第五章　實驗結果分析與評估

第一節　實驗結果統計

本章所述之『對照組』係以相同之實驗素材作為調校範本之實驗結果統計，並與[31]所設計之演算方法(以『方法 $\mathcal{A}$』代稱)之結果做一比較。而所述之『實驗組』係以另外之八篇專利文獻，作為系統評估之依據，用以參照『對照組』。相關結果呈現於下：

表 10：實驗結果【概念 (Concepts)】統計一覽表(對照組：方法 $\mathcal{A}$ vs.本實驗)

	預期正確的 Concept 數		經雛型系統執行後之結果			
			擷取正確的Concept 數		擷取錯誤的Concept 數	
比　較	方法$\mathcal{A}$	本實驗	方法$\mathcal{A}$	**本實驗**	方法$\mathcal{A}$	**本實驗**
第一篇	50		20	46	12	5
第二篇	38		11	36	20	3
第三篇	20		16	19	6	1
第四篇	25		12	24	8	0
第五篇	15		8	15	9	1
第六篇	25		9	23	4	1
第七篇	54		35	52	29	3
第八篇	44		24	40	27	4

表 11：實驗結果【SAO 結構句組】統計一覽表(對照組：方法 $\mathcal{A}$ vs.本實驗)

	預期正確的 SAO數		經雛型系統執行後之結果			
			擷取正確的SAO數		擷取錯誤的SAO數	
比　較	方法$\mathcal{A}$	本實驗	方法$\mathcal{A}$	**本實驗**	方法$\mathcal{A}$	**本實驗**
第一篇	44		6	43	4	2
第二篇	36		1	34	12	2
第三篇	36		14	34	4	4
第四篇	32		12	32	23	0
第五篇	23		1	22	4	1
第六篇	21		2	19	4	2

第七篇	123	44	106	49	17
第八篇	75	20	63	18	9

表 12：實驗結果【概念 (Concepts)】統計一覽表(實驗組)

		預期正確的Concept 數	經雛型系統執行後之結果	
			擷取正確的Concept 數	擷取錯誤的Concept 數
篇幅長	第一篇	93	85	11
	第二篇	48	45	8
	第三篇	31	31	0
	第四篇	45	44	3
	第五篇	22	20	5
	第六篇	18	18	0
篇幅短	第七篇	23	23	1
	第八篇	18	16	1

表 13：實驗結果【SAO結構句組】統計一覽表(實驗組)

		預期正確的SAO數	經雛型系統執行後之結果	
			擷取正確的SAO數	擷取錯誤的SAO數
篇幅長	第一篇	149	138	7
	第二篇	73	67	9
	第三篇	71	66	4
	第四篇	45	40	3
	第五篇	41	39	1
	第六篇	33	33	0
篇幅短	第七篇	28	27	1
	第八篇	22	18	3

第二節　系統評估方法描述

客觀地『評估』自動摘要技術或方法的良窳，乃是自動化資訊摘要研究領域的一大課題。目前，已有不少學者針對自動摘要該如何評估以及評估時所遇到的種種問題做了

廣泛而深入的探討。然而，欲以一套廣為大家所接受、認可且客觀的標準來評價一個自動摘要方法的優劣或是衡量自動摘要系統的表現，這其實並沒有想像中的那樣容易！

在瞭解評估模式之前，我們可以先來思考一下所謂理想的文件『摘要』(Ideal Summarization) 應該具備有哪一些特性，好讓我們透過這些相關特質作為評估的重要依據。I. Mani(2001) 在其經典名著*Automatic Summarization*一書中曾提及自動摘要的目標：

> The goal of automatic summarization is *to take an information source, extract content from it, and present the most important content to the user in a condensed form and in a manner sensitive to the user's or application's needs.*

換句話說，就是要從來源的訊息當中提取相關之內容，並根據使用者的偏好或者是應用領域的特殊需求為考量之著眼點，以扼要簡明的形式將最重要的內容呈現出來 [12]。

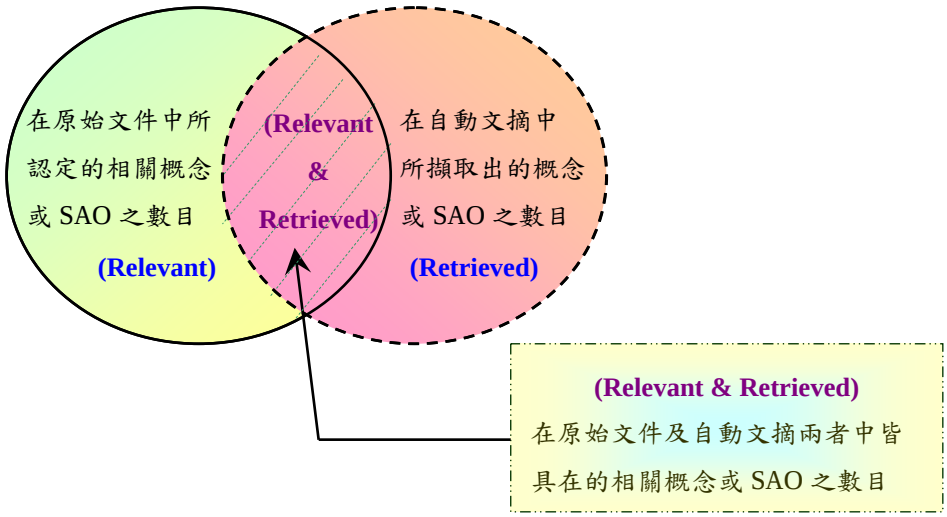
因此，依據上述摘要之目標，我們認為一個〝好 『摘要』(Summarization)〞基本上應該兼備所謂的摘要三性：「簡化性」、「情報性」與「組織結構良好性」(參閱附錄三：良好摘要的三大特性(Characteristics of Summaries)) [19]。

5.2.1 自動化資訊摘要評估(Evaluation)模式

我們可依據上述良好摘要的三大特性來扼要的描述目前常見的評估模式，以及這些模式的優、缺點。一般來說，自動摘要系統的評估可以歸納為兩種主要的模式：一為內部評估(Intrinsic evaluations) 模式，另一則為外在評估(Extrinsic evaluations) 模式 [17] [18][22]。所謂的『內部評估』(Intrinsic evaluations)，顧名思義乃是要直接去評估系統本身的摘要品質。通常內部評估是將系統所產生的自動摘要與藉由具代表性的第三者(比如：領域專家)所撰寫出的人工摘要來加以比對兩者之間的相似程度；或者是直接透過人為閱讀的方式，讓讀者主觀地自評其使用滿意度，以此判定摘要結果的好與壞 [17]

