

文獻檢索에 있어서 Chemical Abstracts와 CA Condensates의 比較*

B. E. Robert** 著
閔 泳 一*** 譯

抄錄

1975年 3月, 4年半 동안의 Chemical Abstracts 索引과 온-라인이 可能한 CA Condensates를 比較하였다. 두가지 데이터 베이스를 함께 利用하여 檢索하는 方法이 가장 効率的이지만 實例에서 보는 바와 같이 CA Condensates를 檢索하는 것이 보다 實用的이다. System Development Corp社(SDC)에 設置되어 있는 온-라인 形態인 CHEMCON과 CHEM7071을 Chemical Abstracts 索引과 比較하였다.

大部分의 Chemical Abstracts 利用者들은 Chemical Abstracts 책자나 累加索引에는 親熟하지만 CA Condensates는 아마도 그리 親熟하지 못할 것이다. CA Condensates는 서지적 사항을 機械로 읽을 수 있는 形態로 되어, 있고 Chemical Abstracts¹⁾에 따라서 索引되므로 매주 발행되는 Chemical Abstracts 책자의 뒷 부분에 있는 索引과 같이 우리에게 가장 親熟한 形態로 되어 있다. Chemical Abstracts가 現在 사용하고 있는 데이터 베이스이지만 本稿에서는 Index와 Condensates를 둘 다 데이터 베이스로 定義한다.

Condensates가 美國의 Chemical Abstracts Service 기관으로부터 商業적으로 利用할 수 있게 되자 여러 情報센터에서는 利用者들의 프로파일을 batch方式으로 處理하여 매주 나오는 磁氣테이프에서 最新情報를 檢索하여 提供하는 서어비스(SDI)를 시작하였다. 어떤 情報센터들은 지나간 磁氣테이프를 모아서 역시 批方式으로 溯及文獻檢索 서어비스를 한다. 磁氣테이프를 직접 取扱하는 사람들을 除外하고는 大部分 Condensates를 아직 잘 모르고 있다. 溯及文獻檢索은 비용이 다소 비싸고 두서없이 이것 저것 文獻을 檢索하는 方法은 實用的이 못된다. 매주 나오는 索引에 대해서 두 개나 그 이상의 概念이나 物質을 組合하여 檢索하는 方法은 어렵고 實用的이 못된다. 오히려 주어진 用語 아래에 있는 모든 引用文들을 보고 抄錄과의 關連性을 決定하는 것이 때때로 더 쉽다.

相互 作用하는 온-라인 檢索을 위한 Condensates의 有用성은 많은 변화를 가져 왔다. 必要한 文獻만을 檢索해 보는 것이 이제 可能하고 어떤 項目에 대해서도 完全히 索引할 수 있게 되었다. 批 시스템으로는 檢索을 시작해서 그 結果를 받아 볼 때 까지 數時間에서 며칠까지 걸리는 번거로운 時間差를 이제는 보통 단 몇 분으로 줄일 수 있다. 그리고 批 시스템과는 달리 부정확하거나 불충분한 檢索方法은 즉시 고칠 수가 있다. 연속적인 批 형태의 檢索方法²⁾에 비해서 順序없이 온-라인으로 檢索하는 方法이 分明하고 正確한 長點이 있다.

CA Condensates를 자주 利用하게 되자 그의

* Searching Chemical Abstracts VS. CA Condensates, J. Chem. Inform.

Comp. Sci. Vol. 15, No. 3, 1975. pp. 174~176.

** Standard Oil Company, Amoco Research Center.

*** KORSTIC 電子計算室

진정한 價値에 대해 論義가 있었다. CA Condensates의 索引方法은 Chemical Abstracts 책자나 累加索引의 方法보다 확실히 덜 體系的이고 徹底하지 못하다. 더우기 두 데이터 베이스는 重複된 것이 많으므로 重複해서 檢索할 價値가 있는지 없는지를 決定해야 한다. 다른 몇 개의 데이터 베이스와 CA Condensates를 比較한 論文들이 여러 번 發表되어 왔는데 一般的으로 CA Condensates는 下位의 데이터 베이스로 나타났다. Buckley는 Chemical Abstracts 索引이 CA Condensates보다 더 좋은 參考文獻(테라마이신의 제법에 관해서)을 提供한 實例³⁾를 引用하였다. 조오지大學의 Search Center는 CA Condensates가 CA Integrated Subject File⁴⁾보다 機能이 못하다는 것을 알았다. CA Condensates의 다른 여러 가지 形態들을 또한 比較하였다. Michaels⁵⁾은 CA Condensates를 온-라인으로 檢索한 것과 매주 나오는 Chemical Abstracts 책자의 索引을 手作業으로 檢索한 것을 比較한 論文을 發表하였다. 그리고 Prewitt⁶⁾은 온-라인으로 蓄積한 두 개의 商業用 CA Condensates를 比較하였다.

