

對話方式의 特徵을 가진 SDI 시스템의 性能*

V. Saskova
J. Kosik 共著
閔 泳 一 **譯

抄 錄

利用者 中心의 對話方式 시스템을 開發하고 試驗하였다. 2年동안 280개의 프로파일로 CA - Condensates 데이터 베이스내에서 情報를 檢索해 보았다. 그 시스템의 作業能率을 使用한 데이터 베이스와 하드웨어의 構造, 소프트웨어 패키지, 利用者의 數, 그리고 情報센터에서 提供한 補助에 따라 評價하였다. 對話式 檢索方法이나 利用者의 要求를 量化하는 것, 그리고 여러가지 分野別로 檢索하는 方法 등 利用者의 要求를 보다 充實하게 滿足시킬 수 있는 여러가지 方法들을 論議하였다.

모든 情報시스템의 目的은 適時에 올바른 情報를 適者에게 提供하는 것이라고 잘 알려져 있다. 다시 말해서, 情報시스템의 主要目的은 利用者의 要求를 充足시켜 주는 것이다. 이러한 觀念의 變化는 情報處理에 큰 變化를 가져와 情報시스템의 適應性을 增加시키기 위해 여러가지 새로운 方法들이 開發되었다. 情報시스템들 사이의 相互作用, 사람과 컴퓨터와의 對話, 그리고 컴퓨터의 助力으로 利用者의 프로파일을 作成하는 것 등은 利用者의 要求를 尊重하고, 必要한 情報를 쉽게 찾기 위해서 考案된 것이다.

한편, 우리들은 利用者의 特性이나 背景, 그리고 文獻에 대한 그의 習性 등에 대해서 얼마나 모르고 있는가를 깨닫게 된다.

上記의 內容을 要略하면, 近代情報 시스템은 다음의 必要條件을 滿足해야 한다.

1) 利用者의 要求에 呼應하기 위해서 適合한 一連의 情報를 提供해야 한다.

2) 利用者와의 關係에서 시스템은 積極的이어야 한다. 즉, 프로파일 檢索公式에서 誘發되는 잘못을 暗示하고 指適할 수 있어야 한다. 나아가서 컴퓨터의 助力으로 프로파일을 作成하고 쉽게 修正할 수 있는 것이 바람직하다.

우리들은 上記한 必須條件에 一致하도록 努力하고 Bratislava에 位置하고 있는 슬로박 科學院의 無機化學研究所와 化學工業經濟研究所가 共同으로 開發한 시스템에 上記의 特殊性을 最少한 몇가지 導入하였다. 시스템은 CACS라는 作業名下에서 運營되었다. 대략 15개월동안 280여개의 프로파일로 CA-Condensates 데이터베이스에서 情報를 檢索하여 利用者들에게 最新情報 周知事業을 해 보았다.

우리들은 시스템의 作業能率을 評價하고 利用者의 要求를 滿足시키는 最善의 方法을 模索하였다. 우선 利用者と 시스템사이를 密接하게 連結하는데 重點을 두었다. 作業能率을 分析해보고 우리들은 効率에 影響을 주는 因子는 다음의 4가지 主要分類에 속한다는 結論을 얻었다.

- (1) 檢索하는 데이터 베이스
- (2) 情報시스템, 소프트웨어 패키지 그리고 하

*Performance of an SDI System, with Interactive Features. Journal of Chemical Documentation. Vol. 14, No. 2, 1974. pp. 85~92

** KORSTIC 電子計算室

드웨어 構造

- (3) 利用者의 特性과 그의 要求
- (4) 情報센터가 利用者에게 提供하는 補助

1. 데이터 베이스

SDF (Standard Distribution Format)로 된 CA-Condensates를 데이터 베이스로 利用하였다. 既成品인 外部의 레퍼런스 서어비스는 分明히 利點이 있다. 高度로 資格이 갖추어진 많은 抄錄者와 索引者, 그리고 編集人들의 作業結果를 쉽게 利用할 수 있어서 우리들은 一次文献을 分析할 필요가 없다. 그러한 外部 레퍼런스 서어비스의 短點은 그것이 설혹 우리에게 適合하지 않다 하더라도 우리는 外部 서어비스 기관의 索引方式을 받아 들여야 한다. 外部에서 情報서어비스用으로 마련한 데이터 베이스가 利用者의 要求를 보다 가깝게 滿足시키리라는 것은 의심할 여지가 없다. 内部의 情報서어비스는 現在의 化學情報量이나 文献들을 分析하고 索引하기 위하여 科學分野에 從事하는 職員들의 인건비로는 非效率的이고 不可能하다.

CA-Condensates는 索引하는데 語彙를 統制하지 않고 使用한다. 結果적으로 한 主題를 여러가지 同義語나 그에 가까운 同義語로 表示할 수 있다. Chemical Titles에 관해 1967년에 出版된 "Search Guide"라는 책에서는, 主要한 用語(基本同意語)에다 별표로 표시하였다. 우리는 索引者들이 이러한 主要用語들을 보다 잘 使用하리라고 期待한다. 이것은 그경우와 다르지만 索引者들은 device-apparatus나 de-watering-dehydration과 같은 同意語들도 함께 使用한다. 물론 이것은 適合한 情報의 損失을 가져올지 모른다. 또한 키판치를 할 때 실수하여 誤記로 인해 適合한 情報의 一部를 檢索하지 못하는 경우도 있을 수 있다. 예를 들면 boundary대신에 "boundry"라고 索引했거나 ethylene대신에 "ethylne"이라고 索引한 抄錄들은 옳게 쓴 말로는 檢索되지 않을 것이다.

이 시스템에서는 각각의 CA-Condensates를 再構成하여 inverted file을 만든다. 이러한 再排列은 CAC의 抄錄을 索引하는데 必要한 重要語辭典을 알파벳順序로 프린트할 수 있다.

록 한다. 우리들은 이 語彙들이 프로파일을 作成할때 대단히 큰 도움이 되리라는 것을 곧 알았다. 이 索引用語辭典을 이곳 저곳, 살펴보고 인쇄가 잘못되고 틀린 곳을, 많이 發見했지만 그보다 더 重要한 것은 抄錄을 索引하는데 必要한 用語들을 광범위하게 볼 수 있었다. 그래서 예를 들어 한 用語의 單수형과 복수형이 얼마나 자주 사용되는가도 調査할 수 있었다. 오른쪽 트런케이손을 시스템에 도입한 이후에는 전보다 자주 辭典을 使用하였다. 몇권의 CAC에서 나온 辭典들을 調査해 보고 얻고 싶은 모든 單語들을 빠뜨리지 않고 얻을 수 있는 트런케이손의 形態를 發見할 수 있었다.

