

以標準化 Metadata 為核心發展金融機構 Ontology 之探討

Developing the Ontology for Financial Institutions based on Standardized Metadata

劉文卿

政治大學資訊管理系副教授

Wenching Liou

Associate Professor, Department of Management Information Systems

National Chengchi University

E-mail: w-liou@nccu.edu.tw

馮國卿

政治大學資管系博士生

Kuo-Ching Feng

Candidate for Doctor's Degree, Department of Management Information Systems

National Chengchi University

E-mail: feng@nccu.edu.tw

關鍵詞 (Keywords)：索引典 (Thesaurus)；知識本體論 (Ontology)；後設資料 (Metadata)；
都柏林核心集 (Dublin Core)

【摘要】

索引典(Thesaurus)與知識本體論(Ontology)兩者皆可提供結構化的詞彙用以描述特定領域知識(Domain Knowledge)。然而想要進一步對領域知識創造出更具有豐富與彈性的知識內涵描述，知識本體論相較於索引典，它包含了比索引典更為廣泛的語意關係(Semantic Relationship)，且其概念定義較為明確。因此，本研究將以某金融機構為例，利用既有之索引典發展出與金融業務相關之 Ontology，並利用都柏林核心集(Dublin Core)為建立文件管理的詮釋資料(Metadata)；此外使用立法院國會索引典中與財務金融相關詞彙索引典作為輸入之

標準，以節省重新開發特定領域 Ontology 所需龐大的時間與成本，並建立個案金融機構的 Ontology。

本文將以(Object Management Group，簡稱為OMG)所提出之MOF(Meta Object Facility)的四層架構為本研究之核心，並參考其他企業模型理論(Enterprise Framework Modeling)如：Zen of Meta-Modeling，並進一步將其擴充為五層之架構。本研究之五層架構由上而下分別為，第一層為 Thing，第二層為資源描述架構 RDF/RDF'S，第三層為網絡層 Network，第四層為 Metadata Ontology，第五層為 Domain Knowledge，並以此架構做為企業知識管理系統之基礎並發展雛型系統。

【Abstract】

Thesaurus can provide structured vocabularies for describing domain knowledge, however, if we want to create a knowledge-rich description of an object, such as required by the semantic web, thesaurus turns out to provide only part of the knowledge needed. In this paper we try to solve the problems related to capture background knowledge for banking and finance resources. We describe a case study in which we attempt to construct an Ontology for a subset of banking and finance, and use the thesaurus of Taiwan congress as input. At the same time, we also follow Dublin core as well as metadata standard.

In this paper, we improve the 4 layers of framework, it is created by meta object facility (MOF) of Open Management Group (OMG), we are extending the framework from 4 layers to 5 layers, including Thing, RDFS, networks, metadata Ontology and domain knowledge, we can apply this framework to develop knowledge management system in the enterprise.

研究動機與目的

在網際網路時代中，知識管理在知識的獲取、儲存及流通的資訊技術上不斷地進步，促使企業得以更有效地使用知識，進而創新知識。一般而言，透過資料倉儲技術進行文件管理，可有效地蒐集與管理企業產生的結構化與非結構化資訊(Liao,2003)；然而，資訊的儲存與檢索，須遵循一定之標準，方能使得資訊的提供者與使用者能夠快速地存取資訊，因此，本研究將著重於建立知識管理流程中之文件轉換格式標準，亦即利用都柏林核心集(Dublin Core)為標準，並以此標準發展金融機構之 Ontology。

索引典(Thesaurus)可提供結構化的詞彙用以描述特定領域知識(Domain Knowledge)。然而，如果想要進一步對特定領域知識創造一個具有豐富的知識內涵的描述，例如產生語意網絡(Semantic web)，但是索引典往往僅能提供部分所需的知識。在企業知識管理系統所需的知識擷取或分享的過程中，經常需要高度結構化的語意描述，知識本體論(Ontology)相較於以往的索引典方法，在語義描述上更為豐富。

因此本研究將以某金融機構為例，利用既有之索引典發展出與財務金融業務相關之 Ontology。並利用都柏林

核心集(Dublin Core)為建立文件管理的詮釋資料(Metadata)，此外使用立法院國會索引典中與所財務金融相關詞彙索引典作為輸入之標準，並建立個案金融機構的 Ontology。

本研究並以(Object Management Group,簡稱為OMG)所提出之MOF(Meta Object Facility)的四層架構為研究之核心，並參考其他理論如Zen of Meta-Modeling進一步將其擴充為五層之架構，五層架構由上而下分別為，第一層為Thing，第二層為資源描述架構(Resource Description Framework,RDF)，第三層為Networks，第四層為Metadata Ontology，第五層為Domain Knowledge，並以此架構做為企業知識管理系統之基礎並發展雛型系統。

文獻探討

都柏林核心集

都柏林核心集是一種能夠依描述主題與資源類型來整理文件的電子資源格式，其特點為簡單且具有彈性，亦能維持相當的延伸性與互通性，因應不同的使用者需求。雖然都柏林核心集允許自訂欄位的存在，但為了資料流通和交換的需要，使用上仍以基本的15個欄位(element)為基礎，然後利用新的修飾語(qualifier)來容納新的需求，有關15個摘要之說明可參考表1。(吳政叡,2003)

索引典(Thesaurus)

“Thesaurus”一字為希臘字，原義為「寶藏庫」，16世紀開始才被當同義詞。直到1950年代，此字正式具有「為了可能蘊藏同義詞的特定概念，因而確定此概念的『敘述語』詞彙」之意涵，(Hans,1995)。美國國家資訊標準組織(National Information Standards Organizations,簡稱NISO)則定義索引典為「利用某種排列順序，藉由等同關係、同形異義字、階層關係與聯想關係，建立控制詞彙之詞間關係，並以標準化的指標來界定」。目的在於增進文件索引的一致性，大多使用在資訊儲存語檢索系統，並藉由入口詞語與敘述語之間的連結以便利檢索(NISO 1993)。理想索引典應是由語意網路所構成的圖形索引典，其主要功能應幫助讀者陳述問題和列舉適當的詞彙，並具備連結功能，以激發讀者在資訊尋求上的聯想能力(黃慕萱,1996)。

索引典之詞間關係包括：等同(equivalence)、層級(hierarchical)和聯想(associative)三種關係，(張嘉彬,1998)參見表2：

表 1 Dublin Core 15 個欄位說明

欄位名稱	說明
名稱 Title	資源名稱，資料創作者或出版者所賦予資源的名稱。
作者 Creator	資源的作者(Author)或創作者(Creator)，資源內容的負責個人或單位，如文件的撰寫者、相片的攝影者。
主題 Subject	主題或與內容相關的關鍵字，以表達資源內容主題，可透過關鍵詞句表達或是利用現有標題表或分類表。
相關敘述 Description	與資源內容有關的文字敘述，例如文獻摘要或者是視覺資料的內容。
出版者 Publisher	出版現行資料的單位。
其他貢獻者 Contributor	其他任何對資料內容有貢獻的個人或是機構，但對資料的貢獻不若作者顯著。
日期 Date	資料以目前型式存在的日期時間，多遵循 ISO 8601 以 8 位數格式，yyyy-mm-dd，著錄；若以其他方式著錄應在記錄中加以說明。
資料型態 Type	資訊類型，如網頁、小說、詩詞等。
資料格式 Format	資料格式，有利於應用軟體的判斷及資料的使用。
識別號 Identifier	足以辨識資源的數字或字串，如:ISBN。
資料來源 Source	足以辨識資源原始資料的數字或字串。
語文 Language	資源內容所使用的語文。
關連資料 Relation	其他關連資源的資訊。
內容範圍 Coverage	資源內容所涉及的時間或地理區。
使用權限 Rights	連結到資料使用說明資訊。

