

語意網時代的網路華語教學： 兼介中英雙語知識本體與領域檢索介面

Chinese Language Education and the Developing Semantic Web:
An Introduction to Chinese-English Bilingual Ontology Interface

黃居仁 張如瑩 蔡柏生

中央研究院語言學所籌備處

churen@gate.sinica.edu.tw, ruyng@gate.sinica.edu.tw, dylan@gate.sinica.edu.tw

壹、背景

提姆·柏納李(Tim Berners-Lee)於 1989 年提出全球資訊網(World Wide Web)的提議書，1990 年底將第一個網頁瀏覽器以及網頁伺服器放置在網際網路上免費提供大家下載使用，在 1991 年夏天全球資訊網開始逐漸在網路上流傳開來，引發之後各種商業瀏覽器的爭奪戰，更引發了網際網路使用習慣的革命。時至今日，電子商務的議題被炒的火熱，線上遊戲社群遍及各年齡層，數位學習的課程囊括各範圍、各層級，這些族繁不及備載的功能與運用，大多數都是透過全球資訊網達到相關目的，從線上教學、從事交易、使用圖書館、獲取娛樂效果、進行社會活動，全球資訊網可說是在我們生活中逐漸扮演著極重要的角色，甚至可說是無所不在。全球資訊網的出現使人們得以達到全球資源分享的願景，閱讀資訊時從以往紙本閱讀的線性或矩陣式瀏覽模式，改為猶如模擬人腦般以直覺、跳躍式的方法，且不論文件、資料庫、圖檔、聲音、影像或某種互動對話螢幕時，可以為任何國家使用任何電腦的任何人所接收、連結和找到（當然，有時須先取得同意）。

就語言學習者來說，傳統語言學習常使用的工具為紙本字/辭典，所提供的資訊包括：定義、拼法、讀音、音節、用法、衍生詞，有些字典還提供字源、詞源、同義詞、縮略詞、俚語以及方言等的相關訊息。隨著全球資訊網的出現，陸續出現線上電子字典。線上電子字/辭典繼承了全球資訊網的特性，提供了不同於傳統字/辭典的功能。在全球資訊網上透過各式各樣的線上字/辭典網站，查詢各種語言字彙、翻譯，不用花一毛錢也不必安裝軟體，隨時都可完成。教育部在網路上推出的四部電子字典，就是線上字典服務最好的例子。由教育部國語推行委員會編修的「重編國語辭典修訂本」[1.]、「國語辭典簡編本」[2.]、「國語小字典」[3.]及「異體字字典」[4.]四本字典，除了已經發行光碟版以外，也開放網友在網路上免費使用。以往查詢中文字典必須要麻煩的計算筆畫或記憶部首，現在透過線上即時查閱，基本的字義與拼音資料等，都可以快速透過網頁顯示出來，不再需要隨時帶著厚重的中文字典。利用網路多媒體的特性，簡編本與國語小字典等提供小朋友或外國人查閱的中文字典，還提供了圖片對照說明，這是一般紙張中文字典所辦不到的事情。[5.]線上電子字/辭典的出現提供了語言學習者多元、多媒體且最便捷的服務。