소프트웨어 패키지와 하드웨어 構造

우리들은 DOS(Disk Operating System)로 된 IBM360/30을 使用했다. 經濟性を 考慮하여 既存 소프트웨어를 利用하되 利用者 中心의 對話方式 情報시스템을 發見하려고 애썼다. 그러나 여러가지 理由로 適當한 시스템을 發見하지 못하여 IRMS (Information Retrieval and Management System)의 一部를 취하여 修正하고 附隨적으로 作成된 프로그램을 利用하여 하드웨어 構造에 作用할 수 있는 시스템으로 고쳤다.

IRMS 패키지 안에 들어있는 7개의 프로그램 가운데 DCT1, DCT4, INV1은 그대로 使用하고 檢索프로그램인 SRCH는 상당히 修正하였으며 CONV와 REF는 새로 作成하여 添加한 프로그램이다. 또한 IBM SORT프로그램도 利用한다. CONV프로그램에 의해서 CA-Condensates의 磁氣테이프 原本을 읽어서 서지적 파일을 만들고 descriptor file과 inverted file을 위한 出力이 나오도록 데이터 베이스를 再構成한다. descriptor file에 必要한 附隨的인 再構成이 SORT와 REF프로그램에 의해서 더 이루어진다.

descriptor file은 DCT1과 DCT4 프로그램에 의해서 管理된다. INV1 프로그램은 디스크 위에 인버티드 검색파일을 준비하는데 使用한다. 質問데이터는 SRCH 檢索프로그램에 의해서 處理된다. 서지적 파일에서의 檢索을 包

숨하여 그 외에 모든 檢索이 直接 이루어진다.

情報시스템의 相互作用

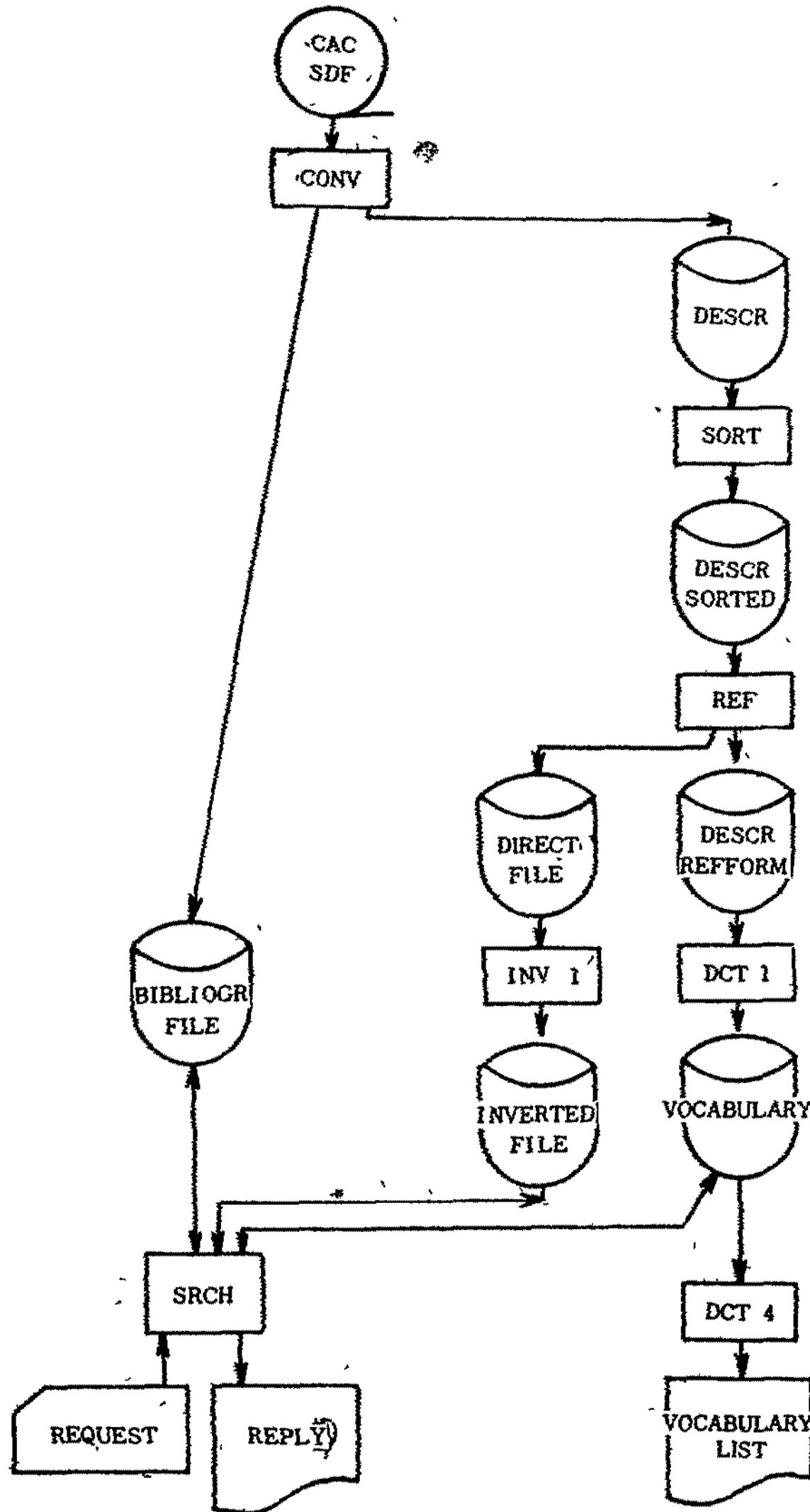


그림 1. 시스템의 全般的인 工程

그림 1에서 본 시스템의 全般的인 工程을 보여준다.

질문과 出力構造는 약간 修正한 것도 있지만 IRMS로 사용한 것과 비슷하다.

機械와 시스템의 構造

시스템은 IBM360을 DOS하에서 運營하도록 되어 있다. 프로그램은 Assembler 言語를 사용했다. 最少單位의 構造는 System360, Model 22인 22Kbytes의 主記憶장치로 되어 있다.

最少의 周邊記憶장치는 3개의 2311 디스크 드라이브와 1개의 테이프 드라이브로 되어 있다.

情報시스템사이에는 相互作用이 많다. 이러한 시스템을 지금까지 무엇이라고 正確하게 定義 하지는 못했지만 King과 Bryant가 指摘한 바와 같이 모두 어떤 共通된 特徵이 있는 것 같다. 특히 質問에 對答하는 速度나 質問에 관한 시스템변수를 주는 能力이 그 特徵이다. 예를 들면 주어진 一連의 索引用語로 索引되는 文獻의 數, 반복하여 檢索하는 方法, 그리고 對話方式의 質問方法 등을 들 수 있다.

