(色材, 54 [9] 560-569. 1981)

化学分野における情報検索オンライン化の現状

本 山 明 子*

1. はじめに

①化粧品に使われている白色顔料の文献が欲しい。

②ジメチルアミノアゾベンゼンの誘導体(既知のもの)をもれなく知りたい。

あなたならこの質問にどうやって答えるだろうか。あな人は Chemical Abstracts (CA) などの索引を使って(マニュアルで)几帳面に調査するだろう。第1問は根気と時間をかければマニュアルでも調査できる。しかし第2問はマニュアルでの完全な調査は難しい。オンラインを使えばいずれも数分でコンピュータが調査を終えてくれる。

本稿では、まずオンラインに馴染みのない方にその威力をわかっていただき、続いて現在オンラインでどのようなことが調査できるか——現状——へと話をすすめていく。

2. オンライン検索の威力

冒頭の問題2題にもどり、マニュアル調査を始めた人の仕事ぶりをのぞいてみよう。

化粧品に用いられる白色顔料には酸化チタン(TiO_2)、酸化亜鉛(ZnO)、酸化ジルコニウム(ZrO_2)がある。一方酸化チタンを例にとってみると、無機顔料として塗料、印刷インキ、製紙のコーティング剤、合成樹脂、化粧品、ゴムにと広く使われている。

CA の Chemical Substance Index (化学物質索引) から始めよう。酸化チタンの CA 索引名 (CA Index Name, CA の巻末索引で正式に用いられている名称) が Titanium Oxide であることは、Formula Index (分

子式索引) などであらかじめ調べておく。実際に CA 92 巻 (1980 年の前半) の Chemical Substance Index を開いてみると 694 件の文献が Titanium Oxide の項に索引されている。CA の索引は analysis, biological studies, occurrence, preparation, properties, reactions, uses and miscellaneous などに分けて探しやすいように並んでおり、また簡単な説明句がついているので根気よく探せば図-1 のように見つけることができる。が、半年で 694 件であるからこの調子でいけば 5 年間で約 7,000 件、他に酸化亜鉛、酸化ジルコニウムの項も見なければならないのであるから一仕事である。一方 General Subject Index (一般事物索引) の Cosmetics (化粧品) の項からでも調べられる(図-2)が、やはり半年当たり 100 件ぐらいの文献の中から説明句をたよりに選ばなければならない。

このように第1問は根気よく探せば見つかるが、時間もかかるし、また数百件あるいは千件以上の中から探し

●
●
●
Titanium oxide (TiO_2) [13463-67-7], biological studies
air pollution by, from welding electrodes, occupational exposure to, health hazards in, 152264b
air pollution by sulfur dioxide from plant manufg., 134421w
beer chillproofing with papain and, 196326c
beverage clouding agent contg. maltodextran and xanthan gum and, P 162416s
diazotized diaminobenzene-copacked, dextranase immobilized of, in reactor-product separator system, 142407m
glucose detn. in body fluids with test strips contg., P 124488j
in human lung tissue, anal. of, 134524g
hydrated, trace element in seawater adsorption by, 152617u
in margarine surface dehydration discoloration prevention, P 40140b
nylon 12 powder modified with, for cosmetic vinyl face masks, 152380z
occupational exposure to, in air of submerged-arc electroslag welding fluxes, health hazard in relation to, 134525h
●
●
●

図-1 Chemical Abstracts 92 巻
Chemical Substance Index より

昭和 56.5.29 受理

Online Information Retrieval in Chemistry

Akiko MOTOYAMA

* (社)化学情報協会

東京都文京区弥生 2-4-16 学会センタービル

•
 •
 •
Cosmetics
 absorption of, by skin, physicochem. factors effect
 on, R 99484v
 acetylsalicylate and indomethacin sunscreens in, P
 82247x
 N-acylamino acid esters with polyoxyalkylene polyol
 ethers for, P 11098p
 •
 •
 •
 erythrosine dye for, neuromuscular junction
 acetylcholine release response to, 196516q
 for eye, hydroxyalkyl polyglycosides for removal of, P
 111273k
 face masks
 enzymes in, R 135087d
 in Japan, R 185696x
 titanium dioxide modified nylon 12 powder form
 vinyl, 152880z
 foaming and cleansing effect of, 152859z
 food-grade preservatives for, R 28374y
 •
 •
 •

図-2 Chemical Abstracts 92 巻
General Subject Index より

出すのであるから“見落し”もあるかもしれない。

第2問に移ろう。ジメチルアミノアゾベンゼン (I)

は N-ジメチルの部分が作用する肝癌の発癌物質といわれているが、この物質の構造をもつ物質は発癌性物質として疑ってみる価値がある。

1980 年夏 CAS 化学物質登録システム (Chemical Registry System) に 500 万番めの化学物質が登録された。この CAS 登録システムについては 5.1 に述べるが、今までに文献に発表された化学物質を片っぱしから登録してある戸籍簿のようなものとお考え頂きたい。500 万——1981 年 4 月現在では約 540 万——の化学物質ファイルを手作業で調査するとどのくらいの時間がかかるのだろうか。仮に化学構造式がきちんと書かれた、理想的なカードとなった化学物質ファイルが手元にあったとして、その中からジメチルアミノアゾベンゼンの誘導体を選び出すことを考えよう。一つの化学物質が該当するかしないかを見るのに 2 秒かかるとすると 1 分で 30, 1 日 7 時間働くとして 1 日 13,000 の化学物質を調べることができる。1 年に 250 日働くとして約 300 万, その間の化学物質数の増加を考慮すると約 2 年かけてこの調査が終わる計算となる。このようなカードファイルをもっている人は現実にはいないであろうし、第一お忙しい読者諸兄から 2 年間という時間を頂くことは不可能だろう。

