数理解析研究所講究録
第525巻 1984年 42-56

12

文献情報検索システムＡＩＲ

篠原　武　(Takeshi Shinohara)

二村祥一　(Shooichi Futamura)

松尾文碩　(Fumihiro Matsuo)

(九州大学大型計算機センター)

１．はじめに

情報検索（ｉｎｆｏｒｍａｔｉｏｎ　ｒｅｔｒｉｅｖａｌ）は，自然言語で書かれた情報を扱うところに本質がある．その情報に対する検索機構として，とくに大量の文献データを扱うときは，現在のところキーワードの転置索引（ｉｎｖｅｒｔｅｄ　ｉｎｄｅｘ）以上のものがなく，したがって検索結果は近似的で正確さに欠ける．しかし，これに代わる機構が近い将来に実用化される見込みはない．反面，この技術の停滞によって，文献情報検索は安定した技術のような印象を与えており，それはまた定型的処理であるかのような誤解を招いている．

著者らは，以上のような認識の上に立って，文献情報検索システムＡＩＲ（Ａｕｇｍｅｎｔｅｄ　Ｉｎｆｏｒｍａｔｉｏｎ　Ｒｅｔｒｉｅｖａｌ　ｓｙｓｔｅｍ）を開発している．ＡＩＲ第一版はディスク領域効率と検索効率を重視して設計・実現したもので，現在これを用いてＩＮＳＰＥＣテープのオンライン検索サービスを行っている．ＡＩＲ第一版では，自然言語情報に対する検索機能自体は従来の情報検索システムの

水準と大差ないが，ディスク使用量は３．５分の１以下，検索処理速度は１０倍以上高速であり，その効率は非常に優れている．

２．開発の背景と設計方針

九州大学大型計算機センター（以下，〃九大センター〃と略す）では，１９７９年から富士通会話型情報検索システムＦＡＩＲＳ－Ｉ［１］によって，ＩＮＳＰＥＣテープ［２］の検索サービスを行ってきた．それは，かなりの数の利用者にとって，研究を遂行するうえで不可欠のものになっている．しかし，２次文献検索は大量のディスク領域を必要とする以外にも，システム維持のために大量の計算時間を必要とし，これらは九大センターにとって少なからぬ負担になっていた．ＦＡＩＲＳ－Ｉは，多くの情報検索システム同様，キーワードの転置ファイルを作る型のシステムである．ところで，大量文献データに対するキーワードの選択は，自然言語情報の単語から不要語（ｓｔｏｐ　ｗｏｒｄｓ）を除去する以外に実用に耐える方法がない．キーワードを選択する対象から抄録を除外すると，不要語除去のための計算及び転置ファイルのディスク領域にかかる負担はかなり軽減される．しかし，情報検索システムとは自然言語情報を扱うシステムであるとの見方からすれば，２次文献中最も豊富な情報を含んでいる抄録からのキーワードを転置ファイルに入れないことは，情報検索の質を著しく損なうことになるであろう．

以上のような背景から，ＡＩＲの設計に当り，つぎの２点に第一の目標を置いた．

（１）良質の文献情報検索サービスの継続をできるだけ軽い負担で可能にする効率の高い検索システムを実現する．

（２）（１）の〃良質〃とは，情報検索では自然言語情報に対する検索能力に対する評価である．キーワードの転置索引による検索には限界があることは明らかなので，必要なときには検索時に直接文章データを〃

読む〃機能を充実させる．

（２）の機能として，将来は自然言語理解機構を開発し，文章データを〃理解〃できるシステムにする計画であるが，ＡＩＲ第一版ではパターン照合機械［３］による文字列探索程度にとどめる．したがって，第一版では，効率に重点を置いたシステム設計を行った．情報検索システムの効率はそのファイル構成によって決まる．ここでは，２次文献などの文書情報を格納したファイルを文書ファイルと呼び，文書ファイルに対する索引を格納したファイルを転置ファイルと呼ぶことにする．以下，この節ではこれらのファイルについての設計方針を述べる．

