

TR-88-010

A COMPUTER-BASED LIBRARY  
CONSULTANT SYSTEM

書	考	參
借	外	不

中研院資訊所圖書室



3 0330 03 00083 5

0083

76-01-010  
A COMPUTER-BASED LIBRARY CONSULTANT SYSTEM

主持人：鄭國揚

研究助理：林進揚

中央研究院資訊科學研究所

INSTITUTE OF INFORMATION SCIENCE ACADEMIA SINICA

中華民國七十七年六月

## 摘 要

---

這篇論文以新的觀念來檢討傳統圖書館所扮演的角色。

傳統圖書館以保留文化、智慧資源為目的。而工業革命之後，研究風氣盛行，研究成果豐碩，刺激圖書館功能的蛻變，加重了圖書館傳播知識的功能。

“新的觀念”導源於新的科技發明。目前電腦的普遍應用，加上網路相結合，使得資訊的傳播無遠弗屆。社會大眾可藉著與網路相連的終端機來作圖書館資料查詢。可免去以往使用者親自跑圖書館的不便。

本篇論文就是要建一個諮詢服務系統，幫助不熟悉圖書館系統與對其需要的資料有模糊概念的使用者，取得有效的資訊。也就是使圖書館系統具有使用者友善 (user-friendly) 的性質。

# 目 錄

誌 謝.....	1
摘 要.....	2
第一章 緒 論.....	1
第一節 研究背景.....	1
第二節 研究目的.....	7
第三節 各章節簡介.....	7
第二章 圖書的 I.D. Card-Catalog .....	10
第三章 時下系統介紹與國內現況	
第一節 OCLC系統.....	25
第二節 國內人工資訊服務系統之探討.....	31
第四章 本系統構建理論基礎	
邏輯表示知識.....	45
THE FIRST-ORDER LOGIC.....	47
Deductive Relation Data Base .....	51
T- Clause .....	52
Natural Language Qutput .....	56
第五章 諮詢系統.....	59
第六章 結 論.....	69

# 第一章 緒論

## 第一節 研究背景

今日之圖書館已非昔日之藏經閣，科技的進步、資訊的快速成長和索求資訊之殷切，已經改變傳統圖書館的功能，由昔日的蒐集、整理、保存資訊，進而提供完善的參考服務功能，幫助使用者快速的找到想要的資訊。

電腦的發明和應用，以及與網路的結合，影響人類活動甚劇。以下引述(註一)關於電腦科技對圖書館資訊事業的影響。

### 電腦科技的衝擊

1. 電腦資料改善了知識的傳遞和運用。
2. 電腦科技的運用於公眾知識方面，擴大了知識傳播的功能。

(註二)歐美資訊學界人士認為資訊世界可依時間、空間兩因素分為三個資訊世界，這三個資訊世界是：

1. 圖書文獻的資訊世界。
2. 資訊處理的資訊世界。
3. 電訊傳遞的資訊世界。

圖書文獻是資訊世界中最古老的世界，圖書館的各種資料及文獻，就是這個資訊世界中的要角。資訊的利用必須借助於資訊媒體的搬移而為之。這是我們所最熟悉的一種資訊世界，因為它的時效性很長，儲存資訊的媒體也是固定的。

資訊的第二世界是以資訊處理為活動領域的世界，各種資訊服務中心，資訊交換及分析中心(clearing house)，資料處理中心和各種記錄中心(recordscenters)，便是這個資訊世界中的主要活動場所。從傳統的圖書文獻的角度去看它時，原始資料在這個資料處理的資訊世界中，被處理而形成有用的資訊產品。第二種資訊世界裡，它的時效性比較短，儲存媒體也具有可變性。

資訊的第三個世界是以電訊傳遞為主要活動領域的資訊世界，也可以說是一個以電腦科技為基礎的資訊世界。在一個高度發展的工業社會，這種資訊也是一種工業資源(註三)。訊息儲存媒體也可以隨著資訊需求者的需要而變化，例如儲存在電腦磁帶，可顯示在電腦終端機，也可列印在紙上，也可錄製在縮影媒體上。

#### 圖書館自動化的資訊角色

在八十年代裡，自動化已是電腦化的同義辭。圖書館自

動化不僅僅是圖書館本身的管理問題，也是一種社會的需要和不可避免的趨勢。圖書管作業的自動化促使三種資訊世界——圖書文獻的資訊世界、資料處理的資訊世界、及電訊傳遞的資訊世界，更加緊密的連結在一起。由於圖書館自動化的作業後，其各種書目記錄，被整理得比傳統的書目作業更加精細、準確，並且透過電腦與通訊的技術結合，能夠迅速地傳達到無遠弗屆。

#### 國內自動化圖書館

(註四)1981年行政院應技小組『生產自動化指導小組』對國內工業自動化調查結果發現國內工業界最迫切需要的便是技術資訊的服務與設立技術諮詢服務機關，於是在1982年底工研院機械工業研就所(Mechanical Industry Research Laboratory)設立"技術諮詢中心"，為全國工業、學術、研究機構提供自動化技術諮詢服務，為達成該任務，是故該中心集資料蒐集、處理、生產與服務於一體的多功能單位。1983年該中心為配合自動化諮詢服務所需，深感欲提供快速完善的服務，須有良好的自動化圖書館在後支援不行。乃於1984年初又規劃了"自動化圖書館系統"。此系統分三期，第一期計劃是與讀者直接相關的流通控制系統(Circulation

Control System )，這也是目前所完成的，其功能包括借還書處理、續借、預約及各種查詢、統計分析、報表列印等。第二期計劃是發展能處理書、期刊、微縮、視聽等書目資料的系統，也就是希望發展能與中文機讀目錄 (Chinese MARC) 溝通的書目系統。第三期計劃是整合自行發展的工廠自動化 (Factory Automation) 技術、MIS、辦公室自動化 ( Office Automation ) 技術於一體的電腦整合圖書館自動化系統。此系統之理想是，讀者從其擁有的個人電腦或終端機，甚至桌上的電話，透過通訊網路，與已有豐富資料的圖書館電腦主機連繫，並檢索到符合所須的資訊，然後在利用電腦命令機器人到書庫、微縮資料庫或磁帶資料庫中選取所需的資料，在利用無人搬運車自動搬運到需用者旁邊，或者利用電子郵遞、電傳視訊將資料傳給讀者，使讀者能即時獲得所需相關資訊。(註四)

(註五)中國圖書館學會與國立中央圖書館合作，於1980年組織圖書館自動化作業規劃委員會，制訂圖書館自動化作業計劃，此一計劃目標為

1. 訂中文機讀編目格式 (Chinese MARC)。
2. 規劃圖書資料自動化作業系統。



3. 建立中文資料庫。

4. 建立全國資訊網。

其中中文機讀編目格式在此一委員會的中國編目規則研訂小組（藍乾章教授主持），已完成編目規則。將使中文編目標準化，清除了圖書館自動化中文編目的障礙。

### 為什麼圖書館要自動化

圖書館為什麼要自動化？R. T. Kimber在其圖書館自動化 (Automations in Libraries) 一書中提出兩大理由（註六）。

1. 現行制度 ( Existing System ) 的效率逐漸減退。
2. 圖書館業務不斷增加。

他提出一連串的問題來強調他的觀點。

- ： 為什麼每一所圖書館必需保存一套自己的目錄？
- ： 為什麼不積極的朝聯合目錄發展？
- ： 為什麼圖書館目錄中所使用的標題索引和摘要學報中所使用的標題索引不能一致？
- ： 為什麼老是保持傳統的作風讓讀者到圖書館來利用圖書資料？為什麼不將圖書資料送到讀者那裡，豈不是對讀者更方便嗎？

由以上例証，可看出圖書館自動化是必然的，而且將走

向國際化，以達資源共享的境界。同時對使用者服務方面，  
將更朝向使用者友善( User Friendly )方向前進。

## 第二節 研究目的

自動化以前的人工處理系統，使用者可能面臨困擾的是，必須親自上圖書館，及在龐大的目錄卡片櫃上，不知從何找起，或是目錄卡上所著錄資料不夠充足。

圖書館自動化之後，使用者亦可能遭遇到一些困擾。

如：

1. 使用者在終端機上不能隨意閱讀到圖書。
2. 使用者對想要資料的基本知識不足。

解決第一個困擾的作法是，藉著電腦化的便利，在目錄上著錄有關圖書的內容的資料。這又會產生第三個不便，

3. 編目太細且太繁，使用者不易使用，圖書館編目人員亦難適當地編目。

本篇論文是想解決第二與第三的不便，構想是參考人工智慧的專家系統，建一個諮詢服務系統，以簡化編目工作，與作一個很好的使用者介面(interface)。

## 第三節 各章節簡介

第二章介紹電腦化國際化的基石—目錄，再介紹運用電腦科技組合資訊的層次圖。第三章介紹 O.C.L.C 系統與圖

書人工資訊服務。第四章介紹本諮詢系統用到的理論基礎。

第五章介紹諮詢系統架構與細節。第六章作個結論。

- 一. 顧敏, "電腦科技對圖書館資訊事業的影響"。  
BULLETIN OF THE LIBRARY ASSOCIATION OF CHINA,  
NO.35 December 18, 1983
- 二. Evelyn Danies, "1980 Forecast Special  
Librarians to Information Managers" Special  
Libraries, 73 (April, 1982), PP. 93~99。
- 三. Daniel Bell, "Welcome to Post-Industrial  
Society " in Libraries in Post-Industrial  
Society ( Phoenix, Arizona : Oxyx Press,  
1977), PP. 3~7。
- 劉錦龍, "邁向自動化圖書館的第一步"。教育資料  
與圖書館學 NO. 36 December 2, 1984
- 沈寶環, "圖書館自動化問題再商榷"。教育資料與  
圖書館學 NO.35 December 18, 1983
- R.T.Kimber: " Automation in Libraries New  
York ", Bergamon Press 1974 P17。

