

# 情報檢索 概說

崔 成 溶\*

< 目 次 >	
1. 머리 말	6.4 蓄積媒體
2. 情報檢索의 뜻	7. 檢索시스템
3. 情報檢索 過程	7.1 檢索過程
4. 情報檢索 手段	7.2 檢索手法
5. 情報檢索시스템의 類型	7.3 파일探索 方式
6. 蓄積시스템	7.4 探索結果의 評價
6.1 情報蓄積 過程	8. 情報提供 시스템
6.2 索引語	9. 맺는 말
6.3 蓄積파일 構成方式	

## 1. 머리 말

“情報檢索” 또는 “IR”이란 말은 이 分野에 관계 있는 일부 人들을 除外하고는 우리나라에서는 아직도 生소하게 들리고, 매우 어려운 것으로 느껴진다.

“情報檢索(Information Retrieval)”이라는 말이 널리 쓰이기 시작한 것은 1960年代 初부터로서, 그 時代的 背景은 “情報의 爆發的인 增加”에 눈을 떠, 닥아오는 “情報化時代”에 對處한 情報의 効率的인 處理技術이 要求되어, 그 中心的인 技術로서 登場한 데 있으며, 따라서 情報檢索은 當初부터 高度의 機械化 혹은 自動化가 前提되었고, 그 代表的인 것으로서 때마침 實用化되기 시작한 電子計算機의 利用이 前提가 되었다.

그러나 情報檢索시스템에는 電子計算機를 利用한 高度로 自動化된 시스템 뿐만 아니라, 카야드選別機를 利用하는 시스템으로부터 종래의 單純카야드 시스템에 이르기까지 여러가지 段階의 시스템이 있어 반드시 어렵기만 한 것은 아니며, 先進諸國에서는 여러 段階의 情報檢索시스템이 널리 普及되어 日常化되어 있고, 아직 낮은 水準에 屬하기는 하나 國內에서도 實施例가 없지 않다.

情報檢索은 情報處理의 中心的인 技術로서, 滿足할만한 結果를 얻으려면 적지 않은 費用과 時間과 技術이 必要하며, 특히 文獻情報活動에 있어서의 情報檢索은

“數值情報”의 處理가 아니고 “言語情報”의 處理이기 때문에 簡單하지 않고, 全面的인 機械化는 아직 不可能하며, 實用化는 되고 있으나 技術的으로는 아직도 完成된 段階는 아니다.

## 2. 情報檢索의 뜻

“情報檢索”(Information Retrieval—IR)이란 “情報의 蓄積과 檢索”(Information Storage and Retrieval)의 略語로서, 入手된 情報를 分析 評價하여 有効한 情報를 미리 蓄積하여 두었다가 必要한 때에 그 中에서 適切한 情報를 迅速히 찾아내어 利用할 수 있도록 하는 一連의 作業을 말한다. 즉 情報檢索은 情報의 蓄積없이 는 이루어지지 않으며, 蓄積된 情報 File로부터 必要한 情報를 찾아내는 일이므로 그 中心은 索引技術이다.

그러나 아무리 精巧한 索引시스템을 만들어도 蓄積하는 情報가 보잘 것 없는 것이라면 거기에서 찾아내는 情報는 쓸모가 없으며, 또 有効한 情報 File과 索引이 만들어져도 使用하는 사람이 그것을 알지 못하여 그 機能이 充分히 發揮되도록 使用하지 못한다면 애써 蓄積한 情報가 死藏되어 버린다.

따라서 情報檢索이란 單純히 情報를 찾아내는 技術이 아니며, 情報의 蒐集으로부터 分析 評價—蓄積(索引化)—檢索—提供에 이르는 一連의 綜合的인 技術이다.

그리고 情報檢索에는 그 對象으로 하는 “情報”의 單位에 따라 事實檢索, 情報檢索 및 文獻檢索의 세 가지

\* KORSTIC 企劃管理室 研究開發次長

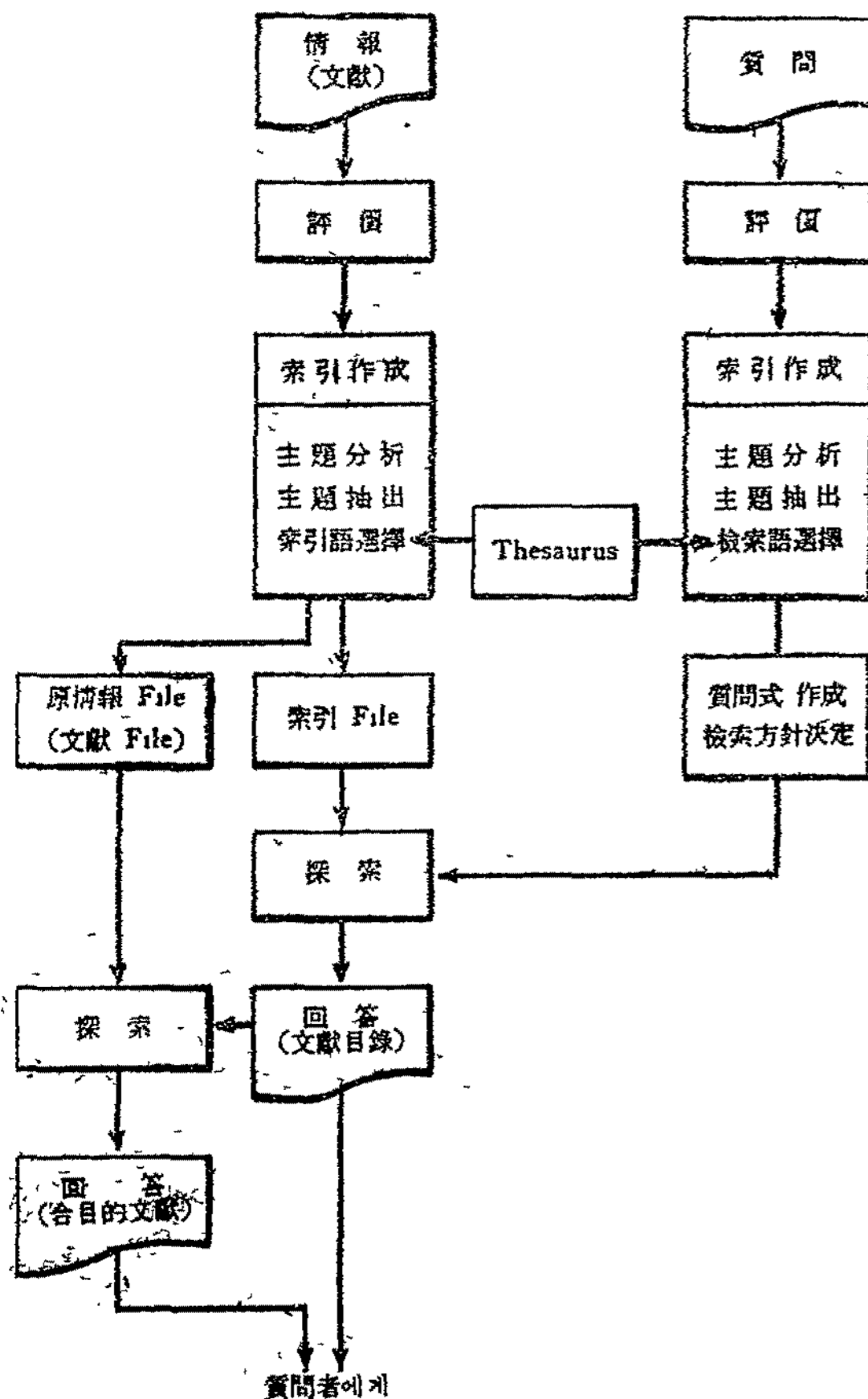
의 形態가 있으나, 文獻情報管理에 있어서의 情報檢索은 該當情報의 所在 즉 “該當情報가 記載되어 있는 文獻”을 指示하고, 또 指示된 文獻을 찾아내는 “文獻檢索 (Document Retrieval—DR)”을 意味한다.

그리고 情報檢索의 手法에는 單純한 카아드를 使用하여 손으로 必要한 情報을 찾아내는 簡單한 方法으로부터 電子計算機 등을 利用하는 高度로 機械화된 複雜한 方法에 이르기까지 여러가지의 方法이 있으나, 現代의 意味의 情報檢索(IR)은 高度로 機械화된 情報檢索 方法, 즉 “情報의 機械檢索” 또는 “電子計算機에 의한 情報檢索”을 뜻하는 경우가 많다. 따라서 最近 文獻情報管理에 있어서의 情報檢索이란 “機械에 의한 文獻檢索”, “電子計算機에 의한 文獻찾기”를 意味하는 경우가 많다.

