OUKA
The University of Osaka
Institutional Knowledge Archive

Title	学術情報の情報科学的研究としての文検索システム SCAT-IR開発とその思想
Author(s)	新国, 三千代
Citation	大阪大学大型計算機センターニュース. 1976, 22, p. 91-103
Version Type	VoR
URL	https://hdl.handle.net/11094/65319
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

学術情報の情報科学的研究としての文検索システムSCAT-IR開発とその思想

北海道大学理学部　新国　三千代

1. はじめに

現在，数値データや文献などを対象とした学術情報の利用に関する様々な情報科学的研究が行われているが，この研究は，文献の内容に立ち入った意味論的研究の第一歩として「研究過程論」に基づいて行われたものである。

この研究は，特定研究「広域大量情報の高次処理」の一つの研究課題として共同研究[注1)]の形で過去3年間行われてきた。

2. 検索システムへの「研究過程論」導入の背景

現在の文献を対象とした検索システムは，検索者の要求に関する資料が存在するか否かその所在を知らせるものであるが，我々は，文献内容に立ち入った意味論的処法について考えたわけである。

ここでは文献の具体的なものとして，論文をとりあげて考えることにする。論文を構成する単位は，文，フレーズ，節，段落，章等が考えられるが，我々はシンタックス的に区分の明確なものとして「文」をとりあげることにした。ところで論文中の文は論文の中で一定の役割を担っていると考えることができる。即ち，論文とはその内容を明らかにしていくことができた研究過程の表現であると考えることができる。そこでこの研究において，まず研究過程を幾つかの段階に分けこれらの段階を特徴付けるカテゴリーを見出し，そのカテゴリーに対応して論文の内容を与える文の論文における役割（位置）を特徴付けるカテゴリーを見出した。今後これを文分類カテゴリーと呼ぶことにする。

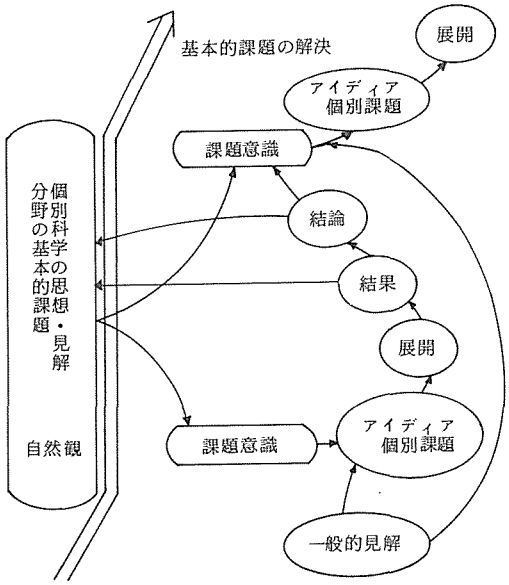
この研究では，原子核の理論的研究をとりあげ論文としてはShort note, Letter等を対象とした。

3. 文分類カテゴリー

研究過程の特徴付けは図3.1で与えられている。

これから得られる文分類カテゴリーとは次の12のものである。即ち，一般的見解，基本的課題，課題意識，課題・現状分析，アイディア，個別課題，展開，結果，評価，結論，次の課題意識及び謝辞である。

図 3.1　研究課程概略図



一般的見解　　各専門分野で，ほぼ確立されていると文献の著者が考えている見解や、専門分野を越えて成り立つと文献の著者が考え，あるいは扱っている広い見解である。この意味で文献の著者が文献の中で前提としてよいと考えているものや，研究過程で個別科学の思想・見解と名付けられたものはこのカテゴリーに入る。

基本的課題　　各専門分野で，比較的長い年月を経て次第に解決されていく課題である。基本的課題の中には，極めて長い年月を経て解決されるものもあれば，それ程の期間を要さないで解決されていくものもある。いずれも基本的課題として扱う事にする。

課題意識　　文献の著者が文献の中で解決しようとして，著者に固有な形で取り上げている課題のことであって，課題意識のままではこれを展開して研究をすすめていく事ができないものである。文献の著者が単に知識として有する場合や，又文献の中で言及しながら解決しようとしていない場合は，課題意識に属さない。異なる著者が同一の課題意識を有することは勿論ありうる。

