

以互動問答方式建構數位典藏知識自我學習系統

黃志泰

德明技術學院資管系助理教授

hgt@takming.edu.tw

摘要

數位典藏內容網站有著許多豐富的資料，若我們想從這些網站裡學習知識，就必須透過網頁瀏覽方式或靠網站提供的檢索機制來查詢相關資料。但是，這種資訊檢索機制，對知識學習者來說，可能很難檢索到所要的資料，這是因為必須先熟悉檢索機制與相關領域知識，而熟悉各個網站的檢索機制與使用檢索機制，對知識學習者來說，在各方面的領域知識還不太充足的情況下，實在是一件很困難的工作。因此，對知識學習者來講，其最希望的知識學習方式，應該是直接告訴他們問題的答案及答案所在，而不是讓他們獨自摸索與彙整答案，這也是最好與最快速的自我學習方式。所以，本研究的目的，便是希冀善用各式數位內容網站知識，透過互動式問題回答機制，來縮短知識學習者學習知識的流程及提高自我學習知識的興趣。

關鍵詞：問題回答系統、自我學習系統、知識擷取。

1. 前言

自從國科會「數位博物館專案計畫」推行以來，歷年來已建置完成包括語文藝術、人文社會、自然生態、生活醫療和建築地理等約一百個主題網站，經過我們深入分析各類數位典藏網站的內容資料後，認為某些數位典藏網站的內容與素材，已足以提供給知識學習者作為知識學習的來源，因為許多豐富的資料與知識皆已隱含在其中，但是，我們若想從這些數位內容網站裡學習知識或查詢資料，則必須透過網站網頁資料瀏覽方式或靠數位內容網站提供的檢索機制來查詢資料，這對知識學習者來說，太困難了，這是因為若是透過網頁資料瀏覽的方式來找尋資料，就必須一頁頁的來瀏覽，其缺點是很花時間，而且可能最後還是沒找到所要的資料，其次，若是透過目前數位內容網站提供的檢索機制來查詢資料，雖然理論上可幫忙快速找到資料，不過必須先熟悉其檢索機制與相關領域知識，否則也很難找到資料，而熟悉各個數位內容網站的檢索機制與使用檢索機制，對知識學習者來說，在各方面的領域知識還不太充足的情況下，使用檢索機制還是一件很困難的工作。這是由於各個網站過於強調專業的知識與過於專業的檢索介面，對知識學習者來說很難與之做資訊互動，所以在橫縱向資訊無法交流的狀況下，其自我學習知識的成效可能相當有限。

不過對這些知識學習者來說，目前最大的問題

是歷年來建置完成的各式數位典藏內容網站只是做到縱向或單一主題的知識整合，對他們在學習知識的過程中，缺乏橫向的知識連接工具來幫助他們串起知識的連結或相關概念，所以從數位典藏內容網站上學習到的可能只是某些片段的知識，很難讓他們能夠窺視到整個知識相關概念的面貌。

因此，就學習理論來說，對知識學習者來講，其最希望的知識學習與問題解決方式，應該是直接告訴他們問題的答案及答案在哪裡，而不是讓他們獨自在摸索答案，而這也是知識學習者最好與最快速的自我學習方式，例如，直接問“什麼是昆蟲？”、“昆蟲有幾支腳？”等。

所以，本研究的主要目標，便是希冀全方位善用各式數位典藏內容網站知識，透過互動式問題回答機制，來縮短知識學習者學習知識的流程及提高自我學習知識的興趣，同時進一步納涵知識管理理念，協助知識學習者逐步建構自我學習環境，達到個人進修、知識增長與進階自我學習之指引。

因此，本研究運用中文問題回答系統技術與基於概念架構的中文資訊擷取技術結合主要之自然與社會科學數位典藏內容網站透過跨網資訊連接等方式來轉置建置成為適合知識學習者的問答式自我學習系統。

2. 相關研究探討

1970 年代強調教導 (instruction) 式學習，但自 1985 年起迄今已改成學習 (learning) 為主。主要的理由是大多數學者認為真正教育的目標在累積有用的知識，培養「個人的研究」態度，建立自我學習能力[2]，而網際網路上的各式數位內容網站正提供了一個自我學習知識的環境，其特點是多種型式的資訊呈現方式(multiple media)，它利用文字、圖形、語音、動畫以及視訊等多種型式作為資訊的表現，不但能吸引使用者的興趣，而且更能充分的表達知識的意義[6]。從教育的角度來看，各式數位內容網站正試圖扮演輔助及改善傳統靜態及序列學習的角色，透過此種學習方式，使用者可以 (1) 建立自我導引及探索性學習 (2) 培養互動學習環境中主動求知方式 (3) 層次性的建構知識 (4) 機動性的智慧性學習[3]。

