

索引作業의 數學的 모델

中井 浩
高 亨 坤 訳
〈KORSTIC 資料管理室〉

1. 頭腦勞動과 技術

事象의 하나 하나를 어떤 觀點에서 記号를 附與하여, 같은 種類의 것을 하나로 모으거나 서로 다른 것을 區別하는 것은 人間이 所有한 能力 가운데 가장 基本的인 것으로서 이것은 人間이 “事物을 생각한다”는 일 중에서 극히 基本이 되는 過程을 意味하므로 人間이, 感覺器官이 受容한 表象을 概念으로 整理하는데 있어 “일” 하는 過程이라고 看做하기 때문이다. 分散되어 脈絡도 없이 받아들여지는 表象을 어떠한 觀念이나 概念으로 모으는 過程은 認識論으로서 永久한 時間哲學의 対象이 되며 하나의 課題로서 아직도 繼續 研究되고 있다.

이것은 人間의 頭腦勞動의 一部이며 그 過程의 研究는 過去의 思辯的 哲學에서 脫皮하여 幼兒意識의, 形成過程의 分析에 의한 實証的, 研究로 轉移되어 哲學에서 科學의 分野로 變遷해가고 있다.

그러나 아마도 科學的 立場에서는 이 過程이 解明되지 못할 것이며, 또한 解明되어서도 안될 性質의 것인지도 모른다. 우리들이 所有한 能力의 가장 基本的인 것으로 變遷하였을 때 人間은 人間에 대해 이제는 人間이라고 할 수 없는 것이 될지도 모르기 때문이다.

그러나 情報라는 것에 關係하고 있는 사람들은 이 過程을 스스로 實踐해야만 한다. 人間의

精神的 勞動의 產物인 論文, 뉴스, 데이터 등의 各種의 記錄된 情報를 分類하고 다른 것과 區別하는 어떠한 指標를 附與하여 파일에 収録하는 作業은 우리들의 頭腦 속에서 행하여지고 있는 것에 意識的인 再現을 試圖하고자 하는 것이다. 그러나 우리들은 그것이 우리들의 頭腦 속에서 어떻게 그리고 왜 可能해지는가를 모른다. 原理를 모른채 實用目的을 充足시키기 위해 그것에 近似하다고 생각되는 것을 행하고 있음에 不過하다.

太古의 人間이 끈적끈적한 粘土를 이겨서 形態를 만들어 불에 구어서 자기가 생각하는 形態의 단단한 돌이 된다는 것을 알았을 때, 크나큰 놀라움이 있었을 것이다. 또 돌덩어리를 불에 태우면 뜨거운 液体가 되고 그 속의 異物質을 除去하며는 번쩍번쩍 빛나는 金屬이 얻어질 수 있다는 것을 알았을 때 神만이 할 수 있는 재주를 부렸다고 생각했을 것이다. 그 理由는 몰라도 神만이 할 수 있는 일을 자기 손으로 해냈다고 생각했을 때, 그것은 太古的 사람에게는 畏怖이었다. 技術에 대해서 太古的 사람은 敬虔하였다. 失敗가 없이 神의 재주를 再現할 수 있도록, 神에게의 祈願이 呪文이 되고 呪文에 리듬이 붙어 施律을 낳고 詩가 되고 노래가 되고 技芸가 되어 思想을 낳은 것이다.

過去의 技術은 精神의 世界와는 不可分이었다. 그러나 近世에 와서 技術이 精神의 世界에서 離脫하고 獨走하는 것은 技術의 墮落이며 惡魔와

의 契約이었다. 이제 技術은 精神世界와의 連結을 되찾지 않으면 안된다.

하나의 論文을 分析한다. 키워드를 붙인다. 分類를 한다. 이것은 그 論文의 內容을 理解하는 過程의 하나이다. 人間은 하나의 것을 보는데 있어 여러 가지 觀點이 있고 많은 立場이 있다. 그리고 많은 사람이 많은 意見을 갖으므로 社會는 偏在하지도 않고, 暴走하지도 않는다. 많은 자유스러운 意見의 主張과 交換 가운데서 行政이나 事業이 遂行되는 社會나 組織은 健全하다. 그러나 그것이 制限되고 많은 사람이 同一한 思考方式으로 強制되었을 때 管理社會를 낳고 社會는 不健全하게 된다.

데이터·베이스産業이 産業으로서 成立하기 위해서는 抄録, 分類, 索引作業(Indexing)은 一定基準下에서 정해진 品質을 갖어야 하며 또 産業으로서 成立하기 위해서는 大量處理가 可能해야 하며 大量의 頭腦勞動力이 動員되어야 한다. 즉 大量의 畫一的인 頭腦勞動을 必要로 한다.

