Library Science No. 4 1966

情報検索の問題点と対策

A Problem in Information Retrieval and the Countermeasure

平山健三

Kenzo Hirayama

Résumé

Analysis of the process of documentation in article retrieval indicates that the greatest problem lies in the process of translation, followed by classification and abstract. It follows, therefore, that the use of a mechanical retrieval system would not offer any fundamental solution to the problem if the machine were to be used solely as a tool for storage and retrieval of information. Hence, for simplifying other procedure, there is an increasing tendency to use facet indexing in mechanical retrieval in place of systematic classification, which takes time.

Performance of the index in three abstract journals using different indexing systems was examined with 13,521 patents. *The Uniterm Index to Chemical Patents* which uses natural language index without any control on terminology had so many retrieval failure that it had low reliability, and was accompanied with much noise. Index of *Photographic Abstracts*, which uses systematic classification of the UDC system, and index of *Abstracts of Photographic Science and Engineering Literature*, which uses descriptors based on systematic classification, were found to give satisfactory retrieval results.

Consideration of a model that explains retrieval functions indicated that natural language index based on the terms collected by merely arranging synonyms, and natural language index or machine language index with functions of 'roles and links' were approximately of the same type as the uncontrolled natural language index. Consequently, retrieval functions of the KWIC index, which extracts index words from the title or from a short phrase that takes its place, was comprehended to be still lower.

Indexing by systematic classification gives good retrieval efficiency but requires technical knowledge for classification process and for finding indexing items at the time of retrieval, so that it is not suited for processing a large amount of information.

Indexing by descriptors determined on systematic classifications is a median form of the foregoing two methods. Although it requires time and effort to determine the descriptor terms for compiling a thesaurus, classification procedure becomes easier once the descriptors have been determined, and this might be a useful indexing for mechanical retrieval.

(Fuji Photo Film Research Laboratories, Ashigara)

I. まえがき

情報検索は情報に関する技術であるにかかわらず，情報検索がかならずしも統一された意味内容で議論されていないことは，いささか皮肉なことである。

ここでは情報検索の定義あるいは分類をすることが目的ではないが，話の行き違いを避けるために，検索の深さないしは情報の広さから情報検索を次にあげる三つに区分し，現在もっとも多く問題になっている第2の場合を中心にして考察を進めていく。

1. 書物単位の情報検索
2. 論文・記事単位の情報検索
3. 論文・記事中のデーターの情報検索

なお，すべての場合がこの三つにちょうど該当するとは言えず，たとえば，記事中の一部分の情報が必要であることには始終出あうのであるが，これとても第2の場合の一事例とみなして話を進めていくのに，なんら妨げとなるものではない。

II. ドキュメンテーションの行程

第2の場合の情報検索に論及するためには，まず情報検索が記事検索のためのドキュメンテーションの全行程中でしめている役割を明らかにしておいたほうがよい。

第1表の左第1欄に示したのは論文・記事単位のドキュメンテーションの比較的複雑な場合の行程であって，IR (Information Retrieval) は主としてHとKの段階の機械化したときを言っているが，まえがきで述べた第3の場合も含んでいることがある点に留意する必要がある。

第1表にあげた各行程に必要な人員数がわかると，人手のかかる行程を機械化して隘路を広げることによって能率をあげることもできるし，またそのような行程に人手を多く要しない種類の資料を処理することによって，ドキュメンテーションの効果的で新しい分野が開拓できる可能性もあるゆえ，類型的な場合の要員を明らかにしておくことは重要である。第1表は論文・記事単位の行程概略である。

第1表　ドキュメンテーションの行程概略

行　程	年間処理件数	要員概数	必要知識
A 原資料検討	2000～3000誌と特許		
↓			
B 資料収集	500 誌(20%)と特許		
↓			
C 情報選択	1.5 万件/年	2人	専　門
↓			
D 処理手配	1.5 万件/年		
↓			
E 翻　訳	0.5 万件/年	12人	専　門
↓			
F 抄　録	1.5 万件/年	6人	専　門
↓			
G 分　類	1.5 万件/年		
↓			
H 蓄　積	1.5 万件/年	2人以下	非専門
↓			
J 質問分析	10 件/日とする	0.5人	専　門
↓			
K 検　索	30 万件×10/日	0.2人	非専門
↓			
L 余剰情報除去		僅　少	専　門
↓			
M 資料入手		方法により種々	非専門

第1表の年間処理件数と要員概数は，化学工業分野の一会社に必要な技術情報を専門誌と特許明細書から取出して，情報検索のための情報処理をし，検索する場合の数である。分野が異なり，会社の規模が違い，または公共の研究機関の場合には，これらの数はもちろん変ってくるが，以下の検討に必要なのは，これらの数の絶対数よりもむしろ相互間の割合である。

この行程を簡単に説明すれば，一化学工業分野では，関連分野も含めて必要な専門誌は二・三千誌もあるが，そのうち重要なものから20%すなわち500誌を情報処理の対象として選ぶと，その中に必要記事が大体70%くらいまで掲載されている。

この500誌から必要記事を選び出し，また必要な特許も選出する。さらにこれらの500誌以外の専門誌に掲載されている必要記事は二次資料を通じて選ぶ。このように情報選択をして得られる記事数（論文と特許の数）は一・二万件である。

このように選択した記事（1.5万件あるとしよう）は，使用国語・内容等に応じてさらに後の処理の手配をする。

上述のような選択をし，手配をするために，あるいはさらに情報の処理をするためには内容を読む必要がある

が，化学技術分野の論文の用語は1/2が英語，1/6がロシア語，1/10がドイツ語・日本語，1/15がフランス語というのが現状である。日本語と英語とドイツ語（またはフランス語）三カ国語は翻訳をしなくてもよいとすれば，残り1/3すなわち約千件の記事は翻訳の行程が必要である。

読める言語で書かれている記事と，翻訳したその他の記事を抄録し(抄録の必要なことはあとで述べる)，分類をきめる。分類はＵＤＣ程度の詳しさの体系的分類をすることと仮定する。記事内容の部分的性質であるファセットの組合わせによる分類も当然考えるべきであるが，これについては第5章以下で詳しく述べたい。

このように加工して得た情報（抄録・分類）を記事の出典・題名・著者名などとともにカードその他の形にして蓄積する。そうして雑誌等の原資料は別に整理保管しておく。

ついで調査をするときには質問内容を分析し，蓄積情報を分類してある分類体系のいずれの分類項目に質問が対応しているかを明らかにし，蓄積情報の中から該当分類項目に含まれている情報を選び出す。

このようにして得られた情報は，多くの場合不要情報まで含んでいるので，記事内容を抄録によって判断し必要情報のみを選出する。この操作をするには多くの場合表題だけでは不十分であって，抄録はこのために必要なのである。

このようにして必要情報の出典がわかれば元の資料，またはその複写入手の手配をすることができる。

これらの行程に必要な人員のうち，少なくとも情報の内容の理解とか，記事の発表に使われている国語の理解を必要とする段階（第1表のＣ～Ｇ）に専門家が20人もいるのに対し，蓄積・検索の行程では僅か2人以下の非専門家でよい。

もっとも，この蓄積・検索の行程は，使う道具によって穴なしカード，ハンドソートパンチカード，機械ソートパンチカード，電子計算機の四方法に大別でき，その方法によって要員を減少させることもできようが，それにしても非専門家2人が0に近づくにすぎないので，記事検索の大きな問題はさらに他のところにある。

その問題を検討するに当たって，まず種々の蓄積・検索方式の長短について述べておこう。

III. 情報検索の手段とその特質

情報検索の道具として現在実用にされ，またされつつあるものを大別すれば次のとおりである。

(a) 記事または情報ごとに記録
- (a1) 穴なしカード
- (a2) ハンドソートパンチカード
- (a3) 機械ソートパンチカード(ＰＣＳ)と類似の方法
- (a4) 電子計算機