특히 外部의 레퍼런스 서비스를 利用할 때에 對話方式의 情報시스템을 갖는것이 가장 重要하다. 利用者는 機關이나 索引하는 方法, 그리고 探索하고자 하는 파일의 內容을 잘 모를 수도 있다. 最適의 質問을 精巧하게 만들기 위해서 利用者는 積極的으로 컴퓨터를 利用해야 한다. 大部分 相互作用하는 시스템에서는 반복하여 檢索한다. 利用者는 以前의 要求에 의해 檢索된 情報를 바탕으로 自己의 質問方式과 探索의 方向을 修正한다. 이와 같이 願하는 情報를 추적해 나가는 方法은 대단히 有用하다.

CACS 시스템內에서의 要求의 處理

이미 言及한 바와 같이 CAC는 索引을 하는데 言語의 制限을 두지 않으므로 質問公式을 作成하는데 여러가지 어려움이 있다. CAC의 體制를 再構成하여 인버티드 파일을 만드는데 알파벳順序로 된 索引用語辭典이 대단히 有用하다는 것은 이미 說明한 바 있다. 用語辭典이나 統制된 語彙를 使用하는 시스템에서 利用者는 컴퓨터에 의해서 自己의 問題에 가장 適合한 用語들을 發見할 수 있다. 그림 2에서 알파벳順序로 된 索引用語辭典의 一部를 본다.

質問에 使用한 重要語는 알파벳順序로 된 辭典에 각각 對照해 본다. 만일 目錄에 그러한 말이 나타나지 않을 경우 컴퓨터는

“...descriptor, wrong or unknown” 이라고 프린트한다. 質問에 使用한 用語에서 綴字가 틀린 것을 자주 發見하는데 이것은 注意를 기

올이지 않은 데서 오는 잘못이다. 또 用語가 틀리지 않았는데도 계속해서 "wrong or unknown"이라고 表示되어 나올 때에는 利用者의 要求가 適切하게 表現되었는지 與否를 檢索할 必要가 있다.

一連의 規則에 따라 作成된 質問은 一連의 檢索프로그램에 의해 處理된다.

質問프로파일은 識別線을 除外하고 最大 15개의 變數로 되어 있는데 그중 10個까지의 記述因子는 論理接續因子인 AND와 OR중의 어느 한가지로 結合되었다. 質問프로파일의 마지막 線은 "Query line" (Qu-line)인데 그 線안에 있는 變數들은 論理接續因子인 AND, OR 그리고 NOT으로 結合되어 있다. 또한 괄호도 使用할 수 있다. 2가지의 質問프로파일을 그림 3에서 본다.

| ALPHABETIC DICTIONARY | PAGE | 3 |
|-----------------------|------------------------|---|
| 301 ANALGESIC | 351 ANTHRACENE | |
| 302 ANALGIN | 352 ANTHRANILIC | |
| 303 ANALOG | 353 ANTHRAPPYRROLE | |
| 304 ANALYSIS | 354 ANTHRAQUINONE | |
| 305 ANALYTICAL | 355 ANTHROSDL | |
| 306 ANALYZABILITY | 356 ANTSAGING | |
| 307 ANALYZER | 357 ANTIANEMIC | |
| 308 ANALYZING | 358 ANTIARRHYTHMIC | |
| 309 ANATASE | 359 ANTIBACTERIAL | |
| 310 ANATION | 360 ANTIBARYON | |
| 311 ANAYZER | 361 ANTIRIOTIC | |
| 312 ANCHOR | 362 ANTICARIES | |
| 313 ANDERSON | 363 ANTICARIOGENIC | |
| 314 ANDESITE | 364 ANTICATHODE | |
| 315 AMORPHOUS | 365 ANTICORROSION | |
| 316 ANEMIA | 366 ANTICORROSIVE | |
| 317 ANEROBIC | 367 ANTICROSSLINKING | |
| 318 ANESTHETIC | 368 ANTICRUSTING | |
| 319 ANGLE | 369 ANTIDEPRESSANT | |
| 320 ANGULAR | 370 ANTIFERROELEC | |
| 321 ANHARMONIC | 371 ANTIFERROMAGNET | |
| 322 ANHARMONICITY | 372 ANTIFERROMAGNETIC | |
| 323 ANHYDRIDE | 373 ANTIFERROMAGNETISM | |
| 324 ANHYDRITE | 374 ANTIPOGGANT | |
| 325 ANILINE | 375 ANTIPOULANT | |
| 326 ANILIND | 376 ANTIPTICITION | |
| 327 ANIMAL | 377 ANTIFUNGAL | |
| 328 ANION | 378 ANTIGEN | |
| 329 ANIOHIC | 379 ANTIGLARE | |
| 330 ANISIDINE | 380 ANTIHISTAMINE | |
| 331 ANISOLE | 381 ANTIHYPERTENSIVE | |
| 332 ANISOTROPIC | 382 ANTIINFECTIVE | |
| 333 ANISOTROPY | 383 ANTIINFLAMMATORY | |
| 334 ANNEAL ED | 384 ANTIKADN | |
| 335 ANNEALING | 385 ANTIMICROBIAL | |

그림 2. 알파벳 順序로 된 索引用語辭典

| | | |
|-------------------------|--|---------|
| 00=SEARCH NUMBER 07 SAV | DEPT A/3 | 0500007 |
| 01.PR=THEMPOD* | THERMOPNS, THERMODYNAMIC, THERMODYNAMICS. | 0500007 |
| 02.PR=HEAT | ENTHALPY, ENTROPY. | 0500007 |
| 03.PR=EQUIL. | ACTIVITY, COMPRESSIBILITY, KP, FUGACITY, THERMAL. | 0500007 |
| 04.AD=MOLAR. | VOL. | 0500007 |
| 05.AD=POTENTIAL. | CHEM. | 0500007 |
| 06.PR=DEHYDRATION. | EVAPN. FORMATION, FUSION, HYDRATION, HYDROXYLATED. | 0500007 |
| 07.PR=MELTING. | SINTERING, SUBLIMATION, VAPORIZATION. | 0500007 |
| 08.PR=DECOMPN. | DISSCN, MIXING, SURFACE, INTERFACIAL, GRAIN. | 0500007 |
| 09.PR=GRAIN. | GROWTH. | 0500007 |
| 10.PR=CARBONATE* | HYDRATE*, HYDROXIDE*, KAOLINITE*. | 0500007 |
| 11.PR=NONITRILLONITE. | | 0500007 |
| 12.PR=OXIDE* | SPINEL*, ALLUV*, MELT, MELTS. | 0500007 |
| 13.PR=IONIZOR* | OROS*ADIC*OROTOR*BIAD(O*ORIO) | 0500007 |
| 00=SEARCH NUMBER 08 SAV | DEPT A/3 | 0500008 |
| 01.PR=DENSITY. | | 0500008 |
| 02.PR=MEASUREMENT. | DETH, APP, CALCN. | 0500008 |
| 03.PR=COMPUTER* | INSTRUMENT. | 0500008 |
| 04.ZO=GIAD02 | | 0500008 |