不過，當知識學習者還沒有明確的相關領域知識基礎時，其知識的學習常常是透過提出質疑(問題)、智能重組等回饋過程的激發，才會獲得成長[4]。因此，目前各式數位內容網站對知識學習者尋求知識的過程中最大的障礙是(1)資訊查尋技能：如

網站索引及分類參考工具的使用問題。(2)系統障礙：線上檢索的設計，往往忽略知識學習者檢索策略的貧乏，造成使用上另一障礙。(3)智能障礙：如檢索策略中的布林邏輯運用[18]。

所以，目前的數位網站內容若直接運用於知識學習者的自我學習上會面臨迷失學習方向的問題，因為多層次多元化的網路超文件／超媒體系統，對於一個初學者來說，容易造成混淆，使人猶如陷入迷宮中。因此建構主義的探索性學習，在學習者不知如何進行或尋找資訊的線索時，這種學習就像瞎子摸象，見樹不見林，迷失了學習的方向，造成學習情緒的沮喪[5]。

因此，本研究提出透過鼓勵知識學習者直接提問的方式與知識智能重組等回饋過程來激發與統合各類知識，以建立自我學習環境與培養主動求知的精神，幫助其在數位學習過程中減少因資訊技能的缺乏與智能障礙造成學習阻礙，進而培養其透過探索(Exploring)和發現(Discovery)式的學習過程來自我獲得知識的成長。

而能讓知識學習者直接提問的系統就是互動式問題回答系統，近年來問題回答系統在資訊檢索、資訊擷取、機器學習、和自然語言處理等研究領域都受到注意[7][8][11][12]。很多系統都正在發展當中，主要的研究方向以網際網路上的資訊為潛在的答案來源為主，而若是以網際網路上的資訊為潛在的答案來源的問答系統，大多會採用現行的網路搜尋引擎如 Google 來作為相關問題答案的網頁資訊蒐集，然後再透過網頁文件分析與答案擷取技術來回答問題。如 Kwok [9]等人發展的 Mulder 問答系統，這個系統在很多步驟上跟我們所提出的方法很類似，針對每個問題，Mulder 問答系統會轉成若干查詢句給網路搜尋引擎並且分析其傳回的結果。Mulder 問答系統會對問句與所有傳回的網頁資料做複雜的語法分析，這會比我們的分析方法更複雜和更強調電腦的計算能力，除此之外，為了答案抽取和選擇，他們也需要先建立整體性的 idf 詞組權重資料，這就需要存儲詞組權重資料的資料庫空間，而我們的系統不同於 Mulder 問答系統因為我們不需要做很深層的自然語言語法分析。

除此之外，Agichtein [14]等人針對問答系統展現一個查詢句的轉換技術，同樣的技術也出現在 Eric [13]，這個技術概念發展是因為目前的網路搜尋引擎例如 Google 等都沒有提供或具備自然語句的查詢介面，他們的方法是將問句轉換成某種特定領域或者含有某些特定資訊的詞組格式，透過這些詞組格式會大大的改善找到答案的機會，然而，他們的研究是專注在如何使網路搜尋引擎能精確地找到使用者所要的文件資訊，所以，沒有提出任何機制可以從這些文件資訊中來找出與回答答案。

最後，Dragomir[10]等人提出以機率方法發展自然語言問答系統，整個系統的處理流程有五個步驟：問句轉換，文件抓取，段落擷取，詞組擷取與答案排序。而他們所提出的一些機率方法則是運用在後三個步驟中，不過，他們所提出的方法可能無法適用在多變的中文語句中，而且跟我們的方法也有所差別，主要是因為我們不使用整個文章段落擷取

來當做問題的答案。

不過，從上述的研究探討，我們可以了解到問題回答系統已成為知識學習者自我學習知識的最佳工具，這已是不可避免的發展趨勢。

3. 數位典藏知識自我學習系統架構說明

本研究之互動問答式自我學習系統，原則上利用我們之前研究所提出的網頁知識擷取技術[15-17]，將典藏機構之數位網站內容資料加以再次組織整理，然後運用概念分析技術將所擷取之資料作知識概念分析與知識連結分析處理，以建構出知識學習網絡資料庫，然後透過問題回答技術建構自我學習系統之問題回答介面與使用者互動，使使用者能直接以中文問題提問之方式達到互動式的自我學習效果。整個系統架構如下圖 1 所示。

