

検索式のインデックスの数と検索される課題数との相関
—研究課題検索システム (REGISTER) における実験—

Correlation between the Number of Codes Used in Searching
Formulas and the Number of Current Research Projects Retrieved

吉村典夫, 五十嵐 瞳
Tuneo Yosimura and Hitomi Igarasi

Résumé

REGISTER is a computer-based system for retrieval of research projects currently conducted in Japan covering all fields of science and technology. has been developed since 1970, and started its actual operation in 1972. Over and beyond the operation of the system, the Division of System Development, Japan Science Foundation, performs various investigations needed for improving the system and developing a new system called RECRAS aimed at retrieving the same kind of information on current research projects in the fields of agriculture, forestry and fishery in Japan.

The main subject of the present paper is one of experimental studies executed by the said Division concerning the quantitative relation between the number of codes in searching formulas and the number of research projects retrieved. For understanding the system, REGISTER, however, the writers describe its outline in the first part of the paper, and then, deal with the experimental study.

The data base of REGISTER is formed by assigning to research project classification codes according to a decimal classification scheme of four digits, CST, using a multi-aspect post-coordinating method. The retrieval is conducted by using and-logical expressions consisting of these classification codes as searching formulas.

Concerning the quantitative relation between the number of codes in searching formulas and the number of research projects retrieved by matching to the searching formulas, the writers assume that "the logarithm of the number of research projects retrieved becomes a linear function of the number of codes used in searching formulas, and the regression coefficient is negative," or our hypothesis is formularized as follows:

$$\log Y = b_0 + b_1 X \quad b_1 < 0$$

吉村典夫：日本科学技術振興財団システム開発室長

Tuneo Yosimura, Director, Division of System Development, Japan Science Foundation.

五十嵐瞳：日本科学技術振興財団システム開発室員

Hitomi Igarasi, Division of System Development, Japan Science Foundation.

where, Y is the number of research projects retrieved, X the number of codes in searching formulas, b_0 a constant term, and b_1 a regression coefficient.

The retrieval experiment was performed, using the data file of 8,515 research projects, having more than four codes selected from the data-base of REGISTER. The searching formulas consisting of four codes of four digits in and-logic relations, of three codes of the same, of two, and of one, formed one set, and a total of 188 sets of these formulas were used for the experiment. Also a total of 133 sets of searching formulas having more generic three-digit codes were formed for the experiment.

The result of the retrieval experiment which supports the hypothesis was obtained. The data obtained from the experiment are approximately agreed with the assumptions which the writers expected at the time of making the hypothesis. The correlation between the number of codes used in searching formulas and the number of research projects retrieved was very high and its coefficient (r) was between -0.857 and -0.902 .

- | | |
|--|---|
| <p>I. まえがき</p> <p>II. REGISTER の概要</p> <p>A. 目的と機能</p> <p>B. 概要と特性</p> <p>C. 原情報</p> <p>D. インデクシングおよび科学技術分類表</p> <p>E. ファイル</p> <p>F. 検索とサービス</p> <p>III. 検索式構成と検索される課題数間の 定量的 仮説</p> <p>A. 理論的モデル</p> <p>B. 現実の場合</p> | <p>C. 仮説の内容</p> <p>IV. 実験</p> <p>A. 実験方法</p> <p>B. 実験結果</p> <p>V. 考察</p> <p>A. 仮説の検証</p> <p>B. 実用ファイルと実験用ファイルにおける 回帰直線の関係</p> <p>C. 回帰直線の傾斜</p> <p>D. 実用ファイル4けた検索のサンプリング 実験</p> <p>VI. あとがき</p> |
|--|---|

I. まえがき

現在、日本で進行中の科学技術に関する研究課題を検索するためのシステムには、科学技術研究情報検索システム (Retrieval System for General Information of Scientific and Technological Research, 以下 REGISTER という)、農林試験研究課題検索システム (Retrieval System for Current Research of Agricultural Sciences, 以下 RECRAS という)、および JICST クリアリングサービスがある。^{1, 2)}

REGISTER³⁾ は、日本科学技術振興財団が開発し運営しているシステムであって、科学技術全分野を対象とし、現在実用に供しつつ、システム改善が行なわれている。

RECRAS⁴⁾ は、農林省が開発しつつあるもので、農林水産技術分野を対象とし、現在データ蓄積とシステム改善が行なわれている。

JICST のクリアリングサービス⁵⁾ は、現在、ダイレクトリーを出版する形で行なわれている。

筆者らは、1970年(筆者のひとり五十嵐は1971年)以来 REGISTER の開発に従事し、1972年度からは、その運営および RECRAS の開発実務を担当している。

REGISTER と RECRAS は、互いに類似の思想によったものであり、システム全体としては、マン・マシン・コオペレーションシステムである。インデクシングおよび検索には、多観点の事後組み合わせ (Post-coordination) 方式で、十進数形式の分類表 (REGISTER では

科学技術分類表—Classification of Science and Technology, 以下, CST という—, RECRAS では農業技術分類表—Classification of Agricultural Science and Technology, 以下 CAST という)を用いている。

インデクシングに当っては, ひとつの研究課題に対して, 大部分の場合, 複数個の分類項目(現実には, CST または CAST の分類項目に対応する分類標数)が指定される。

検索は, 第1次に主として分類標数による検索を機械で行なって課題を出力し, その結果にもとづいて, 人手による第2次の検索が行なわれる。

システム運用の効率上, 機械検索で出力される課題数は, ひきつづいて行なわれる人手による検索に必要なかつ最小限であることが望まれる。そのためには, 検索式の構成と, 検索された課題数との間の定量的関係を知ることが有用である。

REGISTER および RECRAS における検索は, 一般的にいえば, つぎのようになる。すなわち, ある定まったインデクスの集合のなかから複数個のインデクスを任意に選んで, 検索対象に付与してファイルを作り, このファイルから, 複数個のインデクスを構成要素とする検索式で検索を行なう。

検索式が, 複数個のインデクスを含み, かつインデクスが互いに論理積の関係にある場合は, 式に含まれるインデクスの個数が増加すれば, 検索式に該当する検索対象の個数は減少することが当然予想される。しかし, これがどのような定量的関係を持つかについての, 一般化された実証的研究報告については, 筆者らは寡聞にして, 日本国内では, その例を知らない。外国については, 少なくとも米国国立医学図書館(NLM)の医学文献情報検索システム(MEDLARS)においては, このような研究報告はないとのことである。⁵⁾

本報告は, 上記の定量的関係について, 理論的に仮説を立て, REGISTER を使用した実験によって検証を行なったものである。

筆者らは, REGISTER および RECRAS において, システム運用上発生する技術的問題を解明するための研究を行なっていて, 本報告はそのうち今日までに得られた結果の一部である。

II. REGISTER の概要

A. 目的と機能

REGISTER の開発を促したのは, 財団で運営してい

る科学技術館における業務上の必要性であった。

科学技術館は, 1964年に開設された日本最大の理工学博物館であって, 展示を主な活動としている。博物館活動を進めるためには, 常に, 科学技術に関する最新の情報を収集する必要がある。情報収集の対象としては, 図書, 雑誌, カタログ等の文書資料だけでは不十分であり, より新しい情報をつかむための適切な方法が要請されていた。

また, 科学技術館に対しては, 多くの一般市民および科学技術者から, 科学技術上の質問が寄せられる。しかし, それらの質問に対し, 同館のみの力で満足な回答をすることは甚だ困難である。この場合, どこに聞けばよいかを答えるよう努力してきた。そこで, どんな研究がどこで行なわれているかについての情報が必要となった。

科学技術に関する研究機関では, 年報等を発行しているところが多く, それらによりそこで“なに”を研究しているかを知ることができる。また, 各機関の資料を集めて一覽とした資料も, 不完全ながら存在するので, 特定の研究機関で行なっている研究課題を知ることが容易である。

しかし, これらの資料からは, “どこ”で“なに”をに関する情報が得られるのであって, “なに”は“どこ”に関する情報は得られない。REGISTER の開発は, 科学技術全般にわたって[“なに”は“どこ”で研究しているか]に答えるサービスシステムとして進められたのである。

