情報検索におけるマンマシン・インタフェース

村井正和　斉藤栄一　鈴木通博　白石　旭
（電電公社　横須賀電気通信研究所）

1. まえがき

人間と計算機間のコミュニケーションは，リモート・ターミナルを含む入出力装置を介して行なわれることから，マンマシン・インタフェースに関する論議は，入出力装置のハードウェア上の考察と，人間と計算機の対話のための言語を中心としたソフトウェア上の考察とが必要である。　ここでは情報検索（Information Storage & Retrieval　以下 IR と略記する）におけるマンマシン・インタフェースについて特に言語面の考察を行なうこととする。

IR はシステム的にみれば情報の収集，整理，加工，蓄積，検索，提供の全工程を処理する技術である。　全工程のうち情報の収集から蓄積までにおいては，情報提供者と計算機間のインタフェースの問題であり，また検索，提供の工程においては情報利用者と計算機間のインタフェースの問題である。　利用者と提供者とは相異なることから，蓄積された情報の検索段階においては情報利用者と計算機間の対話は，会話形式でかつ試行錯誤に行なわれるところに特徴がある。情報提供者は，計算機システムに情報を加工，蓄積する段階で，検索に必要な索引づけを行なうが，この際システム構成上，機能上および提供者の作業上の都合から制約を受けざるを得ない。　一方，情報利用者が，情報提供者と同一概念の情報を必要とする場合であっても，概念の抽出，分析，用語の相異，検索質問構成上の制約もあって計算機システムに質問を与えても満足な結果が得られない。このため情報利用者は検索質問を変更，拡張して再試行するという過程をくりかえす。　これにより，IR では，人間と計算機とが会話形式でコミュニケートする必要があるとされる所以である。

本稿では，最初に IR におけるマンマシン・インタフェースについて概説し，つぎにマンマシン・インタフェースのための言語に必要な機能および言語形態についてふれる。　さらに具体例として武蔵野電気通信研究所で試作した CIRES（会話形文献検索システム）における言語面の解説を行ない，最後に IR における将来動向について，特にソフトウェアの面から述べることとする。

2. IR におけるマンマシン・インタフェース

IR の目的とするところは，蓄積された大量のデータの中から所望のものを漏れなくしかも無駄なく検索することである。

IR の蓄積・検索段階のマンマシン・インタフェースにはそれぞれ以下の項目が要求される。

① 検索段階：ⓐ高い検索効率*1

② 蓄積段階：ⓐ検索効率の高いデータ・ファイルの作成
　　ⓑ高い蓄積エリア効率
　　ⓒ高い運用性

また，いずれの段階にも操作性の良さが要求される。

2.1. 検索段階におけるマンマシン・インタフェース

検索は蓄積されたデータの中から所望の情報を取り出す過程である。したがって検索段階では，加工蓄積されたデータ・ファイルを持つ計算機と質問者との間のインタフェースがきわめて重要なものになってくる。検索段階で用いられる手法のほとんどは，検索効率を高めることを目的としたものであると云ってよい。したがってインタフェースの良さは，この検索効率がどうであるかによって評価できると考えられる。検索効率向上の手法としては主として

① 会話処理（即時処理）
② 検索質問の自動変更
③ 各種辞書出力

に分けて考えることができよう。

(1) 会話処理（即時処理）

近年，IR 分野にも TSS などの会話処理が導入され，質問者と計算機との会話によって検索対象の的を絞ってゆく検索方式が研究・実用化されるようになってきた。会話処理を応用することによって質問者が，検索結果の出力を見てその適合，不適合を判定しさらに検索の的を絞るための条件をシステム側に返すこと（フィードバック）が可能となった。この過程での計算機からの出力は一般に，最終的な出力を得る前に，質問条件を満足する文献が何件あるかという表示（事前検索統計という）の形で行なわれる。現在のオンラインの IR システムは何らかの形でフィードバックによりマンマシン・インタフェースの改善を計っているものが多い。

会話処理を用いた IR システムのはしりは，MIT の TIP，Lockheed 社の DIALOG などである。米国内では，医学情報を対象とした MEDLINE システムが現在，成功裡に商用に供されている。

(2) 検索質問の自動変更

利用者が最初に入力した検索質問（特に術語，キーワードに関する検索式）をその後，質問者が入力する情報を手がかりとしてシステムが，自動的に変更する手法が各種研究されている。例えば G. Salton の主宰する SMART システム[1]では検索結果に対する質問者の適合，不適合などの応答情報をもとに質問者の元の検索質問を自動的に修正変更して，再度検索を繰り返す適合性フィードバック（Relevance Feedback）と呼ぶ方法を用いている。SMART システムでの実験結果から明らかになったことのひとつとして修正した検索質問による検索の結果が，元の検索質問による検索結果に比べ精度が大幅に向上することが報告されている。

