

KORSTIC 데이터뱅크의 概念

柳 慈 卿*

1. 背 景

요즘은 周邊에서 「데이터뱅크」 혹은 「情報銀行」이라는 말이 자주 들리고 있다. 이 「데이터뱅크」라는 말에 대하여 여러 사람이 定義하였으나, 이를 綜合하여 보면 一般的으로 서로 聯關된 데이터파일 혹은 데이터베이스를 모아놓은 것을 指稱하는 말로 使用되고 있다. 그러나 이것은 單純히 데이터를 集積해 놓은 것에 그치지 않고 모아놓은 데이터 중에서 必要情報를 情報需要者에게 迅速·正確하게 檢索하여 提供할 수 있는 機能을 갖추고 있어야만 한다. 따라서 現代 컴퓨터技術이 發達하면서 이 「데이터뱅크」의 概念도 急激히 成長하였다고도 볼 수 있다.

美國에서는 SDC社의 ORBIT, Lockheed Aircraft Corporation의 DIALOG 혹은 國立醫學圖書館의 MEDLARS와 같은 시스템들이 開發되어 한곳에 蓄積된 데이터를 TYMNET, INFONET와 같은 人工衛星을 使用한 通信網을 통하여 美國全域은 물론 유럽에서까지 온라인(On-line)으로 使用하고 있다. 또한 美國의 ARPANET, 日本의 JIPNET와 같은 프로젝트들은 機種이 서로 다른 컴퓨터를 온라인으로 連結시킴으로써 데이터뱅크 相互間은 물론 하나의 端末(Terminal)에서 여러 데이터뱅크의 資料를 同時に 찾아볼 수 있게 되어 컴퓨터에 의한 資料의 共同利用(Sharing)이 可能하게 되었다.

國內에서도 情報의 効率的인 流通 및 利用을 위하여 1973년 KIST의 綜合데이터뱅크 設置計劃 및 KORSTIC의 國家科學技術情報流通시스템(NASSTI) 등이 發表된 바 있으나 具體적으로 實現된 것은 아직 없다.

KORSTIC은 1975년 美國化學會로부터 컴퓨터 可讀形파일인 CAC를 導入하여 國內에서는 처음으로 컴퓨터에 의한 情報檢索서비스를 實施한 데 이어 1977년에는 物理, 電氣·電子, 컴퓨터 및 制御工學分野의 데이터베이스인 INSPEC과 機械工學分野의 ISMEC을 導入하는 등 提供情報의 範圍를 넓혀가고 있다. 한편 S-5300 시스템을 導入運用함으로써 한글 및 漢字의 컴퓨터 處理化에 대한 研究도 꾸준히 遂行하고 있다.

이번에 樹立한 科學技術情報銀行設置5個年計劃은 이와 같은 KORSTIC서비스를 背景으로 한 技術情報銀行을 設置하여 全國 어디서나 누구든지 必要한 情報를 즉각 찾아 利用할 수 있는 시스템을 開發하는 것을 窮極的인 目標로 삼아 推進하여 NASSTI를 具體化함으로써 世界科學情報網(UNISIST)의 노우드(Node)役割을 遂行하기 위한 것이다.

2. 情報銀行의 範圍 및 運營計劃

2.1 데이터뱅크 構成

데이터뱅크의 構成要素는 ①情報要求者, ②데이터베이스 ③컴퓨터시스템 ④檢索프로그램을 들 수 있으며 이것을 그림으로 表示하면 그림 1과 같다.