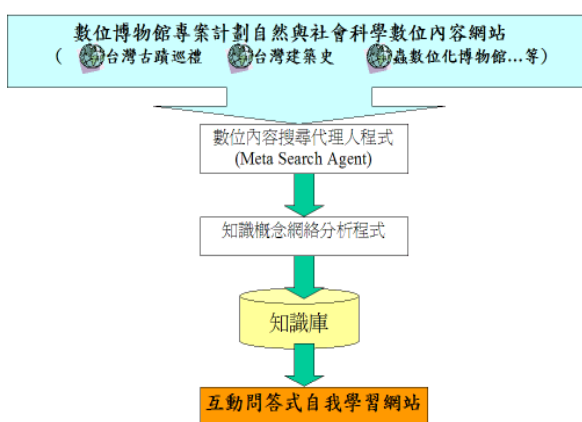


圖 1 數位典藏知識自我學習系統架構

4. 互動式問答系統建構方法說明

目前數位內容網站資料種類可依資料型態概分為近似結構化與非結構化資料。資料形式則包括有文章、圖表、影像及聲音等。數位內容網站上的近似結構化資料大都是以表格格式來展現，其資料形式則包括圖像、文字與數字等資料型態，而數位內容網站上的非結構化資料則大多以圖文描述方式來直接呈現相關資訊。由於數位內容網站有著各類型的資訊，我們需要不同的知識擷取技術與策略。限於論文篇幅，本文僅針對非結構化網頁文件資料提出擷取方法，並透過範例說明與實際系統建立等方式來驗證提問式自主性學習方法的可行性，整個問答系統的運作流程架構如下圖 2 所示。

我們的系統發展步驟主要分為：(1).問句分析(2).網頁擷取(3).網頁分析(4).答案擷取與知識概念關聯等幾大部分，其中問句分析與網頁分析，都必須先完成斷詞與詞性標註等工作，在我們的研究中斷詞是採用中研院詞庫小組所開發的中文分詞技術[1]，至於詞性標註，一般是採用句法分析的文法標註，如名詞，動詞，形容詞等，此種標註的優點是能明確標示句型架構，但缺點是較難表達字詞的概念關係，所以較適合應用在文法或句法分析上，在語義分析與知識概念分析上就較為困難，為了解決

此一缺點，我們就將中文字詞分為一般字詞、文句關係詞與疑問詞等三大類，另外對這些字詞定義一套詞性標註，簡略說明如下。

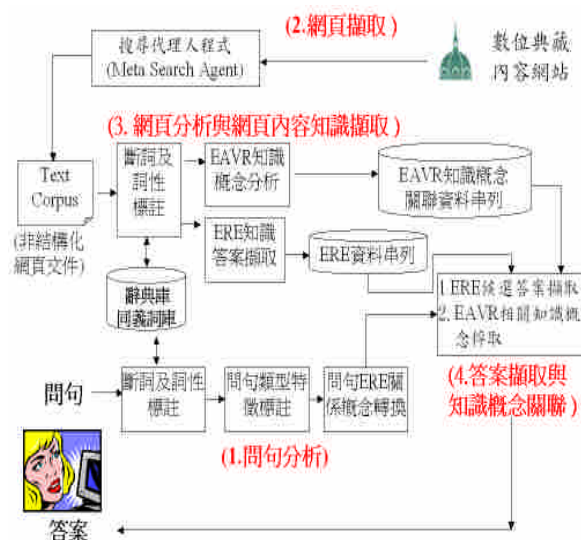


圖 2 數位典藏知識問答系統運作流程架構

(1)一般詞性標註

這類的字詞大多為句法分析上的名詞，在我們所提出的概念關係模式中大部分是歸屬於主體詞(ENTITY)，主要的詞性標註類型為：

詞性標註	詞性標註說明
O	物件主體詞(Object)，如”汽車”
NX	名詞或屬性詞組(attribute)，如”住址”
Q	計量單位詞組，如”公升”
V	數量詞組，當數字資料與計量單位詞合併後就轉為數量詞組，如”一百二十公升”
T	時間詞組，如”今日、昨日”
NXX	新詞或未能判定之詞組
X	虛詞，如”因此”