然而當初柏納李提出以 URI (Universal Resource Identifier)、HTTP (Hypertext Transform Protocol) 以及 HTML (Hypertext Markup Language) 三要素所建構而成的全球資訊網只是資訊的載體，透過超連結將資訊串連，只有人類能讀懂，電腦並不能理解其中所蘊含的語意，無法自動處理資料和資訊。1999 年柏納李在其著作 " Weaving the Web " ([織網]，中譯書名：一千零一網) 中提及他對全球資訊網的夢想之二便是語意網 (Semantic Web) [6.]。在語意網上電腦可以藉超連結找到關鍵詞的定義，用關鍵詞做邏輯推理，從而取得語意資料的意義。語意網並不是另一個獨立的網路，而是當前網路的延伸。語意網上的意義已經明確定義 (well-defined)，讓人機之間能有更好的互動合作，使機器能進一步處理並「理解」資料 (目前只能單純「顯示」資料)。[7.]但表達同一現象可能以不同詞彙陳述，也可能以不同語言表達，所以若是想達到語意網的境界必須獲知單一語言中詞彙和詞彙之間的關係，以及語言和語言之間的對應和轉換。語意網便以 (一) 可擴展標示語言 (XML) 作為標示網頁資訊的標準，讓使用者可彈性擴增描述資訊的詳盡度，(二) 透過描述語言 RDF (Resource Description Framework) 配合數位物件資訊的單一識別碼 URI，使每個網路資源內容不但有豐富的描述，也可經由獨特的識別碼，達成有效且完整的內容知識表達，以及 (三) 運用知識本體 (Ontology) 理解和自動推論並轉換知識，可以說語意網創新關鍵在於以「知識表達」代替檔名，做為網路間訊息交換檢索的依據。而知識內容的瞭解與互通，則要靠每個網路資源上定義知識表達架構的知識本體 (Ontology)。也就是說，未來在語意網上，有無定義完善，而且可由電腦判讀的知識本體，變成了網路資源生存的要件。

本文中將介紹「中英雙語知識本體與領域檢索介面」(網址：<http://ckip.iis.sinica.edu.tw/CKIP/ontology>) 的初步成果。這個成果不但是將來中文網路轉為語意網必須的基本架構。其中的豐富資料亦可馬上應用到當今的網路華語教育。

貳、資料來源

「中英雙語知識本體與領域檢索介面」目前開放的雛型是由「數位典藏」國家型計畫的技術研發分項計畫下「語言座標——參考資源建置與服務」計畫所建構完成。主要計畫成員在中央研究院。所引用的資料除了中央研究院詞庫小組(資訊所)，文獻語料庫 (語言所) 及計算中心開發的資料外。國外的有 IEEE 批准的 SUO 工作小組所建的的 SUMO(Suggested Upper Merged Ontology)知識本體[8.] 及普林斯頓 (Princeton University) 的 WordNet[9.]。國內主要有來自遠見科技股份有限公司 (包括該公司自有資料及與中研院共同開發資料) 以及教育部國語會的辭典...等資料。

WordNet (詞彙網路) 為國際間詞彙知識庫通用的架構，為了結合詞彙編纂資訊和高速運算的所產生更多效益而提出的計畫，由普林斯頓大學認知實驗室在

1985 年開始以現代心理語言學理論所指的人類詞彙記憶為啟發作出的語意式電子字典，將英文的名詞、動詞、形容詞、副詞組織成同義詞集，每個同義詞陳述一種詞彙概念，也就是以字詞意義而非以字形組織詞彙的資訊，並指出詞彙矩陣。[10.]並以同義詞集為對象分析之間包括：反義詞(Antonym)、上位詞(Hypernym)、下位詞(Hyponym)、（整體）部件詞(Part meronym)、（成員）群體詞(Member holonym)、（成分）整體詞(Substance holonym) 等近二十種詞義關係。在多語環境（如 WWW 全球資訊網）中，詞彙網路提供了兼顧概念通性及語言通性的知識表達架構；更因詞彙網路涵蓋了絕大多數富含資訊(information-rich)的語言；正可以擔任不同語言間知識資訊轉換的座標。

SUMO(Suggested Upper Merged Ontology , 建議上層共用知識本體)由 IEEE 標準上層知識本體工作小組所建置。這工作小組的目的是發展標準的上層知識本體，這將促進資料互通性、資訊搜尋和檢索、自動推理和自然語言處理。知識本體(ontology)類似於字典或者術語表，但是能使電腦處理更多內容的細節和其結構。知識本體包括將人們有興趣的領域正規化為一套概念、關係和定理(axiom)。上層的知識本體被限制在 meta 的概念、一般、抽象或者哲學，因此足夠一般提出（在一定水準上）一個涵蓋廣闊範圍的領域區域。特殊領域具體的概念不被包括在上層知識本體中，但是這樣的知識本體確可提供特殊領域(例如：藥、財政、專案...等等)的知識本體結構的建立。[8.]SUMO 希望藉由最高層次的知識本體，鼓勵其他特殊領域知識本體以其為基礎衍生出其他特殊領域的知識本體，並為一般多用途的術語提供定義，另外，目前 SUMO 已經和英語詞彙網路 WordNet1.6 版本作初步的連結。

