

# CA Condensates의 SDI 서비스를 위한 프로파일作成法

関 永 一\*

## 1. 序 論

急増하는 國內 情報需要에 效果的으로 對處하기 위하여 韓國科學技術情報센터(KORSTIC)에서는 化學分野의 著名한 既存 데이터 베이스인 CA Condensates(CAC)와 그 檢索프로그램을 導入하여 컴퓨터에 의한 最新情報檢索서비스(S.D.I.)를 今年 7月 1日부터 本格的으로 開始하게 되었다.

이 서비스를 利用하면 研究開發活動에 必要한 最新 技術情報를 迅速正確하게, 網羅적으로 簡便하게 入手할 수 있어 先進諸國에서도 이 서비스의 利用者가 급격히 增加하고 있다. KORSTIC에서는 앞으로 데이터 베이스를 擴張하여 化學分野뿐만 아니라 科學技術 全分野에 대하여 이 서비스를 實施할 計劃이다.

이 서비스를 利用하려면 利用者마다 자기가 願하는 主題를 컴퓨터가 취급할 수 있는 形態(Profile)로 만들어서 登錄해 두어야 한다. 따라서 本稿에서는 利用자가 CAC 데이터 베이스를 理解하는 것은 물론 프로파일의 作成方法(Profiling)을 習得할 수 있도록 마련하였다. 그러나 프로파일 作成에 必要한 모든 知識을 다 記述한다는 것은 그리 쉬운 일이 아니다. 왜냐하면 紙面관계도 있겠지만 특히 각 개인의 情報要求에 따라 變化하는 因子의 다양성 때문이다. 그러나 本稿에서는 그동안 筆者의 經驗을 토대로 하여 利用者들로 하여금 적은 努力으로 效率적인 프로파일을 作成할 수 있도록 하였다. 되도록 많은 利用者들이 이를 익혀서 研究開發活動에 큰 도움이 되기를 바란다.

## 2. CAC 데이터 베이스

### 2.1 範 圍

CA (Chemical Abstracts)는 世界 各國에서 刊行하고 있는 化學, 化學工學 및 이와 關聯있는 모든 分野의 學術雜誌, 特許, 技術報告書, 政府 刊行物, 單行本 등의 抄錄이 연간 약 40만건이나 收錄되는 世界 最大의 抄錄誌로 週 1回 冊子 形으로 出版되고 있다.

CAC는 CA와 같이 每週 磁氣테이프 形態로 作成되는 것이며, 收錄事項도 同一하나, 다만 抄錄대신 일련의 重要語(Keyword)로 바꾸어 놓은 것만이 다르다.

### 2.2 데이터 베이스 코오드

CAC프로파일에는 다음과 같은 코오드가 쓰인다.

- C1 홀수週 발행분(有機化學/生化學)
- C2 짝수週 발행분(高分子, 物理, 分析과 應用化學, 化學工學)
- C1 C2 홀수와 짝수週 발행분

### 2.3 檢索分野의 選擇

CAC의 檢索領域과 프로파일 檢索項目 카아드에 使用할 檢索領域 코오드는 다음과 같다.

領 域	檢索코오드
著 者	A
作業場所	W <sup>+</sup>
標 題	H
發 行 日	D <sup>+</sup>
發行物 分類코오드	P <sup>+</sup>
CODEN	C <sup>+</sup>

\*KORSTIC 電子計算室

CA 細項 / 細目	X
言 語	L+
重 要 語	K+
標 題 및 重 要 語	T

+표를 한 코오드는 사용하는 일이 극히 드물다.