(2)文句關係模式詞性標註

字詞與字詞之間，基本上是透過連接詞來串接，這類的字詞在概念關係模式中是歸屬於關係詞(RELATION)，主要的詞性標註類型為動作類與集合類：

詞性標註	詞性標註說明
A	動作詞(Action)，如”發佈”，”建立” …
▷、◁、∈、∩、∪、→、= …	集合屬性詞組，如”是”、”位在”、”的”…

(3)疑問詞詞性標註

主要運用在問句分析，用來確定問句類型，以發掘發問者的主要意圖及所想要了解的訊息，相關細節在下面問句分析步驟中將會說明。

至於系統的其他運作流程步驟與方法說明，分

別在下面章節中描述。

4.1 問句分析

我們依據問句的特性，透過斷詞與詞性標註，將使用者的問句關鍵詞組轉換為問題主體(Entity)、限制條件與動作條件(Relation)、答案類型與答案資料型態等項目，然後再將問句關鍵詞組轉換組合為 ERE(Entity-Relation-Entity)的關係概念架構，其中問題主體為問句中的物件名稱或名詞詞組；限制條件與動作條件為問句中的關係詞，此關係詞會影響最後找出之答案；透過問句之疑問詞可以確認問題的答案種類及答案資料型態。

基本上問題的答案型態可能是人名、地名、時間點、數量、定義、原因與方法說明，而答案的長度類型有短答、長答及是否類與選擇類的答案，透過問句分析，確定問句類型以後，大致就可以發掘問句的主要意圖及發問者所想要了解的相關訊息，以利爾後答案之尋找，而問句類型可經由問句的疑問詞來判斷，如下表 1 所示。

表 1 疑問詞對應表

疑問詞	詞性標註	對應的答案型態	對應的答案長度
誰、何人…	XNA	人名詞組	短答案
哪裡、哪邊、何處、哪兒、那裡…	XPL	方位性質的地名詞組	短答案
何時、啥時、何年、何月、多久…	XT	時態性質的時間點詞組	短答案
多少、多高、多長、多重…	XV	數量性質的數量詞組	短答案
什麼、甚麼、何謂、何為…	XD	定義或描述說明詞組	長答案
如何、怎麼…	XE	方法說明詞組	長答案
為什麼、為何…	XR	原因說明詞組	長答案
是否、是不是…	XYN	肯定或否定的答案型態	短答案
還是、或是、或…	XOR	選擇類的答案型態	短答案
那些…	XMA		多個答案以上
那一…	XSA		一個答案

舉例來說，問句”為什麼冬天看不到蝴蝶?”經過問句分析與詞性標註後會得到如下之資訊。

Step1: 詞性標註

為什麼(XR)冬天(NX)看不到(A)蝴蝶(NX)

Step2: 問句分析

問題主體:冬天(NX)、蝴蝶(NX)

限制與動作條件:看不到(A)

答案類型:長答案

答案資料型態:原因說明詞組

Step3: 問句 ERE 關係概念轉換

Entity1	Relation	Entity2
冬天	看不到(A)	蝴蝶

4.2 網頁擷取

此部分的工作是定時由搜尋代理人程式(Meta Search Agent)去搜尋典藏機構之數位內容網站的網頁文件資料，並將其逐一擷取回來，準備交由網頁分析程序處理。

4.3 網頁分析與網頁內容知識擷取

網頁文件資料種類可依目前資料處理方式概分為結構化與非結構化資料兩大類，而所謂結構化的文件資料是泛指表格化的網頁文件資料，其餘沒有固定撰寫順序格式的網頁文件資料，則被稱為非結構化文件資料。在分析非結構化文件資料時，必須先完成斷詞與詞性標註，而網頁分析的目的就是希望能做到各式網頁內容的知識擷取，而所謂的知識擷取是要能辨識或找出隱藏在各式網頁內容的知識，並希望能做到相關知識粹取與整合，因此，本研究針對非結構化的網頁資料，提出 EAVR 知識概念分析與 ERE 知識答案擷取方法，分別說明如下。

4.3.1 EAVR 知識概念分析方法

此方法主要是用來彙整與構築網頁內容的相關知識概念，希望透過概念關係來關聯出知識概念樹，我們透過特定字詞的語意瞭解，將字詞與字詞之間的關係轉為關係(Relation)模式，並仿照集合理論將文件中所隱含的知識概念關係模式化，並定義了如下的基本集合關係(Relation)模式：