(b) 索引項目ごとに記録
- (b1) 一覧表 (Uniterm Index を含む)
- (b2) ピーカブー

これらの諸手段のうち，(a1)～(a4) をドキュメンテーションの立場から見ると次にあげるような機能的差異がある。

手段		機能的差異
(a1)	穴なしカード	一元的索引・常時配列
(a2)	ハンドソートパンチカード	多元的索引・多数同時検索
(a3)	機械ソートパンチカードと類似の方法	多元的索引・逐時検索
(a4)	電子計算機	多元的索引・多重逐次検索または随処検索

上記の機械的差異を簡単に検討しよう。

(a1) 穴なしカードは，一枚のカードを一カ所だけに配列できるから一元的にしか索引できない。したがって一つの記事がＡ，Ｃ，Ｅ，Ｇの4項目のいずれからも検索できることが必要ならば，カードを4枚作ってこれら4項目に配列しておかねばならない。しかし平常から一定順序にカードを配列してあるのだから，検索にもっとも時間がかからない点は大きな特長。これに記事索引を体系的分類でやるときのように，1件当たりの索引数が少なくてよいときには都合がよい。その単純な例は図書の索引カードであって，書名・著者名・内容分類など複数観点からの索引を穴なしカードで作り一定順に配列しておけばよい。

これをハンドソートパンチカードや機械ソートパンチカードや電子機械でやっても，決して早く，あるいは経済的に検索することはできない。

また“Ａ+Ｃ+Ｅ”に該当する情報を検索するときには，項目Ａ，Ｃ，Ｅのうち含んでいる情報数(その項目のところに配列されているカード数)の最少の項目，仮にそれがＥであるとすれば，項目Ｅのカードのうち Ａ と Ｃ も含んでいる情報，すなわち“Ｅ+Ａ+Ｃ”あるいは“Ｅ+Ｃ+Ａ”に該当する分類項目のついたカードを視覚的に捜し出すことになる。この操作は手動でかつ視覚に

たよるゆえいかにも原始的に見えるが，1 件当たりの索引項目が少なく，かつ検索項目を表わすのに多数の索引項目の組合わせを必要としないときには能率的な方法であって，情報が常時一定順序に配列されているから検索時間が非常に短いこと，特殊な検索装置がいらないことは大きな長所である。

ただし副出カードを何枚か作るゆえカードの所蔵場所がやや広く必要であり，副出カードを作る手間がかかるが，後者は最近の複写技術の進歩，たとえばゼロックスの出現により簡単になった。要するに，一般に考えられているより有利かつ実際的な方法である。

(a2)　ハンドソートパンチカードは一枚のカードを複数項目から検索できるから多元的索引の性能を備えている。この点，機械ソートパンチカードや電子計算機と同様である。その索引項目数は概して機械ソートパンチカードより少ないが，コードの種類を自由に選択できる点を考慮に入れると，大きなカードであれば索引項目数は機械ソートパンチカードに匹敵する。

しかし検索速度がおそい欠点があるが，それでも一回に 100～200 枚程度の処理ができるので，機械ソートパンチカードの数分の一の程度の速さである。この点再認識される必要がある。1 人 1 日の検索枚数は数万枚。

またパンチは人手でやるので機械ソートパンチカードよりおそいが，一記事あたりの索引項目数が多いときは穴なしカードで副出カードを多数作るよりも早くかつ経済的にできる。

さらにまた，機械ソートパンチカードや電子計算機のように装置に費用がかかり，装置の設置場所でないと実施できないという制約がない点は，穴なしカードと同様につごうが良い。

記事数，すなわち情報数について言えばあまり多くないとき，精々数万までのときは実用になる。しかし記事数が仮に十万あったとしても，10の大項目に区分しておけば，1 回の検索は 1 万枚程度であるから実用できる。したがってこの方法は，研究所のような比較的大きな団体であっても，その中で主題により分れたグループ（たとえば研究室，研究班）ごとに蓄積・検索するのに実際的な方法である。

(a3)　機械ソートパンチカードを使う方法（ＰＣＳ）は多元的索引機能を備えている点はハンドソートパンチカードや電子計算機と同様である。検索速度がハンドソートパンチカードより早いが，一枚ずつ検索するゆえ 1 分間 500～1000 枚程度である。

パンチ速度がハンドソートパンチカードより早く，パンチの検査もほぼ同速度でできるが，装置のために費用がかかる。さらに仕分けして得られたカードには抄録等の記録がないため，そのカードの記事が要求にあうものかどうかの判定がむずかしい。それでさらに抄録集のようなものに当たって原記事の要否を決定する必要がある。

仕分け機の速度は 1 日 (7 時間) 20～40 万枚。

類似の方法に，穴の機械的検索の代りに，写真フィルム片に記録した黒白のモザイク模様を光学的に検索するのもあるが，ドキュメンテーションの機能の面から見るとまず変りはない。このようなフィルム片を長尺ロールフィルムにしたのもあるが，一情報の記録量を長くできることと，連続体であるため検索速度が早くなるが任意部分の検索が自由でない点以外に，機能の面で本質的な違いはない。

(a4)　電子計算機には種々の形式のものがあるが，いずれも多元的索引機能をそなえており，多重逐次検索（逐次検索をいくつか平行してやる方法）または随処の検索ができ検索速度が非常に速い点，複雑な論理計算ができる点の特長がある。

しかし記事検索にはそれほどの検索速度を必要とせず，またさほど複雑な論理計算も必要としないことが多い。たとえば，“ＡでありＢでありＣであるがＤでなく，ＥまたはＦである”というような形の記事検索はまずないし，よし，かりにあったとしても，“ＡでありＢでありＣである”ものを検索し，それに合った情報の中から“Ｄでなく，ＥまたはＦである”ものを検索しても時間的に十分で，また経済的にはそのほうが引合うことが多い。

全情報のうち“ＡでありＢでありＣである”情報が多すぎるならば，それはむしろ分類体系に問題がある場合が多い。

このように，記事検索には複雑な論理計算ができるよりも，簡単な仕分け能力を要求されることが多く，記事検索では前に述べたように検索速度に最大の問題があるのではなくて，検索以前の段階である蓄積のための記事の分類（検索用の記号づけ）操作に問題があるのである。

以上述べたのはいずれも記事または情報単位に記録する方法についてであるが，次に索引項目ごとに記録する諸手段 (b1)～(b2) の機能的差異をドキュメンテーションの立場からあげる。

手段	機能的差異
(b1) 一覧形式	常時配列・二個併用で二項目の対照検索
(b2) ピーカブー	常時配列可能・多項目同時検索

次に上記の機能的差異を簡単に検討しよう。

(b1) 一覧形式のもっとも簡単なのは一冊の本の巻末索引を考えていただくとよい。ある一定順序（多くはアルファベット順または五十音順）に索引項目が並んでいて，それぞれの項目にはそれに関して記述のあるページがでている。

これは一冊の本でなくて多数の資料の索引としても使える形態である。資料ごとに番号をつけておいて索引にはその番号を出しておけばよい。このようにあらかじめ列挙しておく形式を list-up 方式ともいうが一覧形式ということにほかならない。

もし“秋の食物”について調べたいときには，一覧形式の同じ索引を二つならべ，一方は“秋”を，他方は“食物”の項をあけ，両方に出ている資料番号のうち同じものを捜し出せばよい。*Uniterm Index to Chemical Patents* はこれに類したものである。この方式だと

（1） 索引しようとする概念がどういう索引語で表わされているかわからないことが多い。たとえば“秋”についての情報を捜すために“autumn”を索引したがそういう索引項目がない。そこで“fall”を引いたところ六つの資料番号があったのでそれぞれの資料を引き出して読んでみたところ，秋についての記事二つと滝についての記事四つが見つかった。それで秋についての記事がそれだけだと思っていたら“season”の項目に多数該当するものがあった，ということもありうる。これらの問題点は自然語を使うことによるものである。充実した用語集 (thesaurus) が必要なゆえんであるが，索引語をいくら規定しておいても使用者が思いつくことばまで十分に規定することがむずかしいところに一つの限界がある。

（2） 索引語さえ見つかれば資料番号は早くわかるが，必ず資料を取出してみなければ，はたしてその資料が適当なものかどうかわからない。穴なしカードやハンドソートパンチカードのように索引の手段であるカードに抄録が記載してあれば，それを読むと不要記事が相当精密に淘汰できるし，機械ソートパンチカードでも表題の程度はカードにタイプ印字しておける。