このようにマニュアルでは大変な調査, あるいは不可

能な調査でも, オンライン検索なら第 1 問, 第 2 問ともに数分で検索できる。それにもう一つ, コンピュータは人間がするような“見落し”はしない。それではオンラインではどのように検索し, そして何がわかるのかみてみよう。

3. オンライン検索に必要な道具—— 端末機と電話

オンライン検索はあなたの端末機と, システムを運転している大型コンピュータを電話で結んで行なう。

端末機にはいろいろな種類があり, 英数字用のもの, 英カナ字用のものや, 漢字あるいは英語の小文字が使えるものがある。また感熱紙上に印字されるタイプライタ型のポータブル端末のほか, 文字表示用のディスプレイとプリンタを合わせて使うものもある。また, 図-3 はあとで説明する CAS ONLINE 用のグラフィック端末で, 画面上に化学構造図が表示できる。

電話器は通常家庭用電話器でよい。交換台を通すとノイズが入りやすいので, できれば別の電話を用意するのがよい。欲をいえば専用回線の方がよいが, 費用も余計にかかるので一般にはふつうの電話 (公衆回線) が使われている。

4. CA SEARCH/JOIS を使った文献検索

第 1 問——化粧品に使われる白色顔料の文献——を, JOIS-II システムの CA SEARCH ファイルを使って検索してみよう。

4.1 CA と CA SEARCH と JOIS

Chemical Abstracts Service (CAS) は 1907 年以來世界中の化学文献を収集し, その抄録集 Chemical Abstracts (CA, 週刊) を刊行してきた。これにおきかわ

図-3 CAS ONLINE 検索に用いる端末機

* 対応特許を含む

図-4 Chemical Abstracts の収録文献数
(CAS TODAY より)

る抄録誌がない今日、CA は化学にたずさわる者すべてにとって不可欠のものである。今年はちょうど CA 創刊 75 年めにあたるが、戦争中の文献集めなど苦労話も尽きないようである。編集はもちろん、索引もカードを使って手作業で作成する時代が長いこと続いたが、文献の増加(図-4)に伴いそれでは間に合わなくなってきた。巻末索引の作成に 20 か月もかかるようになってしまったのである。そこで CAS はコンピュータの導入に踏み切り、数年間の研究開発の後、索引作成、本文の版下作成と順次作業をコンピュータに移していった。今でこそコンピュータを駆使した編集は多いが、CA がその草分けである。今日、抄録や索引を作り出す知的作業を除くと、データの編集、版下作成などの作業はほとんどすべてコンピュータがやっている¹⁾。こうしてコンピュータに入力されたデータから、いわゆるデータベースが副生する。

CA に対応するデータベース(またはファイル) CA SEARCH は作成するコンピュータの都合で隔週に発行されているが、対応する 2 号分の CA に収録された文献の表題、著者、書誌事項、キーワード、索引語が収められている。CA SEARCH はいくつかのオンライン検索サービスで検索できるようになっているが、ここでは

①→ ¥JOIS 13 K-0030, AM

s: パ ス ワード ハ?

②→ u: ×××××

③→ s: サービス オ カイシ シマス 1981.04.27

④→ u: ¥FILE 210/ALL

s: CASEARCH (V.086 N. 01-V. 094 N. 12)
1, 845, 633 (17:00 マデ)
シツモン オ カイシ シマス 14:11:
39 カイワ バンゴウ 438
ACS COPYRIGHT

⑤→ [1] u: RN: 13463-67-7

⑥→ s: 5925 ケン

⑦→ [2] u: RN: 1314-13-2

⑧→ s: 4454 ケン

⑨→ [3] u: RN: 1314-23-4

⑩→ s: 2818 ケン

⑪→ [4] u: 1+2+3

⑫→ s: 11850 ケン

⑬← [5] u: COSMETIC &

⑭→ s: 1719 ケン

⑮→ [6] u: 4*5

⑯→ s: 23 ケン

[7] u: ¥P

:

図-5 JOIS の検索

JOIS を例にとって説明してみよう。

JOIS は日本科学技術情報センター(JICST)が作成し、運転しているオンライン検索システムで、1977 年から日本中にサービスされている。はじめのシステムは JOIS-I という名前であったが、この春からシステムの機能が向上し収納ファイル数も増え、名前も JOIS-II となった。一部漢字も使える日本人向けのシステムである²⁾。

(“CA SEARCH ファイルを JOIS システムで検索する”という意味で CA SEARCH/JOIS と書くことにする。)

4.2 化粧品に用いられる白色顔料の文献

CA SEARCH/JOIS で第 1 問、化粧品に用いられる白色顔料の文献を探してみよう。図-5 で下線部がこちらからの入力、その他がシステム側からのメッセージである。

まず、電話で JOIS を呼出し、ピーという発信音が聞こえたら受話器を端末機のカプラという部分にはめこむ。ユーザ番号(①)に続いてパスワードを入力する(②)。パスワードは銀行の暗証番号と同じで、自分のユ