資料來源：吳政叡，2003

表 2 索引典法之詞間關係

詞間關係	符號	種類	意義	範例
等同 (equivalence)	USE UF	同義	意義完全相同，可互相取代的詞	「信用貸款」與「信貸」
		類同義	意義相近或意義不同，但為了索引的目的視為同義的詞	「信用卡」與「簽帳卡」
		組代	索引時以數詞組合代表另一概念	「工商輔導」與「企業輔導」、 「中小企業輔導」
層級 (hierarchical)	BT NT	屬-種	普通概念與特殊概念之間的關係	「銀行」與「商業銀行」
		集-元	集合概念與其所包含的專指概念或元素之間的關係	「房屋貸款」與「首購優惠貸款」、 「勞工購屋貸款」
		整-部	整體概念與部分概念之間的關係	「政大商業銀行」與「政大商業銀行 東台北分行」
		多層級	有些概念從邏輯上看可以同時屬於一個以上的範疇，被稱為具有多層級的關係	「政大商業銀行信託業務」和「政大商業銀行業務」、「銀行信託業務」的關係。「政大商業銀行信託業務」和「銀行信託業務」是屬-種關係；而「政大商業銀行信託業務」與「政大商業銀行業務」是整-部關係。
聯想 (associative)	RT	同一 範疇	涵義有部分重疊但非同義的詞	「人壽保險」與「產物保險」
		不同 範疇	兩詞雖分屬不同的概念體系，但具有強烈的提示關係	「信用貸款」與「保險」

- 一、 等同關係：又稱用代關係，包括同義、類同義和組代關係。為了保證一概念對應一詞的原則，必須選擇其中之一作為描述詞，其它則作為非描述詞。描述詞為其它具有相同概念之非描述詞在索引典中的代表詞彙；而非描述詞則是作為使用者在索引和檢索時，輸入系統的參考詞彙。
- 二、 層級關係：又稱上下或等級關係。凡具有這類關係的詞，彼此是上位與下位概念的關係，相當於分類表中相鄰的上位類與下位類。包括屬-種 (genus-species)、集-元 (set-element)、整-部 (whole-part) 及多層級 (poly hierarchical) 四種關係。若以物件導向的觀念來看，屬-種關係相當於衍生關係、集-元關係為 class-instance 關係、整-部關係類似組合關係。
- 三、 聯想關係：又稱親緣、類緣或相關關係。是指兩個描述詞之間雖無等同或層級關係，但從索引或檢索角度而言，具有相互參照並提醒使用者有另一詞存在的關係。

知識本體論 Ontology

Ontology 之研究

Ontology 可譯成知識本體論、實體組成論或概念化之實體組成論。知識本體論最早是在哲學的領域中被提出來，主要探討什麼是已存在的實體 (entity) 或事物 (Things)，以及這些實體間是以什麼關係來交互作用，並嘗試說明如何定義並規範其表示的方式。依知識系統的構造與複雜度等特徵、術語的關係以及其歷史性功能將知識系統分為：1.術語集 (Term lists) 包含：權威檔 (Authority files)；定義辭典 (glossaries)；地名辭典 (gazetteers)；字典 (dictionaries)；2.分類系統 (Classification and categorization) 包含：標題表 (Subject headings)；分類表 (Classification schemes; taxonomies and categorization schemes)；3.關係群 (Relationship groups) 包含：索引典 (Thesauri)；語意網絡 (Semantic networks) 以及 Ontology (Guarino and Wetly, 2001)。知識本體論可表示為定義一些字 (vocabulary)，這些字彙可以描述這個領域內所有的東西，首先經由定義相關名詞 (term) 來代表領域內的實體，當集合某些名詞就以構成一個字彙，再利用特定機制來分享這些字彙，就形成一個知識本體論。建構知識本體論的功用是分享該領域內有哪些重要的觀念，所定義的名詞間是以什麼關係連結，藉由分享的機制，每一個人都知道這個領域內字彙組織的脈絡。利用這些字彙來發展我們的應用系統，其優點是：字彙的一致性能夠提昇語意關聯性搜尋的準確性，且基

於相同的溝通機能與其它運用相同領域字彙所發展的系統溝通，達到有效的名詞分享。所以知識本體 Ontology 是用於描述及表達特定領域知識的一組概念 (Concept) 或術語，它可以用來研究特定領域知識的類別 (Class)，類別的屬性，以及類別與類別間的關係，進一步達成概念語意的資訊檢索 (Fensel,2000)。

隨著目前電腦科技的發展，知識本體論已漸漸從人工智慧 (Artificial Intelligent, AI) 上應用到知識管理之領域，用以分享知識和再使用知識。例如 William and Austin (1999) 利用 Ontology 描述或表達特定領域知識的一組概念或術語，可使用在組織知識庫中較高層次的抽象知識，也可用來描述與管理特定領域的知識。或者是應用在人工智慧領域中的內容理論 (content theory)，其研究方法是將特定領域知識的對象分類，並描述對象屬性和對象彼此間的關係，它能夠為領域知識的描述提供專業的術語 (Chandrasekaran, et al. 1999)。由以上定義可以歸納出以下 Ontology 的重點：術語或詞彙、術語關係、規則、概念化與形式化的規格說明，描述特定領域的知識，並能清楚的表達知識，使得知識能夠不斷地分享與再使用。

Ontology 分類與應用實例

依照 Ontology 的主題可分為：(1)知識表示的 Ontology (2)通用或一般的常識 Ontology (3)特定領域的 Ontology (4)語言學的 Ontology (5)任務的 Ontology 等。依 Ontology 表示的形式化程度分為：(1)完全非形式化 (2)結構非形式化 (3)半形式化 (4)形式化的 Ontology 等。依上線與否可分為：(1)on-line (2)off-line 等。或依 Ontology 可否分享分為：(1) documenting ontologies (2)sharable ontologies 等。依概念化主題可分為：(1)應用的 ontologies (2)領域的 ontologies (3)衍生的 ontologies (4)表達的 ontologies 等 (Guarino, N., 1997)。

圖 1 即說明通用知識 Ontology 之基本架構圖，此種 Ontology 建構時首先要處理的三個部分，即定義 concept-relation-instances 的關係，因此第一要務為先找出 -- concept，並分析概念與概念間之關係 -- relation，最後再舉出實例 -- instances (Guarino, 1997)。

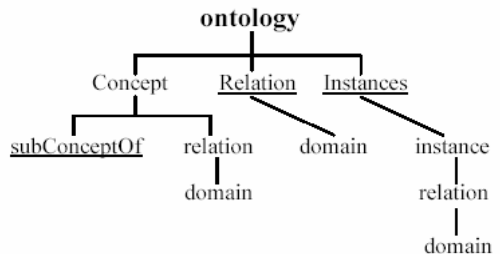


圖 1 Ontology 基本架構圖 (Guarino, 1993)