[18]。而『外在評估』(Extrinsic Evaluations)則是將自動摘要的結果當成其他資訊系統的輸入，然後藉由完成某一種特定工作之成效來間接衡量此摘要的品質(例如：分別依據原始文件和自動摘要內容當作輸入源，並透過同一文件分類資訊系統來做文件的分類，看其結果是否一致？) [17][18]。至於，該採用哪一種評估模式會比較適宜？這答案端視不同之應用情況及需要才有辦法做決定[7][17][18][19][22][23][27]。

本實驗之系統評估想法則如圖 43所示，乃是一種外在評估法(Extrinsic Evaluations)；其目前常用之衡量指標主要有三(如下圖 42 所展示)，分別是：❶召回率/查全率(Recall Rate)、❷準確率/查準率(Precision Rate)以及❸準確率和召回率的調和平均數(F-measure) [14]。



❶ $$召回率/查全率(\text{Recall}) = \frac{(\text{Relevant} \,\&\, \text{Retrieved})}{(\text{Relevant})}$$

❷ $$準確率/查準率(\text{Precision}) = \frac{(\text{Relevant} \,\&\, \text{Retrieved})}{(\text{Retrieved})}$$

❸ $$準確率和召回率的調和平均數(\text{F-measure}) = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$$

圖 42 ：自動摘要系統的三項衡量指標：召回率、準確率及其兩者之間的調和平均數

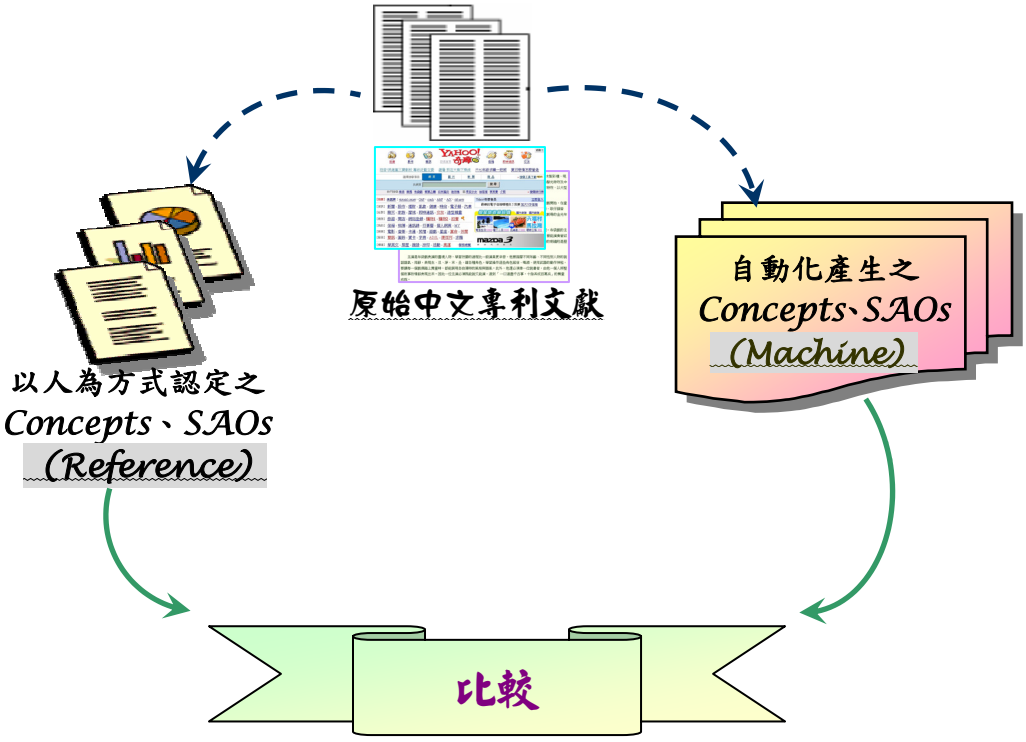


圖 43 ：本實驗系統評估方法示意圖

其中，『召回率/查全率』(Recall Rate)可用以鑑定摘要系統找出所有相關資料的能力，而『準確率/查準率』(Precision Rate)則可以用來鑑定摘要系統過濾不必要資料的能力。在衡量自動摘要系統的成效時，召回率與準確率其實是可以一起並用的，單獨檢視其中某一項指標可能沒有太大之意義。透過召回率與準確率來衡量系統之成效時，可能的最大缺點是在對於原始文件當中相關概念或關鍵詞的認定難有一客觀之評核準則，此種人為之判斷是一種非常主觀的過程，其結果往往會因人而異。儘管如此，此二指標仍是目前最廣為採用的衡量方法。此外，為了能夠有更加客觀的數據以進行上述兩類指標之間的比較，我們可以把召回率與準確率這兩項指標對於自動摘要系統的評估能力視為一樣而將召回率與準確率的資訊合併在一起計算——亦即，將召回率與準確率各占一半的比重，以計算出準確率和召回率之間的調和平均數(F-measure) [14]。

5.2.2 本實驗系統評估方法描述

$$召回率(\text{Recall}) = \frac{(擷取正確的\ Concept\ 數)}{(預期正確的\ Concept\ 數)}$$

以及

$$召回率(\text{Recall}) = \frac{(擷取正確的\ SAO\ 數)}{(預期正確的\ SAO\ 數)}$$

方程式 6：本實驗召回率(Recall)的計算公式

$$準確率(\text{Precision}) = \frac{(擷取正確的\ Concept\ 數)}{(擷取正確的\ Concept\ 數\ +\ 擷取錯誤的\ Concept\ 數)}$$

以及

$$準確率(\text{Precision}) = \frac{(擷取正確的\ SAO\ 數)}{(擷取正確的\ SAO\ 數\ +\ 擷取錯誤的\ SAO\ 數)}$$

方程式 7：本實驗準確率(Precision)的計算公式

$$準確率(\text{Precision})和召回率(\text{Recall})的調和平均數(\text{F-measure}) = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$$

方程式 8：本實驗準確率和召回率之間的調和平均數(F-measure)計算公式

第三節　系統實驗結果評估與分析

我們依照上一節所探討之評估方法，套用方程式 6(召回率(Recall) 計算公式)、方程式 7(準確率(Precision) 計算公式)、以及方程式 8(準確率和召回率之間的調和平均數(F-measure) 計算公式)分別將第一節所呈現之實驗統計成果做一評估，評估結果整理如下：

表 14：實驗結果【概念 (Concepts)】評估一覽表(對照組：方法$\mathcal{A}$ vs.本實驗)

