

文献検索とシソーラス学習のための相互作用  
データベースシステムの開発

A Development of an Interactive Data Base System  
for Document Retrieval and Thesaurus Study

齊 藤 孝  
*Takashi Saito*

*Résumé*

This paper reviews the software techniques of on-line retrieval applications designed to improve the searching strategy and the relevance problems in question and answering procedures by means of a direct interaction between man and computer.

In a practically developed system called IDEAS, Interactive Data Base Easy Accessing System, the conversational language specified for the interface between data bases and end-users at the remote terminals and its sample dialogue methods for document retrieval and thesaurus study are introduced.

はじめに

- I. オンラインIRシステム
- II. オンラインデータベース技術
- III. 言語と対話
- IV. インプリメンテーション
- V. 文献検索とシソーラス学習のための対話

おわりに

はじめに

コンピューターによる IR システムの諸技術の発展も、1950年代の偉大な先駆者である H. P. Luhn が行った KWIC 索引や SDI サービスの開発以来すでに20年近くなる。これまでの歴史を見ると、自動化の試みから自動索引、自動抄録、そして自動分類の研究とい

った、数学と論理から言語学に至る幅広い学術的成果がある。その反面、実用化に基づくコンピュータベースの IR システムそのものの技術革新は不思議なくらい少ない。この原因の一つに、ドクメンタリストなどの実務経験者からの積極的なアプローチや、コンピュータ技術を活用する場が少なかったことがある。また、コンピュータに関しても初期におけるハードウェアとソフトウェア

齊藤 孝：東京芝浦電気株式会社 電算機システム技術部  
Takashi Saito, Computer System Engineering Dept., Tokyo Shibaura Electric Co.

ア上の制約から、そのアプリケーションも貧弱にならざるを得なかった。たとえば、事務計算とか科学計算のための COBOL や FORTRAN などの言語があっても、IR システムに有利なデータベースシステムやオンライン処理を容易にする道具はなかった。

IR システムの理想は、人間-機械系を前提とする知的な情報処理である。この種の処理はバッチ処理が限度であった時代ではコンピュータにとって荷が重すぎた。その時代における自動化の試みは、人間との相互作用 (interaction) の手段を準備しなかったため機械翻訳システムがほとんど実用価値がないと評価された事情に似ている。

適合性 (relevance) や探索戦略 (searching strategy) といった、文献の検索に付きまとう曖昧な概念も、キーワードの品質と主題分析の良否といった人間による前加工を除けば、相互作用の可能性によって IR システムにとって直接の関心事ではなくなる。

現在、コンピュータの進歩は著しく、時分割とオンライン処理、さらにデータベースシステムなどの相互作用の環境を造りだす強力なソフトウェアの道具がある。

本稿では、IR システムの技術的革新はこの相互作用システムに求められるとして、著者のこれまでのアプリケーションの設計と開発に基づく範囲で、文献検索とソーラスの利用を対話によって処理するシステムの技術を紹介する。

## I. オンライン IR システム

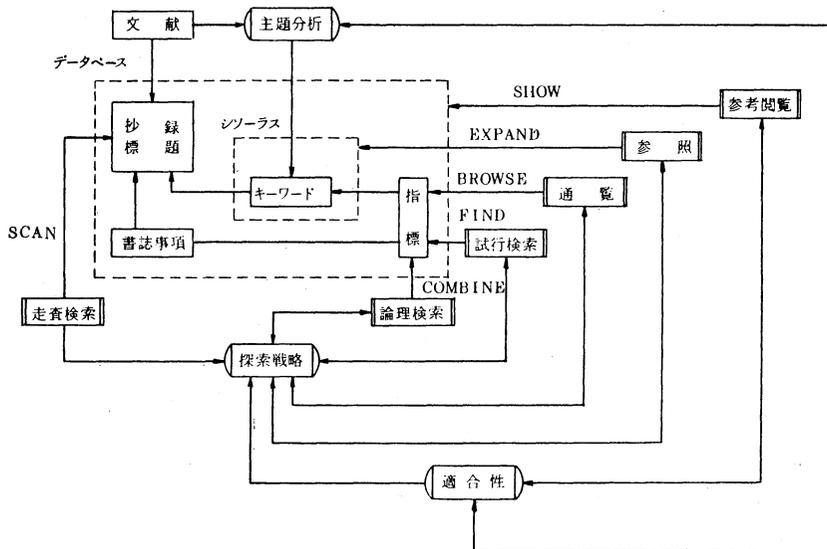
### A. データベース

最近ではデータベースといったコトバが流行しており、その意味が曖昧になっている。文献検索でも文献ファイルをデータベースと呼ぶ傾向がある。このコトバの使用の場には、CODASYL の DBTG<sup>1)</sup> が提案する統合化された重複のないデータの蓄積と、その多目的な利用を原点とする定義や、人工知能のグループ<sup>2)</sup>の証明アルゴリズムを知識のデータベースと称し、それをリスト展開するもの、それに似たりレシヨナルデータベース<sup>3)</sup>の思想などがある。

