*	New Taipei City
Fax	886-2-23803974	Country	Taiwan (R.O.C.)
Remark			

Registration Fee:

Description	Fee (US$)
Non-Academics Late	550
Total Fee	550

This is a computer generated printout. No signature is necessary.