## 第一節 電腦化國際化的基石

自動化的圖書館資訊系統，至少需要牽涉到三種不同層次的電腦資訊技術。這些電腦資訊技術包括：

1. 電腦處理符碼 (code) 技術。
2. 資訊結構及資料庫技術。
3. 資訊分享系統技術 (註一)。

各層次又分為若干層，就其英文書目資訊系統和中文書目資訊系統的資訊組合技術於圖一。

在第一層，英文之內碼與輸入方法，使用者不必擔心其差異性。中文之內碼亦以標準化，但輸入方法尚多分歧，各種輸入法皆有其優缺點，而且最重要的是皆有人支持它，現今之計，即是提供幾種較多人使用的輸入法。

第二層即是討論到資料庫設計的最基本元素—資料格式 (data format)，這是這一章主題。

圖書館是歷史悠久的機構，其組織結構與功能，亦隨時代的腳步與外界需求的刺激而調整。談到目錄的制作，早期的圖書館只是為了清查館藏，可能只用了連續阿拉伯數字，就敷使用了。隨著圖書館功能轉變—由維持館藏到便利讀者

運用電腦科技組合資訊的層次圖

	組 合 層 次	英 文 部 分	中 文 部 分
一 電 腦 處 理 及 符 碼 技 術 層	第一層組合 機讀二進位數元代碼	EBCDIC, ASCII	中文資訊交換碼, 中文資訊 標準碼, 其他內碼
	第二層組合 電腦用字字集	(1)大寫字母 (2)小寫字母 (3)特別符號	(1)常用字 4807 (2)次常用字 (3)罕用字
	第三層組合 文字排檢方式交互索引	無	中文交互索引方法: (1)字形(主字形, 異體字) (2)字音(注音, 羅馬拼音) (3)代碼(四角號碼, 三角號碼等)
資 二 資 訊 結 構 及 資 料 庫 技 術 層	第四層組合 基本資訊元素	索引典詞彙, 或 目錄款目	目錄款目
	第五層組合 資料關及分欄	美國國會圖書館 機讀編格式電腦 圖書館中心 O C L C 機讀編目中心	中國機讀編目格式
	第六層組合 整份電腦資料記錄	每項英文目錄	中文目錄
三 資 訊 分 享 系 統 及 支 流 網 技 術	第七層組合 資料庫	ERIC, INSPEC等	國立中央圖書館機讀編目 資料庫 國立台灣師範大學教育文 摘資料庫等
	第八層組合 資料庫系統  第九層組合 資訊交流網系統	DIALOG系統  Telenet系統	正發展中  待發展中

<圖一>

<11>

查詢，目錄格式就有著者名稱、書名、標題一一出現。而後為了館務運作如採購、統計分析等目的，就相繼出現出版者、出版年、出版地、及分類等相關資料。但因在電腦網路出現之前，館際的流通較少，因此各圖書館編目人員製作目錄時，各依自己圖書館實際運作需要而對目錄的資料內容有所取捨，而且格式亦無一定標準。當電腦網路實際運行時，自然得對資料庫中資料有一標準格式。

以下簡述目錄標準化的歷史(註二)。美國國會圖書館在1965年為應用電腦處理資料之需，集合圖書館界及電腦界專家共同研究，於1966年推出圖書館編目資料記錄(record)結構，當初稱之為機讀編目格式第一代(MARCI)。1968年又推出第二代(MARC II)。MARC即(Machine Readable Cataloging Format)之縮寫，意思是規定第幾欄位(field)是什麼項目(如作者、書名等等)及佔幾個字元(byte)。美國國會圖書館設計的機讀編目格式簡稱為LC MARC。往後各國紛紛模仿，該一模式來規劃自己本國的機讀編目格式，因而有英國的UK MARC，加拿大的Canadian MARC，日本的Japan MARC，澳大利亞的Australia MARC等等。此時因電腦技術網路尚未成熟，因此各國所定編目格式皆各自為政，只是以電腦處理編目來代替人工製作卡式，其處理範圍限於一圖書館。



愈客觀環境成熟，即電腦硬體、軟體進步，與網路的應用。國際圖書館協會鑑於電腦化的最終目的，乃在於資訊的交換與共享，不同的機讀編目格式阻礙了此項目的達成，因而集合各國專家一起研究，終於在1977年推出國際機讀編目格式( Universal Machine Readable Cataloging 簡稱 UNIMARC)作為各國機讀編目資料交換與共享的規範。我國中國圖書館學會與國立中央圖書館首先成立“圖書館自動化作業規劃委員會”，並於民國69年5月組織“中文機讀編目格式工作小組”，經八個月，在民國70年2月完成“中文圖書機讀編目格式”的規劃工作(註三)。

UNIMARC 所規劃的格式為各國交換的標準格式，而各國或各館尚可依自己本身需要而增減欄位，但交換時得轉換為標準格式。

以下是中文圖書機讀編目格式，而依其項目之增減而有三種目錄，即簡略目錄、標準目錄、詳細目錄。如只記載正題名、第一著者敘述、版本敘述、出版地、出版者、出版年、稽核項者為簡略目錄。

標目

正題名 [資料類型標示] = 並列題名：副題名 / 著者敘述；其他次要著者敘述。 - - 版本敘述 / 關係版本之著者敘述。 - - 資料特殊細節。 - - 出版地：出版者，出版年（印製地：印製者，印製年）  
數量（面、葉、冊數）或其他單位：插圖或其他稽核細節；高廣尺寸 + 附件。 - - （集叢名；集叢號，副屬集叢名；集叢號）  
附註項  
標準號碼：獲得方式

1 標題      2 標題      3 標題      I. 副款目      II. 副款目  
III. 副款目

以上概略敘述編目的歷史演進，至於其詳細內容則涉及到編目的專業技術，因各項著錄皆有很多規則可循。而從其著錄項目與讀者檢索較有關的，可分為第一種，作者名、書名；第二種，是標題、分類號。

（註二）西文系統中，採用書名、作者名或兩者合用作為查詢根據，常見形式有如下數種：

1. 全銜形式，即以所屬欄位全部數據作為檢索根據。
2. 節縮形式，即依所屬欄位中前數個字的前幾個字母，組成的單元，作為檢索的依據，稱之為查詢鍵 (search key)，例如採 4, 1, 1, 1 者，即以第一個字的前四個字母，及第二至第四個字的第一個字母組合而成。節縮式的形式很多，須視實際測試情況而定。
3. 採節縮或全銜形式，但介係詞及冠詞不計。

4. 採語音代碼(Phonetic-Code)形式，即將字母數字化以編成查詢鍵。

以上方法中，除第一種是唯一匹配(exact match)之外，其餘各種皆有重覆率——即一查詢鍵(search key)對應到很多資料，必須再次的選擇需要的資料。而其優點是可不必要作者名、書名完全拼對。而其採用何種設計形式，就得作樣本測試。

中文編目在尚未電腦化時，採用筆畫數作為檢索根據。目前中文機讀編目格式所採用的查詢鍵方式，書名、作者名均取前面三個字(註四)。

根據(註二)所作的抽樣記錄，母體是“中華民國出版圖書目錄彙編”第三冊，樣本是從母體上每頁上、下各一書，總計共2746本。算出樣本中文書名的平均用數字為5.6字。而且採前三字作為查詢鍵時，祇有9.79%能夠完全表達出書名(表二)，表二中縱軸類別，是採用杜威十進制分類法，以下有此分類法敘述。採前三字作為查詢鍵時，重覆率高達11.54%(表三)。統計以辭的用字數，發現二字形成一辭者有1664次，三字形成一辭者有392次，四字者有257次，五字者有45次，平均每一辭的字數為2.4字。為了打破辭的形成，因此試著以奇數字及偶數字各取前三字為查詢鍵，其重覆結

表二 中文書名字數統計表

書名字數 類別 (種數)	一	二	三	四	五	六	七	八	九	十	十一	十二	十三	十四	十五	十六	十七	十八	十九	二十	二十一	二十二
000(196)		3	2	49	24	34	28	26	8	8	10	3										1
100(137)		2	9	42	26	23	16	12	2	3	2											
200(99)		3	5	29	22	18	5	8	2	2	5											
300(216)	1	5	17	46	45	39	19	13	4	9	16	1										1
400(374)		6	29	81	68	72	44	25	25	9	13	2										
500(524)		6	18	92	87	105	66	54	34	22	31	9										
600(205)		2	12	77	40	42	13	11	4	1	3											
700(185)		4	14	47	41	25	22	21	5	2	4											
800(663)	8	50	64	189	129	102	47	31	22	12	8											1
900(145)		3	6	22	23	35	19	17	12	2	6	1										1
總計(2,746)	9	84	176	674	505	495	279	218	118	70	98	16										4
百分比	0.32%	3.05%	6.40%	24.50%	18.39%	18.02%	10.16%	7.93%	4.29%	2.54%	3.56%	0.58%										0.14%
累計種數	9	93	269	943	1448	1943	2222	2440	2558	2628	2726	2742										2746
累計百分比	0.32%	3.38%	9.79%	34.34%	52.73%	70.75%	80.91%	88.85%	93.15%	95.70%	99.27%	99.85%										100%

表三 中文書名查尋鍵全銜組合重複情況

字序 數量	第一字	第二字	第三字	第四字	第五字	第六字	第七字
有重複的總量	2266	1216	317	187	34	15	2
有重複的百分比	82.52%	44.28%	11.54%	6.80%	1.23%	0.54%	0.07%

表四 書名查尋鍵奇數位組合重複情況

奇數位組合數 數量	一字組 (1)	二字組 (1,3)	三字組 (1,3,5)	四字組 (1,3,5,7)
有重複的總量	2266	428	47	4
有重複的百分比	82.52%	15.58%	1.71%	0.14%

表五 書名查尋鍵偶數位組合重複情況

偶數位組合數 數量	一字組 (2)	二字組 (2,4)	三字組 (2,4,6)	四字組 (2,4,6,8)
有重複的總量	2252	414	37	7
有重複的百分比	82.01%	15.07%	1.34%	0.25%