	召回率(Recall Rate)		準確率(Precision Rate)		F-measure	
比　較	方法$\mathcal{A}$	本實驗	方法$\mathcal{A}$	本實驗	方法$\mathcal{A}$	本實驗
第一篇	0.40	0.92	0.63	0.90	0.49	0.91
第二篇	0.29	0.95	0.36	0.92	0.32	0.94
第三篇	0.80	0.95	0.73	0.95	0.76	0.95
第四篇	0.48	0.96	0.60	1.00	0.53	0.98
第五篇	0.53	1.00	0.47	0.94	0.50	0.97
第六篇	0.36	0.92	0.69	0.96	0.47	0.94
第七篇	0.65	0.96	0.55	0.95	0.59	0.95
第八篇	0.55	0.91	0.47	0.91	0.51	0.91
總平均	0.51	0.95	0.56	0.94	0.53	0.94

表 15：實驗結果【SAO結構句組】評估一覽表(對照組：方法$\mathcal{A}$ vs.本實驗)

	召回率(Recall Rate)		準確率(Precision Rate)		F-measure	
比　較	方法$\mathcal{A}$	本實驗	方法$\mathcal{A}$	本實驗	方法$\mathcal{A}$	本實驗
第一篇	0.14	0.98	0.60	0.96	0.22	0.97
第二篇	0.03	0.94	0.08	0.94	0.04	0.94
第三篇	0.39	0.94	0.78	0.89	0.52	0.92
第四篇	0.38	1.00	0.34	1.00	0.36	1.00
第五篇	0.04	0.96	0.20	0.96	0.07	0.96
第六篇	0.10	0.90	0.33	0.90	0.15	0.90
第七篇	0.36	0.86	0.47	0.86	0.41	0.86
第八篇	0.27	0.84	0.53	0.88	0.35	0.86
總平均	0.21	0.93	0.42	0.92	0.28	0.93

表 16：實驗結果【概念 (Concepts)】評估一覽表(實驗組)

篇幅長 → 篇幅短

	召回率(Recall Rate)	準確率(Precision Rate)	F-measure
第一篇	0.91	0.89	0.90
第二篇	0.94	0.85	0.89
第三篇	1.00	1.00	1.00
第四篇	0.98	0.94	0.96
第五篇	0.91	0.80	0.85
第六篇	1.00	1.00	1.00
第七篇	1.00	0.96	0.98
第八篇	0.89	0.94	0.91
總平均	0.95	0.92	0.94

表 17：實驗結果【SAO結構句組】評估一覽表(實驗組)

篇幅長 → 篇幅短

	召回率(Recall Rate)	準確率(Precision Rate)	F-measure
第一篇	0.93	0.95	0.94
第二篇	0.92	0.88	0.90
第三篇	0.93	0.94	0.94
第四篇	0.89	0.93	0.91
第五篇	0.95	0.98	0.96
第六篇	1.00	1.00	1.00
第七篇	0.96	0.96	0.96
第八篇	0.82	0.86	0.84
總平均	0.92	0.94	0.93

由表 16 及表 17 的數據來看，經效益評估後的結果顯示，我們在雛型系統中所設計的概念(Concepts)及 SAO 結構句的擷取演算都還有令人滿意的不錯表現。以整體平均

來說，【實驗組】概念(Concepts) 擷取方面的召回率為 95.34%，準確率為 92.13%；而 SAO 結構句組擷取方面的召回率則為 92.45 %，準確率為 93.79%。

