

情報検索システムAIRによるJICST科学技術文献ファイル(情報工学)の検索 : 日本語文献データの検索

二村, 祥一
九州大学大型計算機センター

篠原, 武
九州大学大型計算機センター

永井, 徳仁
九州大学大型計算機センター

入江, 啓一
九州大学大型計算機センター

他

<https://doi.org/10.15017/1468116>

出版情報 : 九州大学大型計算機センター広報. 18 (3), pp.137-144, 1985-05-15. 九州大学大型計算機センター
バージョン :
権利関係 :



情報検索システムAIRによるJICST科学技術文献ファイル（情報工学）の検索

— 日本語文献データの検索 —

二村 祥一，篠原 武，永井 徳仁，入江 啓一，松尾 文碩*

1. はじめに

九州大学大型計算機センター（九大センター）では，1981年11月より富士通製会話型検索システムFAIRS-Iを用いて和文2次文献情報JICST科学技術文献ファイルの検索サービスを行ってきた〔1〕。九大センターでは，INSPECの検索サービスも行っている。これは1979年秋にFAIRS-Iを用いて開始したが，その後1983年11月から1984年1月にかけて検索システムを九大センターで開発したAIRに置き換えた〔2〕。これにより，ディスク使用量，検索応答時間，データベース維持などが，飛躍的に改善された。また，AIRのコマンド系はDIALOGやORBITと比べて特に優れているわけではないが，利用者の感想ではFAIRS-Iより洗練されていると好評である。

今回AIRに日本語処理機能を追加し，JICSTファイルの検索サービスもAIRに切り替えることにした。これにより，INSPECと全く同様にJICSTの検索が可能となった。AIRの日本語処理では，富士通製JEF専用ディスプレイ装置に加えて，公衆電話回線経由で使用可能な，いわゆるJOIS型漢字端末をサポートしている。

2. JICST科学技術文献ファイル

JICST科学技術文献ファイルは，JICSTが刊行している抄録誌「科学技術文献速報」の機械可読版で，周知のようにJICST自身がJOISによってこのファイルのオンライン検索サービスを行っている。本センターでサービスを行っているデータベースは，管理・システム技術編から抽出した情報工学関係だけを入れている。現在の収録量は，昭和56年度からの分であり，月に1回データの追加を行っている。この文献ファイルを検索できる者は，JICSTとの契約により，基本的には九大の利用者だけであるが，九大以外の利用者でもセンターにおいて利用することは許されている。

3. AIRによるJICST文献データの検索

AIRを用いてJICST文献データを検索する方法について説明する。使用できる端末は富士通のJEF専用ディスプレイ端末と交換回線で使用可能ないわゆるJOIS型端末である。JOIS型端末は市販のものを求めなくても，日本語表示が可能なパーソナルコンピュータに端末エミュレータを作成することで実現できる〔3〕。使用できるコマンドは表1に示されているものでINSPECの場合と全く同じである。異なる点は，検索コマンド（FIND, AND, OR, NOT）のオペランドのキーワードや属性値がローマ字→カナ変換されることである。ローマ字→カナ変換には九大センターのライブラリRKHK〔4〕を使用している。またINSPECとJICSTJでは項目の種類，名前が異なっている。表2はJICSTJの項目の一覧である。

* 九州大学大型計算機センター

使用例の説明の前にAIRで用いるローマ字→カナ変換について触れておく。利用者の入力コマンドのうち検索コマンド(FIND, AND, OR, NOT)のオペランドのキーワードや属性値(著者名や雑誌名など)は次の規則に従って変換される。

- i) ローマ字は訓令式, 日本式, ヘボン式いずれを用いても混在してもよい。(ZYOUHOU=JOUHOU)
- ii) 小さいカナはすべて大きいカナに変換する。(HATTEN→ハッテン)
- iii) 撥音(ン)の表現は" N "と母音の間をハイフン(ー)で区切る。(NIN-I→ニンイ)
- iv) ダブルクォートで囲んだものは変換しない。(" AHO "→AHO)
- v) 変換は分ち書きされた語単位で行い, ローマ字でないものは変換しない。(RONRIGATA PROGRAM→ロンリガタ PROGRAM)

(ii)の規則により通常ローマ字では入力できないカナ文字も入力できる。たとえば, エディタは正規のローマ字では表記できないがィをイとするのでEDEITAで入力することができる。