* KORSTIC 電算室 시스템開發擔當 次長

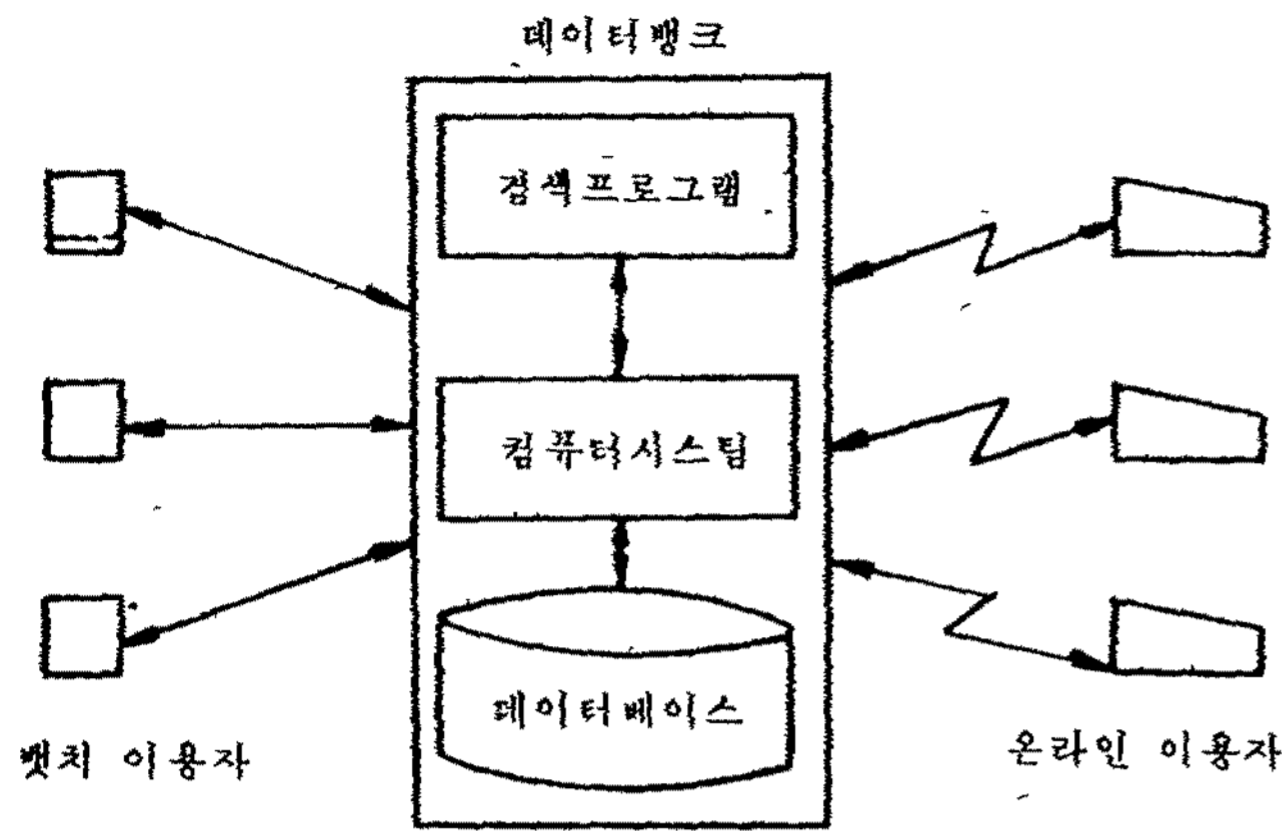


그림 1. 데이터뱅크의 構成圖

2.2 蓄積情報의 範圍

데이터뱅크에 蓄積할 情報은 크게 두가지로 나누어 생각할 수 있는데 하나는 外國의 既成데이터베이스를 購入하여 檢索提供하는 것이고 또 하나는 國內의 文獻 및 製品 등에 관한 情報을 調査, 데이터베이스化하여 蓄積함으로써 필요에 따라 데이터베이스 自体를 他機關에 供給하거나 要求事項만을 檢索提供하는 것이다. 1978년을 初年度로 하여 5個年間 蒐集할 既成 데이터베이스의 種類 및 內容은 表 1과 같으며 目標年度까지 약 600萬件的 情報을 蓄積한다.

表 1.