ーザ番号(銀行の口座番号に当たる)を他人に使われな
いためのものである。パスワードがあっていると JOIS
システムがあなたを迎えてくれる(③)。まずファイルと
範囲を指定する(④)。“210”は CA SEARCH, “ALL”
とは JOIS にのっている分すべて(この場合 1977 年分
から最新号まで)の意味である。

ところで CA SEARCH で文献を探すとき, 化学物質
が特定されている場合はその化学物質の背番号である
CAS の(化学物質)登録番号(5.1 で詳述)で探すの
が最もよい。酸化チタン, 酸化亜鉛, 酸化ジルコニウム
の CAS 登録番号が各 13463-67-7, 1314-13-2, 1314-23-
4 であることはあらかじめ Chemical Substance Index,
Index Guide など調べておく。

酸化チタン, 酸化亜鉛, 酸化ジルコニウムの CAS 登
録番号を入力すると(⑤, ⑦, ⑨), それぞれについての
文献が 5,925 件, 4,454 件, 2,818 件あることをシステム
が答えてくれる(⑥, ⑧, ⑩)。この三つの化学物質の
うち少なくとも一つについて記述されている文献は(⑪)
11,850 件ある(⑫)。ここで+印は論理和の意味である。
一方化粧品に関する文献を探すため “COSMETIC &”
を入力する(⑬)。“COSMETIC &”とは COSMETIC
という文字列のあとにどのような文字(スペースを含む)
が続いていてもよいことを示している(COSMETICS など)。
これらの語が表題, キーワード, あるいは索引語
に含まれている文献が 1,719 件あることをシステムが答
えている(⑭)。さて, 探しているのは COSMETIC &
という語を含み, 先に入力した三つの CAS 登録番号の
少なくともどれかが索引されている文献であるから, そ
れを⑮のように*印(論理積)で指示すると, 該当する
文献が 23 件あることがわかる(⑯)。図-6 はそのうち
の1件のデータをその場で出力したものである。23 件
ぐらいなら 2~3 分で打出してくれる。文献数が多いと
きにはコンピュータから直接打出し郵送で受取るオフラ
インも使える。

ちなみに, JOIS を呼出し, 会話して 23 件の文献を出
力するのに要した時間は 9 分であった。

4.3 CA SEARCH/JOIS で得られる情報

CA SEARCH は文献ファイルであるから, CA SEAR-
CH を検索することは “CA の索引をひくかわり” であ

る。単に “かわり” といっても人間が手で探すよりずっ
と短い時間でしかも見落としなく探せる。ここで例にあげ
たのは最も単純な検索であり, 問題によっては質問式が
ずっと複雑になるが, 手作業より容易なことにかわりは
ない。

ところで CA SEARCH/JOIS で検索すると文献の何
がわかるのか。もう一度出力見本(図-6)をご覧ください。
検索結果が教えてくれるのは, この文献の CA 収録
巻, 号, 抄録番号(①), 表題(②), 著者名と所属機関
(③), 雑誌名(④), 雑誌の巻, 号, 頁, 発行年(⑤),
文献の種類(雑誌, 単行本, 特許など), 言語(⑥), CA
の分類と副分類(⑦), 索引語(⑧), (⑨), 化学物質の
CAS 登録番号(⑩)である。抄録こそついていないが,
この文献が化粧品の顔料について研究したものであり,
3 種の白色顔料のうち酸化チタン, 酸化亜鉛について記
述があることはわかる。そして目的に合った読むに値す
る文献かどうか大体の判断がつくし, 判断しかねたりそ
の文献が手元にない場合は, CA の抄録を読めばもう少し
詳しいことがわかる。

ところで, CA SEARCH/JOIS はどういう時に——ど
ういうニーズがある時——使えるだろうか。

- ① 特定の化学物質(群)に関する文献を探したいと
き。
- ② 特定の事柄についての文献を探したいとき。
- ③ 特定の著者の文献を探したいとき。
- ④ 特定の会社・機関の発表を探したいとき。

実際の質問では ①~④ がいろいろ組み合わせられて
いる。この他, 特定の雑誌に掲載された報文を拾い出すこ
ともできるし, 言語, 年代などで限定することもでき
る。

検索できる範囲は, もちろん情報がコンピュータ処理
可能な形になっていて, しかもシステムでサービス提供
している分である。CA SEARCH ファイルは 1967 年
分からあるが, JOIS にのっているのは(従って CA
SEARCH/JOIS で検索できるのは)1977 年以降の文献
である。

5. CAS ONLINE を使った化学物質検索

第2問——ジメチルアミノアゾベンゼンの誘導体——
を CAS ONLINE を使って調査しよう。コンピュータ
が検索に要する時間は約 5 分である。

5.1 CAS ONLINE と CAS 化学物質登録システ

ム³⁾

CA の Chemical Substance Index は 1907 年に CA
が創刊されたときから続いている(1971 年までは Sub-

⑤→CN= C069016127377
⑥→TI= COSMETICS CONTAINING PIGMENTS IN WATER AND DISPERSERS
⑦→AU= OZAWA TADASHI (FOKA KASEI KOGYO K. K., YOKOHAMA, JAPAN)
⑧→JN= SKYDA (0371-0777) SHIKIZAI KYOKAISHI
⑨→VN= VOL.51 NO.12 PAGE.711-19 '78
⑩→CI= 13 (OR)
⑪→CC= C062001
⑫→KW= COLORING MATERIALS; COSMETICS; DISPERSING AGENTS; KAOLIN, PROPERTIES;
PROPERTIES; BIOLOGICAL STUDIES
⑬→FT= COSMETIC; COLOR; DISPERSIBILITY; AQ; RELATION
⑭→RN= 1308-38-9; 1314-13-2; 1317-61-9; 7758-79-4; 7722-99-5; 7758-29-4; 7778-53-
2; 5011-77-0; 12626-43-6; 13463-67-7; 14607-76-6; 51274-00-1; 69670-73-1