「中英雙語知識本體與領域檢索介面」的目標是以台灣通用的中文為出發點，建立完整的中文知識表達基礎架構；而為了考慮知識的流通性，及擷取以其他語言表達的資訊的方便性，故採用目前資訊先進國家已建構大量知識資料的 WordNet 架構，但其表達形式及內容，必需考慮中文的語意及邏輯結構特性；故另有相當關鍵的理論問題必需先探討。本雛型所使用的基本資料包括：（1）、使用現有機讀資源（包括電子辭典、同義辭典、英文 WordNet、中文知網(HowNet[11.]) 等等）發展抽取語意關係及雙語（特別是英一中）間詞義關係對應、概念領域等知識的擷取與分類的國際數位圖書館合作研究計畫（簡稱 IDLP）所建構之詞彙網路。（2）、已於 2001 年結束，由中研院語言所與「遠見科技」合作，找出所有英文詞網中意義單位同義詞集(synset)（共有 110,000 個以上）的相對中文對譯概念詞；以及兩個語言相對的詞之間的可能語意關係。

參、「中英雙語知識本體與領域檢索介面」目前提供的功能

一、跨語言資訊轉換：

在 WordNet 的架構下針對每個同義詞集給與三個最常對應的中文，提供中英雙語查詢，使用者輸入中文或英文可找到相對應的翻譯以及詞類。例如輸入檢索詞「夢」，英文對譯為 dream 和 dreaming，詞類為 Nac 可數抽象名詞。

EuroWordNet 完成於 1999 年的夏天，也是用 WordNet 相同的架構建構歐語系囊括荷蘭語、義大利語、西班牙語、德語、法國、捷克語和愛沙尼亞語且以同義詞集表達語意關係的資料庫，每一 WordNet 代表一詞彙化的獨一無二語言-內部系統，再透過內部語言索引（Inter-Lingual-Index）連結在一起，經由索引除了可以達到語言內部連結，還可連結至其他語言中表達相似含意的詞彙。此外，經由索引可擷取 63 語意區別的知識本體訊息，這知識本體提供適用於所有語言的一般語意架構，如果有個別語言的特性則在每個詞彙網路中進行維護。目前 EuroWordNet 資料庫設計、語意關係定義、上層知識本體已經凍結，但是若其他機構以 EuroWordNet 相同規格或在 WordNet 資料庫上增加的其他語言訊息，都可藉由索引彼此連結。她們目前正以相同架構發展瑞典、挪威、丹麥、希臘、葡萄牙、巴斯克、加泰隆尼亞、羅馬尼亞、俄羅斯、保加利亞、斯拉夫等語言的詞彙網路。[12.]

目前以中英雙語查詢為基礎，中文與其他語言轉換為長程目標。因為本雛型介面採用了「詞彙網路」(WordNet)的架構，為國際間詞彙知識庫通用的架構。所以找到了英文對譯詞後，可藉由每個同義詞集的索引和 EuroWordNet 等網路上開放的資料庫，對應到 20 幾種語言。

二、詞義的區分與詞義關係的連結

語言的特色之一是同一個詞可能有好幾個意義（中文的「機關」可以指機構組織，也可以指害人的陷阱等），因此正確解讀需辨明詞義。另一個特色是詞與詞之間有複雜的語意關係。因而產生了許多的替代說法與推論判斷（如某人喜歡籃球，是喜歡這個運動，而非喜歡這個球體。這些都是建立在「籃球」這個詞的詞義與詞義關係上的）。詞彙網路的架構，提供了多重詞義與詞義關係的檢索。當使用者輸入檢索詞將出現的訊息包括：詞義區分、該詞義所屬的詞類（分為名詞、動詞、形容詞、副詞）、該詞義所屬的解釋、該詞義所屬的翻譯、詞義關係（包含詞彙網路 WordNet 近二十種的詞義關係）。