(3) 各種辞書出力

検索段階で用いられる辞書類のほとんどは，マンマシン・インタフェースを良くする目的から蓄積段階において，原情報を加工・蓄積する過程において作られたものを流用することが多い。一般に使われているものとしては，①キーワード・リスト，②シソーラス，③分類表，④キーワード出現頻度統計表などがある。

---

*1 検索効率としては，通常下記の量を用いている。

$$\text{適合率} = \frac{\text{検索文献中の適合文献数}}{\text{検索文献数}}$$

$$\text{再現率} = \frac{\text{検索文献中の適合文献数}}{\text{データ・ファイル中にある全適合文献数}}$$

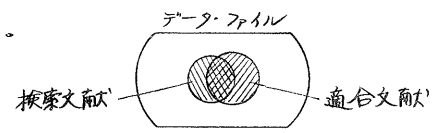


図1 検索効率

2.2 蓄積段階におけるマンマシン・インタフェース

大量のデータを高速に加工し，高いエリア効率で蓄積すること，および，この作業を少ないマンパワーで行なうなど運用効率を高めることは重要な要素である。また，検索時には，蓄積されたデータが直接検索の対象となるため蓄積時に検索時のマンマシン・インタフェースについてどれだけ考慮してあるか，が検索の質の良さにきわめて大きく効いてくる。このため種々の手法が開発されているが，いずれも手作業によるものが多く，マンマシン・インタフェースの観点からは，実験的な試みが行なわれているのが現状である。ここでは，各種手法の概説にとどめることとする。

(1) 内容分析

蓄積の対象となる原情報は，論文などの文章形式のものであり，量が膨大である上に，冗長度も大きいため，そのままを計算機に蓄積することは一部の実験研究的なものを除いてはあまり行なわれていない。このような文章情報を蓄積用に圧縮加工するのが内容分析の技術であり，次のようなものがある。

(i) 抄録作成

抄録は文章としての形をとどめた加工物である。1950年代後半からIBMのLuhnやEdmundsonなどによって計算機による抄録作成の自動化方法が研究されてきた[2]が，経済的に引き合わない上にいずれにしても統計的な手法であって，意味を無視していることや，原文からの単なる文の抽出にすぎないということもあって，現在のところ実用には供されていない。

(ii) 索引づけ

原情報を代表させるものとして，原情報中の特徴的な術語を抜き出し，その集合を作り出すのが索引づけである。計算機による自動化の例としてKWIC（Keyword in Context）などがあげられるが，抄録作成の自動化の場合と同様に，意味の機械的な扱いという点で難点がある[3]。たとえば蓄積のための加工では必ずしも原文中にはないが，内容を表わす適切な語を索引作成者が付加するなどの必要もあって，実用に供されているものは人手により作成されている。

(iii) 分類

蓄積されている情報が適切に分類されていれば検索速度が大幅に向上することが知られている。分類の自動化についても採算性が悪いことと意味内容の処理が困難なことなどから，実験の域を出ていない。

(2) 蓄積時の補助手段

加工・蓄積を行なう場合の補助手段としては，①術語（キーワード）リスト，②シソーラス，③分類表などがある。これらは，抄録や索引を作る場合にその作業者に，指針を示すものである。これらのもの（特にシソーラス）は，作成に要するコストがかなり大きいが，蓄積時と検索時における概念の統一が，計れるため，望ましいマンマシン・インタフェースが期待できるということから現用のシステムではさかんに活用されている。

3. マンマシン・インタフェースのための言語

言語は，何らかの処理を行ないたい人間と，処理を行なう計算機とのコミュニケーションを効率よく行なうための道具である。ここで云う効率とは処理速度，蓄積エリア効率，記述の多様性，記述の容易性などを意味する。これまで開発されたIRシステムが提供する言語は，次のように大別することができよう。

① 非手順形言語 { ⓐ 問い合せ言語 / ⓑ コマンド言語 }
② 手順形言語 { ⓐ IR 向プログラミング言語 / ⓑ 汎用言語 }

ここではまず，これらの言語に共通して必要な機能について述べ，次いで各言語形態での必要機能を述べる。

3.1 共通機能

(1) 検索質問

検索質問に対する条件としては，以下があげられる。

(i) キーワードの一致条件
前方一致，後方一致，任意一致，完全一致などを表現する。

(ii) 論理的条件
キーワードと論理演算子の組み合せにより論理条件を表現する。

(iii) 定量的条件
キーワードに重みづけを行ない，これらを持つ情報に重みづけを行なう。

(iv) キーワードの相対位置関係
キーワード間の前後関係，キーワード間の距離など，相対位置関係を表現する。

(v) 範囲条件
雑誌の出版年度範囲の指定など，必要とする情報の検索範囲を限定する。

主として数値アイテムに対する条件である。

これらの条件を組み合せて，利用者の所望の条件を表現する。 手順形言語においては，これらの条件が容易に表現できることが必要である。

(2) 検索手順

IR の検索手順には，各種の表し方があるが，マンマシン・インタフェース上組み入れておくべき機能として以下があげられる。 なお，これらの機能は会話処理を前提としている。