基本集合關係	關係概念標註	特定字詞	例句與集合關係概念轉換
包含	\supset	的、之...	Ex: 美國的白宮 美國 \supset 白宮
被包含或屬於關係	\subset \in	在、是、位在...	Ex: 日月潭在南投 日月潭 \subset 南投 日月潭 \in 南投
交集	\cap	同位詞(如“;”符號)、和...	Ex: 台北和高雄 台北 \cap 高雄
聯集	\cup	或者、或...	Ex: 台北或高雄 台北 \cup 高雄
延續關係	\rightarrow	“,”符號、其、由...	Ex: 台北, 高雄 台北 \rightarrow 高雄
同義關係	=	“[]”符號...	Ex: 澳洲[澳洲大陸] 澳洲=澳洲大陸

透過以上的基本集合關係模式，當文句中含有某些特定的字或詞時，其代表關鍵詞組(Entity)與關鍵詞組(Entity)之間隱含有某些概念關係，因此，又可定義出下列的概念推導關係模式：

基本關係 1	可推導之概念關係
IF $E1(\text{Entity1}) \supset E2(\text{Entity2})$	THEN $E1 \exists A1(\text{Attribute1})$ $A1 \rightarrow \text{Value}=(E2)$ (Entity1 至少可推導存在一個 Attribute1, 其 Value=Entity2)

例如：“澳洲的天鵝河”這句中的 Entity 為“澳洲”(Entity1)，“天鵝河”(Entity2)，而“的”這個字則轉換為集合關係模式： \supset (包含)，所以可推論出以下的概念關係架構：

澳洲(E1) \supset 河川(A1)
河川(A1) \rightarrow Value=天鵝河(E2)
(河川可透過知識概念樹搜尋替換而來)

基本關係 2	可推導之概念關係
IF $E1(\text{Entity1}) \in E2(\text{Entity2})$	THEN $E2 \exists A1(\text{Attribute1})$ $A1 \rightarrow \text{Value}=(E1)$ (Entity2 至少可推導存在一個 Attribute1, 其 Value=Entity1)

例如：“皇蛾是台灣最大的昆蟲”這句中的 entity 為“皇蛾”(Entity1)，“台灣”(Entity2)，“昆蟲”(Entity3)，而“是”這個字則轉換為集合關係模式： \in (屬於)，所以可推論出以下的概念關係架構：

台灣(E2) \supset 昆蟲(A1)
昆蟲(A1) \rightarrow Value=皇蛾(E1)
(昆蟲可透過知識概念樹搜尋替換而來)

基本關係 3	可推導之概念關係
IF $E1(\text{Entity1}) \rightarrow E2(\text{Entity2})$	THEN $E1 \exists A1(\text{Attribute1})$ $A1 \rightarrow \text{Value}=(E2)$ (Entity1 至少可推導存在一個 Attribute1, 其 Value=Entity2)

例如：“天鵝河由澳洲大陸往西邊流”這句中的 entity 為“天鵝河”(Entity1)，“澳洲大陸”(Entity2)，而“由”這個字則轉換為集合關係模式： \rightarrow (延續關係)，所以可推論出以下的概念關係架構：

天鵝河(E1) \supset 位置(A1)
位置(A1) \rightarrow Value=澳洲大陸(E2)
(位置可透過知識概念樹搜尋替換而來)

所以，結合前面的論述，我們發展一套知識概念分析方法，稱之為“EAVR 概念樹”，我們認為文件，大多在描述或告知某些資訊，所以有意義的文件內容，應該包含下列知識概念特徵因子：

- Entity:** 主體(文件內容的關鍵詞組)
- Attribute:** 描述主體的屬性
- Value:** 屬性的值
- Relation:** 串聯主體與主體或屬性之間的關係

透過知識概念特徵因子組合運算可得出文件的知識概念特徵因子，因此，文件若以向量式表示，可表示如下：

文件的向量式: $D_s=(E_i, A_j, V_k, R_l)$
 (Ds: 第s個文件, E:Entity, A:Attribute, V:Value, R:Relation)

經由文件特徵擷取後所產生的”EAV “串列(list),再透過 E-R Model 或 EAV 串列關係轉成文件的 EAVR 概念樹,最後再經由文件樹間的相似度計算完成樹的合併,進而擴展成為較大的 EAVR 概念樹,透過 EAVR 概念樹的關係索引,就能將相關知識概念粹取與整合。

4.3.2 ERE 知識答案擷取方法

為了能進一步做到問題答案回答,所以另外將文件概念簡化轉換成為 ERE(Entity-Relation-Entity)的關係概念架構,透過以上的詞性標註,我們將文件轉換成為 ERE 的二元關係串列,如下所示:

文件詞句=E+R+E+R...或 R+E+R+E.....