Amoco Research Center에서도 CA Condensates와 Chemical Abstracts 索引의 檢索結果를 比較하고 CA Condensates의 長點과 索引의 長點; 그리고 事實上 서로 同等하다는 實例를 發見하였다. 1975年 3月, 적어도 4年分の CA Condensates와 索引(Vols 72-79, 1970-1973)을 比較하였다. 著者와 一般 主題에 대한 檢索은 Vol 80 (Jan-June, 1974)을 使用하여 比較하였다. CA Condensates는 보통 細分化된 複合物을 檢索하는데 不便하다. Buckley가 提示한 實例가 그 代表的인 例이다. 그러나 다른 形態의 檢索實例(團體著者, 特許受託者, 個人著者, 一般的인/細分化된 化合物 그리고 反應種類)로 實際的인 檢索을 위한 CA Condensates의 利點을 例示하였다. 다음 實例에서 CHEMCON과 CHEM7071은 CA Condensates를 온-라인으로 入力시킨 것이다.

檢索實例

團體著者; 團體著者の 例로서, Chemical Abstracts Service 그 自体를 索引하는 方法과 그의 製品을 알아보는 檢索을 試圖해 보았다. 온-라인 方法은 다음과 같다.

((CHEM OR CHEMICAL) AND (ABSTR OR ABSTRACT# OR CONDENSATES)) OR CAS OR ((CHEM/CS OR CHEMICAL/CS) AND (ABSTR/CS OR ABSTRACTS/CS))
 “#”나 “/” 표시는 System Development Corp社의 새로운 作動프로그램인 ORBIT III에서 단지 하나의 文字를 나타내며 一般的으로 한 用語의 단수와 복수를 동시에 意味한다. Vol.72부터 80까지의 索引에서 “Chemical Abstracts Service” 索引 밑에는 단지 16개의 引用文밖에 없지만 위의 온-라인 方法으로는 더 많은 引用文들이 檢索되므로 CAS와 그의 製品은 다른 주요 제목 밑에 副題目으로 표시됨이 分明하다. CA索引 案內書안에 있는 “情報科學”이란 말 밑에 “參考”라고 해서 “索引”과 “文獻”, 그리고 “定期刊行物”을 모두 包含한다. CAS에 관련된 用語를 찾기 위해 4가지 題目들을 자세히 살펴보면 두 데이터 베이스가 똑같이 效果的이라는 事實을 말해주는 그외의 參考事項을 얻을 수 있다(表 1).

表 1. “Chemical Abstracts”를 위한 引用文 檢索의 比較

	단독 인용문	공통된 인용문
CA Condensates	16	28
CA 索引	17	

각 데이터 베이스에서 檢索된 단독 인용문의 수가 이중으로 檢索할 價値가 있다는 것을 말해준다.

이렇게 二重으로 檢索하는 데는 몇 가지 理由가 있다. CA 索引에서는 CAS나 CA Condensates에 관련된 그외의 引用文들이 主題 밑에 副題目으로 發見될 수도 있겠지만 위의 統計 숫자가 二重으로 檢索하는 것이 가장 實用的이라는 것을 나타낸다. CA Condensates는 CA索引에 있는 引用文들을 다시 CA Condensates에 옮기므로 이따금 너무 一般的으로 索引을 한다. 이 引用文들은 자주 “情報서어비스”로 索引된다.

온-라인에서는 “團體探索要素”를 사용하여檢索效率를 높인다.

반면에·CA 索引에서는 다만 機關著者에 特許만을 引用한다. CHEMCON에서는 “團體 探索要素” 分野가 團體著者, 特許受託者 그리고 作業場 住所 分野에서 나오는 단어들을 모두 포함한다. 그래서 그것은 어떤 種類의 文獻이라도 그의 機關 探索要素를 決定하는 데 重要하고 個人著者나 重要語들과 같은 그밖의 探索要素들과도 함께 使用할 수 있다.

“團體檢索要素”(現在 CHEMCON에서만 有用함)가 特許受託者 機能에만 局限되었다 할지라도 檢索效率를 向上시킨다. CA 索引에서 各機關의 特許들은 抄錄番號대로 인쇄되어 있지 主題別로 인쇄되어 있지 않다. 상당히 많은 수의 特許를 내는 機關에서는 이따금 必要없는 標題를 꼼꼼히 들여다 보아야 될 일들이 굉장히 많다. 온-라인 檢索에 있어서 重要語와 標題는 機關著者名과 함께 많이 使用된다. CA Condensates의 再現率을 試驗하는 데 Vol. 76부터 80까지 CA 索引에서 찾아낸 “Standard Oil Co. 社”의 155개 特許 引用文들이 온-라인에서도 역시 모두 檢索되었다. 온-라인 檢索은 그밖의 다른 이름을 가진 그의 여러 종속회사들이 特許를 내는 경우에는 더욱 價値가 있다.