以英文檢索詞 "Computer"，詞義區分為計算專家或者執行自動計算的機器，這兩種詞義解釋，詞類皆為名詞，若作執行自動計算機器的解釋時，則翻譯成「電子計算機」或「電腦」，同義詞有 information processing system, electronic computer, data processor，而 monitor, chip, data converter, computer circuit, processor, hardware, keyboard...等和 Computer 的詞義關係為「（整體）部件」，machine 是 computer 的上位詞，下位詞包含 digital computer, analogue computer, host, client 等。

三、語言資訊與概念架構（知識本體）的連結

SUMO 和 WordNet 已經對應，而 SUMO 的概念(Concept)和 WordNet 的同義詞集之間包括三種關係：（1）、同義（synonymy）表示 SUMO 的概念和相對應的同義詞集相等，如：Plant 植物的概念同樣為 Plant。（2）、上位（hypernymy）表示 SUMO 的概念包含相對應的同義詞集，如：bank 以銀行作解釋時，則屬於

公司 / 法人的概念，以及 (3)、例子 (instantiation) 表示 SUMO 的概念不相等，也不包含對應的同義詞集，如：Underground Railway 南北戰爭前幫助奴隸逃亡的地下交通網，是一種組織 (Organization)。當 SUMO 和 WordNet 兩者之間沒有可對應的 SUMO 概念，則利用平行相關的概念去推論新的，某些同義詞集是相當廣泛無法縮減的，例如：best 另定義兩個一般性的概念加以表達，而若以語言學觀點產生的些許差異，仍給相同的概念，有些同義詞集屬於複合式的概念，若 WordNet 內容而 SUMO 概念並未包含，則利用 WordNet 去修改 SUMO。[13.]

中文詞彙網路以 WordNet 為基礎架構發展，透過 WordNet 轉換，目前本雛型擷取相關資料連結到該詞義在 SUMO 這個上層共用知識本體的相關資料，可輸入中文或英文詞彙，再由每個詞義查到該詞義在概念架構上的歸屬，並提供其概念的中譯資料，利用知識本體架構可作知識內容分類，與簡單推理，如：風箏為一裝置，根據知識本體架構的上層推論風箏是一種人造物的裝置，又哺乳動物 (mammal) 為哺乳類且是溫血的...等。若往下推論，電腦是機器 (machine) 的一種，可知道相關機器種類包括：工程元件、樂器儀器、傳輸設備以及武器。

另外，透過所屬概念可推論該概念的定理 (Axioms) 在句法上的使用方式，譬如：Computer 歸屬於機器 (Machine) 這概念，an **instance of machine** 表示 Computer 在英文句子中他可以搭配著概念為實例 (instance) 的詞彙一起使用，例如：an example of computer。

四、使用領域

詞語因領域不同而有不同的解釋與用法，例如：資訊界的「病毒」是指一種具有類似生物病毒一樣的潛伏、繁殖、傳染、破壞特性的電腦程式，當它侵入系統後，會改變電腦程式的指令，使原有功能失常；而醫學界的「病毒」則是泛指自己沒有生殖能力，只能在活寄主的細胞內增殖的一種極小的微生物，也是引發人類感冒、麻疹、天花等的過濾性病毒的簡稱。所以我們就現有資料 (遠見字典) 依據不同詞義，利用同義詞、下位詞等詞義關係，進行領域猜測的研究，提供了部分的領域標記[14.]。然而以機器自動判斷其正確率往往有限，且領域的定義因使用目的而產生，因此我們特別提供了回饋機制，鼓勵相關領域的專家學者甚至一般使用者上網提供相關訊息，透過使用者貢獻更多的領域標記 (並提供了小禮品)。

此外，領域的使用，也可藉詞彙在不同領域 (包括時代，區域，學門等) 辭典中的分佈判定。因此我們特別提供了每個詞彙在參考資料中出現分佈的訊息。若是該詞彙存在於該資源中，可點選進一步查詢該資源提出的解釋或實例。譬如：檢索詞「電腦」曾出現在「現代漢語八百詞」、「中文詞知識庫」、「國小課本」、「中研院現代漢語平衡語料庫」以及「教育部國語辭典」，點選作進一步查詢，發現「電腦」一詞並未出現在國編版的國小課本，但出現在南一版課本 4 次，康軒版的國小課本則出現 11 次，而且提供課文例句。