### 3. 情報檢索 過程

情報檢索(文獻檢索)의 過程을 크게 나누어 보면 情報의 蓄積過程과 檢索過程으로 區分할 수 있다.

情報의 蓄積過程은 蓄積할 情報의 選擇, 評價 및 主題分析, 蓄積해야 할 概念(主題語)의 抽出과 거기에 對應하는 索引語의 選擇(索引化)의 順으로 進行되며, 索引 File과 原情報 File(文獻 File)의 두 가지 File을 作成하



<圖 1> 情報檢索 過程

여 됨으로써 情報을 蓄積한다. 索引 File(Term File)은 索引카아드와 같은 文獻 File의 所在(Address)를 包含하는 索引(Index)의 File을 말하며, 原情報 File(Item File)은 學術雜誌, 마이크로필름 등과 같은 文獻資料의 現物 File을 말한다.

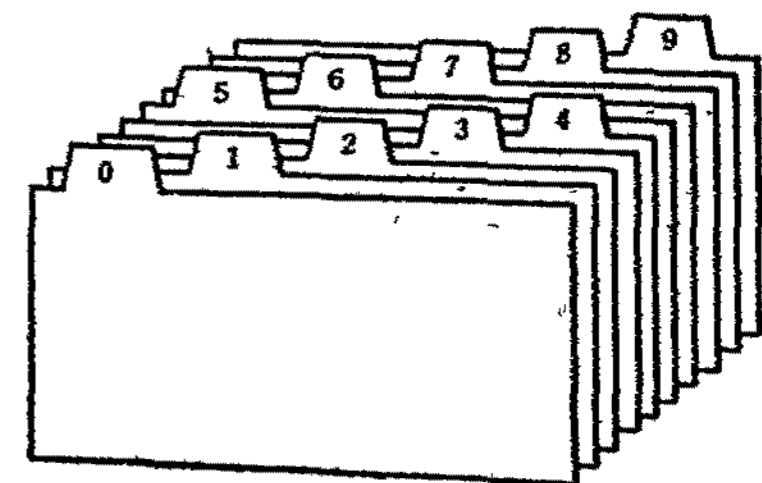
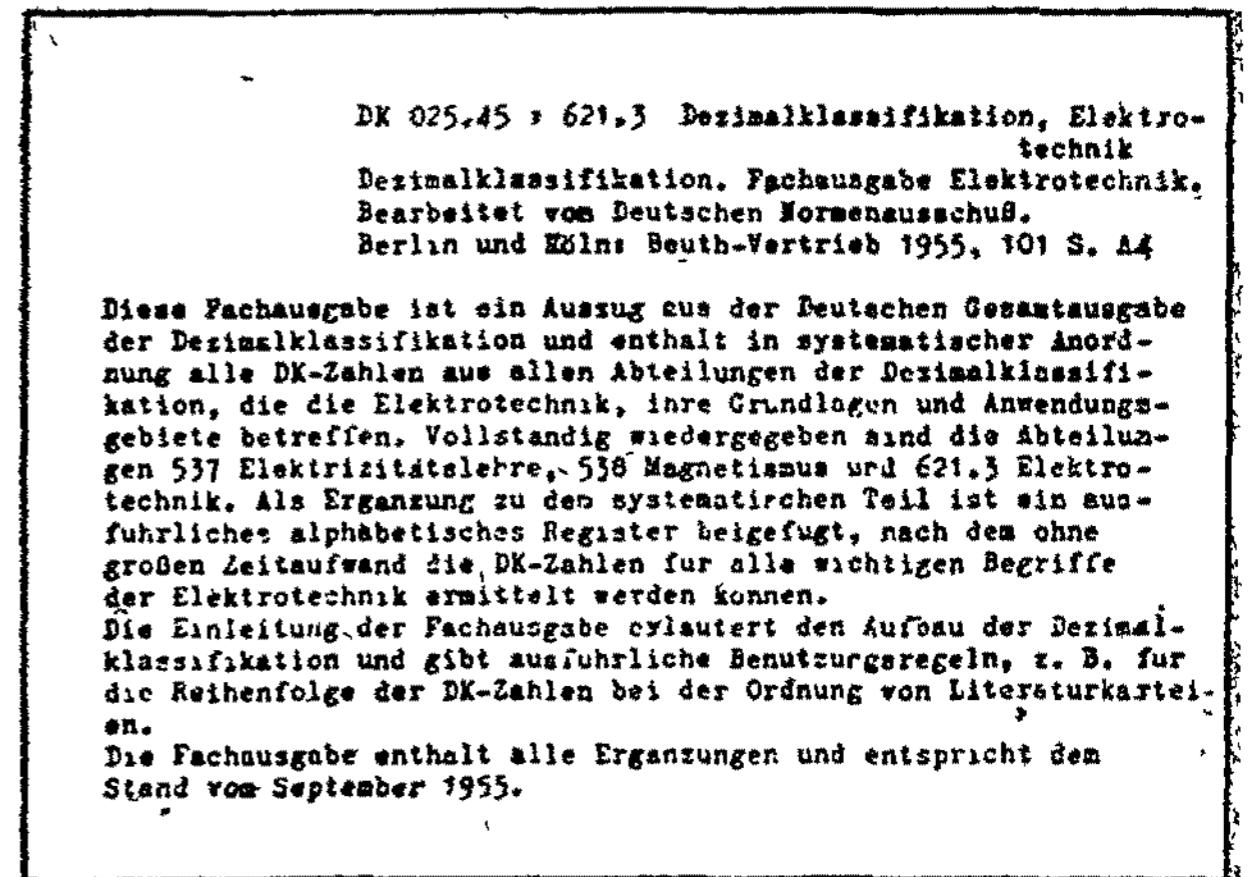
情報의 檢索過程은 情報檢索 要求(質問)의 評價 및 主題分析, 檢索해야 할 概念(主題語)의 抽出과 거기에 對應하는 檢索語의 選擇(索引化), 檢索語 사이의 論理關係의 究明(質問式의 作成) 및 檢索方針의 決定, 情報가 蓄積된 索引 File과 檢索語의 對照에 의한 探索處理, 探索結果 選出된 適合情報의 所在(Address)의 提示, 原情報 提供要求에 따르는 原情報 File 探索, 要求에 適合한 情報의 選出 및 提供의 順으로 進行된다.

이와 같은 情報의 蓄積과 檢索過程中 그 主要部分을 차지하는 索引化까지의 過程은 人間頭腦에 의한 處理過程이며, 機械화가 可能한 處理過程은 一部 蓄積過程과 探索處理 및 提示過程 뿐이다.

### 4. 情報檢索 手段

情報檢索(文獻檢索)이란 索引化하여 蓄積한 文獻情報을 機械的인 手段으로 찾아내는 것을 意味하는데, 機械的인 要素가 어느 程度 物理的으로 加味되느냐에 따라 檢索手段을 分類하여, 보면 대체로 다음의 네 가지의 水準으로 區分된다.

- ① 無孔카아드 方式
- ② 手動式 穿孔카아드 方式
- ③ 카아드選別機를 利用하는 方式

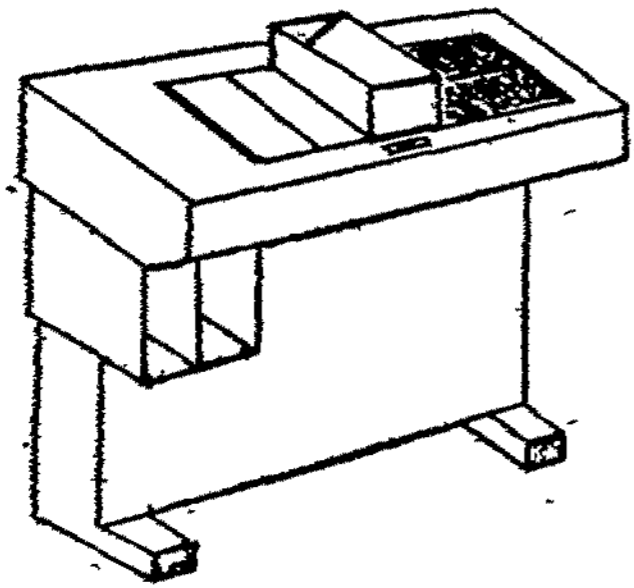


<圖 2> 單純카아드 시스템





것이 特徵이나, 機械裝置의 制約 때문에 設計 및 使用의 自由度가 보다 制限된다.