그림 3. 2개의 質問프로파일

무엇보다 利用者의 要求는 그가 얻고자 하는 情報의 量에 의해 決定되므로 質問프로파일의 出力을 定義하는 變數, 즉 量의 決定이 可能하다. 利用者が 質問프로파일을 새로 作成하여 그가 願하는 文獻이 한권의 CA에서 얼마나 많이 나오는지 모를 때에는 出力을 定義하는 變數를 너무 細分化하지 말고 質問프로파일을 處理하는 것이 바람직하다. 質問을 一般的인 方法으로 處理하면 그에 대한 反應으로 利用者は 프로파일 에 의해 檢索된 모든 抄錄들을 얻게 된다. 그러나 抄錄 그 自体는 프린트되지 않는다. 이 단계에서 情報시스템의 相互作用原理에 따라 利用者は 먼저번 단계에서 얻은 情報에 대한 意思를 한번 더 決定한다. 예를 들면 "칼슘, 마그네슘, 산화규소系의 相平衡"과 같이 아주 細分化된 質問에 대한 反應으로서 파일 內에 85개의 項目이 있을 때 그중에는 얼마간의 잘못된 것이 있다는 것을 알 수 있다. 아마도 잘못된 論理接續因子의 大部分은 위의 경우에 "AND" 대신에 "OR"를 使用했을 것이다. 그렇지 않으면 質問者가 自己의 要求를 正確하게 表現하지 못했거나 프로파일을 修正해야 되는 경우이다. 그다음 단계에서는 出力變數를 明示한다. 즉 質問者가 文獻을 25件까지만 찾고 싶으면 檢索되는 文獻數를 25라고 制限한다. 만일 25件 以上이 檢索되면 컴퓨터는 檢索된 總文獻數만 프린트하고 그 다음 利用者의 意思決定을 기다린다. 利用할 수 있는 모든 可能性을 正確히 알고나면 要求의 定量化, 즉 許容되는 情報量의 決定은 어려운 質問프로파일을 단계적으로 改善하는데 상당히 有用한 道具가 된다. 그의 機能을 더 잘 理解하기 위해서 質問線은 위에서 아랫 level로, 주어진 한 level에서 왼쪽에서 오른쪽으로 處理된다는 것을 記憶해야만 한다. 이 때문에 利用者は 다른 檢索過程으로 自己의 質問에 대한 對答을 얻을 수 있다. 檢索된 文獻의 數가 出力數를 明示하는 變數보다 낮거나 같은 level에서 컴퓨터는 對答을 프린트한다. 實際적인 例를 생각해 보자. 만일 質問者가 "the wear resistance of refractory materials"에 관한 情報를 찾는다면 그의 프로파일은 다음과 같을 것이다.

- 01, OR=refractory, refractories
- 02, AD=wear, resistance
- QU=01 AD 02

첫번째 試圖에서 檢索된 文献의 總數는 다음과 같이 프린트된다.

“Last answer N=3”

다음 단계는 出力을 定義하는 變數를 明示하는 일이다. 만일 質問者가 3件의 文献으로 滿足한다면 出力變數를 다음과 같이 表示한다.

QU, 5=01 AD 02

만일 3件의 文献이 滿足하지 않다면 다음과 같이 더 높은 出力變數를 使用할 수 있다.

QU, 60=01 AD 02

後者の 경우, 첫번째 단계를 끝내고 난 후, 즉 컴퓨터가 “refractory” 나 “refractories”用語로 索引된 모든 文献들을 찾아내고 이 文献들의 數가 60보다 적거나 같을 때 첫번째 索檢 level에서 얻은 對答을 이른바 “temporary answer”라고 프린트한다. 檢索의 level은 항상 temporary answer로 表示된다. 質問을 一般的方法으로 끝까지 處理해서 質問線에 表示된 모든 條件을 滿足시키는 文献들은 “last answer”라고 表示된다.

順序가 組織적으로 된 데이터 베이스나 用語集이 되어 있는 데이터 베이스를 探索할 때에 質問者는 探索하고자 하는 level을 미리 決定한다. CAC와 같이 語彙가 統制되지 않고 組織적으로 順序있게 索引되지 않은 데이터 베이스를 探索할 때에는 檢索을 위해 索引者가 明示해 놓은 것을 알아야 한다. 우리가 文献을 찾을 때에는 最少한 檢索되는 文献과 檢索되지 않는 文献의 두 分類로 항상 나누어진다. 그리고 만일 “properties (epr, nmr, etc) of coordination compounds of copper”에 관한 文献을 찾으려고 한다면 그 프로파일은 다음과 같다.

- 01, AD=COORDINATION, COMPOUNDS.
- 02, OR=COPPER
- 03, OR=EPR, NMR
- QU=01 AD 02 AD 03

위의 프로파일은 文献의 集合을 4가지 分類로 나눈다. 즉 “coordination compound”에 관한

廣範圍한 文献의 集合과 “coordination compound of copper”에 관한 보다 좁은 範圍의 文献의 集合, 그리고 “the epr and the nmr of the coordination compounds of copper”에 관한 文献의 集合으로 나눈다. 이 分類는 데이터 베이스안에 있는 文献의 索引用語와 質問프로파일안에 있는 索引用語가 接續因子의 論理的인 組合으로 서로 重複되는 程度에 따라 適合한 文献을 檢索하는데 基礎를 두고 있다. 위 프로파일에서 3개의 集合은 3단계의 level을 나타낸다. 一般的으로 “coordination compounds”에 관한 첫번째 集合은 가장 廣範圍하다. 이것은 지시한 論理的 作業을 수행해 가면서 점점 좁혀져 간다.

“Temporary answer”의 도움으로 質問者는 細分化 정도를 變化시켜 가면서 자기의 質問에 대한 對答을 얻을 수 있다.

試驗過程에서 여러 角度로 檢索해 보면 가장 有用한 對答을 發見할 수 있다. CAS가 利用者들에게, 아무리 훌륭한 情報資料와 데이터 베이스를 提供한다 할지라도 質問者는 要求를 表現하는 自己의 方法이 시스템이나 索引者의 言語와 習性에 一致하고 있는지 與否를 確信하지 못하고 있다.

트런케이손 (TRUNCATION)

大部分의 檢索시스템은 檢索效率을 높이는 方法으로 單語의 絶斷(트런케이손)을 使用한다. CACS시스템에서는 오른쪽에 트런케이손을 使用한다. 우리는 CA-Condensates의 原本 데이터프를 고쳐서 인버티드 파일을 만들었으므로 우리 시스템에서는 트런케이손을 왼쪽에 使用하지 않는다.