果降至1.7%以下(表四,表五),較取三字者理想很多。

以下談第二種分類號、標題。

### 分類號

(註六)杜威氏圖書十進分類法自1876年問世以來,一百年間不僅美國各類型圖書館紛紛採用,世界各國亦極普遍使用,原因是對知識之組織體系完整。據調查(註七)台灣省八十二所大專院校中,有五十五所採用杜威法整理所藏西文圖書資料,佔66%強。又專門圖書館方面(註八),在八十三所中,有三十七所採用杜威法,佔44%強。

民國六十三年十二月中華民國科技圖書館及資料單位館際合作組織在當年中國圖書館學會第二十二屆會員大會中提出議案,建議中文圖書資料分類法,納入國際分類制度內,於六十四年召開會議,決定將杜威法十八版純粹科學(500)與技術(600)兩類譯成中文,並採分擔方式,工作分配情形為:

- 510 數學, 國立清華大學
- 520 天文, 國立台灣大學
- 530 物理, 國立台灣大學

圖書分類的目的，是希望將所有同類的書集於一處。圖書及其它資料的分類，主要是依其內容的主題分。杜威分類法是以學科為主，同一學科的主題分在一處。如插花的書，分到藝術類裡，植物學的書分到植物學裡，以花為題材的詩，則分到文學類裡，所以同樣以花為主題的書，由於學科不同，分類號碼亦不同(註五)。

杜威十進分類法，是以數字為標記，數字是以十進為原則，所有圖書，依學科性質由0到9分成十個大類，如：

000	總類	500	純科學
100	哲學	600	技術
200	宗教	700	藝術
300	社會科學	800	文學
400	語言	900	史地

原則上每一個大類，可依其學科性質，分成十個小類，每個小類，可再分成十個細目，這樣一層一層的細分下去，實際上視學科的繁簡而定。細分之後的第三與第四位數之間，應加一小數點分開。如：

600	技術(應用科學)
610	醫藥
620	工程總論

## 621 應用物理

### 621.1 蒸器工程

論及二個主題以上的書，則應列入一個較大的類裡。

以下談國內“中文資訊科技圖書資料分類表草案”(由謝清俊、楊鍵樵、黃克東、張仲陶所擬)，(註九)。引用他們的前言，“任何一種科技圖書資料分類表，決不可能由極少數的人及在極短促的時間內完成，尤其對於一種正在萌芽而急速發展的新科技，如欲其分類完備，必需由多數志同道合的同仁們，投入更多的智慧、經驗與時間，更需無數次反覆再三的實際測試與修正，始克完備。以下述及其分類表之目的。

1. 作為擬訂研討中文資訊科技的主題參考。
2. 作為擬訂研討中文資訊科技圖書與資料分類之參考。
3. 作為擬訂研討中文資訊科技人力規劃與訓練項目之參考。
4. 作為從事中文資訊各單位據以研究與發展業務之參考。

以下是分類表一大類

- 100 中文資訊總論
- 200 中文資訊有關標準



- 300 中國文字(有關中文電腦部份)
- 400 中文電腦硬體
- 500 中文電腦軟體
- 600 中文資料通訊
- 700 中文資訊應用
- 800 中文資訊人力資源
- 900 中文資訊研究與發展

以下一大類作說明

- 500 中文電腦軟體
  - 502 系統分析
  - 506 系統設計
  - 512 程式設計
  - 513 編譯程式
    - 521 ADA
    - 522 ALGOL
    - 523 APL
    - 524 ASSEMBLY
    - 525 BASIC
    - 526 COBOL

540	軟體組件
544	介面標準
560	各種機讀格式
:	
:	
570	資料結構
578	資料庫
582	作業系統
586	演算方法

如引言所說，分類工作需要專家學者，群策群力，才能使分類臻於至善境地。

#### 標題：

所謂“標題”(Subject Heading)的意義，是用一名詞、短句、或片語，標出一書本所敘述的主題，同時將相關的參考資料集中在一起。

標題目錄的編製，必需要有標題總錄(List of Subject Headings)做為依據，其功能如分類表上之分類名稱，列出

可用之辭彙。

標題與分類之差異點，分類表較易查詢，使用者根據想要資料對照表上分類，即可查詢。而且分類表可有主題縱的關係，如杜威分類表所謂細分，即是一主題的細分。而標題沒有保有主題縱的關係，即其用語規則，使用者亦不甚清楚且其排列依據字母順序，因此使用者想要查詢就得猜標題用語，再根據字順序去查詢，假如找不到資料，可能標題用語不一致，那可就得大費周章了。

- 註一. 顧敏, "電腦科技對圖書館資訊事業的影響".  
教育資料與圖書館學 NO.35 December 18, 1983
- 註二. 黃鴻珠, "機讀編目格式之認識與應用".  
教育資料與圖書館學 NO.3 SPRING 1983
- 註三. 中文圖書機讀編目格式 第一版及第二版 圖書館自  
動化作業規劃委員會中文編目格式工作小組 台北  
國立中央圖書館發行 民國七十年。
- 註三. James, L linas, " A Method for Evaluating  
Seach Key Performance Dissertation ", State  
University of New York at buffalo, 1970。
- 註五. 梁津南編譯, "杜威十進分類法", II 版。
- 註六. 中國圖書館學會分類編目委員會編譯, "杜威十進分  
類法", 台灣學生書局印行, 民國六十七年。
- 註七. 王振鵠, "台灣大專圖書館現況之調查研究", 圖書  
館與資訊科學第二卷第一期, 民國六十五年四月。
- 註八. 藍乾章, "二十五年來的專門圖書館", 中國圖書館  
學會會報第二十九期。
- 註九. 謝清俊、楊鍵樵、黃克東、張仲陶, " 中文資訊科技  
圖書資料分類表( 草案 ) "  
BULLETIN OF THE LIBRARY ASSOCIATION OF CHINA,  
NO.34 December 5, 1982

### 第三章 時下系統介紹與國內現況

#### 第一節 OCLC 系統

OCLC 是美國四大圖書館網路之一，其成員大約有 4800 個圖書館（在1985年底），且每月以25-30新成員加入。而其成員包括，澳洲、加拿大、丹麥、芬蘭、法國、愛爾蘭、墨西哥、荷蘭等等。以下列述其系統組成。

##### 1. 線上聯合目錄（The Oline Union Catalog）

大約有11.5百萬個目錄（1984年底）。這些目錄是來自 OCLC 的組成圖書館，而每一目錄後有一串記錄擁有該項資料的各圖書館名稱。

##### 2. 目錄查詢子系統（The Cataloging Subsystem）

提供使用者查詢功能。其出現資料格式可依各館運作需要而作編輯。

##### 3. 館際互借子系統（The Interlibrary Loan Subsystem）

透過 OCLC 網路系統，填寫、傳輸借單皆在線上完成。每月平均大約有35,000次，平均六天就可把資料寄到借者。

##### 4. 採購子系統（The Acquisitions Subsystem）

圖書館所有資料的採購工作，包括訂單、驗收、與對書商的抱怨 ( claims ) 等等。

5. The Name-Address Directory
6. The Name-Address Directory 是一個檔，包含有圖書館、出版者、批發商和有關團體的一些資料，以供彼此連繫用。

1983年，OCLC 增加了 CJK Software ( Chinese , Japanese, and Korean )，把非羅馬語系資料納入系統。CJK 系統包括

- a : CJK Library Support
- b : Chinese Language Word Processor
- c : Japanese Language Word Processor
- d : Korean Language Word Processor

這四個系統各自獨立。CJK Library Support 提供使用者與 OCLC 的目錄查詢子系統介面，陸續開發與各個子系統相連。CJK Library Support 有以下特色：

1. 它在 OCLC M300 工作站上操作。
2. 它使用標準英文鍵盤。
3. 它提供有字形輸入法與拼音輸入法。

4. 它提供 OCLC 查詢的功能。

5. 它能印出目錄卡。

CJK 資料庫的資料來源有，一．美國國會圖書館，二．中華民國中央圖書館，三．OCLC 另外之成員。目前約有 21,725 中文，28,217 日文，4,031 韓文資料。(註一)

我國圖書館事業在近三十年來，圖書館數量及典藏資料量成長有相當大的進步。但圖書館提供服務仍停留在書目、索引等工具書之階段，加上圖書館從業人員幾乎全是社會科學之背景，對科技資料，先天即有障礙存在。在有識之士提倡下，於民國六十八年由國際電信局透過人造衛星線路引進了美國洛克希德公司與系統發展公司 (Lockheed and System Development Corporation) 二系統共有一百二十六個資料庫，系統發展公司推出之 "ORBIT" 系統有五十七種資料庫，內含幾乎涵括了全部知識領域，如科學、工藝、醫學、農業、人文社會藝術、商業、公共事務等等。接用 DIALOG 系統者有國科會科資中心，ORBIT 者有師大圖書館。

國科會科學技術資料中心於民國六十九年十月開始對外進行此項服務，推廣服務期間是免費，底下二表是其初期統計資料。其中有 10 種資料庫在這段時間使用次數超過 10 次，還有 50 種使用次數介於一至四次之間。使用者以單位分，以

項目 年月	申請 人數	檢 索 主 件 數	使 用 資 料 種 別	使 用 資 料 次 數	檢 索 得 料 數	線 上 印 出 篇 數	線 印 篇 數	外 出 數	線 上 原 文 篇 數	連 線 時 間 ( 小 時 )	諮 詢
69.10	8	11	18	25	558	247	87			3.929	52
11	8	10	16	25	3,381	93	3,175			2.743	26
12	31	34	26	57	17,100	164	16,578		12	5.243	23
70. 1	8	9	17	33	1,028	2	748			1.346	14
2	8	14	13	26	1,125	30	1,006			1.849	15
3	13	24	28	63	5,832	447	5,452			4.023	19
4	7	12	16	40	5,247	0	5,247			2.072	17
5	12	13	31	54	2,949	34	2,915		30		15
合 計	95	127	165	323	37,220	617	35,208		42	21.21	181