이러한 일은 情報産業이 成立하기 위해서는 非人間的인 管理社會를 前提해야만 함을 意味하는 것인지도 모르겠다.

太古의 人間이 陶工의 技術, 鍛冶師의 技術을 보았을 때 그것이 왜 可能한지를 알지 못했다. 그저 “이렇게 하면 얻을 수 있다”라는 정해진 方法만을 알았던 것이다. 神의 재주의 本質과는 일단 無關한 人間側의 方法을 發生시킨 것이다. 그 方法은 附與된 條件下에서 정해진 目的을 끝내는 것이었다.

어느 前提에서 附與된 條件下에 어떤 目的을 끝내는 方法은 우선 個人中에서 發生한다. 그것은 知識이고 Know how이다. 그 知識은 같은 方法을 模倣하면서 習得함으로써 他人에게 傳達될 수 있다. 그것이 行動에서 分離되어 情報化되고 記号에 의해 傳達可能하게 되었을 때 人間에게서 疎外(entfremden)가 可能하게 된다. 즉 機械化가 可能하게 된다.

技術은 그 發生에 있어서 때때로 自然을 模倣한다. 그 模倣을 위한 必須條件으로 그 自然을 움직이고 있는 法則을 알 必要는 없다. 또 알았다 해도 그것을 그대로 本받을 必要는 없다. 技術은 하나의 black box이다. 같은 前提에서 같

은 結果를 얻는 것이라면 그중에서 使用하는 原理가 무엇이든 좋다. 自然보다도 附與된 條件下에서 보다 效率的이면 그것으로 족하다. 즉 技術은 自然에서 脱却할 수 있다.

分類나 抄録이나 索引作業의 技術도 아마 人間の 認識過程과 無關하지는 않을 것이다. 그러나 그 本質은 아직 모른다. 아마도 가까운 將來에 簡單히 알 수 있는 것은 아닐 것이다. 本質을 모르는채 우리들의 目的을 遂行할 方法을 찾아야만 한다. 人間の 精神活動은 本質적으로 自由로와야 한다. 그러나 우리들이 찾아내야 할 方法은 많은 사람에게 대해 畫一的이어야 하며 될 수 있으면 疎外可能하여야 할 것이다. 그럼으로써 情報産業의 成立이 可能해질 수 있다고 본다.

索引作業은 文献이 入力될 때에 索引을 出力하는 Black box이다. 本稿는 그중에서 作用하고 있는 어떤 하나의 包括的인 記述을 試圖한 B. C. Landry의 Theory of Indexing(오하이오 州立大學의 學位論文)을 紹介하는 것이다.

2. 索引作業의 対象

文献은 索引作業의 結果로서 索引이 附與된다. 이것은 그 文献을 파일에 収録한 뒤 다시 꺼내기 위해 붙여지는 指標이다.

最近의 파일의 理論, 즉 데이터 베이스 모델은 20世紀初頭의 哲學者 카씨러(Ernst Cassirer, 1874~1945)의 概念論에 依據한다. 新칸트派에 속하는 카씨러는 技術論에 基礎한 概念論을 提唱하였다. 하나의 것은 그것을 보는데 여러 개의 觀點(Gesicht Punkt)이 있고, 그 觀點에서 보는 사람이 附與하는 스코어에 의해 規定된다. 그 觀點으로서 그것이 갖는 技能을 設定하는 것이다.

데이터·베이스 모델에 있어서 이 觀點을 “屬性”이라 呼稱하고 各 屬性에 그것을 具體적으로 表現하기 위한 屬性值集合이 準備된다. 하나의 것을 다른 것과 區別하기 위하여 各 屬性에 대해서 그것을 特徵지워 주기 위한 屬性值를 附與한다. 지금 屬性 1, 屬性 2, 屬性 3……의 屬性值集合을 各已 A_1, A_2, A_3, \dots 라 하며는 그것으로 인해 表現可能한 全集合은, 屬性의 種類

수를 n 이라 할 때,

$$A_1 \times A_2 \times A_3 \times \dots \times A_n$$

이다. 즉 属性値의 n 項關係(n -tuple relation)이다. 그리고 그 하나에 대한 파일 내의 표현, 즉 record r 은 그 關係에 屬하여

$$r \in A_1 \times A_2 \times A_3 \times \dots \times A_n$$

파일은 그 record의 集合이다.