これらに共通することとして、大量の種々の性質を持つデータを、情報の主題、データ値の属性とその構造付けに注目し、蓄積したり検索したりする工夫に新しいデータベースの技術が形成されている。

### B. データベースと IR

いわゆる広義のデータベースシステムと、DR (文献検索) システムを含めた IR システムの発生の動機は多少相違するにもかかわらず、共通した目標がある。それは知的学習機能によるヒューリスティックな発想を得意とする人間と、アルゴリズム的なコンピュータ処理の両者が、それぞれの得意とする機能を明確にしておき、オンラインにおいて相互作用からよりの確かな情報処理を期待しようとするもので、人間-機械系の典型とすること



第1図 文献検索の相互作用

である。このインターフェースには、ハードウェアとしてより人間工学的に考案された端末装置と、ソフトウェアとしては、指令を円滑にする言語とが必要である。

C. 相互作用

IR システムの中で DR システムの相互作用<sup>4)</sup>は、第1図に簡単に示せる。このメカニズムを構成する要素として、人間の担当機能とコンピュータの担当機能とが明確に分離されて準備される。前者は、主題分析とか探索戦略や適合性といった人間の知能に依存する機能であり、後者は、データへの高速度なアクセスを中心とする通覧と試行、論理、走査などの検索機能である。第1図に示すそれぞれの機能の関連動作が相互作用となる。この相互作用を必要とする情報処理は、人間が大幅に介入しなければならないので省力化や自動化といったコンピュータ導入の意図からすれば、一歩後退した姿でもある。

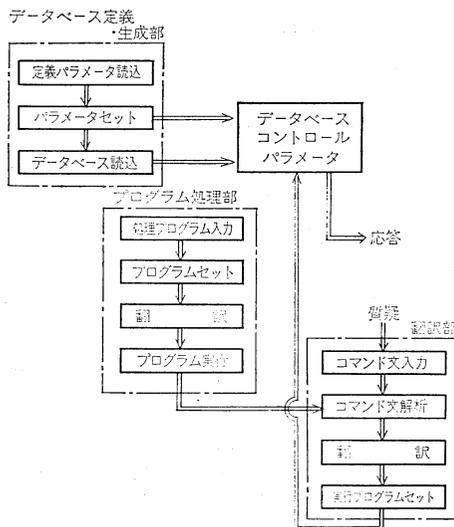
II. オンラインデータベース技術

本章ではデータベースとの相互作用を効果的にするオンラインリアルタイム処理によるデータベースのシステム技術を、IR システムといったアプリケーションの開発や運営に役立つ範囲に限って、ふれることにする。

A. システムの構成

このシステムの要素を概略的に示すと、第2図の基本機能がある。<sup>5)</sup>

(1) データベースの定義とファイルの生成処理部:



第2図 オンラインデータベースの機能

データ記述言語によりレコードのフォーマットやデータの属性と構造を定義し、パラメータカタログを準備する。このカタログに従ってデータが蓄積される。

- (2) 言語入力と翻訳部: 相互作用を指令するために準備されたシステム言語の入力と翻訳をし、実行指示をする。
- (3) 実行プログラム部: 実行指示に従ってカタログを参照しながらプログラムを実行していく。
- (4) データベースとディレクトリ部: パラメータカタログと、データが蓄積されている。

B. システムの条件

オンラインデータベースと呼ぶからには、一般に次の条件を満たしたものでなければならない。

- (1) 応答時間が速いこと。端末装置に人間が我慢できる限度は、30秒以内とされる。
- (2) 同時に複数の利用者にサービスができ、しかもそれぞれが独立に処理できること。同一のデータベースに時分割によって多数の質疑を受けることが可能でなければならない。
- (3) 処理が中断しても、そこまでの状態が記録されていて、継続処理が可能であること。
- (4) 端末装置上の操作ミスが実行に直接的に影響しないこと。
- (5) 人間工学的に設計された装置と言語があること。
- (6) データベースの機密保護ができること。共有ファイルの共同利用が前提であるので個人レコードの無断利用を防止するソフトウェア的保護が必要である。