### 2.4 檢索領域의 構造

#### 1) 著 者

한개의 레코오드에 10명까지의 著者와 共著者를 포함시킬 수 있다. 人名은 姓을 먼저 쓰고 이름은 첫자만 따서 나중에 쓴다. 프로파일은 著者檢索에 있어서 너무 세밀하지 않게 하는 것이 좋다. 예를 들면,

SMITH, D \*

라고 하는 것이 가장 안전한 方法이다. (\*표는 用語줄임을 나타냄)

#### 2) 文獻標題

文獻標題는 英語가 아닌 경우 英語로 번역하여 편집한다. 이 領域과 重要語 領域에서는 美國式 철자법을 사용한다(例 Sulfur). 略號, 化學元素記號 및 化學式은 一般的으로 全名稱으로 확장 사용한다.

#### 3) CA細項 / 細目

CAC에 포함된 文獻은 CA와 마찬가지로 80개 細項(Section)으로 分類되어 있다. 各 文獻은 이들 중 어느 細項 하나에 들어가 있으며 이를 더 細分하여 細目に 分類되어 있다. 細項과 細目は 다음과 같이 수록되어 있다. CAmmnnn 여기서 mmm은 細項을 세자리로 보아 001-080 중의 하나이고, nnn은 細目인데 각 細項別로 그 內容이 다르다. 예를 들면, 合成有機藥品에 關連된 文獻은 細項64와 細目3이기 때문에 CA064003으로 表示되어 있다. \*L파라미터는 檢索을 특정 細項으로 제한할 때 써야 하지만 細項 / 細目 領域은 檢索코오드 X를 쓴 표준파라미터를 쓸 수 있다.

예를 들면 合成有機藥品에 關連되는 記事를 찾아내기 위한 아주 좁은 檢索인 경우 다음과 같이 쓴다.

(A) \*L      C 2      64  
 \*PX                      CA064003

또는 간단히

\* (B) \*PX                      CA064003

이라고 쓰면 된다. 그러나 (A)가 바람직하다. 다른 方法으로도 찾을 길이 많지만 可能한 한 L 파라미터를 사용하여 찾는 길이 확실하다고 할 수 있다.

#### 4) 重要語

이 重要語의 領域에는 숫자적으로 制限은 없다. 重要語는 한개의 單語일 수도 있고 두개 이상의 單語일 수도 있다. 이 重要語들은 標題, 抄錄 및 原文에서 選擇된 것도 있고 追加하는 수도 있다. 重要語에 쓰이는 各種 略號는 美國 化學會가 認定한 것에 限한다. 이러한 낱말은 디소오러스(Thesaurus)나 사전에서 선택한 것이 아니기 때문에 利用者는 스스로의 利用經驗을 통하여 檢索項目을 판단하여야 한다.

#### 5) 標題와 重要語

T檢索이란 檢索項目으로 標題와 重要語群에서 찾도록 되어 있다.

### 3. 프로파일의 構造

CAC프로파일의 예를 表 1에서 紹介하였다. 프로파일을 作成하기 전에 利用者는 프로파일의 構造와 檢索項目 사이에 適用되는 論理에 대해서 充分히 알아야 한다.

#### 3.1 制御카드와 檢索項目카드

프로파일은 7種의 制御카드(\*R, \*N, \*D, \*M, \*T, \*E, \*Z)와 2種의 檢索項目 카드(\*L, \*P)로 되어 있는데, 모든 制御 카드는 반드시 순서있게 나열해야 한다.

##### 1) \*R카드

欄 1-2	*R
3-4	版數
5-12	利用者 코오드
13-22	데이터 베이스의 種類
23	프로파일의 形式
24	出力部數
25	프로파일 狀態
41	프로파일 作成者의 略字

##### 2) \*N카드(3장까지 쓸 수 있음)

欄 1-2 \*N(첫장에 한함)

- 5-76 利用者の住所, 姓名
- 3) \*D카아드(3장까지 쓸 수 있음)
  - 欄 1-2 \*D (첫장에 한함)
  - 5-76 要求主題에 관한 說明
- 4) \*M카아드
  - 欄 1-2 \*M
  - 5-8 檢索文獻의 最大 프린트 部數
- 5) \*T카아드
  - 欄 1-2 \*T
  - 5-8 基準加重值
- 6) \*E카아드
  - 欄 1-2 \*E (프로파일의 끝을 나타냄)
- 7) \*Z카아드
  - 欄 1-2 \*Z (프로그램의 끝을 나타냄)
- 8) \*L카아드
  - 欄 1-2 \*L
  - 7-8 데이터 베이스의 코오드
  - 12-14 一般的으로 空欄, 그러나 특수한 경우 NOT를 記入함
  - 16-75 데이터 베이스에서 細分한 項目의 코오드