우리들도 索引作業의 理論을 構成함에 있어서 그 対象 즉 “文献”을 보는데는 그 概念論에 依據한다.

定義 1 対象 : 属性値測定의 対象은 觀測可能量의 集合이다.

定義 2 属性 : 属性은 対象에 대하여 모든 觀測可能量中에서 다른 것과 區別하는데 必要한 部分集合이다.

定義 3 属性値測定 : 属性値測定이란 어느 属性에 대하여 可能한 属性値의 集合中에서 가장 바르게 対象을 特定하는 것을 選定하는 일이다.

以上과 같이 지금부터 우리들이 하고자 하는 対象은 모두 属性의 集合으로서 보는 것이 된다. 즉 対象은 觀測可能한 量의 集合이고, 그 가운데서 다른 것과 區別하는데 必要한 것에 의해서 特定지워진다. 그것이 属性이다.

지금 索引作業의 過程을 생각하면 入力되는 것은 文献이다. 文献은 觀點에 의해 各種의 形象이 취해지나 어느 觀點에 의하면 文字와 스페이스와 句讀點의 配列이다(数学記号도 文字의 一種이라고 한다). 또 어떤 觀點에 의하면 單語의 配列이다. 어느 觀點에 의하면 標題, 著者名, 章이나 節의 標題, 그리고 節의 配列이다. 또는 標題, 著者名, 章의 標題, 그리고 節의 配列이다. 또한 論文의 集合이다. 또는 圖書의 集合이다. 등등 많은 레벨로서의 觀點이 있을 수 있다.

여기서 “單語”라는 対象을 취하면 이것은 스페이스로 구분된 文字의 配列이다. 이 때의 属性은 文字의 配列이다. 어느 文字의 配列이 “單語”이기 위해서는 그것이 스페이스로 区分되어 있는 것이 條件이다. 즉 属性値測定은 “공이 아닌 2개의 스페이스” 사이에서 행하여 지고 이것을 属性値測定의 單位라 呼稱한다. 그 属性値測定은 스페이스나 句讀點이나 文字를 하나 하나

읽으면서 행하여진다. 그러므로 그 精密度는 “하나의 文字에 대한 可讀의 與否”이다. 또는 機械可讀파일에 있어서 “文字”는 4, 또는 6, 또는 8, 또는 16비트의 配列이다. 標題란 文献의 始作에서 몇 行인가의 스페이스 다음에 著者名까지 사이의 몇 行인가의 스페이스로 둘러싸인 單語의 配列이다. 標題의 属性値測定은 그 몇 行인가의 스페이스 사이에서 單語의 하나 하나를 읽으면서 행하여 진다. 즉

定義 4 測定의 單位 : 測定이란 測定対象에서 2개를 選定하여 그것에 대하여 어느 實數를 対応시키는 것이다.

$$\rho : A \times A \rightarrow R$$

R : 實數의 集合

그리고 $a, b \in A$ 라 할 때는 $\rho(a, b)$ 는 距離關係(metric)를 充足시킨다.

- (i) $\rho(a, b) = \rho(b, a)$
- (ii) $a=b$ 이면, $\rho(a, b) = 0$
- (iii) $\rho(a, b) + \rho(b, c) \geq \rho(a, c)$

定義 5 測定의 精度 : 附與된 $a, b \in A$ 에 대하여

$$\rho_1(a, b) < \rho_2(a, b)$$

일 경우에 ρ_1 쪽이 ρ_2 보다도 精密하다고 한다.

以上과 같이 対象은 属性과 그 測定의 單位와 精度를 附與함으로써 規定할 수가 있다. “1文字씩 읽어감으로써 (精度), 0가 아닌 文字의 길이를 가진 2개의 스페이스의 사이(單位)의 文字의 配列”이 單語의 属性値이다. 그래서

定義 6 데이터·엘리먼트(data element) : 指定된 測定의 單位와 精度에 의해 測定되는 属性에 의해 名命되는 것중에서 個個의 要素로서 識別되는 最小의 것을 데이터·엘리먼트라 한다.

3. 데이터·엘리먼트, 文献, 文献空間

定義 6에 의한 데이터·엘리먼트의 表現은 不分明해지기 쉽다. 데이터·엘리먼트는 本 理論 構成中에서 가장 基本的인 概念이므로 多小 相異한 角度에서 問題를 다루어 보고자 한다.