名 稱	分 野	提 供 處	年間 收錄件數	頻 度
CAC (Chemical Abstracts Condensates)	화학 및 관련분야	미국화학회	400,000	週 刊
INSPEC (Information Service for Physics, Electrotechnology, Computer & Control)	물리, 전기전자, 컴퓨터 및 제어공학	영국 전기 공학회	150,000	月 2回
ISMEC (Information Systems for Mechanical Engineering)	기계공학	영국 기계 공학회	20,000	月 2回
COMPENDEX (Computerized Engineering Index)	엔지니어링 전반	미국 엔지니어링 인덱스社	100,000	月 刊
INPADOC (International Patent Documentation)	세계특허	오스트리아 INPADOC	800,000	週 刊
INIS (International Nuclear Information System)	원자력공학	국제원자력기구 (IAEA)	100,000	月 2回
AGRIS (International System for Agricultural Science and Technology)	농업, 수산, 식품과학 및 관련분야	식량농업기구 (FAO)	200,000	月 刊
GRA (Government Report Announcement)	美國 정부보고집. 과학기술전반 및 사회과학	미국상무성 (NTIS)	50,000	隔月刊
METADEX (Metals Abstracts Index)	금속에 관한 공학, 물리학, 화학	미국금속학회	25,000	月 刊
JICST-IR (JICST Information Retrieval)	이공학분야 전반	일본과학기술정보 센터	400,000	分野에 따라 다름
MEDLARS (Medical Literature Analysis and Retrieval System)	의학 및 약학	미국 국립 의학 도서관	230,000	
국내문헌특허정보파일	과학 기술 전반	KORSTIC	5,000	미 정
국내제품정보파일	국산품 전반	KORSTIC	100,000	미 정
국내진행중연구과제파일	과학 기술 전반	KORSTIC	5,000	미 정

國內에서 製作하고자 하는 데이터베이스로는 다음과 같은 것이 있다.

(1) 國內文獻特許情報파일 : 國內에서 研究되어 國內외의 學會誌, 論文集, 專門技術誌 등에 發表된 科學技術記事 및 特許에 대한 情報파일로 目標年度까지 약 30,000件的 情報를 調査, 蓄積한다.

(2) 國內製品파일 : 國內製品의 製造會社名, 價格, 關聯特許 및 技術 등에 대한 情報를 收錄한 것으로 目標年度까지 약 500,000件的 情報를 蒐集整理한다.

(3) 進行中 研究課題파일 : 國內 大學, 研究所 등에서 遂行中인 研究課題 및 研究者, 研究期間 등에 대한 情報를 目標年度까지 약 30,000件 蒐集한다.

이 외에도 科學技術人力의 效果的인 活用을 위한 國內科學技術人力情報파일 등의 製作을 考慮중에 있으며, 이러한 國內데이터베이스 製作을 위한 國內資料調査 뿐만 아니라 國際協力에 의하여 國際機構에서 만들어지고 있는 INIS, AGRIS, INPADOC을 위하여 國內에서 發生하는 原子力情報, 農業情報, 特許情報를 調査, 發送하여야 할 것이다. IAEA(國際原子力機構), FAO(食糧農業機構), INPADOC(國際特許廳)에 入力시킬 年間 데이터 件數는 各各 500件, 1,500件, 5,000件으로 推定하고 있다.

KORSTIC은 國內데이터베이스의 製作을 위하여 S-5300시스템을 活用하여 한글·漢字의 코우드組合시스템을 開發하여 ORRS(進行中 研究課題 檢索시스템) 등에 活用하고 있으며 이를 發展시킨 文字像 組合프로그램과 한글·漢字 自動變換시스템을 開發중에 있다.

2.3 檢索시스템

檢索시스템은 Lockheed의 DIALOG나 SDC社의 ORBIT와 같은 시스템을 모델로 하여 自體內에서 開發하는 것을 原則으로 삼고 있다. 데이터베이스別로 別個의 檢索프로그램을 使用할 수도 있으나, 그보다는 하나의 프로그램으로 全 데이터베이스를 檢索하는 시스템을 開發한다.

檢索서비스에는 現在를 中心으로 하여 계속 入

手되는 磁氣테이프에서 새로운 情報를 찾아 提供하는 SDI(Selective Dissemination of Information; 現況追跡調査)와 一定期間동안 過去分の 情報에 계속 入手되는 새로운 情報를 累積하였다가 情報要求가 있을 때마다 檢索提供하는 RS(Retrospective Search; 遡及調査)가 있다. 現在 KORSTIC이 CAC, INSPEC 및 ISMEC으로 提供하고 있는 서비스는 SDI서비스이다.