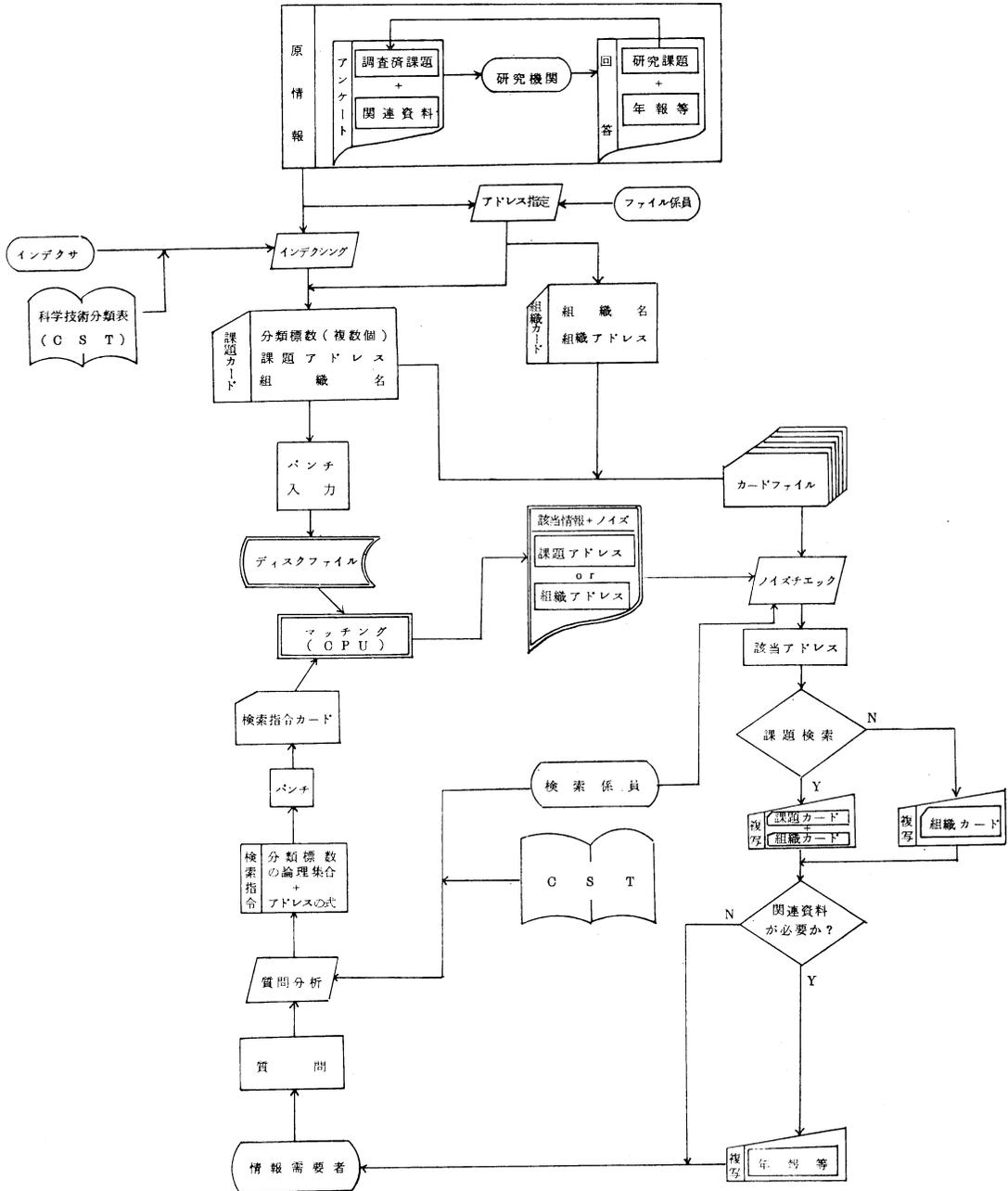
現在, REGISTER によって次のようなサービスが可能である。

1. 特定の内容をもつ, 現在進行中の研究課題名を知らせる。
2. 特定の問題について, 現在研究を実施中の組織の名称その他を知らせる。
3. 特定の問題について研究を実行する能力をもつ(と考えられる)組織名を知らせる。
4. 上記の情報に関連する諸資料(の複写物)を提供する。

B. 概要と特性

REGISTER の概念図は図 II-1 に示す通りである。この図でも明らかなように, REGISTER は, 研究課題(またはそれを実施している機関部名)を検索するシステムであり, 一種の fact retrieval である。原情報の内容(研究課題名等)は直接コンピュータにはファイル

検索式のインデックスの数と検索される課題数との相関



図II-1. REGISTER 概念図

されておらず、カードファイルになっている。コンピュータによる出力は課題および組織アドレスであって、最終的なノイズチェックは検索係員が直接課題名を参照して行なう。

REGISTER の特性としては次の3点をあげることができる。

1. 原情報は、直接研究機関へアンケートをだすことによって、hot なものを集める。

2. 情報の分析・検索には、多観点事後組み合わせ方式の十進数分類——科学技術分類表(CST)——を用いる。
3. 情報の蓄積・第一次検索にはコンピュータを用いるが、システム全体としては、マン・マシン・コオペレーション方式とする。

C. 原情報

情報源である国内の主な科学技術研究機関に対して、毎年1回、直接アンケート調査を行ない、そこで実施中の研究課題、組織、研究者数等を把握する。その際、要覧や研究年報など、研究情報に関する資料も収集する。

調査にあたっては、白紙の調査用紙を送付して記入を求めることを避け、可能な限り、調査先で行なっている研究課題をあらかじめ調査し、確認と訂正を求めるようにする。前年も調査を行なった調査先には、前年度に得た回答を調査用紙に記入して送付し、加除訂正を求める。

この方法は調査先機関の手間を減らすこととなり、アンケートの回収率向上のうえで大きな効果が認められる。なお、昭和48年度においては研究機関の約60%からは、年報等の関連資料が送られてきた。

昭和48年度における原情報の数は、表Ⅱ-1に示す通りである。

表Ⅱ-1. 原情報数(昭和48年度)

| 研究機関種別 | 機関数 | 組織数 | 課題数 | アンケート未回収機関数 |
|--------------|-----|-------|--------|-------------|
| 国立研究所 | 92 | 580 | 5,014 | 0 |
| 公社・公団等の研究所 | 36 | 193 | 1,030 | 1 |
| 国立大学付置研究所 | 61 | 726 | 3,683 | 2 |
| 国立大学学部付置研究施設 | 121 | 204 | 1,005 | 2 |
| 都道府県立研究所* | 164 | 274 | 2,039 | 0 |
| 合計 | 474 | 1,977 | 12,771 | 5 |

* 東京都を除く各道府県の研究所については、昭和47年度の研究課題を、昭和47年度公設試験研究機関現況(中小企業庁、工業技術院)の記載にもとづいて原情報とした。なお、同書所載の研究所は、大部分工業技術関係のものである。

D. インデクシングおよび科学技術分類表

1. インデクシング

a. インデクサおよび作業方法

インデクサは、研究課題をCSTを用いて内容分析を行なうのであるから、その作業の適否はREGISTERの性能と最も深いかかわりをもつといえる。そこで、インデクシングは、大学卒業後10年以上の研究実務歴をそなえた科学技術専門家、または科学技術情報処理の経験者に依頼している。またREGISTERは科学技術全分野の課題を対象としているためインデクサの専門も、物理、化学、機械、電気、生物学、農学、医学と各分野にわたっている。

インデクシングにあたってはこれらの人々が、CSTの項目のうち自己の専門分野に該当する部分(大項目)(たとえば、医学の人は「7 医薬の技術」)を担当する。

そして課題のすべてを必ず上記の各専門のインデクサ(6名)が目を通すように編成する。

この方法により、科学技術各分野の研究を、科学技術全分野の観点からインデクシングすることが可能となり、REGISTERの特性である、多観点事後組み合わせ方式の長所が充分発揮されることが期待される。

また、インデクシングの効率をあげるため、インデクサにはインデクシングの方法を具体的に解説した「REGISTER インデクサの手引き」、およびCSTの各項目に、その項目の示す概念の下位概念にあたる項目およびその例を付した「CST 作業版」を交付する。

2. 科学技術分類表(CST)

CSTは、REGISTERの原情報を整理し、検索する目的で作成された。そのため、主として、現在実施されている科学技術の研究に関する情報を対象として考えら

検索式のインデックスの数と検索される課題数との相関

れている。整理および検索にあたっては、コンピュータを用いて、比較的簡易なシステムで行なうことを想定した。

情報の整理および検索のためにその内容を表示する方法としては、キーワード方式と分類方式の2つが通常考えられる。REGISTER は、後者の方式が採られている。

キーワード方式によらず分類方式を採用した理由は、つぎのとおりである。

キーワード方式を採用して、自然語に何等の統制を加えずに使用するとき、同語異義がしばしばあって混乱をきたすおそれがあり、しかも、コンピュータに多大の負担をかけ、漢字が使用できないときは、機械の出力を読むのにいちじるしく不便である。

コトバに統制を加えるときは、シソーラスを使用するわけであるが、現在、既成のシソーラスで実用に耐えるものが少なくとも日本語のものでは存在せず、新たに作することは多大な費用と時間を要する。

また、シソーラスにとりあげられているコトバは、既に報文等に実用されているコトバである。しかし、REGISTER は、現在進行中の研究課題を取り扱う関係上、まだ、報文等に表われてないコトバ、すなわち、シソーラスに加わっていないコトバの意味が、しばしば発生することが予想される。したがって、シソーラスのコトバを用いて、現在進行中の研究課題の内容を表示することは不適當である。概念分類表を用いればこれらの問題は、ほとんど起こらない。