데이터·엘리먼트는, (1) 하나 하나를 獨立하여 取扱할 수 있으며, (2) 그 自体를 더욱 작은 單

1-1 專文
820 EFFECT OF A SELECTIVE BETA-ADRENERGIC BLOCKER IN PREVENTING FALLS IN ARTERIAL OXYGEN TENSION FOLLOWING ISOPRENALINE IN ASTHMATIC SUBJECTS.
Palmer K H V, Legge J S, Hamilton W F D, Diamant M L :
Dept Med Unit

6673354 4-(2-hydroxy-3-isopropylaminopropoxy)acetanilide (practolol) (20 mg subject, i.v.), a BETA-ADRENERGIC BLOCKING agent selective to the HEART prevented the decrease in ARTERIAL OXYGEN TENSION in 11 ASTHMATIC patients following

7683592 isoprenaline (0.1 mg/subject, aerosol inhalation) treatment without significantly decreasing the BRONCHODILATOR action of isoprenaline

1-6 s_d : Multi-term entries

- 820 Beta-blocking Drug
- 820 Myocardial Contractility
- 820 Bronchial muscle
- 820 Oxygen tension
- 820 Bronchodilator activity
- 820 Blood-gas tension
- 820 Bronchial Asthma

1-7 s_e : Articulated entries

- 820 beta-adrenergic blocker
 - effect of, in preventing falls following isoprenaline in asthmatic subjects
 - in preventing falls following isoprenaline in asthmatic subjects, effect of
- 820 asthmatic subjects,
 - effect of beta-adrenergic blocker in preventing falls in, following isoprenaline
 - following isoprenaline, effect of beta-adrenergic blocker in preventing falls in

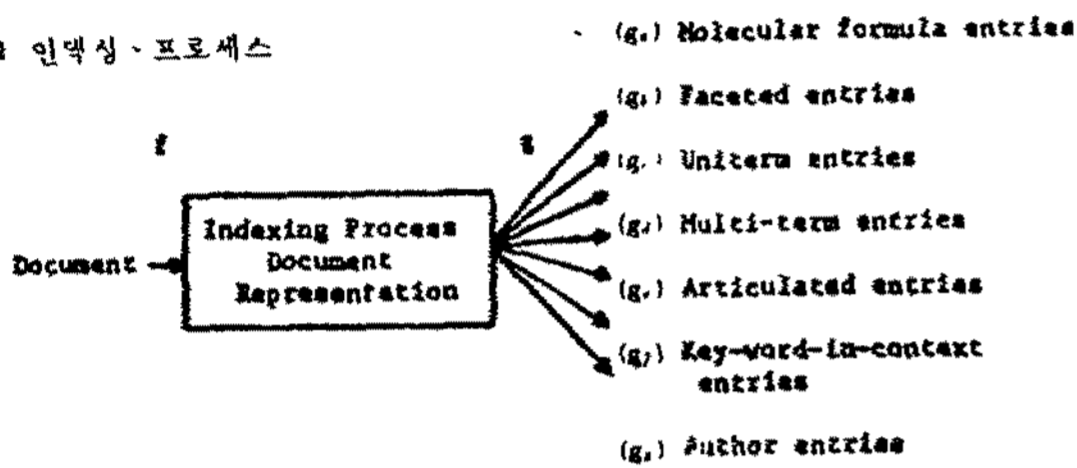
1-8 s_f : Key-word-in-context entries

4-(2-Hydroxy-3-isopropylaminopropoxy) acetanilide (practolol) (20 mg/subject, i.v.) OF A SELECTIVE BETA-ADRENERGIC BLOCKER IN PREVENTING FALLS IN ARTERIAL OXYGEN TENSION FOLLOWING ISOPRENALINE IN ASTHMATIC SUBJECTS. EFFECT OF A SELECTIVE BETA-ADRENERGIC BLOCKING agent selective to the HEART, prevented the decrease in ARTERIAL OXYGEN TENSION in 11 ASTHMATIC patients following ISOPRENALINE (0.1 mg/subject, aerosol inhalation) treatment without significantly decreasing the BRONCHODILATOR action of isoprenaline. 4-(2-HYDROXY-3-ISOPROPYLAMINOPROPOXY)ACETANILIDE (PRACTOLOL) (20 mg/subject, i.v.), a BETA-ADRENERGIC BLOCKER IN PREVENTING FALLS IN ARTERIAL OXYGEN TENSION FOLLOWING ISOPRENALINE IN ASTHMATIC SUBJECTS.