（3） “秋”と“食物”の両方の項目に出ている資料番号の資料を調べると，前半には“秋”の天候によって農作物が影響をうけること，後半には，したがって冬の“食物”が変ってくることが記載してある。しかし“秋の食物”については何も書いてないかもわからない。“秋の食物”について記述がないということは丹念に読んでみないとわからない。一般に“ない”ことの証明は非常にむずかしい。

上記の難点はファセット分類法に起因するものであって，これを解決するために一資料の記事をいくつかに分けて扱うこともあるが，程度は軽減されるにしろ各部分については上記の難点がやはり残っている。

また“秋”と“食物”で索引された資料が実は“秋にない食物”について記載したものであるかもわからない。索引語に否定か肯定かを必ず結びつけて表わすことは普通されていない。

さらに“教育法の歴史”を“教育法”と“歴史”で索引できるようにしても，それが“歴史の教育法”や“歴史と教育法”あるいは“歴史または教育法”，“歴史でなくて教育法”などと解釈して索引されることもある。role and link の考えはこのような場合の解決に役立つ。

（4） 本の索引のような一覧形式であるから，新しい資料番号を適当な索引項目の所に追加し，また一定の順に配列記載してある索引項目の間に新しい索引項目を追加記入することは一般にできない。いつも最新の情報まで索引できるようにしておくためには始終改訂せねばならない。

（5） A項目の索引の網羅度が 50% (洩れが 50%)，B項目も同様だとすると“A の B”（たとえば“秋の食物”）という項目の索引の網羅度は $0.5\times0.5=0.25$ しかない。すなわち必要情報の 1/4 しか見つからないことになる。これはファセット分類法に内在する欠点である。

Uniterm Index 形式の索引の問題点は上記のようなことである。

なお Uniterm Index 形式を他の検索手段，たとえば磁気テープに記憶させて機械検索することもおこなわれている。しかし単に索引手段を変えただけであるならば得られる情報はまったく同じであって，費用をかけて機械検索という形をとっても Uniterm Index の欠点は相変らずそのまま残っているわけであるし，自然語索引に原因する問題点も解決されない。原情報と，検索の結果えられる情報との関係は次の模型図のとおりであって（ロ）→（ハ）の過程をいくら変えても（ロ）が表現していない情報は（ハ）では得られない。

（イ） 原情報
↓
（ロ） Uniterm Index の情報
一覧形式 ↓↓ 機械検索
（ハ） 得られる情報

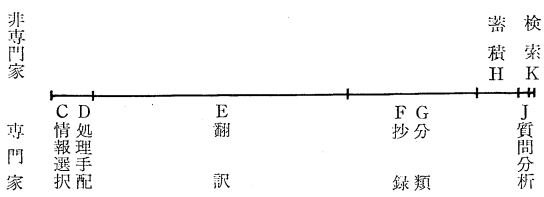
(b2) ピーカブー。Uniterm Index にはいくつかの問題点があるが，そのうち (4) の改訂の不便さを除くためには索引項目ごとに別のカードに記入配列しておけばよい。カード式索引である。

この形式にするとき，カードに縦横各 100 目盛の方眼を印刷し座標をきめておくと，位置で 0 から 9999 までの番号を示すことができる。上記のカード式索引に資料番号を書く代りに，この方眼カードの該当する目盛の位置に穴をあけるようにしたのがピーカブーである。“秋”のカードと“食物”のカードを重ねて見て，裏までのぞける穴にあたる番号の資料が求めるものだというのであるから，前記の Uniterm Index の問題点は (4) 以外すべてピーカブーにもあてはまる。

IV. 情報検索の問題点と対策

第 1 図は第 1 表の各行程を必要人員·時間(man·hour)の割合のグラフにしたものである。第 1 表または第 1 図からわかるように，記事検索の全行程のうち，専門知識を要する行程 C～G に約 20 人を必要とするに対して，記事内容に関する専門知識を要しない蓄積・検索行程 (H, K) に必要な人数は，蓄積·検索の方法が違えば変るにしても，せいぜい 2 人程度でよい。したがって相当の費用をかけてこの蓄積・検索行程を機械化しても 2 人以上の人員減を望むことはできないから，この点だけを見ると機械化が有利だとは言えない。

だからといって機械化が不要だというわけではない。情報量の増大に対処するために機械化は是非必要であるが，上記の点を理解したうえで機械化の効果を発揮させる方法を考えるべきであると同時に，ドキュメンテーションの全行程から考えて隘路の解決をはかるべきである。次に情報検索に関する目下の大きな問題と対策について述べよう。



第 1 図 ドキュメンテーション各行程の要員比較

（a） 第 1 図からわかるように翻訳(E)の行程の短縮が最大の問題である。現在この問題点が痛感されていないとすれば，必要な情報であっても読解しがたい国語で書かれた記事は黙過されているからであろう。翻訳行程の短縮には自動翻訳機の実用化が必要であるが，翻訳の完全自動化は現段階ではまだむずかしい。翻訳の際に最終的に加えねばならぬ人手をなるべく減らし，またその人に要求される専門知識（記事内容と使用国語との両方の専門知識）の水準をなるべくさげられるように翻訳機を使うことを目標にするのが実際的であろう。

（b） 翻訳についで抄録(F)と分類(G)の自動化が大きな問題であるが，自動抄録は現在初歩的な試験段階である。自動化がまだ実用化していない現段階では，抄録なり分類の方法を簡易化する工夫も必要である。

抄録にしても，抄録雑誌の抄録や，論文につけられる抄録と記事検索のための抄録とは目的が異なり，その形式もおのずから違ってもよいはずである。記事検索のための抄録は，いわゆる電報文型抄録で十分であり，またそのほうが作りやすい。

分類の簡易化については，体系的分類のほかファセット的分類法を使うことも考えられるが，これについては次章以下でくわしく検討する。

（c） 情報利用者側からいえば，上記のE～Gの行程を短縮するために，他所でこの行程の処理加工をした情報で利用できるもの，たとえば抄録誌・索引誌等をできるだけ利用することは有意義である。ただし，このときは自家処理する人手が節減できる代りに，抄録誌等が手にはいるまで情報の利用ができないという犠牲をはらっていることになるが，日本の企業体などで必要情報を全部自家加工できるだけの人員を擁しているところはまずないといってよい。

（d） 記事検索ではE～Gの段階に多くの人手がいる事が情報処理上の一つの隘路になっている。したがって機械検索をするならば，E～Gの行程になるべく人手を要しないような種類の情報を扱えば相対的に機械検索の効率を高めることになる。数値データのようにすぐ蓄積に持ちこめる情報は，機械検索への適性を備えていることになる。

また，記事検索の場合でも F～G の段階を (b) (c) で述べたような方法で簡易化すれば，蓄積・検索行程の機械化の意義を相対的に高めることになる。

（e） 単なる記事検索の場合，蓄積・検索の機械化は

必ずしも有利でないことは前述した。したがって蓄積・検索の機械化をするには単なる記事検索ではなく，機械の特長がいかせるような場合にすべきである。

たとえば，一情報につける索引項目の多いものほど検索手段 (a1)～(a4) のあとのほうが良い。一元的索引である穴なしカードでは蓄積行程で副出カードを多く作る必要があり，またハンドソートパンチカードと機械ソートパンチカードでは索引の多元性に関しては比較的低いところに限度があるのに反して，電子計算機は多元性の高いものに適している。

したがって，有機化合物を全構造とあらゆる部分構造のいずれからも検索できるようにするには電子計算機が良い。ただし，全構造のみのときは穴なしカード，全構造と限られた少数の部分構造のみから検索するときは穴なしカードやハンドソートパンチカードのような単純な方法で解決できる。

さらにまた分類法との関連を考えるならば，ファセット分類によると一記事につけられる索引項目が概して多く，したがって検索方式も複雑にならざるをえない。

蓄積·検索の機械化の特長をいかすもう一つの方式は，機械のもっているその他の性能も生かして使うことである。たとえば作表・複写などである。次に作表をうまく使った例をあげよう。