以上の点を考慮した上で、REGISTER のための分類表として CST が作成された。

CST は、全科学技術分野および総合技術、関連分野を対象として、概念分類の立場に立った 10 進数 4 けたの分類表である。

特徴は 2 つあるが、そのひとつは「多観点事後組み合わせ方式」をとっている点である。

従来の分類は、生物分類や図書分類のように「事前組み合わせ方式」をとっているのが殆んどである。この方式では、分類の対象として予想されるものの位置を、あらかじめ表の上に定めておかねばならない。したがって、詳しく分類するには階層を多くしなければならず、十進数の場合、標数のけた数を増やす結果となる。

その点 CST は「事後組み合わせ方式」であるから、科学技術および関連領域における技術等の諸問題をあらわすために、多くの観点で分類を行ない、分類標数を決

めておき、その標数の組み合わせを利用して研究の内容を表示できる。つまり、詳しく分類するには、組み合わせる標数（観点）の数を増やせばよい。

CST のもうひとつの特徴は、「間接定義方式」の採用にある。

分類表の各項目の表わす意味があいまいであっては、分類表を使用できない。この意味の指定をする方法には、各項目の定義づけを直接行なう形式と、間接に行なう形式とが考えられる。

CST のとっている「間接定義方式」は、分類項目に該当する具体的な用語や下位概念にあたる語をあげ、分類項目の表わす意味を間接に定義する方法である。この方法は、理論的な厳密さには欠けるが、必要に応じて逐次用語を補なっていけるので、実際に利用するには便利な場合が多い。現在 CST は、本表の記述は、「直接定義方式」をとっている。しかし 50 音順索引には間接定義をするためのコトバが項目全体の約半分についてあげられており、今後改善を加えて、「間接定義方式」の分類表として完全なものにする予定である。CST は、最初作られて以来、REGISTER において使用された結果にもとづいて改訂が続けられており、現在のファイルでは改訂 5 版が使用され、目下改訂 6 版が作られつつある。

表 II-2 に CST 改訂 5 版の大項目名および項目数を示し、図 II-2 には、CST の一部を示す。

E. ファイル

1. カード・ファイル

REGISTER の原情報である研究課題名等は、組織カードおよび課題カードの 2 種類のカードに記入され、アドレス番号順にファイルされている。

課題カードは、インデクシングカードおよびデータシートを兼ねている。

2. コンピュータ・ファイル

a. ファイルの構成

ファイルは、マスター・ファイルおよびインバーテッド・ファイルの 2 種類である。

マスター・ファイルは、入力データそのままの形で構成され、課題コードを検索のキーとして、各課題につけられた標数を引きだすものである。

インバーテッド・ファイルは、標数順にその標数が与えられた課題コードが配列され、標数からその標数を持つ課題コードを引きだすものである。

マスター・ファイルおよびインバーテッド・ファイルの形式を、図 II-3 および図 II-4 に示す。

表 II-2. 科学技術分類表 (CST)

| 大項目名 | 中 項 目 名 | | | | | | | | | 中項 目数 | 小項 目数 | 細項 目数 | |
|----------------------|---------|------------|-----------------|---------|---------------|--------------------|------------|--------|----------------|----------|----------|----------|-------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | | | 9 |
| 場所, 時, 量, 度量衡, 規格 | 一般 | 気候帯, 陸地 | 海, 洋, 淡水域, 咸水水域 | 宇宙空間 | 海域 | 世界の各地 | 日本の各地 | 時の観点 | 量, 標準, 度量衡, 規格 | | 9 | 46 | 138 |
| 自然科学 (生物科学を除く) | 1 | 一般 | 数学 | 物理学 | 化学 | 結晶, 物品 | 地学 | 氣象, 氣候 | | | 8 | 57 | 354 |
| 物質の技術 | 2 | 一般 | 資源の探索 | 使用原料 | 物質の製造 | 物質製品 | 不潔物・有害物の技術 | | | | 7 | 22 | 99 |
| エネルギーの技術 | 3 | 一般 | エネルギー源 | エネルギー処理 | エネルギー利用の最終形態 | エネルギー輸送 | エネルギー貯蔵 | | | | 7 | 36 | 107 |
| 情報の技術 | 4 | 一般 | 情報形態 | 情報媒体 | 情報処理 | | | その他 | | | 5 | 35 | 100 |
| 加工・運搬・貯蔵の技術 | 5 | 一般 | 加工の総合的問題 | 被加工物 | 総合完成 | 加工の結果 | 運搬の技術 | 貯蔵の技術 | | | 8 | 42 | 209 |
| 生物科学 (農林水産技術含む) | 6 | 一般 | 生物学 | 動物 | 農林水産加工産物 | 農林業における技術 | 畜産業における技術 | | | | 7 | 53 | 366 |
| 生物科学 (医薬の技術を含む) | 7 | 一般 | 組織, 器官, 系統 | 診断, 治療 | 衛生学, 公衆衛生, 疫学 | 薬学, 薬理学, 薬物療法, 毒物学 | | | | | 6 | 45 | 179 |
| 総合問題, 関連分野 | 8 | 総合問題, 関連分野 | 開発 | 安全, 福祉 | 警察, 消防, 防衛 | 経済, 営 | 法律, 制度, 行政 | 学術, 教育 | その他 | | 9 | 32 | 115 |
| | | | | | | | | | | (合計) | 67 | 369 | 1,668 |

検索式のインデクスの数と検索される課題数との相関

| | |
|--|------------------------|
| 32 エネルギー変換 (理論的, 基礎的なものを扱う。 実用のデバイス→5533 又は 583) | 325 電力から電磁エネルギーへの技術 |
| 320 エネルギー変換一般 | 3250 電力から電磁エネルギーへの技術一般 |
| 3200 エネルギー変換一般 | 3251 変圧 |
| 321 熱発生・熱交換の技術 | 3252 周波数変換 |
| 3210 熱発生・熱交換の技術一般 | 3253 電磁石 |
| 3211 化学反応 | 3254 交直変換 |
| 3212 原子力 | 3259 その他 |
| 3213 低温発生 | 326 熱から電力への技術 |
| 3214 電熱 | 3260 熱から電力への技術一般 |
| 3215 電波, 電気振動 | 3261 熱電子発電 |
| 3216 摩擦 | 3262 熱電発電 |
| 3217 熱交換 | 3263 光電発電 |
| 3218 熱ポンプ, 冷凍機 | 3264 熱誘電発電 |
| 3219 その他 | 3265 熱磁気発電 |
| 322 熱から動力への技術 | 3266 MHD発電 (MFD発電) |
| 3220 熱から動力への技術一般 | 3267 EHD発電 (EFD発電) |
| 3221 往復動蒸気機関 | 3269 その他 |
| 3222 回転蒸気機関 | 327 エネルギー源から電力への技術 |
| 3223 往復内燃機関 | 3270 エネルギー源から電力への技術一般 |
| 3224 回転内燃機関 | 3271 燃料電池 |
| 3225 ジェット機関 | 3272 原子力電池 |
| 3226 ロケット機関 | 3273 一次電池 |
| 3229 その他 | 3274 生物電池 |
| 323 動力から電力への技術 | 3275 太陽電池 |
| 3230 動力から電力への技術一般 | 3279 その他 |
| 3231 発電機 | 328 エネルギー源から動力への技術 |
| 3232 圧電変換 | 3280 エネルギー源から動力への技術一般 |
| 3239 その他 | 3281 水力利用 |
| 324 電磁力から動力への技術 | 3282 風力利用 |
| 3240 電磁力から動力への技術一般 | 3289 その他 |
| 3241 電動機 | |
| 3249 その他 | |

図 II-2. C S T 本表 (例)

| *カソコード* | ソシキ コード* | カソイ コード* | REGISTER | | | | | | | | | | 49/03/18 | PAGE 1 6 | | |
|---------|-------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|----------|----------|--|--|
| | | | ヒヨウス ウ 1. | ヒヨウス ウ 2. | ヒヨウス ウ 3. | ヒヨウス ウ 4. | ヒヨウス ウ 5. | ヒヨウス ウ 6. | ヒヨウス ウ 7. | ヒヨウス ウ 8. | ヒヨウス ウ 9. | ヒヨウス ウ 10. | | | | |
| | | | 02 | ソシキ | ノ | カソイ | スク | 9 | | | | | | | | |
| 0 2 01 | 7, | 03 | 1, | 001 | 0115 | 5581 | 5501 | | | | | | | | | |
| | | | 2, | 002 | 0115 | 5581 | 5610 | | | | | | | | | |
| | | | 3, | 003 | 0115 | 5581 | 5610 | 8311 | | | | | | | | |
| | | | 4, | 004 | 0115 | 5212 | 5581 | | | | | | | | | |
| | | | 5, | 005 | 5581 | 5451 | | | | | | | | | | |
| | | | 6, | 006 | 1618 | 6520 | 5451 | | | | | | | | | |
| | | | 7, | 007 | 5242 | 5212 | 5451 | 5456 | | | | | | | | |
| | | | 8, | 008 | 5456 | 5504 | 5572 | 8262 | | | | | | | | |
| | | | 9, | 009 | 0115 | 5212 | 5581 | | | | | | | | | |
| | | | 10, | 010 | 5581 | 5506 | 5212 | | | | | | | | | |
| | | | 11, | 011 | 5581 | 0115 | 5506 | 1767 | 5610 | 5501 | 0123 | 8311 | | | | |
| | | | | | 03 | ソシキ | ノ | カソイ | スク | 11 | | | | | | |
| 0 2 01 | 8, | 04 | 1, | 001 | 0115 | 1623 | 5451 | 8264 | 0145 | 0123 | | | | | | |
| | | | 2, | 002 | 1618 | 6520 | 5451 | 8110 | 8264 | | | | | | | |
| | | | 3, | 003 | 0120 | 1719 | 6000 | 8113 | | | | | | | | |
| | | | 4, | 004 | 5572 | 5210 | 5455 | 5456 | 8261 | 8262 | 6552 | | | | | |
| | | | 5, | 005 | 6524 | 6552 | | | | | | | | | | |
| | | | 6, | 006 | 6550 | 6523 | 6519 | | | | | | | | | |
| | | | 7, | 007 | 1650 | 5451 | 8261 | | | | | | | | | |
| | | | 8, | 008 | 1629 | 5451 | 5453 | | | | | | | | | |
| | | | 9, | 009 | 1629 | 1665 | 1623 | 5451 | 8264 | | | | | | | |
| | | | | | 04 | ソシキ | ノ | カソイ | スク | 9 | | | | | | |
| | | | | | 0201 | *カソ | ノ | カソイ | スク | 43 | | | | | | |