オンラインシステムは、ここに挙げたシステムの条件を考慮しなければならない。

III. 言語と対話

オンライン化の最大の特徴は、これまでの専門家のみによるコンピュータへの接触を、遠隔地の現場に端末装置を置くことにより実務者の直接使用を可能にすることである。このためには、キーボードとディスプレイをメディアとした記号の送受信の形をとる対話と、そのための言語がいる。

A. コマンド言語

端末利用者 (end-user) のための言語がこれまでの言語と違う点は、コマンドと呼ぶオペレータとオペランド

第1表 コマンド言語の文法の一例

- 1) <定義コマンド> ::= DESCRIBE <項目番号> <項目名> <データタイプ>
- 2) <データタイプ> ::= A | N | K
- 3) <数コマンド> ::= TALLY | TALLY <条件>
- 4) <条件> ::= <述項> | <述項> <論理記号> <条件>
- 5) <述項> ::= (<項目名 | 項目番号> <関係記号> <値>)
- 6) <論理記号> ::= AND | OR | NOT
- 7) <関係記号> ::= GT | EQ | GE | LT | LE | INC | NEQ
- 8) <印刷コマンド> ::= PRINT <項目名 | 項目番号> | PRINT <項目名 | 項目番号> <条件>
- 9) <集計コマンド> ::= SUMMARY <式>, . . . , <式> | SUMMARY <式> <条件>
- 10) <式> ::= <項目名 | 項目番号> | <項目名 | 項目番号> <算術演算子> (<式> | <項>)
- 11) <算術演算子> ::= | - | \* | /
- 12) <項> ::= <項目名 | 項目番号> | <常数>
- 13) <計算コマンド> ::= CALC <項目名> EQ <値> <組込関数>, . . . , <組込関数>

から構文する簡単な文法があり、特に次の事項が考慮されている。

- (1) 言語は非プログラマ向きのものである。
- (2) 覚えやすく、適当な省略形ができる。
- (3) 誤りの解釈は即時応答され、再度試みられる。
- (4) 入力形式はフリーフォーマットとする。

**B. コマンドの機能**

オンラインデータベースのコマンドを分析すると次に示す系統が抽出される。

- (1) 開始終了系： 目的のデータベースファイルを開き、対話の申込をしたり終了の手続きをする。
- (2) 準備系： 本格的な検索に移る前にキーワードやカタログの内容を閲覧する。
- (3) 回答と編集系： 検索されたデータをソートしたり表を作成する。
- (4) 計算系： 数値データを対象に算術演算をする。
- (5) 検索系： データ項目と値とその論理条件を指定して、データベースに質疑する。
- (6) 補助系： 対話をより円滑にするサービス機能を果す。
- (7) マクロ定義系： 手順化した処理手続きをサブ

ルーチンとして定義し、マクロコマンドとして使用する。

コマンド言語の文法は単純であるが、その記述方法によっては高級な手続きも可能となる。第1表に文法<sup>9)</sup>の一例をBNF記法(Backus Normal Form)で示した。

**C. 対話**

コマンド言語による対話の方法は、指令順固定(closed sequence)による場合と、指令順任意(open sequence)によるものがある。

1. 指令順固定の対話

この対話は、概してこれまでのバッチ処理の言語の記述と似ている。たとえば、検索対象のファイルを指定し検索条件を前後の文脈に依存して記述する。この対話の欠点は、暗中模索的な機会がないことである。そこで処理の手続きが前もって明確な検索以外には使いにくい。

しかし、探索戦略も事前に立案されて短期間に結果だけを要求する対話に有利である。

数値データを取扱う検索手続きにバリエーションのあまり必要としない事務処理用のデータベース<sup>8)</sup>が、この種の対話を採用している。

2. 指令順任意の対話

この対話は相互作用を忠実に指令するのに有効であ

第2表 対話の方法と特性

対話の方法	能力	用途	データの属性	処理形式
指令順固定	非相互作用	アルゴリズムック	数値	コンパイラ
指令順任意	相互作用	ヒューリスティック	非数値	インタプリタ