Ontology 被應用於許多研究領域之中，(Visser, P et.al,1996) 分類出五個實用性的研究領域：

1. 領域理論之發展 Ontologies (Domain theory development) :可用在將文件集合分析及比較出不同的 domain theories。
2. 知識之獲取 Ontologies (Knowledge acquisition) : 描述及組織在 domain 中需要被取得的實體及關係。
3. 知識系統設計 Ontologies (Knowledge system design) : 是在知識系統設計中可重覆使用的結構。
4. 系統化文件 (System documentation) : Ontologies 提供了一個包含詞彙及架構的 meta-level 的概觀，以協助完成系統文件化的工作。
5. 知識的交換 (Knowledge exchange) : Ontologies 能夠被使用來定義不同 Agents 間知識交換 assumptions。本研究將特別注重於第 4 個部分，及發展了一個包含詞彙及架構的 meta-level 的概觀，以協助完成系統文件化的工作。

以下介紹 Ontology 之應用實例「中英雙語語言座標」目前開放的雛型是由「數位典藏國家型計畫」下的「語言座標計畫」所建構完成。主要計畫成員分屬中央研究院語言所與資訊所。目前提供的功能為：

1. 跨語言資訊轉換；目前發展主要為中英雙語語言之轉換。
2. 語言資訊與概念架構及知識本體的連結；目前連結到 SUMO 之全文，簡單介紹 SUMO 這個上層共用知識本體 Ontology。也就是說可以由每個詞查到該詞在概念架構上的歸屬。利用知識本體論架構作知識內容之分類，與簡單的推理，例如：哺乳動物為溫血。



圖 2 中英雙語語言座標之檢索畫面

3. 詞義的區分與詞義關係的連結

圖 2 即為中英雙語語言座標之檢索畫面，由「Ontology」點進去，可看見「樹」這個詞在 WordNet 裡有兩個對譯詞義 (Sense 1、Sense 2) 由「Sense 1」或「Sense 2」點進去，可看到該對譯詞義的「SUMO 對應」，再由此點入，可查看該對譯詞義在 SUMO 的概念，包括英譯與中譯。

從知識組織的角度來觀察，Ontology 的確比分類表與索引典更適用於 Web 環境之下做知識表達，然而此種表達最終的結果仍是為了資訊檢索或是溝通做準備。因而，如何將此一知識組織工具與已被定義之檢索標準或規範作結合，將是知識組織與資訊檢索專家未來更進一步的研究方向。同時，建構 Ontology 需要很專精的領域知識專家、語言學家及資訊學家等的跨領域合作，才能建構出更好、更適用的 Ontology。但也有不少研究者質疑 Ontology 與分類表及索引典的建構目的以及處理概念的方式相近，認為與其移植知識工程界的 Ontology，不如修改補充現有的分類表或索引典的形式或功能，或是應致力於各種知識組織系統之整合 (阮明淑與溫達茂，2002)。因此，本研究即利用既有的立法院國會索引典 (網址 <http://lis.ly.gov.tw>) 做為本研究輸入之標準，節省重新建立 Ontology 之時間與成本。

發展 Ontology 的設計方法

Ontology 編輯器可以幫助研究知識工程的專家與學者建立 Ontology，在 Ontology 編輯器中必須有概念階層、屬性與限制，而這些編輯器也需提供圖形使用者介面並且符合基於應用 Ontology 軟體發展的標準。Protege-2000 是目前較多國家使用的發展 Ontology 的工具軟體，可以讓領域專家建構、修改 Ontology，並且可藉由修改重複使用 Ontology 來建立以知識為基礎的系統。此工具軟體是由史丹福大學醫學院的史丹福醫藥資訊中心所發展，並由美國國家醫藥圖書館、美國國家科學基礎中心與國防高級研究計劃署所贊助。此項軟體可以定義類別、類別階層、資料槽 (slot) 以及資料槽、資料槽中資料限制、類別間的關係與關係的相關屬性，並且可以直接輸入類別實體 (Kamel et. al, 2000)。因此本研究即利用 Protege-2000 為發展 Ontology 的工具，Protege-2000

表 3 發展 Ontology 之工具比較表

	Protege2000	Ontoprise	Ontopia	OilEd
Version	1.8	3.05	1.3	3.4
Type	Editor	Suite	Suite	Editor
Ontology Storage	Database, files	Database, files	Database	Files
Ontology Languages	Flogic, OIL, Ontolingua	Flogic, RDFS	Topic Map	DAML+OIL, RDFS
Open Source	Yes	No	No	No
Plug-in	Yes	Yes	Yes	Yes
Inference Engine	Yes	Yes	Yes	Yes
API	Yes	Yes	Yes	Yes

可以正規化外顯分享概念的敘述，概念 (concept) 有時被稱為類別 (class) 槲架 (frame) 或者專門詞彙 (term)；而描述每個概念中的性質 (property)，有時被稱為屬性 (attribute) 資料槽 (slot) 或者角色 (role)；而每個資料槽中的限制 (restrictions) 有時被稱為關係 (relations) 或者琢面 (facets)。說法雖然不同，然而卻是指向同一件事。因此人與電腦溝通時可以利用 Ontology 讓電腦瞭解人們的意向為何。電腦與電腦 (或者機構與機構) 的溝通亦可以透過 Ontology 來降低溝通的模糊度與增加知識的再使用。以上表 3 為整理不同之 Ontology 發展工具，例如 Protege-2000、Ontoprise、Ontopia、OilEd 等工具之比較表。

研究架構

發展個案機構詮釋資料之規格 (Metadata Schema)

本研究將對個案金融機構內部文件進行都柏林著錄實務的工作，亦即利用都柏林核心集進行電子資源格式制定，藉由都柏林核心集簡易的描述格式，將個案金融機構內部電子化文件，並做一資料格式的制定。首先，利用個案金融機構文件資料庫中所取得的格式，進一步訂立 metadata 的架構格式，每一欄位下之修飾語則根據金融機構的行銷資料庫與公佈的規章制度資料庫的 form 中萃取出其以下的修飾語，而修飾語的部分則依照該銀行不同需求進行調整性增加刪減，制定出個案金融機構專屬之 metadata 規格，見表 4：

表 4 個案金融機構專屬之 metadata 規格

編號	基本欄位 (Element)	欄位說明	修飾語 (Qualifier)	修飾語說明
01	題名 (Title)	文件題名或名稱。	題名	
02	創作者 (Creator)	文件的創作者或單位。	建立者單位	
			建立者姓名	
			建立者代號	
03	主題 (Subject)	文件的分類、標題、關鍵詞、索引典詞彙。	關鍵詞 (Keyword)	一般是從文章內容且其主要是文獻題名中抽出的關鍵性詞彙。
			索引典 (Thesaurus)	採用概念分解與組配的原理，以規範化的描述詞 (Descriptor) 作為文獻主題標幟的主題法。
04	簡述 (Description)	文件的摘要或影像資源的內容敘述。	內容摘要 (Abstract)	
05	出版者 (Publisher)	負責發行作品的組織。		
06	其他參與者 (Contributor)	除了創作者外，對作品創作有貢獻的其他相關人士或組織。	修改人單位	
			修改人姓名	
			修改人代號	
			公佈人單位	
			公佈人代號	