２．１　文書ファイルのデータ圧縮

筆者らは，ＱＯＣ［４］と名付けた単語を要素とするデータ圧縮技法を考案した．ＱＯＣは，理論的にはＺｉｐｆの法則が成立するすべての言語に対して９７．２１％という高い圧縮効率を有し，文書データを１／４以下に圧縮する．ＱＯＣは，最適符号ではないが，Ｈｕｆｆｍａｎ符号に比べると符号器と復号器の実行速度の点で決定的に勝っている．ＩＮＳＰＥＣの抄録における単語の生起頻度特性はほぼＺｉｐｆの法則に従っており［５］，ＱＯＣは抄録，標題などの項目を１／４に圧縮する．２次文献データ中には，ＱＯＣに適さない項目があり，これらを文字単位で圧縮すると，ＩＮＳＰＥＣテープに関しては全体として圧縮率は３．６程度となる［６］．ＩＮＳＰＥＣテープのような大量の文献情報を取り扱う場合には，文書ファイルの大きさを１／３．６に圧縮することによる利益は非常に大きい．また，ＡＩＲのようなオンライン検索システムでは，復号器の実行速度が重要であるが，ＦＡＣＯＭ　Ｍ－３８２上では復号器は７２５ナノ秒／字の速度で走るので，ＱＯＣを採用しても検索結果の出力に及ぼす影響は全くないと言ってもよい．そこで，ＡＩＲではＱＯＣなどの圧縮技法によって文書ファイルを圧縮することにした．

2．2　転置ファイルと不要語選択方式

ＡＩＲの転置ファイル，すなわち文書ファイルに対する索引は，キーワード転置ファイルと属性転置ファイルから成る．前者は，抄録，標題などからのキーワードを索引の見出し語とする．後者は，文献に対する属性，たとえば著者，雑誌などについての索引であり，見出し語は，著者名，誌名などである．属性転置ファイルは，本質的に通常のデータベース・システムの転置ファイルと同じであり，文献情報検索ではこの部分の構造はそれほど効率に影響を与えないと思われるため，ＡＩＲではこれをＢ木に基づく標準的な索引技法によって実現することにした．

キーワード転置ファイルの設計にあたっては，まずキーワード索引の対象について考慮しなければならない．キーワードを抽出する項目としては，標題，自由索引句，抄録が考えられるが，抄録を索引の対象とすると，キーワード転置ファイルが非常に大きくなる．しかし，最も多くの情報を含んでいる抄録を索引の対象から外すことは好ましくない．すでに筆者らが開発した不要語選択方式［７］によれば，検索の質をそれほど落とすことなく，キーワード転置ファイルの大きさを約半分にできる．この不要語選択方式は，自由索引句と抄録における単語の生起頻度の比を指標とするもので，自由索引句にはあまり現われず抄録に多く現われる単語が不要語とされる．

さらに，キーワード転置ファイルの設計において問題となるのは，転置ファイルで隣接演算（ａｄｊａｃｅｎｔ　ｏｐｅｒａｔｉｏｎ）［８］を行うかどうか，すなわちキーワードの生起位置情報を転置ファイル内に持つかどうかである．著者らのＩＮＳＰＥＣテープについての調査では，生起位置情報を持つ場合，ファイルの大きさは８０％増加する．また，生起位置情報を持たない転置ファイルの場合，句による検索を，句を構成する単語を含む文献集合の積で近似すると，質の悪化は検索結果が平均３倍増加するという形で現われる［９］．自然言語文にお

ける単語の順序が持つ情報量の大きさ，ならびに情報検索の評価としては精度（ｐｒｅｃｉｓｉｏｎ）より再現率（ｒｅｃａｌｌ）の方が重要であることなどを考えれば，この３倍という数字はそれほど悪いものではないと思われる．そこで，ＡＩＲでは隣接演算は，転置ファイルでは行わないこととし，必要なときには，パターン照合機械によって文書ファイルの順探索を行うこととした．

キーワード転置ファイルの索引部にはＱ木［１０］を採用した．これは，根（ｒｏｏｔ）と葉（ｌｅａｆ）からなる多岐平衡木であり，一般に根と葉の大きさは異なる．これにＱＯＣの符号表を高頻度単語の索引としても使えるように拡張することにより，１単語検索に要する索引部の平均ディスクアクセス回数を０．０５回程度にすることができる．