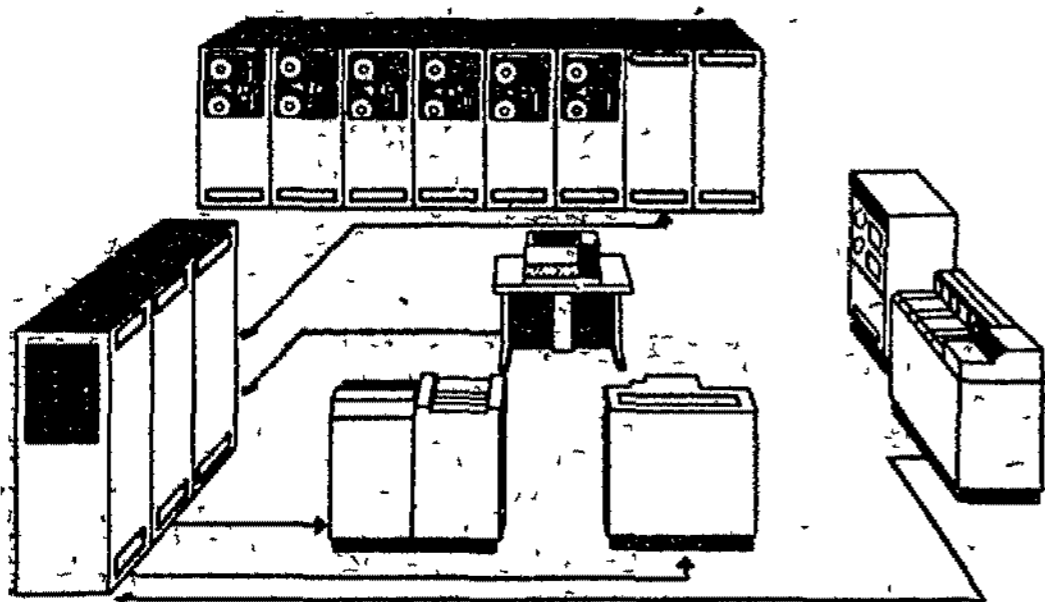


<圖 6> 多欄카드 選別機에 의한 檢索시스템

펀치(穿孔) 또는 마아크判讀(Mark sense)方式의 多欄 카드選別機(Multi-Column Card Selector)를 利用하는 檢索시스템이 여기에 屬하며, IBM 規格 카드의 一定한 位置에 窓을 내고 마이크로필름片(Chip)을 붙인 Aperture Card System과 마이크로필름에 文獻像과 함께 적어 넣은 格子무늬 2進法 Code에 의하여 檢索하는 MIRACODE System, Filmorex System 등 마이크로 필름을 利用하는 情報檢索시스템도 이에 包含된다.

4.4 電子計算機를 利用하는 方式

電子計算機의 高度로 迅速하고 龍大한 記憶能力과 判斷能力 및 制御能力 등을 利用하여 情報의 蓄積과 檢索過程의 많은 部分을 機械化(自動化)한 것으로서, 情報의 變換, 加工 處理 能力 및 速度에 있어 다른 方式이 따를 수 없는 뛰어난 能力을 갖추고 있으나, 索引化 및



<圖 7> 電子計算機에 의한 檢索시스템

入力까지의 過程은 機械化가 아직 不可能하며, 運營費用이 많이 들고, 接續面에 있어서의 制約이 많아 周邊 시스템도 全部 電算機에 맞도록 構成해야 하는 難點이 있다.

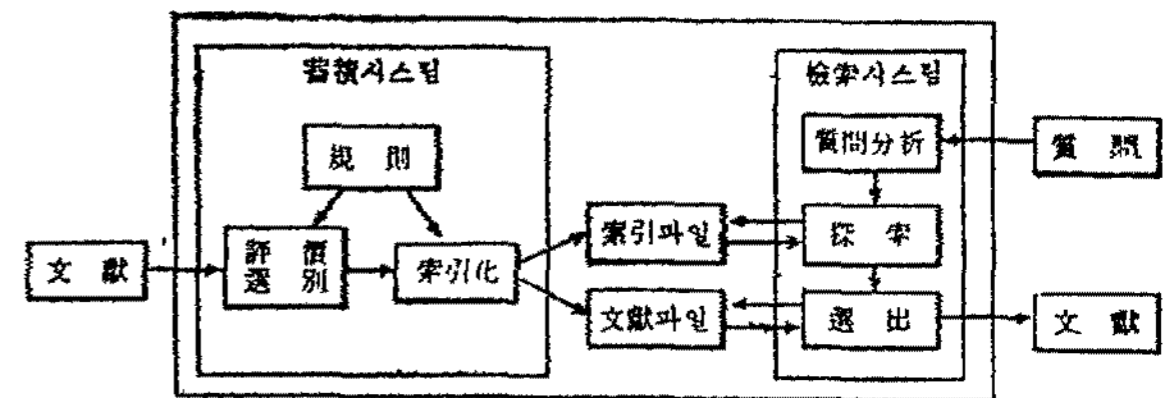
最近 先進諸國에서는 電算機時代에 便乘하여 처음부터 電算機를 前提로 한 情報檢索시스템을 設計하는 傾向이 있으나, 어떤 경우에나 電算機를 利用한 시스템이 가장 效果的인 것은 아니다.

5. 情報檢索시스템의 類型

情報檢索시스템은 그 水準에 따라 다음과 같은 세 가지 類型으로 區分된다.

(1) 蓄積-檢索型

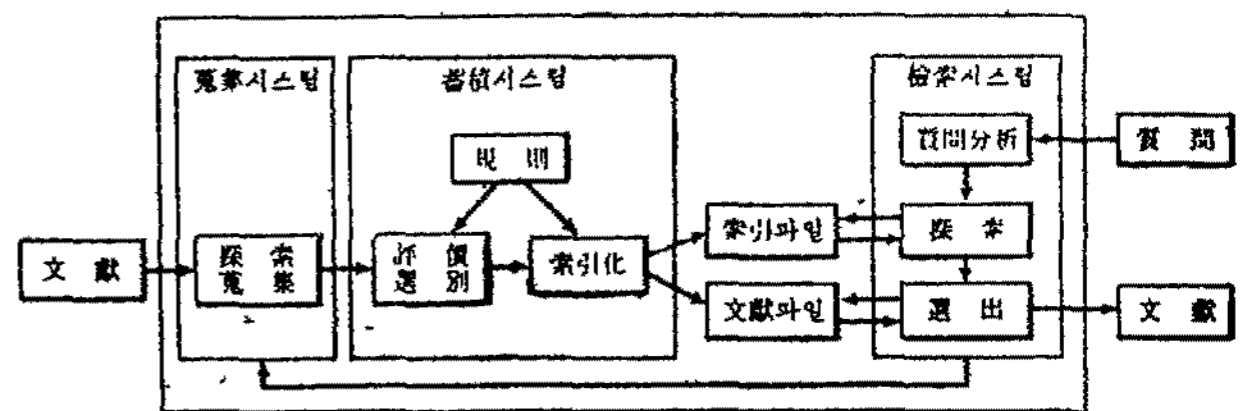
蓄積과 檢索(選出)의 두 시스템으로 構成된 가장 單純한 낮은 水準의 情報檢索시스템이다.