(i) 事前検索統計出力機能
検索質問（通常はキーワードによる検索式）に合致した情報の件数を出力する機能。

(ii) フィードバック機能
計算機からの応答内容によって，前段の検索過程にもどる機能

(iii) 検索結果出力件数の指定機能

(iv) キーワード・リスト，シソーラスなどの出力機能

(v) 検索結果の編集出力機能

3.2 非手順形言語

(1) 問い合せ言語

システムが，あらかじめ決められた検索手順にしたがって，その都度，利用者に問い合せを行ない，利用者がその問に一つ一つ答えてゆきながら目的の検索業務を実行する言語形態である。 初心者には便利であるが，使い慣れた利用者にはわずらわしい欠点がある。 問い合せ言語の特徴を以下にあげておく。

(i) システムから質問者への要求内容

① 検索手順の選択情報 --- 検索手順の方向を，その程度表示し，利用者に選択させる。

② 検索質問 --- 利用者の検索要求情報に対する条件内容

(ii) 問い合せ言語の機能

前述3.1の共通機能の他に，問い合せ言語特有の機能として，途中打ち切り機能が必要である。 固定化した手続きから抜け出すために，フィードバック機能を充実させる必要がある。

(2) コマンド言語

ある程度まとまった機能単位をコマンド形式で提供するものであり，

命令文　パラメータ1，パラメータ2，---- パラメータn

という形式を取る。 システムからのコマンド促進記号に続いて，利用者は，命令文とパラメータを入力する。

コマンド言語に必要な機能として前述3.1の他に，以下があげられる。

(i) コマンド登録機能

あらかじめ，検索手順をファイルに登録しておき，任意の時点で呼び出してこの手順を利用する機能である。

(ii) 指定内容の簡単化機能

命令文の短縮形の表現，パラメータの省略形などを可能とすることが望ましい。

(iii) 入力誤り修正機能

例えば，入力文字の一部の修正，削除，挿入も存ることが望ましい。 また1字程度の誤りならば自動修正することが望ましい。

(iv) ガイダンス機能

システムの利用方法，ファイルの蓄積状況などの問い合せに応答する機能である。

3.3 手順形言語

手順形言語を利用するのは，主に利用者システムを設計するためであり，プログラマが，その利用者である。 手順形言語を用いて，利用者個人のための検索を行なうこともありうるが，通常は，複数回使用する，あるいは，他の複数利用者に使用させることを目的とすることが多く，結局は，利用者システム設計の範囲に属することが多い。

IRを対象と考えると手順形言語は

① 汎用言語

② IR向プログラミング言語

に分けることができる。

汎用言語としては COBOL，FORTRAN，PL/I などが，あるが，ここでは，IR業務を対象とした問題向き言語の必要機能について述べ

表1 IR向プログラミング言語の機能

機能	内容
ファイル処理	ファイル定義，データの蓄積，検索，更新，ソート/マージ，ガベージ・コレクション
ストリング処理	ストリングの切り出し，結合，挿入，圧縮，置換，移動
演算処理	算術演算，論理演算，定量演算
比較判断制御	論理判断，文字列の比較，スキャン，ルートコントロール(GOTO)，繰り返し(DO)
テーブル処理	コア内ソート，テーブル・サーチ，最大値/最小値の検出
通信処理	端末入出力，オペレータ・コンソール入出力
その他	リスト処理，プログラム停止，デバグ・ステートメント，注釈，サブルーチン定義