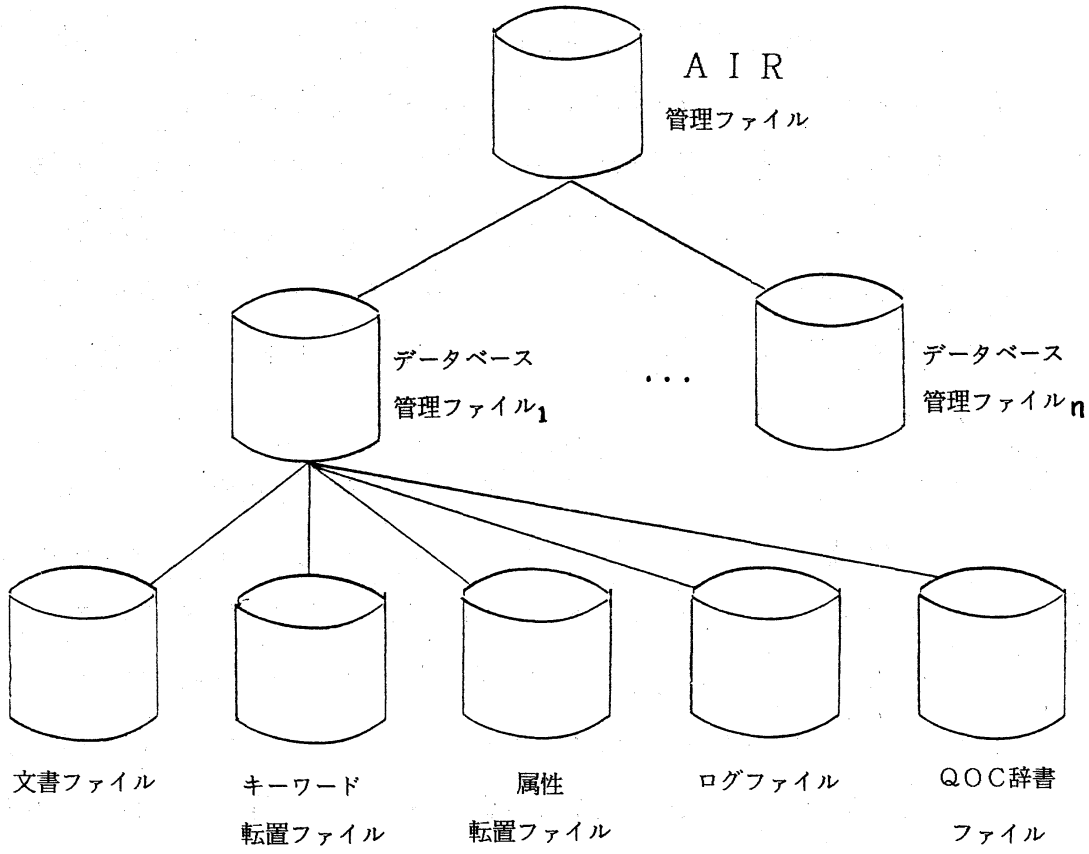

図１．ＡＩＲのファイル構成

3．ＡＩＲのファイル構成

ＡＩＲのファイル構成を図１に示す．ＡＩＲでは，検索対象の単位をデータベースと呼ぶ．ＡＩＲの情報検索者は，ある時点では一つのデータベースだけを利用できる．データベースは，文書ファイル，キーワード転置ファイル，属性転置ファイル，ＱＯＣ辞書ファイル，ログファイルから構成される．ＱＯＣ辞書ファイルは文書ファイルの圧縮を行うために必要な辞書である．これらのファイルの所在やデータベースにあるレコードの件数等はデータベース管理ファイルにある．データベース管理ファイルは，このほかレコードを構成する項目の定義情報（項目名，符号化法，転置ファイル作成の有無，出力の指示など）やデータベースの機密保護情報をもっている．データベース管理ファイルの所在情報はＡＩＲ管理ファイルにある．