\*L카아드를 한장 이상 쓸 수 있다.  
\*L표시는 첫장에만 쓰되 각 카아드의 7-8欄에 데이터 베이스 코오드는 반드시 記入하여야 한다.
- 9) \*P카아드
  - 欄 1-2 \*P
  - 4 檢索分野 코오드
  - 5-8 純論理 프로파일인 경우엔 빈칸, 加重 프로파일인 경우 項目의 加重值
  - 12-14 NOT論理項目인 경우 NOT, 그밖의 경우 빈칸
  - 16-75 檢索項目
  - 21-75 IGNORE論理項目

3.2 프로파일 論理

表 1. 純論理 프로파일

0000	*R01CNAPH1	C1C2	IIP	GLJ
0010	*N	MR_B B	BLACK	
0020		CSIRO DIVISION OF ANIMAL PHYSIOLOGY		
0030		BLACKTOWN N. S. W.		
0040	*D	INFORMATION RELATING TO THE LOSS OF HAIR, WOOL OR		

다음의 4가지 論理가 適用된다.

1) OR論理

파라미터內的 各 檢索項目 사이에 適用된다. 項目의 어느 하나라도 해당되면 파라미터는 만족하게 된다.

2) AND論理

파라미터間에 적용된다. 全파라미터가 滿足되어야만 프로파일이 滿足되어 檢索된다.

3) NOT論理

檢索項目 카아드의 12-14欄에 制限되어 있다. 어떤 파라미터內的 어떤 項目에 NOT論理가 附與되어 있다면 이 파라미터안의 全項目이 컴퓨터에 의하여 NOT論理가 附與된다. NOT論理項目이 해당된다면 레코오드는 즉각 無視된다.

4) IGNORE論理

프로파일의 IGNORE論理 構造는 다음과 같다.

CHROM \*

IGNORE

CHROMATOGRAPHY

NOT論理보다 약한 것이며, IGNORE論理는 CHROMATOGRAPHY와 관계없이 된다. 그러나 계속해서 레코오드를 檢討한다. 그래서 CHROMATES의 CHROMATOGRAPHY에 관한 文獻은 重要語群에서 CHROMATES란 項目이 있는 경우에만 檢索된다. IGNORE 아래에 오는 項目은 관련이 있는 檢索項目을 포함하여야 한다. 例를 들면,

GENE

IGNORE

GEN \* 라고 하면 안된다.

4. 프로파일의 作成

4.1 作成方法의 種類

프로파일을 作成하는 方法에는 다음의 2가지가 있다.

0050		FLEECE AR A RESULT OF ANY CHEMICAL, METABOLIC,
0060		NUTRITIONAL OR ENVIRONMENTAL INFLUENCE
0070	*M 200	
0080	*T 0	
0090	*L C1	1 2 3 4 5 6 7 8 9 13 14 15 17 18 19 29 30 31 32 33 34
0100		C2 41 62 63
0110	*P T	*HAIR*
0120		IGNORE
0130		ROOT HAIR*
0140		*WOOL*
0150		*FLEEC*
0160		*KERATIN*
0170		FUR
0180		*MITOTIC*
0190		DEPILAT*
0200	*P T	DEPILAT*
0210		COSMETIC*
0220		REVIEW
0230		*NUTRI*
0240		*FOLLIC*
0250		HORMON*
0260		METAB*
0270		ROOT*
0280		IGNORE
0290		ROOT HAIR*
0300		COPPER
0310		*PLOIDY
0320		SHEARING
0330		ALOPECIA
0340		DISULFID*
0350		AMINO ACID*
0360		RADIATION
0370		ZINC
0380		CHELAT*
0390		ENZYM*
0400		*PROTEIN*
0410		SULFHYDRYL*
0420		CHEMOTHERAP*
0430		*MOULT*
0440		DERMAT*
0450	*E	
0460	*Z	