ることとする。　IR処理を記述する主な言語機能を表1にあげる。

3.4　言語の機能比較

表2は，IRを対象とした言語の機能比較表である。

表2.　IRを対象とした言語の機能比較表

	非手順形言語		手順形言語
	問い合せ言語	コマンド言語	IR向プログラミング言語
利用者	一般利用者	一般利用者	一般利用者あるいはプログラマ
利用の方法	システムからの問いかけに対して応答することによって処理が進行する。	システムからの促進記号出力のあと，所望の処理を示すコマンドを入力することによって処理が進行する。	処理したい手順をプログラミングし，ソース形式あるいは実行形式でファイルに登録しておき，自分あるいは，他人が利用する。
利用手続きの可変性	固定的	やや可変的	可変的
利用の容易性	容易。しかし慣れるとゆずらわしい。	やや容易	やや難
利用の目的	主に検索	検索および蓄積	IR業務を行なうシステムの作成
適用業務の広さ	狭い	やや狭い	やや広い。
要求される機能	高い検索効率および高速検索。	高い検索効率および高速検索。	左記機能を満足するようなIRシステム記述の容易性，および処理効率の良さ。
言語処理方式	インタプリタ	インタプリタ	インタプリタあるいはコンパイラ
処理形態	即時処理	主に即時処理	主に一括処理
システム例	CIRES（通研）	TIP(MIT), DIALOG(ロッキード), JOLDOR（情開センタ），DOOR（JICST）	GIS（IBM） （注．事務処理向である。）

4.　具体例 ── CIRES ──[4),(5)]

問い合せ形式のIRシステムの例として会話形文献検索システムCIRES（Conversational Information Retrieval System on Documents）をあげる。CIRESでは，検索効率を高めるためのマンマシン・インタフェースとして以下の特徴を持っている。　なお，図2にCIRESの検索処理の流れを示す。

① 検索式（キーワードに関する条件）としては，論理検索式と定量検索式が使用可能である。
② 検索質問の入力モードとして，項目別質問と一括質問がある。
③ 検索結果の出力形式は，標題のみおよび全項目出力が可能である。
④ 検索式の自動変更
⑤ 関連語（CIRESシソーラス）の出力

4.1　CIRESの検索機能

(i)　検索質問の作成方法

部門（12部門），分野（のべ84分野），年度，雑誌名，著者名および検索式が指定可能である。　なお検索式は，次の2形式が指定可能である。

(i)　論理検索式

例：L = INFORMATION * RETRIEVAL * (TSS + REAL * TIME)

(ii)　定量検索式

例：W = COMPUTER #5, DESIGN #15,
HARDWARE #11

(2) 検索質問の入力モード

検索質問に限らずCIRESでは利用者の入力をできるだけ少なくかつ簡単にした。また，システム・メッセージは，可能なかぎり平易なカナ文を用いた。

(i) 項目別質問

入力質問を1項目ずつCIRESの指示にしたがって入力する方法で初心者向きである。

(ii) 一括質問

入力質問を一度にまとめて指定する方法であり慣れた利用者向きである。

例： P = 01; F = 03, 10; Y = 71, 72;
J = CACM; W = DATABASE #10,
IR #10, FILE #5, REAL-TIME #5

(3) 検索結果の表示形式

(i) 標題のみの出力

(ii) 全項目の出力

検索された文献について，標題の他，著者名，雑誌名，国名，発行年度などを出力する。

(4) 検索式の変更

(i) 利用者による変更

(ii) CIRESによる変更

利用者がもとの検索式に使用した各キーワードに対し，補足追加したいキーワードを，すでに出力されている関連語キーワードを見て略号で指定する。CIRESは追加されたキーワードにより次の2つの形式の検索式を表示し，利用者に適当と思うものを選択させる。

