

CAC—SDI 서어비스의 利用現況 分析

李 靜 玉*
申 鉉 雄*

1. 머리말

最近 科學技術이 急激히 發達함에 따라 發表되는 學術雜誌 및 研究報告書, 特許明細書 등도 매년 爆發적으로 增加되고 있다. 따라서 研究者 個個人이나 小集團의 힘만으로는 氾濫하는 새로운 情報의 洪水속에서 그들이 必要로 하는 모든 情報를 適時에 個別的으로 入手하기가 거의 不可能하게 되었다. 따라서 이와 같은 狀況에 對處하고 大量情報속에서 要求되는 情報를 迅速 正確하게 檢索하기 위하여 韓國科學技術情報센터(KORSTIC)에서는 1975년 7월 1일부터 化學 및 化學工學分野의 既成 데이터베이스인 CA Condensates(CAC)와 그 檢索프로그램을 導入하여 컴퓨터에 의한 最新情報檢索서어비스(SDI: Selective Dissemination of Information)를 開始하게 되었다. 이것은 現在를 中點으로 하여 未來의 情報를 定期的으로 찾아주는 것으로서 RS(Retrospective Service: 遡及文獻檢索) 서어비스와 區別된다. 그동안 研究活動에 從事하는 研究者, 技術者, 科學者들을 對象으로 이 서어비스의 必要性和 利用方法 등을 周知시키기 위한 세미나를 여러번 가진 바 있어 이제는 200여명의 利用者들이 이 서어비스를 받고 있을 뿐만 아니라 이 事業은 또한 본 센터의 비중 큰 事業의 하나로 發展되었다. 本稿에서는 이 서어비스에 대해 간단한 案內와 함께 약 2년간 遂行하여온 서어비스의 定量的인 分析을 해보고자 한다.

2. CAC 데이터베이스란?

CA(Cheical Abstracts)는 世界 各國에서 刊行하고 있는 化學, 化學工業 및 이와 關聯된 모든 分野의 學術雜誌, 特許, 技術報告書, 論文, 政府刊行物, 單行本 등의 抄錄이 年間 약 40萬件이나 收錄되는 世界 最大의 抄錄誌로 週 1回 冊字形으로 出版되고 있다. CAC(Cheical Abstracts Condensates)는 CA와 같이 每週 磁氣 테이프 形態로 作成되는 것이며 收錄事項도 CA와 同一하나 다만 抄錄대신 一連의 重要語(Keyword)로 바꾸어 놓은 것만이 다르다.

2.1 收錄範圍와 內容

CAC는 1주일에 1회 홀수週와 짝수週로 나뉘어져 作成되는데 홀수週에 發行되는 分野는 有機化學과 生化學으로 C1이라 불리워지며, 짝수週에 發行되는 分野는 高分子, 物理, 分析化學과 應用化學, 化學工學으로 C2라 불리운다.

테이프에 收錄되는 內容으로는 著者名, 標題 資料名, 卷, 號, 페이지, 資料의 種類, 使用語, 發行年月日 등 書誌的 事項과 文獻의 內容을 表示하는 Keyword가 있다.

3. 서어비스의 種類

이 서어비스를 利用하려면 利用者 各者가 必要로 하는 主題에 대해서 電子計算機가 情報를 檢索해 낼 수 있는 形態(Profile)로 만들어 줌

*KORSTIC 電子計算室

錄해 두어야 하는데 이 Profile의 作成은 利用者 스스로가 必要한 主題에 따라 作成하는 方法도 있고 또 본 센터의 擔當者와 充分한 의견을 교환하여 利用者의 要求가 充分히 反映된 Profile을 본 센터에서 作成하는 경우도 있다. 現在 본 센터에서 하고 있는 서비스에는 標準主題서비스와 要求主題서비스의 두가지 形態가 있는데 標準主題서비스(Standard SDI Service)는 본 센터에서 利用者의 要求가 많을 것이라고 생각되는 標準主題리스트를 미리 選定하여 그 리스트 중에서 利用者が 要求하는 主題를 選擇하여 檢索서비스를 받는 것으로서, Profile의 內容을 修正할 수 없는 반면 價格이 低廉하다.

要求主題서비스(Request SDI Service)는 利用者が 要求하는 主題가 標準主題리스트 속에 包含되어 있지 않은 경우에 利用者と 各 分野 擔當者間에 充分한 意見을 교환하여 要求가 充足된 Profile을 作成하는 것으로서 만일 檢索된

情報가 利用자가 要求하던 情報가 아닌 경우는 擔當者와의 手摺 對談으로 Profile 內容의 修正이 가능하다.