課題・現状分析　　基本的課題と課題意識の関係，及び課題意識とこれを展開していく方法との関係に関する検討の部分である。このような検討は，文献の著者の課題に関する現状の分析である。この分析には，課題意識に関連する研究の現状認識が必要であることは言うまでもない。現状認識も現状分析に含める事にする。現状の分析には，「一般的見解」を用いることが普通である。このような一般的見解は「一般的見解」として分類するものとする。

アイディア　　課題意識がそのままの形では研究の進展が見られないが，新しい手段を展開する上での鍵となるものを見出したり，又すでに見出されたことをここで用いることに気付くなど，課題の展開はこの展開を可能にする契機を見出すことによって行われていく。このような課題展開の契機を獲得した時，課題意識をアイディアという。課題の展開を可能にする契機について述べた「文」をも，アイディアとして分類する。今まで既に見出されている事実を，課題の展開を可能にするものとして認めて取り上げた時これをもアイディ

アに含めることにする。

個別課題　課題の展開を可能にする契機を新しく見出すという非連続的な過程を経なくても，多くの研究の連続として課題を設定する事がある。このような場合には，課題の展開を可能にする契機が既に他の場合に得られていることが多い。このような課題を個別課題として分類する。

展　開　アイディアあるいは個別課題の後，これが展開された結果を得るまでの過程をいう。

結　果　展開の後に得られた事実，グラフ等は結果に分類される。結果に関する説明であって評価を含まないものは結果に含めるが，結果に対する議論であってその結果の評価に関するものは次の評価に入る。分類上疑わしい場合は評価に含める。

評価：結論　結果を課題意識あるいは個別課題にてらして評価する過程を凡て評価といい評価の結果得られたものを結論という。結論は課題意識に答えるものであるため，概念的な命題の形をとっていることが多い。結果の単なる繰り返しは結論ではない。結論に含めるべきか否か疑わしい場合は評価に含めることにする。

次の課題意識　結果や結論に基づいて導かれる次の課題意識のことである。

謝　辞　研究を開始し，遂行し，完成するまでの過程において，その研究に対して行われた協力について感謝する文で論文の終わりにまとめておこなっている部分を意味する。注として謝辞を述べている場合は注として扱う事にする。

4. 文分類カテゴリーの客観性の検証

表4.1は，6人の原子核物理理論の専門家がR. Frosh Phys. Letter 37 B(1971)P140を用いて，この論文の各文を独立にカテゴリー分類したものである。

表4.2は同様な試みを北大理学部大学院在学者6名に対して数回行い，その一致率を示したものである。

表 4.1

カテゴリー＼人	A	B	C	D	E	F	標準解
分野の一般的見解			1	1	1		
分野の基本的課題							
課題意識							
現状分析	1~4	1~4,9	2~4	2~4	2~4	1~4	1~4
アイディア	16			5. 16	5	5	
個別課題	5	5	5	5	5	5	5
展開過程	6~15 16~24	4~8 10~19 20~24	6~24	6~15 17~24	6~8 10~19 20~24	6~15 16~24 30	6~24
結果	25. 28	25. 28	25	5. 25. 28	25. 28	25. 28	25. 28
評価過程	26. 27. 29 30	26. 27 29. 30	26~30	26~30	9. 26. 27 29. 30	26. 27. 29	26. 27 29. 30
結論	31	31	5. 31	31	31	31	31
次の課題意識							
標準解と一致した数	31/31	30/31	28/31	29/31	29/31	30/31	

表 4.2

%

	A	B	C	D	E	F	平均
1回	72	66	83	83	72	72	75
2回	73	90	73	80	80	83	80
3回	94	88	94	81	94	94	91

以上のことより，文分類カテゴリーは客観性をもつという結論を得た。そこで次に述べる文検索システムの開発にとりかかった。

5. 文検索システムＳＣＡＴ—ＩＲ第一版開発

5.1 文検索システムSCAT-IRの特徴

このシステムは研究者に研究過程の各段階において必要とする情報を提供することを目的としている。この必要とする情報は，研究者が検索要求としてテーマを任意の単語からなる論理式で，更に必要とする情報の所謂叙述の述部にあたる部分をカテゴリー番号（表5.1）で与えることができる。