<圖 8> 蓄積-檢索型 情報檢索시스템

(2) 蒐集-蓄積-檢索型

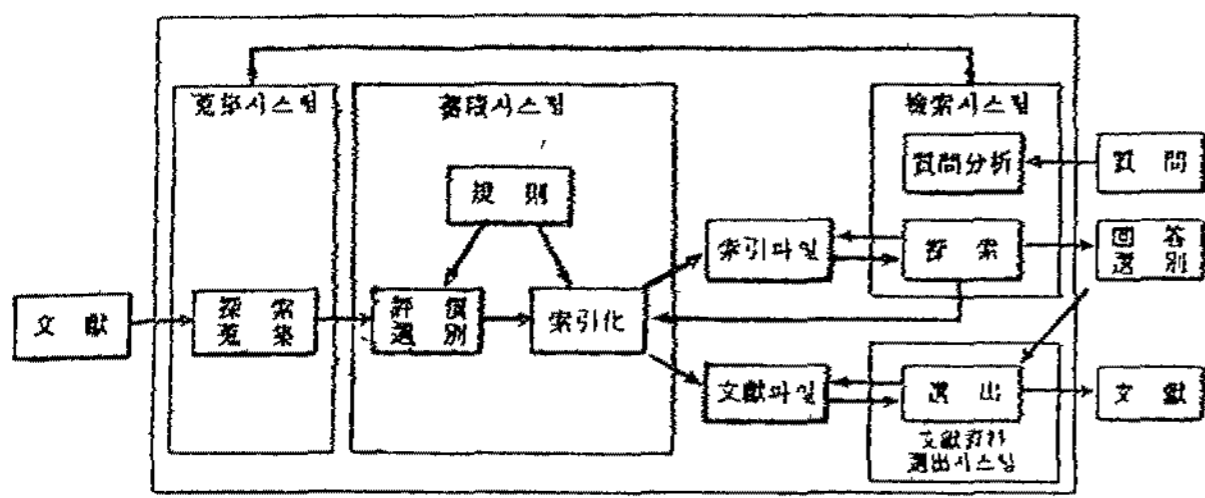
蓄積 및 檢索시스템 外에 良質의 文獻 File을 만들기 위하여 蒐集시스템을 追加시키고, 檢索시스템과의 사이에 feed back 回路를 만든 向上된 水準의 시스템이다.



<圖 9> 蒐集-蓄積-檢索型 情報檢索시스템

(3) 蒐集-蓄積-檢索-提供型

蒐集, 蓄積 및 檢索시스템 外에 시스템의 能力이 充分히 活用되도록 하기 위하여, 利用者와 시스템 사이에 中介役割을 하는 提供시스템을 追加시키고, 利用者와 提供시스템과의 사이와, 檢索시스템과 蓄積, 蒐集시스템 사이에 各各 feed back 回路를 만든 가장 複雜한 높은 水準의 시스템이다.



〈圖 10〉蒐集—蓄積—檢索—提供型 情報檢索시스템

## 6. 蓄積시스템

情報檢索시스템의 中心은 蓄積과 檢索이다. 蓄積과 檢索은 密接하게 相互 關聯되어 어느 한쪽 方式이 決定되면 다른 한쪽은 이에 따라 거의 決定되어 버린다. 시스템 設計의 경우에는 目的과 範圍가 決定되면 먼저 適切한 檢索方式을 決定하고 이에 따라 蓄積方法을 決定한다.

### 6.1 情報蓄積 過程

情報의 蓄積過程은 다음과 같다.

(1) 蓄積할 情報의 選擇과 受入

- ① 蓄積할 文獻의 選擇
- ② 登錄臺帳 記帳 및 文獻 一連番號 附與
- ③ 文獻所在 記號(文獻이 File되는 位置) 附與

(2) 文獻分析

情報檢索을 單純化하기 위하여 文獻의 確認, 檢索에 있어서의 接近點, 內容의 表示, 文獻의 代用으로 쓸 수 있는, 文獻을 濃縮시켜 集約적으로 表現한 特性語를 文獻으로부터 抽出한다. 그 過程과 內容은 다음과 같다.

① 書誌的 事項의 抽出

著者名, 標題名, 雜誌名, 卷, 號, 페이지, 年度 등을 抽出한다.

② 主題分析

文獻의 內容(主題)을 分析하여 그 主要 構成要素가 되는 몇개의 概念(文獻內容을 集約적으로 表現한다고 認定되는 適切한 語句 즉 主題語)을 抽出한다.

主題語의 抽出에는 그 基準이 必要하나, 主題語의 數를 限定하는 方法보다는 分野別로 一定한 Category를 定하여 各 Category에 該當하는 主題語를 抽出하는 方法이 바람직하다.

(3) 索引語의 標準化

情報檢索에 있어서의 用語 使用의 混亂을 막고 檢索을 效率化하기 위하여, 主題分析에 의하여 文獻에서 抽出한 主題語를 標準化된 索引語로 變換한다. 文獻中에 使用되는 用語는 著者에 따라 다른 것을 使用하여 統一

性이 없으므로, 文獻用語의 同義語와 包含關係를 整理하여, 한 概念에 대하여 한 用語를 그 概念의 代表로 定한 標準化된 “索引語”로 바꾸어 놓을 必要가 있다. 이때 이와 같은 索引語를 모아서 表로 만든 索引語辭典, Thesaurus, Keyword List, 主題名標目表, 分類表 등을 使用할 必要가 있다.

(4) 索引語의 Code化

索引語의 表現의 簡潔化, 配列 順序의 決定, 相互間의 關係의 表示, 異義語의 區分 등을 위하여 索引語를 數字 또는 글자로 된 記號(Code)로 變換시킨다.

많은 시스템들은 이것을 省略하고 自然語를 직접 索引語로 使用하고 있어 반드시 必要한 過程은 아니나, 경우에 따라서는 Code化가 극히 有利한 경우도 있다.

(5) 파일作成

主題分析이 끝난 文獻은 文獻所在(Address)記號에 따라 파일(File)되고, 索引語와 書誌的 事項은 Filing에 適合한 記號로 變換되어, 文獻所在記號와 함께 索引파일에 收錄된다.

### 6.2 索引語

現在 情報檢索용으로 使用되고 있는 “索引語”에는 ① 體系的 分類, ② 主題名索引, ③ Keyword索引, ④ Facet分類的 네 가지 種類가 있다.

(1) 體系的 分類

나무가지가 갈라지는 것처럼(Tree 構造로) 主題內容을 體系的으로 細分하여 가서 索引語를 定하는 方法으로서, 그 長點은 索引作成者나 索引利用者가 自己가 알려진 主題가 어떤 말로 表現되어 있는지 알지 못하여도 그 上位概念 또는 下位概念을 알면, 求하는 索引語에 容易하게 到達될 수 있도록 되어 있는 點이며, 複數의 主題를 가진 文獻을 미리 定해져 있는 體系的 分類項目의 하나에 無理로 適用시켜야 하는 것과, 分類項目을 任意로 追加시킬 수 없고, 새로운 變化에 대한 適應이 늦는 것 등이 缺點이다. (일반적으로 이와 같은 缺點을 補完하기 위하여 複數의 分類項目을 索引語로 選擇하는 것이 보통이다). 國際 10進分類法(UDC)과 各國 特許分類 등이 여기에 屬한다.

(2) 主題名索引

主題를 表現하는 用語로서 使用하도록 미리 定해진 用語리스트(主題名標目表)中에서, 主題內容을 表現하는데 適合한 말들을 必要한만큼 選擇하여 索引語로 使用하는 方法으로서, 文獻의 主題內容에 대한 表現能力이 豊富하고 새로운 主題에 대하여 새로운 主題名을 任意

로追加시킬 수 있는長點이 있으나, 使用法에 熟達되지 못하면 混亂을 招來하기 쉽다.

主題名이 알파벳順으로 配列되어 있어 體系的 分類法과 크게 다른것 처럼 보이나, 實際로는 主題名索引에도 體系的 分類方式이 加味되어 있고, 體系的 分類法에도 알파벳順으로 分類項目이 配列된 相關索引이 있어 兩者의 差異가 보기 보다는 크지 않다. 美國 議會圖書館(LC)의 主題名標目表가 有名하다.

(3) keyword 索引

主題를 表現하는 用語(Keyword)를 그 文獻 속에 包含된 自然語의 用語中에서 任意로 選擇하여 使用하는 方法으로서, 表現能力이 豊富하고 索引語의 選擇에 制限이 없어 索引作成이 比較的 容易하나, 實際로는 使用하는 用語의 同義語, 異義語, 上位概念(BT)과 下位概念(NT), 關聯語(RT) 등을 整理하고, 한 概念에 대하여 한 用語를 그 概念을 代表하는 用語로 指定하여 使用하는 用語를 制限하고, 文獻中에서 抽出한 用語를 指定된 用語로 바꾸어서 標準化된 用語를 使用하지 않고서는 用語 使用의 混亂으로 索引이 不確實하여져서 쓸모가 없어진다.