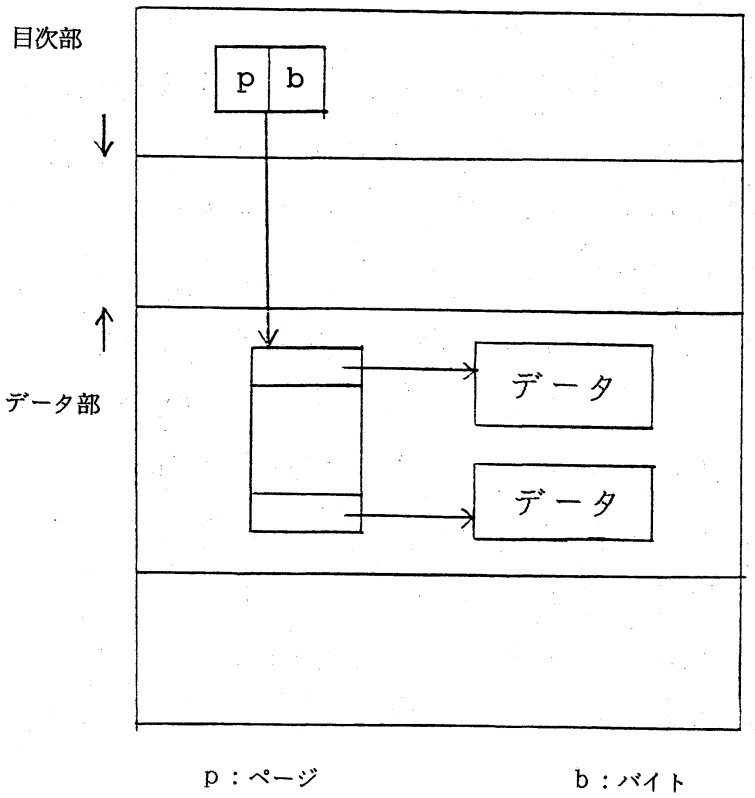

図２．文書ファイルのデータ構造

3．1　文書ファイルのデータ構造

文書ファイルは1個あるいは複数のデータセットから構成される．文書ファイルのレコードには，各レコードを一意に識別するために1から順に参照番号（24ビット；最大16，777，215までを管理）が与えられる．各データセットにはある範囲の参照番号のレコードが入れられる．どのデータセットにどの範囲の参照番号のレコードがあるかは，データベース管理ファイルに記録されている．文書ファイルを構成する各データセットは，データ部及びその目次部に分けられる（図2参照）．データ部には，符号化された圧縮レコードが，レコードを構成する項目の索引とともに入れられる．目次部には，圧縮レコードの格納番地（ページ番号，先頭バイト位置の組；それぞれ2バイト）が設定され，参照番号が与えられた場合に，2回以下のディスク参照で該当レコードを見つけることができる．

3．2　転置ファイルのデータ構造

キーワード転置ファイルには標題，抄録などの文章情報の項目から不要語フィルタを通して得られたキーワードが，その生起情報（参照番号，項目番号）とともに記録される．図3にキーワード転置ファイルのデータ構造を示す．キーワード転置ファイルは索引部とデータ部からなり，索引部にはキーワードとその関連情報（該当文献数，データ部へのポインタ；計7バイト），データ部にはキーワードの生起情報（項目番号，参照番号の組のリスト；計4×nバイト）が設定される．

索引部には高さ1の多岐平衡木であるQ木を採用した．Q木の根は常に主記憶に置かれて処理される．Q木の葉へのアクセスには1回のディスク参照を必要とするが，この部分にQOCの符号表（生起頻度の高い8191単語が登録されている）を併用することにより処理の高速化を図っている．