図 II-3. マスター・ファイルの形式

b. ファイルの内容 (インデクシングの結果)

現在コンピュータには、総数 12,768 の研究課題が格納されている。それらの課題のインデクシング結果を示すことにより、ファイルの持つ内容の特性を示すと次のようになる。

まず、表 II-3 に、1 課題に与えられた CST 標数 (4けたの場合) の数を示す。

表 II-3 のデータで計算すると、1 課題あたり平均 4.7 個の標数が与えられていることになる。

表 II-4 に、大項目別分類標数使用結果を、表 II-5 に、分類標数度数分布表を示す。

F. 検索とサービス

1. 機械による検索

a. 検索式の構成

機械検索は A で述べたように、第 1 次的には標数の組み合わせを使用する。2 次的には、課題コードによることも考えるが、この部分は当面検索係員によることとする。

検索式は、標数と論理記号によって表示する。使用する論理記号は、* (論理積)、+ (論理和)、Y (否定——これは必ず前に * を伴う) の 3 種類とする。

論理記号の使用については次のような制限を設ける。
論理積…… 1 質問中において 9 個以内とする。

否定—— 1 質問中において 1 回までとするほか、式においては最後におく、すなわち A*B*C*D*F*G*H*Y I の形とする (ただし A, B…… I は標数を表わす)。

論理和—— 1 質問中において 1 回までとし、論理和は

検索式のインデクスの数と検索される課題数との相関

| REGISTER | ヒヨウスウ | カタガ | イコト | タイ | イテラン | ヒヨウ | 49/03/18 | PAGE1 | 1 | | |
|----------|-------|-----|-----------|----|-----------|-----|-----------|-------|-----------|----|-----------|
| NO. | ヒヨウスウ | | | | | | | | | | |
| 1 | 0100 | 1 | 060105013 | 2 | 081406004 | 3 | 082008001 | 4 | 100503006 | | |
| 2 | 0110 | 1 | 082008002 | | | | | | | | |
| 3 | 0111 | 1 | 060105001 | 2 | 060105005 | 3 | 060105006 | 4 | 060105007 | 5 | 060105008 |
| | | 6 | 060105009 | 7 | 060105010 | 8 | 060105011 | 9 | 060105013 | 10 | 060105014 |
| | | 11 | 060105015 | 12 | 060105016 | 13 | 060105017 | 14 | 060105018 | 15 | 060105021 |
| | | 16 | 060105022 | 17 | 060105023 | 18 | 060105024 | 19 | 060105025 | 20 | 060105026 |
| | | 21 | 060105027 | 22 | 060105028 | 23 | 060105029 | 24 | 060105030 | 25 | 060105031 |
| | | 26 | 060105032 | 27 | 060105033 | 28 | 060105034 | 29 | 110403008 | 30 | 201701002 |
| | | 31 | 346499001 | 32 | 410101006 | 33 | 410101010 | 34 | 533799002 | 35 | 533799003 |
| | | 36 | 547003002 | | | | | | | | |
| 4 | 0112 | 1 | 083102001 | 2 | 083102001 | | | | | | |
| 5 | 0113 | 1 | 040603004 | 2 | 040603005 | 3 | 081302008 | 4 | 082306008 | 5 | 083102001 |
| | | 6 | 083102001 | 7 | 083102002 | 8 | 083102003 | 9 | 100402001 | 10 | 100402002 |
| | | 11 | 110416005 | 12 | 110416004 | 13 | 410108006 | 14 | 453303002 | 15 | 483001005 |
| | | 16 | 483004006 | 17 | 483007002 | 18 | 533799004 | 19 | 547901003 | 20 | 588001001 |
| | | 21 | 588001002 | 22 | 588202001 | 23 | 588202002 | | | | |
| 6 | 0114 | 1 | 080605004 | 2 | 080605005 | 3 | 080605006 | 4 | 081201004 | 5 | 081201004 |
| | | 6 | 081202001 | 7 | 081202002 | 8 | 081299001 | 9 | 081301001 | 10 | 081301008 |
| | | 11 | 081301009 | 12 | 081301011 | 13 | 081301012 | 14 | 081301013 | 15 | 081301016 |
| | | 16 | 081303007 | 17 | 081304007 | 18 | 081304010 | 19 | 081305007 | 20 | 081307001 |
| | | 21 | 081307002 | 22 | 081307005 | 23 | 081399005 | 24 | 083003018 | 25 | 083003022 |
| | | 26 | 372099007 | 27 | 533799004 | | | | | | |
| 7 | 0115 | 1 | 020102002 | 2 | 020102003 | 3 | 020103001 | 4 | 020103002 | 5 | 020103003 |
| | | 6 | 020103004 | 7 | 020103009 | 8 | 020103011 | 9 | 020104001 | 10 | 080701002 |
| | | 11 | 080701005 | 12 | 080701008 | 13 | 080702007 | 14 | 080704003 | 15 | 080704004 |
| | | 16 | 080705006 | 17 | 080709007 | 18 | 080710001 | 19 | 080710002 | 20 | 080802010 |
| | | 21 | 083003018 | 22 | 091201002 | 23 | 200707001 | 24 | 200707003 | 25 | 200707004 |
| | | 26 | 200707005 | 27 | 310499002 | 28 | 310499005 | 29 | 310499006 | 30 | 310499007 |
| | | 31 | 310499009 | 32 | 324003004 | 33 | 343003002 | 34 | 350299001 | 35 | 410108003 |
| | | 36 | 410109002 | 37 | 510399006 | 38 | 510399007 | 39 | 547003001 | | |
| 8 | 0119 | 1 | 060105007 | 2 | 060105027 | | | | | | |
| 9 | 0120 | 1 | 020104003 | 2 | 070116004 | 3 | 080701013 | 4 | 083101016 | 5 | 100401005 |
| | | 6 | 100401008 | 7 | 100403003 | 8 | 100404014 | 9 | 100404015 | 10 | 410107002 |
| | | 11 | 410107004 | 12 | 422101007 | 13 | 434810004 | 14 | 453517003 | 15 | 453525001 |
| 10 | 0123 | 1 | 020103011 | 2 | 020104001 | 3 | 040407004 | 4 | 080901009 | 5 | 080901010 |
| | | 6 | 081301012 | 7 | 100405011 | 8 | 130209004 | 9 | 200707001 | 10 | 200707003 |
| | | 11 | 200707005 | 12 | 410103003 | 13 | 410107001 | 14 | 410107005 | 15 | 410107006 |
| | | 16 | 410107007 | 17 | 453529001 | 18 | 560103001 | | | | |
| 11 | 0124 | 1 | 560103001 | | | | | | | | |
| 12 | 0125 | 1 | 100407007 | 2 | 410103001 | | | | | | |
| 13 | 0129 | 1 | 070117011 | | | | | | | | |
| 14 | 0131 | 1 | 100301009 | 2 | 100499008 | 3 | 454008001 | 4 | 454008002 | 5 | 480302006 |

図 II-4. インバーテッド・ファイルの形式

最後におく。すなわち

A*B*C*D*E*F*G*(H+I)

また否定を用いるときは *Y のひとつ前におき、

A*B*C*D*E*F*(G+H)*Y I とする。

このほか、課題検索か組織検索かを識別する命令も検索命令に含まれている。

b. 上位標数検索

ファイルが、課題アドレスと4けたの標数の対応となっているため、検索も4けたの標数の組み合わせとするのが自然である。

しかし、「エネルギー変換のための熱発生における公害防止に関する研究」を検索する場合は、3210 から3219 までの9種と8250 から8259 までの10種の標数の組み合わせを取らなければならない(図 II-2 参照)。