図-6 JOIS の出力

ject Index として現在の General Subject Index と合
わさって発行されていた)。この索引は長年手作業で作
成されてきたが、索引される化学物質数の増加は著し
い。たとえば創刊号の CA Subject Index は全部で
46,000行であったが、1976年前半の Chemical Substance
Index はベンゼン関連項目をカバーするだけで同じ行数
を使っている。また化学物質はいろいろな名称と呼ばれ
ることが多く、一貫した系統的名称をつけて管理してい
くことは大変な努力が必要であった。従って CAS の作
業にコンピュータが導入されたとき、この化学物質情報
処理の部分が真っ先に機械化された。CAS のコンピュ
ータが最も活かされているのはこの処理部分であり、そ
の中核となっているのが CAS 化学物質登録システム
(Chemical Registry System) である。このシステムは
もともとは CA の Chemical Substance Index 作成の
ために作られたが、現在は事実上すべての既知物質が登
録されている世界唯一の化学物質辞書として権威もあり
また重宝がられている。

文献に出てきた化学物質は、まず文献中に書かれてい
る“名称”により CAS 化学物質登録システム中にすで
にあるかどうかチェックされる。そこで“ない”とされ
たものは、さらに構造式を描いて照合される。それ
でも“ない”となるとその化学物質ははじめて新規物質と
して登録される。以上の照合はすべてコンピュータで行
なわれる。

ところで、CAS 化学物質登録システム中の化学物質
には背番号として CAS 登録番号がついている。この番
号は、化学物質が CAS 化学物質登録システムに登録さ
れるとき順番につけられるので、化学的な意味はない。
CAS 登録番号は今や CAS 化学物質登録システムにおけ
る整理番号であるのみでなく、4. でおわりのとおり
文献検索でもなくてはならぬ存在である。米国環境保護
庁(EPA)の毒性物質規制法(TSCA)の目録(Inventory)
や Merck Index (第9版)など多くの化学物質ファ
イル、出版物にも使われており、化学界の共通語にならう
としている。

CAS 化学物質登録システムにはひとつひとつの化学
物質に対して CAS 登録番号、CA 索引名、慣用名、商
品名などのほか、構造式や立体化学の情報が収録されて
いる(図-7)。この CAS 化学物質登録システムと、コ
ンピュータが得意とする自動編集(並びかえや選出)の
手法により、CA の Chemical Substance Index は迅
速かつ正確に作られるようになった。

CAS ONLINE はこの化学物質登録システムをオン
ラインで一般に使えるようにしようというサービスで、

図-7 CAS 化学物質登録システムに登録されているデータ

```

① → @C 614 21 S
      //LOGIN ××××××××
      ×××××××× IS STARTING
② { ENTER PASSWORD : ××××××××
    { ENTER PERSONID : ×××××
      PASSWORD COMMAND COMPLETE-NEW
      PASSWORD / PERSONID ACCEPTED
      ENTER COMMAND : SCREEN
③ → SC 12 : AZOCAN=172 342 1121 1665 1609
      1399 1868
      SC 12 : END
      SCREEN COMMAND COMPLETED
④ → ENTER SRCHSTART OR *END : SRCHSTART
      1 : 36 : 14 SEARCH COMPLETE (1 : 34 : 30)
⑤ → SEARCH TIME (0. 04. 42)
      TOTAL TIME (0. 06. 26)
⑥ → SUBSTANCES SCREEND : 5357396 (100.000%)
      ANSWERS WRITTEN TO FILE : 2511
      SUBSTANCES PASSED : 2511 (0.046%)
      ANSWERS TRANSMITTED : 4
  
```

図-8 CAS ONLINE の検索

1980 年末にスタートした。日本の場合、KDD の ICAS、
米国の TELENET などの通信回線を介して、CAS の大
型コンピュータとあなたの端末機とをつないで直接検索
ができる。構造はわかっているが名前がわからない、あ
るいは文献検索の結果、CAS 登録番号で示された化学
物質の構造や名称を調べたいという場合にはもってこ
いのシステムである。またその化学物質が索引された文
献の CA 抄録番号を表示する一種の文献検索の機能もも
っている。近い将来は化学物質の名称から構造を調べ
ることもできるようになり、ますます便利な“オンライン
化学物質辞書”となる予定である。