RT	STORAGE	RT	SEPARATION
RT	UPGRADING	RT	SWEATING
RT	WATER REMOVAL	RT	VAPORS
CONCENTRATIONS		RT	WATER REMOVAL
		RT	WETNESS
RT	COMPOSITIONS	CONDENSERS /ELECTRICAL/	
RT	PROPORTIONS	RT	ALTERNATING CURRENT
RT	TITRATION	RT	CAPACITORS
CONCENTRATORS		RT	CAPACITY
		RT	CHARGES /NOT ACCOUNTS/
RT	CENTRIFUGES	RT	CONDUCTORS
RT	CLASSIFIERS	RT	CURRENTS
RT	DECANTER	RT	DIELECTRICS
RT	DEHYDRATORS	RT	ELECTRICITY
RT	DEMISTERS	RT	IMPEDANCE
RT	DEPHLEGMATORS	RT	RESISTORS
RT	DEW FILTERS	CONDENSERS /PROCESS EQUIPMENT/	
RT	EVAPORATORS	RT	DEPHLEGMATORS
RT	EXTRACTORS	RT	COLLECTORS
RT	FILTERS	RT	COOLERS
RT	HELVIXES	RT	DEHUMIDIFIERS
RT	HILMPHREY SPIRALS	RT	DEMISTERS
RT	JIGS	RT	HEAT EXCHANGERS
RT	PRESSES	RT	PRECIPITATORS
RT	SCREENS	RT	SEPARATORS
RT	SEPARATORS	RT	SPRAY
RT	SETTLERS	RT	STILLS
RT	STILLS		
RT	STRAINERS		
RT	THICKENERS		
RT	TRAPS		

<圖 11> Thesaurus의 例

따라서 일반적으로는 文獻中에서 抽出한 自然語의 用語(Keyword)를 標準化한 索引語(Descriptor)로 바꾸어서 索引化하고, 複合主題의 索引에는 主題語의 論理的인 組合에 의한 索引法(Coordinate Indexing)을 利用하고(例: Uniterm Card System), 索引語의 相互關係의 役割(Links와 Roles)을 表示하여 檢索效率을 높이고 있다.

規制되지 않은 自然語의 用語를 使用하는 Keyword 索引의 例로는 標題를 對象으로 하여 電子計算機에 의하여 自動處理하여 만드는 KWIC(Key Word in

Context)索引이 있으며, 標準化된 索引語辭典으로는 EJC Thesaurus (Thesaurus of Engineering and Scientific Terms), DDC Thesaurus(Thesaurus of DDC Descriptors) 등이 有名하다.

(4) Facet分類

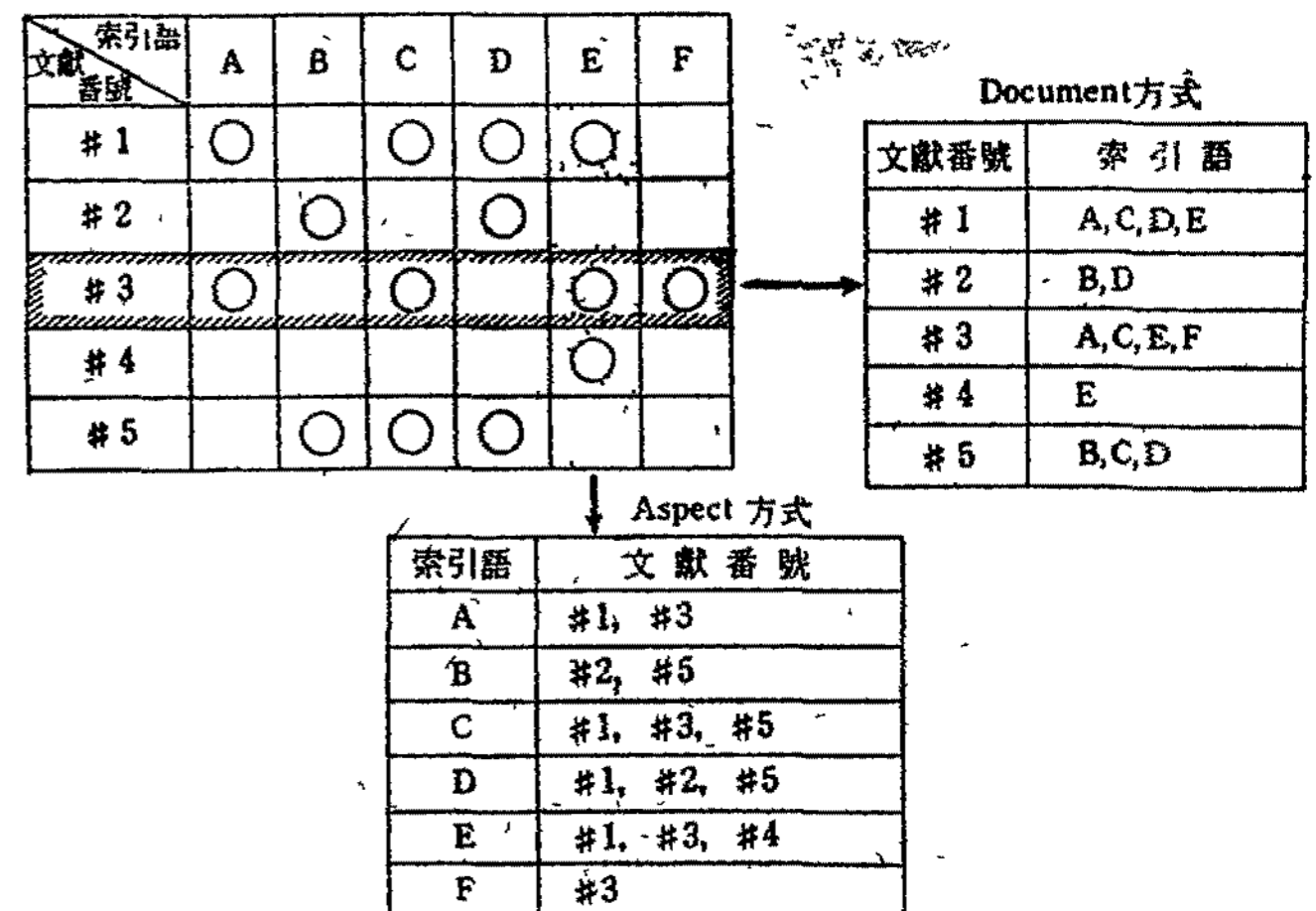
文獻에서 抽出한 主題語들을 그 屬性別로 나누어 보면 몇개의 큰 Category로 區分되는데, 이것을 利用하여 各 分野別로 몇개씩의 Category를 定하고(例: 化學 分野 ① 生産物, ② 原料, ③ 反應, ④ 觸媒, ⑤ 溶液, ⑥ 操作, ⑦ 條件, ⑧ 裝置, ⑨ 測定, ⑩ 性質, ⑪ 構造), 이 Category에 따라 主題를 分析하고 主題語를 抽出하여 索引語로 定하는 方法으로서, 一定한 基準에 따라 主題를 分析하기 때문에 누가 分析하여도 큰 差異가 생기지 않으므로 均質的인 索引을 作成할 수 있고, 檢索時에도 이 Category에 맞추어서 質問을 構成하면 適合한 文獻을 틀림 없이 얻을 수 있으나, Category의 設定과 이 Category에 의한 主題語의 抽出이 쉽지 않다.

6.3 積蓄 File 構成方式

情報 蓄積 File(索引 File) 構成方式에는 다음과 같은 두 가지 방식이 있다.