表 II-3. 1 課題に与えられた CST 標数の数

| 標数の個数 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 合計 |
|-------|-------|--------|--------|--------|--------|--------|-------|-------|-------|-------|--------|
| 課題数 | 469 | 1474 | 2,310 | 2,423 | 2,088 | 1,507 | 1,053 | 644 | 416 | 384 | 12,768 |
| (%) | (3.7) | (11.5) | (18.1) | (19.0) | (16.4) | (11.8) | (8.3) | (5.0) | (3.3) | (3.0) | (100) |

表 II-4. 大項目別分類標数使用結果

| CST の 項 目 | 分類表にある 項 目 (a) | 使用された 項 目 (b) | 項目のべ使用 回 数 (c) | 項目使用回数 比率% (100 ci/m) | 項目使用率 % (100 bi/ai) |
|--------------|-----------------------|-----------------------|---------------------------|--------------------------|-----------------------------|
| 0 | 138 (46) | 129 (44) | 2,522 (2,520) | 4.2 (4.2) | 93.4 (95.7) |
| 1 | 354 (57) | 335 (56) | 14,586 (14,484) | 24.5 (24.4) | 94.6 (98.3) |
| 2 | 99 (22) | 96 (21) | 4,363 (4,347) | 7.3 (7.3) | 96.8 (95.5) |
| 3 | 107 (36) | 78 (32) | 612 (603) | 1.0 (1.0) | 72.9 (88.9) |
| 4 | 100 (35) | 85 (35) | 2904 (2,900) | 4.9 (4.9) | 85.0 (100) |
| 5 | 209 (42) | 204 (39) | 14,881 (14,846) | 24.9 (25.0) | 97.6 (92.9) |
| 6 | 366 (53) | 336 (52) | 11,439 (11,412) | 19.2 (19.2) | 91.8 (98.1) |
| 7 | 179 (45) | 156 (44) | 4,521 (4,501) | 7.6 (7.6) | 87.2 (97.8) |
| 8 | 115 (32) | 103 (28) | 3,808 (3,761) | 6.4 (6.3) | 89.7 (87.5) |
| 9 | 1 (1) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| 合 計 | (k) 1,668 (369) | (l) 1,522 (351) | (m) 59,636 (59,374) | 100 (100) | (100 l/k) 91.2 (95.1) |

注) 上欄 細項目 (4けたの標数)
下欄 小項目 (3けたの標数)

さらに、「エネルギー変換における公害防止であれば 3200 から 3289 までの 49 種と 8250 から 8259 までの 10 種の標数の組み合わせが必要となる。

これらの検索式の数、前者では 90 回、後者では 490 回におよぶ。論理和を使うこともできるが、この場合はいたずらに検索式を長くすることになり、いずれにしても非常に不便である。

このような場合に備えて、たとえば 3210 から 3219 までをまとめて検索する方式(以下、上位標数検索という)ができるようにしてある。上位標数検索は実際上の必要性から考えて、2けたまでにとどめる。

2. 検索の実例

次に REGISTER による検索の実例を示す。

「金属素材の腐食・防食についての現在実施されている研究課題にはどのようなものがあるか」という質問があ

ったとする。

これを内容分析すると、取り扱っているものは、加工の素材としての金属であるから

- 52 被加工物の中の
- 5202 金属原素材かまたは
- 5211 金属素材があてはまる。

腐食防止技術は

- 53 単位加工, 加工の手段
 - 533 表面処理, 塗装, 表面加工のところを見ると (腐食・防食 → 1493) という参照がついている。これは、腐食・防食は加工の技術であっても、
 - 149 その他の化学
 - 1493 腐食・防食の化学を用いるということを示している。
- 従って検索式は、

検索式のインデックスの数と検索される課題数との相関

表 II-5. 分類標数使用度数分布表

| 使用頻度 | 標数の数 | | | | |
|-----------|------|-----|-------------|------|-----|
| | 4けた | 3けた | | | |
| | | | 601 ~ 650 | 0 | 2 |
| | | | 651 ~ 700 | 1 | 5 |
| 1 ~ 10 | 610 | 69 | 701 ~ 750 | 0 | 2 |
| 11 ~ 20 | 242 | 30 | 751 ~ 800 | 1 | 1 |
| 21 ~ 30 | 163 | 25 | 801 ~ 850 | | 2 |
| 31 ~ 40 | 103 | 19 | 851 ~ 900 | | |
| 41 ~ 50 | 77 | 17 | 901 ~ 950 | | |
| 51 ~ 60 | 49 | 15 | 951 ~ 1000 | 1 | 2 |
| 61 ~ 70 | 49 | 8 | | | |
| 71 ~ 80 | 32 | 9 | 1001 ~ 1100 | | 2 |
| 81 ~ 90 | 20 | 11 | 1101 ~ 1200 | | 3 |
| 91 ~ 100 | 28 | 5 | 1201 ~ 1300 | | 1 |
| | | | 1301 ~ 1400 | | 1 |
| 101 ~ 150 | 64 | 33 | 1401 ~ 1500 | | |
| 151 ~ 200 | 36 | 26 | 1501 ~ 1600 | | 1 |
| 201 ~ 250 | 19 | 20 | 1601 ~ 1700 | | |
| 251 ~ 300 | 14 | 13 | 1701 ~ 1800 | | |
| 301 ~ 350 | 5 | 8 | 1801 ~ 1900 | | |
| 351 ~ 400 | 3 | 6 | 1901 ~ 2000 | | |
| 401 ~ 450 | 4 | 1 | 2001 ~ | | 2 |
| 451 ~ 500 | 0 | 3 | | | |
| 501 ~ 550 | 0 | 6 | | | |
| 551 ~ 600 | 1 | 3 | | | |
| | | | 合計 | 1522 | 351 |

1493*(5202+5211)ということになる。しかし、金属と指定されなくても金属を対象として含んでいる課題があるおそれがある。その場合、520 および 521 として上位検索をすれば、金属としての標数が与えられていなくても適合する課題は検索される。検索式に4けたが1つ含まれているから、回答として多量に打ち出されることはないと思われる。

そこでこの検索式は
1493*(520+521) とする。

図 II-5 に課題検索出力を、図 II-6 に組織検索出力を示す。(この例では質問者は、課題を求めているので、実際には組織検索はしない。)

機械出力された 27 件について、検索担当者がカードファイルによってチェックする。

ノイズチェックの結果、質問に該当する課題は、25件であり、実施している部(組織)数は 19 である。

各部毎に、組織カードと課題カードを組み合わせてコピーしたものが、最終的な回答である。

以上は、1 質問に 1 検索式の場合であるが、実際には 2, 3 の検索式を要することが多い。

なお、要望があれば、該当研究課題に関する資料(当館で所蔵しているもの)の複写を提供することも可能である。

3. REGISTER サービスの利用状況

REGISTER は、1972 年度からテストサービスを行なっているが、研究課題サービスとして 1974 年度から本格的にサービスを開始した。

利用者は主に、企業の研究管理者・情報担当者であって、営業種目拡大に伴う新技術の開発のための利用が多い。そのため、関連主題を広くカバーする検索が要求される。その点、多観点事後組み合わせ方式の分類表によって情報を整理し、検索している REGISTER は大いに利用価値があるものと考えられる。

1974 年 4 月から 7 月までの、検索依頼の総件数は約 40 件である。また、依頼先は、企業 23、大学の図書館 1、公立の研究機関 1、計 25 か所である。

検索式のインデクスの数と検索される課題数との相関

この仮定のもとで、分類表は、その分類標数の個数を n 、ひとつの研究課題に与える分類標数の個数を r としたとき、 nCr 個の Cell を用意したことになる。そして、インデクシングは、研究課題をその Cell のいずれかに配架することになる。

この場合、各 Cell にそれぞれ1個の（平均頻度が1として考える）研究課題が配架されたと仮定すれば（仮定3）、マスター・ファイルに格納できる研究課題の数は nCr 個となる。

分類表の分類標数の個数が、表 II-2 から4けたでは、1688、3けたでは369あるから、ひとつの課題に与える分類標数を4個とすれば、

$$4\text{けたの場合} \quad 1688C4 \doteq 3.2 \times 10^{11}$$

$$3\text{けたの場合} \quad 369C4 \doteq 7.6 \times 10^8$$

となる。

2. 検索のモデル

1で述べたファイルを用いて検索を行なう場合、検索式のなかの互いに論理積の関係にある分類標数の個数は、1個、2個、3個、4個の4通りであって、検索式によって検索される課題の数は、一般的には、

$$y = \frac{(n-m)!}{(n-r)! (r-m)!}$$

y : 検索される課題数

n : 分類表の分類標数の数

m : 検索式を構成する互いに論理積の関係がある分類標数の個数

r : ひとつの課題に与えられる分類標数の個数で表わされる。

いま、 r を A. 1 に従って4とし、 m を1、2、3、4として、 n の各種の値について y の値を計算すれば、表 III-1 に示すようになる。

さらに x の値は、 m を特定したとき、 n についてそれぞれの場合、3次、2次、1次の階差を示し、差分方程式で表される性質の関数となる。

したがって x の値を示す曲線（以下、効率曲線という）を作れば、曲線は n の増加に従って、対数曲線に漸近して行く。また、一定の n に対して m の増加につれて減衰する。