表 1. AIRのコマンド

コマンド	オペランド	機能	例
AIR	[データベース名]	AIRシステムの起動。 検索するデータベースを指定することもできる。	AIR AIR JICSTJ
BEGIN	[データベース名]	データベースを切り替える。	B JICSTJ
END		AIRシステムの終了。	END
FIND	キーワード ... キーワード [/TI /KW] {AU AA JN PY AN} = 属性値...属性値	指定された検索用語に該当する文献集合を検索する。	FIND JOURNAL SHORI F AU=TANAKA? F APL/TI
AND OR NOT		指定された検索用語に該当する文献集合と直前の検索結果との集合演算を行う。	AND SYSTEM OR SYSTEM NOT SYSTEM
COMBINE	集合演算式	すでに求められている文献集合を組合せて集合演算を行う。 *, +, - は積, 和, 差を表す。	COMBINE 1*2 C 1*(2+3) C SET1-SET2
KEEP	[集合番号] 集合名	文献集合に名前をつける。	KEEP 5 IRS
REMOVE		直前の文献集合を取り除く。	
TRACE		文献集合のトレースをとる。	T
DISPLAY	[集合番号 集合名] [SYNOUT(クラス名)] [ELEMENT(ALL MIN 項目名)] [RECORD([開始番号:]終了番号)] [MODE(BREAK NOBREAK)]	文献集合を端末やプリンターに表示する。	DISPLAY D S(K) D E(MIN) D E(TI, AU) M(N) D R(10:20)
LOOK	[ELEMENT] [ELEMENT(項目名, ..., 項目名)] [DATABASE]	データベースの説明, 項目名の一覧, 項目の説明を表示する。	LOOK L E(TI) L D

表 2. JICSTJ の項目

項目名	内 容	日本語	索 引
TIJ	日本語で書かれた論文の標題または 外国語で書かれた論文の和文標題	○	
TI	和文標題のフリガナ		キーワード
OTJ	外国語で書かれた論文の標題	○	
AUJ	著者名	○	
AU	著者名のフリガナ		属性値
AAJ	著者所属機関	○	
AA	著者所属機関のフリガナ		属性値
JNJ	雑誌あるいは資料に関する情報	○	
JN	雑誌名のフリガナ		属性値
PY	文献の発行年		属性値
ABJ	文献の主題に関する日本語の要約	○	
AN	抄録番号		属性値
KWJ	文献から抽出された主要概念を表わす用語（キーワード）	○	
KW	キーワードのフリガナ		キーワード
BN	国際標準図書番号（ISBN）		
CO	逐次刊行物を一義的に識別するためのコード（CODEN）		