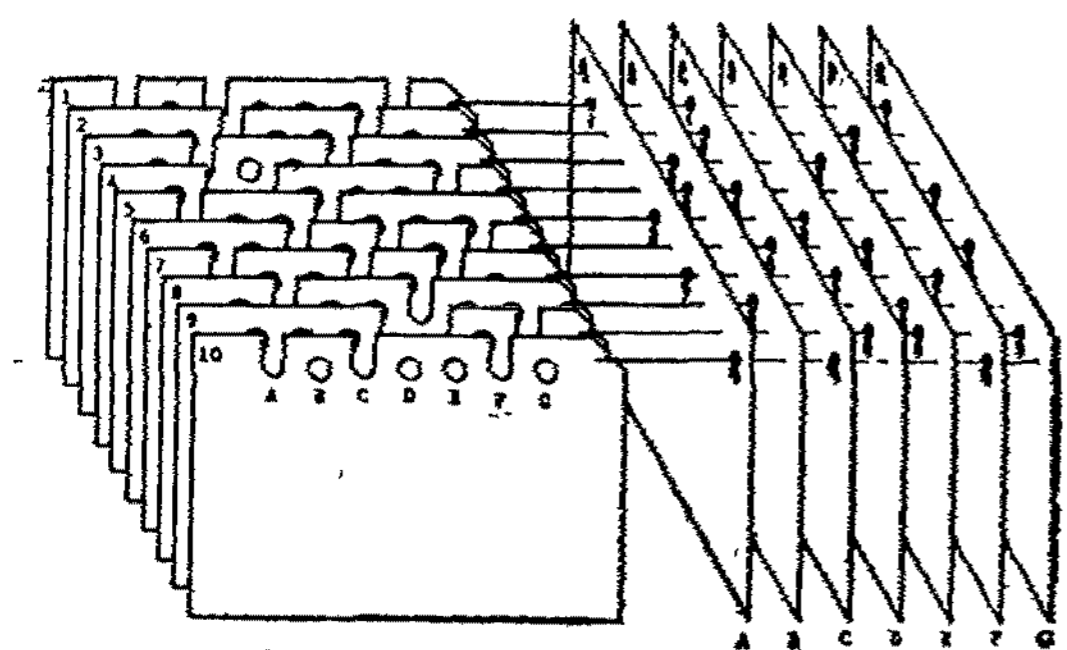
(1) 文獻單位의 File 構成方式

索引語를 文獻單位로 記載하여 File해 두는 방식(Serial File)으로서, Document方式이라고도 하며, 檢索時에 File을 全部 調査하여야만 必要한 情報를 찾아 낼 수 있게 되어 있어(Search 方式), 많은 勞力과 時間이 所要되나, 電子計算機의 發達로 實用性이 높아졌다. 機械 檢索시스템에서 주로 이 방식이 利用된다.



<圖 12> 蓄積파일 構成方式





<圖 13> File 構成方式의 比較(手動式 펀치 카아드 시스템과 Peek-a-boo Card System)

(2) 索引語單位의 File 構成方式

文獻(番號)을 索引語 單位로 記載하여 File해 두는 방식(Inverted File)으로서, Aspect 方式이라고도 하며, 檢索時에 File의 該當部分만 調査하면 되므로 檢索이 簡便하나, File 作成에 많은 勞力과 時間이 所要된다. 주로 매뉴얼(Manual) 카아드 시스템에 이 방식이 利用된다.

6.4 蓄積媒體

情報 蓄積媒體에는 다음과 같은 것들이 使用된다.

(1) 原文獻 File用

- ① 原文獻 그대로.
- ② 마이크로필름(Roll Film 또는 Microfiche).
- ③ 磁氣테이프 또는 磁氣 Disk.
- ④ 其他 Video-Tape 등.

(2) 索引 File用

- ① 冊子形態: 各種 索引誌(KWIC, KWOC 등).
- ② 카아드形態: 單純카아드, 手動式 펀치 카아드(HSPC), Uniterm Card, Peek-a-boo Card, 機械檢索 펀치 카아드(MSPC), Microfiche 등.
- ③ 테이프形態: 磁氣테이프, 마이크로필름(Roll Film).
- ④ Disk 形態: 磁氣 Disk.

7. 檢索시스템

7.1 檢索過程

檢索要求에 따라 蓄積된 情報를 檢索하여 提供하는 過程은 다음과 같다.

- (1) 檢索要求(質問)의 接受.
- (2) 檢索要求(質問)의 分析(主題分析).

(3) 檢索해야 할 概念(主題語)의 抽出.

(4) 抽出된 主題語와 對應하는 檢索用語(索引語)의 選擇.

檢索漏落과 不必要한 情報가 檢索되는 것을 防止하기 위하여서는 檢索用語는 蓄積에 使用된 索引語와 同一한 것이 使用되어야 하며, 이를 위해서는 標準화된 檢索用語集(Thesaurus, 主題名標目表, 分類表 등)을 使用할 必要가 있다.

(5) 檢索語사이의 論理關係의 究明

不必要한 情報(Noise)가 檢索되는 것을 막고 適合한 情報만이 檢索되도록 하기 위하여서는 檢索語사이의 論理關係를 밝힐 必要가 있다(機械檢索의 경우에는 論理記號를 使用하여 이와 같은 關係를 表示하는 質問式을 正確하게 作成할 必要가 있다).

(6) 檢索方針의 決定

回答의 要求程度(包括적인 답이 要求되는가, 詳細한 답이 要求되는가)에 따라 檢索方針을 決定한다.

(7) 檢索指示와 探索處理

檢索方針과 正確한 質問式에 따라 檢索裝置에 檢索指示를 하여, 情報가 蓄積된 索引 File과 檢索語의 對照에 의하여 探索處理하여, 檢索指示와 一致된 情報를 選出한다.

(8) 選出된 適合情報의 所在(Address) 提示.

(9) 原情報(文獻) 提供要求의 接受 및 原情報(文獻) File의 探索.

(10) 要求에 適合한 原情報(文獻)의 選出 및 提供.

7.2 檢索手法

檢索手法으로는 概念(主題語)의 論理的 關係를 利用하여 蓄積된 情報中에서 合目的情報를 찾아내는 論理檢索法이 흔히 利用된다.

즉, 效率적인 檢索을 하려면 質問內容을 分析하여, 檢索하려는 主題를 나타내는 概念들의 論理關係를 밝혀 正確한 質問을 作成하고, 이것을 論理記號로 表示하여 質問의 條件을 滿足시키는 正確한 質問式을 作成할 必要가 있다.

그리고, 檢索結果 包括적인 답이 要求될 경우에는 可及的 包括적인 檢索用語를 使用하고 많은 關聯語를 論理和(OR)로 組合시켜야 하며, 詳細한 답이 要求될 경우에는 詳細한 檢索用語를 使用하고 몇개의 質問項目을 論理積(AND)으로 組合시켜야 한다.

論理檢索法에서 흔히 利用되는 論理記號와 論理式은 다음과 같다,

[論理記號]

- 一致 : A A인 것.
- 否定(NOT) :  $\bar{A}$  A가 아닌 것.
- 論理積(AND) :  $A \cdot B$  A이면서 B인 것.
- 論理和(OR) :  $A+B$  A 또는 B인 것.

[論理式]

(例) 『自動抄錄(A) 또는 自動分類(B)의 精度(C) 또는 費用(D)에 관한 文獻으로서 日本語文獻(E)이 아닌 것』

$$(A+B) \cdot (C+D) \cdot \bar{E}$$

[論理式的 分解]

- $A \cdot C \cdot \bar{E}$
- $A \cdot D \cdot \bar{E}$
- $B \cdot C \cdot \bar{E}$
- $B \cdot D \cdot \bar{E}$

7.3 파일探索 方式

質問의 分析에 의하여 作成된 檢索指示에 따라 蓄積情報파일(File)로부터 合目的情報를 찾아내는 파일探索方式에는 다음과 같은 方法들이 있다.

(1) 逐次探索法(Sequential Search法)

蓄積情報파일의 첫머리로부터 차례로 더듬어서 合目的情報를 찾아가는 方法으로서, 蓄積順序가 一定치 않고 파일이 적을 때, 파일의 變更이 頻繁할 때에 有効하나, 處理時間이 길어지는 것이 缺點이다.

(2) 2分探索法(Binary Search法)

蓄積情報파일이 一定한 順序(數值 또는 文字順)로 配列되어 있을 때, 파일 한가운데에 位置한 情報를 檢索項目과 比較 對照하여, 그 大小關係에 따라 다음 檢討範圍를 上半分 또는 下半分으로 判定하고, 이와 같은 일을 合目的情報가 檢出될 때까지 되풀이하여 探索하는 方法으로서 極히 高速으로 目的情報를 探索할 수 있으며, 그 效果는 파일이 클수록 顯著하나, 蓄積情報의 追加가 번거로워서 파일의 變更이 적을 때에 Inverted File에 適用된다.

(3) Random Search 法

1回 또는 最少回數의 反復 探索으로서 直接 目的情報에 到達하도록 하는 方法으로서, 蓄積情報量이 많고 蓄積順序가 一定치 않을 때 有効하며, Random Access 裝置(磁氣 drum, 磁氣 Disk 등)에 適用된다.