1-9 s_g : Author entries

- Diamant ML, 820
- Hamilton WFD, 820
- Legge JS, 820
- Palmer KHV, 820

1-2 인덱싱 - 프로세스



1-3 s_a : Molecular formula entries

- 820 $C_{14}H_{22}N_2O_3$ 6673354
- 820 $C_{11}H_{17}NO_3$ 7683592

1-4 s_b : Faceted entries

- 820 Receptor, beta(1)
- 820 Receptor, beta(2)
- 820 Drug, beta-blocking
- 820 Muscle, Bronchial
- 820 Asthma, Bronchial

1-5 s_c : Uniterm entries

- 820 Isoprenaline
- 820 Receptors
- 820 Drug
- 820 Myocardium
- 820 Myocardial
- 820 Contractility
- 820 Oxygen
- 820 Tension

그림 1.

位로 区分하지 않고, (3) 確定된 意味나 解釈을 갖은 것이다.

지금 하나의 文献과 거기에서 만들어지는 몇 개의 索引의 例를 그림 1에 提示한다.

索引作業의 過程은 現象的으로 보면 하나의 文献에 많은 代替的 시스템이 存在할 수 있다. 그러나 이것을 包括的으로 記述하면 다음과 같은 抽象 automaton으로 表現할 수가 있다.

$$M = (K, \Gamma, \Sigma, \delta, \sigma, F, q)$$

K : automaton이 취해질 수 있는 狀態의 有限集合.

Γ : 入力를 構成하는 記号의 有限集合.

Σ : 出力을 構成하는 記号의 有限集合.

δ : 어느 狀態의 入力記号를 읽었을 때 automaton이 취하는 다음 狀態를 決定하는

關係数.

$$\delta : K \times \Gamma \rightarrow K.$$

σ : 어느 狀態에서 어느 記号를 읽었을 때 出力하는 出力記号를 決定하는 關係数.

$$\sigma : K \times \Gamma \rightarrow \Sigma$$

F : 最終狀態 $F \subseteq K$.

q : 開始狀態 $q \in K$.

Γ 과 Σ 는 同一國語의 경우 $\Gamma = \Sigma$ 이다. 그러나 日本人의 경우 Γ 는 많은 國語를 包含하고, Σ 는 通常 日本語이므로 $\Sigma \subseteq \Gamma$ 이다.

이 記述에서

觀測可能量 $a \in \gamma$ 또는 $a \in \Sigma$.

屬性 $A \subseteq \Gamma$ 또는 $A \subseteq \Sigma$.

이다. 그리고 入力文献을 構成하는 데이터·엘리먼트는 Γ 의 部分集合이고, 出力되는 索引을

構成하는 데이터·엘리먼트는 Σ 의 部分集合이다. 그리고 后述하는 바와 같이 索引作業은 入力 데이터·엘리먼트로 出力 데이터·엘리먼트로 變換하는 過程이다.

文獻은 어느 觀點에서 보면 文字의 全順序集合이다.

定義 7 文獻 : 文獻은 데이터·엘리먼트의 全順序集合이다.

그리고 파일에 収録되는 文獻의 集合은 一般的으로 어떠한 順序가 附與된 集合이다. 예를 들자면 受入窓口的 到着順序, 혹은 任意로 附與된 ID番号, 혹은 資料의 種類別로 附與한 一連番号 등이다. 그러므로

定義 8 文獻空間(document space) : 文獻空間은 文獻의 順序集合이다.

4. 커뮤니케이션

索引作業은 将来 그 索引作業이 附與한 索引에 의해 그 情報가 檢索되어 그것에 의해 情報가 傳達되는 것을 目的으로 행하여 진다. 즉 情報의 蓄積과 檢索시스템은 情報源에서 發生하는 情報가 各種의 過程을 經過함으로써 窮極적으로 受信체에 傳達됨으로써 完結된다. 이것은 커뮤니케이션의 一種인데 커뮤니케이션은 다음과 같이 定義된다.

定義 9 커뮤니케이션 : 커뮤니케이션은 情報源과 受信체를 通信經路와 피드·백經路로써 成立시킨다.

이 情報의 流通中에서 傳達되는 것은 데이터·엘리먼트이다.

定義 10 伝送速度 : 伝送速度는 單位時間當 伝送되는 데이터·엘리먼트의 數로 表示된다.

情報蓄積과 檢索시스템에서는 正確하고 效果的인 커뮤니케이션이 窮極적인 目標이다. 情報源과 受信체에 있어서 效果的인 커뮤니케이션이 행해지기 위해서는 서로 알 수 있는 것이어야 한다. 이를 다음과 같이 表現한다.