4. 使用例

次の使用例では下線部が利用者の入力である。

```
(1) LOGON TSS F9999/ハスワートン
KOS70001I F9999 LAST ACCESS AT ...
READY

(2) AIR
AVAILABLE DATABASE(S)
      INSPECA      INSPECB      INSPECC      JICSTJ
SELECT DATABASE (OR TYPE "END" TO EXIT AIR): JICSTJ

(3) .LOOK DATABASE
DATABASE REMARKS
JICSTJ (INFORMATION TECHNOLOGY)          57,708 RECORDS (1981- )
      THE JICST-J DATA BASE CONSISTS OF SECONDARY RECORDS FROM JOURNAL
      ARTICLES, TECHNICAL REPORTS, CONFERENCE PROCEEDINGS, BOOKS AND
      THESE IN THE FIELD OF INFORMATION TECHNOLOGY. EACH RECORD INCLUDES
      TITLE, AUTHORS, ABSTRACT, INDEXING, AND THE OTHERS IN JAPANESE.
      CURRENTLY THIS DATA BASE IS EXPECTED TO GROW AT THE RATE OF SOME
      15,000 ITEMS PER YEAR.

(4) .LOOK ELEMENT
LISTING OF ELEMENT(S)
&AN  ABSTRACTNO      &AA  AAFFILIAT      &AU  AUTHOR
BN   ISBN            CO   CODEN          &JN  JOURNAL
*KW  KEYWORD         &PY  YEAR           *TI  TITLE
AAJ  AAFFILIATJ      ABJ  ABSTRACTJ      AUJ  AUTHORJ
JNJ  JOURNALJ       KWJ  KEYWORDJ      OTJ  OTITLEJ
TIJ  TITLEJ
      * ... KEYWORD INDEXED,      & ... ITEM VALUE INDEXED

(5) .LOOK E(TI KW)
REMARKS OF ELEMENT (TI,TITLE)
和文標題のフリガナ
REMARKS OF ELEMENT (KW,KEYWORD)
キーワードのフリガナ

(6) .FIND JINKOUTINOU
1: 2065 DOCUMENT(S) FOUND

(7) .AND PATA-NNINSIKI
1642 = ハーターニンシキ
2: 141 DOCUMENT(S) FOUND

(8) .F AU=IIJIMA?
#1: 3 = イシマ カスヒコ
#2: 1 = イシマ カスユキ
#3: 11 = イシマ シュンイチ
#4: 25 = イシマ タイソウ
#5: 1 = イシマ タカシ
#6: 1 = イシマ テツシ
#7: 1 = イシマ マサキ
#8: 1 = イシマ マサル
#9: 1 = イシカ アケヒコ
SELECT NUMBERS: 3
+E
3: 11 DOCUMENT(S) FOUND

(9) .COMBINE 2*3
4: 1 DOCUMENT(S) FOUND

(10) .DISPLAY
4: COMBINE 2*3
1/ 1
TIJ= 自立ロボット「山彦」の画像認識, JA
AUJ= 飯島純一, 油田信一, 金山裕
AAJ= 電通大, 筑波大
JNJ= 情報処理学会全国大会講演論文集, Vol. 21 ST, P. 793 - 794, 1980, JPN
ABJ= ロボットが移動する際必要となる基本的情報は、障害物の有無と障害物までの距離である。このため、超音波を利用してこれらの情報を得る機構を作成した。この方法は、単純な機構で済む、消費電力や重量が小さいという利点があるが、詳細な情報を得るには限界がある。そこで、テレビカメラを利用して画像認識の実験を実施
```

```

(11) .F JOUHOUKENSAKU?
#1: 244 = ショウホウケンサク
#2: 1 = ショウホウケンサクカンリ
#3: 241 = ショウホウケンサクシステム
#4: 1 = ショウホウケンサクヨウセンモンカシステム
#5: 40 = ショウホウケン
#6: 1 = ショウホウケンツウシンロケツコウフコウカ
#7: 1 = ショウホウケンキノウ
#8: 1 = ショウホウケンクワン
#9: 53 = ショウホウケンフコウカ
+E
(12) .F JOUHOUKENSAKU*/TI
5: 36 DOCUMENT(S) FOUND
(13) .KEEP IR
(14) .F DE-TABE-SU
6: 3429 DOCUMENT(S) FOUND
(15) .KEEP DB
(16) .TRACE
1: FIND JINKOUTINOU
2065 DOCUMENT(S)
2: AND PATA-NNINSIKI
141 DOCUMENT(S)
3: F AU=IIJIMA?
11 DOCUMENT(S)
4: COMBINE 2*3
1 DOCUMENT(S)
5: F JOUHOUKENSAKU*/TI
36 DOCUMENT(S), NAME=IR
6: F DE-TABE-SU
3429 DOCUMENT(S), NAME=DB
(17) .C IR*DB
7: 7 DOCUMENT(S) FOUND
(18) .D
7: C IR*DB
1/ 7
TIJ= 専門システムと情報検索 伝記データ管理における実験, EN
OTJ= Expert systems and information retrieval
eval: An experiment in the domain of
biographical data management.
AUJ= ZARRI G P
AAJ= Lab. Informatique pour les Sciences
de l'Homme CNRS, Paris
JNJ= Int J Man Mach Stud, Vol. 20, No. 1, P. 87
- 106, 1984, GB
ABJ= 伝記データからなる事実データベースに働く人工知能的管理システムであるRE
SEDA計画で記述されている。このシステムでは、データは格文法に基づいた知
識表現された質問に対する直接的な回答がデータベース中に見つけられないとき、RE
SEDAは変換と呼ばれる第1段階の推論手続きを使って間接的に答える。データベ
ースの因果リンクを確立する。この場合、推論操作の結果は、データベースの
新しい内容を確認したものである。RESEDAの試作システムは、すでにか動
している
+
2/ 7
TIJ= 情報検索用専門システムにおける不確実性の計算法の比較, EN
OTJ= A comparison of uncertainty calculi
in an expert system for information
retrieval.
AUJ= TONG R M, SHAPIRO D G, DEAN J S, MCCUNE
T B
AAJ= Advanced Information & Decision Syst
ems, CA
JNJ= Proc 8th Int Jt Conf Artif Intell 19
83 Vol 1, Pt 1, P. 194-197, 1983, USA
ABJ= 83年度に生起する確率演算(AND, OR, detachment,
不確実性をとらえる。しかし理論して、その良否を4種合計12種の評価法の
implification)に對して、その良否を4種合計12種の評価法の
設定が可能である。この12種の良否を4種合計12種の評価法の
に關する知識を言明化した後、30種の簡単な話題を選び、それらがテロとど
のような関係をもつかという真実度を計算した。その結果を評価する二つの方法