5.2 ジメチルアミノアゾベンゼンの誘導体

それではこの CAS ONLINE を使って第2問を実際に解いてみよう (図-8)。端末の前にすわり、第1問と同じ要領で電話で ICAS を呼出し続いて TELENET と接続する。ただし、ここで使う端末は文字だけでなく図の描けるグラフィック端末が好ましい。CAS のコンピュータを呼出し (①), LOGIN ID (ユーザ番号), パスワードなどを入力すると (②), もう CAS ONLINE があなたの質問を待っている。探している構造は数個のスクリーン (番号) という数字で表現して質問するので、ここでスクリーンについて説明しておこう。スクリーンとは化学構造を組立てているさまざまな要素をコード化したもので、いわばこのオンラインシステムの要である。たとえばアゾ基 $N=N$ には 1665 というスクリーン番号がついていて、この 1665 を入力することにより、CAS ONLINE は $N=N$ を含む化学物質を一つ残らず見

図-9 CAS ONLINE のスクリーン

図-10 CAS ONLINE で出力される化学構造式

つけてくれる。他のスクリーン例を 図-9 に示す。このスクリーンは別に用意してある印刷版のスクリーン辞書で簡単に調べることができる。実際の検索にもどるがこの AZOCAN という質問式 (③) は 172, 342, 1121, 1665, 1609, 1399, 1868 という7個のスクリーンで表わされる構造を満足する化学物質を選び出せという意味である。質問が組立てられたのち、次の SRCHSTART で実際の検索が始まる (④)。

検索が行なわれている間は OUT というコマンドで見つかった化学物質を表示することができるので、グラフィック端末で見てみよう (図-10)。これらの化学物質が、注目している構造を含んでいることは、表示された構造図によりすぐわかる。

そうこうしているうちに検索は終了する。この検索にかかった時間は4分42秒で (⑤), 2511 個の化学物質が見つかった (⑥)。ここで注意しなければならないのは、ここで得られた化学物質は検索式 AZOCAN に該当するもので注目しているジメチルアミノアゾベンゼン骨格をもっているものはすべて含まれているが、それ以外のもの——いわゆるノイズ——も入っている可能性がある。このノイズは質問式の作り方により減らすことができるが、手間とのかねあいもあるので適当なところで妥協するのがよい。

5.3 CAS ONLINE で得られる情報

現在 CAS ONLINE で検索キーとして使えるのは、上記のスクリーンの他に CAS 登録番号がある。従ってある部分構造を含む化学物質をすべて選び出したいとき、あるいは文献検索で見つかった CAS 登録番号をもつ化学物質が何であるか知りたいとき、CAS ONLINE が問題を解決してくれる。また、近々化学物質名、分子式からも検索できるようになる予定である。さらに計画では、スクリーン番号をいれるかわりにグラフィック端末画面上に自分で化学構造を組立て、その化学構造から CAS のコンピュータがスクリーンを自動的に選び出す機能がつけ加えられるはずである。

回答の内容は例 (図-11) に示すとおりであるが、CAS 登録番号 (a), 分子式 (b), CA 索引名 (c), 慣用名や商品名 (d), 化学構造式 (e), その化学物質について記載のある文献の CA 抄録番号 (f) で、立体異性体がある場合には立体化学の情報も表示される。

ところで、CAS ONLINE のデータベースになっている CAS 化学物質登録システムは、1965 年以降の世界中の化学文献 (特許を含む) に記載された化学物質がすべて登録されており、事実上全既知物質が網羅されていると考えられるので、この例のように化学構造と活性・毒

```

a → REC 68-11-7
b → FOR C14 H15 N3
c → IND Benzamine, N,N-dimethyl-4-(phenylazo)- (9CI)
d → SYN C.I. 11020
   SYN Brilliant Fast Oil Yellow
   SYN Brilliant Fast Spirit Yellow
   SYN Brilliant Oil Yellow
   SYN Cerasine Yellow GC
   SYN 4-(N,N-Dimethylamino)azobenzene
   SYN Dimethyl Yellow
   SYN DMAB
   SYN Enial Yellow 2C
   SYN Fast Oil Yellow B
   ADDITIONAL NAMES - UNAVAILABLE FOR DISPLAY
DEN 55964-95-9;

```

c →

```

① { REFERENCES
    CA94(16):127972q
    CA94(13):97665x
    CA94(12):91287d
    P CA94(12):85120c
    CA94(9):59501h/AMD
    CA94(7):42408k
    CA94(7):42398g/AMD
    P CA94(2):8080n
    CA94(2):7722m
    CA94(1):894k
    718 MORE

```

図-11 CAS ONLINE の出力

性の相関の研究をする上で化学物質辞書として使えるほか、ある化学物質が新規であるかどうか正確に知ることができるので、特許関係の調査では非常に強力な手段である。

6. データベースと検索システム

オンラインサービスはどんなものでもデータベースと検索システムとで成立っている。データベースはデータ要素の集まりで、いわば商店に並べられた商品、検索システムはそれらを売る商店のようなものである。従ってデータベースに入っていないければ、入荷されていない商品を買うことができないように、そのことばを検索のとき使うこともできないし、出力もできない。逆にいくらデータベースに入っているシステムがそれを検索あるいは出力するようになっていなければ、倉庫に入れられたままの商品と同じでやはり検索不能である。

第1問で使った CA SEARCH/JOIS は、CAS が提供する CA SEARCH というデータベースを JICST の JOIS システムにのせたものである。それに対して CAS ONLINE は CAS 化学物質登録ファイルをデータベースとする化学物質検索システムで、ともに CAS が提供するものである。