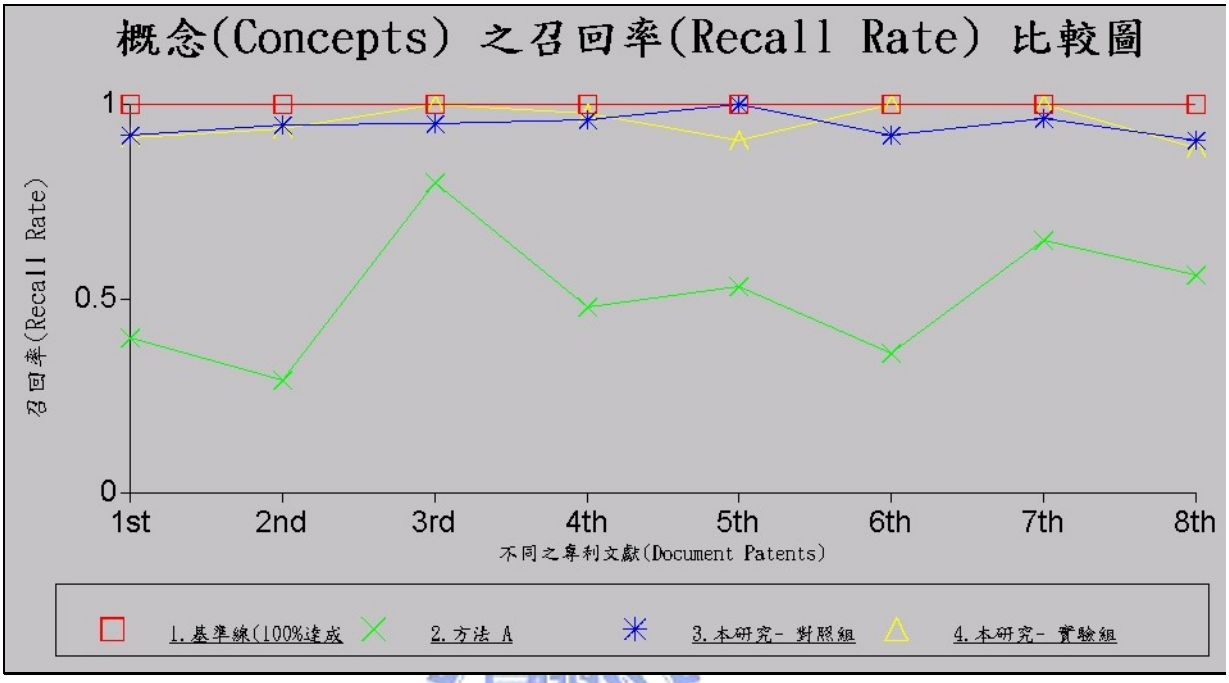


圖 44 ：概念(Concepts) 之召回率(Recall Rate) 比較圖

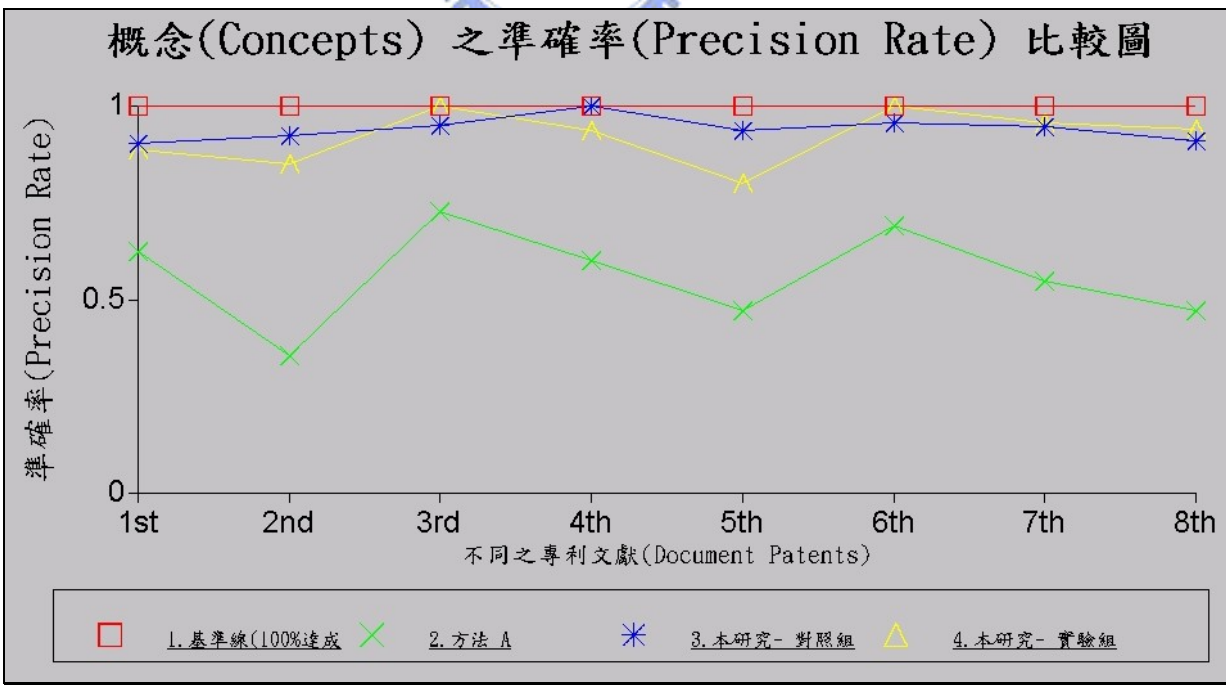


圖 45 ：概念(Concepts) 之準確率(Precision Rate) 比較圖

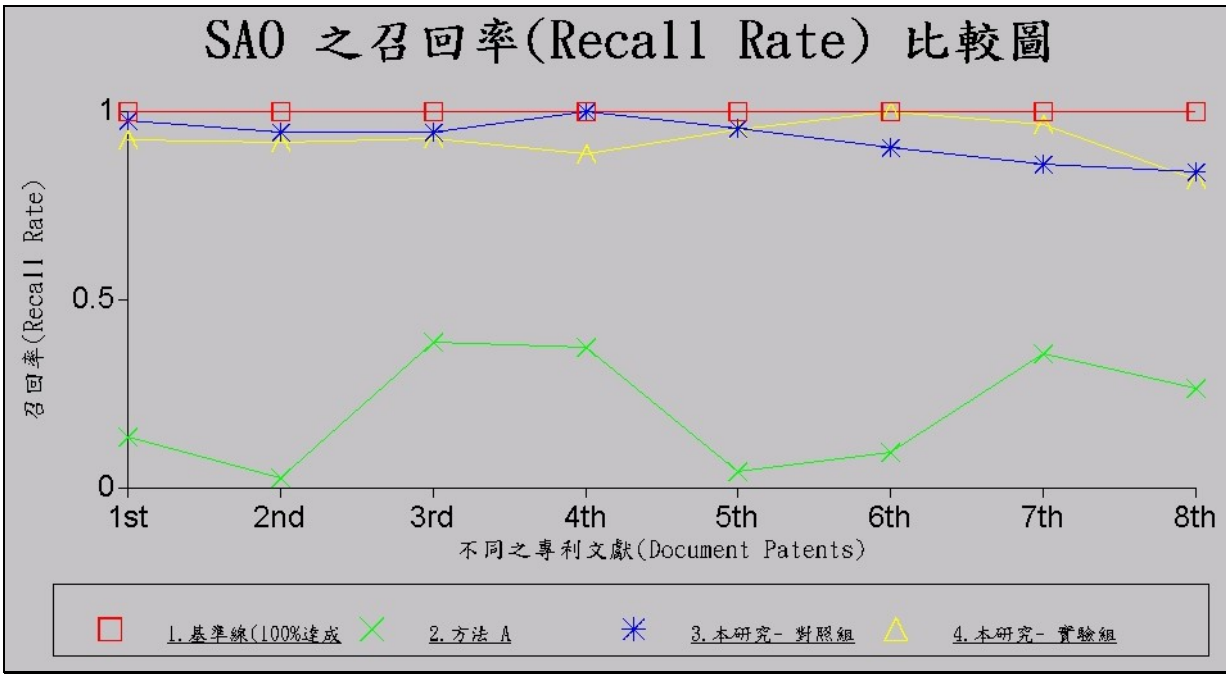


圖 46 ：SAO 之召回率(Recall Rate) 比較圖

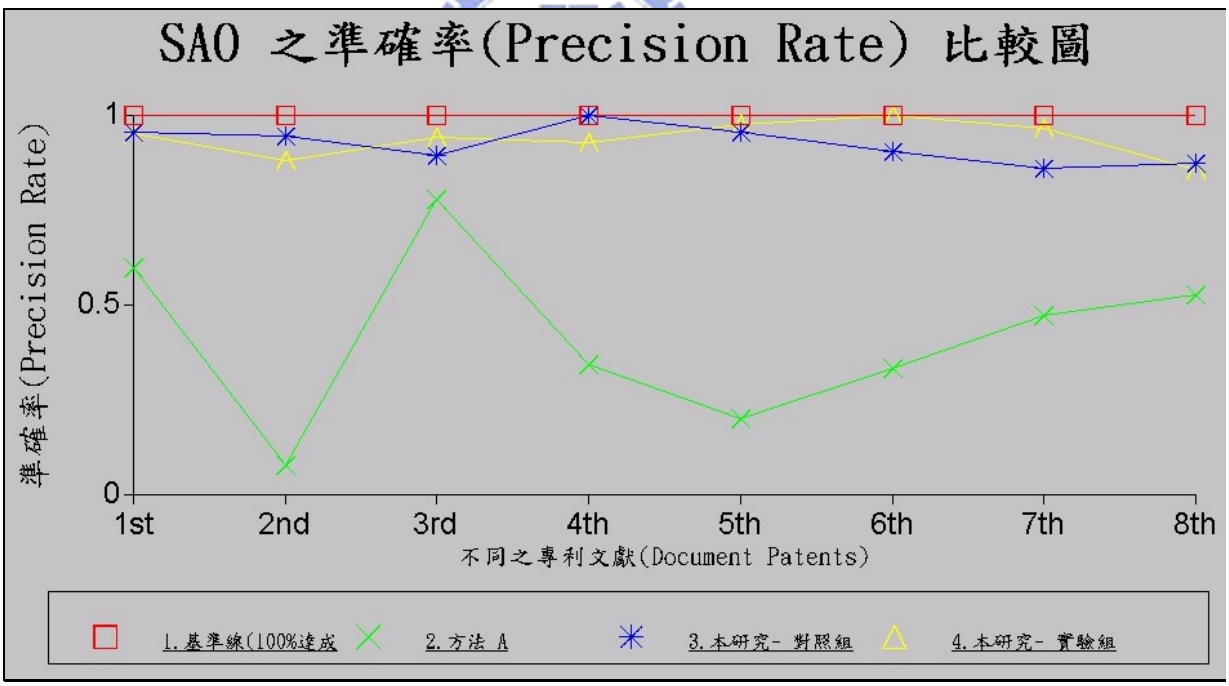


圖 47 ：SAO 之準確率(Precision Rate) 比較圖

值得一提的是，【實驗組】的八篇中文專利文件乃是依據申請專利範圍(Claims)之篇幅由長至短所作之排序(參考表 12 及表 13)。由實際所呈現的召回率(Recall Rate)、準

確率(Precision Rate)、F-measure 等評估數據來看(參考圖 44、圖 45、圖 46 及圖 47)，本研究所設計之概念(Concepts)、SAO 擷取演算不會因專利篇幅的長短而有差異懸殊之結果表現，顯見本研究所施行的演算方法具有某一定程度之穩定度與可靠度。而由表 16 及表 17 中召回率(Recall Rate)、準確率(Precision Rate)所呈現的數據來看，其數值相對來說較低的部份，究其原因後可以發現：乃是經由 CKIP 此一工具進行中文自動斷詞以及詞性標記後之部份謬誤現象所造成之影響。儘管我們已針對【對照組】的八篇中文專利文獻所出現攸關於斷詞(Tokenization)及詞性標記(Tagging) 的部份謬誤現象嘗試探索出一些通則作為 Heuristic Rules 來做一些調校與修正，但仍無法對其所有可能之謬誤現象予以窮舉。也就是說，倘若未來經 CKIP 工具所協助進行之中文自動斷詞以及詞性標記之正確率愈高，那麼透過本研究所設計之 Concepts (概念)、SAO 擷取演算之擷取結果之正確率亦將隨之愈高。

第六章　結論與未來研究方向

本章歸納及探討個人在實驗過程中的一些心得與啟發，以及說明未來可行的研究方向。第一節先總結本論文將英文 SAO 的結構句型運用在中文專利文獻自動摘要系統上的效益與可行性；第二節則說明未來可能的研究發展方向。

第一節　結論