データ部の要素は，項目番号（8ビット），文献番号（24ビット）

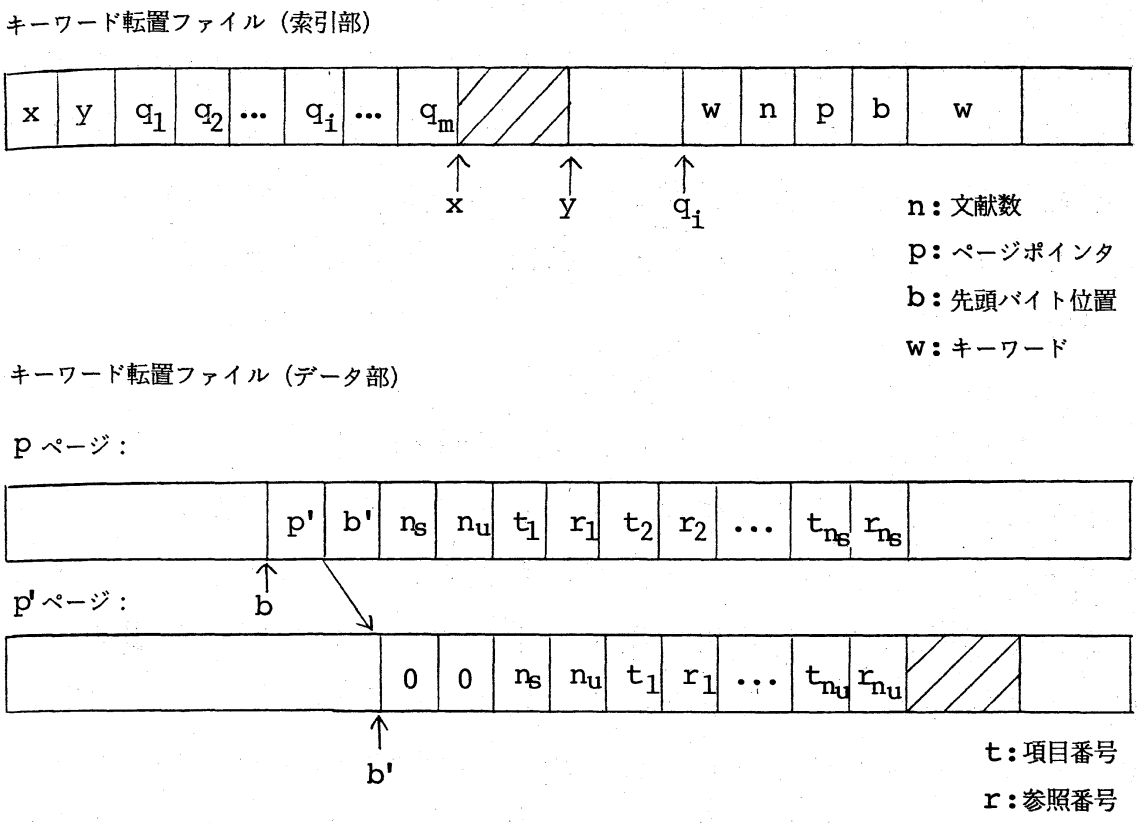

図３．キーワード転置ファイルのデータ構造

の対で構成される．項目番号の各ビットによって項目が指定される．したがって，最大８項目までのキーワード転置ファイルの作成を指示できる．転置データはリストとして，データ部の各ページの連続領域に置かれる．ページサイズを超える場合には，ページポインタを使って，さらに次の連続領域が割り当てられる．ディスク参照回数を減少し効率を上げるため，データ部のページサイズはディスクのトラックサイズに一致させている．また，転置データの追加による効率の低下を防ぐため，データ部には創成時に，各キーワードに対してその出現頻度に応じた空領域が自動的にとられる．

キーワードが与えられ，転置データを取り出すのに必要なディスク参照回数は，転置データの最初の部分に対しては１～２回であり，残りの

部分についてはページの境界で１回ずつ必要である．

属性転置ファイルには項目の値そのものを登録する．属性転置ファイルは項目別に作成され，それは一つのデータセットにまとめて入れられる．属性転置ファイルは索引部とデータ部からなる．索引部にはＢ木を採用した．データ部は，キーワード転置ファイルのデータ部と同様のものである．

４．ＡＩＲによるＩＮＳＰＥＣデータベースの構築

ＡＩＲにより，１９７３年以降のＩＮＳＰＥＣテープからデータベースを構築した．ＩＮＳＰＥＣテープの各文献は，Ａ（物理学），Ｂ（電気・電子工学），Ｃ（計算機・制御工学）の３つの分野に分類できる．それぞれに対応してＩＮＳＰＥＣＡ（約１１０万件），ＩＮＳＰＥＣＢ（約５７万件），ＩＮＳＰＥＣＣ（約３７万件）の３個のデータベースを作成した．複数の分野に属する文献が約２０％あり，ディスク領域は約２０％余分に必要になるが，検索精度と検索効率の点から３個に分けた．

ＩＮＳＰＥＣデータベースの文献レコードは，標題，著者，著者所属機関，雑誌名，雑誌の巻・号，発行年月日，出版社，抄録，自由索引句などの３１項目から構成される．このうち標題，抄録，自由索引句はＱＯＣを用いて圧縮した．ＱＯＣ辞書には，生起頻度の高い８１９１単語を登録するが，これを分野別に作成し圧縮率の向上を図った．著者所属機関，出版社などの圧縮にもＱＯＣを用いた．ただし，この場合のＱＯＣ辞書は，標題，抄録，自由索引句のものとは別に作成し，各分野とも共通のものを用いた．他の項目については，１字あたり４ビットあるいは５ビットの符号化を行った．