(一) 詞彙分部情況：

據詞彙分佈情況可分為三層次依序為核心詞彙、通用詞彙與參考詞彙。英文區分核心、通用與參考詞彙的原則為：(1) 核心詞彙是 BNC, Brown, CIDE 等幾個語料庫取累計頻率 66% 時的所有詞，再把各語料庫的個別詞表交集，得到的結果便是核心詞彙。(2) 通用詞彙：同上，但累計頻率取到 80%。(3) 參考詞彙：各英文資源有收的詞。中文區分的原則參考「資訊用中文分詞標準」[15.]，[16.]，(1) 核心詞彙是指所參考的五本辭典都列的詞且出現在中研院平衡語料庫語料庫 10 次以上。(2) 通用詞彙則收錄在任意三本辭典以上的詞且出現在中研院語料庫 4 次以上。(3) 參考詞彙指的是收錄在三本以上辭典的詞，或收錄在五本辭典中任一本且出現在中研院語料庫 1 次以上，或者是同義詞詞林的標題詞。參考詞彙包含了通用詞彙與核心詞彙，通用詞彙則包含了核心詞彙。

(二)、詞彙比對的參考資料：

進行詞彙比對的參考資料又分為中、英兩種比。中文參考資源有：(1) ABC 辭典(ABC Chinese-English Dictionary)，美國夏威夷大學資深教授 John DeFrancis 編，第一本完全依據羅馬拼音字母順序排列加注拼音且電腦化的辭典，適用時機為聽到不熟悉的詞條時，可經由其發音得知其字元如何書寫。本辭典的詞頻參考了兩岸的語料庫。[17.](2) 「現代漢語八百詞」是呂叔湘的鉅作，現代漢語語法詞典的經典，是學漢語時最可靠的文法參考書之一，也可供一般語文工作者和方言地區的人學習普通話時參考。[18.](3) 「文國尋寶記」針對語言文學學習者設計的，重點對象在華語文學習者，所提供包括國編館、南一書局、及康軒書局三個版本的國小課本。[19.](4) 教育部國語辭典，包括教育部所公布的常用、次常用、及常用異體字共計約一萬四千個詞條。[1.](5) 由中央研究院中文詞知識庫小組執行研究的中文詞知識庫，包含一般用詞、常用專有名詞、成語、慣用語、常用派生詞、異體詞、合併詞以及少數特殊領域用語和古漢語詞語。[20.](6) 「中央研究院現代漢語語料庫」蒐集了分配在不同的主題和語式上的現代漢語，是現代漢語無窮多的語句中一個代表性的樣本。[21.]

英文參考資源，除了上述的「ABC 辭典」，還包括：(1) 英國國家語料庫 (British National Corpus, BNC) 以涵蓋各種範圍的手寫和口語資源中抽樣的一億字，呈現據跨區域特性的現代英式英語語料庫，提供編撰詞典、人工智慧、語音辨識 文學研究 等相關研究的進行。[22.](2) 劍橋國際英語字典 (Cambridge International Dictionary of English)：英語語言的學生和學者可以找到涵蓋美式、英式以及澳洲英語，每個詞條區分詞義、提供在俚語和一般英語使用情況的例句。[23.](3) 布朗語料庫 (Brown Corpus) 包含各種不同文體並加上詞類標記的第一個平衡語料庫，在各種語言學研究上有其不可取代的價值。[24.](4) 知網 HowNet 以漢語和英語的詞語所代表的概念為描述對象，以揭示概念與概念之間以及概念所具有的屬性之間的關係為基本內容的常識知識庫。[11.]