編號	基本欄位 (Element)	欄位說明	修飾語 (Qualifier)	修飾語說明
07	日期 (Date)	與文件形成、修改、公告等相關的日期。	製作日期	
			公告日期	
			取得日期	
			修改日期	
08	資源類型 (Type)	文件的類型或所屬的抽象範疇，例如會議記錄、新聞稿、簽呈、調查報告等。	資源類型	
			文件層級	文件層級訂定建議考量個案內部需要，並參酌台積電公司所訂定的五個層級：Quality Manual、Guide-line / Policy、Procedure、Operation instruction / Work instruction、Records等(陳玫炎等人,2001)。
09	資料格式 (Format)	資訊的實體形式或者是數位特徵，也用來告知檢索者在使用此作品時，所須的電腦軟體和硬體設備。如果是電子檔案，建議使用MIME格式的代表法。	檔案格式	
10	識別資料 (Identifier)	字串或號碼可用來唯一標示此作品者，例如URN、URL等。	文件編號	
			版本	
			版本日期	
			機密等級	
			伺服器名稱	
11	來源 (Source)	資源的衍生來源，例如同一作品的不同媒體版本，或者是翻譯的來源等。		
12	語言 (Language)	作品本身所使用的語言。		
13	關連 (Relation)	與文件為不容輕易分離的附件或夾檔；以及與其他文件（不同內容範疇）的關連，或所屬的系列和檔案庫。	附件	與文件為不容輕易分離的附件或夾檔。
13	關連 (Relation)		相關參照	與文件屬不同內容範疇，但可供參考的其他文件。
			包含	文件所包含的子文件。
			包含於	文件所被包含於的母文件。
14	內容範圍(Coverage)			
15	使用權限(Rights)	Process(文件處理)文件的相關處理人員和進度等。	處理人單位	
			處理人姓名	
			處理人代號	
			處理時間	
			時間限制	
			保存年限	
			歸檔方式	

MOF 四層架構

OMG 所提出之 MOF (<http://www.omg.org>) 的四層架構為本研究之核心架構基礎，四層架構分別說明如下：

1. M0 層：即為所謂的 Information 或 Instance 層，在我們的實作中，做為存放實體資料的地方。
2. M1 層：是 model 或 meta-data 層，這一層的功用就是將資料分類，並加以描述資料所擁有的一些相關的屬性值。例如：Dublin Core 中的 15 項欄位可以描述文件的 Title、Date、Format ..等。
3. M2 層：是 meta-model 或 meta-class 層，這一層的功用主要是描述 model 層 (亦即 M1 層) 的內容，也可以用語言描述資料，並可制定 M1 層的 meta-data 所將展現的 model。例如：可以使用統一模式化語言 (Unified Modeling Language, 簡稱 UML) 和 (Common Warehouse Metadata, 簡稱 CWM) 定義 M1 層物件 (Objects) 中的 Slots 與 Classes 彼此間的關係。
4. M3 層：是 meta-meta-model 或稱為 MOF 層，這一層主要目的為描述 meta-model 層的內容，也以用語言去描述 meta-data。下圖 3 即為 Meta Object Facility 的四層架構的一個實例。

本研究之五層架構

本研究將以上述 MOF 的四層架構為本研究之核心基礎，並參考其他企業模型架構理論 (Enterprise Framework Modeling) 加以擴充，例如學者 Tannenbaum 在 1994 年所提出的「禪的企業模型架構」(Zen of Meta-Modeling)，說明任何企業模型架構經過不斷的分析與歸納，最上一層終究會抽象化為「事物層」(Thing 層)。本研究結合不同理論，進一步將研究架構擴充為五層之架構。五層架構由上而下分別為；第一層為 Thing 層，第二層為資源描述架構 (Resource Description Framework, RDF) 層，第三層為網絡層 Networks，第四層為知識本體層 Metadata Ontology，第五層為領域知識層 Domain Knowledge。本研究將以此架構做為企業知識管理系統之基礎並發展離型系統。下圖 4 即為本研究之五層架構圖。本研究之五層架構分別說明如下：

1. 第一層 Thing：依據 MOF 架構亦可稱為 (M4, meta-meta-model layer)，此層為五層架構的最上層，我們參考 Zen of Meta-Modeling 的企業模型架構理論，所以可用 Thing 來表示 (Tannenbaum Adrienne, 1994)，而 Thing 所表現的模型架構適用於不同企業，所以能夠管理企業處在動態環境中相當重要的架構。Zen of Meta-Modeling 的企業模型架構

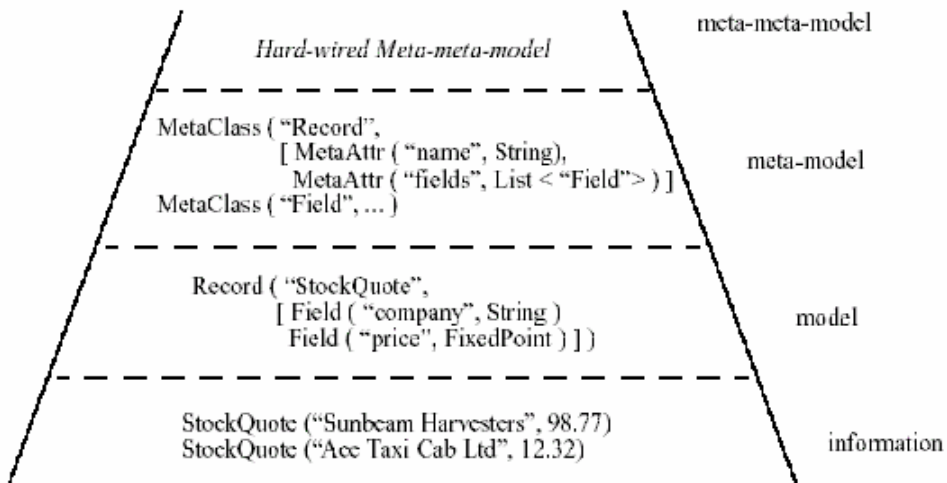


圖 3 Meta Object Facility 的四層架構

中的內容包含了五個層級，如下圖 5 並說明在企業中常見的例子；第一層最主要的就是 Thing，一切的事物的皆可以抽象化為 Thing 的架構。第二層之目的是將企業中的所有商業行為與流程實體化，例如交易處理流程，包括了交易處理程序、交易的地點，和交易所需的資料元素等，共同組成了一個商業行為的實體，也就是將 Thing 的實體化 (Instances of Thing)。然後在第三層可依照第二層的商業行為與流程的實體，又可細分成多個實體，像是原料的清單、提款、功能說明書等，組成了一個零件這一個實體 (Instances of Business Entity)。第四層又依零件實體再細分為多個實體，像托架、螺絲釘、螺帽等實體，組成了一個門鎖實體 (Instances of Part)，此層的實體已經和門鎖這個產品有關的實體，此實體又可自行產生出不同的實體 (Instances of Bolt)。

2. 第二層為資源描述架構層(RDF/S): 依據 MOF 架構亦可稱為(M3, meta-model layer), 此層的主要功能在於能夠清楚定義 meta-data 的結構及語意，並可利用一種語言描述不同種類的資料，例如 XML 語言。然而