트런케이손을 使用할 때에는 注意를 해야 한다. 우리는 願하지 않는 用語가 檢索되는 것을 防止하기 위하여 몇권의 CAC에서 뽑은 重要語들을 알파벳順序로 편집한 辭典에 대하여 絶斷한 用語들을 調査해 보았다. 그리고 우리들은 絶斷한 用語와 完全한 用語로 몇가지 探索을 調節해 보았다. 그 結果 絶斷한 用語를 使用한 프로파일과 完全한 用語를 使用한 프로파일의

結果가 거의 같았다. 그러므로 檢索者는 可能한 모든 變化를 考慮하는 것보다 차라리 絶斷한 用語를 使用하는 것이 훨씬 편리하다.

利用者의 個性과 그의 要求

科學技術 問題에 關하여 利用者와 對話를 해서 무엇이 重要한 問題라는 것을 잘 알아도 利用者의 研究分野에 광범위하게 참여한다는 것은 不可能하다.

우리는 利用者의 數와 그들 要求의 特性에 대해서 서술하려고 한다. 利用者의 소속과 근무년수, 화학분야, 그리고 직위별로 利用者의 數를 調査해 보았다. 利用者들이 願하는 情報는 研究의 種類(基礎 또는 應用)와 目的(物質, 方法, 理論) 등에 따라서 여러가지 特性이 있다. 그래서 檢索質問書인 利用者의 프로파일을 이러한 特性에 關하여 檢討한다.

利用者들은 學術研究機關인 슬로박 과학원과 이 機關에서 하고 있는 研究에 공동으로 참여하고 있는 研究機關들에서 근무하는 사람들이다. 全體的으로 88개의 연구팀들이 그들의 必要한 情報를 280개의 檢索프로파일로 作成하였다. 한 연구팀내에 평균 4~5名의 研究員들이 있어서 약 450名의 研究員들이 情報를 利用해 보고 그를 評價하였다.

機關별로 소속되어 있는 研究팀들의 숫자를 表 1에서 본다. 그리고 化學의 전공분야에 따른 研究팀數의 분포는 表 2에서 본다.

表 1.

| | 연구팀 數 | % |
|----------|-------|-----|
| 學術 研究 機關 | 34 | 39 |
| 産業 研究 機關 | 45 | 50 |
| 技術 專門 學校 | 9 | 11 |
| 合 計 | 88 | 100 |

研究에서 基礎와 應用部門사이의 뚜렷한 境界線은 없지만 우리는 利用者들을 研究의 一般的인 特性에 따라서 다음과 같이 두 그룹으로 分離하였다. 즉 基礎部門研究가 48%이고 應用部門研究가 52%이다. 그와 비슷하게 化學專攻分野에 따라 研究員들을 表 2와 같이 대충 分類했는데

총 280개의 檢索프로파일 가운데 120개의 정도 프로파일은 物理化學이나 理論化學으로 分類할 수 있다.

表 2. 化學의 專攻分野에 따른 研究팀數의 分布

| | % |
|-----------|-----|
| 無 機 化 學 | 25 |
| 有 機 化 學 | 23 |
| 生 化 學 | 11 |
| 高 分 子 化 學 | 12 |
| 石 油 化 學 | 14 |
| 化 學 工 學 | 15 |
| 合 計 | 100 |

利用者들에 關해서 자세히 說明하고 正確하게 分類하려면 利用者들의 상당한 協助가 必要한데 그들 모두에게 이것을 期待할 수는 없으므로 그중에서 기꺼이 協助하고 質問에 對答해 주는 10개의 研究팀을 選定하였다.

우리들은 한 그룹안에 從事하고 있는 研究員數와 그들의 教育程度, 勤務年數, 從事하고 있는 化學分野, 研究形態, 그리고 研究팀의 水準 등을 알아내기 위하여 10개 研究팀의 代表者들과 面談하였다. 研究팀의 水準은 專門機關의 會員加入과 學術會議에 積極的인 參與度, 論文發表 그리고 專門學校에서의 講義나 專門雜誌의 編輯能力 등과 같은 그밖의 活動에 비추어 評價한다. 이와 같은 利用者의 特性을 表 3과 4에서 보여준다.

表 3에서 보는 바와 같이 한 研究팀은 平均 4名의 研究員으로 構成되어 있다. 그룹은 보통 經驗이 많은 研究員(20~25年間の 勤務經歷)의 指導下에 運營한다. 45名의 研究員들중에서 23名의 研究員들이 Ph. D나 教授와 대등한 學位를 가지고 있으며 22名의 研究員들이 大學卒業生들이다. 研究의 性格에 關해서는 3個의 팀이 基礎研究이고 1個의 팀이 應用, 그리고 6個의 팀이 基礎와 應用을 접한 研究를 하고 있다.

表 3. 利用者들의 特性

| 研究 그룹 | 그룹내의 연구員數 | 教育程度 | 勤務年數 | 研究팀의 水準 M = 中級 H = 上級 VH = 最上級 |
|-------|-----------|---------------------|--------------------------------|---|
| 1 | 6名 | Ph. D. 3名 大學卒 3名 | 25. 20. 12 8. 5. 5 | H |
| 2 | 6名 | Ph. D. 5名 大學卒 1名 | 25. 20. 12 16. 15. 5 | H |
| 3 | 3名 | Ph. D. 1名 大學卒 2名 | 25. 20. 5 | M |
| 4 | 4名 | Ph. D. 3名 大學卒 1名 | 17. 15. 14 10 | VH |
| 5 | 5名 | Ph. D. 3名 大學卒 2名 | 23. 10. 7 5. 3 | M |
| 6 | 7名 | Ph. D. 5名 大學卒 2名 | 20. 18. 15 20. 10. 9. 8 | VH |
| 7 | 3名 | Ph. D. 1名 大學卒 2名 | 25. 10 5 5 | VH |
| 8 | 3名 | Ph. D. 2名 大學卒 1名 | 20. 18. 5 | VH |
| 9 | 7名 | Ph. D. 3名 大學卒 4名 | 20. 18. 15 15. 12. 12 10 | H |
| 10 | 1名 | Ph. D. 1名 大學卒 0名 | 9 | VH |

表 4 利用者들의 特性

| 研究그호 | 機 關 | 化 學 分 野 | 研究形態 |
|------|--------|-----------|-------|
| 1 | 學術研究機關 | 珪酸塩化學 | 基礎+應用 |
| 2 | " | 物理化學+珪酸塩 | 基礎 |
| 3 | " | 合成珪酸 | 基礎+應用 |
| 4 | " | 天然珪酸 | " |
| 5 | " | 分光學, 分光化學 | " |
| 6 | " | 電氣化學 | " |
| 7 | " | 無機化學 | " |
| 8 | " | 結晶學 | 基礎 |
| 9 | 産業研究機關 | 펄프와 製紙化學 | 應用 |
| 10 | 技術專門學校 | 等位化學 | 基礎 |

이 作業을 시작한 이래 우리들은 利用者들의 情報要求에 關係 可能한 限 많이 배우려고 努力했다. 그 理由는 分明하다. 왜냐하면 情報要求의 內容을 알지 못하면 그것을 檢索프로파일에 表現하기 힘들기 때문이다. 어떤 種類의 質問書도 大部分 人氣가 없고 對答하기를 대단히 싫어하므로 우리는 이 目的을 위해 조심성있게 對談하는 方法을 취했다. 특히 아래의 事項들을 確實히 알고 싶었다.