```

を設定し、モデルを評価した。結論として、明らかに不良なモデルを除くと、各モデルの絶対評価は困難であることを示すにすぎない

- (19) $\frac{+E}{-D \ E(MIN)}$
7: C IR*DB
1/ 7
TIJ= 専門家システムと情報検索 伝記データ管理における実験, EN
OTJ= Expert systems and information retrieval
eval: An experiment in the domain of
biographical data management.
AUJ= ZARRI GP
AAJ= Lab. Informatique pour les Sciences
de l'Homme CNRS, Paris
JNJ= Int J Man Mach Stud, Vol. 20, No. 1, P. 87
- 106, 1984, GBR
- (20) $\frac{+E}{D \ E(ALL)}$
7: C IR*DB
1/ 7
TIJ= 専門家システムと情報検索 伝記データ管理における実験, EN
OTJ= Expert systems and information retrieval
eval: An experiment in the domain of
biographical data management.
AUJ= ZARRI GP
AAJ= Lab. Informatique pour les Sciences
de l'Homme CNRS, Paris
JNJ= Int J Man Mach Stud, Vol. 20, No. 1, P. 87
- 106, 1984, GBR
ABJ= 伝記データからなる事実データベースに働く人工知能的管理システムであるRE
SEDA計画について述べて、このシステムでは、データは格文法に基づいた知
識表現言語で記述されており、計算の核部分では、推論解釈プログラムでと
RESEDAは変換と呼ばれる第1段階の推論手続を使って間接的に答えよ
うとする。この場合、推論操作の結果は、データベースの新
しい因果リンクを確立する。この場合、推論操作の結果は、データベース
もとの内容を修正したものとなる。RESEDAの試作システムは、すでに動
作している。
KWJ= データ管理, 人工知能, 人工知能推論, 知識表現, データベース, 情報検索, エ
キスパートシステム
AN = 4249691
CO = IJMMB
- (21) $\frac{+E}{-D \ E(TI \ AU) \ M(N)}$
7: C IR*DB
1/ 7
TI = センモンカシステム ト ショウホウケンサク テンキテマカンリ ニ オケル シツケン
AU = ZARRI GP
2/ 7
TI = ショウホウケンサクショウモンカシステム ニ オケル フカクシツセイ ノ ケイサンホウ ノ ヒカク
AU = TONG R M, SHAPIRO D G, DEAN J S, MCCUNE B P
3/ 7
TI = ショウホウケンサク ハノ ルールソツカンタシユホウ ケツカト コメント
AU = TONG R M, SHAPIRO D G, MCCUNE B P, DEAN J S
4/ 7
TI = ショウホウケンサクシステム (FAIRS II) ト イメーシショウホウ
AU = スナモト タカオ
5/ 7
TI = ショウホウケンサク ト シンコウチノウ ト ノ カカワリアイ
AU = DEJONG G
6/ 7
TI = テマヘースキシツ ト ショウホウケンサク
AU = ナカイ ヒロシ
7/ 7
TI = タマヨウリョウキョクシステム (MSS) オ モチイタ ショウホウケンサク ノ セイノウ ニ ツイテ
AU = オサワ ヨシアキ, ホリイキ ヒロミ, ワタナベ トヨヒテ, ホシノ サトシ

```
(22) .D SY(K)
      7: C IR*DB
```

```
(23) .END
      READY
```

```
(24) LOGOFF
      RETURN CODE : 0000
      KEQ56470I F9999 LOGGED OFF AT ...
```

- (1) TSSのセッションを開設する。
- (2) AIRコマンドを用いてAIRシステムを起動する。データベース名を省略するとシステムが利用可能なデータベース名を一覧表示するのでその中から検索したいデータベースを選択する。
- (3) AIRの初期設定が終了すると端末に，' '（ピリオド）が表示されAIRのコマンドが入力可能になる。LOOKコマンドを用いてJICSTJデータベースの説明を表示する。JICSTJには、現在57,708件の文献があることがわかる。
- (4) LOOKコマンドを用いてJICSTJデータベースの項目定義を表示する。項目名、別名が表示されている。*、&はそれぞれキーワード索引、項目値索引が作られている項目名を示す。
- (5) LOOKコマンドを用いて項目名TI,KWの説明を表示する。
- (6) FINDコマンドを用いてキーワードにジコウチノウをもつ文献を求めている。

```
1: 2065 DOCUMENT(S) FOUND
```