今月集った情報を情報ごとに機械ソートパンチカードで索引カードにしておき，索引項目のアルファベット順にカードを配列させてアルファベッド索引を印刷配布する。ついで来月末には来月分の索引カードと今月分の索引カードを一括して改訂アルファベット索引を印刷し今月のと取換える。さらに来々月も1カ月分のカードを追加するだけで3カ月分の総合索引を作り前月のと取換えさせる。このように索引カードは1カ月分作るだけで毎月総合索引を改訂して前月のと取換えさせると商売としてはうま味がある。

12月に年間索引ができるまでこれを続けるとすれば，1月分の索引は12回，2月分のは11回，……というふうに売れるわけだから年間に78カ月 (6年半分) の索引を売りつけることになるが，索引カードはただ12月分作るだけでよい。

最近米国製で市販されている索引にこの形態のものがある。*Uniterm Index to Chemical Patents* (2カ月ごと)やピーカブー形式の *Termatrex* はこの例であって，消費させる米国経済界の行き方の良い例である。

V. 索引法の比較検討

第IV章で述べたように記事検索では分類の簡易化が一つの問題である。機械検索が高速化するほど，機械に記憶させるべき情報の作成，たとえば情報の分類が急がれるようになった。その結果，記事情報全体として，既成分類体系のどの概念に該当するかによって記事を分類する従来の体系的分類法と違って，記事情報中の一部分の概念を表わすことば(またはその概念を表わす記号)いくつかを組合わせる分類法（ファセット分類法）がおこなわれるようになった。ディスクリプタによる分類はその一種である。また概念に方向性をもたせるためにファセットに role indicator をつけたり，概念の結びつき方を示す link の考えを導入する事もおこなわれているが，根本的には組合わせ分類法である。

このような分類法の効率については従来しばしば検討されてはいるが，検索で得た情報のうちの余分な情報(ノイズ) の検討が大部分であって，検索洩れの検討は母集団の全数検査が必要なためほとんどされていない。

それで，ここでは自然語を使った索引の性能を調べ，さらにそれと，体系的分類法を使った索引と，相当統制された用語集による索引との性能を比較した。

自然語による索引としては *Uuiterm Index to Chemical Patents* を使い，調査対象とした母集団は1962年の米国特許のうち上記 *Uniterm Index* に採録されている13,521件とした。

まず，“ハロゲン化銀を使ったカラー写真”に関する特許を調べるために *Uniterm Index* で使っている索引語を調べると“color”と“photography”と“silver compounds”と“halides”があるので，この4語いずれでも索引されている特許を調べたところ3件だけあった。

第2表の keywords の欄のうち，color photography と silver compounds と halides の4欄とも ○ 印のものがそれであって，U. S. Patent No. 3,062,647 と 3,062,652 と 3,062,653 (特許番号は第1欄右列) である。そこでこの3件に当たってみたところ，該当するものは最後の1件だけで前二者は不適当な特許であった。ノイズが 2/3=67% である。

しかし上記の主題の特許が年間1件とは少なすぎるので，母集団 13,521 件の全数調査をしたところ該当するものが37件もあった。第2欄 color photography us-

第 2 表　Uniterm Index による“ハロゲン化銀カラー写真法”の検索

U. S. Patents Class	No.	COLOR PHOTOGRAPHY USING SILVER HALIDES	KEYWORDS* Color	Photography	Silver compounds Halides (Decision)	Silver halide (Decision)	Silver chloride	Silver bromide	Silver iodide	Silver bromoiodide	TITLE
96-29	3,015,561	yes	o	o	1	o +					Novel Photographic Color Process and Products
96-29	3,019,104	no	o	o		o n		o			Photographic Products, Processes, and Compositions
117- 8	3,019,124	yes	o	o	1	o +					Multicolor Photosensitive Film and Process of Making the Same
96-30	3,022,164	yes	o	o	1	o +					Reproduction of Color Drawings, Film Transparencies and Photographs
96- 9	3,028,237	yes	o	o	1	o +					Masking of Cyan Images in Color Photography
96-53	3,028,238	yes	o	o	1	o +					Color Photography
96- 3	3,032,413	yes	o	o	1	o +					Color Photographic Processes and Materials
96- 3	3,034,890	yes	o	o	1	o +					Color Formation [in Diffusion Transfer Process]
96-55	3,034,891	yes	o	o	1	o +					Procedure for the Production of Yellow Dye Images by Color Development
96-55	3,034,892	yes	o	o	1	o +					Magenta-colored Cyan-forming Couplers
96-99	3,034,894	no	o	o		o n					Hardening of Gelatin
96-27	3,035,912	no	o	o	o						Process of Recording
96-27	3,035,913	yes	o	o	1	o +					Photographic Tone Correction
96-55	3,035,914	yes		o	1	o 1					Prevention of Cyan Dye Fading in Color Developed Prints and Films
96-73	3,038,802	yes	o	o	o 1	o +					Photographic Color Element with Novel Cyan Dye
96- 3	3,039,869	yes	o	o	1	o +					Photographic Color Processes and Compositions
96-60	3,042,520	yes	o	o	1	1					Bleaching Bath for Processing Color Film
96-29	3,043,689	yes	o	o	1	o +					Novel Photographic Products and Processes
96-29	3,043,692	yes	o	o	1	o +					Photographic Products and Processes
96-55	3,043,694	no		o							Novel Class of 3-Indazolinone Developing Agent
96-29	3,044,873	yes	o	o	1	o +					Photographic Products and Processes
96-67	3,044,874	no	o	o		o n		o	o		Photographic Materials
96-55	3,046,129	yes	o	o	1	o +					Sensitization of Photographic Silver Halide Emulsions Containing Color-forming Compounds
96- 9	3,047,385	yes	o	o	1	o +					Production of Color Photographic Images
96-29	3,047,386	yes	o	o	1	o +					Anthraquinone Dye Developers [in Diffusion Transfer Process]
96-55	3,047,388	yes	o	o	1	o +				o	Color Photography
96-82	3,047,390	no	o	o		o n					Method for Optical Bleaching Coated Papers
96-84	3,048,487	yes	o	o	1	o +					Basic Mordants Derived from the Reaction Between Maleic Anhydride Interpolymers and Disubstituted Diamines
96-84	3,050,393	yes	o	o	1	o +					Filter Layer for Photographic Elements
96-100	3,050,394	yes	o	o	1	o +				o	Method of Incorporating Color Couplers in Hydrophilic Colloids
96-20	3,053,655	yes	o	o	1	o +					Photographic Material and Process
96-55	3,056,674	yes	o	o	1	o +	o	o			Color Formers for Producing Yellow Dye Images by Color Development
96-55	3,056,675	yes	o	o	1	o +	o	o			Benzoyl Acetanilide Couplers
96-111	3,058,827	no	o	o		o n					Dialdehyde Starch as Gelatin Hardener
96-66	3,060,028	no	o	o		o n					Stabilized Photographic Silver Halide Emulsions Containing Iodine Complexes of Poly-N-vinyl-2-oxazolidinones
96-77	3,060,029	yes	o	o	1	o +					Photographic Ultraviolet Absorbers
96-29	3,061,428	yes	o	o	1	o +				o	Photographic Products and Processes Using Alkali Permeable Co-polymeric Layers
96-55	3,061,432	yes	o	o	1	o +					Pyrazolinobenzimidazole Color Coupler
117-34	3,061,453	yes	o	o	1	o +					Process for Incorporating Photographic Reagents in a Photographic Element using a Common Solvent and a Preferential Solvent
96-66	3,062,646	no	o	o		o n					Sensitization of Silver Halide Emulsions with Macrocyclic Compounds
96-66	3,062,647	no	o	o	o o n						Photographic Emulsions Containing Colloidal Material and Alkylene Oxide Polymers
96-99	3,062,652	no	o	o	o o n						Hardening of Gelatin with Oxy Plant-gums
96-100	3,062,653	yes	o	o	o o +	1					Photographic Emulsion Containing Pyrazolone Magenta-forming Couplers
96-29	3,065,074	yes	o	o	1	o +					1,4-Benzoquinone Oxidizing Agents for Color Transfer Processes
96-22	3,068,097	yes	o	o	1	o +					Developers for Color Photography Containing Sulfite Ester Polymers
96-29	3,069,262	yes	o	o	1	o +				o	Processes for Forming Dye Developer Images Having Stability in Sunlight
96-29	3,069,263	yes	o	o	1	o +					Photographic Products and Processes Using Alkali Permeable Polymeric Layers
96-29	3,069,264	yes	o	o	1	o +					Photographic Products and Processes Using Alkali Permeable Copolymeric Layers
When coordinated with Color and Photography	Retrieval failure	Ratio of no.			36/37	3/37					
		Percentage			97%	8%					
	Noise	Ratio of no.			2/3	7/42					
		Percentage			67%	17%					