1) 純論理 프로파일 (Pure Logic Profile) 純論理 프로파일의 例를 表 1에서 紹介하였다. 項目이 每 파라미터마다 매치된다면 純論理 프로파일은 滿足된다. 파라미터 內에서의 探索

은 매치가 됨으로써 끝나고 컴퓨터는 다음 파라미터로 옮긴다. 그리고 全 파라미터가 滿足되는 경우 레코오드가 檢索된다. 純論理 프로파일은 一般的으로 加重 프로파일보다 빠르게 檢索되며,

明確히 定義된 狹少한 關心事의 檢索에 適切하다.

2) 加重 프로파일 (Weighted Profile)

加重 프로파일의 構造는 純論理 프로파일과 같으나 다른 것은 每 探索項目마다 同 項目 카아

드의 5~8欄에 加重值(플러스, 零 또는 마이너스)를 준 것이다. 加重 프로파일의 例를 表 2에서 紹介하였다. 매치過程에서 全 項目에 對한 加重值를 合算한다. 純論理 프로파일의 경우

表 2. 加重 프로파일

	*R01CQAH2	C1	IIP	GLJ
10				
20	*N	DR G G GREEN		
30		CSIRO DIVISION OF ANIMAL HEALTH		
40		INDOOROPILLY QLD		
50	*D	PATHOPHYSIOLOGY OF BABESIA ARGENTINA AND B. BIGEMINA		
60		INFECTIONS OF CATTLE PARTICULARLY PHARMACOLOGICAL		
70		ASPECTS OF THE DISEASE AND THE VASCULAR SYSTEM		
80	*M	200		
90	*T	110		
100	*P T	80	BABES *	
110		90	KININ *	
120		80	BRADYKININ *	
130		90	KALLI *	
140		80	SEROTONTI * TONI *	
150		80	HISTMAIN *	
160		50	INFLAMMAT *	
170	*P T	50	PATHO *	
180		-250	ANTAGON *	
190	K	30	*PEPTID *	
200		30	ANGIOTENSIN	
210		20	REVIEW	
220		30	SHOCK *	
230		20	PYRETIC	
240		30	BLOOD	
250		30	EXUDAT *	
260		30	VASO *	
270		20	*GLOBULIN *	
280		20	ANAPHYL *	
290		30	VASCULAR *	
300		30	TRYPTAMIN *	
310		30	ASSAY	
320		30	PROTOZOA *	
330		30	TRYPANOSOMA	
340		30	PLASMODIUM *	
350		0	*KININ *	
360		0	*HISTAMIN *	
370		80	BABES *	
380		30	BOVINE	
390		-30	DRUG *	
400	*E			
410	*Z			



와 같이 매치되기 前까지는 파라미터의 探索이 끝나지 않으므로 留意할 必要가 있다. 加重 프로파일의 全 파라미터가 滿足된다면 즉 每 파라미터에서 적어도 하나의 매치가 있다면 加重值를 合算하고 미리 設定해 둔 基準値와 같거나 이보다 크면 檢索하게 된다. 加重檢索에서는 探索項目에 加重值를 -250으로 定하면 NOT 論理와 같은 結果가 되며, 이 項目으로서 매치되는 경우의 文獻은 無視된다. 다른 論理는 모두 定해진 대로 適用된다.

加重探索에서 出力되는 文獻은 檢索 加重值가 높은 順序대로 프린트 되는데 純論理 探索에서는 文獻番號의 順序대로 프린트된다.

### 4.2 檢索效率을 높이는 方法

프로파일을 作成할 때에 다음의 몇가지 方法을 使用하면 檢索效率을 높일 수 있다.