그러나 데이터베이스안에 아무리 좋은 情報가 들어있다 하더라도 利用者と Profile作成者와의 사이에 여러가지 문제점, 즉 專攻分野가 다르던가 意見의 誤傳 등으로 인해 必要로 하는 情報의 檢索이 제대로 되지않아 Profile이 잘못 作成되었으면 願하는 情報를 提供받을 수 없다. 따라서 主題를 잘 把握하고 있는 利用者 自信이 Profile을 作成하는 것이 가장 바람직하다. 이를 위하여 본 센터에서는 수시로 Profile作成法에 관한 세미나도 갖고 產業界와 學界를 直接 訪問하여 Profile作成法에 관한 講義를 하고 있다.

4. CAC-SDI의 利用現況

4.1 地方別 利用現況

表 1. 地方別 利用現況

건수	지역	서울	부산	경기	경북	경남	충북	충남	전북	전남	강원	제주	합계
접수건수		111	35	20	7	4	0	7	2	3	0	0	189
백분율 (%)		58	19	11	4	2	0	4	1	1	0	0	100

表 1에서 보는 바와 같이 利用者의 大多數가 서울에 集中되어 있고 그 다음이 부산과 경기도로서 비교적 대도시에서 利用자가 많다. 위의 表에서 현저히 나타난 것은 大都市일수록 經濟, 文化, 技術의 中心이 되고 있다는 것이다. 또한 서울과 가까운 지역에서 利用자가 많고 강원도 및

제주도에서는 아직 1件도 接受되지 않은 점으로 보아 정보교환에는 거리에 의한 교통통신의 문제점이 있음을 알 수 있다.

4.2 分野別 利用現況

表 2. 分野別 利用現況

건수	분야	식품	화학	원자력	섬유	전기전자	제약	종이	피혁	제강	페인트	타이어	시멘트	합계
접수건수		19	76	19	18	17	11	5	4	6	9	3	2	189
백분율 (%)		10.5	40.2	10.3	10	9.5	5	2	2	3	4	1.5	1	100

分野別 利用은 表 2에 나타난 바와 같이 CAC가 化學分野인만큼 分類하기 어려운 化學, 化工

分野가 主종을 이루고 있으며 食品, 원자력, 섬유계통의 分野는 한국원자력연구소가 활발히 利

用하고 있고 또 식품이나 섬유분야의 業체가 大企業들인 만큼 利用件數도 많음을 알 수 있다.

제약과 페인트 分野 등은 다른業種에 비해 새로운 情報를 쉽게 실험에 옮겨 볼 수 있음에도 불구하고 지난 해에 비하여 현저히 이용자수가 감소하였다. 주목할 만한 일은 CAC가 化學分野에 주종을 이루고 있음에도 불구하고 원자력분

야와 전기, 전자, 물리분야의 利用이 활발한 것으로 볼 때 INIS(원자력분야 데이터베이스), INSPEC(전기, 전자, 물리, 컴퓨터분야의 데이터베이스) 등의 國內 서어비스도 시급히 서둘러야 할 것이다.

4.3 利用機關 現況

表 3. 利用機關 現況

이용기관 건수	KAERI	KAIS	홍능기계	KIST	기타 연구소 검사소	대 학	협 회	생산업체	개 인	기 타	합 계
접 수 건 수	18	7	5	6	5	12	7	117	9	3	189
백분율 (%)	10	3.7	2.6	3.2	2.6	6.1	3.7	62	4.7	1.4	100

表 3에서 보는 바와 같이 生産業체가 利用者의 주종을 이루고 있다. 그러나 利用生産業체의 數가 前年度에 비해 減少하고 있다는 것은 CAC-SDI 서어비스 자체에도 問題가 있겠지만, 最新 해외기술정보에 대한 最高經營者들의 인식부족 및 기술정보관리자들의 설득력 부족에도 있다 하겠다. 우리나라 科學技術研究의 核心을 이루고 있는 韓國科學技術研究所(KIST)를 包含한 其他 연구기관의 利用數가 작년도에 비해 減少 내지는 踏步狀態에 있는 것도 海外 科學技術情報에 대한 消極的인 姿勢를 말해주고 있다.