1. 必要로 하는 情報의 種類; 예를 들면 다만 現在의 專攻分野에 關係된 最新情報만을 願하는지, 그렇지 않으면 앞으로 自己 分野에 새로운 着想을 얻기 위한 것인지 또는 어떤 새로운 分野의 研究를 위한 情報인지를 알고 싶었다.

2. 必要로 하는 情報의 量; 예를 들면 다만 自己의 研究에 密接하게 關係된 情報만을 願하는지, 그렇지 않으면 약간의 背景情報나 보다 넓은 範圍의 情報를 願하는지를 알고 싶었다.

3. 必要로 하는 情報의 目的; 예를 들면 物質이나 方法이나 또는 理論적인 問題인지를 알고 싶었다.

첫번째 質問에 대해서 利用者의 95%가 最新情報과 새로운 着想, 그리고 새로운 分野의 研究를 위해 자극이 되는 情報를 함께 期待한다고 對答했다. 우리가 實驗하고 있는 10개 팀 가운데 작은 그룹에 있는 利用者들은 모두 시스템이 이 機能을 다 할 수 있기를 願한다.

두번째 質問에 대해서 基礎部門의 研究員들은 完全한 一連의 情報를 얻고 싶어하며 시스템이 適合한 文獻을 데이터 베이스안에 남기지 않고 모두 檢索만 한다면 出力가운데 適合하지 않은 文獻이 60%까지 차지하고 있는 것은 기꺼이 許容한다.

그와 반대로 應用部門의 研究員들은 그들의 要求를 正確하게 滿足시키는 最新情報를 얻기 願한다. 出力가운데 適合하지 않은 文獻이 나타나면 適合한 文獻을 檢索하지 못한 이상으로 不快하게 생각한다.

利用者は 物質이나 方法, 또는 理論적인 問題에 關係한 情報를 찾을지도 모른다. 우리는 實驗對象으로 하고 있는 그룹의 22개 프로파일을

위의 세가지 分類로 나누고 프로파일當 重要語의 數와 프로파일當 平均 檢索文獻數, 그리고 프로파일의 細分化 정도를 表 5에 표시하였다. "OR" 接續因子를 한번만 使用한 프로파일은 "S" (單純한 프로파일)이라고 表示하고 "AND"나 "NOT" 接續因子를 한번만 使用한 프로파일은 "R", (制限된 프로파일)이라고 表示하였다. 만일 "AND"나 "NOT" 接續因子를 2번, 또는 그 以上 使用했으면 "HR" (극히 制限된 프로파일)이라고 表示했다.

우리의 經驗에 의하면 物質에 관한 프로파일은 理論的인 問題나 現象에 관한 프로파일보다

훨씬 作成하기 쉽다. 예를 들어, 熱力學이나 動力學에 관한 問題에 관심이 있는 科學者는 基礎理論에 관한 情報를 원하는 것이지 一般的인 計算의 應用에 관심이 있는 것이 아니다. 問題는 "그러한 프로파일을 어떻게 作成하느냐"이다. 아마도 利用者는 "thermod*"나 "Kinetic*"으로 索引되는 用語들, 또는 이들 分野에 보다 細分化된 索引用語들을 모두 檢討해야 할 것이다. 利用者는 실제로 願하는 情報를 몇개 選擇하고 그 나머지는 버려야 될 것이다. 우리는 經驗을 통해서 "基礎的, 基本的, 理論的, 새로운 開發, 새로운 傾向"과 같은 用語를 添

表 5. 檢索프로파일로 表現한 利用者의 要求

| 연구그룹 | 검색 프로파일의 수 | 검색프로파일의 형태 | 프로파일당 중요어의 수 | 프로파일당 검색된 문헌 수 | 프로파일의 세분화 정도 |
|------|------------|---------------------------|---------------------------|--------------------------|---------------------------|
| 1 | 3 | { 理論 1件 方法 1件 物質 1件 | 10 4 2 | 2 1 33 | { HR HR S |
| 2 | 5 | { 理論 2件 方法 2件 物質 1件 | { 38 34 7 9 5 | { 11 2 3 8 1 | { HR HR R R R |
| 3 | 2 | 物質 2件 | { 22 9 | { 2 6 | { HR R |
| 4 | 1 | 物質 1件 | 41 | 50 | R |
| 5 | 2 | { 方法 1件 理論 1件 | 8 3 | 29 3 | { R HR |
| 6 | 2 | 物質 2件 | { 5 18 | { 47 45 | { R HR |
| 7 | 2 | { 理論 1件 物質 1件 | 9 35 | 6 77 | R R |
| 8 | 2 | 理論 2件 | { 11 12 | { 1 2 | { R R |
| 9 | 2 | 方法 1件 物質 1件 | 18 6 | 11 8 | { HR R |
| 10 | 1 | 理論 1件 | 55 | 6 | HR |

加할 必要는 없다는 것을 안다.

方法에 관한 情報를 컴퓨터로 檢索하는 것은 책자로 된 Chemical Abstracts를 手作業으로 檢索하는 것보다 훨씬 쉽다.

方法들 그 自体는 大部分의 경우, 그들 명칭으로 確實히 定義되므로 SDI프로파일을 適當한 말로 表現하는 것은 比較的 쉽다. 이 方法들에 관한 抄錄들은 CA의 여러 分類에 흩어져 있어 手作業에 의한 檢索은 대단히 불편하지만 컴퓨터를 利用하면 단 몇초밖에 걸리지 않는다.

表 6.

| | 프로파일당 평균 중요어의 수 | 프로파일당 평균 檢索 文獻의 수 |
|------------------------|--------------------|----------------------|
| 8개의 理論에 관한 프로파일의 그룹 | 23 | 4 |
| 5개의 方法에 관한 프로파일의 그룹 | 9 | 9 |
| 9개의 物質에 관한 프로파일의 그룹 | 16 | 29 |

表 5와 6에서 보는 바와 같이, 理論에 관한 프로파일의 그룹에선 프로파일당 평균 23개의 重要語로 4件的 文獻을 檢索했다. 方法에 관한 프로파일의 그룹에선 9개의 重要語로 9件的 文獻을 檢索하고 物質에 관한 프로파일의 그룹에선 16개의 重要語로 29件的 文獻을 檢索하였다.