例：追加キーワードがA1, A5の時
形式1： AをA+(A1+A5) で置換
形式2： AをA*(A1+A5) で置換

(5) 関連語の出力

利用者は，関連語（CIRESシソーラス）を，オンライン即時で見ることができる。

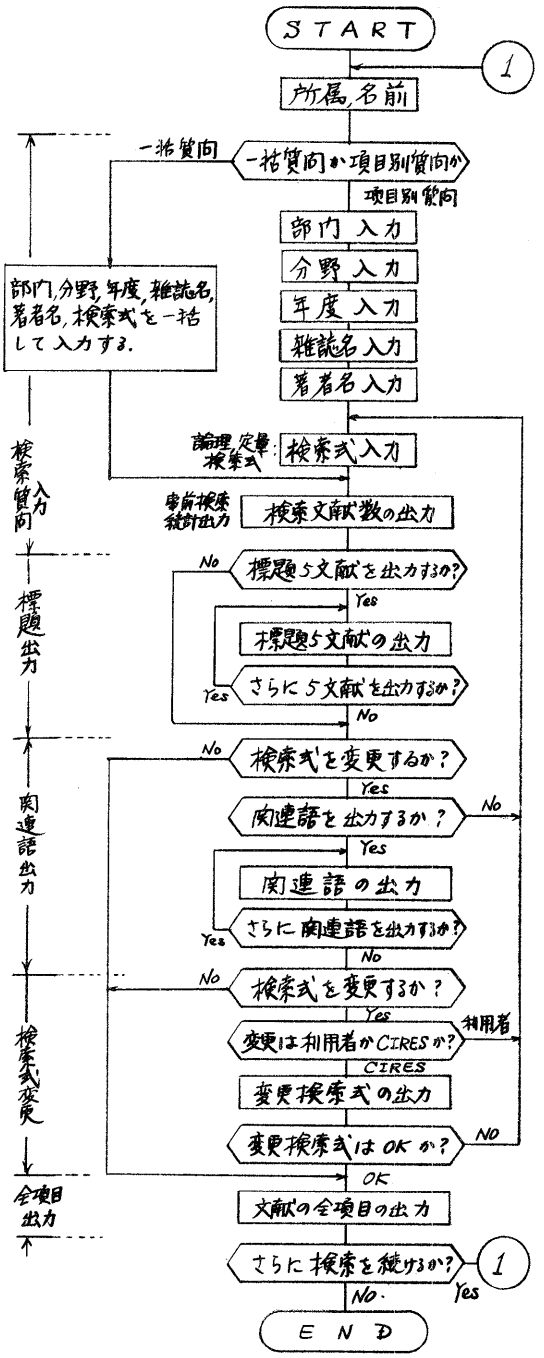


図2 CIRES検索処理概略フロー

関連語とはキーワードの中でお互いに意味的に深い関係を持つものを云い，各キーワード毎に10個ずつCIRESのファイルに蓄積されている。

例： A(INFORMATION); A1(RETRIEVAL), A2(STORAGE), A3(DATA), ----

ここで，AはINFORMATIONというキーワードにCIRESが検索時に与えた略号であり，A1, A2, ---はAの関連語の略号である。以後この略号が使用できる。

```
/ DIPS-0 FINAL VERSION 0201
¥ メ オ フミ,カクニンブシ テ゛ ハジメ
/00057 カ# ハ?
¥ / メ オ カクニンブシ,フミ,ノーマルモード゛ テ゛ 70.11.10 ニチ 09.59.14 ニ ハジメタ
¥ ブンケンケンサク オ ジッコウ

*****************************************************************************************************************************
*** DIPS-0 ブンケンケンサク(CIRES) ハジマリ ***
/ アナタ ノ ショゾク ト ナマエ ハ?
¥ ECL ナカムラ
/ ケンサクシソモン オ 1(コウモクヘ゛ツ シテイ) カ 2(マトメテ シテイ) カ?
¥ 1
/ ブ゛モン(PART) ハ                          01(INFORMATION PROCESSING) カ          02(COMMUNICATION SYSTEMS) カ
  03(APPLIED MECHANICS) カ                   04(SOLID STATE TECHNOLOGY) カ          05(PARTS & MATERIALS) カ
  06(SEMICONDUCTORS & DIELECTRICS) カ        07(PARTS & MATERIALS FOR COMPUTER) カ  08(POLYMER SCIENCE) カ
  09(PHYSICAL PROPERTY) カ                   10(ELECTRONIC ENERGY CONVERSION) カ    11(MEASUREMENT & ANALYSIS) カ
  12(COMPUTER & AUTOMATON) カ?
¥ 01
/ ブ゛ンヤ(FIELD) ハ              00(GENERAL) カ          01(THEORY) カ              02(COMMUN THEORY) カ
  03(AUTOMATON) カ                10(COMPUTER) カ         15(DATA TRANSMISSION) カ   16(AUTOMATIC CONTROL) カ
  21(EXCHANGE SYSTEM) カ          23(CARRIER SYSTEM) カ   31(ELECTRONIC CIRCUIT) カ  32(ACOUSTICS) カ
  45(MICROELECTRONICS) カ?
¥
/ ネンド゛(YEAR) ハ  69(1969) カ  70(1970) カ?
¥ 69,70
/ サ゛ッシ(JOURNAL) ハ?
¥
/ チョシャ(AUTHOR) ハ?
¥
/ ケンサクシキ(RETRIEVAL EXPRESSION) ハ?
¥ L=COMPUTER*INFORMATION
/ ケンサク ブンケンスウ カ゛  44ケン. ブ゛ンケン 5ケン ノ ヒョウタ゛イ オ ウチタ゛シ 1(スル) カ 2(シナイ) カ?
¥ 2
/ ケンサクシキ オ ヘンコウ 1(スル) カ 2(シナイ) カ?
¥ 1
/ カンレンコ゛(ASSOCIATIVE WORD) ノ ウチタ゛シ オ 1(スル) カ 2(シナイ) カ?
¥ 1
/ カンレンコ゛ ウチタ゛シ オ シテイスル KEYWORD ハ A(COMPUTER) カ B(INFORMATION) カ?
¥ B

/ B(INFORMATION);B1(RETRIEVAL),B2(STORAGE),B3(DATA),B4(PROCESS),B5(CODE),B6(THEORY),B7(SYSTEM),B8(MANAGEMENT),
    B9(CHANNEL),B0(BIT)
/ サラニ カンレンコ゛ ノ ウチタ゛シ オ 1(スル) カ 2(シナイ) カ?
¥ 2
/ ケンサクシキ オ ヘンコウ 1(スル) カ 2(シナイ) カ?
¥ 1
/ ケンサクシキ オ 1(アナタ カ゛ ヘンコウスル) カ 2(CIRES カ゛ ヘンコウスル) カ?
¥ 2
/ ツイカスル KEYWORD ノ リアクコ゛ウ ハ ?
¥ B1
/ ヘンコウマエ ノ ケンサクシキ ハ  L=A*B      アタラシイ ケンサクシキ ハ
    (1)   L=COMPUTER*(INFORMATION+RETRIEVAL)
    (2)   L=COMPUTER*INFORMATION*(RETRIEVAL)
  ケンサクシキ ハ 1カ 2カ  ソレトモ 3(アナタ カ゛ スル)カ?
¥ 2