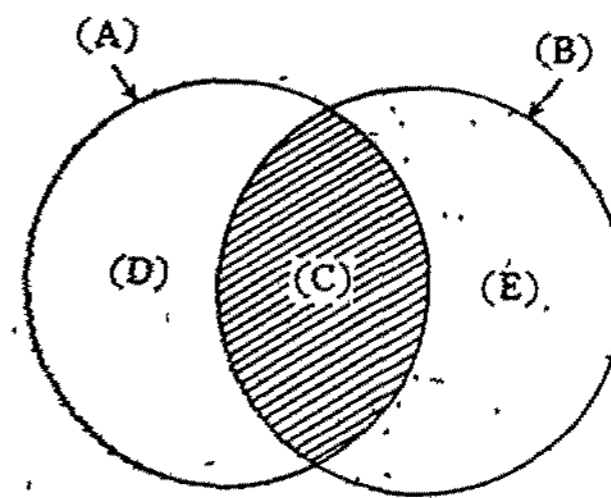
7.4 檢索結果의 評價

檢索結果는 檢索效率에 의하여 評價된다. 일반적으로 適合한 情報가 漏落없이 檢索되는 比率(Recall Factor, 再現率)과 不必要한 情報가 檢索되지 않는 比率(Relevance Factor, 適合率)에 의하여 檢索結果가 評價된다.

$$\text{再現率} = \frac{[\text{檢索된 適合文獻數}]}{[\text{檢索對象 文獻中の 適合文獻數}]} = \frac{C}{A}$$

$$= \frac{C}{D+C}$$

$$\text{適合率} = \frac{[\text{檢索된 適合文獻數}]}{[\text{檢索된 文獻數}]} = \frac{C}{B} = \frac{C}{E+C}$$



- (A) 檢索對象 文獻中の 適合文獻
- (B) 檢索된 文獻
- (C) 檢索된 適合文獻
- (D) 適合文獻中에서 檢索되지 않은 文獻(檢索漏落) (第1種의 誤差)
- (E) 檢索된 不適合文獻 (Noise) (第2種의 誤差)

<圖 14>

理想的인 檢索은 A, B, C가 完全히 一致되고 再現率과 適合率에 1(100%)이 될 때이며, 再現率을 높이려면 索引을 깊이하고 Thesaurus를 活用하는 것이 效果的이며, 適合率을 높이려면 索引語의 關聯表示(Links)와 役割表示(Roles) 등이 效果的이다.

適合率과 再現率의 어느 것에 重點을 두느냐 하는 問題는 Noise를 적게 하느냐, 檢索漏落을 적게 하느냐 하는 問題로서, 일반적으로는 Noise가 多少 많아져도 檢索漏落을 적게 하는 것이 바람직하다. Noise의 除去는 利用段階에서도 處理可能하나 檢索漏落은 救濟의 길이 없기 때문이다.

8. 情報提供 시스템

8.1 情報提供 方式

情報檢索을 情報서비스의 立場에서 보면, 最新情報를 繼續적으로 迅速히 利用者에게 提供하는 速報的 서비스(Current Awareness)와 利用者의 要求에 따라 一定期間 過去에 遡及하여 情報를 檢索하여 提供하는 遡及的 檢索서비스(Retrospective Search)로 大別할 수 있다.

한편, 이것을 利用對象者別로 나누어 보면, 不特定多數 利用者를 對象으로 하는 서비스와 特定個人을 對象으로 하는 서비스로 區分할 수 있다.



利用對象	不特定多數人	特定個人
速報的	索引誌	SDI (情報選擇提供)
邇及的	累積索引	個別檢索 (Q-A)
提供內容	標題	標題, 抄錄, 原文

8.2 不特定多數人을 對象으로 하는 情報檢索서어비스

不特定多數人을 對象으로 하는 情報檢索서어비스로는 索引誌와 累積索引(Cumulative Index)등 冊子體의 情報檢索 Tool을 作成하여 配布하는 서어비스가 있다. Keyword 또는 著者名에 의하여 利用者 自身이 必要

한 情報를 직접 檢索하여 利用하는데 便利한 索引을 定期的으로 作成하여 配布하고, 이것을 一年分 또는 多年分을 邇及的으로 檢索하여 再編輯한 累積索引을 作成配布하는 서어비스로서, 電子計算機에 의하여 自動적으로 作成 提供하는 KWIC(Key Word in Context)索引誌, KWOC(Key Word out of Context)索引誌 등이 代表的이다.

KWIC索引은 標題中에서 選擇한 Keyword를 標題의 文脈中에 단체 페이지의 中央에 配列하고, KWOC索引은 標題中에서 選擇한 Keyword를 標題머리에 내놓고 같이 配列하는 索引이다.

FORMATION SYSTEM IN THE USSR.*	/A DESIGN OF HIGH SPEED KANJI PRINTER.*	FID-67-3D5
ESE KEYWORD INDEXING SIMULATOR (JAKIS) SYSTEM. 3. STATISTICAL ASSOCIATION METHOD.*	/DEVELOPMENT AND STATE-OF-THE-ART OF THE SCIENTIFIC AND TECHNICAL INFORMATION SYSTEM IN THE USSR.*	FID-67-3A1
LATOR (JAKIS) SYSTEM. 2. DICTIONARY STRUCTURE AND STATISTICAL DATA.*	/STATISTICAL ASSOCIATION METHOD.*	FID-67-3E3
LM CORPORATION.*	/JAPANESE KEYWORD INDEXING SIMULATOR (JAKIS) SYSTEM. 2. DICTIONARY STRUCTURE AND STATISTICAL DATA.*	FID-67-3E2
EVALUATION OF A STUDY IN THE FIELD OF INFORMATION STORAGE AND RETRIEVAL.*	/DOCUMENT STORAGE AND RETRIEVAL SYSTEM FOR GENERAL ANILINE & FILM CORPORATION.*	FID-67-3C4
CULUM.*	/EVALUATION OF A STUDY IN THE FIELD OF INFORMATION STORAGE AND RETRIEVAL.*	FID-67-3G2
ECHNOLOGY B/THE U.S. PATENT OFFICE CLASSIFICATION SYSTEM: AN OPERATIVE, UNIVERSAL, CLASSIFICATION OF TECHNOLOGY B/THE U.S. PATENT OFFICE CLASSIFICATION SYSTEM: AN OPERATIVE, UNIVERSAL, CLASSIFICATION OF	UNIVERSE OF SUBJECTS: ITS STRUCTURE AND DEVELOPMENT* IN THE CURRICULUM.*	FID-67-2-9
ND CLASSIFICATION (REWDA)AN INFORMATION RETRIEVAL SYSTEM BY EDPS BASED ON THE USE OF NATURAL LANGUAGE AND CLASSIFICATION (REWDA)AN INFORMATION RETRIEVAL SYSTEM BY EDPS BASED ON THE USE OF NATURAL LANGUAGE	AN AUTOMATIC ABSTRACTING SYSTEM BASED ON STRUCTURAL SEMANTICS.*	FID-67-3H2
AGMENTATION CODE SYSTEM.*	MECHANIZED RETRIEVAL SYSTEM FOR ORGANIC COMPOUNDS. AN EVALUATION OF THE FRAGMENTATION CODE SYSTEM.*	FID-67-3E1
THE DEVELOPMENT AND USE OF OPTICAL COINCIDENCE SYSTEM IN THE USA.*	MECHANIZED RETRIEVAL SYSTEM FOR ORGANIC COMPOUNDS. AN EVALUATION OF THE FRAGMENTATION CODE SYSTEM.*	FID-67-3C3
E-ART OF THE SCIENTIFIC AND TECHNICAL INFORMATION SYSTEM IN THE USSR.*	MECHANIZED RETRIEVAL SYSTEM FOR ORGANIC COMPOUNDS. AN EVALUATION OF THE FRAGMENTATION CODE SYSTEM.*	FID-67-3C4
CT TRANSFORM: A COMPUTER BASED MACHINE INDEXING SYSTEM WITH SIMPLE MACHINE TRANSLATION AND THE INTERN	MECHANIZED RETRIEVAL SYSTEM FOR ORGANIC COMPOUNDS. AN EVALUATION OF THE FRAGMENTATION CODE SYSTEM.*	FID-67-3D2
	MECHANIZED RETRIEVAL SYSTEM FOR ORGANIC COMPOUNDS. AN EVALUATION OF THE FRAGMENTATION CODE SYSTEM.*	FID-67-3C7
	MECHANIZED RETRIEVAL SYSTEM FOR ORGANIC COMPOUNDS. AN EVALUATION OF THE FRAGMENTATION CODE SYSTEM.*	FID-67-3F6
	MECHANIZED RETRIEVAL SYSTEM FOR ORGANIC COMPOUNDS. AN EVALUATION OF THE FRAGMENTATION CODE SYSTEM.*	FID-67-3A1
	MECHANIZED RETRIEVAL SYSTEM FOR ORGANIC COMPOUNDS. AN EVALUATION OF THE FRAGMENTATION CODE SYSTEM.*	FID-67-3D1

<圖 15> KWIC 索引

8.3 特定個人을 對象으로 하는 情報檢索서어비스

特定個人의 要求에 따라 特定主題에 관한 情報를 檢索하여 提供하는 方式으로는 速報的인 SDI (Selective Dissemination of Information)서어비스와 邇及的인 個別檢索(Q-A, Question-Answering) 서어비스가 있다.