另還有一資料庫使用統計情形超過十次者如下：

檔號	資料庫名稱	使用 次數	檢 索 篇 數	印出篇數 (線上)	印出篇數 (線外)
2	CA Search 67-71	13	819	14	805
3	CA Search 72-76	25	2316	30	2053
4	CA Search 80-pres.	32	3753	93	3386
5	BIOSIS preview 74-pres.	15	898	49	849
6	NTIS	16	16592	6	16586
13	INSPEC 78-pres.	13	729	87	642
55	BIOSIS preview 69-73	10	150	4	146
72	EXCEPRTA MEDICA 80-present	21	161	21	140
73	EXCEPRTA MEDICA in process	17	87	1	86
104	CA Search 77-79	22	2589	1	2588
172	EXCEPRTA MEDICA 74-79	21	1257	17	1240

榮民總醫院、長庚醫院、國防醫學院單位使用次數較多，所以醫學界較常利用此系統檢索資料。

師大所接之"ORBIT"系統，目前共有五十七個資料庫，每週新增四萬五千筆資料，但是據師大在中國圖書館學會中所作使用情形報告，"有時竟月無一人使用"，可見系統之推廣亦是不易之事。（註二）

## 第二節 國內人工資訊服務系統之探討

“資訊服務”是針對讀者的各別興趣，從新資料中選出與其興趣相關的資訊，主動且適時地提供給讀者。這也是圖書館除了傳統的服務方式之外，積極的展開主動而新穎的“現況知曉”(Current Awareness Service)服務之一種。而推行“諮詢服務”必須包含六個要素：

1. 讀者：針對一般從事研究工作或政策的決定者，為了工作或志趣必須不斷獲得新知，可能無暇常到圖書館查閱資料。
2. 讀者興趣檔(Interest Profile)：記載讀者興趣之資料檔案。一人一卡，註明讀者姓名、單位、研究主題、研究計劃名稱、相關的術語、語文限制、必讀期刊等等基本資料。
3. 資料過濾人員(Scanner)或機器：
  - a：機器式：利用電子資料處理設備將讀者興趣檔中的術語事先儲存在電子計算機中，日後與資料庫中的文獻術語，配對選出符合讀者興趣的資料。
  - b：人工式：由資料過濾人員，熟記興趣檔資料，在廣

泛的資料中，研判選擇讀者所需的資料。

4. 資料的通知與分送：定時將彙集之資訊，通知或寄送方式，交予讀者。

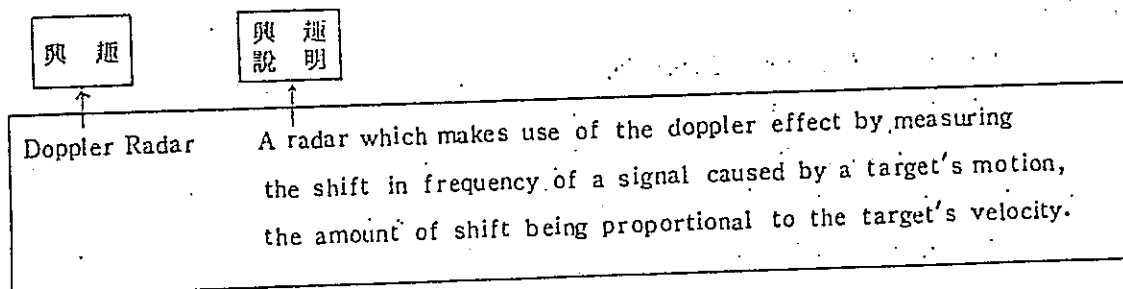
5. 後續服務 ( Back-Up Service )：依據讀者勾選之需求提供讀者需要的原件。

6. 回饋 ( feed back )：讀者參閱所供應的資訊後，一種評估反應。

關於讀者，可以透過圖書館通報、圖書館指南、訪談、專題介紹，向讀者推介。

至於讀者興趣檔，建立方式可採取不同方式，同時得依據需要更新。可有如下：

1. 要求讀者將其興趣以一段字來描述，如此可對其興趣能有較深入的瞭解，如下：



讀者以一段文字來描述其興趣

2. 提供讀者一份主題表，給讀者選勾。

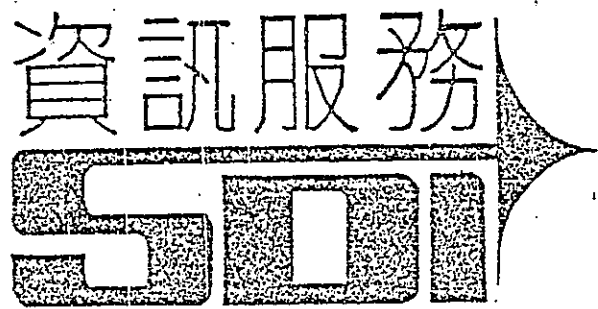
關於資料過濾人員，應對讀者興趣有相當程度瞭解，且

對某一主題應有基本的知識背景。

通過後續服務、回饋，可決定讀者興趣檔是否應更新，  
以下是關於遺漏重要資料、所選資料與讀者不相關、讀者已  
經閱讀過的分析流程圖。

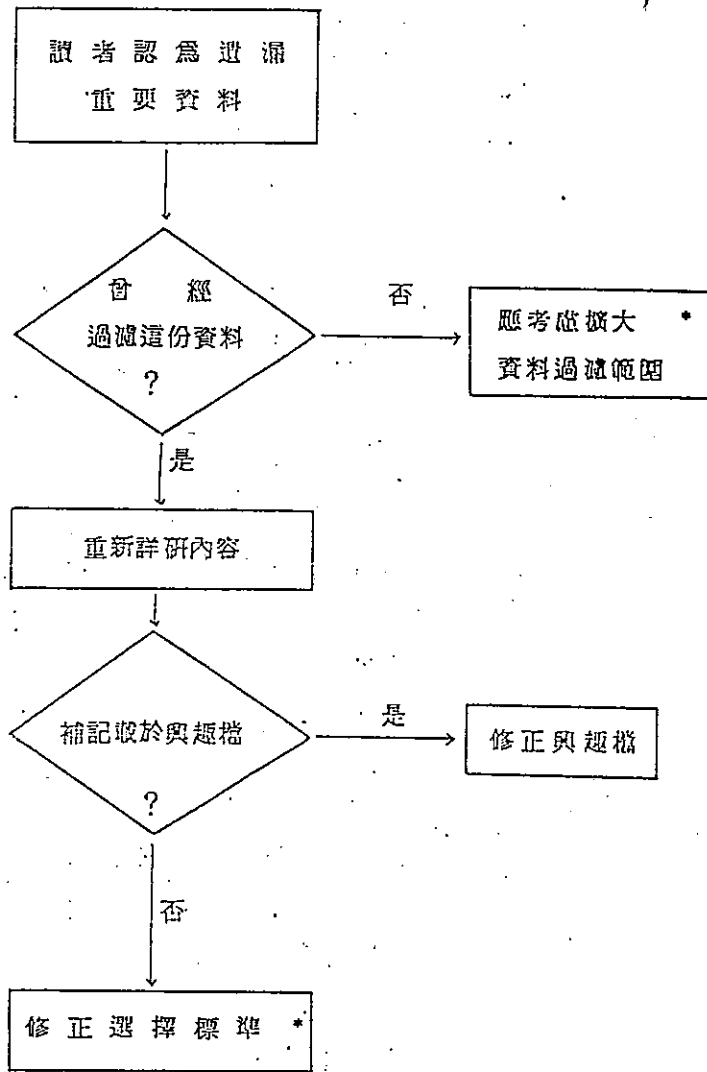
中山科學研究院於民國六十五年八月籌備“圖書館提供  
資訊服務計劃”，成立資訊服務組。同年十月依計劃先行對  
高階層決策主管服務。因試行成效良好，於民國六十六年八  
月，先以圖書館顧問委員為對象，進行推廣，採取自由申請  
加入方式，有以個人為單位，亦有以團體為單位。根據上述  
人工資訊服務系統之六大要素，以下圖表敘述中山科學研究  
院資訊服務組之工作內容。（註三）