B. 現実の場合

以下、A で定めた仮定をはずした場合を考える。

まず、仮定2をはずした場合は、分類表において、実際に起こり得ない分類標数の組み合わせがあることになる。現実には CST は、この場合に属する。これは、A で述べた Cell のなかから、実際に起り得ない組み合わせの分を除いて考えればよい。

つぎに、仮定3をはずす。REGISTER のファイルでは、ひとつの Cell にあてはまる研究課題の数には、ばらつきが当然存在する。この場合、理論的にみちびかれ

表 III-1. 検索される課題数 (y)

| n \ m | m | | | | |
|-------|-----------------------|--------------------|--------------------|--------------------|---|
| | 0* | 1 | 2 | 3 | 4 |
| 5 | 5 | 4 | 3 | 2 | 1 |
| 6 | 15 | 10 | 6 | 3 | 1 |
| 7 | 35 | 20 | 10 | 4 | 1 |
| 8 | 70 | 35 | 15 | 5 | 1 |
| 9 | 126 | 56 | 21 | 6 | 1 |
| 10 | 210 | 84 | 28 | 7 | 1 |
| 100 | 3.92×10^6 | 1.57×10^5 | 4.75×10^3 | 9.7×10 | 1 |
| 500 | 2.57×10^9 | 2.07×10^7 | 1.24×10^5 | 4.97×10^2 | 1 |
| 1000 | 4.15×10^{10} | 1.66×10^8 | 4.98×10^5 | 9.97×10^2 | 1 |
| 1500 | 8.41×10^{11} | 7.48×10^8 | 1.12×10^6 | 1.50×10^3 | 1 |

$$y = \frac{(n-m)}{(n-r)! (r-m)!}$$

$m=0, 1, 2, 3, 4$

$r=4$

$n > r$

$r \geq m \geq 0$

* 分類表から考えられる Cell の全数

た効率曲線を中心に、利用頻度の高い Cell を対象とした検索を行なえば、効率曲線の回帰係数の絶対値が大きくなり、利用頻度の低い Cell を対象として検索すれば、回帰係数の絶対値が小さくなる。

仮定1をはずした場合は、分類標数の間に仮定2をはずした場合で述べたことと、分類標数のグループ化、すなわち、上位標数で検索を行なった場合との2つが考えられる。前者については既に述べたことと同質である。後者では CST を3けたの分類表として考えることになる。したがって、分類標数の数が減少し、その結果各

Cell に配架される研究課題の数が増加したと考えればよい。結果的には、効率曲線の m のそれぞれの価での x の価が平均的に見て大きくなり、分類標数の数が大きい場合に比べて検索される課題数が増加するから、曲線は上方へとシフトする。

C. 仮説の内容

A の理論的モデルにもとづけば、B で述べた現実の問題を考慮しても、検索式を構成する互に論理積の関係にある分類標数の個数と検索される研究課題数との間には、分類表における分類標数の数が相当大きな場合、近

表 IV-1. 1 課題に与えられた CST 標数の数 (実験ファイル)

| 標数の個数 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 合計 |
|-------|--------|--------|--------|--------|-------|-------|-------|-------|
| 課題数 | 2,423 | 2,088 | 1,507 | 1,053 | 644 | 416 | 384 | 8,515 |
| (%) | (28.4) | (24.5) | (17.7) | (12.4) | (7.6) | (4.9) | (4.5) | (100) |

表 IV-2. 大項目別分類標数使用結果 (実験ファイル)

| CST 大項目 | 分類表にある項目 (a) | 使用された項目 (b) | 項目の使用回数 (c) | 項目使用回数比率% (100 ci/m) | 項目使用率% (100 bi/ai) |
|---------|-----------------------|----------------------|---------------------------|----------------------|-----------------------------|
| 0 | 138 (46) | 126 (44) | 2,186 (2,160) | 4.4 (4.4) | 91.3 (95.7) |
| 1 | 354 (57) | 322 (55) | 11,921 (11,852) | 24.2 (24.2) | 91.0 (96.5) |
| 2 | 99 (22) | 94 (21) | 4,012 (3,993) | 8.1 (8.1) | 95.0 (95.5) |
| 3 | 107 (36) | 75 (30) | 526 (521) | 1.1 (1.1) | 70.1 (83.3) |
| 4 | 100 (35) | 81 (35) | 2,326 (2,321) | 4.7 (4.7) | 81.0 (100) |
| 5 | 209 (42) | 201 (38) | 12,882 (12,838) | 26.1 (26.2) | 66.2 (90.4) |
| 6 | 366 (53) | 332 (52) | 8,638 (8,617) | 17.5 (17.6) | 90.7 (98.1) |
| 7 | 179 (45) | 153 (42) | 3,708 (3,695) | 7.5 (7.5) | 85.5 (93.3) |
| 8 | 115 (32) | 98 (26) | 3,081 (3,069) | 6.3 (6.3) | 85.2 (81.3) |
| 9 | 1 (1) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| 合計 | (k) 1,668 (369) | (l) 1482 (343) | (m) 49,280 (49,066) | 100 (100) | (100 l/k) 88.9 (93.0) |

注) 上欄: 細項目 (4けたの標数)
下欄: 小項目 (3けたの標数)

検索式のインデクスの数と検索される課題数との相関

表 IV-3. 分類標数使用度数分布表 (実験ファイル)

| 使用頻度 | 標数の数 | | | | |
|-----------|------|-----|-------------|------|-----|
| | 4けた | 3けた | | | |
| | | | 601 ~ 650 | | 1 |
| | | | 651 ~ 700 | | |
| 1 ~ 10 | 654 | 75 | 701 ~ 750 | 1 | 4 |
| 11 ~ 20 | 249 | 33 | 751 ~ 800 | 1 | 1 |
| 21 ~ 30 | 149 | 27 | 801 ~ 850 | | 2 |
| 31 ~ 40 | 94 | 17 | 851 ~ 900 | | |
| 41 ~ 50 | 63 | 13 | 901 ~ 950 | | 2 |
| 51 ~ 60 | 46 | 11 | 951 ~ 1000 | | |
| 61 ~ 70 | 45 | 12 | | | |
| 71 ~ 80 | 18 | 10 | 1001 ~ 1100 | | 3 |
| 81 ~ 90 | 30 | 14 | 1101 ~ 1200 | | |
| 91 ~ 100 | 18 | 9 | 1201 ~ 1300 | | 1 |
| | | | 1301 ~ 1400 | | |
| 101 ~ 150 | 57 | 27 | 1401 ~ 1500 | | |
| 151 ~ 200 | 25 | 31 | 1501 ~ 1600 | | |
| 201 ~ 250 | 14 | 14 | 1601 ~ 1700 | | |
| 251 ~ 300 | 8 | 8 | 1701 ~ 1800 | | |
| 301 ~ 350 | 5 | 6 | 1801 ~ 1900 | | 1 |
| 351 ~ 400 | 3 | 2 | 1901 ~ 2000 | | |
| 401 ~ 450 | | 7 | 2001 ~ | | 1 |
| 451 ~ 500 | | 6 | | | |
| 501 ~ 550 | | 1 | | | |
| 551 ~ 600 | 2 | 4 | | | |
| | | | 合計 | 1482 | 343 |

似的に一般的関係が成立つことが予想される。そこで、つぎの仮説を立てることとする。

仮説. 検索式で検索される研究課題数の対数は、検索式を構成する互いに論理積の関係にある分類標数の個数の1次函数となり、1次の係数は負の値をとるであろう。

この仮説を回帰方程式の形であらわせば、

$$\log Y = b_0 + b_1 X$$

ここに

X: 分類標数の個数

Y: 検索される課題数

b_0 : 定数

b_1 : 回帰係数 $b_1 < 0$

となる。

IV. 実験

A. 実験方法

1. 概要

III で述べた仮説を検証するため、検索式を構成する互いに論理積の関係にある分類標数(以下、標数という)を1個から4個まで変化させ、検索実験を行なって、検索される課題数を求め、標数の個数と検索される課題数との関係をあらわす回帰方程式を計算した。(以下、本報告では、ひとつの検索を構成する標数はすべて互いに論理積の関係にあるものとする。)

検索に用いた標数は、CSTの細項目をあらわす4けたのもの、および細項目の属するすぐ上の上位概念の小項目をあらわす3けたのもの2種類である。

なお、実用の検索における上記の関係をj知るため、上記と同様の方法で実験を行なった。

2. 実験に用いたファイル

実験AおよびBでは、IIで述べたREGISTERのファイル(以下、実用ファイルという)から、4個以上の標数が与えられている8515の研究課題をすべて抽出したファイル(以下、実験ファイルという)を用いた。