キーワード転置ファイルには，標題，抄録，自由索引句からのキーワードを登録し，属性転置ファイルには著者，雑誌名などの索引を用意した．

```
(1)   LOGON TSS F0026
      + PASSWORD ?
      ********
      KDS70001I F0026    LAST ACCESS AT 21:43:17 ON 83.341
      KEQ56455I F0026 LOGON IN PROGRESS AT 21:03:44 ON DECEMBER 9, 1983
      JOB NO = TSU8024 CN(01)
      KEQ56951I NO BROADCAST MESSAGES
      READY
(2)   AIR INSPECC
(3)   .FIND INFORMATION RETRIEVAL
             36879 = INFORMATION
              5023 = RETRIEVAL
        1:    3486 DOCUMENT(S) FOUND
(4)   .AND INDEXING
        2:     297 DOCUMENT(S) FOUND
(5)   .FIND AU=SALTON?
        #1:     33 = SALTON, G.
        #2:      2 = SALTOR, F.
        #3:      4 = SALTSBURG, H.
        #4:      1 = SALTSMAN, M.R.
        #5:      1 = SALTSMAN, S.
        #6:      1 = SALTSMAN, T.
        #7:      2 = SALTUS, G.E.
        #8:      1 = SALTYKOV, A.I.
        #9:      1 = SALTYSOV, I.V.
      SELECT NUMBERS: 1
      +E
        3:     33 DOCUMENT(S) FOUND
(6)   .COMBINE 2*3
        4:      8 DOCUMENT(S) FOUND
(7)   .DISPLAY
        4: COMBINE 2*3
             1/          8
      TI = A BLUEPRINT FOR AUTOMATIC INDEXING
      AU = SALTON, G. (DEPT. OF COMPUTER SCI., CORNELL UNIV., ITHACA, NY, USA)
      SO = SIGIR FORUM (USA), VOL.16, NO.2, 22-38, FALL 1981
      AB = SUMMARIZES SOME OF THE CURRENTLY AVAILABLE INSIGHTS IN AUTOMATIC
           INDEXING. THE EMPHASIS IS ON ASPECTS THAT ARE EXPECTED TO BE USEFUL IN
           A PRACTICAL AUTOMATIC INDEXING APPLICATIONS. THE DISCUSSION IS
           NECESSARILY CURSORY, BUT THE REFERENCES WILL LEAD INTERESTED READERS TO
           A DEEPER TREATMENT OF THE INDEXING PROBLEM
      +Z
             2/          8
      TI = EFFECTIVE AUTOMATIC INDEXING USING TERM ADDITION AND DELETION
      AU = YU, C.T., SALTON, G., SIU, M.K. (UNIV. OF ALBERTA, EDMONTON, ALBERTA,
           CANADA)
      SO = J. ASSOC. COMPUT. MACH. (USA), VOL.25, NO.2, 210-25, APRIL 1978
      AB = IN INFORMATION RETRIEVAL INDEXING IS THE TASK CONSISTING OF THE
           ASSIGNMENT TO STORED RECORDS AND INCOMING INFORMATION REQUESTS OF
           CONTENT IDENTIFIERS CAPABLE OF REPRESENTING RECORD OR QUERY CONTENT. IF
           THE INDEXING IS PERFORMED AUTOMATICALLY AND THE RECORDS ARE WRITTEN
           DOCUMENTS, AN INITIAL SET OF INDEX TERMS MIGHT BE CHOSEN BY TAKING
           WORDS EXTRACTED FROM DOCUMENT TITLES OR ABSTRACTS; THIS INITIAL TERM
           ASSIGNMENT MIGHT THEN BE IMPROVED BY ADDING RELATED TERMS CHOSEN FROM A
           THESAURUS, BY DELETING EXTRANEOUS OR MARGINAL TERMS, AND BY REPLACING
           SINGLE TERMS BY TERM COMBINATIONS AND PHRASES. FORMAL PROOFS ARE GIVEN
           OF THE RETRIEVAL EFFECTIVENESS UNDER WELL-DEFINED CONDITIONS OF
           INDEXING POLICIES BASED ON THE USE OF SINGLE TERMS, TERM ADDITIONS AND
           DELETIONS, AND TERM COMBINATIONS OR PHRASES
      +N
(8)   .END
      READY
(9)   LOGOFFC
      RETURN CODE : 0000
      CPU TIME(   1.85SEC.) USE TIME(   5MIN.) REGION SIZE(1056KB)
      INPUT(    12LINES) OUTPUT(      54LINES) EXCP(     166TIMES)
      SESSION CHARGE(TSU8024,21:03:44)                              24YEN
      TOTAL CHARGE SINCE 04/01/83 (EXCEPT THIS SESSION'S)      97,711YEN
      KEQ56470I F0026 LOGGED OFF AT 21:07:53 ON DECEMBER 9, 1983+
      KEQ54100I SESSION ENDED
```