所以文件內容,就轉變為以二元語義關係串列來描述整篇文章主體與主體之間的關係,例如:”皇蛾是台灣最大的昆蟲”這句話就可轉成如下的 ERE 串列:

Entity1	Relation	Entity2
皇蛾	是(ε)	台灣
台灣	最大的(A)	昆蟲

因此,第s個文件的 ERE 串列集可以表示如下:
 $D_s=(E_i, R_k, E_j)$

經由文件特徵擷取後所產生的”ERE “串列,可做為問題回答的答案來源,這就是我們所提出的 ERE Model 方法。

4.4 答案擷取與知識概念關聯

在此步驟中我們將問句的 ERE 關係串列與網頁的 ERE 關係串列做相似度比對搜尋來找尋候選答案,若候選答案超過一個以上時,我們就另外依據候選答案的權重來決定出問題的最佳或者近似答案,相關演算法說明可參閱我們之前之研究[16]。另外透過 EAVR 概念樹的關係索引,亦可將相關知識概念與資訊連結一併粹取呈現。詳細之運作效果,我們在下一章節中以實際範例來說明。

5. 問答系統運作範例說明

透過網站與網路來學習或獲得知識是未來自我學習的趨勢,但是現存之數位內容網站建構時未以人類學習知識之概念來思考及架構,因此造成知識分散而難以連結。為簡單說明與比較問答系統與傳統數位內容網站運作的差異性,所以,我們以昆蟲數位化博物館(<http://insect.cc.ntu.edu.tw/91insect/>)的網站內容做對照分析。以下圖3這個部分網頁內容為例,隱藏許多有關昆蟲的知識,有文字與圖像

資訊,甚至整個網頁檔案內容也是知識的來源,但問題是圖像跟什麼資訊有關?網頁內容到底描述了哪些重要的知識?然後這些知識跟哪些知識有關?或者跟哪些網頁檔案內容相關?傳統的數位內容網站是靠知識學習者自我摸索與彙整,而自我學習系統,則靠系統的自我知識學習分析。

昆蟲是一種節肢動物,到目前為止已經發現了一百萬種左右,佔地球上所有動物種類的百分之七十以上。昆蟲在三億五千萬年前即出現在地球上,而人類出現的時間,也不過是一百萬多年而已,由此可知昆蟲出現年代的久遠。

一般而言,昆蟲的體型並不大,以臺灣最大的昆蟲是皇蛾,當翅膀展開時也只有二十五公分左右。

昆蟲大部份生長在氣候溫暖的地方,寒帶地區種類較少,大致而言,只要種物容易生長的地方,就能夠找到昆蟲。



皇蛾是台灣最大的昆蟲,翅膀展開有25公分以上。

昆蟲的特徵

昆蟲生長的過程有四個時期包括卵、幼蟲、蛹及成蟲。昆蟲是節肢動物的一類,而節肢動物門的動物一般有以下特徵:

圖 3 昆蟲數位化博物館部分網頁內容

就自我學習系統來講,以上這些問題,必須靠知識概念關聯分析才能做到,在我們的研究裡是透過 EAVR 概念樹分析方法,將上述網頁內容的知識抽取轉化為 EAVR 知識概念關聯樹來表示,下圖4為我們節錄部分 EAVR 知識概念樹來展現其知識概念的組成架構。

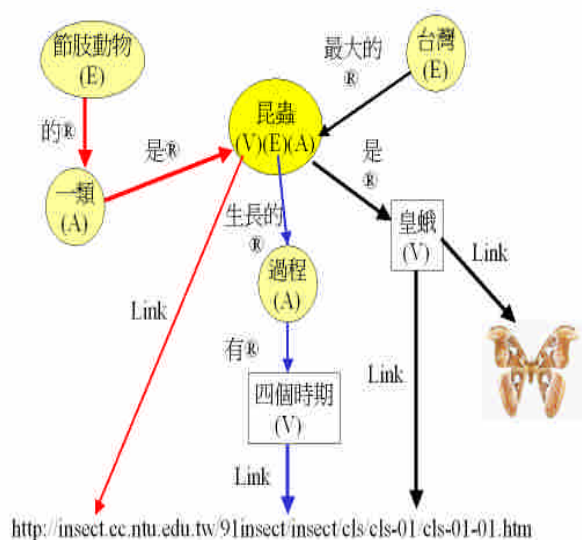


圖 4 昆蟲數位化博物館網頁內容 EAVR 概念樹化

之後的動作,就是將網頁內容再轉換為 ERE 的二元關係串列,以方便之後的問題答案的查找,同樣的我們以上述網頁內容為範例轉化為 ERE 關聯串列,下表2為我們所節錄的部分 ERE 二元關係串列內容。