SDI는 한번 지나간 情報는 다시 찾아 볼 수 없다는 短點이 있는 반면, 情報의 網羅性이 있는 RS서비스는 많은 量의 情報를 한꺼번에 컴퓨터의 記憶裝置에 記憶시켜 놓아야 하기 때문에 費用이 많이 든다는 短點이 있다. 예를 들면 CAC 1년分の 데이터를 記憶시키기 위하여는 약 100MB의 記憶容量이 必要하다.

또한 電算處理方式에는 一定期間동안의 作業量을 모았다가 一括處理하는 batch(Batch)서비스와 中央컴퓨터에 端末을 連結하여 利用者가 願하는 때에는 언제 어디에서든 情報를 檢索할 수

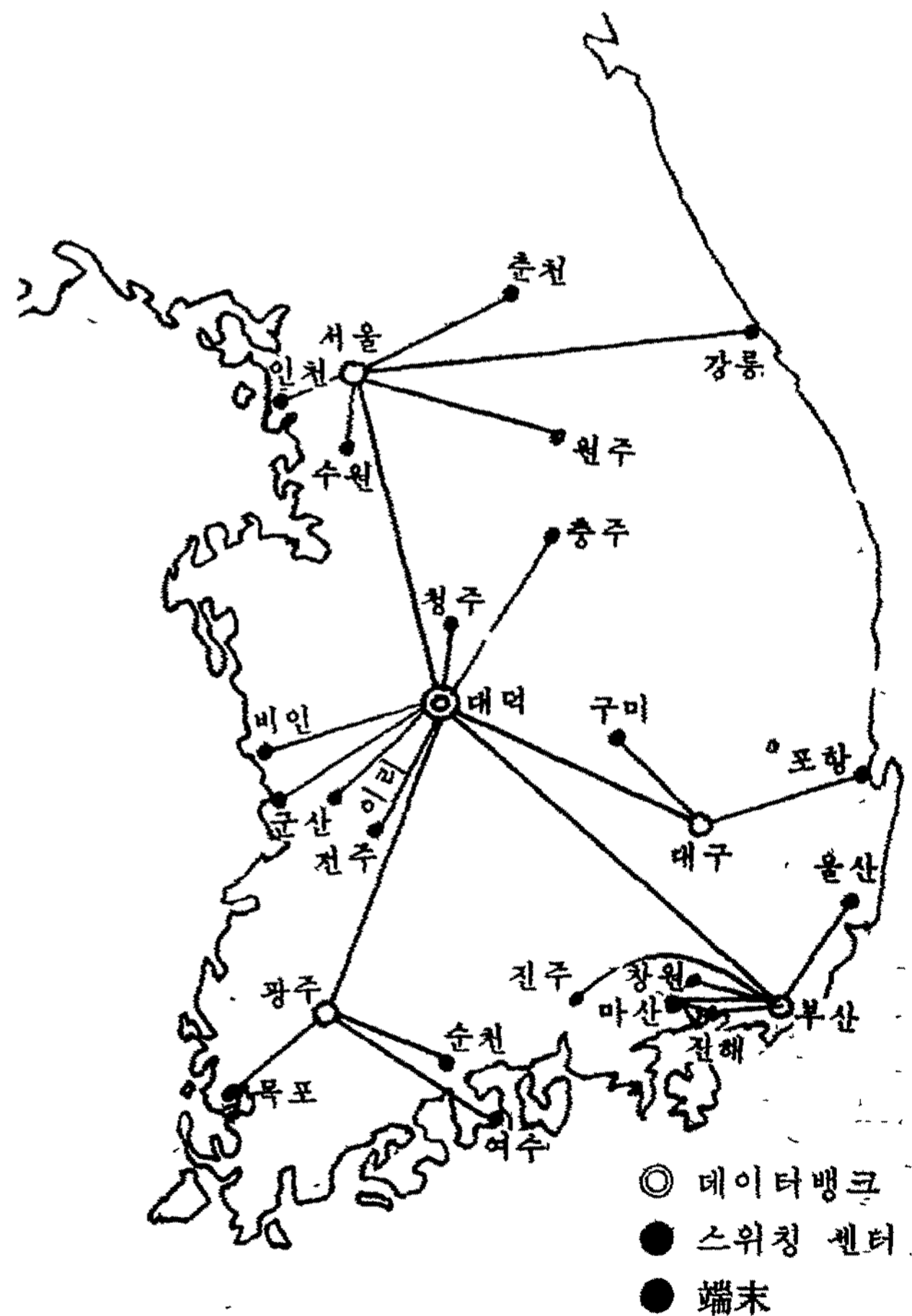


그림 2. 全國 技術情報 流通網

있는 온라인(on-line)서비스가 있다. 배치方式은檢索式의作成 및修正을效果的으로遂行할 수 없고檢索結果提供에도時間,距離上의 문제로遲延되므로窮極的으로는 온라인方式을指向하여야 한다. 따라서 배치方式에 의한 SDI서비스만을 위주로 하고 있는現 시스템을發展시켜 배치RS, 온라인 SDI 등의段階를 거쳐 온라인 RS시스템을開發하여 5個年計劃이 끝나는 1982년에는 KORSTIC負擔으로全國에 25臺의 端末을 그림 2와 같이設置하여全國 어디에서든지中央情報銀行에 있는資料를 찾아볼 수 있도록 할方針이다.

檢索서비스에 사용할 컴퓨터시스템은 IBM370-138이며 主記憶容量 512KB, 初年度の補助記憶容量 700MB로 RS시스템開發에對備하고 있다.

2.4 運營計劃

데이터뱅크의設置計劃은年次別로 나누어推進하며各段階의推進內容과時期는 다음과 같다.

(1) 第1期(1978년~1979년 中半)