#### 1) 項目의 切斷 (Term Truncation)

項目을 切斷해서 檢索하는 方法을 터엄 트런케이션이라고 하는데 單語의 語幹만 가지고 찾으면 檢索能力이 向上된다. 잘려진 檢索項目에는 표를 하는데 語頭나 語尾에 붙인다. 機械로 處理하는 過程에서 單語의 切斷表示만 하여 주면 컴퓨터는 \*표한 자리에 다른 文字群이 와 있었던 것으로 取扱해 준다. 項目이 잘려지 있지 않으면 項目의 뒤에 빈칸이나 點이 올 때까지 處理한다. 例를 들면,

Polymer		Polymer
*Polymer	} 라고 쓰면 {	Polymer, Copolymer 등
Polymer *		Polymer, Polymers 등
*Polymer *		Polymer, Copolymerization 등

이 檢索된다.

그러나 單語의 길이가 짧아진 만큼 不適合한 文獻이 檢索될 위험성이 커진다. 例를 들면 \*A-SE와 \*OSE는 酵素나 設糖 등의 語尾와 같고 해서 이들로서 酵素나 設糖에 관한 文獻을 檢索할 수는 없다. 왜냐하면 Base, Phase, Horse와 같은 일반적인 낱말들이 檢索되어 버리기 때문이다. 또 Carbon \*Oxide와 같이 中間에 \*표를 使用할 수 없다.

#### 2) 필터 파라미터 (Filter Parameter)

探索效率은 프로파일에서 첫번째 파라미터의

檢索項目의 數가 적을수록 더욱 最適化되는 것이다. 一般的인 규칙으로서 첫째 파라미터는 10個의 探索項目 以內로 制限하여야 한다. 그러나 때때로 不可能한 때가 있다. 이때 첫째 파라미터를 試驗하기 前에 필터 파라미터를 使用하여 出力에 逆效果가 없이 프로파일效率이 向上된다. 例를 들면,

```
*PT NITROGEN
NITROGEN FIX *
NITROGENOUS
AMMONI *
NITRAT *
NITRIT *
NITRIF *
IGNORE
*NITRIFICAN *
15N *
```

이것은 牧場에서의 질소 경제에 관한 프로파일의 첫째 파라미터이다. 이런 경우에 아래와 같은 필터 파라미터를 挿入함으로써 出力에 逆效果가 없이 프로파일效率이 向上된다.

```
*PT *NITR *
AMMONI *
15N *
```

이 필터 파라미터에서는 3개의 探索項目만으로 不適合한 레코오드를 일단 除去하고 餘分の 레코오드는 위의 파라미터로서 對照하면서 除去해 나간다. 이런 形式의 필터 파라미터를 導入하면 探索時間을 40%까지 節約할 수 있다.

#### 3) \*L 파라미터의 使用

이미 說明한 바와 같이 CAC는 다음과 같이 80個 細項으로 分類되어 있다.

```
홀수週 (C 1) 細項 1 ~ 34
짝수週 (C 2) 細項 35 ~ 80
```

\*L을 使用하면 特定 細項만 찾기 때문에 시간절약이 可能하다.

```
*L C 1 17 20 26 31
```

라고 쓰면 C 1에 있는 34개 細目中서 4개 細項만 찾는 것이 되며,

```
*L C 1 NOT 33
```

라고 쓰면 細項 1 ~ 34까지에서 33만 除外하고 모두 찾는다는 뜻이 된다.

4) 파라미터의 數와 順位

프로파일에서 사용되는 파라미터의 數와 그 順位는 效率에 미치는 영향이 크다. 파라미터의 數가 많을수록 要求主題를 具體적으로 正確하게 찾을 수는 있겠지만 너무 細分化되어서 出力이 적어지고 때로는 아무 出力도 얻지 못하는 경우가 있다. 따라서 프로파일 作成者는 되도록 적은 파라미터를 사용하여 要求主題를 正確하게 찾도록 하는 것이 最善의 方法이다. 또한 2개 이상의 파라미터를 사용할 경우 어느 파라미터를 먼저 사용할 것인가 하는 問題도 重要하다. 이 경우에 使用빈도수가 적은 探索項目이나 그외 파라미터를 먼저 사용하는 것이 原則이다.