CA SEARCH の場合、ここで取上げた JOIS システムのほか、米国 Lockheed 社の DIALOG システム、同

じく SDC 社の ORBIT システムでもオンラインサービスされている。データベースが同じでも各システムで商品配列が違うので、システムに合った検索をしなければならない。つまり同じ化学文献検索システムでもそれぞれに特徴がある。

7. その他のオンライン

限られた紙面で各種オンラインを深く説明することはできないので、詳細は各文献に譲ることにし、おもなものについて“何を調べることができるか”みていこう。

7.1 文献情報

7.1.1 TOXLINE/JOIS⁴⁾

TOXLINE は異なる機関から発行されている 11 のサブファイルを米国国立医学図書館 (NLM) で編集している毒性情報に関する文献ファイルで、11 のサブファイルの提供者と各の立場、見地はつぎのとおりである。

CBAC (CAS, 化学)

TOXBIB (NLM, 医学)

PESTAB (EPA, 農学)

IPA (米国病院薬剤師協会, ASHP, 薬学)

HEEP (Biosciences Information Service, BIOSIS, 環境化学物質)

HAYES (Dr. Hayes, 農薬)

EMIC (Oak Ridge National Laboratory, EMIC/ORNL 他, 環境内突然変異誘発物質)

TMIC (ORNL, 毒性物質)

TERA (奇形学)

ETIC (ORNL, 環境物質による奇形発生)

RPROJ (Smithsonian Science Information Exchange, SSIE, 毒物学, 疫学)

収録データ項目はサブファイルにより異なるので検索法は一概にはいえないが、文献の表題、書誌事項、キーワード、化学物質の CAS 登録番号などから検索できる点は CA SEARCH/JOIS とよく似ている。出力内容も CA SEARCH/JOIS とほぼ同じであるが、サブファイルによっては抄録が収録されているので、それが出力されるのが特徴である。

7.1.2 JICST 理工学文献ファイル/JOIS⁵⁾

理工学全般にわたる日本語ファイルで抄録が入っている。JICSTで発行している科学技術文献速報をオンライン化したもので、JOIS には 1975 年以降のものがのっている。化学、化学工学のほか物理学、機械、建築などの分野も含まれており、年間収録文献数約 50 万件のうち約 3 割が化学、化学工学である。検索は書誌事項、キーワードを入力語とするが、出力としてはこれらの他に抄録

が含まれる。キーワードは専用のソーラスに基づいてつけられており、検索はそのソーラス中の語を使うのが特徴である。日本人向けのファイルで、最近開発されたポータブルの漢字端末を使うと、漢字かなまじり文で回答がかえってくる。

7.1.3 Surface Coatings Abstracts/DIALOG⁶⁾

英国塗料研究協会が表面塗装とその関連事項についての世界中の雑誌記事、特許、単行本を収録しているファイルで、収録文献数は年間約 8,500 件である。

7.1.4 その他の文献ファイル

化学の関連分野のおもなファイル名/システム名を以下にあげる。詳細は他の資料を参照されたい。

AGRICOLA/DIALOG, OPBIT (農学)^{6,7)}

BIOSIS/DIALOG, OPBIT (生物学)⁸⁾

CAB/DIALOG, ORBIT (農学)⁹⁾

COMPENDEX/DIALOG, ORBIT (工学)¹⁰⁾

INSPEC/DIALOG, ORBIT (物理学, 電気工学, 電算機)¹¹⁾

Science Citation Index/DIALOG, ORBIT (科学技術)^{6,7)}

7.2 特許情報専用のオンライン

7.2.1 PATOLIS

日本特許情報センター (JAPATIC) が提供する特許情報専用のシステムで、データベースは日本の特許、実用新案、米国特許を収録するファイルである。検索、出力はともに日本語で、漢字、カナ端末を使う。収録内容は特許文献、特許出願人、発明者に関する書誌的事項や審査経過情報のほか、特許公開の場合キーワードや要約も入っている。米国特許は International Patent Documentation Center (INPADOC) とのデータ交換により入手しており、収録内容は書誌的情報のみである。

このオンラインシステムでは、いわゆる検索のほかにデータ照会、番号照会の機能がある。このうち検索ができるのは日本のファイルのみで、検索項目は出願人コード、IPC/JPC (分類) コードと、JAPATIC がつけた分類コードとキーワードである。また日本、米国のファイルでデータ照会ができ、日本特許の出願番号、公開番号、公告番号、登録番号、米国特許の出願番号、特許番号のいずれからでもアクセスできる。検索、データ照会の結果個々の特許について前述したファイル収録内容を見ることができるが、これとは別に特許の番号の対応関係 (出願番号と公開番号など) を見る番号照会もある。あるテーマについての特許を見つけないとき、ある出願番号の特許の審査経過が見たいとき、ある公開番号の特許の内容が知りたいとき、PATOLIS が使える。

7.2.2 WPI/ORBIT (SEARCH-J)⁷⁾

Derwent Publications, Ltd. が提供する、世界各国の特許を集めたファイルである。すでに収録されていた医薬、農業関係の特許に加え、高分子関係が 1966 年から、その他の化学特許が 1970 年から、そして 1974 年からは全分野の特許が収録されるようになった。書誌的事項、キーワード、Derwent 社による分類、化学構造式により検索できる。