컴퓨터 시스템의導入으로外部 컴퓨터에 의해處理하고 있는 SDI서비스를 自体에서處理하도록 시스템을開發하며, 既存의 CAC, INSPEC, ISMEC에 이어 原子力工學分野의 INIS, 農學分野의 AGRIS, 世界特許인 INPADOC, 美國政府報告書인 GRA 등을追加導入하여 데이터베이스의範圍를 넓힌다.

INIS, AGRIS, INPADOC의 서비스開始와 함께 이들 데이터베이스를 위한國內資料를蒐集, 索引하여送附하는體系를 갖춘다.

既存데이터베이스의 RS서비스를 위하여는繼續하여入手되는資料를累積, 檢索할 수 있는 시스템을 배치方式으로 이時期에開發한다. 開發중인 KIPS-2(한글·漢字處理시스템)의完成과 함께國內進行中 研究課題를調査, 整理하여 데이터베이스化 한다. 또한國內情報파일의製作과收錄情報의檢索을 위한 우리말 디조오러스(The-saurus)研究도隨伴한다.

(2) 第2期(1979年後半-1980년)

第1期에完成된 RS시스템을活用하여第1期에使用하였던 데이터베이스 1~2年分에 대하여本格的으로 RS서비스를實施한다.

既存데이터베이스로는 COMPENDEX, JICST-IR을追加하고한글·漢字自動變換시스템을完成시킴으로써國內製品파일 및國內文獻, 特許파일을製作, 配布하며 S-5300시스템活用으로 JICST-IR 및國內情報파일들에 대하여 우리말로檢索할 수 있는 시스템을 갖춘다.

또한政府의人口分散政策에呼應하여 컴퓨터 센터를現位置에서大德으로移轉하고 온라인 시스템研究를 위하여 컴퓨터센터 内部에 2-4臺의 端末을設置하여記憶裝置의規模가比較的 작아도 가능한 데이터베이스에 한하여 온라인 SDI서비스를提供한다.

(3) 第3期(1981년)

小規模의 온라인시스템을完成한다. 우선大德의 데이터뱅크本部를中心으로하여서울, 大邱, 釜山, 光州에 端末을設置한다. 이4個의 端末은디스플레이型 端末에高速프린터를附着함으로써 미래의스위칭센터의役割을 말도록 한다.