このSCAT-IRシステムの特徴は次のようなことである。

1. 蓄積される情報は，全文献内容である。但し，各文献の本文には研究者の手で文分類カテゴリー番号が付加されている。
2. 検索・出力の対象は全文献内容の凡てにわたるが，原則としてその最小単位は「文」である。

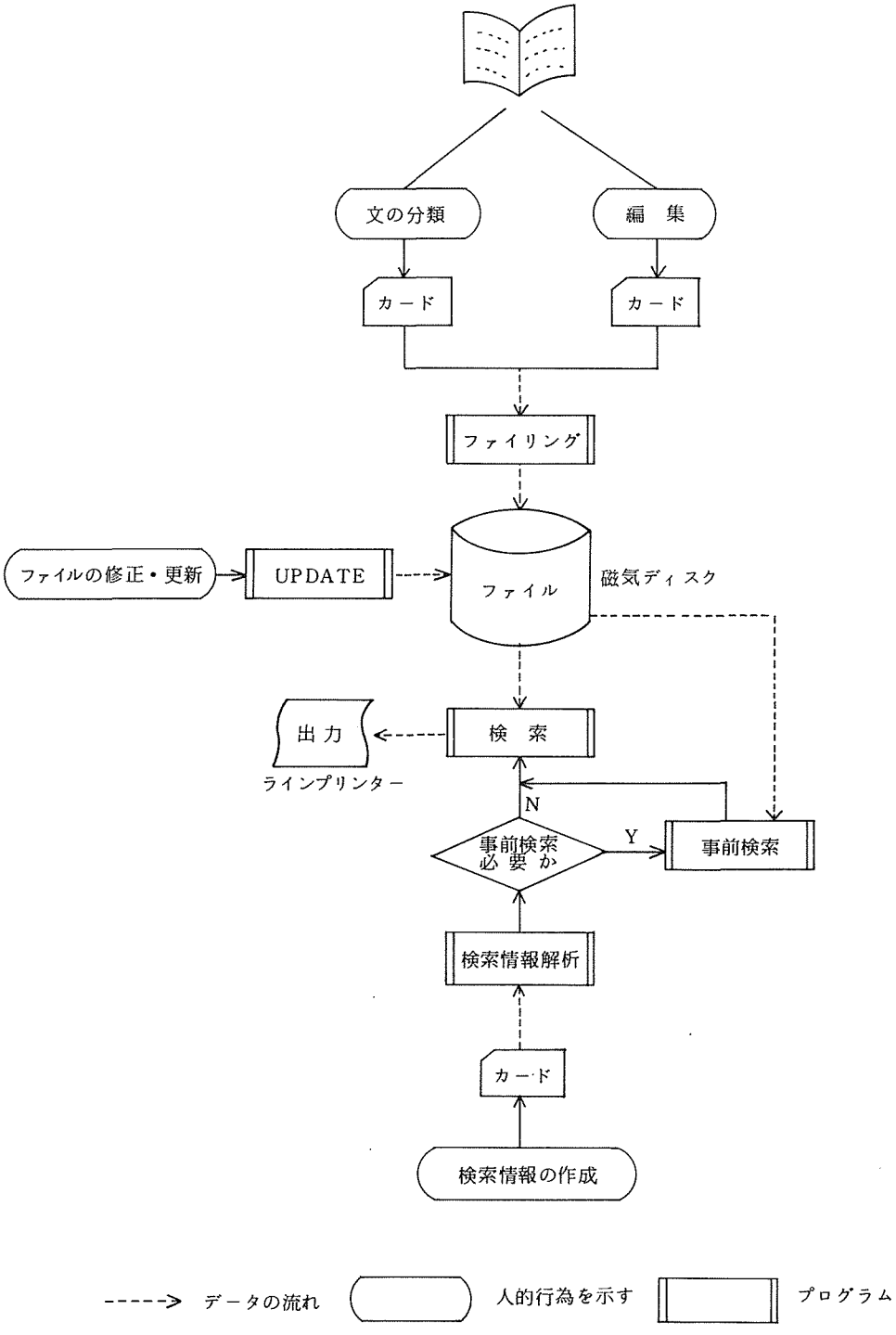
表5.1 文分類カテゴリーとカテゴリー番号表

番号	文分類カテゴリー	番号	文分類カテゴリー
01	一般的見解	07	展開過程
02	分野の基本的課題	08	結果
03	課題意識	09	評価過程
04	現状分析	10	結論
05	アイディア	11	次の課題意識
06	個別課題	12	謝辞