肆、應用範圍

(一) 華語文學習

(1) 詞義區分和詞彙使用：本系統提供中英雙語詞義與知識訊息，因此有助於華語學習。其最大的特色在於詞義的區分清楚；而且對於每個英語詞義與用法，均提供了最準確的中文對譯。例如：學習者可知道中文詞「電腦」是指自動執行計算的機器，相對應的英文有 computer, data processor 等對譯，若針對 data

processor 進行再查詢，並沒有其他詞義，所以在英文句子中，可以以 data processor 代換 computer 使用，但是 Computer 卻不一定可以替換成 data processor，因為英文 computer 有「電腦」和「評估家」兩種中文用法，但是在中文句子中「電子計算機」和「電腦」是完全可以替代使用的，因為兩個詞彙皆只有一個詞義，且互為同義詞。若不清楚在句子中實際運用的情況，透過與其他資源比對，他可以了解該詞彙在不同參考資料中的解釋和用法，這些訊息在華語文學習上可幫助學習者清楚了解每個詞彙的用法，甚至增進寫作以及翻譯能力。

(2) 詞義關係和概念學習：藉由詞義關係，學習者可得知「電腦」是由(整體)部件詞 - monitor (螢幕) chip (晶片) 以及 hardware (硬體) 等所組成，電腦種類包含 digital computer (數位電腦) analogue computer (類比電腦) 等且可能是 host (主機) 或 client (客戶端) 等，「電腦」的上位詞是 machine (機器)。透過 SUMO 再查詢，可知「電腦」是一種機器，而機器是人造的設備。所獲得的知識將不限於語言學門，而是詞彙網路綜合概念發展，如此將有助於學習詞義、領域、概念之間的關係、概念的延伸。

(3) 紀錄學習歷程：比如說，如果某個詞在國小課本中出現，應該是簡單基本詞彙，若是分布情況又屬於核心詞彙，表示為極常用的基本詞彙，則學習的過程可以從國小課本的核心詞彙、通用詞彙到參考詞彙依次進行。除了依照詞彙使用頻率，還可依照出處、概念、領域、詞類，或混合搭配所產生的詞彙表進行教師教授或自習，如此，可確保學習過程的系統性和完整性。

(二)、其他應用：除了華語文學習之外，本離型系統以及產出資料還可運用成為人工或自動翻譯的參考資源、自然語言處理的基礎資料、各異質性系統間的資料交換、電子商務和知識管理包括不同領域資料的應用，因而所需的資料和語意層次的內部運作...等。

伍、結語和未來發展

本離型和一般線上電子字典最大的不同在於針對每個詞彙進行清楚的詞義區分和詞義關係，而藉由詞彙網路與知識本體的介面，更可由英語轉換到其他語言。領域知識的標記，更有助於限定領域的華語學習。最重要的是，這個資源提供了華語文與知識本體間的整體對應，為將來語意網上的華文教育與其他應用，建立了不可或缺的基礎。

未來，除了提供使用者線上提供領域，此外將繼續進行詞義的自動領域猜測的研究。而由於目前的中文限於以英文詞彙網路對應的中文詞彙，因此所提供的中文詞彙有限，以後將針對台灣地區使用的中文詞彙特性發展完整的中文詞彙網路。以及針對語言學習者提供加值詞彙庫。此外，藉由與「數位典藏」各主題計畫專家學者的合作，繼續發展各個不同知識領域的知識本體架構。也可進一步藉知識本體連結到其他語言。整體來說，近程目標：(1) 建立「完整」精確的中英對譯資料庫及檢索介面，作為數位典藏知識國際化的基礎。(2) 逐步建立各領域之雙語領域辭典，達成專門(學術)領域詞彙翻譯之標準化，作為該領域/典藏雙語控制詞彙的參考標準，及具領域判斷能力資訊檢索之依據。(3) 建立帶領域標記之雙語辭典及檢索介面，以加值成為知識加值雙語電子辭典，翻譯資料庫之精準雙語檢索。長程目標則是：(1) 建立精確的領域知識架構，以作為

高加值知識產業的基礎(2) 建立完整的知識本體架構,做為下一代網路(如「語意網」)之語意骨幹(3) 建立以知識為經緯的中英雙語訊息交換平台,作為多語知識處理的憑藉。