重化學工業의 持續的인 발전과 플랜트 輸出에 必要한 最新情報를 迅速하고 多樣하게 받아들이기 위하여는 業체와 연구기관이 혼연일체가 되어 最新技術情報에 積極的인 姿勢로 임하여야 하며 본 센터에서도 活潑한 弘報活動과 꾸준한 說得으로 利用者 擴大에 努力을 아끼지 말아야 할 것이다.

4.4 標準主題와 要求主題別 利用現況

表 4. 標準主題와 要求主題別 利用現況

서어비스종류 건수	요구주제	표준주제	합 계
접 수 건 수	119	70	189
백분율 (%)	63	37	100

위의 표에서 보는 것과 같이 서어비스 種類別

로 分類하여 보면 전체 利用者中에 63%가 要求主題를 選擇하였다. 이것은 研究主題에 따라 利用者 各者의 Profile 內容이 다르고 標準主題 리스트에도 있지 않은 細分된 特定分野에 대한 研究가 많음을 알 수 있다. 그러나 標準主題 接受件數도 적지않은 數를 차지하고 있는데 그것은 要求主題보다 우선 가격이 저렴하다는 이유도 있겠지만 광범위하게 情報를 檢索하여 점차 원하는 情報의 形態로 범위를 좁혀가며 접근해 가는 方法의 하나로 使用되어 가고 있기 때문이다. 그러므로 본 센터에서는 많은 실무자와 연구진들이 利用者들에게 보다 적합한 情報를 提供해 주기 위해서는 標準主題의 리스트를 많이 만들고 利用者들과 충분한 대화를 통해 檢索해 나가는 것이 바람직한 일이라고 볼 수 있다.

5. 檢索소요시간에 대한 분석

表 5는 CAC Vol. 83, 84, 85의 모두 78卷의 테이프를 對象으로 서어비스한 結果를 평균 Profile數, 檢索시간, 테이프當 包含된 기사수 및 檢索기사수 별로 작성해 본 統計이며 表 6은 CAC Vol. 83의 26卷 테이프에 대해 서어비스한 結果를 테이프當 수록기사수, 檢索에 使用된 프로파일 수 및 檢索에 소요된 시간별로 도표화한 것이다.

구입한 테이프를 標準 Format으로 바꾸어 주

表 5.

Vol.	내용 분류	(A)	(B)	(C)	(S)	(D)	(E)	(F)	(%)
		기 사 수	프로파일수	테이프변환 소요시간 (초)	검색 소요시간 (초)	검색건수	프로파일당 검색건수 (건) <D/B>	프로파일당 평균 검색시간 (초) <S/B>	전체데이터 베이스에 대한 프로파일당 검색건수의 비율 <D/A>
83	C1	69,328	28.1	623.4	373.1	655	23.6	13.9	9.4
	C2	9,832	120.5	757.4	1,121.3	5,039	42.5	7.6	53.1
84	C1	6,281	44.3	534.5	478.3	1,321.3	32.4	10.8	21.1
	C2	8,207	153.7	713.3	1,348.7	5,050.8	33.0	8.5	61.9
85	C1	6,862	35.7	605.8	472.3	1,356.7	38.2	13.27	19.5
	C2	8,327	141.9	801.5	1,245.7	5,109.8	36.2	11.1	64.4

는 테이프 변환 시간을 (C)라 하고 실제 검색에 소요된 시간을 (S)라 한다면 실제 검색을 위해 소요된 기계 시간은 (C)+(S)가 된다. 表 5의 (C)에서 나타난 바와 같이 컴퓨터를 CDC機種으로 使用하기 때문에 거쳐야 하는 테이프

변환과정에 소요되는 시간이 전체 기계사용시간의 상당부분을 차지하고 있다. 테이프 변환은 테이프로 收錄된 모든 記事를 하나씩 읽어 標準樣式으로 바꾸어 주는 過程이기 때문에 이에 소요되는 시간은 收錄된 기사수에 비례하고 있다.

表 6.