* Note　o indexed keyword
+ required information
1 retrieval failure
n noise

ing silver halides が yes となっているものがそれである。

したがって先の調査では37件中36件，すなわち97%も検索洩れがあったことになり，これでは索引としての機能が非常に悪いというほかないであろう。（第4欄右列に1とあるのがこの検索洩れ，＋が要求にあったもの，nがノイズである。）

しかし検索洩れが97%もあるということは，他の索引語が使われている可能性がある事を示しているので，該当する特許，すなわち第2欄が yes となっている37件の特許につけられている索引語を逆に調べた結果，大部分が“color”と“photography”と“silver halide”で索引されていることがわかった。

この3索引語で検索される特許は，第3欄の color と photography と第5欄の silver halide のいずれにも○印のついたものであって，その結果は第5欄右列にあるとおり検索洩れは3/37＝8%，ノイズは7/42＝17%である。これだとある程度検索に使用できる。

しかし上記の事実でもわかるように自然語による索引では検索したい概念が何という語で索引されているかわからない。仮に，ある索引語から必要情報が100%検索でき，ノイズが0%であるように索引ができていても，その索引語がわからなければその索引は役立たないわけである。したがって上記の例のように一つの概念に対する索引語が奔放に分散していては，利用者としては施すすべもないのである。

上例のように“silver halide”が“silver compounds”＋“halides”という索引語の組合わせにおきかえられているものがあるところから，さらに他の語に索引されている可能性も当然考えられるので全索引語を調べたところ，必要な37件の特許のうちには“silver chloride”や“silver bromide”や“silver bromoiodide”で索引されているものもあり，そうかといって，これらを組合わせたものが必ずしも必要情報だけを示していないこともわかった（第2表第6欄）。

このように索引語のとり方が一定していないので，それでは一つの特許にどのように索引語がつけられているかを調べたのが第2図である。

左の original text の部分は特許明細書の原文のうち索引に関係のある部分を抜き出したもので，右欄の keywords はこの特許につけられている *Uniterm Index* のキーワード32個である。

これらのキーワードのうちA番号のもの5個は特許番号・発明者・特許権者のように機械的につけられるもの5個である。

B番号のもの14個は包括概念が広すぎて索引語として不適当なものである。たとえば solvent（溶剤），reaction（反応）または reactivity（反応性），dispersion（分散），heterocyclic（複素環という化合物の大きな種類），layer（層）などで，母集団13,521件のうちこれらの索引語がつけられているのは，それぞれ1735件，897件，890件，548件，501件に及んでいる。13,521件中1735件すなわち13%も同じ索引語がついているようなことで，はたして索引としての機能が十分果たされるのであろうか。もっとも多数の特許につけられている索引語は“water, aqueous”というので13,521件中実に3,626件（27%）に及んでいる。

C番号のもの7個も，それ自体では包括概念が広すぎるが，他の索引語と組合わせて索引に使いうると思われるもの。photography（写真），emulsion（乳剤），color（色）または coloration（着色），silver halides（ハロゲン化銀）または silver salts（銀塩），halides（ハロゲン化物），stability（安定性），thermostability（熱安定性）。

D番号のキーワード3個は化合物名であって，これを索引語に採用するならば，同程度あるいはそれ以上に重要なF番号のもの10個も採用しなければならない。しかし，一般に化合物名は一化合物に対して数個，ときには数千個も違ったものをつけることができ，索引の利用者はどの名称から索引するか予想がつかないゆえ，化合物名索引というものの価値は非常に低い。ことにこの特許にあるような複雑な化合物の場合はそうである。たとえばF37の化合物に普通つけられる可能性のある名称が少なくとも1,120個はあり，そのうち国際的命名規則からみて正しいと思われるものでも48個ある。

最後に，E番号のついた3個はたしかにこの記事の内容を索引するのに適当と思われるものである。pyrazolone（ピラゾロンという化合物の種類），magenta（マゼンタ色素），coupler（発色剤の一種）。しかし一方これと同程度に必要と思われるもので索引語になっていないものがある。G番号の4個，coupling（カプリング反応），subtractive process（減色法），anti-printout（抗プリントアウト性），anti-yellowing（抗黄色化）であって，このうち前二者は索引項目があるにかかわらずこの特許の番号があげられていない。後二者はこの特許としては特長としてあげられるべき性質である。

情報検索の問題点と対策

ORIGINAL TEXT

U.S.Pat. 1 3,062,653

2 PHOTOGRAPHIC 3 EMULSION CONTAINING 4 PYRAZOLONE 5 MAGENTA-FORMING COUPLERS

Arnold 7 Weissberger, Anthony 8 Loria, and Ilmari F. 9 Salminen, Rochester, N.Y,, assignors to 10 Eastman Kodak Company, Rochester, N.Y., a corporation of New Jersey

Filed Feb. 18, 1960, Ser. 9,467
9 Claims. (Cl. 96-100)

This invention relates to 11 color photography and particularly to substituted 12 1-phenyl-3-amino-5-pyrazolone magenta-forming couplers and photographic 13 14 silver halide emulsions containing these couplers.

Formation of colored photographic images by 15 coupling the development product of primary aromatic amino developing agents with color forming or coupling compounds is well known. In these processes the 16 subtractive process of color formation is ordinarily used and .....

.....

.....

.....

It is, therefore, an object of our invention to provide a new class of magenta-forming couplers which have good 17-19 stability to heat and light and which can be incorporated in photographic emulsions that will therefore 20 not be subject to print-out or 21 yellowing.

Another object of our invention is to provide a new class of magenta-forming couplers which are readily incorporated in photographic silver halide emulsion 22 layers with standard coupler 23 solvents over a broad span of coupler to coupler solvent ratios, and which 24 react during 25 development with the oxidized 26 developing agent to produce a magenta 26 dye having improved light 27 absorption characteristics.

Another object is to provide couplers which upon development with oxidized developing agents produce dyes whose light absorption curves are shifted toward the shorter wavelengths in such a way as to enable one to avoid excess 28 red absorption and inadequate 29 green absorption of the dyes in current use.

Another object is to provide a new class of magenta-forming couplers which are not only readily soluble in coupler solvents but give 30 dispersions which are free from coupler crystallization at high coupler to coupler solvent ratios.

.....

.....

The following representative couplers will illustrate our invention. However, it is to be understood that our invention is not limited to these specific couplers

Coupler 1

31 1-(4,6-dichloro-2-methoxyphenyl)-3-[α-(m-pentadecylphenoxy)butyramido]-5-pyrazolone

Coupler 2

32 1-(4,6-dichloro-2-methoxyphenyl)-3-[α-(m-pentadecylphenoxy)acetamido]-5-pyrazolone

Coupler 3

33 1-(2,6-dichloro-4-methoxyphenyl)-3-[α-(m-pentadecylphenoxy)butyramido]-5-pyrazolone

Coupler 4

34 1-(4,6-dichloro-2-methylphenyl)-3-[α-(m-pentadecylphenoxy)butyramido]-5-pyrazolone

Coupler 5

35 1-(6-chloro-2,4-dimethylphenyl)-3-[α-(m-pentadecylphenoxy)butyramido]-5-pyrazolone

Coupler 6

36 1-(2,4-dimethyl-6-chlorophenyl)-3-[[α-(m-pentadecylphenoxy)butyramido]-m-benzamido]-5-pyrazolone

Coupler 7

37 1-(2-methoxy-5-methyl-3,4,6-trichlorophenyl)-3-[α-(m-pentadecylphenoxy)-acetamido]-5-pyrazolone

.....