# 資訊服務



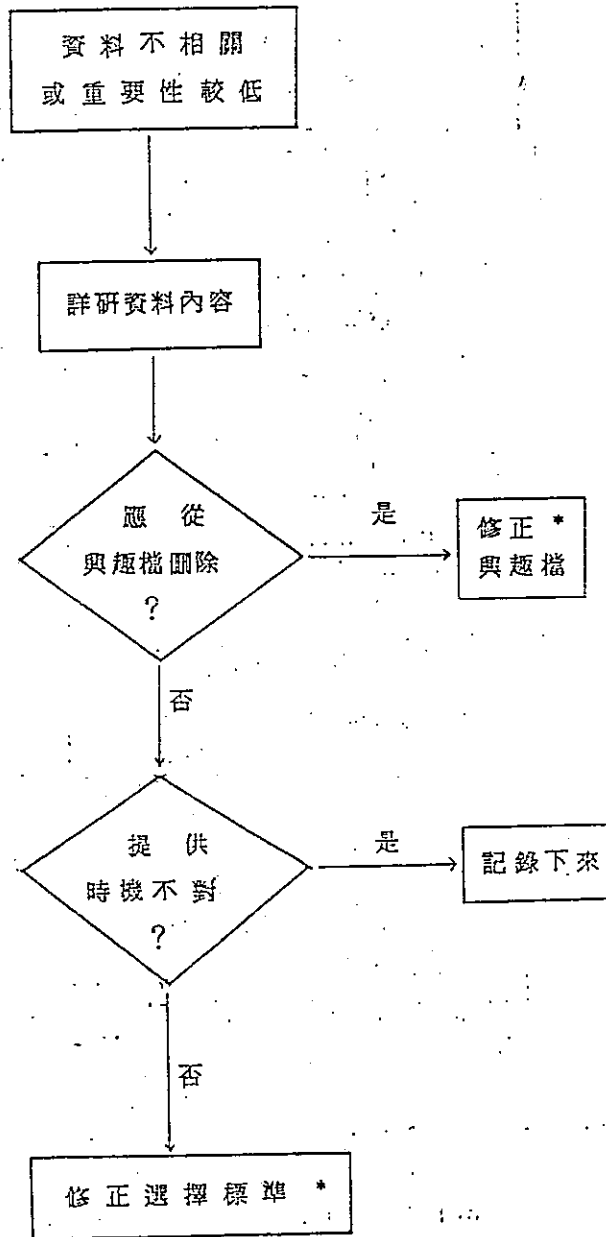
簡介

中山科學研究院圖書館  
資訊服務組



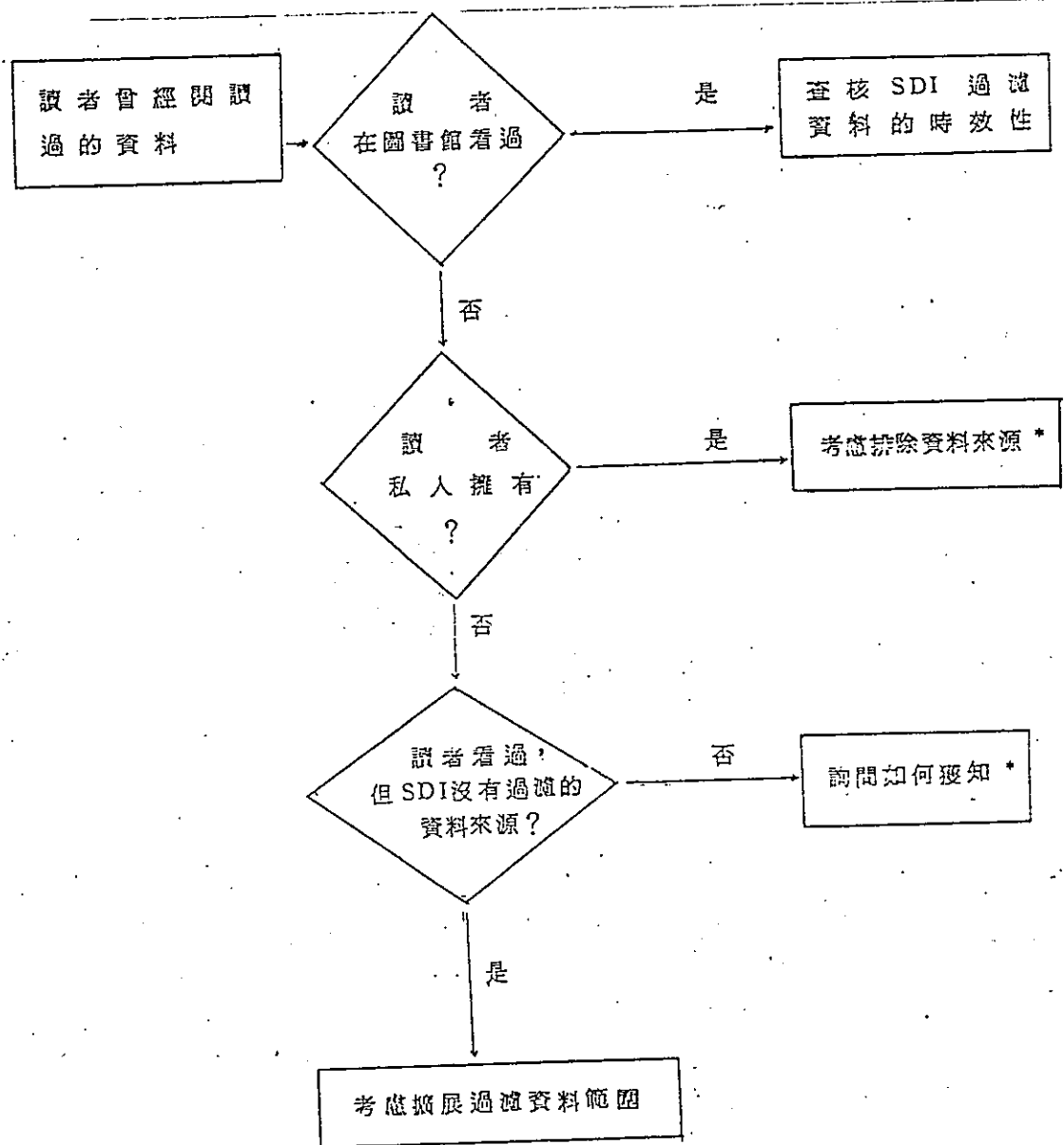
\*與讀者討論

遺漏重要資料的分析流程圖



\*與讀者討論

所選資料與讀者不相關的分析流程圖



\*與讀者討論

讀者已經閱讀過的分析流程圖



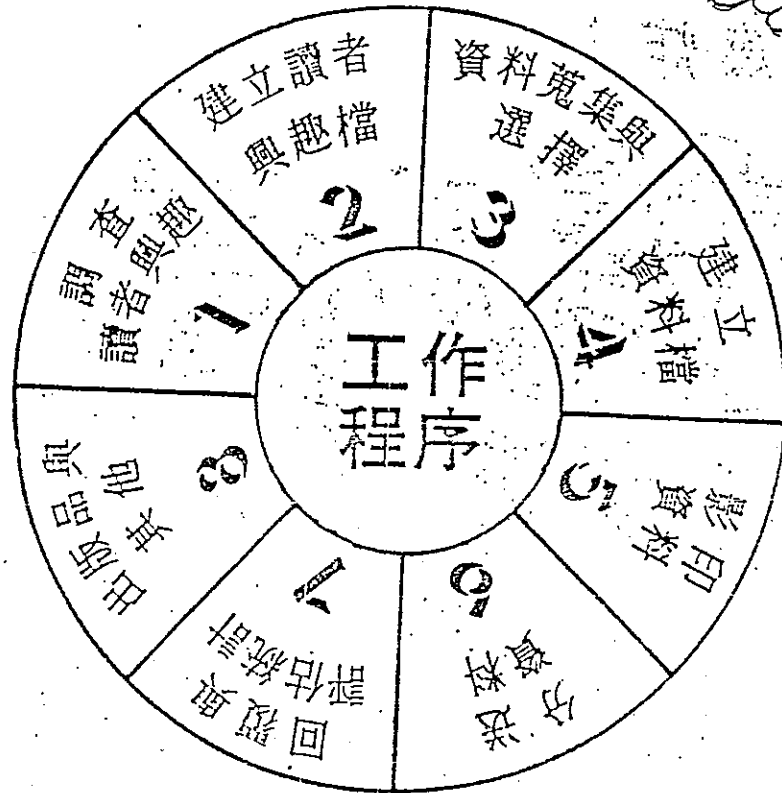
# 資料爆炸！



## 資訊服務 (*Selective Dissemination of Information*)

宗旨：配合讀者研究需要，主動且適時地提供資料。

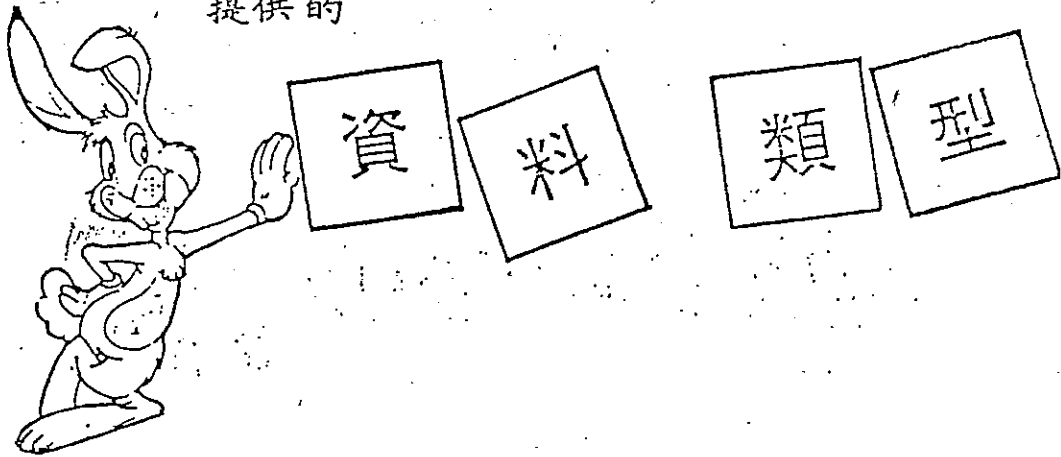
目的：節省讀者檢索資料的時間，加強館藏的流通利用。



收費

工本費 + 服務費

提供的



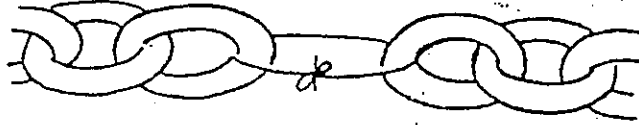
**原件資料：**工作人員確定某篇文獻與讀者研究工作相關時，即直接影印原文並附回覆單。