る。質疑応答のくり返しと試行錯誤の検索により探索戦略を学習していきける。

そこで文献検索やシソーラス中のことばの処理に最適である。しかし、場合によってはこのような処理が冗長になり時間の無駄ともなる。

これらの対話の方法は、データベースの主題の性質や利用者の水準に依存して選択することになる。第2表に両者の特性を示す。

#### IV. インプリメンテーション

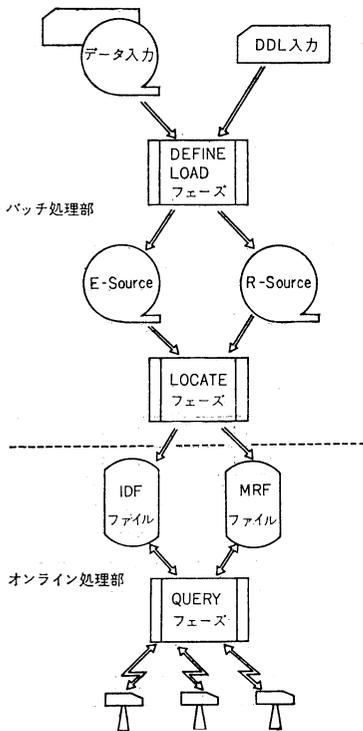
##### A. IDEAS について

このアプリケーションは IDEAS (Interactive Database Easy Accessing System)<sup>9)</sup> と呼ぶ。

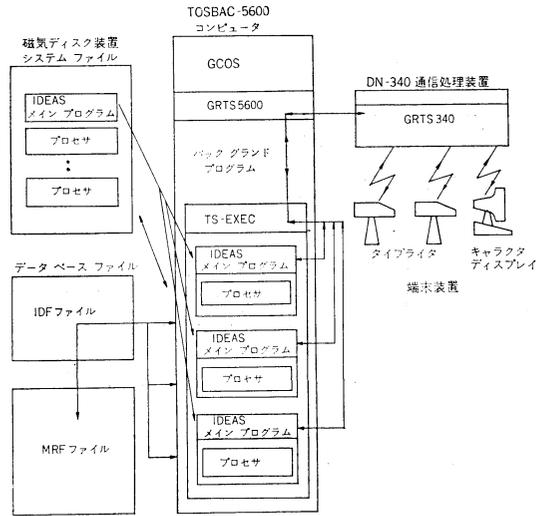
多くのデータベースシステムが採用しているように IDEAS も、データベースファイルの生成と更新を目的とするバッチ処理部と、質疑応答を目的とするオンライン処理部を分離し、それぞれ独立に実動できるようにしている。

##### 1. バッチ処理部

データの作成、収集後に一括して蓄積する次の処理フ



第3図 IDEAS のシステム構成



第4図 IDEAS の時分割処理の動作

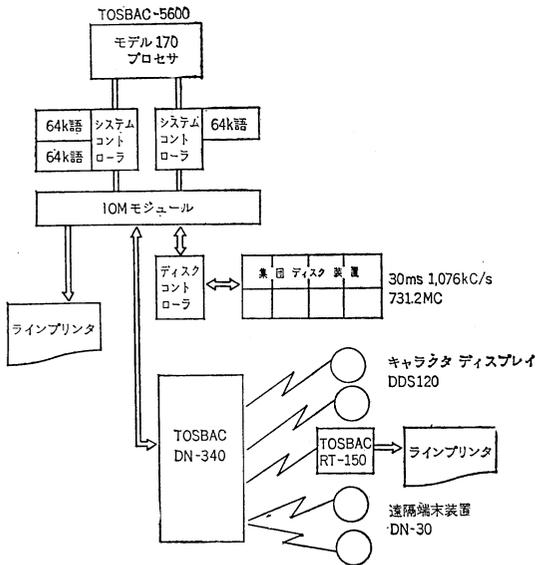
ーズからバッチ処理部は構成される。

- (1) DEFINE フェーズ: 入力レコードのフォーマット、項目名とデータの属性を定義する。
- (2) LOAD フェーズ: カードやテープ上のデータを読み取り、原始ファイルを生成する。
- (3) LOCATE フェーズ: 原始ファイルから抽出すべき項目を分離し、インバーテットファイル (IDF) とデータファイル (MRF) として展開する。

第3図は各フェーズの関係を示すものである。

##### 2. オンライン処理部

コマンド言語のプロセッサがオンライン処理部である。このプロセッサのコントローラが第3図の QUERY フェーズである。QUERY フェーズの時分割動作を第4図を中心に簡単に説明する。QUERY フェーズはオペレーティングシステム GCOS の下で、リアルタイム時分割に稼動するサブオペレーティングシステム TS-EXEC にコントロールされる。まず遠隔地にある端末装置から IDEAS の起動の要求があると、メッセージ処理専門の DN-340 コンピュータを介し GCOS に指令が行く。ここで TS-EXEC が入りこみ、そのコントロールによってパーマメントファイル上にある IDEAS のメインプログラムがスワップインされる。この状態でコマンド文の入力があれば、その種類に応じ言語プロセッサが呼ばれ解釈し実行に移る。もし第4図のように3台の端末装置によって同時に起動要求があったとすれば、3個の IDEAS メインプログラムが転写されて同一の TS-EXEC の下