7.2.3 CLAIMS/DIALOG⁶⁾

IFI が作成している米国特許のファイルで米国特許庁の Official Gazette 化学部門および関連部門の全特許を収録している。化学の特許ではキーワード、CAS 化学物質登録番号や英国、西ドイツ、フランス、ベルギー、オランダの対応特許も収録されている。

7.3 データ情報システム

Chemical Information System (CIS)¹³⁾

米国の EPA, National Institute of Health (NIH) が開発した、化学物質の安全性に関する数値情報検索システムである。化学物質の辞書ファイルである SANSS ファイル (7.4.2) を中心に、41 のデータファイルで構成されており、これらのデータファイルの提供、サポート機関は世界各所に散らばっている。データファイルのおもなものを 2, 3 あげておこう。

a. MSSS (Mass Spectral Search System) 英国の Mass Spectrometry Data Centre (MSDC) 他 1 機関、米国の NIH, EPA 他 3 機関から入手したデータをもとに作られているファイルで、マスマススペクトルと関連文献が収録されている。マスマススペクトルのピークから該当する化学物質 (群) を拾い出すこともできるし、化学物質を指定してスペクトルを出力することもできる。

b. RTECS (NIH-EPA-NIOSH Registry of Toxic Effects of Chemical Substances) 米国の National Institute of Occupational Safety and Health (NIOSH) により収録された化学物質の毒性、生物学的効果に関する基礎的データのファイルである。化学物質を指定してこれらのデータとその出典を見ることができ、特定の化学物質の各動物に対する致死濃度、致死量など調べることができる。

c. OHMTADS (Oil and Hazardous Materials Technical Assistance Data System) EPA, NIH が協同でサポートしているファイルで、油や有害物質などの安全性に関する広範なデータ——たとえば引火点、発火点、残留性、発癌性、取扱い基準など 126 種——を収録している。化学物質を指定してデータを見ることができただけでなく、特定のデータ項目の内容を限定してそれ

に合った化学物質を選び出すこともできる。

7.4 化学物質辞書

化学物質辞書としてすでに CAS ONLINE を紹介した。CAS ONLINE は独立した化学物質辞書として多様な目的に使われているが、多くの場合、オンライン化学物質辞書は特定の文献情報システム、あるいはデータ情報システムのサポートシステムとして動いている。ここにいくつかの例をあげよう。

7.4.1 CHEMLINE/JOIS⁽¹⁴⁾

NLM の要請により CAS が作成し更新している化学物質辞書ファイル CHEMLINE が JOIS システムにのっている。TOXLINE 文献ファイルに対応する辞書ファイルで、各化学物質の CAS 登録番号、CA 索引名、慣用名、商品名、分子式などの他、TOXLINE のどのサブファイルにその化学物質が収録されているかが記載されている。従って TOXLINE を検索するための化学物質名から CAS 登録番号の調査、逆に文献検索の結果出力された CAS 登録番号から名称の調査の他、一般的化学物質辞書として、また CA 索引名を利用した部分構造検索にも使える。

7.4.2 SANSS/CIS⁽¹⁵⁾

7.3 の CIS システムの一部で、CIS の各種データファイルを使うための化学物質辞書ファイルである。化学物質の CAS 登録番号、各種名称、分子式の他に構造式も収納されている。また CIS 中のどのファイルにその化学物質の記載があるかののっている。

7.4.3 CAS NAME/JOIS⁽¹⁴⁾

CHEMNAME, CHEMSIS, CHEMSEARCH/
DIALOG⁽⁶⁾

CHEMDEX/ORBIT⁽⁷⁾

化学文献ファイル CA SEARCH に対応する化学物質辞書ファイルで、各提供システムごとに幾分形が異なっている。CAS NAME, CHEMSEARCH には CAS 登録番号、CA 索引名、分子式が、また CHEMNAME, CHEMSIS, CHEMDEX にはこの他に慣用名、商品名が収録されている。

7.5 ビジネス情報

“化学”の見方には科学、技術としての立場の他に、ビジネスというもう一つの立場がある。

7.5.1 CIN/DIALOG, ORBIT⁽¹⁵⁾

世界の主要な新聞、雑誌から化学界のビジネス情報を収集し、その抜粋を収録しているのが、CAS が発行する Chemical Industry Notes (CIN) である。もともと印刷版で利用されているものであるが、磁気テープの形で発行され、DIALOG, ORBIT システムにのったのがオンラインで使える。83 の新聞、雑誌にくまなく目を通すかわりに CIN を検索することにより、化学界のようすが容易にかつ正確につかめるわけである。特定の企業の動向、特定の分野の業界の動向、各国政府の動きをみるのに最適である。

7.5.2 NEEDS-IR⁽¹⁶⁾

日本経済新聞社が自社で発行する新聞の記事をオンラインで検索できるようにしたシステムである。

8. その他のオンラインの可能性

オンラインで調査ができるためには少なくとも情報ベースがコンピュータ可読になっていなければならない。コンピュータ可読となっているファイルをコンピュータシステムにのせれば、原理的にはオンライン検索が可能である。