本論文主要是在探究以英文句型中的主詞、動詞與受詞結構(SAO結構)輔以探索性經驗法則(Heuristic Rules)來擷取並彙整出中文專利文獻中的重要語句並加以綜合整理後使之成為一專利摘要。首先，我們透過中央研究院CKIP 工具的使用來快速地協助我們完成中文自動斷詞以及詞性標記的工作。接著，我們以『長詞優先法則』作為我們擷取專利文獻當中重要“概念”(Concepts)之指引。然後再以基本的動詞作為『候選關聯』(Candidate Relations)，以利我們將概念(Concepts)、概念(Concepts)與概念(Concepts)之間的關聯(Relations)橋接成為一SAO的結構句(概念----關聯----概念)。最後，以人性化的圖形使用者界面如“樹狀階層結構”方式或是“自然語言之形式”來將此專利的摘要依照資訊量大、中、小的不同來加以呈現，並將擷取自“申請專利範圍”(Claims)中較為抽象的上位用語－“概念”(Concepts)映射到“發明說明”(Detailed Description of the Invention) 中的內容，以尋求意思較為具體、貼切的下位用語作為輔助閱讀之參考。而經效益評估後的結果顯示，我們在雛型系統中所設計的概念(Concepts)及SAO結構句的擷取演算，其結果尚可令人感到滿意。顯示本研究所提之演算構想，深具可行性。

第二節　未來可行的研究方向

未來，我們可以針對不同的主題、不同之專利文獻，先建構出不同的詞彙庫，然

後在不同的詞彙庫上再建構出不同的專利詞彙網路。若能累積一定數量的專利文獻及其對映之 SAO(主詞、動詞和受詞)的結構句組後，然後再透過機器學習(Machine Learning)的方式，我們可迅速地據以分析並構建出屬於該技術領域的知識本體(Ontologies) 出來。一旦建立好了某個領域知識體系的平台，我們就可以透過『問題解決』(Problem-Solving）的方法(如下表 18及圖 48所示)來呈現結構化的資訊索引，讓企業研發部門、專利工程師、產業分析師或智權人員能夠以最經濟、最有效率的方式來尋求問題的解答，以驚人的速度找出屬於該領域的專利技術趨勢，以及所對映之各種技術的分類與專利文獻。此外，亦可更進一步地來製作完整的專利技術功效矩陣圖，幫助企業決定未來可行的研發方向並且提供寶貴的攸關侵權之警訊，不致落入了陷阱而不自知。

表 18：擷取自“申請專利範圍(Claims)”中的 SAO結構句之應用概想

問句句型	主詞(Subject)	動詞(Action)	受詞(Object)
❶ ? + V. + O.	☞ 參考解答	✍ 輸入參數	✍ 輸入參數
❷ S. + ? + O.	✍ 輸入參數	☞ 參考解答	✍ 輸入參數
❸ S. + V. + ?	✍ 輸入參數	✍ 輸入參數	☞ 參考解答
❹ ? + V. + ?	☞ 參考解答	✍ 輸入參數	☞ 參考解答
❺ S. + ? + ?	✍ 輸入參數	☞ 參考解答	