図４．ＡＩＲの使用例

5．ＡＩＲによるＩＮＳＰＥＣの検索

ここでは，図４に示した検索例従ってＡＩＲの使用法を簡単に説明する．下線部は利用者の入力である．

(1)ＡＩＲを使用するためＴＳＳセションを開設する．

(2)ＡＩＲを起動する．データベースとしてＩＮＳＰＥＣＣを指定している．データベース名を省略した場合や指定したデータベースが登録されていない場合は，利用可能なデータベースの一覧が表示されるのでメニュー選択する．

(3)ＩＮＦＯＲＭＡＴＩＯＮ及びＲＥＴＲＩＥＶＡＬをキーワードとして持つ文献を検索している．システムは，ＩＮＦＯＲＭＡＴＩＯＮ，ＲＥＴＲＩＥＶＡＬをキーワードとして持つものがそれぞれ３６８７９，５０２３件あり，結果として３４８６件見つかったことを表示している．この文献集合には集合番号１が割り当てられている．

(4)ＡＮＤコマンドにより，キーワードＩＮＤＥＸＩＮＧを持つ文献集合と直前の検索集合との積集合を求めている．２７９件見つかった．この文献集合には集合番号２が割り当てられている．直前の検索集合との集合演算を行うコマンドは，ＡＮＤのほかにＯＲ（和），ＮＯＴ（差）がある．

(5)ＦＩＮＤコマンドにより著者名がＳＡＬＴＯＮ？のものを検索している．ＳＡＬＴＯＮ付近の著者名が一覧表示され，１を選択することにより〃ＳＡＬＴＯＮ，　Ｇ．〃を検索している．３３件見つかった．最後の〃Ｅ〃の入力は一覧表示を止めることを指示している．

(6)ＣＯＭＢＩＮＥコマンドにより集合番号２の２９７件と集合番号３の３３件の積集合を求めている．８件見つかった．ＣＯＭＢＩＮＥコマンドで用いることができる演算記号は＊（積），＋（和），－（差）である．

(7)ＤＩＳＰＬＡＹコマンドにより見つかった文献を表示している．ＡＩ

Ｒでは標準的には複数の文献は１件ずつ新しいものから順に表示される．１件分の表示が終わると〝＋〟が出力される．続けて表示するには復改キー（ＲＥＴＵＲＮキー）を押す．表示をやめる場合は任意の文字を入力する．

(8)ＡＩＲを終了している．

(9)ＴＳＳを終了している．

６．ＡＩＲの評価

６．１　ディスク使用量

ＡＩＲでは文献データをＱＯＣなどのデータ圧縮技法を用いて圧縮表現にして文書ファイルに格納する．文献データそのものは１／３．６程度に圧縮されるが，文献単位に項目見出しを付けるので，全体としては１／３程度に圧縮される．また，ＡＩＲではキーワード転置ファイルに生起位置情報を持たない．このことによりキーワード転置ファイルの大きさはほぼ半分になる．