/ ケンサク ブ゛ンケンスウ カ゛  10ケン. ブ゛ンケン 5ケン ノ ヒョウタ゛イ オ ウチタ゛シ 1(スル) カ 2(シナイ) カ?
¥ 1

*** C I R E S .     ロンリケンサク       . . . T I T L E S  ON  R E T R I E V E D  P A P E R S . . .        11/10/70 ***

** ケンサクシキ **   L=COMPUTER*INFORMATION*(RETRIEVAL)

001 USING THE TIP SYSTEM IN THE ASIS FILE MANAGEMENT EXERCISE (( DISPLAY, QUERY, TECHNICAL INFORMATION PROGRAM, REMOTE-A
    CCESS CTSS COMPUTER, COMPUTER COST FIGURES, RETRIEVAL, ON-LINE ))

002 A MULTIVARIATE STATISTICAL ANALYSIS OF THE USE OF A SCIENTIFIC COMPUTER BASED CURRENT-AWARENESS INFORMATION RETRIEVA
    L SYSTEM

003 THE POTENTIAL USEFULNESS OF CATALOG ACCESS POINTS OTHER THAN AUTHOR, TITLE, AND SUBJECT (( LIBRARY, NONSTANDARD INFO
    RMATION, COMPUTERIZED CATALOGS, MEMORY EXPERIMENT, RETRIEVAL ))

004 A COMPUTER-CONTROLLED MICROFILM SYSTEM (( INFORMATION STORAGE AND RETRIEVAL, SHARED-TIME, GENERAL-PURPOSE COMPUTER )
    )

005 MICROFILM - A NEW DIMENSION FOR COMPUTERS (( MEMORY, INFORMATION STORAGE AND RETRIEVAL SYSTEMS, COM, COMPUTER OUTPUT
    MICROFILM ))
/ サラニ ブ゛ンケン ノ ヒョウタ゛イ オ ウチタ゛シ 1(スル) カ 2(シナイ) カ ?
¥ ¥T
/ ケンサク ブ゛ンケンスウ カ゛  10ケン. ウチタ゛シ ブ゛ンケンスウ ハ?
¥ 2

*** C I R E S .     ロンリケンサク  **  ケンサクシャ :                    ECL ナカムラ ト゛ノ ＊ケンサクケッカ ＊  11/10/70 ***

** シテイコウモク **  ショウリョクシテイスウ :   2  ケンサクコ゛ウケイブ゛ンスウ :  10  ネンド゛ : 1969 1970
ブ゛モン : INFORMATION PROCESSING

** ケンサクシキ **   L=COMPUTER*INFORMATION*(RETRIEVAL)

MATHEWS,W.D.      MIT               USING THE TIP SYSTEM IN THE ASIS FILE MANAGEMENT E   J.AM.SOC.INF.SCI.      @   001
                                    XERCISE (( DISPLAY, QUERY, TECHNICAL INFORMATION P   21,3     204/08     A
                                    ROGRAM, REMOTE-ACCESS CTSS COMPUTER, COMPUTER COST   70.05/06             7008001354

ANICK,D.J.        UNIV.PITTSBURGH   A MULTIVARIATE STATISTICAL ANALYSIS OF THE USE OF    J.AM.SOC.INF.SCI.      @   002
                                    A SCIENTIFIC COMPUTER BASED CURRENT-AWARENESS INFO   21,3     171/78     A
                                    RMATION RETRIEVAL SYSTEM                            70.05/06             7008001349
/ ケンサク ブ゛ンケン ウチタ゛シ オワリ. アナタ ハ サラニ ケンサク オ 1(ツヅ゛ケル) カ 2(ヤメル) カ?
¥ 2

*** ブ゛ンケンケンサク(CIRES) オワリ ***
*****************************************************************************************************************************
/ プ゛ログ゛ラム カ゛ シュウリョウシタ
¥ オワリ
/ 70.11.10 ニチ 10.23.15 ニ オワッタ.ショウシ゛カン カ゛ 00.01.12
```

図3　CIRES の検索実例

図3は，CIRES の検索実例である。

4.2 CIRES の運用機能

(1) CIRES 形インバーテッド・リスト・ファイル作成機能
(2) 関連語表の作成機能 ---- 4.3 参照のこと.