RDF/S 載明了特定應用領域的綱要 (Schema)，以宣告該領域自己的資源描述語，利用此綱要該領域的應用層面將可以由語法(Syntax)的層級提昇至語意 (Semantic)層面。故 RDF/S (M3)為一個 meta-model，它可以加以描述本研究架構下一層網絡層 Network (M2)中所必需使用的關係，並可利用 RDFS 將關係抽象化為類別(Class)及其屬性(Slot)，以明確表示其類別之間彼此的關係，讓 Network 層可以利用此層所描述的結構清楚展現其類別間的關係。

3. 第三層為網絡層(Network): 依據 MOF 架構亦可稱為 (M2, model layer), 利用上一層 RDF/RDFS 所定義好的抽象化類別，因此我們建置了 Network (M2)層，此層主要的核心為是以都柏林核心集為基礎，並根據前一節中表 4 個案金融機構專屬之 metadata 規格，進一步去描述資源的類別與語意。有關資源的類別，即是在圖 5 本層中之方框「資源 (Resource)」。在描述資源的語意方面，則是透過 Ontology 來描述知識領域的字彙以表示類別之間的階層架構，讓此架構能夠定義我們所要描述文件的內容，並能夠清

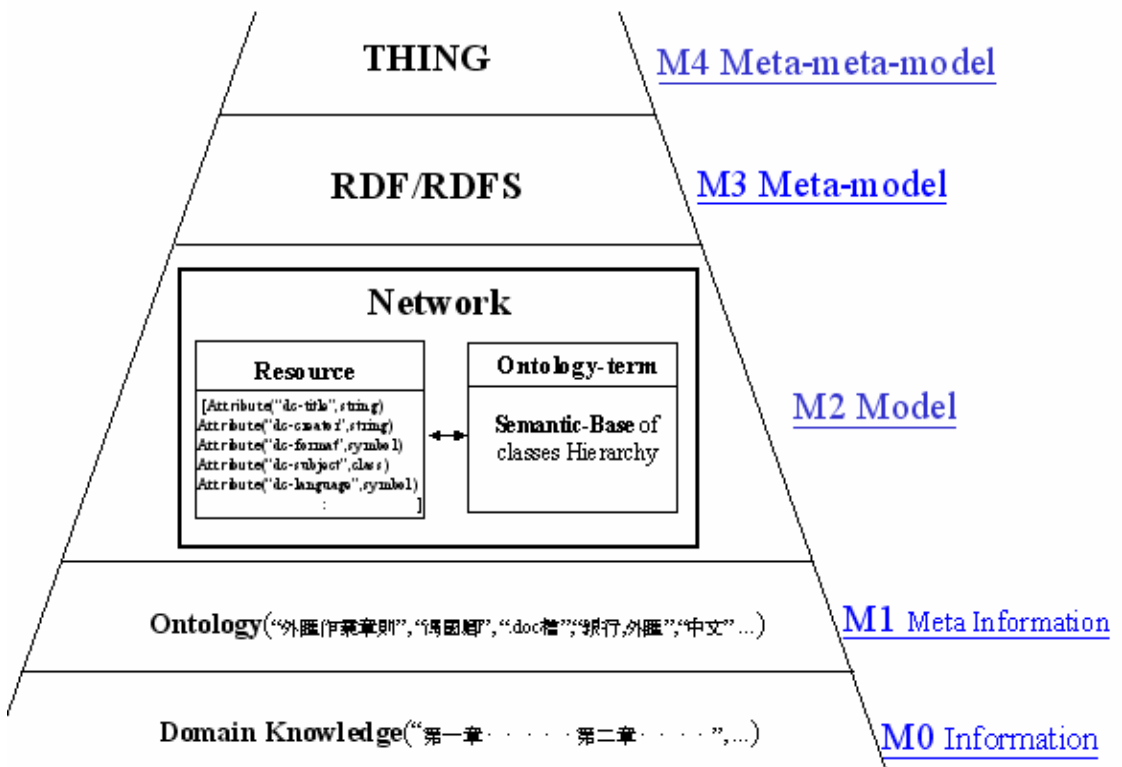


圖 4 本研究之五層架構圖

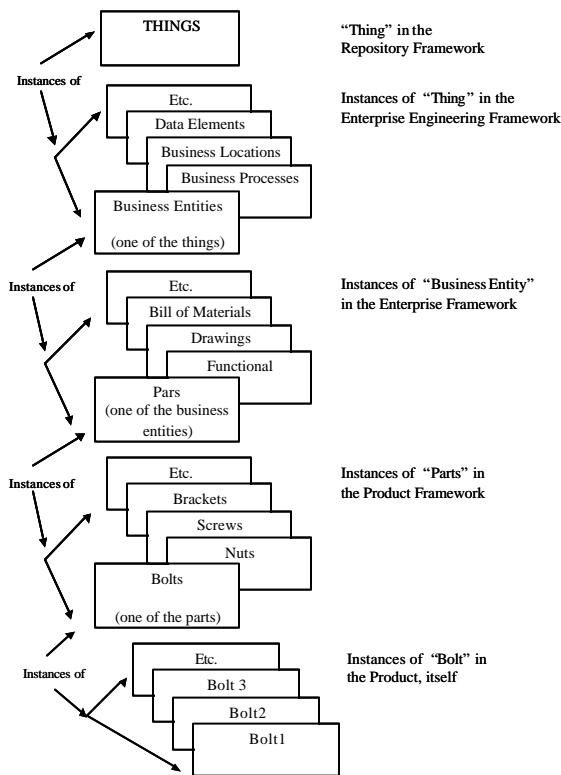


圖 5 The Zen of Meta Modeling(Adrienne Tannenbaum,1994)

楚的加以分類，即在本層中之方框「 Ontology-term 」。藉由 (Resource) 與 (Ontology-Term) 的結合形成了一個包含不同文件類型與語意內容的網絡架構，藉此可以幫助描述下一層 M1 層實體化後的 Meta Information。

4. 第四層為知識本體層(M1, Metadata Ontology)：此層是利用 M2 層所定義的類別及屬性將其實體化，構成一個可以去描述真正資料的 Meta Information，即根據表四個案金融機構專屬之 metadata 規格中所定義的都柏林核心集(Dublin Core)。例如：在圖五中“外匯作業章則”為一份文件的主題(Title)；“馮國卿”是描述一份文件的作者(Creator)；“銀行，外匯”為這份文件的標題(Subject)，標題可為多個值；“.doc 檔”為這份文件的格式。等。並利用上一層網絡層中的(Ontology Term)將資訊清楚且明確的定義其關鍵詞，使所有的不同的資源類型，能夠藉由此架構擁有了清楚的語意階層的觀點同時又具有 Ontology 物件導向(Object-Oriented) 彈性的優點，成為一個描述特定知識領域的 Ontology。

5. 第五層為領域知識層(M0, Domain Knowledge)：依據 MOF 架構亦可稱為 (information layer)，此層為一個真正的實體，也就是一份真實的文件或資料，內容並未經整理及分類，只知道其為一個特定領域的知識，在本研究中是從個案公司資料庫中所取的文件，如：外匯作業章則。