個人著者; 온-라인 檢索에서 個人著者로 찾는 경우에 CA 索引과는 달리 여러 사람의 이름으로 된 文獻을 이름의 順序에 관계없이 직접 찾을 수 있기 때문에 檢索速度가 훨씬 빠르다. Princeton大學의 Paul Schleyer 교수가 쓴 文獻들을 檢索하는 例가 그 좋은 實例이다. P. Schleyer에 “Neighbor” 命令語를 使用하면 CA Condensates(1972-Feb. 1975)에서 다음과 같은 이름들이 나온다.

Schleyer P. R. Schleyer P. Von Rague
Schleyer P. V. R. Schleyer Paul R.
Schleyer P. Von R. Schleyer Paul V. R.
 Schleyer Paul Von R.

위의 모든 이름들은 “Schleyer P:”에 의해서 모두 檢索된다. 온-라인에서 서지적 사항들을 좀더 자세히 주면 數分내에 모든 引用文들을 調査할 수 있다. Paul Schleyer가 비록 CA 索引에

서 단지 2 가지 方法으로 적혀져 있다 할지라도 모든 논문마다 그의 첫 이름이 나오는 것이 아니므로 그밖에 적당히 이름의 첫자만 만 관련된 모든 文獻들도 檢討해야 한다. CA 索引에서 抄錄番號를 알아도 더 자세한 서지적 사항까지 알도록 調査해 보는 것이 좋다.

複合物; 特定한 複合物을 檢索하는 데는 CA 索引이 CA Condensates보다 자세하고 세밀하게 索引하기 때문에 CA 索引이 보다 월등하다. 索引하는 方法에 대해서 Chemical Substance File과 CA Condensates를 徹底히 比較해 보았더니 별로 심한 差異點은 없었지만(4), CA Condensates는 관심이 있던 複合物을 索引하지 않은 文獻들이 많았고 CA 索引에 비해서 눈에 띄게 索引數가 적었다. CA Condensates의 包括的인 特性은 特定한 複合物에 대해서 特有性의 喪失(例: 모든 키실렌 이성체는 “키실렌”으로 索引)과 複合物의 種類別(例: 키실렌을 “알킬벤젠” 種類로 索引)로, 索引하므로 正確度가 問題視 된다. 그러나 이 包括的인 索引은 檢索要求가 細分化하지 않고 다소 一般的인 때에는 利點이 있다. 例를 들면, P-키실렌의 反應에 關心이 있는 利用者는 다른 키실렌의 反應이나 심지어 一般的인 알킬벤젠의 反應까지 關心이 있을지 모른다. 위에서 보는 例 이외에 環줄기나 다른 이름의 語幹들은 보통 단독으로 나타난다. 例를 들면, BIPHENYL# (biphenyl 또는 biphenyls)는 CA Condensates에서 biphenyl 環조직을 다루는 大部分의 引用文을 檢索한다.

CA 索引에서 主要 題目으로 나타나는 이름 語幹(例: biphenyl)으로 複合物을 實體로 또는 그 複合物의 種類안에서 檢索하는 일은 지루하지만 모든 題目들을 자세히 훑어보면 찾아낼 수 있다. 그러나 어떤 複合物의 種類 이름이나 이름의 語幹은 “에텔”이나 “캐톤”과 같이 特定한 複合物의 主要 題目으로 이따금 나타나지 않는다(또는 결코 나타나지 않는다). 또 다른 例는 前に 主要 題目으로 나타났던 “nitron”이 이제는 轉化되지 않은 形態로 “N-oxide”가 따르는 amine의 이름 밑에 索引되는 例이다. 이것은 General Subject Index에서 “Nitrones”라고 하는 題目에 관한 단지 一般的인 考察에 불과하다. “Nitron”

이나 Nitrones"에 관한 引用文의 統計 숫자는 8th 와 9th Collective 索引기간으로 나뉘어 表 2 에 나타나 있다. 一般的인 檢索目的을 위해서는 分明히 索引이 實用的이다.

表 2. "Nitrones"의 引用文 檢索의 比較

	단독 인용문	공통된 인용문
CHEM 7071	58	
8th Collective (1970-71)	19	34
CHEMCON	113	
Volume Indexes (1972-June 1974)	5	22

이따금 無機化合物의 包括的인 檢索이 必要할 때가 있다. 例를 들면 sodium hydroxide나 sodium bicarbonate 또는 sodium carbonate 등을 檢索할 경우, 온-라인 檢索으로는 "sodium"만 의 간단한 調整으로 充分할 수 있다. 그러나 索引에서는 관심사에 따라 각각 다른 主要 題目과 副題目으로 3 가지 化合物에 대해서 모두 檢索해야 한다. 모든 sodium 化合物에 대해서 關心이 있다면 온-라인 檢索은 보다 더 效率的이다. 陰이온 (例: Sulfate나 Sulfates)에 관한 全般的인 檢索을 할 경우, 索引題目은 보통 매우 一般的이므로 CA Condensates의 檢索이 索引을 檢索하는 것보다 더 效率的이다.