4.3 キーワードと関連語

マンマシン・インタフェースを向上させるために, CIRES で検討した内容分析は, キーワードの選定と関連語の分析である。

(1) キーワードの選定

CIRES ファイルに登録するキーワードの選定を, 以下の手順で行なった.

(i) 部門編成

部門の編成は, 当研究所情情部門で行なっている REWDAC 文献検索依頼データから 3~4 年間の分について質問主題の分析と指定された分類項目を基礎に同部門で行なったものである。 各部門に属する文献数は, ほぼ等しくなるようにしており, 平均値は約 7,000 文献である。

(ii) 術語リストの作成

部門ごとに文献に含まれる全ての術語を, 抽出しソートする。 次に各術語の頻度を集計して術語リストを作成する。

(iii) キーワードの抽出

(ii) で作成した術語リスト中から CIRES ファイルに登録するキーワードを次の規則にしたがって抽出した。

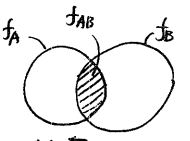
① 不要語 (OF, THE, ABOUT など), 1文字語, 数値語などを除いた。
② 抽出語数を部門毎に 1,500 を上限とした。
③ 語幹を同じくする術語は, 1つのキーワードで表わした。
④ 同義語, 省略語があるときには, これも記録する。

(2) 関連語表

一般に検索効率の向上にはたすシソーラスの役割は大きいが, これを作成する作業量は非常に大きい。 CIRES では, Maron および Stiles[6] などによって検索効率向上のために利用されていた関連語をシソーラスとして用いることにした。 CIRES ではインバーテッド・リスト・ファイルのキーワードになっている全ての術語について, 関連語を統計処理で抽出し, CIRES ファイルに蓄積した。 あるキーワードの関連語とは, 関連係数の大きなキーワードである。 関連係数としては種々の定義があるが, CIRES では, 比較的計算の容易な次式を用いた。

$$\text{キーワード } A, B \text{ の関連係数} = \frac{f_{AB}}{\sqrt{f_A \cdot f_B}}$$

ここで
f_A, f_B : 対象文献データ・ファイルにおけるキーワード A, B の出現頻度
f_{AB} : A, B が同一文献中に出現する頻度



4.4 CIRES の評価

マンマシン・インタフェースの観点から, CIRES の評価項目は以下のとおり.

① 検索効率 (検索式自動変更に対する評価も含む)
② 会話性
③ 選定キーワードの妥当性
④ CIRES シソーラスの妥当性

(1) 検索効率

一般の論理検索式の場合，約70～80％の適合率を得ることができた。（表3参照）

また，利用者の指定したもとの検索式に，検索式自動変更機能によって関連語を補足した場合の検索効率の変化は以下のとおりである。すなわち，4.1(4)(ii)で述べた形式1の新検索式では適合率が数％の増加，再現率も1.5～2.5倍の増加，一方，形式2の新検索式では，適合率が10％程度の増加，再現率は数分の1に減少した。

表3 論理検索の適合率

検索式 ＼ 適合率	検索文献中の有効文献数/検索文献数
例 1	74% (17/23)
〃 2	82% (56/68)
〃 3	86% (13/45)
〃 4	70% (14/20)

表4 自動変更検索式の検索効率

検索効率 ＼ 検索式	元の検索式の適合率	形式1による変更検索式		形式2による変更検索式	
		適合率	再現率の増加率	適合率	再現率の増加率
例 1	86% (13/15)	89% (17/19)	1.31 (17/13)	100% (3/3)	0.23 (3/13)
2	67% (16/24)	69% (34/49)	2.13 (34/16)	78% (7/9)	0.44 (7/16)
3	67% (16/24)	71% (37/52)	2.31 (37/16)	100% (2/2)	0.13 (2/16)

(2) 会話性

検索式の改善に会話処理が非常に有効であることがわかった。

(3) 選定キーワードの妥当性

CIRESでは部内毎に約7,000件の文献中から主として出現頻度によって1,300語の術語をキーワードとして登録したが，入力された検索式中に含まれる非登録キーワードは10％程度であった。