**書目資料：**資料的篇幅太長或無法判別讀者是否需要，即列出資料的篇名、作者、出處及摘要等書目性資料，製成書目單，供讀者按興趣程度選勾。

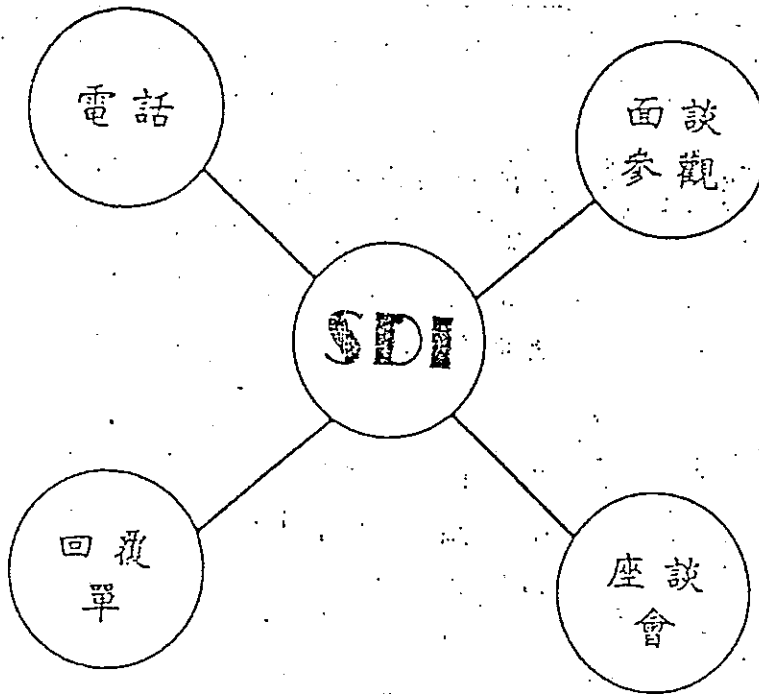
**目次表：**各學科重要的書刊資料及讀者指定需要閱讀的期刊，即影印目次表，供讀者選勾需要影印原文者。

**消息性資料：**凡是報導性的短文，直接影印分送讀者。

SDI



讀者



研究主題：		代 號：F 222
1. MIS (Management Information System)		人 名：趙 世 強
2. Data Base		組 別：六所五組
3. Quality Control		電 話：7962
		聯 絡 人：劉 明 德
		語 文：中、英、日
術 語	同 義 字 及 相 關 語	說 明
Computers	Software in Management	
MIS (Management Information System)		
Data Processing	Data Base	
Quality Control		The application of the computers

個人興趣檔

<41>

資訊服務提供書目資料及評核表

編 號 71,-5,20

姓名代號 趙世強

回 覆 單

編 號 71, -5, 20

姓 名 趙世強 F222  
(代號)

注意事項：

1. 請於興趣程度與資料需要欄內，各擇一項作“√”記號，  
撕下回覆單寄回。
2. 參研興趣若有更改，或有疑問，歡迎以電話聯絡。(2214 資訊服務小組)

項 目	書 目 資 料	興 趣 程 度			資 料 需 要			
		極 度 相 關	有 興 趣	略 有 興 趣	無 興 趣	需 要 原 件	需 要 摘 要	不 需
	FINITE ELEMENT METHOD  On a finite element method for a certain class of timedependent problems. Nemethy, A. Numerical Functional Analysis and Optimization, V2N5, 1980, p. 331-345. (QA1/NFAODU).  Abstract : The finite element method is applied to solve a linear initial-boundary value problem. The basic idea is to combine this method for a discretization in space varivables with the Laplace transform technique for a time variable. Formulation, existence and uniqueness of a weak solution is investigated. The conver- gence and the rate of convergence of the proposed approximate solution is disucussed.							

書目資料回覆單

資訊服務評估統計表

讀者代號	回覆日期	供應數	原 件					書 目					供應數	需要數	影印篇數	影印頁數
			興 趣 程 度				供應數	興 趣 程 度								
			3	2	1	0		3	2	1	0					
A111	71.5.25	5	1	3	1	0	0	0	0	0	0	2	2	4	152	
A113	71.5.27	3	2	0	1	0	1	0	1	0	0	3	2	3	65	
A141	71.5.25	2	1	0	0	1	2	1	1	0	0	5	4	10	78	
A202	71.5.24	2	0	1	1	0	0	0	0	0	0	4	4	6	50	
A212	71.5.30	6	3	2	1	0	1	0	0	0	1	3	1	1	70	
A213	71.6.-2	4	2	0	1	1	2	1	0	1	0	0	0	0	75	
A214	71.5.27	4	3	1	0	0	5	3	1	1	0	1	1	3	45	
A231	71.5.26	7	4	3	0	0	3	1	1	1	0	10	8	18	186	
A234	71.6.-3	1	0	0	1	0	5	2	1	1	1	5	5	9	35	
A311	71.6.-3	15	10	2	1	2	2	1	1	0	0	7	7	18	190	
A323	71.5.30	1	1	0	0	0	7	3	3	1	0	2	1	1	31	
A333	71.5.29	35	2	33	0	0	5	5	0	0	0	12	10	25	255	
A432	71.6.-8	25	15	5	3	2	4	2	1	1	0	3	3	6	150	
A511	71.5.26	1	0	0	1	0	10	4	4	0	2	5	5	12	53	
A532	71.6.-7	6	5	0	0	1	9	3	4	2	0	1	1	2	70	
B121	71.5.23	3	3	0	0	0	1	0	0	1	0	7	5	9	75	
B212	71.6.-2	13	10	1	2	0	2	0	2	0	0	0	0	0	66	
B315	71.6.-4	0	0	0	0	0	6	3	3	0	0	2	2	5	16	
B413	71.6.-1	4	1	2	1	0	3	1	1	0	1	8	7	16	95	
B523	71.5.22	11	7	1	3	0	5	5	0	0	0	20	20	61	333	
B531	71.5.31	1	0	0	1	0	0	0	0	0	0	0	0	0	10	
B551	71.6.-5	8	7	1	0	0	2	1	1	0	0	3	3	8	56	
C132	71.6.20	9	9	0	0	0	3	1	2	0	0	15	15	71	401	
C202	71.6.-2	2	0	1	1	0	2	2	0	0	0	1	0	0	44	
C344	71.6.-1	4	1	2	1	0	5	2	2	0	0	5	5	12	67	
小 計																
累 計																

月統計表

註一： " OCLC AND ITS CJK SOFTWARE ", David Y. Hu  
Journal of Educational Media and Library  
Sciences 23:1 ( Autumn 1985 )

註三： 魏瓊釵、王愛玲、曾淑賢、陸敏， " 資訊服務人工  
系統之探討 "



## 第四章 本系統構建理論基礎

### 邏輯表示知識

The Propositional Logic: A proposition is a declarative sentence that is either true or false, but not both.

邏輯可視為一種語言，不僅處理數學問題，日常事物亦可。

例： " Snow is white. "

" Sugar is a hydrocarbon. "

" Smith has a Ph.D. degree. "

皆是命題 ( Propositions )

底下以大寫字母代表命題

p : Snow is white,

Q : Sugar is a hydrocarbon,

R : Smith has a Ph.D. degree.

P, Q, R 叫做 atomic formulus 或 atomics.

Definition: Well-formed formulas in the propositional logic are defined recursively as follows:

1. An atom is a formula.
2. If  $G$  is a formula, then  $(\sim G)$  is a formula.
3. If  $G$  and  $H$  are formulas, then  $(G \wedge H)$ ,  $(G \vee H)$ ,  $(G \rightarrow H)$ , and  $(G \leftrightarrow H)$  are formulas.
4. All formulas are generated by applying the above rules.

Table 1 是  $G, H$  命題與組合命題的真值表。

$G$	$H$	$\sim G$	$(G \wedge H)$	$(G \vee H)$	$(G \rightarrow H)$	$(G \leftrightarrow H)$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

Logical Consequences: 根據幾個命題推導一個命題的真假。

例:  $P$ : 利率升高

$S$ : 股票下跌

$U$ : 大部分的人不高興

1. 如果利率升高, 那麼股票下跌。  $\implies P \rightarrow S$

2. 如果股票下跌，那麼大部分的人不高興。  $S \rightarrow U$
3. 利率上升。  $P$
4. 大部分的人不高興。  $U$

我們想知道 1,2,3 命題皆真的情況下，4 命題是否真。

$$\begin{aligned}
 ((P \rightarrow S) \wedge (S \rightarrow U) \wedge P) &= ((\sim P \vee S) \wedge (\sim S \vee U) \wedge P) \\
 &= (P \wedge (\sim P \vee S) \wedge (\sim S \vee U)) \\
 &= ((D \vee (P \wedge S)) \wedge (\sim S \vee U)) \\
 &= (P \wedge S) \wedge (\sim S \vee U) \\
 &= (P \wedge S \wedge \sim S) \vee (P \wedge S \wedge U) \\
 &= P \wedge S \wedge U
 \end{aligned}$$

因為命題 1,2,3 為真，那麼  $P \wedge S \wedge U$  為真，那  $P, S, U$  皆真。命題4為真，命題4由命題1,2,3推導而來，所以稱之 (logic consequences) 邏輯結果，對命題 1,2,3 而言。

The First-Order Logic :

很多推導，命題還是有不足之處，如下例：

" Every man is mortal. "

Since Confucius is aman, he is mortal.

上面的推導應是正確的，但如果我們這樣表示

$P$  : Every man is mortal.

Q : Confucius is a man.

R : Confucius is mortal.

R 無法由 P, Q 推導而來。

我們引進 first-order logic, first-order logic 包括三個要素

1. terms
2. predicate
3. quantifiers

先解釋 predicate, predicate 是一個關係 (relation)。

例如, " X is greater than Y "

可表示 Greater (X, Y), Greater 是一個關係, X, Y 是變數 (variables)。P 是一個 Predicate, 如果 P 有幾個參數 (arguments),  $P(t_1, \dots, t_n)$  叫作 n-place predicate symbol。

Definition : Terms are defined recursively as follows :

1. A constant is a term.
2. A variable is a term.
3. If  $f$  is an n-place function symbol, and  $t_1, \dots, t_n$  are terms, then  $f(t_1, \dots, t_n)$  is a term.