定義 11 經驗集合 : 情報源 또는 受信체의 記憶構造는 데이터·엘리먼트와 그 關係의 順序集合으로서 모델화 된다. 이것을 經驗集合이라 하며 ES로 表記한다.

이 定義와 定義 7에서 다음의 定理가 導入된다.

定理 : 經驗集合은 文獻이다.

定義 12 效果的 커뮤니케이션 : 情報源의 經驗集合(ES)_s와 受信体の 經驗集合(ES)_r의 交換이 空集合이 아닐 때 커뮤니케이션은 效果的이라 한다.

이 定義에서 다음의 定理가 導入된다.

定理 : 效果的 커뮤니케이션은

$$(ES)_n = (ES)_r$$

일 때 最大가 된다.

이 情報源과 受信体 사이에서 행해지는 커뮤니케이션에 있어서 使用되는 데이터·엘리먼트와 그 사이의 關係는 그 兩者間에 interface를 만든다.

定義 13 interface經驗集合 : 情報源과 受信体の 사이에서 커뮤니케이션에 實際로 使用되는 데이터·엘리먼트와 그 사이의 關係의 集合을 interface經驗集合이라 하며, (IES)라 表記한다.

즉 情報源中の 受信体에 傳達하고자 하는 데이터·엘리먼트와 그것들 사이의 關係는 受信体에 傳達될 수 있는 形態로 變換되어야 한다. 그 變換을 뒷받침하는 中間媒体를 interface經驗集合이라 呼稱한다. 예를 들면 機械와 人間間의 커뮤니케이션에 있어서 interface經驗集合은 프로그래밍 言語이다.

定義 14 經驗集合變換 : 情報源과 受信体間의 커뮤니케이션은 (ES)_s에서 (IES)로의 變換과 (IES)에서 (ES)_r로의 變換을 통해서 행하여 진다.

$$(ES)_s \xrightarrow{S} (IES) \xleftarrow{R} (ES)_r$$

이 定義에서 다음의 定理가 얻어진다.

定理 : 情報蓄積과 檢索시스템의 利用者는 그 시스템과의 사이에 效果的인 커뮤니케이션을 행하기 위해서는 시스템의 데이터·엘리먼트의 構成과 表現에 대한 知識을 갖고 있어야 한다. 이것은 시스템을 情報源으로 하고 利用者를 受信体로 할 때에 利用者の 시스템에 대한 知識은 interface의 經驗集合을 構成함으로써 明確해 진다.

Text:

Effect of a selective beta-adrenergic blocker in preventing falls in arterial oxygen tension following isoprenaline in asthmatic subjects.

(Input)	(System Representation)	D_a
Data Element	Data Element	
Effect	R_{19}	Role
beta-adrenergic adrenergic blocker	beta-receptor adrenergic-receptor blocking-drug	BT USE BT
prevention	drug	BT
falls reduction	R_6	Role
arterial arterial oxygen arterial oxygen tension following after	reduction R_3	USE Role
isoprenaline	cardiovascular system	BT
asthmatic	arterial oxygenation airway resistance after R_{30}	USE RT USE Role
	$C_{14}H_{22}N_2O_3$	FORMULA
	asthma	USE

그림 2.

서, 또는 그 사이에서 데이터·엘리먼트 관계를 찾아나간다. 타임·슬라이스는 2개의 文献에 걸쳐 있을 수도 있다. 즉 索引作業시스템은 하나의 文献, 또는 文献間의 데이터·엘리먼트 관계를 찾아낸다.

索引作業시스템은 文献의 흐름을 入力하여 그 索引을 出力하는 시스템이다. 그러므로 커뮤니케이션·프로세스 중에서는 伝送経路는 受信体 사이에 位置하게 된다.

索引作業시스템이 作用하는 文献에는 2개의 種類가 있는데 하나는 入力文献이고, 다른 하나는 分析用 文献이다.

定義 26 入力文献: 入力文献 D_i 는 伝送経路를 經過하여 索引作業시스템에 入力되어 處理를 받는 文献이다.

定義 27 分析用文献: 分析用 文献 D_a 는 恒常 索引作業시스템 중에 常駐하여 變換 S를 記述하는 文献이다. 이 2개의 定義에서

定理: 文献 $D = D_i \cup D_a$

이 D_a 는 디스오러스이거나 分類表이거나 各種의 典據파일일 때도 있다. 그러나 이 役割은 D_i 와 複合하여 入力文献의 데이터·엘리먼트에 關係를 附與한다. 이것을 다음과 같이 表現한다.