第5図 IDEAS のハードウェアシステム

で動きます。この動きは時分割になるので利用者にとっては、あたかも自分専用に IDEAS がサービスしているように感じられる。

なお IDEAS のハードウェアシステムは第5図に示すとおりである。

**B. システムの特色**

IDEAS のインプリメンテーション上の特色は次のようになる。

- (1) 大規模になりがちなオンラインプログラムを平均8K語(32ビット/語)としている。
- (2) 他の無関係のプログラムと平行処理ができる。
- (3) 応答スピードを3秒以内とした。
- (4) 検索結果を一時保存する物理上のワークファイルを一切使用していない。

**C. 言語の仕様**

**1. QCL と DML**

IDEAS には対話の2方法、つまり指令順任意と、指令順固定に対応したQCL(Query Command Language)と、DML(Data Manipulation Language)がある。これらのコマンド名と機能の一覧を第3表に示した。ここでは第1図に示した相互作用に必要なコマンドの機能を説明する。

- (1) 試行検索(FIND): 指標を入口としてデータベースのレコードを探索し、セットを準備する。

- (2) 論理検索(COMBINE): 探索済みの2つ以上のセットの論理演算をし、サブセットを準備する。
- (3) 走査検索(SCAN): セットやサブセットをシケンシャルに走査する。自然語などの任意の文字照合に使用する。
- (4) 通覧(BROWSE): 指標中のキーワードのアルファベット順のリストを準備する。
- (5) シソーラスの学習(EXPAND): シソーラスエントリの構造を連想的に参照する。
- (6) 閲覧(SHOW): セットやサブセットのレコードやデータ項目の内容を出力する。

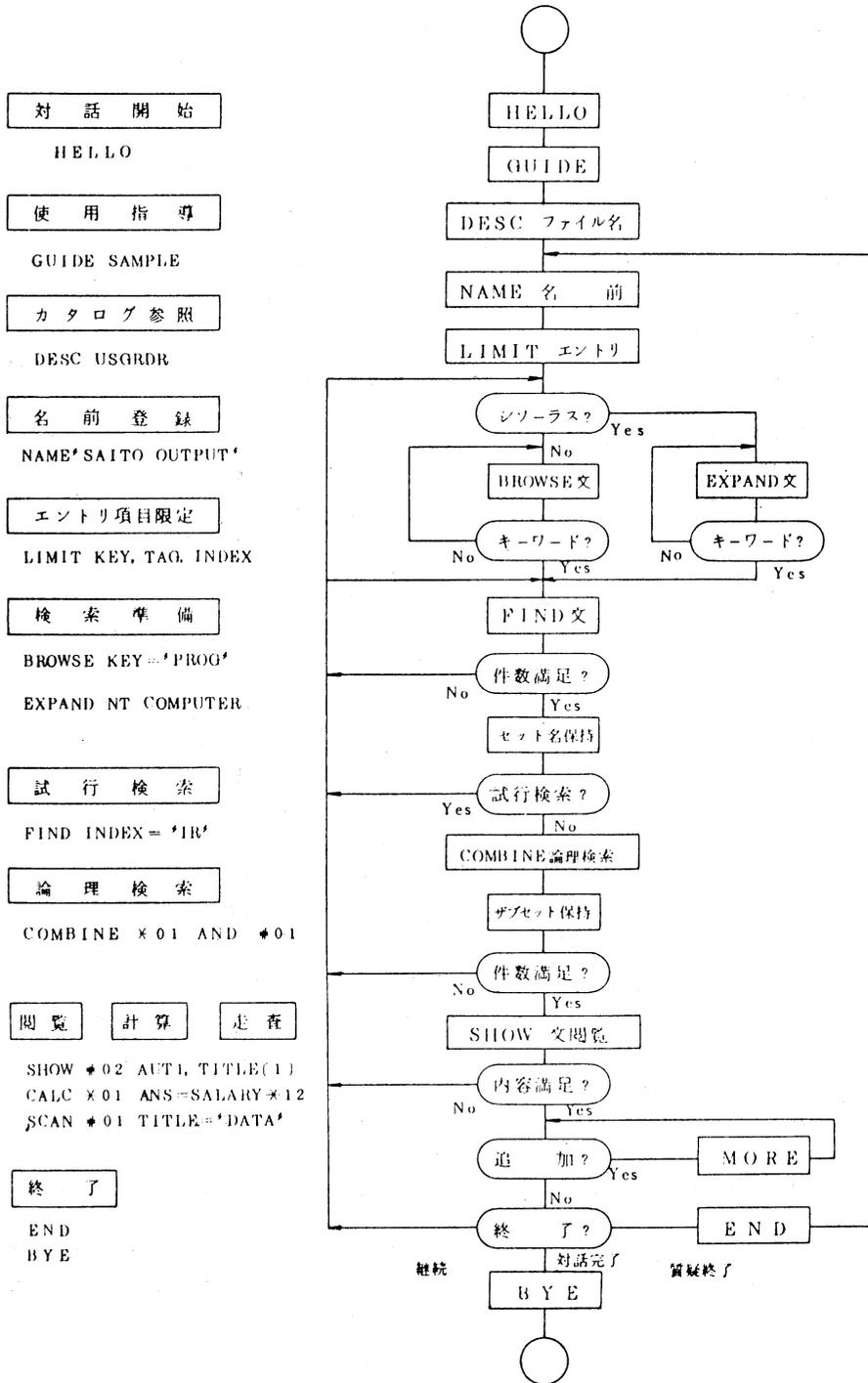
このほか第3表に示した IDEAS のコマンドについては文献<sup>10)</sup>を参照されたい。