表 2 ERE 二元關係串列內容表

Entity	Relation	Entity
昆蟲	在	三億五千萬年前
三億五千萬年前	出現在	地球
地球	→(延續關係)	人類
人類	出現的	時間
時間	是	一百萬多年
一百萬多年	→(延續關係)	昆蟲
昆蟲	出現	年代
年代	的	久遠
臺灣	最大的	昆蟲
昆蟲	是	皇蛾
皇蛾	當	翅膀
翅膀	開展只有	二十五公分左右

待完成上述的知識概念關聯分析後，自我學習系統只要透過問句分析，將使用者的問句轉化為對應的問句 ERE 關聯串列後，透過 ERE 與 EAVR 關聯串列相似度比對搜尋或資料型態與權重分析後，就有能力來回答使用者的問題。

事實上，透過我們的 ERE 與 EAVR 概念關聯分析方法，以此網頁範例為例，可以立即回答下列問題：

- Q1: 台灣最大的昆蟲？
 Ans: 台灣最大的昆蟲是皇蛾。
 Q2: 昆蟲生長的過程？
 Ans: 昆蟲生長的過程有四個時期。
 Q3: 昆蟲什麼時候出現在地球？
 Ans: 昆蟲在三億五千萬年前出現在地球。
 Q4: 皇蛾翅膀開展有幾公分？
 Ans: 皇蛾當翅膀開展只有二十五公分左右。

6. 雛形系統建置功能展示說明

為了實現與驗證互動問答式自我學習系統的運作成效，我們建置了自我學習雛形系統，目前我們所建置完成之互動問答式自我學習系統具有自然語句提問與知識概念導引的功能，如下圖 5 所示。

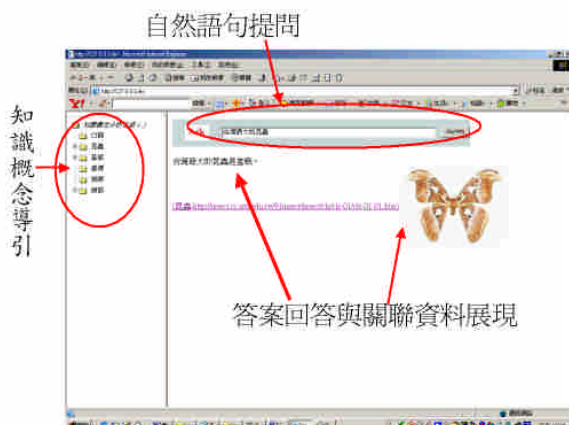


圖 5 自然語句提問與知識概念導引功能展示圖

系統允許知識學習者以自然語句來提問問題，例如“冬天的主要星座有那些?”，“故宮博物院位於哪裡?”，“為什麼兒童不能喝咖啡?”，“圖書館何時誕生?”，“全世界最大的拱橋?”，“台灣有哪些國家公園?”等，系統都會以最接近的答案來回答問題，除此之外，亦可透過知識概念導引來加速知識概念的學習與縮小資料搜尋範圍。若透過知識概念導引更可達到下列功能：

(1) 概念式分類知識提示：當知識學習者對提問的問題沒有概念時，提示概念式分類架構。

例如知識學習者對昆蟲沒有什麼提問的概念時，可提示此類資料的主要內容概念架構，如“種類”、“體長”等，以幫助知識學習者快速蒐尋所要的資料。如下圖 6 所示。



圖 6 概念式分類知識提示

(2) 動態知識概念分類擴展：依據知識學習者的提問問題概念詞，做即時動態的知識概念分類擴展，以提示知識學習者及幫助知識學習者快速蒐尋所要的資料。

例如知識學習者欲搜尋“蚜蟲”的相關資訊時，可即時動態提示此詞之概念分類架構，如“體長”、“翅鞘”等，如下圖 7 所示。



圖 7 動態知識概念分類擴展提示

(3) 蒐尋資料收斂: 依據知識學習者的提問問題概念詞及即時動態的概念分類架構, 以做到蒐尋的資料量自動過濾縮減, 以幫助知識學習者快速蒐尋到所要的資料。

例如知識學習者欲搜尋”昆蟲”相關的網頁時, 透過分類概念延續為”昆蟲, 蚜蟲”後, 即可將原本的搜尋網頁量從七篇收斂為一篇, 以幫助使用者快速蒐尋到正確的網頁資料。如下圖 8 所示。