5. 프로파일의 評價와 修正

5.1 評 價

프로파일의 檢索結果는 다음과 같은 事項을 檢討하여 評價할 수 있다.

1) 手作業으로 檢索된 文獻중에서 檢索되지 않은 文獻의 數

2) 手作業으로 檢索되지 않았던 文獻중에서 檢索된 適合文獻의 數

3) 檢索된 不適合 文獻의 數

探索범위의 깊이는 이 단계에서 決定되어야 하며, 이것은 要求情報의 用途에 따라 左右된다. 즉 利用者가 特定分野의 研究에 從事한다면 그는 部分的으로만 適合하다고 할지라도 그 分野에 관하여 出版된 것이라면 모든 文獻을 要求할 것이다. 그러나 높은 再現率을 요구하는 探索은 많은 不適合 文獻을 檢索한다는 것을 잊지 말아야 한다. 그러므로 꼭 필요한 주제만을 요구할 때에는 높은 精度의 探索을 함으로써 不適合한 文獻의 檢索을 줄일 수 있다.

5.2 修 正

프로파일을 修正하는 경우는 檢索結果의 상태에 따라 다음과 같이 4가지로 나눌 수 있다.

1) 檢索된 文獻이 없을 경우; 檢索項目의 範圍를 넓혀 프로파일을 再構成한다.

2) 檢索된 文獻의 數가 手作業에 의한 檢索

文獻의 數보다 너무 적은 경우; 手作業에 의해 抽出된 檢索項目의 論理的 組合에 留意하여 프로파일을 再構成한다.

3) 약간의 不適合한 文獻과 함께 手作業에 의해 檢索되었던 文獻이 檢索되는 경우; 修正할 必要없이 實際적으로 사용할 프로파일의 形態로 바꾼다.

4) 手作業에 의해 檢索되었던 文獻이 많은 부적합한 文獻과 함께 檢索되는 경우; 不適合한 文獻을 檢討하여 서로 어떤 共通되는 現象(예를 들면 모두가 特許라든가, 모두가 하나의 重要語나 Chemical Abstracts의 特定分類項目에 포함되었다든가 하는 點 등)을 갖고 있는가를 考察하여 不必要한 文獻을 除外시킬 수 있도록 프로파일을 再作成한다. 이러한 과정은 效果的인 프로파일을 作成할 때까지 反復한다.

6. 結 論

今年 7월부터 韓國科學技術情報센터에서 처음 시작하는 CAC SDI서비스는 科學技術情報 利用者들에게 놀랄만한 喜消息이다. 暴增하는 情報속에서 자기가 必要한 最新技術情報를 迅速正確하게, 網羅적으로, 簡便하게 入手할 수 있기 때문이다. 그러나 이 서비스는 만드시 훌륭한 프로파일을 作成하여야만 成功的인 效果를 얻을 수 있다. 그러므로 利用者는 上記한 프로파일의 構造와 作成法을 熟讀하고 여러번 經驗을 쌓아 自己의 要求主題에 대한 훌륭한 프로파일을 作成할 수 있도록 해야 한다.

앞으로 본 센터에서는 化學分野의 CAC 외에 INIS(原子力), INSPEC(物理, 電氣, 電子分野), CAIN(農業分野), FSTA(食品工業分野), BA PREVIEWS(生物, 農業分野), STIS(織物分野) 등의 다른 데이터 베이스도 導入하여 科學技術 全分野에 걸쳐 서비스할 예정이다.

本稿에서 記述한 것보다도 상세한 내용에 관하여 아시고자 하는 분은 본센터 전자계산실로 문의하여 주시기 바란다.