上頁圖 6 說明將金控-Ontology 的類別圖展現為 RDFS 的格式，RDFS 是用來定義每一個抽象化的類別，包括類別的來源與類別之間彼此之間的關係，使得當我們在輸入與檢索實際資料內容時，能夠取用 RDFS 事先定義好的類別來描述其資料的內容。例如：“dc-creator”為馮國卿，而 dc-creator 為一個先前在網絡層中的 Resource 類別中之屬性，其屬性的限制及來源都在設計 RDFS 時已定義完成，所以 RDF 可以去取用 RDFS 所定義的內容，不需要再重新設計。所以 RDF 是一種知識內容的表示模型，能夠去支援網路上的知識共用 (information sharing) 與知識交換 (information exchange)；表 5 即說明本研究所使用之工具 Protege-2000 與 RDF/S 之差異。

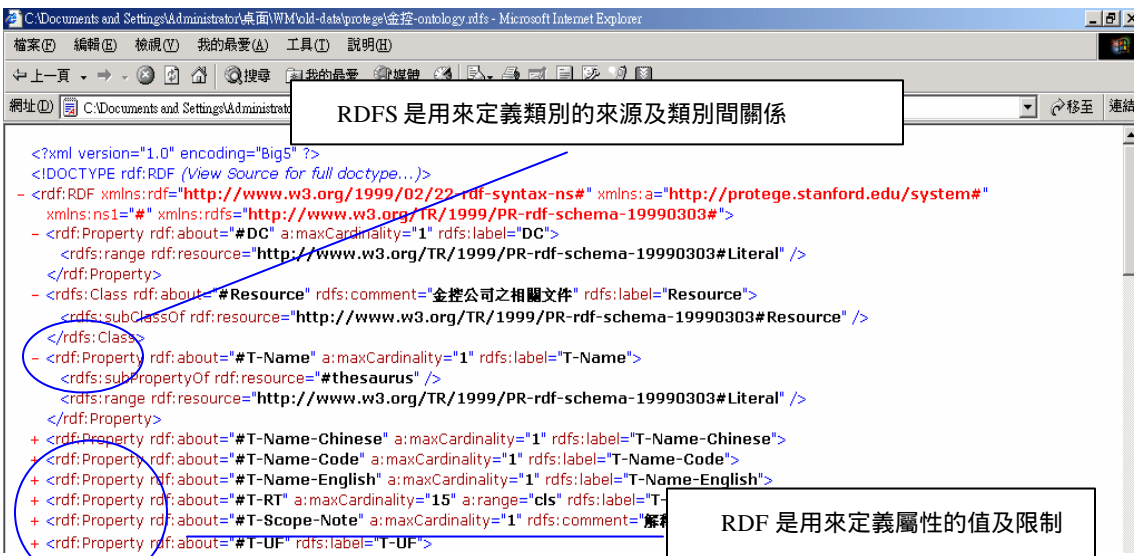


圖 6 金控-Ontology 的 RDFS 圖

表 5 Protege -2000 與 RDF/S 之比較表

Protege-2000	RDF and RDF Schema
一個資源可以只用一個類別來實例	一個資源需使用多個類別來實例
一個屬性的值可以是一個簡單類別的值或是一個類別的實例	一個單一類別屬性的範圍，是限制物件的特性使得類別能夠實體化
一個屬性在一個單一的實例中不能是另一個屬性的子類別，即屬性之間無繼承關係	一個屬性可以是另外某一個屬性的特殊化
容器 (Container) 是做為執行的清單	有三種的容器物件: bag, sequence and alternative.

資料來源：Oscar C. et.al,(2003)

以 Protégé -2000 展現知識結構

以下將以立法院國會圖書館所制定的索引典做為本研究輸入之標準(<http://lis.ly.gov.tw/lgh.html>)，選取其中有關投資、財務、金融、保險等四個部分，並經過查詢結果總數如下表：

以下以「金融」類之檢索詞彙「外匯」為例加以說明：詞彙下面有 SN 範圍註、UF 異形詞、NT 狹義詞、BT 廣義詞、RT 相關詞等不同的屬性；而我們所需用到包括：UF 異形詞(Used For)：可以指出不同的同義詞或近似詞。例和與「外匯」相同義意的字詞 “國際匯兌”。

表 6 立法院國會索引典

序號	類別	名稱	英文類名	總數
1	A	經濟	Economy	124
2	A01	經濟政策	Economic Policy	29
3	A02	經濟建設	Economic Development	8
4	A03	商業	Business	56
5	A04	貿易	Trade	55
6	A05	投資	Investment	62
7	A08	營利事業	Profitable Enterprise	8
8	A15	工商普查	Industrial and Business Census	5
9	A19	其他	Others	34
10	E	財務	Finance	39
11	E01	國庫	Treasury	5
12	E02	主計	Audit	7
13	E03	稅賦	Tax & Revenue	0
14	E04	金融	Monetary Situation	41
15	E07	保險	Insurance	102
16	E09	國有財產	State-owned Property	8
17	G14	財政	Financial Administration	39
18	G15	教育	Education	258
19	G16	經濟	Economics	124

資料來源：(<http://lis.ly.gov.tw/lgh.html>)