実験ファイルの内容を、表IV-1、表IV-2および

IV-3 に示す。実用ファイルの内容は、表 II-3、II-4、および II-5 に示されている。

3. 検索式

a. 標数 4 個を含む検索式の選択

実際に用いた検索式は、REGISTER で使用可能なも

ののうち、式のなかに含まれる標数が互に論理積にあるものだけとした。

CST で用いられている標数の種類は、4 けたで 1.668、3 けたで 369、うち現在のファイルで実際に使用されているものは、4 けたで 1,522、3 けたで 343 ある。この

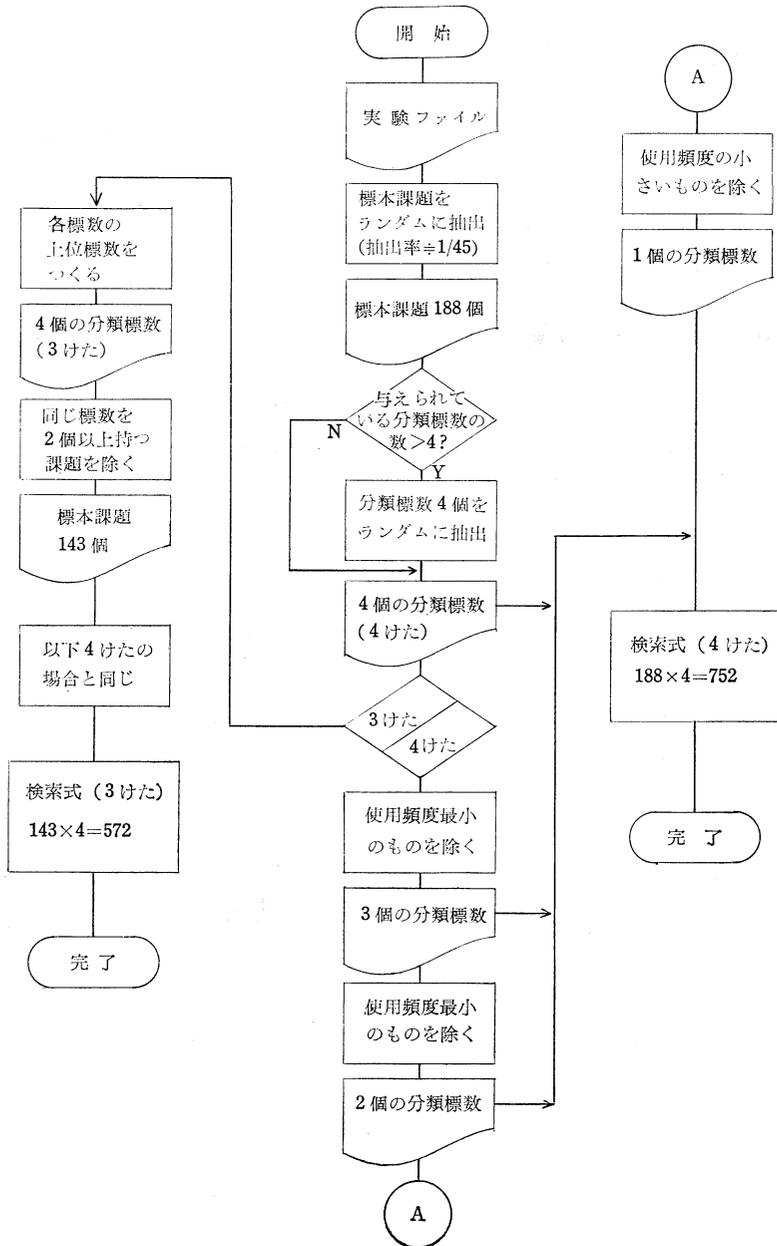


図 IV-1. 検索式作成の手順

検索式のインデクスの数と検索される課題数との相関

なかから任意に4個の標数を選んで、それらが互に論理積の関係にある検索式を作れば、その数は、4けたで ${}_{1525}C_4 \doteq 2.3 \times 10^{11}$ 、3けたで ${}_{342}C_4 \doteq 5.7 \times 10^8$ である。一方、ファイルに格納されている研究課題の総数は、実用ファイルで 1.3×10^4 個である。したがって、上記の検索式の大部分は、検索される課題がファイルのなかには実在しない。

そこで、実験に用いる検索式はすべて、それを用いて検索を行なったとき、検索される研究課題が、ファイルのなかに少なくともひとつは存在するようなものを選択した。

b. 3個以下の標数を含む検索式の選択

aで決まった標数4個を含む検索式から、順次標数を除いて、標数を3個から1個まで含む検索式を作る方法は、いろいろ考えられる。これを大別すると、

- ファイル中での使用頻度の小さいものから順次除いて行く。
- ファイル中での使用頻度の大きいものから順次除いて行く。
- ファイル中での使用頻度の無関係に、任意に除いて行く。

の3通りになる。

実際の検索では、まず標数1個をえらび、順次2個、3個とふやす考えかたで検索式を作る場合が多い。したがって、現実には、上記のa, b, cは、

- ファイル中での使用頻度の最も大きいものを選び、使用頻度の順につけ加える。
- ファイル中での使用頻度の最も小さいものを選び、使用頻度の逆順につけ加える。
- 全く任意に行なう。

となる。

いずれの方法によるかは、質問の性格が包括的であるか、一点指定的であるか等の条件によって決まる。そして、効果曲線は、それぞれ異なるものとなる。

筆者らは、まず、a)の方法による実験を行なうこととした。b), c)については、現在実験中であり、本報告ではa)の場合を取り扱う。

c. 検索式作成の手順と実例

検索式作成の手順を図IV-1に示す。

実例

標本課題 No. 2

与えられた標数：5544, 5504, 5502, 3226, 5533, 3119

ランダムに抽出された4個の標数：5502(415), 5533(311), 5544(229), 3119(19)。()内はファイル中での使用頻度。

検索式(4けた)：5502*5533*5544*3119 5502*5533*5544, 5502*5533, 5502

上位標数：550(2672), 553(2082), 554(485), 311(42) ()内はファイル中での使用頻度。

検索式(3けた)：550*553*554*311, 550*553*554, 550*553, 550

注) *は論理積をあらわす演算子

標本課題 No. 6

与えられた標数：1435, 1437, 2523, 5201

検索式(4けた)は前例と同様に作成。

上位標数：143, 143, 252, 520

この場合同じ標数143が2回出現するので、3けたの検索式作成には、この標本課題は使用しない。

この方法で、検索式を作成したため、4けたの場合と3けたの場合の実験回数が異なることになった。3けたの場合除去された標本課題と同数の課題を新に抽出して補充し、実験回数をそろえることが望ましいが、理論的考察には支障がないので、時間の関係上、補充は行なわなかった。

4. 検索実験

3の方法で作成した検索式で、実験ファイルおよび実用ファイルを用いて、検索実験を行なった。(実用ファイルについては、4けたの場合のみ)ただし、標数1個の検索は、ファイルから作成した分類標数別該当課題数表のデータを用い、機械検索は行なわなかった。

検索実験の結果の1例を、表IV-4に示す。実験の所要時間は、表IV-5のとおりである。

5. 回帰方程式の計算

4の検索実験の結果得られた検索される課題数と検索に用いた検索式中の分類標数の個数との関係からⅢにあげた回帰方程式

$$\log Y = b_0 + b_1 X$$

の回帰係数 b_1 、定数項 b_0 、 $\log Y$ と X の相関係数 r および回帰係数検定($H_0: b_1=0$)のための t 値を算出した。

なお、実用ファイルを用いて、4けたの分類標数で検索を行なった場合について、188種の検索式のなかから、25例ずつ任意に抽出して、上記と同様に b_1 , b_0 , r , および t を算出した。

B. 実験結果

実験の結果得られた回帰方程式の b_0 と b_1 , r および

表 IV-4. 検索実験の結果(例)
標本課題 No. 2

| 標け 数た の数 | 検 索 式 | 検 索 さ れ る 課 題 数 | |
|----------------|---------------------|-----------------|--------|
| | | 実験ファイル | 実用ファイル |
| 4 | 5502*5533*5544*3119 | 2 | 2 |
| | 5502*5533*5544 | 3 | 3 |
| | 5502*5533 | 40 | 40 |
| | 5502 | 361 | 415 |
| 3 | 550*553*554*311 | 4 | / |
| | 550*553*554 | 71 | |
| | 550*553 | 576 | |
| | 550 | 2365 | |

表 IV-5. 検索実験の所要時間

| 標数の けた数 | 検 索 式 中 の 標 数 の 個 数 | 所 要 時 間* (秒) | |
|------------|------------------------|--------------|--------|
| | | 実験ファイル | 実用ファイル |
| 4 | 4 | 4.5 | 5.3 |
| | 3 | 4.0 | 4.2 |
| | 2 | 3.6 | 3.8 |
| 3 | 4 | 182.5 | / |
| | 3 | 87.7 | |
| | 2 | 63.5 | |

* 1 回の検索についての平均値

t の値は、表 IV-6 に示したとおりである。回帰方程式のあらわす回帰直線を図 IV-2 に示す。

表 IV-6. 実験結果

a. 実験ファイルにおける結果

| 検索式の分類 標数のけた数 | 回帰係数 b_1 | 定数項 b_0 | 相関係数 r |
|------------------|----------------------|--------------|-------------|
| 4 | -0.732 (-45.532)* | 2.841 | -0.857 |
| 3 | -0.882 (-49.815)* | 3.737 | -0.902 |