１９７３年から１９８４年現在までのＩＮＳＰＥＣ－Ｃの文献データからＡＩＲ，ＦＡＩＲＳ－Ｉによってデータベースを構築した経験では，ＡＩＲはＦＡＩＲＳ－Ｉに比べて文書ファイルで１／３．５，転置ファイルで１／３．８，全体として１／３．６のディスク領域で済んだ．ＡＩＲ，ＦＡＩＲＳ－Ｉの使用量は，それぞれ約２００，７００メガバイトであった．ＩＮＳＰＥＣ－Ａではこの３倍，ＩＮＳＰＥＣ－Ｂでは１．５倍のディスクが必要であるので，ＡＩＲによってもたらされるディスクの節約の効果は非常に大きい．実際，ＦＡＩＲＳ－Ｉを用いていた場合には，最大のＩＮＳＰＥＣ－Ａで３年分，ＩＮＳＰＥＣ－Ｃで５年分程度しかディスク上に置くことできなかったが，ＡＩＲを用いるようになって１０年分以上のＩＮＳＰＥＣデータをすべてディスク上に置いて検索サービスを行えるようになった．

5

6．2　システム応答速度

ＡＩＲでは，文献キー集合の表現を，キーワード転置ファイル，属性転置ファイル，演算結果を格納する作業用ファイル上ですべて同一の形式としている．従って，ＡＮＤ演算などを必要としない〝ＦＩＮＤ　ＳＹＳＴＥＭ〟のようなコマンドに対する処理は，Ｑ木の索引を参照するだけであるので，極めて高速である．

ＡＩＲでは，文献集合を１０キロバイト以上の大きな単位（ブロック）でディスクに格納することによって，集合演算処理中のファイルアクセス回数を減らしている．ＦＡＩＲＳ－Ｉではブロックの大きさは２０４８バイトに固定されている．ＡＩＲでは，集合演算方式それ自体にも多少工夫をこらし，またＦＡＩＲＳ－Ｉと比べると，文献集合の表現の大きさが１／３以下であり，ブロックの大きさが４倍以上であるので，ファイルアクセス回数は１／１０以下である．システム応答速度は，まだ正確な計測は行っていないが，確実に１０倍以上高速である．

7．おわりに

九州大学大型計算機センターでは，ＩＮＳＰＥＣ文献検索サービスのための情報検索システムを１９８４年１月に，従来のＦＡＩＲＳ－ＩからＡＩＲ第１版に置き換えた［１１］．情報圧縮により１０年分以上のＩＮＳＰＥＣデータベースの遡及検索が可能となり，検索の応答時間も格段に向上していることが確認できた．ＡＩＲには，近い将来日本語処理機能を追加し，現在ＦＡＩＲＳ－Ｉで行っているＪＩＣＳＴファイルの検索サービスをＡＩＲで行う予定である．

文章データを〝読み〟，〝理解する〟機能については，今後研究開発を進め，ＡＩＲを真に高度な情報検索システムにしたいと考えている．

参考文献

[1] 計算機マニュアル FACOM OS Ⅳ FAIRS－I解説書，富士通（株）．

[2] Aitchison, T. M., Martin M. D. and Smith, J. R.：Developments towards a Computer Based Information Service in Physics, Electrotechnology and Control, Inform. Storage and Retrieval, 4, 2, 177－186（1968）．

[3] Aho, A. V. and Corasick, M. J.：Efficient String Matching：An Aid to Bibliograpic Search, Commun. ACM, 18, 6, 333－340（1975）．

[4] 松尾，二村，吉田：準最適テキスト圧縮符号，九大工学集報，55，2，103－106（1982）．

[5] 松尾，二村，吉田：科学技術論文抄録における単語の統計的性質，九大工学集報，54，4，411－416（1981）．

[6] 二村，松尾：オンライン検索のための文献データ圧縮技法，情報処理学会データベース・システム研究会資料34－5（1983）．

[7] 松尾，二村，高木，吉田：INSPECデータベース転置ファイル生成における不要語選択法，九大工学集報，54，2，99－105（1981）．

[8] Salton, G. and McGill M. J.：Introduction to Modern Information Retrieval, xv＋448, McGraw－Hill, New York（1983）．

［９］篠原，二村，松尾：転置ファイル内のキーワード生起位置情報について，情報処理学会第２６回全国大会講演論文集（Ⅱ），７Ｆ－３（１９８３）．

［１０］松尾，二村，篠原：高速単語索引，同上，７Ｆ－４（１９８３）．

［１１］二村，篠原，松尾：情報検索システムＡＩＲによるＩＮＳＰＥＣの検索，九州大学大型計算機センター広報，１７，１，１－２２．（１９８４）．