**2. 対話の手続き**

DML は問題設定型なので試行錯誤による探索戦略の対話には乏しい。それに反し、QCL ではその手続きが流動的であるが、概して第6図に示すフローチャートになる。

第3表 QCL と DML のコマンド機能の一覧

機能	言語	QCL	DML
開始終了	会話開始	HELLO	・DML/I
	終了	END	END
	了会話	BYE	・ENDJOB
準備	対象制限	LIMIT	LIMIT
	通覧	BROWSE	
	シソーラス参照	EXPAND	
検索	試行	FIND	FIND
	論理	COMBINE	FIND
	走査	SCAN	SCAN
回答出力	閲覧表示	SHOW	
	追加	MORE	
	印刷標準	PRINT	PRINT
	印刷表		TABLE
計算	統計		LIST
	数値演算	CALC	CALC
	氏名登録	NAME	NAME
補助	使用指導	GUIDE	
	カタログ参照	DESCRIBE	
	通話記録	ECHO	
プロファイル定義	無効	CLEAR	
	生成手順書		DEFINE~EXIT
	生成主題書		SUBJECT~ENDSUB
	保存		SAVE
特殊	利用		USE/CALL
	関数		AVG/SUM/MAX/MIN



第6図 QCL の対話のフローチャート

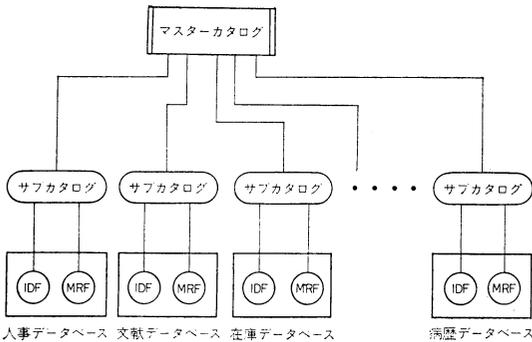
3. 対話の能力

QCL と DML によってどの程度の対話が可能であるかを次に整理した。

- (1) 指令順任意にも固定にも方法が取れる。
- (2) マイペースで納得のいくまで対話ができる。
- (3) 単純な質疑の積み重ねで複雑な検索ができる。
- (4) 言語自身の種類と機能を指導できる。
- (5) シソーラスの学習ができる。
- (6) データベースのカタログの参照ができる。
- (7) 数値データを対象に計算ができる。
- (8) 対話の記録を保持し再生できる。

D. ファイルの構造

IDEASのデータベースファイルはディレクトリ(IDF)とデータ(MRF)が対になって構成される。このファイルはバッチ処理により主題別や用途別に準備し、第7図に示すようなファイル名と機密保護のパスワードを持つカタログ化した共有ファイルコントローラにより管理される。