圖 8 動態蒐尋資料收斂

7. 結論與預期發展

本研究運用知識概念分析與問題回答技術嘗試讓分散之知識能加以重組與連結, 使知識學習者能真正利用網站知識來達到快速自我學習之目的, 除此之外, 本研究亦嘗試建構網路知識仲介者的角色, 將數位內容網站視為知識提供者, 一般大眾則視為知識學習者, 而本研究所建構之互動問答式自我學習系統網站則扮演知識轉仲介與撮合之角色, 使需要獲得知識與學習知識之使用者能快速得到所要之知識, 這將是我們最重要的預期研究發展成果, 未來整個網路自我學習系統的角色扮演架構希望能如下圖 9 所示。

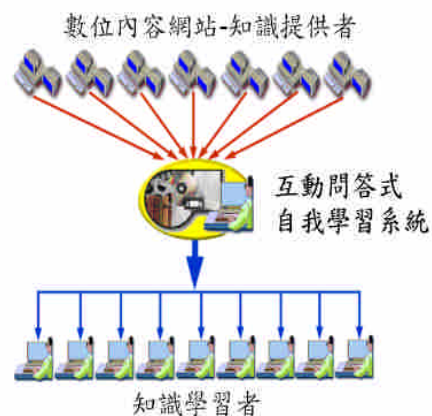


圖 9 自我學習系統網絡架構圖

參考文獻

- [1] 中央研究院中文計算語言研究小組中文詞知識庫小組, <http://godel.iis.sinica.edu.tw/CKIP/>
- [2] 卡拉絲姬原著, 王文科譯, ”兒童的認知發展導論”, 台北市: 文景, 1990, 頁 319。
- [3] 朱錦鳳, ”超文件於較學上應用的特性”, 資訊與教育雜誌, 1994, pp 31-35。
- [4] 陳正乾, ”皮亞傑與維高斯基的對話”, 台北: 台北市立師院, 1996, 頁 22。
- [5] 張史如, ”從建構主義的觀點套討網路超文件/超媒體應用於教學上的意義”, 資訊與教育, 58 期, 1997, pp 39-48。
- [6] 楊家興, ”隔空教育下的傳播科技: 新舊教學媒體的省思”, 教學科技與媒體, 1995, pp 5-12。
- [7] AAAI Spring Symposium Series, Mining Answers from Text and Knowledge Bases, 2002.
- [8] ACL-EACL, Workshop on Open-domain Question Answering, 2002.
- [9] C. Kwok, O. Etzioni and D. Weld, ”Scaling question answering to the Web”, In Proceedings of the 10th World Wide Web Conference (WWW'10), 2001, pp.150-161.
- [10] Dragomir Radev, Weiguo Fan, Hong Qi, Harris Wu, and Amardeep Grewal, ” Probabilistic Question Answering on the Web”, WWW2002, Honolulu, Hawaii, USA, May 7-11, 2002.
- [11] E. Voorhees and D. Harman, Eds., Proceedings of the Ninth Text Retrieval Conference (TREC-9), NIST Special Publication, 2001, pp.249.
- [12] E. Voorhees and D. Harman, Eds., Proceedings of the Tenth Text Retrieval Conference (TREC 2001), NIST Special Publication, 2002, pp.250.
- [13] Eric J. Glover, Gary W. Flake, Steve Lawrence, William P. Birmingham, Andries Kruger, C. Lee Giles, and David M. Pennock, ” Improving category specific web search by learning query modifications”, In The Proceedings of Symposium on Applications and the Internet, SAINT 2001, San Diego, California, 2001.

- [14] Eugene Agichtein, Steve Lawrence, and Luis Gravano, "Learning search engine specific query transformations for question answering", In the Proceedings of the 10th World Wide Web Conference (WWW 2001), Hong Kong, 2001.
- [15] Gai-Tai Huang and Hsiu-Hsen Yao, " A Non-Training Approach to Generate High Semantic Title for Chinese Documents", to appear in Journal of Information Science and Engineering, Institute of Information Science, Academia Sinica, 2004.
- [16] Gai-Tai Huang and Hsiu-Hsen Yao, "Chinese Question-Answering System", Journal of Computer Science and Technology, Vol.19, No.4, July 2004, pp. 479-488.
- [17] Gai-Tai Huang and Hsiu-Hsen Yao, " A System for Chinese Question Answering", The 2003 IEEE/WIC International Conference on Web Intelligence (WI 2003), 2003.10.15, pp. 458-461.
- [18] Pauline Heather, "research on information skills in primary-school", The school librarian,1984,p214.