권	호	(A)	(B)	(C)	(S)	(D)	(D/B)	(S/B)	(D/A)
		기 사 수	프로파일수	테이프변환 소요시간 (초)	검색소요 시간 (초)	검색건수	프로파일당 검색건수 (건)	프로파일당 평균 검색 시간 (초)	전체데이터 베이스에대한 프로파일당 검색건수의 비율 (%)
83	1	10,891	22	850.3	523.2	694	31.5	23.8	6
	2	11,178	86	919.3	1,065.2	4,339	50.5	12.4	38
	3	6,477	22	508.2	333.8	498	22.6	15.2	7
	4	8,719	86	698.3	858.2	3,581	41.6	10	41
	5	6,466	22	503.2	299.8	344	15.6	13.6	5
	6	9,243	97	737.2	910.2	3,566	36.8	9.4	38
	7	6,264	22	495.2	299.8	388	17.6	13.6	6
	8	11,939	97	968.3	1,298.2	6,197	63.9	13.4	51
	9	8,404	21	658.2	374.1	438	20.9	17.8	5
	10	10,839	110	877.3	1,201.2	4,950	45	10.9	45
	11	7,478	21	591.2	343.9	460	21.9	16.4	6
	12	10,040	110	790.2	1,107.2	4,366	39.7	10.1	43
	13	6,953	41	547.2	427	565	13.8	10.4	8
	14	10,960	135	891.3	1,639.3	7,450	55.2	12.1	67
	15	6,016	41	479.1	361.8	465	11.3	8.8	7
	16	9,606	135	773.2	1,397.2	6,100	45.2	10.3	63
83	17	6,191	30	482.2	374.9	922	30.7	12.5	14

	18	9,847	145	793.2	1,467.2	6,019	41.5	10.1	61
	19	7,011	30	551.2	414.1	969	32.3	13.8	13
	20	7,563	145	610.2	1,325.2	4,896	33.8	9.1	64
	21	7,378	30	600.2	435.1	1,114	37.1	14.5	15
	22	8,421	135	679.2	1,257.2	5,563	41.2	9.3	66
	23	5,697	32	454.1	363.7	946	29.6	11.4	16
	24	7,914	143	649.2	1,182.2	4,929	34.5	8.3	62
	25	4,901	32	384.1	299.7	711	22.2	9.4	14
	26	5,696	143	459.1	869.1	3,550	24.8	6.1	62
	C1	90,127	366	7,104.9	4,850.9	8,514	307.1	181.2	122
	C2	121,965	1,567	9,846	15,577.6	65,506	553.7	131.5	701
	C1+C2	212,092	1,933	16,950.9	20,428.5	74,020	1,173.5	312.7	823

表 6에서 보면 Vol. 83의 Issue 1, 3, 5, 7은 각기 모두 22개의 프로파일을 대상으로 檢索을 수행하였으며 이에 소요된 檢索時間은 각각 523.2초, 333.8초, 299.8초, 299.7초(Profile當 檢索時間은 23.8, 15.2, 13.6, 13.6)였다. 이 수치들을 이들 테이프에 收錄된 기사수와 비교하여 보면 같은 수의 프로파일로 檢索하는 경우 테이프에 收錄된 기사수가 많을수록 檢索時間이 많아짐을 알 수 있다. 또한 기사수가 거의 비슷한 Vol. 83 Issue 1과 Vol. 83 Issue 10의 프로파일당 평균 檢索時間(表 6의 (F))을 보면 Issue 10은 Issue 1에 비해 프로파일 수가 약 5배임에도 檢索時間은 약 1/2에 불과하다. 이것은 檢索프로그램이 檢索효율을 높이기 위하여 200개의 프로파일을 單位로 하여 探索을 행하기 때문이다. 따라서 檢索 시스템의 효율은 프로파일 수가 200의 培數일 때 가장 높다. 같은 프로파일이라 할지라도 데이터베이스에 包含된 內容에 따라 檢索件數가 달라지기 때문에 包含된 기사수와 檢索건수는 비례하지 않는다. 이것은 表 6중에서 收錄記事數 및 프로파일수가 비슷한 C1 파일들의 통계를 비교하면 알 수 있다.

그러나 프로파일당 檢索기사수와 이에 소비된 檢索시간을 比較하여 보면 檢索件數(Hit 수)가 많을수록 그 소요시간이 길다. 이것은 檢索된 레코드에 대해 一定한 處理를 기계 内部에서 해주기 때문이다.