情報은 이따금 全般的이면서 또한 細分化된 檢索能力을 必要로 한다. 有機半金屬에 관한 情報를 자주 CA Condensates에서 便利하게 檢索한다. 例를 들면, lead alkyl分野에 관한 一般的인 檢索은 온-라인 檢索을 위하여 "OR" 節의 方法으로 다음과 같이 一部 作成하였다.

ORGANOLEAD# OR TETRAALKYLLEAD
OR TETRAARYLLEAD# OR TRIAL-
KYLLEAD# OR TRIARYLLEAD# OR DI-
ALKYLLEAD# OR DIARYLLEAD# OR
ALKYLLEAD# OR ARYLLEAD#

그외에 tetramethyllead, trimethyllead, tetraethyllead 등의 用語들을 必要하다면 더 添加할

수 있다. 결국 "lead"가 들어 있는 위 用語들을 "OR" 節로 作成하여 添加하면 正確度는 떨어져지만 再現率은 훨씬 증가시킬 수 있다. CA 索引에서 그런 化合物들은 例를 들면 Plumbane 가 같은 題目 밑에 들어갈 수 있다. 또 그런 目錄들은 累加索引에서 여러 페이지에 계속될 수 있다. 이런 경우에 CA Condensates로 檢索하면 아주 實用的이다.

反應種類; 끝으로 두 데이터 베이스에서 工程이나 反應種類에 관한 情報를 檢索해 보면 각 데이터 베이스의 效用을 알 수 있다. 그러나 두 데이터 베이스를 합쳐서 檢索한 再現率은 어느 한 데이터 베이스만을 사용한 경우보다 顯著하게 높다. "decarbonylation"의 경우에 CA Condensates보다 CA 索引에서 단독 引用文들이 더 많이 檢索되었지만 두 데이터 베이스를 합쳐서 檢索한 引用文들은 그 數가 훨씬 더 많다. 그 例를 表 3에서 본다.

表 3. "Decarbonylation"에 관한 引用文 檢索의 比較

	단독 인용문	공통된 인용문
CHEM 7071	18	
8th Collective (1970-71)	21	44
CHEMCON	15	
Volume Indexes (1972-June 1974)	17	61

結論

지금까지의 實例로 完全하지는 못하지만 그것이 代表的이고 그것에 비추어 다른 檢索의 結果를 미루어 짐작할 수 있다. CA Condensates의 短點은 特定한 複合物을 檢索하거나 다른 곳에서 發見되는 特定한 項目들 (抄錄番號나 著者名 등에 의해서)을 찾는 일, 또는 索引한 모든 레코드를 프린트하지 못하는 점 등에서 나타난다. 그러나 이와 같은 약간의 缺點이 있음에도 불구하고 CA Condensates는 現在 가장 實用的인 檢索方法으로 알려져 있으며, 다소 重複된 點이 있다 하더라도 一般的으로 CA 索引과 兼用

하여 檢索하면 顯著하게 높은 檢索效率을 나타낸다.

参 考 文 獻

- 1) "Information Tools 1975. Book One," Chemical Abstracts Service, Columbus, Ohio p.18, 1975.
- 2) Lancaster, F. W., and Fayen, E. G., "Information Retrieval On-line," Melville Publishing Co., Los Angeles, Calif., 1973.
- 3) Buckley, J. S., Jr., "Planning for Effective Use of On-line Systems," 168th National Meeting of the American Chemical Society, Atlantic City, N. J., Sept 1974, Abstract CHLT 14.
- 4) Zipperer, W. C., Park, M. K., and Carmon, J. L., "The CA Integrated Subject File, II. Evaluation of Alternative Data Base Organization," J. Chem. Doc., 14, 15~23, 1974.
- 5) Michaels, C. J., "Searching CA Condensates On-line VS. the CA KWIC Indexes," 169th National Meeting of the American Chemical Society, Philadelphia, Pa., April, 1975, Abstract CHLT 6; J. Chem. Inf. Comput. Sci. 15, p. 172, 1975.
- 6) Prewitt, B. G., "Searching the Chemical Abstracts Condensates Data Bass Via Two On-line Systems," 169th National Meeting of the American Chemical Society, Philadelphia, Pa., April 1975, Abstr. CHLT 8; J. Chem. Inf. Comput. Sci., 15, p. 177, 1975.