陸、附圖



圖 1、首頁



圖 2、中英跨語詞義及詞義關係查詢



圖 3、相關資源比對



圖 4、SUMO 概念查詢

參考文獻

- [1.]教育部重編國語辭典修訂本。<http://140.111.1.22/clc/dict/>
Accessed:2003/04/15
- [2.]教育部國語辭典簡編本。<http://140.111.1.22/clc/jdict/main/cover/jianbian.html>
Accessed:2003/04/15
- [3.]教育部國語小字典。<http://140.111.1.43/> Accessed:2003/04/15
- [4.]教育部異體字字典。<http://140.111.1.40/> Accessed:2003/04/15
- [5.]線上查字典 不花半毛錢。
http://news.yam.com/udnnews/today_focus/news/200203/2002032002150221.htm
Accessed:2003/04/15
- [6.]提姆·柏納李(Tim Berners-Lee)著;張介英 徐子超譯,一千零一網(Weaving

- the Web)(臺北市：台灣商務印書局館，1999 年)。
- [7.] 柏納李，亨德，拉斯萊 (Tim Berners-Lee, James Hendler, Ora Lasslia)；高虹譯，黃居仁審，語意網 (Semantic Web)(台北市：科學人，2001 年)。
- [8.] SUMO (Suggested Upper Merged Ontology) . <http://ontology.teknowledge.com/>
Accessed:2003/04/15
- [9.] WordNet. <http://www.cogsci.princeton.edu/~wn/> Accessed:2003/04/15
- [10.] George A. Miller, Richard Beckwith, Christian Fellbaum, Derek Gross, and Katherine Miller. Introduction to WordNet: An On-line Lexical Database. Cognitive Science Laboratory, Princeton University, August 1993.
- [11.] HowNet Knowledge Database. <http://www.keenage.com/>
- [12.] EuroWordNet. <http://www.illc.uva.nl/EuroWordNet/> Accessed:2003/04/16
- [13.] Niles, I. "Mapping WordNet to the SUMO Ontology." (Teknowledge, 2003.) Technical Report.
- [14.] Echa Chang, Chu-Ren Huang, Sue-Jin Ker, Chang-Hua Yang. "Induction of Classification from Lexicon Expansion :Assigning Domain Tags to WordNet Entries". 19th COLING 2002 Post-Conference Workshop--SEMANET: Building and Using Semantic Networks Processing. Center of Academia Activities, Academia Sinica. Taipei. Taiwan. September 1, 2002.
- [15.] 中文分詞標準計劃相關資料下載區 - 四、成果效益檢討內容
<http://ckip.iis.sinica.edu.tw/ROCLING/4.htm> Accessed:2003/04/15
- [16.] Huang, Chu-Ren, Zhao-ming Gao, Claude C.C. Shen, and Keh-jian Chen. 1998. Quantitative Criteria for Computational Chinese Lexicography: A Study based on a Standard Reference Lexicon for Chinese NLP. Proceedings of ROCLING XI. 87-108.
- [17.] John DeFrancis, Alphabetically-Based Computerized Chinese-English Dictionary (Honolulu: University of Hawai'i Press ,1996)
- [18.] 呂叔湘，現代漢語八百詞 (北京：商務印書館，1980)
- [19.] 「文國尋寶記」 - 「國小課本」。
<http://www.sinica.edu.tw/wen/textbooks/intro.html> Accessed:2003/04/15
- [20.] 「中文詞庫資料 (八萬目詞)」。
<http://ckip.iis.sinica.edu.tw/ROCLING/corpus98/decform.html>
Accessed:2003/04/15
- [21.] 「中央研究院現代漢語平衡語料庫」。
<http://www.sinica.edu.tw/SinicaCorpus/>
Accessed:2003/04/15
- [22.] British National Corpus, BNC. <http://info.ox.ac.uk/bnc/> Accessed:2003/04/15
- [23.] CIDE: Cambridge Internal Dictionary of English.
<http://dictionary.cambridge.org/> Accessed:2003/04/15
- [24.] The Brown Corpus.
http://clwww.essex.ac.uk/w3c/corpus_ling/content/corpora/list/private/brown/bro

wn.html Accessed:2003/04/15