情報센터와 그의 利用者에 대한 協助

利用者は 利用하려고 하는 情報시스템을 자세히 알기 힘들 것이다. 그러므로 檢索하기 위해서는, 시스템과 利用者의 要求를 모두 잘 알고 있는 情報專門家의 도움이 꼭 必要하다. 우리들은 시스템에 대한 專門的인 知識은 알고 있지만 利用者 각각에 대한 그들의 要求를 철저히 알아낸다는 것이 問題이다. 質問書를 作成하는 式의 方法은 그리 滿足스럽지 못하고 대단히 興味없는 일이다. 우리의 經驗으로는, 利用者와 情報專門家 사이에 끊임없는 對話가 必要한 것 같다. 利用者가 무엇을 研究하고 있는지를 情報專門家가 보다 正確하고 完全하게 알고 있으면 그는 利用者가 自己의 要求를 作成하는데 보다 効

果的으로 도와줄 수 있다.

利用者의 要求와 앞에서 言及한 規則(파라미터의 수, 使用할 수 있는 重要語數의 制限, 適當한 形態의 質問線 등)에 따라서 이 要求를 컴퓨터가 處理할 수 있는 形態로 만든 프로파일과의 사이에는 分명한 差異點이 있다. 우리는 시스템의 作業과 컴퓨터가 處理할 수 있는 프로파일의 作成에 관해서 利用者들에게 教育시켰다. 만일 그들이 시스템 規則에 따라서 그들의 要求를 作成할 수 없다면 우리는 억지로 그들이 프로파일을 作成하도록 要求하지 않는다. 그대신에 그들이 새로운 情報를 얻고자 하는 主題에 관해서 可能한 限 正確하고 完全하게 描寫해 주기를 要求한다. 즉, 研究의 對象인 物質이나 現象, 研究에 利用한 方法(研究에 重要한 것이면 實驗道具까지), 그리고 必要한 情報量에 관해서 우리는 利用者마다 正確하게 定義해야 한다고 強調한다. 특히 重要한 것은 "Temporary answer"을 얻느냐, "Last Answer"을 얻느냐를 左右하는 出力파라미터를 明示하는 일이다. 그림 4와 5에서 "Last Answer"과 "Temporary Answer"로 나온 出力들을 각각 보여준다.

```

CACS SYSTEM
00=SEARCH NUMBER 181 SAV DEPT W 050181
01,OR=CRYSTALLOG*,X, STRUCTURE. 050181
02,AD=STINGLE,CRYSTALS. 050181
03,OR=POWDER. 050181
04,OR=DIFFRACTION,DIAGRAM. 050181
05,OR=SILICATE*. 050181
99
0U,50=(010R020R(03A004)1)A005
    
```

```

LAST ANSWER No. 3
4308 FELSCHE, J.
RARE EARTH SILICATES WITH THE APATITE STRUCTURE
CAS PUB.CIT.:CA07718119197Z PUB.CL.CO.:1J PUB.DATE:000072
J. SOLID STATE CHEM.
SER./VOL.NO.:15 PATENT NO.:
ISS./REP./PART.NO.:12
PAGES:266-75
APATITE STRUCTURE RARE
EARTH SILICATE
4492 SMOLIN, YU. I.
SHEPELEV, YU. F.
TITOV, A. P.
REFINEMENT OF THE CRYSTAL STRUCTURE OF THORTVEITITE SC251207
CAS PUB.CIT.:CA07718119382F PUB.CL.CO.:1J PUB.DATE:000072
KRISTALLOGRAFIYA
SER./VOL.NO.:17 PATENT NO.:
ISS./REP./PART.NO.:14
PAGES:857-8
THORTVEITITE STRUCTURE SCANDIUM
SILICATE
4520 MERLINO, STEFANO
NEW TETRAEDRAL SHEETS IN REYERITE
CAS PUB.CIT.:CA07718119410P PUB.CL.CO.:1J PUB.DATE:000072
NATURE (LONDON), PHYS. SCI.
SER./VOL.NO.:1238 PATENT NO.:
ISS./REP./PART.NO.:86
PAGES:124-5
REYERITE STRUCTURE POTASSIUM
CALCIUM SILICATE HYDROXIDE
    
```

END OF REQUEST

그림 4. "Last Answer"로 나온 檢索出力

```

CACS SYSTEM
00=SEARCH NUMBER 180 SAY DEPT W
01,OR=ZEOLITE,FAUJASITE,MORDENITE,LINDE.
02,OR=MOL.
03,OR=STEVE.
04,OR=PREPARATION,SYNTHESIS,STRUCTURE.
05
06,SN=(01OR02A003)A004
FAUJASITE DESCRIPTORS WRONG OR UNKNOWN
MORDENITE DESCRIPTORS WRONG OR UNKNOWN
PREPARATION DESCRIPTORS WRONG OR UNKNOWN
    
```

TEMPORARY ANSWER N= 25 QU=01,02,03

```

47 EGINA, S. P.
ALIEV, A. M.
MFSTEROVA, L. A.
STRUCTURAL STUDY OF POLYMERIZATION PRODUCTS ON DIFFERENT CAT
ALYSTS
CAS PUB.CIT.:CA07718214931Z PUB.CL.CO.:J PUB.DATE:00007Z
HEFTEPERERAB. NEFTEKHIM. (MOSCOM)
ISS./REP./PAR 7,NO. 13
PAGES:55
BUTANE BUTENE POLYMER
CATALYST MOL SIEVE
ALUMINUM CHROMIUM NICKEL
    
```

그림 5. "Temporary Answer"로 나온 檢索出力

많은 경우에 化學情報를 컴퓨터로 處理하는 것은 이제 틀에 박힌 일이 되었지만 대체로 利用者들은 이 方法을 利用하는데 익숙해야 하고 아직도 그 利用法에 대한 教育이 必要하다. 科學者들은 情報시스템이 그들을 도와야 하고 文獻을 檢索하는데 지루하거나 時間을 消費시켜서는 안된다고 생각하므로 利用者의 教育은 조심스럽게 다루어야 하고 적게 해야 한다.

시스템의 效率에 대한 評價

最近에 情報시스템을 評價하는데 많은 관심이 쏠리고 있다. 情報시스템의 作業能率을 數値로 測定할 수 있는 適當한 方法을 發見하려고 많은 試圖를 해 왔다.^{2, 3, 6)} 情報시스템의 價値는 利用者가 그 시스템으로 얻을 수 있는 目的 如何에 따라 評價할 수 있는 것이지만 시스템 자체에 絶對적인 價値가 있는 것은 아니다. 그리고 다른 시스템들과 比較해 볼 필요가 있다.

앞에서 여러 利用者그룹과 함께 시스템의 目的을 論議했다. 또한 시스템의 內容과 우리의 意見으로 그의 效率에 影響을 주는 여러가지 因子들에 관해서도 論議했다. 그런 절차는 비용뿐만 아니라 時間消耗도 되므로 우리는 評價하는 方法을 대단히 조심스럽게 選擇해야 했다.