.....

KEYWORDS

A1) 3,062,653 (1)
C2) PHOTOGRAPHY (251)
C3) EMULSION (604)
E4) PYRAZOLONE, 5-PYRAZOLONE (13)
E5) MAGENTA incl. Fuchsin (12)
E6) COUPLER (17)
A7) WEISSBERGER, ARNOLD (1)
A8) LORIA, ANTHONY (1)
A9) SALMINEN, ILMARI F. (2)
A10) EASTMAN KODAK CO. (193)
C11) COLOR, COLORATION (489)
F12) [not indexed]
C13) SILVER COMPOUNDS, SILVER SALTS (37)
C14) HALIDES (396)
G15) [not indexed in COUPLING, COUPLED (141)]
G16) [not indexed in SUBTRACTIVE (2)]
B17) LIGHT, ILLUMINATION, LUMINESCENCE, LUMINOSITY (424)
C18) STABILITY (844)
C19) THERMOSTABILITY (272)
G20) [not indexed]
G21) [not indexed]
B22) LAYER (501)
B23) SOLVENT (1735)
B24) REACTION, REACTIVITY (897)
B25) DEVELOPER, DEVELOPMENT (142)
B26) DYE (498)
B27) ABSORPTION, ABSORBENCY, ABSORBING (126)
B28 RED (78)
B29) GREEN (61)
B30) DISPERSION (890)

D31) 1-(2,5-DICHLORO-4-METHOXYPHENYL)3/ALPHA-(M-PENTADECYLPHENOXY)BUTYRAMIDO/-5-PYRAZOLONE (1) (erroneously indexed)
F32) [not indexed]

D33) 1-(2,6-DICHLORO-4-METHOXYPHENYL)-3-[ALPHA-(M-PENTADECYLPHENOXY)BUTYRAMIDO]-5-PYRAZOLONE
F34) [not indexed]

F35) [not indexed]

F36) [not indexed]

F37) [not indexed]

.....

.....

第 2 図　特許明細書原文とキーワード (1)

EXAMPLE I
Coupler 1
STEP 1. 4,6-DICHLORO-2-METHOXYANILINE HYDROCHLORIDE
In a 3-liter flask, fitted with a stirrer, a condenser, a thermometer, and a dropping funnel was placed 160 g. (1.0 mole) of o-anisidine hydrochloride. The flask was cooled in an ice water bath and 500 cc. of [38]sulfuryl chloride was added in one portion through the condenser with efficient stirring. A stirrable suspension was obtained which changed to a fluffy [39]mass after a few minutes of stirring. .....

B38) SULFONYL CHLORIDE (32) (erroneously indexed. read SULFURYL CHLORIDE (2))

B39) MASS (122)

STEP 3. [40]1-(4,6-DICHLORO-2-METHOXY)PHENYL-3-AMINO-5-PYRAZOLONE

D40) 1-(4,6-DICHLORO-2-METHOXY)-PHENYL-5-PYRAZOLONE (erroneously indexed)

EXAMPLE IV

STEP 4. [41]1-(4,6-DICHLORO-2-METHYLPHENYL)-3-AMINO-5-PYRAZOLONE

F41) [not indexed]

EXAMPLE V

STEP 4. [42]1-(6-CHLORO-2,4-DIMETHYLPHENYL)-3-AMINO-5-PYRAZOLONE

F42) [not indexed]

EXAMPLE VI
STEP 1. [43]1-(6-CHLORO-2,4-DIMETHYLPHENYL)-3-(3-NITROBENZAMIDO)-5-PYRAZOLONE

F43) [not indexed]

EXAMPLE VII

STEP 5. [44]1-(2-METHOXY-5-METHYL-3,4,6-TRICHLOROPHENYL)-3-AMINO-5-PYRAZOLONE

F44) [not indexed]

Not in the text {
B45) EXPOSURE (209)
B46) HETEROCYCLIC (548)

第2図　特許明細書原文とキーワード (2)

索引語として適・不適の判断にはたしかに主観的要素が含まれるにしても，採用されている索引語のうち必要なものの割合が少なすぎることと，一方必要なもので洩れているものが少なくないことは言えるであろう。

次に自然語による索引 *Uniterm Index* と，比較的統制のとれたディスクリプタによる索引を使っている *APSE* (*Abstracts of Photographic Science and Engineering Literature*), 体系的分類である UDC を使っている *Photographic Abstracts* の索引効率を比較した。この *APSE* は前身誌 *Monthly Abstracts Bulletin from the Eastman Kodak Research Laboratories* ではUDCを索引に使っていたのをディスクリプタに切換えたのであるが，そのディスクリプタの選択に当たってはUDCを参照し，ディスクリプタとUDCの6ケタ程度の分類項目との間の対応もつけられている。

調査対象としたのは前記自然語索引の検討と同様に1962年の米国特許のうち *Uniterm Index to Chemical Patents* に採録されている 13,521件であって，“ピラゾロン系マゼンタ色素形成用カプラー”を検索した結果を第3表に示した。

第3欄のIは *Uniterm Index*, II は *APSE*, III は *Photographic Abstracts* の検討結果である。

Iの検索に使った索引語は“pyrazolones”と“fuchsin (including magenta)”と“magenta color formers,” III の検索に使った分類項目は 778.625 “color coupler”と，最近改訂されて 771.726 になったものである。

その結果は

第 3 表　索引効率の比較（ピラゾロン系マゼンタ色素形成用カプラーについて）

U. S. Patents Class	No.	PYRAZOLONE MAGENTA-FORMING COUPLER	I: Pyrazolones	I: Fuchsin incl. Magenta	I: Coupler (Decision)	II: Pyrazolones	II: Magenta color formers (Decision)	III: 771.726 or 778.625 (Decision)	TITLE
88-65	3,015,989	no		o					Light-polarizing Film Materials and Process of Preparation
8-83	3,018,155	no	o						Process for the Manufacture of Solutions [of Azo Dyestuffs]
106-32	3,019,115	no	o						Surface Coloration of Perfluorohaloolefin Polymers and Colorant Composition Therefor
117- 8	3,019,124	no		o	o				Multicolor Photosensitive Film and Process of Making the Same
167-65	3,019,166	no	o						Antiphlogistic and Choleretic Compositions and Process of Therapeutically Using Same
260-163	3,019,217	no	o						Azopyrazolone Dye for Polyester Fibers
96-100	3,021,215	no	o	o	o n				Polythiaalkylenediols as Sensitizers for Photographic Silver Halide Emulsions
96-30	3,022,164	no		o					Reproduction of Color Drawings, Film Transparencies and Photographs
96- 9	3,028,237	no			o				Masking of Cyan Images in Color Photography
260-147	3,030,353	no	o						Conversion Products of Azo Dyestuffs Containing Heavy Metal Bound in Complex Linkage
96- 3	3,034,890	no			o				Color Formation
96-55	3,034,891	no			o				Procedure for the Production of Yellow Dye Images by Color Development
96-55	3,034,892	no		o	o				Magenta-colored Cyan-forming Couplers
96- 3	3,039,869	no		o					Photographic Color Processes and Compositions
96-29	3,044,873	no		o					Photographic Products and Processes
204-12	3,046,203	no			o				Method of Making Copper Cathode Starting Sheets
260-271	3,046,271	no	o						Metal-containing Azo Dyestuffs
96-84	3,050,393	no		o					Filterlayer for Photographic Elements
96-100	3,050,394	no			o				Method of Incorporating Color Couplers in Hydrophilic Colloids
8-46	3,051,542	no	o						Process for Dyeing or Printing Polyhydroxylated Materials
96-55	3,056,674	no			o			o n	Color Formers for Producing Yellow Dye Images by Color Development
96- 1	3,057,720	no		o					Xerographic Color Reproduction
96-99	3,057,723	no			o				Hardening of Gelatin with Oxystarch
167-55	3,057,776	no		o					Schistosomiasis Treatment
260-146	3,057,845	no	o						Complex Copper Compounds of Azo Dyestuffs
96-11	3,058,827	no		o	o				Dialdehyde Starch as Gelatin Hardener
260-173	3,060,167	no	o						Process for the Manufacture of New Metallizable-4-Hydroxy-5-carboxyphenyl-2,2'-dihydroxy-azo-dyestuffs
96-55	3,061,432	yes			o l	o	o +	o +	Pyrazolinobenzimidazole Color Coupler
117-34	3,061,453	no			o				Process for Incorporating Photographic Reagents in a Photographic Element Using a Common Solvent and a Preferential Solvent
96-55	3,062,645	no	o		o				Sensitization of Photographic Emulsions to be Developed with p-Phenylenediamine Developing Agents
96-90	3,062,650	no			o	o			Photographic Printout System Comprising an Organic Azide
96-99	3,062,652	no			o				Hardening of Gelatin with Oxy Plant-gums
96-100	3,062,653	yes	o	o	o +	o	o +	o +	Photographic Emulsion Containing Pyrazolone Magenta-forming Couplers
260-160	3,066,134	no	o						Disazo-pyrazolone Dyestuffs
260-302	3,068,238	no						o n	3-Cyano-4-hydroxy-1,2,5-thiadiazole, Derivatives and Process
Retrieval failure	Ratio of no.		1/2			0/2		0/2	
	Percentage		50%			0%		0%	
Noise	Ratio of no.		1/2			0/2		2/4	
	Percentage		50%			0%		50%	