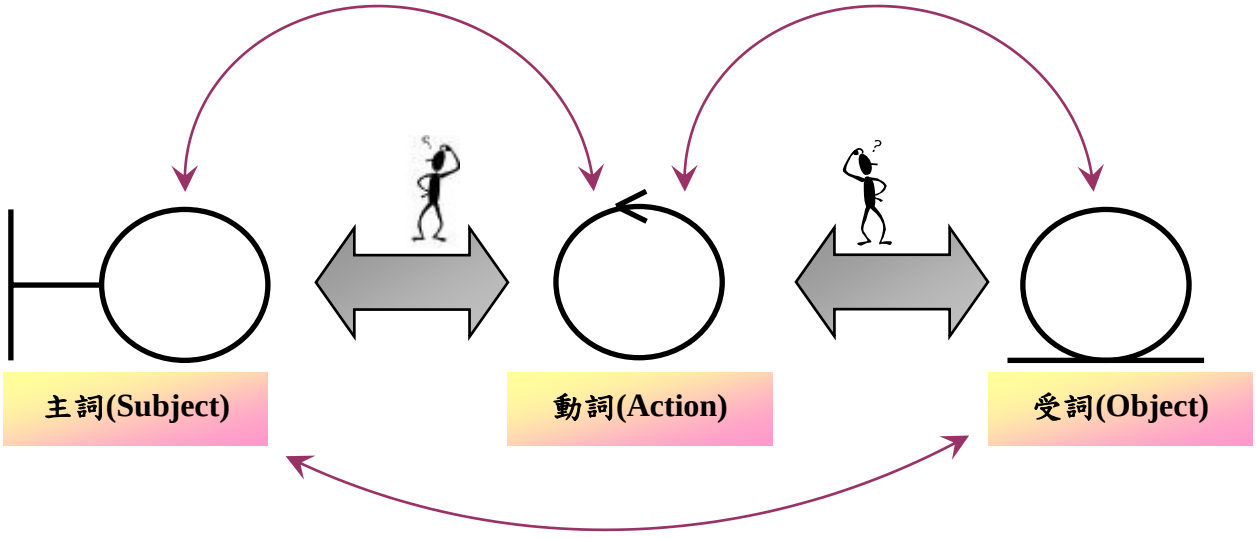


圖 48 ：SAO 關係求解示意圖

附錄一：專利說明書(Document Patent)的主要結構及其文件規範

專利說明書（Document Patent）的主要結構
(一)**發明名稱 (Title of the Invention)**
1. 應與其申請專利範圍內容相符，不得冠以無關之文字。 2. 台灣及大陸規定發明名稱必須反映發明內容。 3. 台灣特別要求發明名稱必須與專利申請範圍之標的相呼應。
(二)**摘要 (Abstract)**
1. 應敘明發明或新型所揭露內容之概要，並以所欲解決之技術問題、解決問題之技術手段及主要用途為限；亦即，摘要敘述發明特徵。 2. 一般，均由Claims中的獨立項所改寫；字數，以不超過 250字為原則。 3. 有化學式者，應揭示最能顯示發明特徵之化學式。 4. 不得記載商業性宣傳詞句。 5. 一般而言，專利文獻上所述之“**摘要**”(**Abstract**)，其資訊並不足以代表此篇專利的全文內容。亦即，此“**摘要**”(**Abstract**)可能埋有伏筆，其敘述可能並非全然是發明者的真心話語、也非發明內容的真實縮影。
(三)**發明說明 (Description)**
發明說明應明確且充分揭露，使該發明所屬技術領域中具有通常知識者，能瞭解其內容，並可據以實施；亦即，使同行之人能瞭解。 1. 發明或新型所屬之**技術領域(Technical Field)**。 2. **先前技術(Background of the Invention/Related Art)**： 記載申請人所知之先前技術(Prior Art)，並得檢送該先前技術之相關資料。在英文說明書中，會使用現在完成式，特別說明先前技術的演進。 3. **發明或新型內容(Summary of the Invention)**： 敘述發明或新型所欲解決之技術問題、解決問題之技術手段及對照先前技術之功效。通常，此部份的摘要內容是將申請專利範圍內容改寫至此，少部份則是從實施方式的內容摘要下來。 4. **實施方式(Detailed Description of the Invention)**： 就一個以上發明或新型之實施方式加以記載，必要時得以實施例說明；有圖式者，應參照圖式加以說明。 5. **圖式簡單說明(Brief Description of the Drawings)**： 其有圖式者，應以簡明之文字依圖式之圖號順序說明圖式及其主要部份之代表符號。
(四)**申請專利範圍 (Claims)**
申請專利範圍應明確記載申請專利之發明、申請專利之標的、技術內容及特點，各請求項應以簡潔之方式記載，且必須為發明說明及圖式所充分支持。 1. 獨立項應載明申請專利之標的、構成及其實施之必要技術內容、特點。 2. 附屬項應敘明所依附之項號及申請標的，並敘明所依附項目外之技術特點。

附錄二：淺層摘要的研究取向的一些重要方法及其實體參考特徵

淺層摘要研究取向(Shallower Approaches)		
	方 法 描 述	重 要 參 考 特 徵
❶表層的方法	Surface-level Approaches傾向於透過一些出色的演算法將原始文件當中的一些淺顯特徵*(Shallow Features)*選取出來並加以重新組織、合成以表達冗長之原文訊息。	▻ 主題特徵(Thematic Features) ▻ 位置特徵(Location Features) ▻ 背景特徵(Background Features) ▻ 線索字詞/提示片語(Cue Words/Phrases)
❷個體元素層的方法	Entity-level Approaches傾向於引入額外的知識體系來表達原始內文的連結樣式(Patterns of Connectivity，例如：Graph Topology) 、分析文件的結構及其所代表的意義，以塑模出文件當中所包含的個體元素(Text Entity)與個體元素之間的關聯性，從而建立此文件內部的知識表示模型，最後以某種演算法擷取出具代表性的部分，達到自動化摘要的目的。	▻ 相似度(Similarity) ▻ 鄰近度(Proximity) ▻ 同時出現(Co-occurrence) ▻ 語彙在詞典中的關係(Thesaural relationships among words) ▻ 同指涉/共同參照關係(Coreference) ▻ 邏輯上的相關性：如同意(Agreement)、矛盾性(Contradiction)與一致性(Consistency)等等； ▻ 句法關係(Syntactic relations) ▻ 以意義表述為本的關聯(Meaning representation-based relations
❸語段層的方法	Discourse-level Approaches傾向於建構出全文內容的整體結構原型，及其之間的關聯。	▻ 文件的格式(Format of the Document) ▻ 在原文中關於同一主題的相關訊息串連(Threads of topics as they are revealed in the text) ▻ 原文的修辭結構(Rhetorical Structure of the Text)

整理自 I. Mani, and M. Maybury (1999), “Introduction”, In Advances in Automatic Text Summarisation (Ed. Mani and Maybury), MIT Press, pp. x-xv,1999. [18]

附錄三：良好摘要的三大特性(Characteristics of Summaries)