(4) CIRES シソーラスの妥当性

CIRES シソーラスで関連係数の高い術語程，一般のシソーラスに含まれる率が多く，シソーラスとして意味あることが分かった。図4は，ある範囲の関連係数を持つ関連語毎にEJC，あるいはNASA シソーラスに含まれる割合を示したものである。

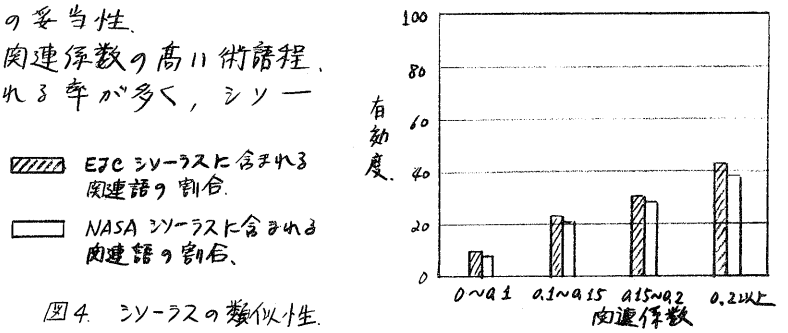


図4 シソーラスの類似性.

5. 将来動向.

5.1 内容分析.

(1) 内容分析の自動化

第2章で述べたように，情報蓄積に際してのデータの内容分析に要するマンパワーは非常に大きい。自動化の研究もいくつか成されているが，いずれも実験の段階であり，実用に耐えるものが出ていない。

しかし，将来不動産情報など膨大な情報を扱う国家的な規模のIR システムの開発がいくつか要求される可能性もあり，内容分析の自動化は，いずれは必要となってくると思われる。将来は，計量言語学，自動翻訳の分野に用いられている構文分析，形態分析，意味構造変換などの手法をIR の内容分析にも応用することになろう。このためには早いうちに，両分野の研究者の密接な情報交換，意見交換が成されなければならない。

(2) 自然言語処理

EURATOM の ISPRA 研究所では，機械翻訳（露語→英語）とIR を統合して，一つのシステムを作っている。検索条件を入力すると，これに対応する翻訳論文が出力される。充分な需要があれば，日本においてもこのような自動翻訳とIR

とが結びついたシステムが開発されよう。　また，検索条件を自然言語で入力し，解析して条件式を作成し検索を開始するようなことも可能になってこよう。しかし，自然言語は冗長性およびあいまいさが大きいので，用途は限られよう。

5.2　日本語処理

日本語の文献をそのままの漢字かなまじり文で扱う場合の問題点は，元来使用される漢字の字種が多いことから，入出力装置が高価かつ操作性が良くないこと，さらに装置ごとに漢字コードが異なっていてコードの標準化が成されていないなどの点にある。　しかし，将来パターン認識をベースにした光学的文字読取り装置などが利用できるようになれば，日本語データ処理技術が急速に要求されよう。

6．むすび

情報検索システムにおけるマンマシン・インタフェースの良さは，究極的にはいかに必要なものを漏れなく無駄なく，しかも能率よく検索することができるかにかかっている。　TSS システムを用いた検索者自身とシステムとの会話では，システムへの直接のフィードバックが繰り返し可能であるために，検索の漏れと無駄を小さくするための一手段としてきわめて有力である。　会話の方法としては基本的にはシステムから問いかけてくる形とコマンドなどによりシステムへ指示を与えてゆく形とがあり，現在各種の方式が研究されているが，真に実用に供しうるシステムとするためには，今後は，実験室を出て実際の運用システムによる稼動環境のもとで研究をさらに進めてゆくことが必要であろう。

今後解決しなければならない問題としては，内容分析の自動化，自然言語処理及び日本語処理などがあるが，これらの問題に共通して言えることはいずれも〝意味〟の処理技術が，確立されれば，大きな発展が期待されるであろう。

本稿を終えるにあたり，日頃御指導を頂いている柴山室長ならびに高橋調査役に深く感謝致します。

参考文献

(1) Salton, G. (ed.): The SMART Retrieval System, Prentice Hall, 1971
(2) Schultz, C.K. (ed.): H.P. Luhn: Pioneer of Information Sciences, Spartan Books, 1968
(3) 笹森勝之助「情報検索の現状と動向」，情報処理，Vol. 11, No. 12, 1970.12
(4) 藤田・村井・小島「会話形文献検索システム CIRES プロセッサの概要」，昭和45年度情報処理学会大会
(5) 村井他「DIPS-0 会話形文献検索プログラム」，研実報 Vol. 20, No. 1, 1971
(6) H.E. Stiles: The Association Factor in Information Retrieval, ACM, Vol. 2, No. 8, pp. 271～279, 1961