5.2 文検索システムSCAT-IRの概略図と計算機システムの構成

以下に文検索システムSCAT-IRの概略図を示す。

この第1版を開発するために使用した計算機システムは，HITAC 8250(コア容量64KB)磁気デスク装置2台，（1517Bのディスク使用），カード・リーダー，ラインプリンター各々1個使用している。尚プログラミング言語はPL/Iである。

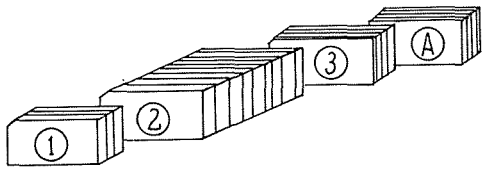

文の分類
編　集
カード
カード
ファイリング
ファイルの修正・更新
UPDATE
ファイル
磁気ディスク
出力
ラインプリンター
検　索
N
事前検索必要か
Y
事前検索
検索情報解析
カード
検索情報の作成
データの流れ
人的行為を示す
プログラム

5.3　入力データ

入力データは，論文の全内容を所定の format に従いカードにパンチしたものである。1論文のカード枚数は約80～300枚であり現在64件が蓄積されている。入力データは，図5.3.1の形でカード・リーダーにセットされる。この末尾Ⓐは研究者が論文にカテゴリー番号をふった情報を所定の format でカードにパンチしたものである。

尚本文の各文の末尾には ˇ が2個うってある。

図 5.3.1　入力データ単位



① 書誌的事項 { 雑誌名、発行年、ページ / タイトル / 著者名、所属 / アブストラクト }
① 本　文
③ 表・図の説明、リファレンス、注

5.4　検索情報の構文則

簡略化したバッカス法で示す。

〈単語〉:: =任意の単語（何字でも可）

〈区切り記号〉:: =〈 | 〉|=| , | ; | ˇ

〈演算子〉:: =＊ | + | − | (|) | 〉| 〈 | 〉= | 〈= | =

〈カテゴリー番号〉:: =01～12の数字

〈文番号〉:: =4ケタの数字，ケタ数が満たない場合右につめて左に0を補う

〈リファレンス番号〉:: =文中のリファレンス番号

〈通し番号〉:: =5ケタの数字

〈単語の論理式〉:: =＊，+，−，(，)，で構成される。但し，カッコは一重まで。

〈検索項目〉:: =CAT | STW | TLW | ABW | CSN | STN | JNL | YER | ATH |
RFN

〈検索要素〉:: =〈単語の論理式〉|〈カテゴリー番号〉|〈文番号〉|〈リファレンス番号〉|〈通し番号〉|年代（西暦）|著者名|雑誌名

〈検索単位〉:: =〈検索項目〉=〈検索要素〉|とくに例外として，YER =(〉，〈，〉=，〈=)年代

〈検索情報単位〉:: =〈〈検索単位〉〉＊〈〈検索単位〉〉|
〈検索情報単位〉＊〈〈検索単位〉〉

〈INPUT　DATA〉:: =〈検索情報単位〉，|〈検索情報単位〉；

・検索項目が指定するもの

CAT→カテゴリー番号，

STW，TLW，ABW→文（タイトル，アブストラクト）中の単語の論理式

CSN，STN，RFN→通し番号，文番号，リファレンス番号

JNL→雑誌名

YER→年代（西暦）

ATH→著者名（family nameで5人まで'+'で結んでよい。字数は問わない）

注）RFNとSTNは上昇順に番号を'+'で10個まで指定出来る。

・演算子と区切り記号

＊→ and（論理積），＋→or（論理和），－→ not（否定），→他の検索情報が後続，；→後続しない場合。

5.5 検　索

検索は現在バッチ処理で行われている。検索の種類は次のとおりである。

◦事前検索；検索情報中に書誌的事項の指定があった場合，事前検索を行い文献をしぼる。

◦文検索；該当する文を検索し出力する。

◦文献検索；タイトル・アブストラクトを利用した検索。

◦リファレンス検索；リファレンスを出力する。

5.6 検索例

文検索システムSCAT-IRは，種々の検索が可能であるがここにその代表例を挙げることにする。ワクで囲まれた部分は，検索情報を示す。

例1

```
〈CAT=10〉*〈ABW=22NA*NUCLEAR*PROPERTY〉*〈YER〈1974〉;
```

カテゴリー番号の指定　アブストラクト中のwordの指定　年代の指定

```
**GAITOU KENSUU OF PREIR=38 → 年代についての事前検索結果
```