效率을 評價하는 것으로만 그치지 말고 檢索 方法을 더 잘 理解해서 이 方法을 改善하는데 도움이 되어야 한다.

어떤 프로파일에 대해서 正確도와 再現率을

比較하는 方法이 자주 利用되었다. 우리는 이 方法이 여러가지 理由로 아주 滿足하지 못하다는 것을 알았다. 왜냐하면 이 方法은 利用者의 要求量을 定하는 것이나 檢索된 文獻의 適合與否를 定全히 無視하기 때문이다. 그리고 再現率에 관해서도 制御目的으로 CA — Condensates와 같은 데이터 베이스에서 한 프로파일에 適合한 모든 文獻들을 手作業으로 檢索한다는 것은 거의 不可能하기 때문이다.

컴퓨터화된 情報시스템의 導入으로 보다 나은 情報서비스를 提供하고 努力과 時間, 그리고 돈을 節約할 수 있어야만 그 意義가 있다.⁵⁾

시스템의 作業效率을 測定하는 文獻⁷⁾에 실린 여러가지 方法들 가운데 經濟적인 理由로 Cooper²⁾가 提案한 '豫期하는 檢索길이를 簡單하게 測定하는 方法을 選擇했다. 그 基本原理는 必要한 適合文獻이 發見될 때까지 順序없이 文獻들의 集合을 探索할 때에 얼마간의 쓸데없는 努力이 浪費된다는 것이다. 우리의 경우에, 컴퓨터화된 檢索시스템을 利用하여 浪費되는 이 努力의 量을 줄이는 것이다. 우리는 3개의 프로파일을 취하여 檢索된 適合文獻의 總數와 그 適合文獻들이 나온 CA分類내에 있는 文獻의 總數를 比較하였다. 만일 우리가 이 모든 文獻들을 檢索하려고 했다면 관련된 CA의 모든 分類를 探索했어야 했을 것이다. (表 7 參照)

豫期되는 檢索길이를 評價해 보니 우리의 質問프로파일에 適合한 對答이 CA의 여러 分類에 흩어져 있다는 것을 알고 상당히 놀랐다. 컴퓨터로 찾아낸 適合文獻이 나온 CA의 모든 分類

表 7.

| 質問프로파일 | R | S | S/R |
|--------|-----|--------|-----|
| 原子的 吸收 | 172 | 14,006 | 81 |
| 病原体的構造 | 140 | 8,798 | 62 |
| 耐火物質 | 265 | 15,497 | 58 |

*R = 檢索된 適合文獻의 數

S = 適合文獻들이 나온 CA分類내에 있는 文獻의 總數

S/R = 檢索還元因子

를 科學者가 결코 手作業으로 檢索하지 못한다는 것은 너무나 明白한 일이다. 表 8에서 한

要求에 應答하고 尊重해 주어야 한다. 이러한 目的을 위해서는 對話方式의 컴퓨터化된 情報시스템이 가장 有用한 것 같다.

表 8. CA 分類에 대한 適合文獻의 分布

CA에 한 分類에서 각 프로파일에 대한 檢索文獻數

| 耐火物質 | 原子的吸收 | 病原体的構造 |
|----------|---------|---------|
| 158(1 X) | 68(1 X) | 52(1 X) |
| | | 43(1 X) |
| | 25(1 X) | |
| 23(1 X) | | |
| 18(1 X) | | |
| 13(1 X) | | |
| 9(1 X) | 11(1 X) | 11(1 X) |
| | | 8(2 X) |
| 7(1 X) | 7(1 X) | |
| | 6(1 X) | 6(1 X) |
| | | 5(1 X) |
| 5(1 X) | 5(2 X) | |
| 4(2 X) | 4(3 X) | |
| 3(3 X) | 3(3 X) | 3(1 X) |
| 2(4 X) | 2(4 X) | 2(1 X) |
| 1(7 X) | 1(6 X) | 1(2 X) |

* ()속의 숫자는 각각 明記한 숫자의 適合文獻이 CA의 얼마나 많은, 分類에서 나오는가를 표시한다.

分類當 檢索되는 文獻數를 보여준다. 表에서 보는 바와 같이 原子的 吸收에 관한 適合한 抄錄들 가운데 54%가 CA의 2個의 分類에서 發見된다. 耐火物質과 病原体的 構造에 관한 情報檢索에 있어서는 適合文獻의 68%가 2個 分類에 集中되어 있다. 만일 利用者가 데이터 베이스에서 54%와 68%의 適合文獻을 받는데 滿足한다면 이 適合文獻들이 密集하여 있는 分類들만 手作業으로 檢索하는 것도 靚찮을 것이고 아마 컴퓨터化도 問題視되지 않았을 것이다. 반면에, 完全한 情報檢索이 必要한 경우 — 같은 種類의 研究를 똑같이 반복해서 努力을 浪費할 必要가 없는 경우—에는 컴퓨터化된 檢索시스템의 應用이 必要하다. 앞에서 強調한 바와 같이 利用者의

参 考 文 獻

1. Balke, S., "Benutzerprobleme der Dokumentation und Infrmation" Nachr. Dok., v. 24, n. 2(1973).
2. Cooper, W.S., "Expected Search Length :A Single Measure of Retrieval Effectiveness Based on the Weak Ordering Action of Retrieval Systems," Amer. Doc., v. 19, n. 30(1968)
3. King, D. W., and Bryant, E. C., "The Evaluation of Information Services and products," Information Resources Press, Washington, D. C., 1971.
4. Meyer, R. L., Meskin, A. J., Mracek, J. J., Schwartz, J. H., and Wheelihan, E. C., "A Systematic Approach to Current Awareness and SDI," J. Chem. Doc., v. 11, n. 19(1971).
5. Skolnik, H., "The What and How of Computers for Chemical Information Systems," J. Chem. Doc., v. 11, n. 185(1971).
6. Stumpf, W., "Entwicklung und Erprobung von Methoden zur Answertung in-und ausländischer Datenbänder für Retrievalzwecke im Bereich ber anorganischen Chemie und ihrer Grenzgehiete" Chem-zt., v. 96, n. 301(1972).
7. Swets, J. A., "Information Retrieval Systems" Science, v. 141, n. 245(1973).
8. Thompson, D. A., "Interface Design for an Interactive Information Retrieval System : A Literature Survey and a Research System Description" J. Amer. Soc. Inform. Sci., v. 22, n. 361(1971).
9. Wersig, G., "Zur Systematik der Benutzerforschung," Nachr. Dok., v. 24, n. 10(1973).

<p. 158의 계속>

Evaluation", Scott, Foresman, 1969

- 4) John J. Donovan, "Systems Programming", McGraw Hill, 1972

- 5) Charles T. Meadow, "The analysis of Information Systems", Melville, 1973
- 6) "AUERBACH Timesharing Reports", AUERBACH Info, Inc. 1969