V. 考 察

A. 仮説の検証

表 IV-6 に示されたとおり、相関係数 r の値は $-0.856 \sim -0.902$ であっていずれも高く、 t 値の絶対値もすべて大きい。また b_1 の値はすべて負である。このことは、検索式を構成する標数の個数の増加につれての検索される課題数の減少は、対数関係で近似される可能性を示している。

図 IV-2 に示された回帰直線の A (実験ファイル 4 けた検索)、B (同じく 3 けた検索) を比較すると、B は A に比べて上方にシフトしている。このことは、III 章の

検索式のインデクスの数と検索される課題数との相関

b. 実用ファイルにおける結果

| 検索式の分類 標数のけた数 | 回帰係数 b_1 | 定数項 b_0 | 相関係数 r |
|------------------|----------------------|--------------|-------------|
| 4 | -0.774 (-47.651)* | 2.986 | -0.867 |

c. 実用ファイル4けた検索の
サンプリング実験結果 (25 例)

| 実験番号 | 回帰係数 b_1 | 定数項 b_0 | 相関係数 r |
|------|----------------------|--------------|-------------|
| 1 | -0.754 (-16.688)* | 3.033 | -0.856 |
| 2 | -0.799 (-16.733)* | 2.998 | -0.856 |
| 3 | -0.743 (-17.956)* | 2.924 | -0.872 |
| 4 | -0.777 (-22.055)* | 2.942 | -0.902 |

* () 内は t 値

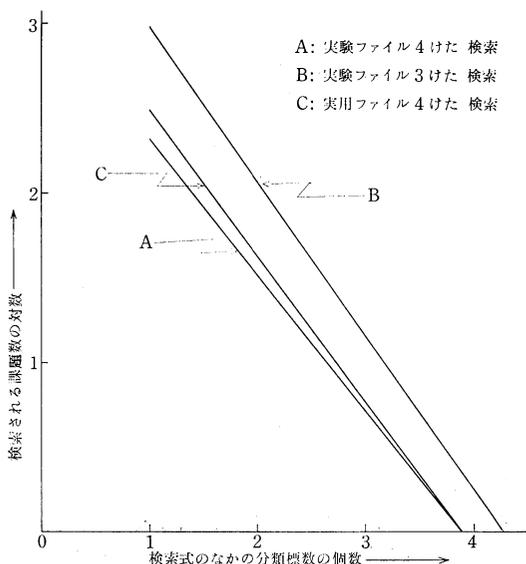


図 IV-2. 回帰直線

- B. 現実の場合、で述べられたことを支持している。したがって、仮説は検証されたことと考えられる。
- B. 実験ファイルと実用ファイルにおける回帰直線の関係

図 IV-2 に示された回帰直線の A (実験ファイル 4

けた検索)、C (実用ファイル 4 けた検索) を比較すると、検索式を構成する標数が少ない場合は、それが多い場合に比べて回帰直線上の点は上方にあり、標数の個数の増加とともに両直線上の点の差が小さくなっている。これは、実験ファイルが、実用ファイルのうち、3 個以下の標数が与えられた課題を除いたものであることから考えて当然のことである。

実用ファイルを用いての 3 けたの標数による実験は行っていないが、おそらく同じ傾向をとることと考えられる。

C. 回帰直線の傾斜

図 IV-2 の回帰直線 A (実験ファイル 4 けた検索) と B (実験ファイル 3 けた検索) との傾斜を比較すると、傾斜は B の方が大きい。(表 IV-6 a において、4 けたの $b_1 = -0.732$, 3 けたの $b_1 = -0.882$)

その原因は、III-3 で述べたように、標数 3 個以下で構成される検索式を作る方法として、a) の方法を採用し、標数 4 個のものから、ファイルにおける使用頻度の小さいものから順次取り除いて、標数 3 個、2 個および 1 個で構成される検索式を作成したこと起因すると考えられる。

これを検証するためには、III-3 における b), c) の方法による実験を行なう必要がある。

D. 実用ファイル 4 けた検索のサンプリング実験

表 IV-6 b および c の結果を比較すると、少数例を抽出して行なったサンプリング実験の結果は、多数例で行なった実験結果と良い一致を示している。このことは、実験を少数例で行なっても実際上支障がないことと考えられ、今後実験を行なう際の計画立案上有意義である。

E. 実用検索における検索される課題数

表 IV-6 b に示された結果から、実用の検索において、検索式を構成する標数の個数を決めた場合、検索される課題数を現在程度のファイルサイズにおいて推定することが可能となる。

たとえば、図 IV-2 によれば、平均的に約 20 個の課題が検索されるようにするためには、4 けたの標数 2 個を論理積の関係で結んだ検索式を用いればよいことになる。

REGISTER や RECRAS に対する情報需要は、需要者によってさまざまな形がある。たとえば、「海洋に関する研究のすべて」といった場合は、質問に該当する課題の数の多少は問わず、そのすべてを提供しなければならない。しかし、研究者からの質問には、このような包

括的な質問の例は少ない。範囲のせまい情報を求める場合、システムの出力にもとづいて、更に詳細な調査を行なうことを考えると、あまり多数の課題を出力することは実用的でない。

そこで、検索を行なう前に、機械が出力するであろう課題の数を大体限定し、それに適切な検索式を作る必要がある。このためには、検索に用いるファイルについて、検索式の標数の個数と検索される課題数との量的関係を知っておかなければならない。したがって上述の例をあげたような関係は、REGISTER を運用するうえで有用である。

VI. あとがき

REGISTER を効果的に運用し、情報需要者の要求に応えるサービスを行なうためや、RECRAS の開発のためには、種々の技術的・情報学的知見が必要である。

その知見を得るための研究の例としては、本報告で指摘しながら未だ研究途中の検索に関する問題のほか、分類表の構成に関する問題、インデクスを行なう際のインデクサの思考態度に関する問題などがあげられる。これらの諸問題を解決するためには、REGISTER および RECRAS のデータを利用した各種の実験および統計的取扱いが必要となる。これらについては、今後の研究にまつことにする。

REGISTER の開発に当っては財団外部の多くのひとびとや機関から指導、援助を受けた。特に、押田勇雄（上智大学理工学部）、古賀晴人（三菱開発株式会社）、佐々木敏雄（東京大学農学部）、沢本孝久（慶応義塾大学文学部）および中村幸雄（日本通信協力株式会社）の各位（五十音順）からは、開発の当初から引きつづいて懇篤な指導を得ている。筆者らはここに上記の各位に対して最も深厚な敬意と謝意を表す。

慶応義塾大学文学部沢本孝久教授は、本研究に対して最も貴重な示唆を与えられ、また原稿の校閲を賜わった。筆者らは、同教授に対し重ねて厚く敬意と謝意を表

する。

電力中央研究所経済研究所員浜田宗雄氏は、本研究のうち統計的理論の取り扱いについて極めて貴重な指導を与えられた。同氏に対して深く感謝する。

日本科学技術振興財団システム開発室員皆川敏一氏、同沖永哲哉氏、同財団情報管理部電算機課員鈴木浩氏からは、本研究に必要な計算プログラムの作成および計算ならびにデータ整理に関して、熱心かつ有効な援助を得た。3氏に対して厚く謝意を表す。

- 1) 中村幸雄. "研究課題検索サービス", 薬学図書館, vol. 18, 1973. 11, p. 1-13.
- 2) 津田良成. "クリアリング・サービス", 学術月報, vol. 26, 1973. 12, p. 18-22.
- 3) REGISTER に関する文献
吉村典夫. "科学技術研究情報システムの開発について," 情報管理, vol. 14, 1971. 12, p. 567-78.
吉村典夫. "なにはどこで," 化学と工業, vol. 26, 1973. 8, p. 468-70.
吉村典夫. "「科学技術分類表」について," ドキュメンテーション研究, vol. 22, 1972. 12, p. 399-409.
日本科学技術振興財団. 科学技術研究情報機械検索システム (REGISTER). 東京, 1974. 82 p.
- 4) RECRAS に関する文献
吉村典夫, 五十嵐瞳. "「農林水産試験研究課題機械検索システム」について," 農学図書館協議会会報, no. 24, 1974. (印刷中)
吉村典夫, 五十嵐瞳. "農学研究の課題検索—RECRAS について," 科学技術文献サービス, no. 41, 1974. 11. (印刷中)
農林省農林水産技術会議事務局, 日本科学技術振興財団. 農林水産試験研究課題機械検索システム概要. 東京, 1973. 123 p.
- 5) 谷 昌博, 広田広三郎, 岩崎忠雄. "JICST のクリアリングサービスについて," 情報管理, vol. 16, 1973. 9, p. 436-43.
- 6) 慶応義塾大学医学情報センター後藤敬治氏に为する, 米国立医学図書館書誌サービス部長 William H. Caldwell 氏の 1974 年 5 月 14 日づけ書信による。