該当分献　この文献の通し番号（システム側でつける）

```
**BIBLIOGRAPHY OF (00059)**
  1. JOURNAL              =PHYS.LETTERS
                           VOLUME=B 47 YEAR=1973 PAGE=219
  2. TITLE                =BAND STRUCTURE IN 22NA
  3. AUTHOR & POSITION=M.R.GUNYE(TATA INSTITUTE OF FUNDA-
                           MENTAL RESEARCH,BOMBAY-5,INDIA)
  4. ABSTRACT             =THE NUCLEAR PROPERTIES OF THE LOW-
                           LYING POSITIVE PARITY STATES IN 22
                           NA ARE SATISFACTORILY EXPLAINED IN
                           TERMS OF THE KL=3(TL=0),KL=0(TL=0)
                           AND KL=0(TL=1) BANDS OBTAINED FROM
```

A MULTISHELL SELF-CONSISTENT CALCULATION.

(STN 0041) THIS GOOD AGREEMENT SUPPORTS THE INTRINSIC BAND STRUCTURE IN 22NA

この文献中でCAT=10, i.e. 結論について述べた文

この検索情報中〈CAT=10〉を除くとこれは所謂文献検索である。これはabstract 中にNa22, nuclear; propertyといった単語を含むもので1974年より前に書かれた論文が欲しいことを示している。その結果出力された論文のabstractには確かにnuclear , properties, 22Naが含くまれている。このシステムでは指定された単語の前方8文字を有効として，蓄積されている論文中の文が，それを文字列として含むか否かによってマッチングの合否を判定している。この時，文中の単語の前方8文字とmatchingされるとは限らない。又8文字のwordのmatchingの場合はそのうちの一文字の違いは成功したものとしている。その結果propertiesは property と同一視されたわけである。次に出力されている文は，この論文の中で，CAT＝10 即ち結論について述べたものである。これから述べる例2 ，例3 ，例5はあるテーマについてのいろいろなapproachの仕方を示したものである。あるテーマとは「有効相互作用を用いた原子核の構造の研究 」である。各例の検索情報中の単語の論理式はそのような内容を示している。

例2

課題意識　　本文のセンテンス中のwordの論理式の指定

```
〈CAT=03〉*〈STW=EFFECT*(FORCE+POTENTIAL+INTERACT)
          *(STRUCT+SHELL+CLUSTER+PARTICLE+HOLE)〉
  *〈ATH=ROWE〉,
```

次に同一検索希望者の別の検索情報が続くことを示す。

著者名の指定

##GAITOU KENSUU OF PREIR=1 （著者名についての事前検索）

SENTENCE NUMBER　0005(00009)

該当文　OUR OBJECTIVE IS TO EXAMINE THE SYSTEMATIC,PREDICTABLE EFFECTS OF SELF-CONSISTENCY IN PARTICLE-HOLE CALCULATIONS;TO FIND OUT WHEN IT IS OR IS NOT NECESSARY TO TAKE IT INTO ACCOUNT;TO SEE HOW THE EFFECTS DEPEND ON THE INTERACTION STRENGTH AND WHAT FRACTION OF THE CORRECTIONS IS DUE TO CHANGES IN THE SINGLE-PARTICLE ENERGIES AND WHAT TO WAVE FUNCTION CHANGES

BIBLIOGRAPHY OF (00009)

1. JOURNAL　=PHYS.LETTERS
   VOLUME=B 44 YEAR=1973 PEGE=155

書誌事項

2. TITLE =IS HARTREE-FOCK SELF-CONSISTENCY IMPORTANT IN PARTICLE-HOLE CALCULATIONS?#

3. AUTHOR & POSITION=D.J.ROWE(DEPARTMENT OF PHYSICS, UNIVERSITY OF TURONTO,TURONTO, ONTARIO,CANADA)

4. ABSTRACT =THE SESTEMATIC PREDI-CATABLE EFFECTS OF SELF-CO NSISTENCY IN PARTICLE-HOLE CALCULATIONS ARE EXAMINED. IT IS SHOWN THAT SELF-CONSISTENT RADIAL WAVE FUNCTIONS ARE ESSENTIAL FOR MONOPOLE EXCITATIONS AND THAT SELF-COSISTHENCY IN DEFORMATION IS VITAL FOR THE QUADRUPOLE AND OTHER VIBRATIONS OF DEFORMED NUCLEI.