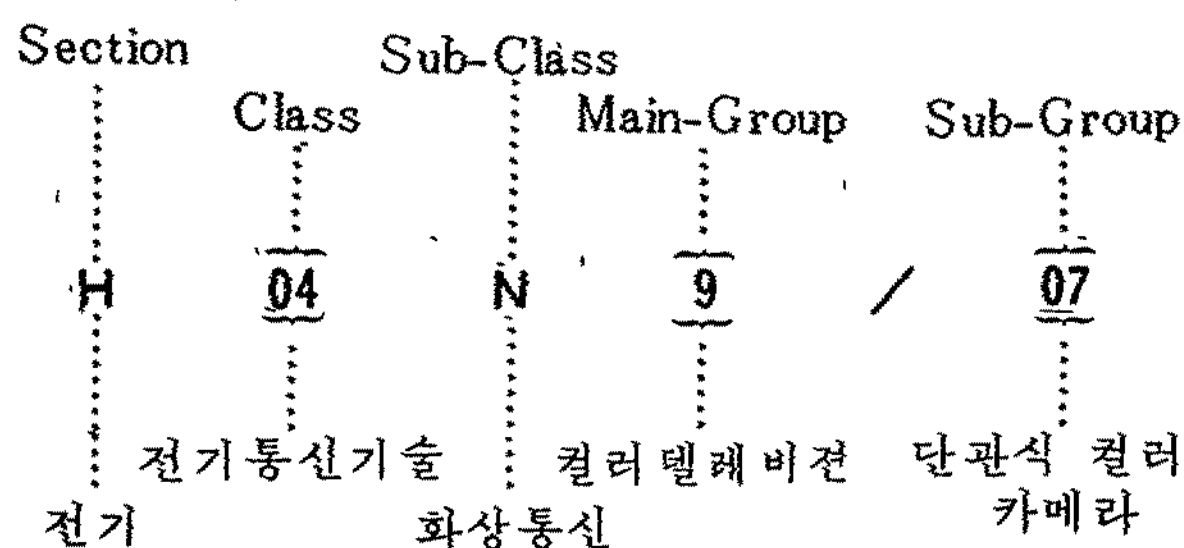
參 考 文 獻

<p. 82에 계속>

表 1. IPC의 구성

Section	산 업 분 야	기술·제품분야
A	생활필수품	농업, 음식품, 가정용품, 가구, 보건 등
B	처리, 조작	분리, 혼합, 성형, 인쇄, 이송 ..... 등
C	화학, 야금	무기화학, 유기화학, 수처리, 도로, 유리, 석유, 세제, 야금 등
D	직물	섬유, 방직, 염색, 종이, 펄프 ..... 등
E	고정구조물	도로, 개도, 다리, 수공, 급배수, 전조물, 열쇠, 광업 등
F	기계, 조명, 가열, 무기	기관, 펌프, 공학일반, 조명, 가열, 폭발, 무기 등
G	물리	측정, 시험, 광학, 사진, 시계, 제어, 계산, 교육, 악기 등
H	전기	전기소자, 발전, 전자회로, 전기통신 등

IPC의 실제



으로의 특허정보관리에 있어서 불가결한 요소가 되겠으므로 특허에 종사하는 사람은 각국의 특허분류와 아울러 IPC에 대해서도 각별한 공부

가 있어야 할 것이다.

6. 맺는 말

이상으로 특허정보의 개요를 주마간산격(走馬看山格)으로 살펴 보았다. 현재 기업에서 기술부분에 종사하면서 특허부문의 일을 맡고 계신 분이냐, 앞으로 특허분야에 관심을 갖고자 하는 분들에게 특허정보가 무엇인가 하는데 대해 다소나마 이해가 되셨으면 하는 마음 간절하다.

<p. 75의 계속>

1. Garrow C., Gilmore J.H. and Jackson G. L. "SDI IN CSIRO"
2. Jackson G. L.

3. "Profiling for CA Condensates" CSIRO SDI User's Manual 1974 (CSIRO Information Service)