表 7은 特定한 4개의 프로파일에 대하여 Vol. 86 Issue 13~Issue 16까지 모두 4개의 테

이프를 使用하여 檢索한 結果이다. 이중 A1과 A2, B1과 B2는 사실상 같은 內容의 프로파일로 단지 A2와 B2에서는 *L 카아드를 使用하여 探索 범위에 制限을 주었을 뿐이다. 여기에서 나타난 바와 같이 *L 카아드를 使用하는 경우의 檢索時間은 使用하지 않는 경우의 약 1/2~1/3로 현저히 짧다. 檢索結果面에서 보면 B1과 B2의 檢索건수의 경우와 같이 *L 카아드로 制限하는 경우 除外되는 기사가 많이 생길 가능성은 있으나 이것은 몇번의 實驗을 거쳐 有效하다고 생각되는 除外된 기사들의 해당 分類를 追加시키다던가 하여 補完할 수 있다. Vol. 86의 Is-

表 7.

Vol/Issue	Profile	檢索시간(초)	檢索건수
8613	A1	1.308	3
	A2	0.726	3
8614	A1	3.156	2
	A2	1.026	0
	B1	2.31	56
	B2	0.588	16
8615	A1	1.308	6
	A2	0.81	6
8616	A1	2.82	7
	A2	0.822	2
	B1	1.476	46
	B2	0.45	16

sue 13, 15에 대한 A1과 A2는 똑같은 檢索結果를 얻으면서 *L 카아드로 檢索時間을 단축시킨 좋은 例이다. 이중 A1은 2개의 Parameter에 모두 29개의 Term을 使用한 多少 複雜한 프로파일이고 B1은 7개의 Term만의 單一 Parameter로 된 프로파일이다. A1에서 檢索된 기사수가 B1의 그것에 비하여 상당히 적음에도 불구하고 더 많은 檢索時間이 소요된 것은 프로파일의 複雜性이 檢索時間에 影響을 주는 좋은 例이다.

以上에서 살펴본 것을 整理하여 檢索을 위한 기계사용 시간에 影響을 미치는 要因을 나열해보면

1. 데이터베이스 內의 기사수
2. 檢索에 使用된 프로파일의 수
3. 檢索된 文獻의 수
4. 프로파일에서의 探索分野 制限
5. 프로파일의 복잡성

등을 들 수 있다. 따라서 프로파일 作成者는 이러한 점을 감안하여 檢索結果 및 시스템 使用面에 있어서 共히 효율적인 프로파일을 作成해야 한다.

6. 맺는말

以上에서 살펴본 것은 現在 提供하고 있는 서어비스의 利用者 및 檢索시스템 自体에 대한 의

형적인 分析에 불과하다. 따라서 이 서어비스에 대한 利用者의 만족도는 測定하지 못하고 있는 事實이다. 물론 서어비스 출력에 檢索文獻數를 프린트하여 주고 이중 요구가 만족된 文獻의 數를 利用者가 記入하여 KORSTIC에 보내도록 함으로써 이 서어비스의 만족도를 測定하고는 있으나 이에 呼應하는 利用者의 數가 적을뿐만 아니라 이 두 숫자만 가지고 만족도를 表示하는 것에는 問題가 있다.

本稿에서 測定이 불가능했던 것의 또 한 가지는 關聯된 文獻의 除外率 및 이러한 情報들이 除外되도록 한 檢索시스템 및 프로파일 技法 등에 관한 定性的인 分析이다. 이것은 프로파일 담당자 혹은 KORSTIC 內部에서 獨立的으로 수행할 수 있는 일이 아니며 이 서어비스의 利用者와 의 緊密한 협조에 의하여야 할 것이다.

제외율을 줄이고 효과적인 檢索을 할 수 있는 시스템으로는 RS와 On-line Service가 있다. RS시스템에서는 遡及檢索이 가능하므로 프로파일의 수행에 의하여 일단 除外되었던 情報를 다시 찾을 수 있고, On-line에서는 質問應答式으로 프로파일을 작성하도록 하여 보다 나은 프로파일 작성을 가능하게 함으로써 SDI의 약점을 보완한다. 본 센터에서는 데이터베이스의 확장과 더불어 RS 및 On-line으로 시스템을 확장함으로써 명실공히 迅速正確한 情報서어비스를 하루속히 이루어야 할 것이다.

<p. 125의 계속>

- 2) FAO : FAO/AGRIS 3 (Rev. 1) (1974)
- 3) FAO : FAO/AGRIS 3 (Rev. 2) (1976)
- 4) FAO : FAO/AGRIS 5 (1973)

- 5) FAO : Agrindex, v.2, n.8 (1976)
- 6) AIBA : AgriAsia, v.1, n.1 (1977)
- 7) 金石章, 情報管理研究, v.7, n.6 p.150 (1974)