Note　o indexed keyword
+ required information
l retrieval failure
n noise

	I	II	III
検索洩れ	1/2=50%	0/2=0%	0/2= 0%
ノイズ	1/2=50%	0/2=0%	2/4=50%

であって，Ⅰは必要特許（第3表第2欄が yes のもの）2件のうち1件しか検索できないが，他の二者は検索洩れがない。しかもⅠはノイズが 50% あるが，Ⅱはノイズもなく完全回答である。Ⅲは 2/4=50% のノイズがあるが，これは同誌ではＵＤＣをもっとくわしいところ 771.726.23 “magenta coupler” まで使わないで 771.726 “coupler” で止めているからであるが，索引された4件のうちノイズが2件であるからこの程度のノイズをなくするために全記事のＵＤＣ標数を一ケタまたは二ケタくわしくするより，この程度のノイズを見分ける現状のほうが方法としては実際的であろう。

以上の検討で自然語による索引がきわめて危険であることがわかる。上例では検出された記事数が2件程度であるし，その前の自然語索引の検討では該当特許が37件であるから例としては少数すぎるように見えるが，実は母集団13,521件という多数についての検討結果である。実際に検索する概念の包括範囲が一般に狭く，したがっ

て該当する情報数が少ないのである。

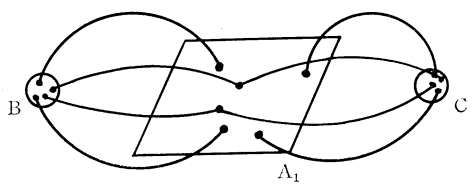
また以上述べた検討結果は単に索引形式に関するもののように見えるが，*Uniterm Index* は自然語索引による機械検索の結果を一覧形式の索引にしたものである点に注意する必要がある。いくら機械検索をしても分類体系(索引形式)の欠点を機械が補うことはできない。

自然語索引に類するものとしてKWIC索引(Keyword-in-context Index)がある。自然語索引は索引語をアルファベット順などの一定順序に配列したものであるのに対して，ＫＷＩＣ索引は一定順に配列した索引語の前後につながることばも一覧できるようにしたもの，言いかえるとコンテキスト(文脈)の中にあるままで索引語を一定順に配列したものである。この点のみに関して言えば自然語索引よりＫＷＩＣ索引のほうがすぐれていることになるが，問題は索引語がどこから抽出されているかという点にある。記事全体から全目的的に抽出された自然語索引でさえ目的達成にはほど遠いことを *Uniterm Index* が示していた。まして必ずしも記事内容を十分に表わしていない記事表題，あるいは表題を若干改めたものから自然語のままで索引語を抽出したＫＷＩＣ索引がどの程度要求を満たしてくれるかは大体推察できよう。筆者のこの小論の表題“情報検索の問題点とその対策”の中の自然語から今まで述べてきたことが推察できるか，あるいは索引できるか考えていただくとよい。

ＫＷＩＣ索引は作成が容易でほかの索引形態のものより速時性がある点を考えると，綿密な調査用というよりは，一定順に配列された索引語のある速報と考えるほうが妥当であろう。

VI. 索引法の機能的差異の原因

次に，自然語索引と，統制されているディスクリプタ索引と，体系的分類による索引との間の機能的差異の原因について定性的な解釈を加え，今後の索引形式について考えてみたい。



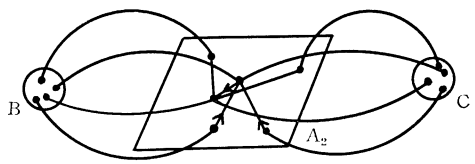
第３図　自然語による索引

第３図の A_1 は人間の全概念を表わすフィールドであって，その中に全自然語が一定順序に配列されており，Ｂは索引をつけようとしている記事が包括している概念とする。自然語による索引とは，Ｂの記事の中のファセットである自然語のうち記事の概念を表わすのに有効と思われるのを索引語として選び，A_1 の全自然語のフィールド中にある同じ語に結びついている記事を捜すことにほかならない。

記事Ｂの含む概念と検索したい概念Ｃとが同一であれば，両者は A_1 の自然語を通じて完全に一致するはずであるが，*Uniterm Index* の検討でわかるようにまず完全には一致しない。それはどうしてであろうか。

Ｂの中からの自然語の選び方と，Ｃの中からの自然語の選び方とに一貫性がない，一貫性をつけられないからである。Ｂにある多数のファセットである自然語の中から少数の索引語を選び出し，同じ概念であるＣからまた少数の自然語を選び出す。語数が少ないほど両者の自然語のくいちがいの可能性が大きい。その偏差を少なくするためにはＢにつける索引語を多くするとよいが，そうすると今度はノイズを多くする結果となる。さらにそれのみか同一の概念が，いくつか違った索引語または索引語の組合わせ方で表わされるという結果になる。

組合わせ方に起因する過誤の例としては“歴史”と“教育法”から“教育法の歴史”と“歴史の教育法”と“歴史と教育法”などの組合わせがあってどれとも決定できないことを前に述べた。これを解決するには単語相互間の関係を明らかにする，言いかえると接続詞的機能を各単語に付与しておけばよい。

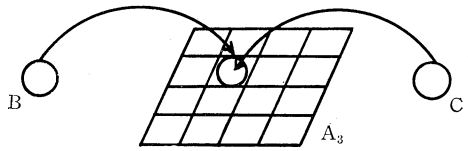


第４図　role と link による索引

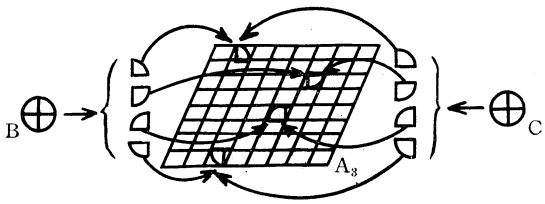
第３図のようにフィールド A_1 で自然語を点としてとらえるのでなく，第４図のように方向性をもった点（ベクトル）としてとらえ，あるいはまた点相互間の結合状態まで示すとよい。Role indicator はこのベクトルに，link は結線に相当するものであり，Ranganathan の colon classification といわれているものも機能的にみればこれに類するものである。

もっとも，機械語なり記号は自然語とは違うが，自然

語と常にある一定の対応が成立つときは，概念作用的にはその間に差がないと考えてよい。



第 5 図　体系的分類による索引 (1)



第 6 図　体系的分類による索引 (2)