第7図 共有データベースの構成

1. インバーテット方式

IDF と MRF の関係は、指標とその存在を示すマトリックス状のインバーテット方式による。第8図を参考にしてサブファイルの機能を次に説明する。

- (1) DCOM: データベースファイルが開かれると同時にコアに常駐し、各サブファイル間の連絡情報を収録する。
- (2) KHF: 指標として項目の値がソートされている KEF の最大値のポインタが表になっている。
- (3) KEF: 指標の値とその頻度が入っている。
- (4) BMF: MRF と KEF がレコードとキーワー

ドのマトリックスであるが、この存在を示す表である。

- (5) KAT: MRF のレコードのアドレスを入れたポインタの表である。
2. ファイルの動作
- プロセッサによって動作する IDF と MRF の関連を第8図を例に説明する。
- (1) FIND コマンドにより探索するキーワードが入力されると、DCOM 中の KED を見て存在を確認する。
  - (2) KVD により KHF のポインタを得る。
  - (3) KHF では最大値が入っているので範囲を限定することができ、キーワードの頻度と BMF のポインタが得られる。
  - (4) ビット演算の後 BMF のビットの立った位置に対応する KAT から論理演算後、MRF のポインタが得られる。

V. 文献検索とシソーラス学習のための対話

A. 文献検索における対話

1. 対話の開始

IDEAS が起動すると ENTER COMMAND と「?」マークが打たれる。これはシステムが入力を要求していることを示す。そこで、第9図(a)のように HELLO を入力する。ここでデータベースのファイル名とパスワードの明示がある。

2. コマンドの指導

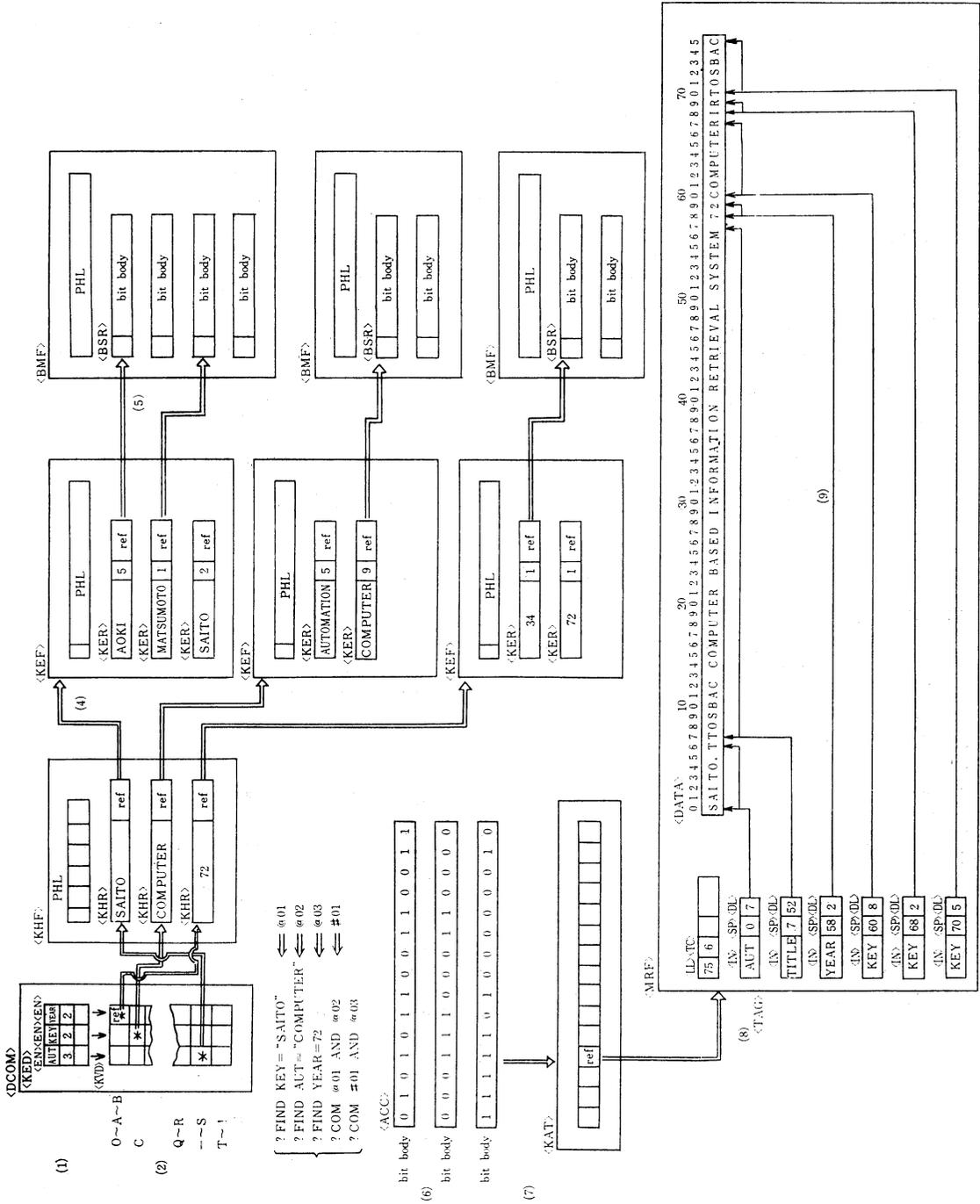
IDEAS のコマンドの種類と文例を知りたいければ第9図(b)のように GUIDE を入力する。