定義 28 入力文献表現은 $D_i \otimes D_a = (D_i) \cdot R-EL$

이의 具體的인 例는 그림 2를 참고하면 된다. 여기에서 D_a 는 롤, 인디케이터(T), 類種關係(G), 化學式表現(E)에서 成立되어 있다. 이것을 使用하여 索引이 定義된다.

定義 29 索引: 索引은 入力文献에 대해 움직이는 關係決定關係數의 像이다.

$$I = g(D_i \otimes D_a)$$

定義 30 索引의 更新: 이미 얻어진 索引 I_0 를 새로 入力된 文献 D_a 에 의해 更新하는 過程은 다음과 같이 定義된다.

$$I_n = g(D_i \otimes D_a) \cup I_0$$

6. 效 率

여기서의 檢索過程을 說明하면 檢索質問은 데이터·엘리먼트의 順序集合이다.

定義 31 檢索質問: 檢索質問은 데이터·엘리먼트의 順序集合이며,

$$Q \subset I$$

이다.

여기서 다음의 定理가 얻어진다.

定理: Q는 文献이다.

그러면 마지막으로 索引作業을 ダイナ믹하게 把握하는 것을 試圖해보자. 우선 索引作業이 엔트로피(entropy) 減少의 過程임을 提示한다.

入力文献 D_i 는 그 自体가 完全히 無秩序하다. 文字와 스페이스와 句讀記호의 配列에 不過하여 데이터·엘리먼트도 그들 사이의 關係를 明示하고 있지 않다. 索引作業을 經過하고 文献은 그것들 사이의 關係까지 包含하여 데이터·엘리먼트의 集合으로서 表現된다. 一般的으로 엔트로피는 対象의 構造에 관한 情報의 缺如로 보며 그것이 갖는 秩序의 未知의 尺度를 表示한다. 즉 無秩序할수록 엔트로피의 값은 크다.

여기에서 새로운 空間概念을 定義한다.

定義 32 位相空間(phase space): 位相空間은 文献空間中の 座標空間이며 2개의 座標系를 갖는다.

(1) 位置空間(Configuration space): 位置空間은 데이터·엘리먼트의 길이의 空間에서 文献은 그 중의 1개의 벡터(vector)로서 表現된다.

(2) 運動量空間(momentum space): 運動量空

間은 檢索오퍼레이션의 길이의 空間에서 1회의 探索에 어느 만큼의 位置空間內的 點을 經過하느냐를 表示한다.

이 位相空間은 索引空間과 同型이다. 이 位相空間에 位置를 정함에 있어 文獻空間은 秩序化되어 엔트로피는 減少한다.

定理: 關係決定 其他에 의해 索引作業시스템은 (文獻-空間)/(文獻-空間-探索)을 構成하는 시스템 全體의 엔트로피를 減少시킨다.

커뮤니케이션이 閉鎖된 體系이면 그중에서 에너지는 保存되어 全體로 볼 때 엔트로피도 不變한다. 고로

定理: 索引作業에 의해 減少한 엔트로피는 位相空間中에서 데이터·엘리먼트의 座標를 구하는 努力에 의해 相殺된다.

다음에는 情報의 實際的인 定義를 導入해 보자.

定義 33 情報: 情報란 決定過程에 價値를 갖는 데이터·엘리먼트이다.

그리고 情報를 求하는 사람의 行動을 다음과 같이 假定한다.

- (1) 假定을 세워 그것을 테스트한다(H).
- (2) 決定을 행하고 結果를 얻는다(D).
- (3) 目的을 達成한다(G).

여기에서 利益의 概念을 導入한다.

定義 34 利益: 目的達成을 위하여 행한 決定의 數와 얻어진 情報의 關係로 正해진다. 이것에 의해 다음의 概念이 定義된다.

定義 35 最大利益의 行動: H-D-G(즉 1회의 假定-檢証과 1회의 決定에 의해 目的을 達成한다)의 行動經路를 最大利益의 經路라 한다.

定義 36 最小利益의 行動: H-H-H.....(즉 假定-檢証만의 反復으로 決定過程에도 到達하지 못하는 行動經路)를 最小利益의 行動이라 한다.