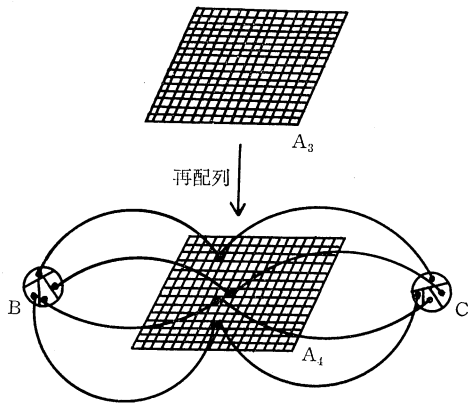
次に，体系的分類の模型を第5図と第6図に示す。人間の有する全概念を表わすフィールド A_3 (第3図の A_1，第4図の A_2 に一致する) を概念の上位·下位関係にしたがっていくつかの小フィールドに分けておく。ついで分類しようとしている記事Bが包括している概念に対応する概念をフィールド A_3 の中の小フィールドに見つけだしてそれに結びつけておく。

このとき A_3 の分け方があらければBの概念を小フィールドに結びつけることがやさしく，また過誤も少ない。A_3 を分けるあらさ，すなわちこまかさは記事の総数によって大体きまる。一つの小フィールドに数百もの記事が結びつくようではあらすぎるし，数個以下というのではこまかすぎるであろう。

全概念を適度に分けた小フィールドの概念の大きさが記事Bそのままを表わす概念の大きさにくらべて小さすぎるときは，第6図のようにBの概念をいくつかに細分してその小概念を A_3 の小フィールドに結びつけることになる。

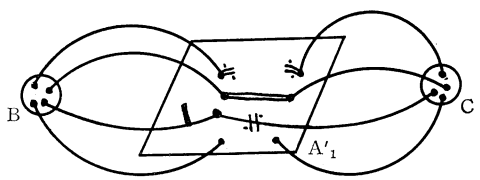
この操作は第3図のに似ているが，第3図の自然語の場合は，記事Bの多数あるファセット中から若干を抽出するゆえ全数抽出しないかぎり洩れがある。ところが第6図の体系的分類のときはBの概念をいくつかに分割してその小概念を小フィールドに結びつける建前であるから原則としては洩れがない。

APSE のように体系的分類法から出発した用語集 (thesaurus) によるときは，索引語の抽出と検索の方法は第3図とまったく同様であるが，全概念フィールドが A_1 のような無制限自然語フィールドでなく，第7図のように，あらかじめ体系的分類フィールド A_3 を通じて規制したディスクリプタを一定順序に配列したフィールド A_4 である点が違っている。



第 7 図　体系的分類を基礎にしたディスクリプタによる索引

しかし用語集による方法がすべて第7図の $A_3 \rightarrow A_4$ の形式かというと必ずしもそうでなく，同義語・類語を統一した程度の用語集だと概念の集合関係についての規制がないゆえ，むしろ第3図の A_1 が使われていて，そのフィールド内のところどころで同義語·類語を等号 (=) などでつないで統合あるいは参照 (see also) がしてある程度だといってよい (第8図)。



第 8 図　同意語・類語の統合をした用語集による索引

したがって一概に thesaurus と言い，descriptor と言っても機能に大きな差がある。

VII. 記事機械検索のための索引法

以上のように考えてくると A_3 型の概念フィールドを使う方法が良いことは明らかであるが，体系的分類法を使うためには分類体系を記憶していなければならない点が分類操作上の難点であり，A_1 型の自然語は原文中から任意抽出していくように簡単にはいかない。

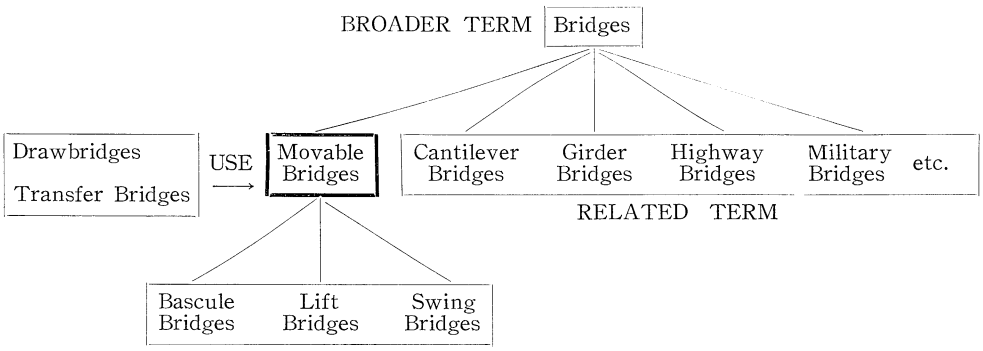
しかし第1図でわかるように，ドキュメンテーションの全行程中で分類の行程を簡易化することは翻訳の機械化とともに重要なことであり，事実，蓄積・検索行程の機械化が採用されていくにつれ，分類の簡易化の方向として A_1 型の自然語フィールド，A_1' 型の類語統合自然語フィールド，A_2 型のベクトルフィールド，あるいは，これらの機械語・記号を使う方法がとられてきている。

ところが既述のように自然語索引は体系的分類法および *APSE* のごとくよく規制されたディスクリプタ索引にくらべて信頼度が非常に低い。

以上のことを考慮にいれると，今後の記事機械検索のためには体系的分類から出発したディスクリプタ，あるいは概念の集合の面から規制を加えたディスクリプタによるのが有望であるが，これらのディスクリプタは一朝一夕に決定できない点，注意すべきである。

この点から紹介しておく必要があるのは米国の Engineering Joint Council の *Thesaurus of Engineering Terms* である。これは体系的分類と関連させたディスクリプタをアルファベット順に配列したものであるが，一つの体系が概念フィールド全体にわたるようなものではなく，きわめて多数のごく限られた小分野ごとの体系的分類体系で全フィールドをおおうようにしたものである。

"Movable Bridges" を例にとれば



のような小さい概念ピラミッドを各ディスクリプタごとに作ってあって，用語集には全ディスクリプタをアルファベット順に配列してある（第9図）。

索引をつけるとき，あるいは検索するときは思い当た

```
MOUNTAINS
  RT  FORMATIONS
      GEOMORPHOLOGY
      TOPOGRAPHY
      VOLCANISM

MOUNTING
  RT  BRACKETS
      HANGERS
      JOINING
      SUSPENDING (HANGING)

MOVABLE BRIDGES
  UF  DRAWBRIDGES
      TRANSFER BRIDGES
  NT  BASCULE BRIDGES
      LIFT BRIDGES
      SWING BRIDGES
  BT  BRIDGES (STRUCTURES)
  RT  CANTILEVER BRIDGES
      GIRDER BRIDGES
      HIGHWAY BRIDGES
      MILITARY BRIDGES
      PORTABLE BRIDGES
      RAILROAD BRIDGES
      SKEW BRIDGES
      TRUSS BRIDGES

MOVABLE DAMS
  BT  DAMS
      HYDRAULIC STRUCTURES &
  RT  CHECK STRUCTURES

MOVEMENT
  USE MOTION

MOVING
  USE MOTION

MOVING AVERAGE
  BT  AVERAGE
  RT  AUTOCORRELATION
      CURVE FITTING
      EXTRAPOLATION
      PERIODIC VARIATIONS
      TIME SERIES ANALYSIS
      TRENDS

MOVING PICTURES
  USE MOTION PICTURES

MOVING TARGET INDICATORS
  RT  DOPPLER TRACKING
      FIRE CONTROL
      RADAR TRACKING
```

第 9 図　Thesaurus of Engineering Terms の一部

ったディスクリプタを中心とした概念ピラミッドの中のディスクリプタのうち適当なものを選ぶことによって，簡便に一層適切なディスクリプタを，しかも洩れることが少なく使えるようになっている。

Engineering Joint Council の下部団体の一つである Engineering Index Inc. では *Engineering Index* という抄録雑誌を編集しながら，それの索引に必要なディスクリプタを前記の概念小ピラミッドの考え方で選択し，その Thesaurus (非公開)を毎週改訂して最終的なものにする努力を数年間続けている。

Engineering Joint Council では下部団体のこのような資料をもとにして，さらに広範囲の分野の用語集の制定を目ざしており，1964 年版に続いて 1966 年には改訂版を出すことにしている。

機械検索を成功させ，ひいては技術情報のドキュメンテーションを効果的にやるためにはこのような地道な努力が必要であって，電子計算機を買えばあとは他所で作った用語集を借用してうまくやっていける状態に現状は達していないのである。

(富士写真フイルムＫＫ足柄研究所)