好摘要三大特性	代表意義及量測依據
❶**簡化性** **(Reduction)**	✍指將龐大的資訊內容縮減，化繁為簡--講重點，取菁華。 ✍ *Compression complicates evaluation.* ➔以『壓縮率』(Compression Rate)、『濃縮率』(Condensation Rate)或是『減少率』(Reduction Rate)為量測的準則，所謂的『壓縮率』乃是指摘要文件長度與原始文件長度之間的比率，是評估文件摘要系統優劣的重要指標之一，公式如下： c = Summary Length / Source Length (0 < c < 100)。 ✍或是以目標長度(Target Length)為量測之依據(例如：僅取原文中的最重要四句出來當作文摘)。
❷**情報性** **(Informativeness)**	✍ *Modeling user needs.* ➔意指所擷取出的摘要必須含有足量而豐富的資訊內容，且儘可能地吻合使用者真正的需求(Relevance to User's Interests)，以提供相關閱聽人作為決策、參考之依據。 ✍必須忠於原始文件之內容(Fidelity to Source)，不可以畫蛇添足。
❸**組織結構良好性** **(Well-formedness)**	✍ *Avoid incomplete reference set.* ➔不管是在語法或是語段的層次(Syntactic-level or Discourse-level)要分別能夠符合句法結構以及語意的連貫性，亦即需考量摘要內容的流暢度及其可讀性等等因素。 ✍對於節錄(Extracts)要避免語意上有落差、語句的重複等等情事(Need to avoid gaps, dangling anaphors, ravaged tables, lists, etc.)。 ✍對於摘述(Abstracts)要產生合於文法的、言之有物的內容輸出(Need to produce grammatical, plausible output)。

引述及整理自 I. Mani, and M. Maybury (2001), "Automatic Summarization", in ACL/EACL 2001 Meeting [19] 及 I. Mani(2001), "Automatic Summarization" 一書 , John Benjamins Publishing Co.

附錄四：三種主要的關鍵詞自動擷取技術比較一覽表

	方法特徵	優點	缺點
詞庫比對法	詞庫比對法主要是利用已建構好的詞庫，將輸入的文件(或文句)來加以比對，以擷取出文件(或文句)當中出現在詞庫裏的字詞或是片語。	1. 製作簡單，只要將詞庫中的每個詞，去比對是否出現在所輸入的文件當中即可。 2. 其結果都是詞庫中的正確詞彙。	1. 不保證所有關鍵詞都能被擷取出來。 2. 需要耗費人力、時間維護詞庫以容納各個領域的專業用語與新生詞彙，無法應付未曾預料的人名、地名、機構名等專有名稱。 3. 若所建構之詞庫越大則其比對速度越慢。
文法剖析法	將詞庫比對法輔以構詞之規則，透過自然語言處理(NLP)技術的文法剖析程式，剖析出文件中的名詞片語，再運用一些方法與準則，過濾掉不適合的詞彙。	1. 其結果幾乎也都是有意義的名詞片語。	1. 大部份的剖析程式，需要藉助已經建立的詞典或語料庫，因此其缺點也和詞庫比對法一樣。 2. 有些文法剖析法甚至只能剖析合乎文法的完整文句，使得書目、標題等資料裡的關鍵詞無法被擷取出來。
統計分析法	透過對文件的分析，累積足夠的統計參數後，再將統計參數符合某些條件的片語擷取出來。最簡單的統計參數就是去計算詞彙發生的頻率，亦即詞頻，將詞頻落在某一範圍的詞彙取出，小於此閾限值而未被取出的詞彙，即是所謂的停止字元(Stop Word)。	1.較不受語文國別與句型的限制。 2.可以擷取出未曾被詞庫、語料庫網羅的專業用語、新生詞彙與專有名稱等片語。	1.由於沒有用到詞庫或是語料庫，往往會有擷取錯誤的情況發生，出現了毫無意義或是不合乎語法邏輯的詞彙。 2.有可能因統計參數的不足，而致某些關鍵詞錯失了選取機會。 3.比較難以應付出現新詞時的情況，亦較無法呈現出語言的意義。

整理自 曾元顯(1997)，“關鍵詞自動擷取技術與相關詞回饋”，中國圖書館學會會報，第59期，1997. available at http://blue.lins.fju.edu.tw/~tseng/papers/feedback.htm . [24]

附錄五：中央研究院 CKIP 現代漢語詞類標記及其對應之意義

項	現代漢語詞類標記	所 對 應 意 義 之 中 英 文 說 明
1	A	非謂形容詞(Non-predicative Adjective)
2	D	副詞(Adverb)
3	Da	數量副詞(Quantitative Adverb)
4	Dfa	動詞前程度副詞(Pre-verbal Adverb of Degree)
5	Dfb	動詞後程度副詞(Post-verbal Adverb of Degree)
6	Dk	句副詞(Sentential Adverb)
7	Di	時態標記(Aspectual Adverb)
8	Caa	對等連接詞，如：和、跟(Conjunctive Conjunction)
9	Cbb	關聯連接詞(Correletive Conjunction)
10	Nep	指代定詞(Demonstrative Determinatives)
11	Neqa	數量定詞(Quantitative Determinatives)
12	Nes	特指定詞(Specific Determinatives)
13	Neu	數詞定詞(Numeral Determinatives)
14	FW	外文標記(Foreign Word)
15	Nf	量詞(Measure)
16	Na	普通名詞(Common Noun)
17	Nb	專有名稱(Proper Noun)
18	Nc	地方詞(Place Noun)
19	Ncd	位置詞(Localizer)
20	Nd	時間詞(Time Noun)
21	Nh	代名詞(Pronoun)
22	P	介詞(Preposition)
23	Cab	連接詞，如：等等(Conjunction)
24	Cba	連接詞，如：的話(Conjunction)
25	Neqb	後置數量定詞(Post-quantitative Determinatives)
26	Ng	後置詞(Postposition)
27	DE	的，之，得，地
28	I	感嘆詞(Interjection)
29	T	語助詞(Particle)
30	VA	動作不及物動詞(Active Intransitive Verb)
31	VB	動作類及物動詞(Active Pseudo-transitive Verb)
32	VH	狀態不及物動詞(Stative Intransitive Verb)
33	VI	狀態類及物動詞(Stative Pseudo-transitive Verb)

項	現代漢語詞類標記	所 對 應 意 義 之 中 英 文 說 明
34	SHI	是
35	VAC	動作使動動詞(Active Causative Verb)
36	VC	動作及物動詞(Active Transitive Verb)
37	VCL	動作接地方賓語動詞**(Active Verb with a Locative Object)**
38	VD	雙賓動詞(Ditransitive Verb)
39	VE	動作句賓動詞(Active Verb with a Sentential Object)
40	VF	動作謂賓動詞(Active Verb with a Verbal Object)
41	VG	分類動詞(Classificatory Verb)
42	VHC	狀態使動動詞(Stative Causative Verb)
43	VJ	狀態及物動詞(Stative Transitive Verb)
44	VK	狀態句賓動詞(Stative Verb with a Sentential Object)
45	VL	狀態謂賓動詞(Stative Verb with a Verbal Object)
46	V_2	有

引用自 CKIP AutoTag, available at http://godel.iis.sinica.edu.tw/CKIP/ .