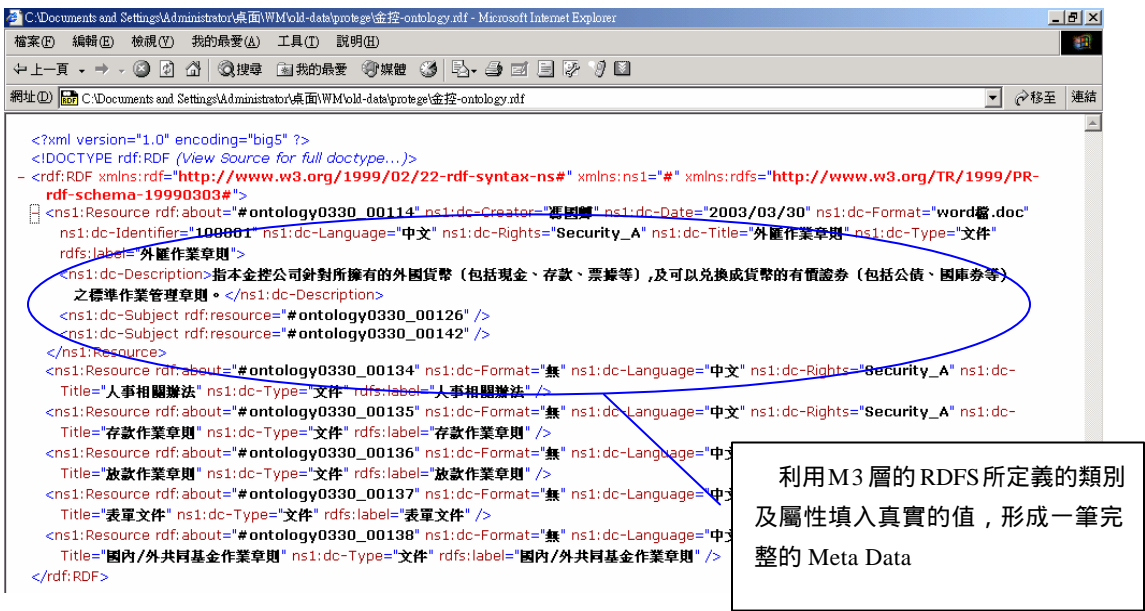


圖 7 金控-Ontology 的 MetaData 圖

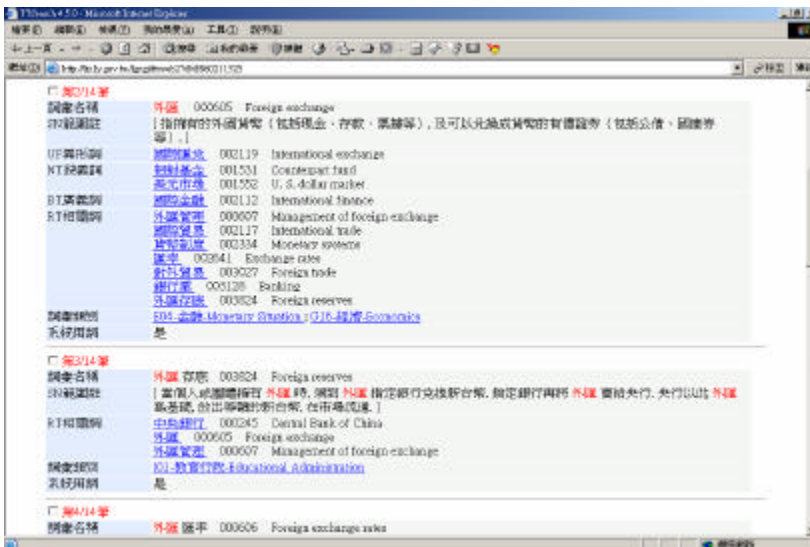


圖 8 立法院國會索引典檢索詞彙「外匯」圖

NT 狹義詞 (Narrowed Term)：可以指引至更專精的詞彙。例如「外匯」下更範圍更小的字詞 “相對基金”。

BT 廣義詞 (Broader Term)：指引至更廣泛概念的詞彙。例如「外匯」之上更範圍更大的字詞 “國際金融”。

RT 相關詞 (Related Term)：可指引至其他相關詞彙。例如與「外匯」有關的字詞 “外匯管理”，並可參見上圖 8 立法院國會索引典檢索詞彙「外匯」圖。

以「金控 MetaClass」做為 Protege-2000 的 MetaClass，當做在下頁圖 9 中之「金控-Ontology-term」的 Class 的

輸入樣版 (Template)，凡是繼承「金控-Ontology-term」的 Class 都必須在建立時皆需輸入 BT、NT、RT、UF 等相關詞彙，其中 BT、NT、RT、UF 等屬性，在 Protege-2000 稱為 MetaSlot。

下頁圖 10 說明由本研究所自行定義的 MetaClass 中的「金控 MetaClass」所產生的 Template，引進並輸入立法院國會索引典中與「外匯」相關的 BT、NT、RT、UF 等。其中可輸入 Term 中英文名稱與此 Term 所具

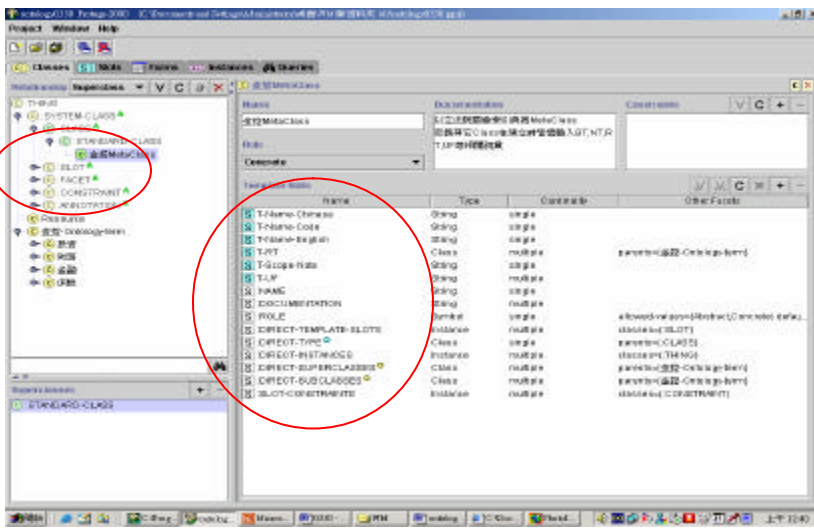


圖 9 「金控 MetaClass」之 MetaClass 圖

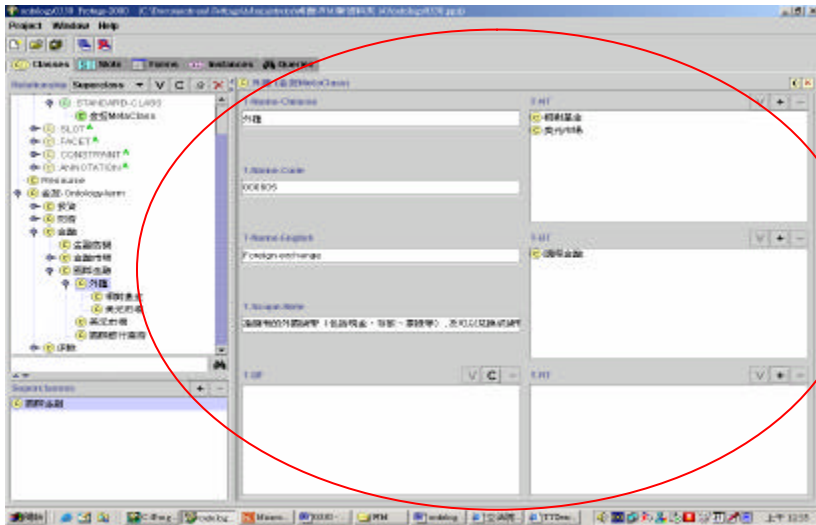


圖 10 MetaClass 「金控 MetaClass」所產生的 Template 圖

有特定的代號(Code)及填入 SN 以解釋此詞之適用範圍。此時 BT 會自動設為 SuperClass 的類別，如：國際金融，NT 為自動設為 SubClass 的類別，如：相對基金，RT 則必需參照立法院國會索引典以 “+” 號加以選取，UF 則是建立同義字的功能。

下頁圖 11 則說明在「Resource」類別之下為例，如何新增一個 Instance，此類別(Class)皆依據 Dublin Core 為基礎訂定其 Slots，所以其產生的 Instance 皆依其訂定的 Slots 填入相關的資料，例如：Dc-Title 外匯作業章則、

Dc-Creator 馮國卿、Dc-Subject 銀行與外匯 等，以實際描述一份文件。

Ontology 目前研發的最廣泛的就是語意網頁內容檢索之相關研究 (Karp et.al,1999; Decker, S. et.al, 2000; Oscar C. et.al, 2003)。開發一套使用 Ontology 之觀念與人工智慧的推理方式，並建構在 XML 與 RDF 的標準下，為網頁內容建立共通的詞彙與概念；如此一來可以對網頁內容進行有意義的搜尋，而不再只是關鍵字詞的尋找。