使用 데이터베이스로는 2期の 데이터베이스에金屬分野의 METADEX와醫學分野의 MEDLARS를追加하고 RS서비스의範圍를 데이터베이스 3~5년분의 데이터로擴張한다. 이와 같은 RS서비스의擴張을 위하여서는補助記憶裝置의容量을 700MB에서 1,500-3,000MB로增設할必要가 있을 것으로豫想된다. 그러나蒐集하고 있는 14種의 데이터베이스에 대한 RS를實施하기에容量이不足한境遇에는時間單位別로 데이터베이스를 나누어使用하는方法도考慮할 수 있다.

온라인서비스에 사용되는命令語數는 약 10~15個로比較的 간단한會話의機能을 가지며 1個의命令에 대한平均應答時間은 약 10秒로推定된다.

(4) 第4期(1982년以後)

全國에 모두 25臺의 端末을設置하고 第3期の 端末이 스위칭센터의役割을 맡아 네트워크를 이루는, 온라인시스템이本格的으로稼動되는時期이다. 端末裝置의設置優先順位는情報要求

頻度에 따라 選定하며 選定된 團體 以外의 團體 또는 個人의 加入要求가 있을 때에는 모든 經費를 그 團體 또는 個人이 負擔하도록 하여 데이터뱅크의 資料를 使用할 수 있도록 한다.

이때의 端末은 비용이 저렴한 TTY型 端末로 하여 檢索에만 使用하고 檢索結果는 이 端末이 接屬된 스위칭센터에서 프린트 되도록 한다.

使用하는 데이터베이스는 第3期와 같으나 補助記憶裝置의 容量을 擴張하여 同時에 모든 데이터베이스에 대한 온라인서비스가 可能하도록 한다.

온라인에 使用하는 會話命令語數도 약 30個로 擴張함으로써 多様な 檢索式을 作成할 수 있도록 하여 檢索效率를 높이고 主記憶裝置 容量도 1MB로 增設하여 1個命令當 平均應答時間이 5秒以內가 되도록 性能을 높인다.

3. 問題點

데이터뱅크 計劃을 成功的으로 遂行하기 위하여 解決하여야 할 問題點으로는 다음과 같은 것을 들 수 있다.

3.1 네트워크의 構成方案決定

既成의 데이터베이스를 購入하여 國內 네트워크를 통하여 情報網을 構成하는 것이 비록 技術的으로는 可能하다고 하더라도 15~20種의 데이터베이스를 每年 購入, 管理하는 것이 外國 데이터뱅크의 데이터베이스를 TYMENET와 같은 人工衛星을 利用한 通信網을 통하여 使用하는 것과 比較할 때 과연 經濟性이 있는가에 대한 檢討는 아직 充分하지 못한 實情이다. 그러나 KORSTIC의 科學技術情報銀行이,

① 英文으로 構成된 科學技術情報 뿐만 아니라 國內에서 製作하는 한글·漢字를 包含한 情報를 데이터通信하여야 한다는 點

② 外國의 데이터뱅크를 利用할 경우 時差때문에 생기는 運營上의 問題

③ 人工衛星을 통하더라도 이것을 받아 處理하기 위한 中央컴퓨터시스템과 이를 國內 各地에 보내는 네트워크가 必要하다는 點

④ 國內通信技術向上 및 情報시스템 育成이라는 諸問題點을 勘案할 때 데이터베이스의 購入費는 데이터뱅크 全体運營費 중에서 큰 比重을 차지할 것 같지 않으며 오히려 衛星通信을 使用하는 費用에 相殺될 것으로 보이나 이에 대하여는 좀 더 具體的인 調查檢討가 이루어져야 할 것으로 생각된다.

3.2 우리말 데이터通信

外國의 境遇에는 英數字 및 特殊文字를 包含한 文字種數가 64-128種 밖에 되지 않아 端末에서 文字像을 發生시키는데 別로 問題가 없으나 우리나라의 경우는 한글과 漢字를 合하여 그 數가 數千種에 달하여, 文字코우드를 傳送하는 경우 各 端末에 對應하여 高價의 文字發生裝置를 附着하여야 한다. 이를 補完하는 方法으로는 文字코우드 대신 文字像을 직접 傳送하는 方法이 있으나 6~8bit면 充分한 英文字 傳送到에 비하여 1,024bit나 되는 文字像을 傳送한다는 것은 一般的으로 使用하고 있는 4,800~9,600bps의 通信回線을 使用할 때 速度面에서 問題가 있다. 近來에는 國內에서 開發된 한글용 디스플레이端末 및 日本등지에서 開發된 한글·漢字用 端末들이 있어 이를 活用하는 方案도 있으나 經濟的, 技術的으로 最善의 效果를 얻기 위하여서는 데이터코우드 및 데이터通信에 관한 充分한 研究가 必要하다.