SDI서어비스는 要求되는 情報를 미리登錄하여 둔 利用者에게 個別的으로, 要求에 合致되는 最新情報가 入手되는 대로 定期的(例: 每週一回)으로 檢索하여 提供하는 速報的인 서어비스로서, 한번 登錄만 해두면 要求되는 特定主題에 관한 最新情報가 自動적으로 迅速히 入手되어 便利하므로 근래에 널리 利用되기 시작하였으며, 일반적으로는 電子計算機에 의하여 處理되어 提供되는 경우가 많다.

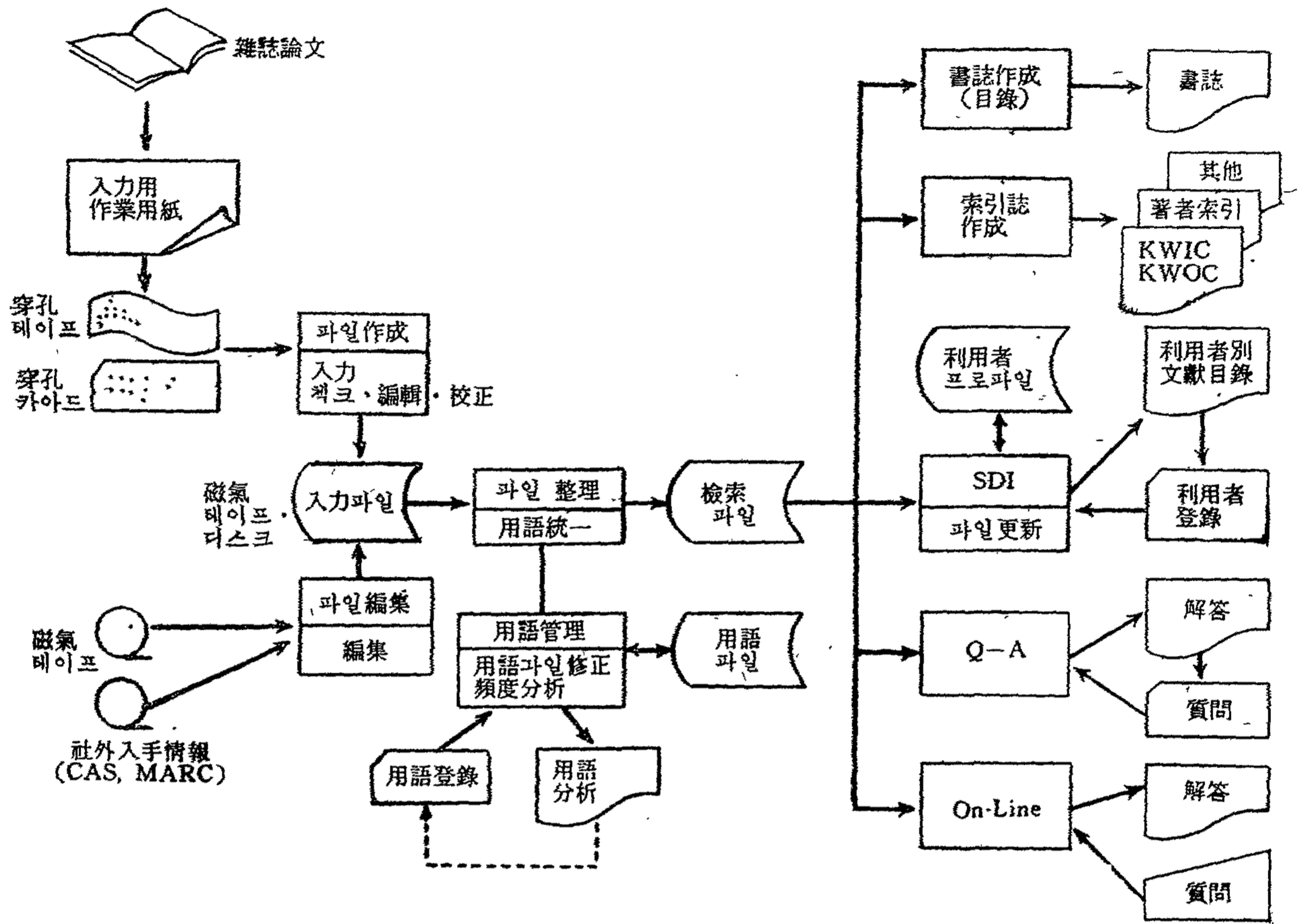
個別檢索(Q-A)서어비스는 特定個人의 要求에 따라 個別的으로, 一定期間 過去에 邇及하여, 蓄積된 情報과 일속에서 合目的情報를 檢索하여 提供하는 邇及的 檢索서어비스(RS)로서, 좁은 意味의 情報檢索서어비스는 이것을 말한다. 電子計算機를 利用하는 시스템에서는 一定期間마다 質問을 모아서 處理하는 Batch處理가 일

반적이며, 最近 質問과 同時에 問答式으로 即時 處理되어 回答을 얻을 수 있는 On-Line檢索시스템도 實用化되고 있다.

利用對象	不特定多數人	特定個人
速報的	速報的 서어비스	邇及的 檢索提供
性 格	能動的	受動的
서어비스對象 (利用者)	不特定多數人	特定個人
서어비스期間	長期繼續的, 定期的	單發的, 不定期的
파일 內容	短期間의 最新情報(週~月)	長期間의 蓄積情報(半年~數年)
파일 整備 (追加更新)	不必要	必要
檢 索	Batch 處理, 再檢 索 없음	Batch 處理 On-Line檢索 再檢索 있음
質問의 整備	必要	不必要(再檢索時 必要)

9. 맺는 말

以上으로 情報檢索技術 전반에 걸쳐 概略的으로 記述하였다.



<圖 16> 電子計算機를 利用한 綜合的인 情報시스템

前述한 바와 같이 情報檢索은 當初부터 機械化를 前提로 하였고, 궁극적으로는 電算化에 의한 高度로 自動化된 시스템을 目標로 하고 있다. 그리고 그것은 書誌目錄의 自動作成과 索引誌 編輯 및 組版의 自動化, SDI 서비스, 個別檢索(Q-A)서비스, On-Line檢索서비스 등을 包含하는 綜合的인 시스템이어야 한다. 이와 같은 시스템은 先進諸國에서는 이미 實用化되고 있다. 따라서 우리도 이와 같은 目標아래, 우선 實情에 맞는 簡單한 시스템부터 早速히 導入하여 活用하고, 궁극적으로는 電算化된 시스템의 導入 活用을 위하여 技術的인 檢討를 早速히 開始할 必要가 있다.

< 參考文獻 >

1. 高橋達郎 外. 情報檢索의 手法와 機器. (株)南江堂, 東京.

pp.3~16, 1969.

2. 河野德吉. 情報檢索의 知識. 日本經濟新聞社, 東京. p. 199. 1968.  
 3. 中井 浩 外. 情報檢索시스템. (株)日本出版經營會, 東京. p. 221. 1971.  
 4. 橋本昌幸. 情報檢索의 ABC. 日本放送出版協會, 東京. p. 187. 1971.  
 5. 櫻井宣隆. 情報管理入門(5), (完結). 마이크로寫眞—技術と應用(追補版). (社)日本マイクロ寫眞協會, 東京. pp. 20~39. 1971.  
 6. (社)日本能率協會. 技術情報管理의 手引き. (社)日本能率協會, 東京. pp. 17~38. 1971.  
 7. 粟田 洵. SDI와 其及檢索(IR講座 第18回). 情報管理. 15(1) pp. 46~56. 1972.