4. All terms are generated by applying the above rules.

Quantifiers :

符號  $\forall$ ,  $\exists$  分別叫做 universal 與 existential quantifiers,  $(\forall X)$  意思是全部  $X$ ,  $(\exists X)$  意思是存在一個  $X$ 。

Quantifiers Scope :

(a)  $(\forall x) (P(x) \rightarrow R(x))$

(b)  $(\forall x) Q(x)$

(a) formul,  $(\forall x)$  範圍是  $(P(x) \rightarrow R(x))$ , (b) formula,

$(\forall x)$  範圍是  $Q(x)$

Definition: Well-formulas in the first-order logic  
are defined recursively as follows:

1. An atom is a formula.
2. If  $F$  and  $G$  are formulas, then  $(\sim F)$ ,  $(F \vee G)$ ,  
 $(F \wedge G)$ ,  $(F \rightarrow G)$ , and  $(F \leftrightarrow G)$  are formulas.
3. If  $F$  is a formula and  $x$  is a free variable in  
 $F$ , then  $(\forall x)F$  and  $(\exists x)F$  are formulas.
4. Formulas are generated only by a finite number  
of applications of 1, 2, and 3.

例: 轉換 " Every man is mortal. Confucius is a man.

Therefore, Confucius is mortal. "

$\text{Man}(x)$ :  $X$  is a man

$\text{Mortal}(x)$ :  $X$  is mortal

所以 Every man is mortal 可被表示如下

(i)  $(\forall x) (\text{Man}(x) \rightarrow \text{Mortal}(x))$ .

(ii)  $\text{Man}(\text{Confucius})$ : Confucius is a man.

由 (i), (ii) 可推導  $\text{Mortal}(\text{Confucius})$ . (註一)

Deductive Relational Data Base :

Predicate 可以看作 relation,

例 : Produce(ABC,Chalk)是一個 Predicate, 在 relation Base 中是一個 relation

Produce

company	Product
ABC	Chalk

EDB ( extensional data base ) 是 facts, real relation

IDB ( intensional data base ) 是 general statement,

virtual relation.

例 : EDB : Produce( ABC, Chalk) 是一個 fact

IDB : Produce(x,y)  $\rightarrow$  Distribute(x,y)

問 : goal : Who distributes chalk ?

Distribute(x, Chalk) Produce(x, Chalk)

{ Chalk/y }

Distribute(ABC, Chalk) Produce(ABC, Chalk)

{ Chalk/y, ABC/x }

所以 ABC distribute Chalk 推導過程是先至 EDB 找

，如沒有匹配之 relation，再至 IDB 去找匹配之 virtual relation.

$\Pi$ -clause notation:  $(T, \phi)$ , where  $T$  is a template and  $\phi$  is a set of substitution sets.

example:  $((x, x, y), \{ [Produce]/\alpha, [ABC, XYZ]/x, [Pencils, Paper, notebook]/y \})$

$((\alpha, x, y) \rightarrow (\beta, x, y), \{ [Produce/\alpha, [Distribute]/\beta, [company]/x, [Product]/y \})$

代表  $Produce(x, y) \rightarrow Distribute(x, y)$

T-clause:

一般 clause 的 conjunction 消去:

$$A1 \wedge A2 \wedge A3 \rightarrow A1A2A3$$

一般 \* 表示此 atomic 已完成，所以 \*，括號可被消除，而 conjunction 置於 atomic 之間

$$A1A2 * (A3A4A5) A6 * \rightarrow A1 \wedge A3 \wedge A4 \wedge A5$$

an T-Goal clause with an EDB clause:



Let  $\alpha$  be a T-goal clause with selected atomic formula  $G_i$ , where  $G_i$  is not followed immediately by an  $\ast$ , and let  $G_1$  be an EDB clause.

Let  $G_i$  and  $G_1$  match in that they share the same predicate name, and a substitution  $\theta$  can be found to make the two atomic formula match in every position.

for example :

If  $\alpha = G_1(x,y) \ G_2(x,a) \ G_3(y,z)$

and  $G_1 = G_2(b,a)$

then  $\theta = \{ b/x \}$  and  $\alpha_1 = (G_1(x,y)G_2(x,a) \ast$   
 $G_3(y,z)) \{ b/x \}$

hence  $\alpha_1 = G_1(b,y) \ G_2(b,a) \ \ast \ G_3(y,z)$

an T-Goal clause with an IDB clause :

IBD  $A_1 A_2 \dots A_n \rightarrow G_i$

$\alpha = G_1 G_2 \dots G_{i-1} G_i \dots$

$\alpha_1 = (G_1 \dots G_i * (A_1 \dots A_n) G_{i+1} \dots G_n) \theta$

EDB : Produce ( ABC, Pencils )

IDB : Produce (x,y)  $\rightarrow$  Distribute(x,y)

Goal clause : Distribute(x,Pencils)

1. Distribute(x,pencils)
2. Distribute(x,pencils) \* (produce(x,pencils))
3. Distribute(ABC,Pencils) \* ( Produce(ABC,  
Pencils) \* )

Distribute(ABC,Pencils)  $\rightarrow$  answers

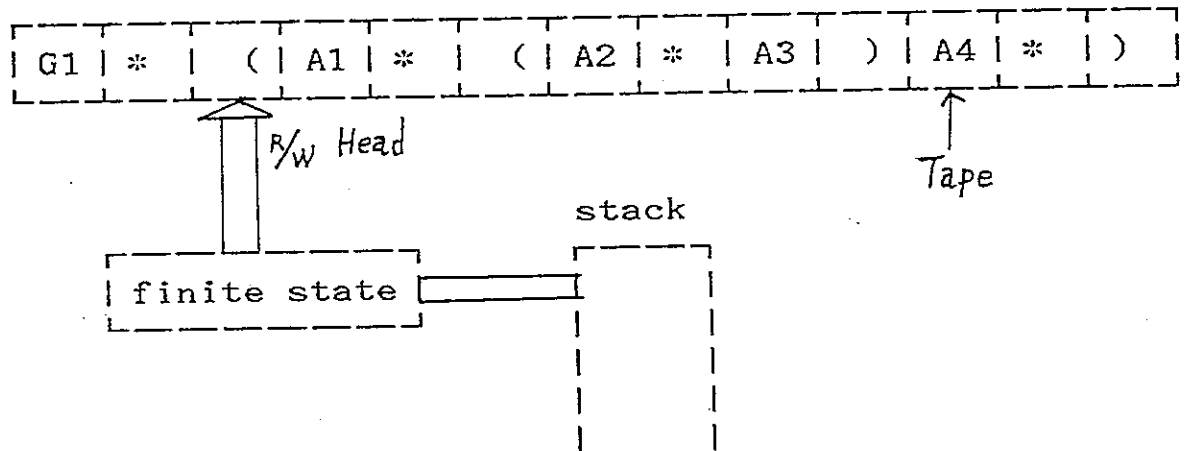
Produce(ABC,Pencils)  $\rightarrow$  Distribute(ABC,Pencils)

$\rightarrow$  implication

Produce(ABC,Pencils)  $\rightarrow$  assertion

Answer and Reason Extraction :

可建一個 pushdown automata 執行回答與推理



level 0 是 answer

其他 level 是 reason

level=0	1.	G1	answer
level=1	2.	$A1 \wedge A4 \rightarrow G1$	implication
level=2	3.	$A2 \wedge A3 \rightarrow A1$	implication IBD
level=2	4.	A2	assertion EDB
level=2	5.	A3	assertion EDB
level=1	6.	A4	assertion EDB

Natural Language Output :

作法是每一Predicate 有一Pattern相對應。

$\Pi$ -clause :

$((\alpha, x, y), \{ [s]/\alpha, [ XEROX ]/X, [ COPIERS ]/y \})$

Pattern: % @1 sells @2%

所以 sentence: XEROX Sells compiers.

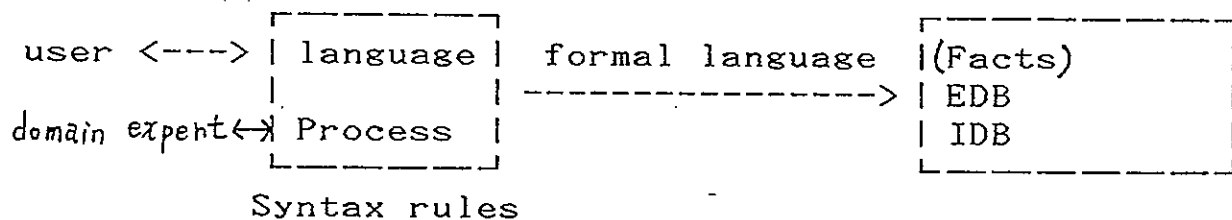
% 符號把 Patter 包含

@1 代入第一個參數

@2 代入第二個參數

Knowledge Base System :

Relation data Base

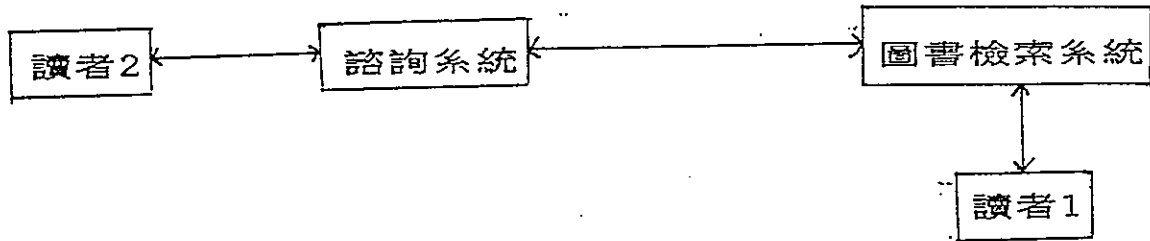


1. User 和 domain expert 與系統溝通，如上所述 Natural language output 方式，因此是一些 Pattern 組合。
2. language process 所處理是一些 pattern，因此祇要 Syntax rules 就夠了，而其轉換成 formal language，是 first-order logic。
3. First-order logic 是一些 Predicate 或其組合，而 Predicate 可視為一個 relation，故可用 Relation Data Base 來儲存，其內存 EDB, IDB。
4. 當 User 一個問題提出時，產生一個 goal clause，然後在 data base 中的 EDB, IDB 去匹配。
5. 建一個 Pushdown automata 抽出 answer and reason。

- 註一： " Symbolic Logic And Mechanical Theorem Proving ", Chin-LIANG CHANG, RICHARD CHAR-TUNG LEE.
- 註二： " ANSWER AND REASON EXTRACTION, NATURAL LANGUAGE AND VOICE OUTPUT FOR DEDUCTIVE RELATIONAL DATA BASES ", Jack Minker and Patricia B. Powell, Natural Language Base Computer Systems.
- 註三： " FORMAL LANGUAGES AND THEIR RELATION TO AUTOMATA ", JOHN E. HOPCROFT and JEFFREY D. ULLMAN.