3.3 檢索用語의 統一

外國의 경우에는 각 데이터베이스別 혹은 分野別로 用語辭典(Thesaurus)이 있어 檢索시스템 開發, 데이터베이스作成 뿐만 아니라 檢索效率를 높이기 위한 手段으로 有效하게 使用하고 있으나 우리나라에서는 科學技術用語에 대한 体系的인 整理가 되어 있지 못하다. 國內 데이터베이스의 作成 및 檢索시스템의 開發을 위하여는 國內의 科學技術用語 標準化가 時急히 이루어져야 한다.

4. 結 論

推進中에 있는 데이터뱅크는 우리나라 初有의 가장 現代化된, 그리고 全國的인 네트워크를 가진

은 1973年이래 더욱 惡化되었다. 司書나 出版業者들은 資金의 不均衡을 革新이나 協同하는 技術로 克服할 진정한 能力이 없다. 出版業者들은 더욱 더 많은 價格引上의 形式을 通해서 그들의 價格引上을 通過시키고 이것은 個人 購入者들이 圖書館보다 더 높은 비율로 取消하는 原因이 되며 圖書館들에 대한 壓迫은 增加된다. 圖書館들도 革新을 통한 解決方法은 별로 쓰지 않는다. 製本을 하지 않거나 마이크로필름으로의 代替가 약간 있으나 財政的으로 크게 도움이 되지 못한다.

價格節減의 可能性이 가장 많은 것은 技術過程의 協同이다. 또한 價格節減의 가장 좋은 方法은 自動化分野라고 主張하나 재미있는 것은 應答者中 절반이 自動化를 통한 價格節減을 報告하는 한편 나머지는 自動化 除去를 통한 價格節減을 主張했다.

이것은 1973年 以後 무엇이 繼續해서 發生해 왔나하는 見地에서 보면 明白하다. 經濟的 類型의 適應할 수 없는 것이고 만약 우리가 모든 主題分野의 出版物이 사라질 수도 있는 自由放任

主義風土를 싫어한다면, 그 制度는 連邦政府로 부터 支援을 받아야 한다. 支援範圍와 그것을 極小化하는 段階, 그리고 가장 重要한 應用分野 (著者, 出版社, 圖書館, 利用者)는 아직 더 研究가 必要하다.

引用文獻

- 1) Fry, Bernard M. and Herbert S. White / Economics and Interaction of the Publisher-Library Relationship in the Production and Use of Scholarly and Research Journals Bloomington, In., Indiana University Graduate Library School, November 1975. (Available NTIS P. B 249108) hard copy \$11.00, Microfiche \$2.25.
- 2) Goffman, W./Bradford's Law Applied to the Maintenance of Library Collections. In Saracevic, T./Introduction to Information Science, New York, Bowker, 1970.
- 3) Gore, D./The View from the Tower of Babel. Library Journal 100:1601-1604(Sept. 15, 1975).

< p. 13의 계속 >

情報銀行으로 全國 工業團地를 中心으로 한 26 個의 大都市에서 卽刻的인 檢索이 可能하도록 함으로써 迅速正確한 意思決定을 할 수 있게 하여 企業運營에 크게 기여할 것이며, 全世界의 最新 科學技術情報와 國內의 技術人力, 製品 및 研究課題에 대한 迅速한 情報는 重複研究의 防

止와 人的 技術資源을 有效適切하게 活用可能케 함으로써 國內技術發展에 功獻할 것이다.

또한 索引 및 檢索시스템에 대한 研究, 특히 우리말에 의한 檢索시스템 및 데이터通信의 開發은 國內의 情報處理技術向上을 가져올 것으로 期待된다.