これは「ある研究者があるテーマについてどのような内容の仕事をしたか」を知りたい場合の検索例である。この場合はRowe氏について，彼の課題意識を調べたものである。その結果，1個の文と，その文を含む論文の書誌的事項が出力されている。

例3

個別課題

```
〈CAT=06〉*〈STW=EFFECT*(FORCE+POTENTIAL+INTERACT)
          *(STRUCT+SHELL+CLUSTER+PARTICLE+HOLE)〉
  *〈JNL=NUO.CIM.〉,
```

雑誌名の指定

```
##GAITOU KENSUU OF PREIR=9
  SENTENCE NUMBER  0004(00054)
      IN ORDER TO LOOK INTO THE CAUSE OF THIS,WE HAVE
      RECARRIED OUT TALMI-TYPE EFFECTIVE-INTERACTION
      SHELL-MODEL CALCULATION 1〉 FOR THE NUCLEI OF MASS
      NUMBERS 6 TO 9.
```

リファレンス番号

```
**BIBLIOGRAPHY OF  (00054)**
  1. JOURNAL =NUO.CIM.
     VOLUME=6 YEAR=1973 PAGE=224
  2. TITLE =ENERGY SHIFTS OF GROUND-STATE
     ROTATIONAL BAND IN 8BE AND 9BE
```

3. AUTHOR & POSITION=N.KUMAR(*)(INTERNATIONAL CENTRE
FOR THEORETICAL PHYSICS TRIESTE
RESEARCH SCHOOL OF PHYSICAL
SCIENCES,THE AUSTRALIAN NATIONAL
UNIVERSITY-CANBERRA)

これは，「あるテーマについて著者は誰でもよいがどのような内容の仕事がなされてきたか」，ヌーボーチメントという雑誌から調べた例である。カテゴリーとしては個別課題を指定した。

例 4

文献の通し番号の指定　リファンレス番号の指定

```
〈CSN=00054〉*〈RFN=1〉,
```

OUTPUT

*** REFERENCE OF (00054) ***

1 〉 S.COHEX AND D.KURATH:NUCL.PHYS.,73,1(1965);101A, 1
(1967);141A,145(1970)

例３の出力文中のリファレンスの内容が知りたい場合の検索例である。

例 5

結果

```
〈CAT=08〉*〈STW=EFFECT*(FORCE+POTENTIAL+INTERACT)*(STRUCT+
+SHELL+CLUSTER+PARTICLE+HOLE)〉
*〈JNL=NUO.CIM.〉,
```

##GAITOU KENSUU OF PREIR=9

SENTENCE NUMBER 0007(00054)
THE PARAMETERS OF THE SECOND COMBINATION GIVE THOSE
CHARACTERISTICS OF THE EFFECTIVE TWO-BODY NUCLEAR
INTERACTION,WHICH LOOK VERY UNREASONABLE ON
COMPARISONS WITH THE ONES OBTAINED FROM EARLIER
SHELL-MODEL CALCULATIONS,AND WITH THE ONES OBTAINED
FROM REALISTIC TWO-HODY INTERACTIONS.