コンピュータ可読の形になっていながらもまだ一般にオンラインサービスされていないファイルの例として、CAS の発行する CAS Source Index (CASSI) がある。印刷版 CASSI は CAS が CA に収録した雑誌の辞書として化学系の図書館で広く利用されている。図-12 は CASSI に掲載された色材協会誌の項であるが、雑誌名 (a), CODEN (b), ISSN (c), 雑誌名の英訳 (d), 本文の言語 (e), 抄録の言語 (f), 創刊年月 (g), 発行頻度 (h), 発行者とその所在地 (i), 世界のおもな所蔵図書館 (j) が載っている。現在 CASSI は一般のオンラインで公開されていないが、化学情報協会が筑波大学と協同で試験的に動かしている。雑誌名から CODEN の調査、CODEN から雑誌名の調査に文献検索の補助ツールとして使えるほか、雑誌の発行状況の調査や、雑誌名中のことばで検索し関連雑誌を見つけるなど多様な利用が可能である。

また、オンライン化学用語辞書はないものかと思われる方もあるだろう。この種のものとしては、CA 巻末索

図-12 CASSI の内容

引を利用するための辞書であり、印刷版はオンラインによる文献検索、化学物質調査に不可欠なツールとしてすでに広く使われている CA Index Guide が、コンピュータ可読なデータとなっている。DIALOG では CA SEARCH のファイル中に検索補助機能としてこの Index Guide を一部内蔵させているが、独立した化学用語辞書としてはまだオンラインサービスされていない。

この他にもいろいろな発想があるかと思うが、こんなことがオンラインでできたらということがあればぜひお伺いしたい。

9. 海外のオンラインと日本

コンピュータの発達と同時に通信技術も発達し、今や国際通信網も普及している。米国 Lockheed 社の DIALOG システム、SDC 社の ORBIT システムはすでに国際通信回線 ICAS を使って日本からも利用されているが、この他に米国には BRS 社や NLM のまだ日本では利用されていないオンラインサービスもある。一方、ヨーロッパには EURONET の DIANE をはじめとするシステムがある。海外のシステムには日本で利用できるものとできないものがあるが、ファイルという立場から見るとほとんどのファイルを日本で使うことができるというてよいと思う。すなわち、今や日本で利用できない情報はほとんどないのである。

ところで、このように通信技術が発達した今日、海外のデータベース、システムをそのまま利用するのは容易であり、また一時的には安上りである。しかし、それでは日本は情報の単なるユーザに終わってしまい、その処理技術はたち遅れ、とり残される一方である。すでにご紹介した 5 分間で 500 万以上の化学物質を調べ終わってしまう CAS ONLINE は、1965 年から運転し続けている CAS 化学物質登録システムで鍛えられた高度の化学物質情報処理技術の上に成立つもので、初期の研究開発をはじめシステムの維持には莫大な資金が投じられている。こういう技術は自然にわくものでもないし、一夜にしてでき上るものでもない。すぐれた技術の上には、さらにすぐれた技術が実る可能性があるが、何も無い所でただ周囲に頼ろうとするのではそこには何も生れてこない。この意味で長期的にみて、また国際協力の立場から、わが国も単なる情報の use に終らず化学界が協同で情報処理技術の開発を行なうことが必要なことは、読者諸兄も賛成して下さるだろう。

10. おわりに

以上、オンラインでどのような情報が得られるかという観点から、種々のシステムを駆逐でみてきた。実際にオンライン検索をされる時には改めて専用のマニュアルをご覧頂きたいが、最後に 2, 3 の注意事項をあげてこの稿を閉じたい。

まず、オンラインというのは便利なもので慣れると容易なものである。しかし、よい検索をするためにはしっかりした予備調査が必要である。検索語の選択一つを例にとってみても、ファイルの性格を理解し適確なことばを選んでおかないと正確な検索はできない。

第二に、システムには特徴があり同じファイルでも処理方法が少しずつ異なることに注意しなければならない。システムによりコマンドが違うのはもちろんであるが、同じキーワードでもシステムによって扱い方が異なるし、従って検索手法がちよっとずつ違う。

最後に、オンラインという手軽な調査方法が浸透するにつれ、忘れられがちな印刷版サービスを上手に併用するのもまた大切である。調査によっては印刷版サービスの方がうまくいくこともあるので。

筆者の勤務する(社)化学情報協会は、皆様が研究あるいは開発のため適確な情報をスムーズに手に入れられるのをお手伝いすることを主目的としているが、オンライン検索をはじめ化学情報調査については何なりと化学情報協会にご相談いただきたい。

参考文献

- 1) 時実象一：化学の領域, 32 [11], 69 (1978)
- 2) 江里口将夫：色材, 53 [5], 290 (1980)
- 3) 広田勇二：ドクメンケンキュウ, 31 [5], 201 (1981)
- 4) 森川夏子：情報管理, 23 [5], 451 (1980)
- 5) 片岡彰博：情報管理, 23 [2], 145 (1980)
- 6) DIALOG マニュアル
- 7) ORBIT マニュアル
- 8) 森岡祐二：情報管理, 23 [8], 721 (1980)
- 9) 中山和彦, 上田修一：情報管理, 23 [9], 838 (1980)
- 10) 根岸正光：情報管理, 23 [7], 624 (1980)
- 11) 山崎 昶：情報管理, 23 [6], 536 (1980)
- 12) 大山勝弘：情報管理, 23 [2], 132 (1980)
- 13) 石原好一郎：情報管理, 23 [5], 433 (1980)
- 14) JOIS テキスト
- 15) 化学情報協会サービスご案内 (1981)
- 16) NEEDS-IR マニュアル