索引作業이거나 檢索이거나 옳은 答을 얻기 위해서는 H와 D와를 몇번이든 反復할 必要가 있다. 이 試行錯誤의 行動經路를 最端으로 하기

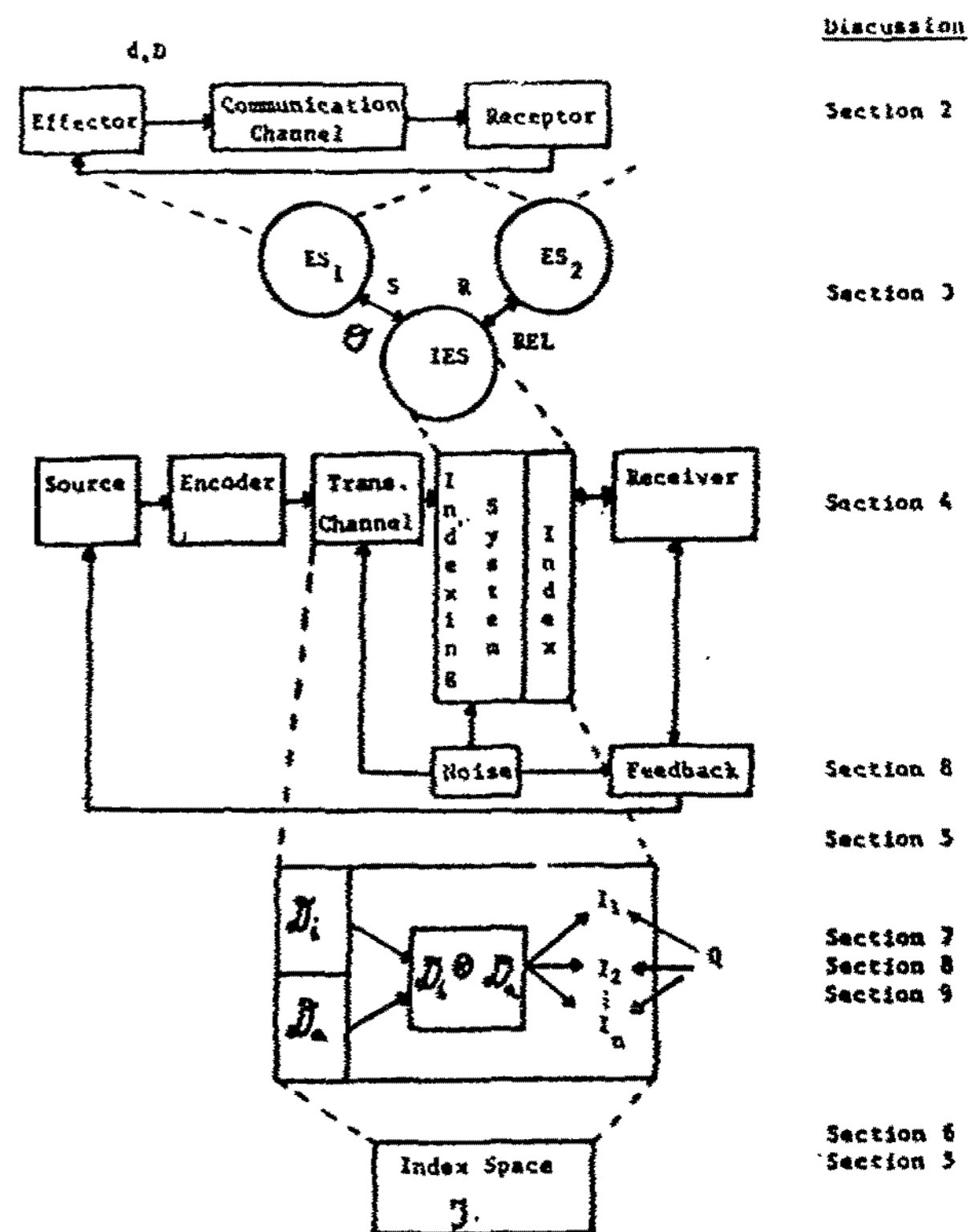


그림 3.

위해서는 시스템과 人間과의 사이에 共通經驗集合(IES)의 整備가 必要하다. 이것이 支援시스템, 利用者 援助, 索引者 援助 등으로 呼稱되는 것이다.

索引作業의 全般的인 概要는 그림 3에 提示하였다.

7. 結 言

以上으로 索引作業過程에 대해 數學的 記述을 하였지만 많은 點에서 精密化를 必要로 한다. 특히 索引作業에 따르기 마련인 에러, 個人差 등에 대해 品質管理를 어떻게 할 것인가, 또한 그것은 어떻게 檢索하는 쪽에서 補完할 것인가, 그리고 그것에 대해 支援시스템이 어떻게 機能할 것인가 등의 理論構成이 앞으로의 重要한 課題이다.