これはジコウチノウをキーワードとする文献が2065件見つかり、この文献集合に集合番号1が与えられていることを示している。AIRでは検索した集合に対して1から順に集合番号が割り当てられる。

- (7) ANDコマンドを用いて直前の検索集合とパターンニシキをキーワードとして持つ文献集合との積集合を求めている。この他、集合演算のために、NOT（差集合）、OR（和集合）の二つがある。
- (8) FINDコマンドを用いて著者名がIIJIMA?のものを検索している。ここでFはFINDの省略形で各コマンド、およびコマンドのオペランドは他のものと区別できる範囲で省略できる。ただし、AIR、ENDコマンドは省略できない。?は著者名のつづりを正確に知らないときなどに指定する。このコマンド入力により、システムはIIJIMA付近の著者名を一覧表示するので、この中より適当なものを選び指定する。ここでは3を指定し著者名がイイジマ ジュンイチのものを検索している。一覧表示の中のを複数指定する場合は、2,3,4のように指定する。この状態で一覧表示を継続するか聞いてくるのでD、U、Eのいずれかを指定する。D、Uはそれぞれ一覧表示を下方、上方に続けることを、Eは一覧表示をやめることを指示する。
- (9) COMBINEコマンドを用い、文献集合2と3の積集合を求めている。COMBINEコマンドはすでに求められている文献集合を用いて集合演算を行うコマンドで、演算子には、*（積）、+（和）、-（差）の三つがありオペランドに1*（3-4）のような複雑な式も書くことができる。
- (10) DISPLAYコマンドで標準の編集出力（標題、原文標題、著者名、所属機関名、雑誌、抄録文）を行っている。
- (11) FINDコマンドを用いてJOUHOUKENSakuの近くの索引語を表示している。
- (12) FINDコマンドを用いて標題にJOUHOUKENSaku*をもつ文献を求めている。*は任意の文字列である。この場合ジョウホウケンサク、ジョウホウケンサクカンリ、ジョウホウケンサクシステム、ジョウホウケンサクヨウセンモンカシステムが該当する。また、この例では検索を標題に限

定するために ' /TI ' を付加している。

- (13) KEEP コマンドを用いて集合番号 5 に IR という名前をつけている。
- (14) FIND コマンドを用いてデータベースをキーワードとしてもつ文献を求めている。
- (15) KEEP コマンドを用いて集合番号 6 に DB という名前をつけている。
- (16) TRACE コマンドを用いて文献集合をトレースしている。文献集合を作り出したコマンド、該当文献数、集合名が表示される。
- (17) COMBINE コマンドを用いて文献集合 IR と文献集合 DB との積集合を求めている。
- (18) DISPLAY コマンドを用いて標準の編集出力を行っている。AIR では標準的には複数の文献は 1 件ずつ表示され、1 件分の表示が終わると '+' が表示される。さらに続けて表示するには RETURN キーを押せばよい。空白以外の任意の文字を入力した場合は DISPLAY の処理は終了する。
- (19) DISPLAY コマンドを用いて MIN を指定した編集出力(標題, 原文標題, 著者名, 所属機関名, 雑誌)を行っている。
- (20) DISPLAY コマンドを用いて全項目 (ALL 指定) の出力を行っている。
- (21) DISPLAY コマンドを用いて著者名と標題を出力している。MODE (NOBREAK) を指定し文献を連続して出力させている。
- (22) DISPLAY コマンドを用いて標準の編集出力をラインプリンタに出力させている。
- (23) END コマンドを用いて AIR の処理を終了する。
- (24) TSS セッションを終了する。

4. おわりに

今回追加した AIR の日本語処理機能は、FAIRS-I のものときほど違わない。しかし、AIR は九大センターで開発したものであるから、FAIRS-I と異なり、機能の強化・改善を自由に行える。今後日本語文献検索の研究を進め、また広く利用者の意見を取り入れて、機能の強化・改善に努めたい。

参考文献

1. 二村, 平野, 入江, 高木, 鬼塚, 古城, 松尾 FAIRS-I / JEF による JICST 科学技術文献ファイル (情報工学) の検索, 九大大型計算機センター広報, **15**, 1, 1982, 13-20.
2. 二村, 篠原, 松尾 情報検索システム AIR による INSPEC の検索, 同上, **17**, 1, 1984, 1-22.
3. 入江, 永井, 篠原, 松尾 JOIS 型漢字端末エミュレータの作成について, 同上, **18**, 3, 1985, 222-224.
4. 石原 ローマ字-カナ変換用サブプログラム RKHKN について, 同上, **12**, 1, 1979, 20-28.