參考文獻

[1] R. Barzilay, and M. Elhadad(1997),“Using Lexical Chains for Text Summarization”, In ACL/EACL Workshopon Intelligent Scalable Text Summarization, pp. 10-17, 1997.

[2] B. Boguraev, and C. Kennedy(1997),“Salience-Based Content Characterisation of Text Documents”, In Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization, pp. 2-9, 1997. available at http://www.research.ibm.com/talent/documents/bran_salience.pdf (2005/01/26).

[3] R. Brandow, K. Mitze, and L. F. Rau(1995),“Automatic Condensation of Electronic Publications by Sentence Selection”, in *Information Processing and Management: an International Journal*, 31(5), pp. 675-686, 1995.

[4] H. P. Edmundson (1969), “New methods in automatic extracting”, Journal of the Association for Computing Machinery, 16(2), pp.264-285, 1969.

[5] E. Hovy, and C. Y. Lin (1997), “Automated Text Summarization in SUMMARIST”, In *ACL/EACL-97 Workshop on Intelligent Scalable Text Summarization*, pp. 18-24, 1997.

[6] H. Jing(2002) , “Using Hidden Markov Modeling to Decompose Human-Written Summaries” , Computational Linguistics, 28(4), pp.527-543 , 2002.

[7] H. Jing, R. Barzilay, K. McKeown, and M. Elhadad(1998), “Summarization Evaluation Methods: Experiments and Analysis” ,*In Working Notes of the AAAI-98 Spring Symposium on Intelligent Text Summarization*, pp.60-68, 1998.

[8] H. Jing and K. McKeown(1999) , “The decomposition of human-written summary sentences” , In Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR'99), pp.129-136,1999.

[9] J. Kupiec, J. Pedersen, and F. Chen(1995),“A Trainable Document Summarizer”, In *Proceedings of the 18th annual international ACM-SIGIR conference on Research and development in information retrieval*(SIGIR '95) , pp. 68-73, 1995.

[10] C. Y. Lin(1999), “Training a Selection Function for Extraction” ,In *Proceedings of the eighth international conference on Information and knowledge management***,** pp.55-62, 1999.

[11] H.P.Luhn (1958), “The Automatic Creation of Literature Abstracts”, *IBM Journal of Research and Development*, 2(2), pp. 159-165, 1958. available at http://www.research.ibm.com/journal/rd/022/luhn.pdf (2005/01/26).

[12] I. Mani(2001),“Preliminaries”, In Automatic Summarization , John Benjamins Publishing Co. , pp. 1-25, 2001.

[13] I. Mani(2001),“Professional summarizing”, In Automatic Summarization , John Benjamins Publishing Co. , pp. 27-44, 2001.

[14] I. Mani(2001),“Extraction”, In Automatic Summarization , John Benjamins Publishing

Co. , pp. 45-75, 2001.
[15] I. Mani(2001),“Discourse-level information”, In Automatic Summarization , John Benjamins Publishing Co. , pp. 91-128, 2001.
[16] I. Mani(2001),“Abstraction”, In Automatic Summarization , John Benjamins Publishing Co. , pp. 131-167, 2001.
[17] I. Mani(2001),“Evaluation”, In Automatic Summarization , John Benjamins Publishing Co. , pp. 221-259, 2001.
[18] I. Mani, and M. Maybury (1999), “Introduction”, In Advances in Automatic Text Summarisation (Ed. Mani and Maybury), MIT Press, pp. x-xv,1999.
[19] I. Mani, and M. Maybury (2001), “Automatic Summarization”, in ACL/EACL 2001 Meeting, 2001.
[20] M. T. Maybury(1995),“Generating Summaries From Event Data”, in *Information Processing and Management: an International Journal*, 31(5), pp. 735-751, 1995.
[21] K. Spärck Jones (1998), “Automatic summarising: factors and directions”, available at http://arxiv.org/PS_cache/cmp-lg/pdf/9805/9805011.pdf (2005/01/26).
[22] K. Spärck Jones and J. R. Galliers(1996), “Evaluating Natural Language Processing Systems: An Analysis and Review”, Springer-Verlag, 1996.
[23] 吳家威(2003), “新聞自動摘要方法之研究與探討”, 碩士論文, 國立政治大學資訊科學研究所, 台北, 2003.
[24] 曾元顯(1997) , “關鍵詞自動擷取技術與相關詞回饋” ,中國圖書館學會會報,第59期, 1997. available at . http://blue.lins.fju.edu.tw/~tseng/papers/feedback.htm (2005/01/26).
[25] 盛玉麒 ,“基於語料庫的漢語字詞相關性研究”, available at http://www.yyxx.sdu.edu.cn/content/xueshuyanjiu/xueshu2-syq2.htm (2005/01/26).
[26] 黃居仁(2003),“語意網、詞網與知識本體：淺談未來網路上的知識運籌”, 佛教圖書館館訊 , 第33期 , pp. 6-21, 2003. available at http://corpus.ling.sinica.edu.tw/paper/churen/c_onto-intro-0304.doc (2005/01/26).
[27] 葉鎮源(2002), “文件自動化摘要方法之研究及其在中文文件的應用”, 碩士論文, 國立交通大學資訊科學研究所, 新竹, 2002.
[28]曾志軒、陳意芬、余明哲、蒙以亨、柯皓仁、楊維邦(2001),“中文結構化文件之語意索引”, available at http://www.ccu.edu.tw/TANET2001/scheduel/paper_abs/P103.html
[29] 蔡志浩(1996) ,“淺論中文斷詞”, available at http://bbs.ee.ntu.edu.tw/boards/Programming/15/16.html (2005/01/26).
[30] 廖鼎銘、劉昭麟(2003) ,“賭博案件起訴書文句分類之研究”, available at http://www.cs.nccu.edu.tw/~chaolin/papers/liao03.pdf (2005/01/26).
[31] 廖嘉新等,“以SAO 物件為基礎之中文專利文件摘要方法及架構”, 第十五屆物件導向技術及應用研討會, 2004.
[32] 謝宛諭,“快速深度分析專利文獻”, 勢流科技 , available at http://www.patent104.idv.tw/pat10.htm (2005/01/26).