**BIBLIOGRAPHY OF (00054) **

1. JOURNAL =NUO.CIM.
VOLUME=6 YEAR=1973 PEGE=224
2. TITLE =ENERGY SHIFTS OF GROUND-STATE
ROTATIONAL BANDS IN 8BE AND 9BE
3. AUTHOR & POSITION=N.KUMAR(*)(INTERNATIONAL CENTRE FOR
THEORETICAL PHYSICS-TRIESTE RESERCH
SCHOOL OF PHYSICAL SCIENCES, THE
AUSTRALIAN NATIONAL UNIVERSITY-CANBERR)

これは例３と同様であるが，結果を指定した例である。

例６

```
                          文番号の指定
                          ↓
〈CSN=00054〉*〈STN=0006+0008〉,
OUTPUT
 **STN=0006 WE FIND THAT WE CAN OBTAIN GOOD FITS FOR THEM
            IN THEIR ONLY TWO COMBINATIONS OF TWO AT A TIME,
            VIZ.FOR THE COMBINATION OF AL=6 AND 7,AND OF
            AL=8 AND 9 NUCLEI SEPARATELY.
 **STN=0008 THIS INCIDENTALLY IS NOT THE CASE WITH THE
            PARAMETERS OF THE FIRST COMBINATION,INFUSING
            US MORE FAITH IN THEM
```

例５で出力された文の前後の文を出力したい場合の例である。

例７

```
          謝辞
          ↓
〈CAT = 12〉*〈STW=HORIUCHI〉*〈YER〉=1974〉;
  ##GAITOU KENSUU OF PREIR=26
     SENTENCE NUMBER   0029(00028)
        THE USEFUL DISCUSSIONS BY THE MEMBERS OF THE
        PROJECT,ESPECIALLY BY DRS.H.HORIUCHI,A.TOHSAKI AND
        MR.S.OHKUBO ARE GREATLY ACKNOWLEGED,
  **BIBLIOGRAPHY OF (00028) **
      1. JOURNAL              =PROG.THEOR.PHYS.
                               VOLUME=51 YEAR=1974 PAGE=1621
      2. TITLE                =STRUCTURE OF TL=0 STATES IN 160
      3. AUTHOR & POSITION=YASUYUKI SUZUKI AND KIYOMI IKEDA
                               (DEPARTMENT OF PHYSICS NIIGATA
                               UNIVERSITY,NIIGATA)
```

これは，謝辞中にHORIUCHI氏の名前が書かれているものを検索した例である。これはある著者（グループ）とつながりのある研究者（グループ），更にテーマ等を知りたい場合に利用できる。

以上であるがこれらの検索は蓄積論文数が64件ということもあり検索に要する時間（検索情報の入力から，出力し終わるまで）は，平均して10秒前後である。

6. 結　果

表6.1は，5.6の例2，例3，例5で扱った単語の論理式を用いタイトル（T），アブストラクト（A）から検索したものと，任意のカテゴリー番号を指定して検索した結果である。

表中の○は求めるものを検索したことを示し，⊗はそれが目的とするものでなかったことを示す。ここでカテゴリー番号{ 03, 06, 08, 10 }は{課題意識，個別課題，結果，結論 }であって量的には一論文中に該当する文は各々高々１～２個であるが重要な部分である。この表より明らかなようにタイトル・アブストラクトからは１件しか文献がひき出せなかったが先程述べた{ 03, 06, 08, 10 }は10件とりだすことができた。このことは内容によって検索する検索方式が極めて有効であることを示唆していると思われる。

表６－１

通し番号	T	A	01	02	03	04	05	06	07	08	09	10	11
2						○							
5												○	
7					○								
9					○	○							
14						○		○					
15								○					
23											○		
24									○			○	
26												⊗	
27											⊗		
28									⊗				
37					○			○					
38												○	
39			○								○		
43						○							
45		○				○						○	
49						⊗							
54						○		○		○			
61		⊗	⊗										
62									○				
64						○			○				

7. 今後の課題

今後は蓄積データの増加をはかり文検索によって得られた文を用いて総合報告，研究動向など高次学術情報を形成することを試みることを主眼としている。

更にシステムに関しては，「研究者の様々な動的な研究活動に対して有効に対処できるようなシステムとはどのようなものであろうか ？ 」というquestionに対するanswerも，これらの研究の進行に付随して得られてゆくのではあるまいかと，自ら期待しているわけである。

以　上

注１）　研究代表者　　田　中　　一（北大理）

共同研究者　　玉垣良三（京大理），永田忍（宮崎大工），樋浦順（岩手大教養），坂東弘治（福井大工），赤石義紀（北大理）

詳しくは，「原子核文単位検索システム報告書」

昭和51年3月

特定研究「広域大量情報の高次処理」（Ｄ－２班）をごらん下さい。

各大型計算機センターでごらんになることができます。