## 第五章 諮詢系統

### 第一節 系統架構



： 圖書館檢索系統；檢索項包括作者、書名、主題、分類號等等。

： 讀者大略分兩類

1. 讀者 1 類： 確定找那一本書（包括知道作者或書名），或哪一主題的書，這一類讀者直接使用圖書館檢索系統。

a： 查詢作者名或書名，圖書館檢索系統為了解決讀者可能無法記住全名，故有 4,1,1,1 檢索法，即第一字取前四個字母，第二、三、四取第一個字母，檢索到資料如有多筆，則顯現在螢幕上，讓讀者選所要的資料。

b： 查詢主題或分類號，讀者對其想找資料有足夠的背景知識，故圖書館檢索系統，可依書本的借出次數

或出版年的前後，做一順序排列之後，顯現螢幕讓讀者選所要的資料。

2. 讀者 2 類：對想找的資料主題，沒有足夠的背景知識。例如作者大二時已修完電腦概論、高階程式語言、Assembly、與語言的比較，自己想看編譯程式，發覺看不下去。當然是看編譯程式的背景知識不足（沒有自動機理論基礎）。這一類讀者可透過諮詢系統，瞭解到這一領域學科的因果關係，讓讀者有可詢之道，不要一開始就給讀者在學科中迷失，遭受挫折感，而打消其好奇之心。我們這裡不考慮讀者的動機，與背景知識。

：諮詢系統；介於讀者與圖書館檢索系統之間，因此須考慮到兩方面的溝通。

1. 圖書館檢索系統分類號的討論：參考（註一）的中文資訊科技圖書資料分類表（草案）。

513	編譯程式
520	程式語言
521	ADA



522	ALGOL
523	APL
524	ASSEMBLY
525	BASIC
526	COBOL
527	FORTRAN
528	LISP
529	MACHINE LANGUAGE
530	PASCAL
531	PL/1
532	RPG

分類號是一書一個號碼，編的原則是如一書擁有多於一個分類主題時，理論上取一個大於這些分類的分類主題。例如編譯程式，其內容有高階語言、低階語言、自動機理論（formal language），但從此分類表上無法看出編譯程式有這些內容，也無法看出其關係，因分類號 513，與 52～沒有大小類之分，這是實際編類的難處。所以諮詢系統要從現有分類系統整理出分類號之間的關係，而且幫助圖書館檢索系統建某些分類號其內容。

a：建分類號關係，採取學校排課方式，建立學科之間

先後關係。

b：如上編譯程式，其編目資料應增加關於使用何種高階語言，與使用何種低階語言。

2. 讀者的溝通：因無法預測讀者使用詞與圖書館檢索系統所用詞是否一致。因此諮詢系統當遇到不同於圖書館檢索系統用詞時，採取一些詢問步驟，來求得讀者本意，然後把讀者用詞附加於系統用詞之下，以擴大其與讀者溝通能力。

：系統資料結構；底下所考慮限於同一學術領域，如資訊，機械就得分開。

1. 檢索系統用詞檔：（以下簡稱系統用詞）

系統用詞	總類	level	附加讀者用詞		
					... n
⋮					

系統用詞：是分類號名稱

附加讀者用詞：是用來擴充系統與讀者溝通能力，其附加個數，與一些附加策略，實

際 implement 應考慮到。level 與  
總類待學科關係檔時就明白其用意。

2. 學科關係檔：其用詞同系統用詞。以下展示其結構如  
predicate, 所以亦可以用 relation  
data base 來儲存。

(a) 條件敘述

IF you want to read compiler, then you have to  
read firstly formal language and high level  
language and assembly

predicate 表示：

Compiler ( formal, high, assembly )

(b) 總類：如 High level language, 包括 C, PASCAL,  
LISP, COBOL 等等。我們亦以 Predicate 表  
示,

high ( C, PASCAL, LISP, COBOL, ... )

屬於總類名詞, 不能作檢索用, 故應在系統詞  
中註明。

(c) 一個 predicate 的 attribute, 如 compiler 關係中  
之 high, 亦可成為另一個 predicate。

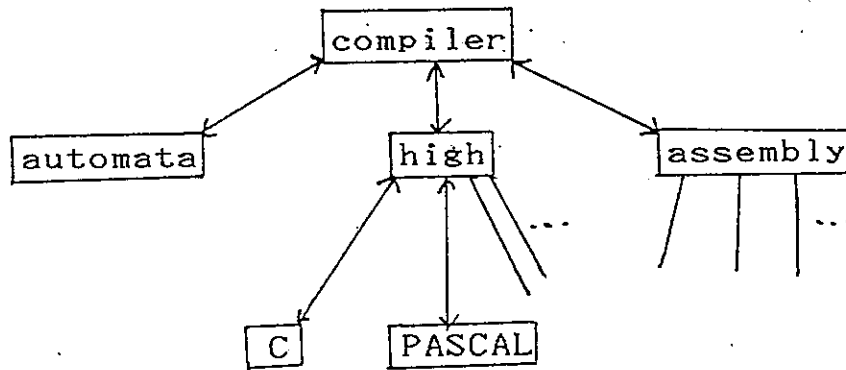
level(P) : 就是 P attribute 中 level 最多的, 再加 1  
 即是 level(P), 而所有 predicate 的 level  
 都大於等於 一; 系統用詞檔中, 則包括 level  
 等於 0, 即是最先修課程。其學科關係檔結構  
 如下:

predicate	level	attribute(variable length)			
					...
⋮					

檢索系統用詞檔與學科關係檔可使用 hash 方法建檔,  
 加速查詢時間。

存在於關係檔中的關係, 可以以圖表示。

例:



TERMINAL NODE : C, PASCAL, automata 等, 他們的  
 level 等於零, 也就是他們是系統  
 檢索的一個檢索點。complier 的

level 為 2。

圖中 node 之間是雙向的，因在關係表的建立，即有此種性質。level 高的 node, 指向 level 低的 node，是唯一，而 level 低的 node, 指向 level 高的 node, 不唯一，也就是可能同時指到好幾個 level 高的 node。因為 level 低的 node 表示較先修課程，所以可能為構建好幾個不同但較後修的課程。

: 關係表既可看成如上圖之樹狀，那我們知道整個關係檔就是好多樹的集合。

: 系統運作

STEP 1: 諮詢系統利用系統用詞，試圖匹配讀者用詞。假設讀者用詞屬於同一學術領域。因不同學術領域用詞可能相同，但意義不同，且對檢索系統亦構成不知檢索哪個資料庫。讀者用詞分為兩種情況，一種是單一用詞，就跳到 STEP 3，一種兩個用詞以上，那就跳到下一步。

STEP 2: 兩個用詞以上，我們根據 level 的資料，由高的往其 attribute 推，看能否包含其它用語，如可全部包含，那就屬於最高 level 的單一用詞，那

就跳到 STEP 3。如不能全包含於最高 level 的用詞下，那是屬於兩棵以上的樹，這時可告知讀者其用詞，那些不相關，徵求讀者意見，是否從新輸入用詞，那就跳到 STEP 1 從新開始。或是不相關樹分開處理，這樣準備一串列儲存待處理單一用詞，同時把不相關用詞，記錄下來，待專家往後作參考。

STEP 3：從最高 level 的 node 開始以 preorder 順序展開。每個 node 須測試其是否為總類，如是那把總類所含內容顯示出來，讓讀者挑選一個，來代替總類。如不是，那就繼續展開至 terminal node 為止。展開同時亦印出如第四章介紹 T-clause 的樣式，存入 T-clause 串列。檢查待處理用詞串列是否空了，如不是，繼續 STEP 3 展開未處理完之用詞。如是則跳到 STEP 4，作第四章所述 Answer and Reason Extraction 的工作，同時亦把須檢索資料交付圖書館檢索系統去執行。

STEP 4：T-clause 樣式就是如同有關數學式子之以括號來表示先做，是故是一 context-free language, 故

可用一 pushdown automata 來作處理，抽出所有 implication 與 STEP 3 所檢索的資料！一起作 Natural language output 給讀者做解釋。

結論：此諮詢系統充分應第四章之理論。我們可以這麼解釋，關係表即是一 predicate 的集合，亦是 deductive relation data base 的 EBD，而 STEP 3 的處理規則即 IBD，而 STEP 4 即是 Answer and Reason Extraction。

註一：謝清俊、楊鍵樵、黃克東、張仲陶，" 中文資訊科技圖書資料分類表(草案) "，BULLETIN OF THE LIBRARY ASSOCIATION OF CHINA, NO. 34  
December 5, 1982。

註二：Denning、Dennis、Qualitz, " Machine, Languages , and Computation 。

註三：JEFFREY D. ULLMAN, " Principles of DATABASE SYSTEMS "，Second Edition。



圖書館是知識的寶庫，經過前人的努力經營來提供適用的寶庫之 KEY。這把 KEY 會因時間不同、科技文明不同，風貌會有所不同，但其功能是不變的，造福追求知識者。

最近消息，國科會科技資訊中心與王安電腦接洽，籌建全國資訊網，屆時使用者可透過電話線查詢科技中心資料。我想本諮詢系統可以考慮其適用於類似科技中心，其 knowledge 可能須更多專家貢獻了。知識內容如包括計劃題目，使用何種文獻，使用何種方法，須有何種背景知識等等，這將加速研究人員蒐集資料與收集方向。