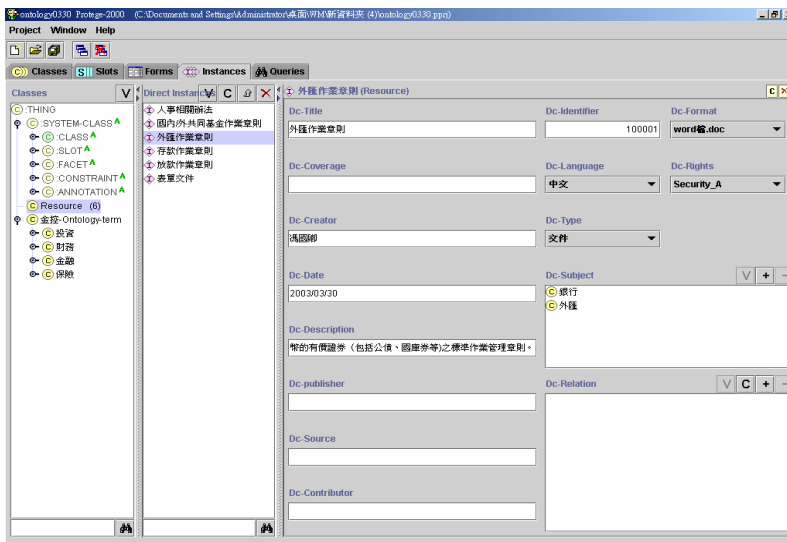


圖 11 在「Resource」類別為例新增一個 Instance

結論

本研究主要針對個案金融機構之文件管理相關議題進行個案探討，有鑑於文件管理範疇廣泛，本文主要針對如何建立文件管理流程中文件轉換格式之標準進行探討，在技術面上以都柏林核心集、索引典與 Ontology 等方法，試著解決目前文件管理所遇到的問題。因此，本研究將以某金融機構為例，利用既有之索引典發展出與金融業務相關之 Ontology；並利用都柏林核心集 (Dublin Core) 為建立文件管理的詮釋資料 (Metadata)，此外使用立法院國會索引典中與財務金融相關詞彙索引典作為輸入之標準，以節省重新開發特定領域 Ontology 所需龐大的時間與成本，並建立個案金融機構的 Ontology。

本研究以 (Object Management Group) 簡稱為 OMG 所提出之 MOF (Meta Object Facility) 的四層架構為本研究之核心，並參考其他企業模型理論 (Enterprise Framework) 如 Zen of Meta-Modeling，進一步將其擴充為五層之架構。本研究之五層架構由上而下分別為，第一層為 Thing，第二層為資源描述架構 RDF/RDF'S，第三層為網絡層 Network，第四層為 Metadata Ontology，第五層為 Domain Knowledge，並以此架構做為企業知識管理系統之基礎並發展雛型系統。一套完善之知識管理制度，有賴於整體組織成員共同努力，至於後續之知識處理流程、知識管理控管等機制礙於研究限制，將作為未來研究之建議，以臻完善。

參考文獻

- 吳政毅 (2003)。都柏林核心集 (Dublin Core) 簡介。
http://dimes.lins.fju.edu.tw/ucsiv/eng/dublin_intro.html
- 張嘉彬 (1998)。索引典及其於資訊檢索上之探討。
書苑, 36, 46-59。
- 張慧鈺 (2001)。主題分析。圖書館資源組織與整理研習手冊。中國圖書館學會編。
- 黃惠株 (1996)。淺談索引典。佛教圖書館館訊, 5, 2-7。
- 黃慕萱 (1996)。資訊檢索。台灣學生, 229-236。
- 阮明淑、溫達茂 (2002)。Ontology 應用於知識組織之初探。在佛教知識組織管理研討會會議論文集，國立中央圖書館臺灣分館編，7-1~7-16。
- 資訊科學學會臺北分會、農業科學資訊服務中心、國家圖書館主編 (1994)。索引典理論與實務。中國圖書館學會。
- 陳玫言、吳蕙如、蔡銘箴 (2001)。科技產業下知識管理系統運作模式與探討-以台積電為實例。第十二屆國際資訊管理學術研討會論文集。台灣大學。
- Chandrasekaran, B., Josephson, J. R., Benjamins, V. R. (1999). What are ontologies, and why do we need them?, 20-25.
- Decker, S. et al (2000). The semantic web - on the roles of XML and RDF. IEEE Internet Computing.
- Fensel, D. (2000). Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce. Springer-Verlag Berlin.

- Guarino, N. & Welty, C. (2001). Identity and Subsumption. In, Green, R., Ed. Semantic Relations. Kluwer.
- Guarino, N. & Welty, C. (2001). Supporting Ontological Analysis of Taxonomic Relationships. Data and Knowledge Engineering 39(1), 51-74
- Guarino, N.(1997). Understanding, building and using ontologies. International journal of human and computer studies 46(3/4), 219-310.
- Hans. H. Wellisch(1995). Indexing from A to Z (2nd ed.). New York: H. W. Wilson, 472.
- ISO2788, Documentation-guidelines for the development of monolingual thesauri, second edition.
- Jean Aitchison, Alan Gilchrist, & David Bawden(1997). Thesaurus Construction: A Practical Manual (3rd ed.). London: Aslib, 38, 67-69
- Karp, P. D., Chaudhri, V. K. & Thomere, J.(1999). XOL: An XML-based Ontology exchange language. (V. 0.3).
- Megill, K. A., and H. Schantz. (1998). Document Management : New Technologies for the Information Services Manager. London : Bowker-saur.
- NISO(1993). National Information Standards Organization, ANSY/NISO Z39.19, Guidelines for the Construction, Format, and Management of Monolingual Thesauri, (Bethesda, Maryland: The American National Standards Institute, 1993)
- Oscar C. et. al.(2003).Methodologies, tools and language for building ontologies. Where is their meeting point? Data and Knowledge Engineering . 46, 41-64.
- Santosus, M., and J. Surmacs. (2000). The ABCs of Knowledge Management. <http://www.cio.com/research/knowledge/edit/kmabcs.html> (Accessed May, 2001)
- Soergel, D.(2001). Thesauri and ontologies in digital libraries tutorial. joint conference on digital libraries (JCDL 2001) Roanoke, VA, USA.: 694.
- Sowa, J. F. (2001). Agents. Retrieved April 11,2002 from World Wide Web: <http://users.bestweb.net/~sowa/Ontology/agents.htm>. (Accessed June, 2001)
- Tannenbaum, Adrienne(1994). Implementing a Corporate Repository: The Models Meet Reality. New York: John Wiley.
- Purba, S. (2000). Data Management Hand BOOK (3rd ed.). London : Boca Raton.
- V. Balasubramanian, V., and A. Bashian. (1998). Document Management and Web Technologies. Communications of The ACM (41)7, 107-115.
- Visser, P. and Bench-Capon, T.(1996). On the reusability of ontologies in knowledge-system design. Database and Expert Systems Applications, Proceedings., Seventh International Workshop, 256 – 261.
- W.H. Inmon, John A. Zachman, Jonathan G Geiger(1997) Data Stores, Data Warehousing, and the Zachman Framework Managing Enterprise Knowledge. McGraw.Hill.
- William, S., Austin, T.(1999). Ontologies. IEEE Intelligent systems,18-19. <http://lis.ly.gov.tw>. (Accessed Feb,2003). [http:// www.omg.org](http://www.omg.org). (Accessed May,2003).