3. カタログの参照

データベースの項目名とその属性を登録したカタログを参照するには、第9図(c)のように DESCRIBE を入力する。

4. 指標の通覧と検索

BROWSE コマンドによりキーワード“elect”という文字を含む指標の通覧を要求すると、第9図(d)のようにアルファベット順のリストが出力される。次に第9図(f)で FIND ? \$07 として試行検索を指令すると応答件数は13件あって、それが \*02 セットに保存される。なお、\$07の意味は、(e)でリストされた“electrical resistance”のことである。そして、先の質疑で求められた \*01 と \*02 の論理和の検索を行なうと、その結果がサブセット #01 に保存される。



第8図 データベースファイルの構造

文献検索とシソーラス学習のための相互作用データベースシステムの開発

```

ENTER COMMAND
..2HELLO
REQUEST ACCEPTED.

GREETINGS. WELCOME TO IDEAS/1.

(a) PLEASE TYPE YOUR NAME AND SECURITY CODE OF DATABASE.
WHAT IS YOUR DATABASE NAME?
..7USERDR
WHAT IS YOUR SECURITY CODE?
..7HELLO
YOUR DATABASE WAS LAST UPDATED IN 73/07/19.
LET'S DIALOG WITH IDEAS/1.
NOTE THAT YOUR REQUEST SHOULD END WITH A CARRIAGE RETURN.

REQUEST COMPLETE.
    
```

```

ENTER COMMAND
..7GUIDE SAMPLE
REQUEST ACCEPTED.

** EACH COMMAND AVAILABLE TO THE IDEAS/1 WITH AN EXAMPLE OF ITS
FUNCTION IS DISPLAYED. **

1 BROWSE(BROW) BROW KEY = "INFORMATION RETRIEVAL " / BROW #01
2 BYE BYE
3 CALC CALC #08 SUM = TOTAL + COST
4 CLEAR CLEAR
5 COMBINE(COMB) COMB #01 AND #01 / COMB #01 OR #02
6 DESCRIBE(DESC) DESC ALISS / DESC ENTRY
7 ECHO ECHO
8 END END
9 EXPAND EXPAND BT = " COMPUTER SOFTWARE "
10 FIND FIND KEY = " DATA BASE SYSTEM " / FIND #02
11 GUIDE GUIDE / GUIDE SAMPLE / GUIDE COMMAND
12 HELLO HELLO
13 LIMIT LIMIT KEY,COST, YEAR, REF, CODE
14 MORE MORE 4
15 NAME NAME SAITO/TOSHIBA/(580)7111-2209
16 PRINT PRINT #05
17 SAVE SAVE
18 SCAN SCAN #05 COST =< 543
19 SHOW SHOW #05 TITLE,KEY, YEAR L,J:

REQUEST COMPLETE.
    
```

```

ENTER COMMAND
..7DESCRIBE USGRDR
REQUEST ACCEPTED.

** A FILE DESCRIPTION SUCH AS RECORDS, ITEMS AND ENTRIES DEFINED
FOR A DATABASE IS DISPLAYED. **

FILE NAME: USGRDR
CREATION DATE: 73/07/19

LVL ITEM TYPE SIZE TIMES
01 CPDEX
02 ID A 12
02 AUT1 A 36
02 AUT2 A 36
02 AUT3 A 36
02 AUT4 A 36
02 AUT5 A 36
02 TITLE A 360
02 KEY A 36
02 INDEX A 36
02 TAG A 36
02 TERM A 36
(c) 02 DESC A 36 6
02 BIB A 36
02 OPU A 36
02 UN A 36 5
    
```

```

ENTRY NAME
AUT1
KEY
INDEX
TAG
TERM
REQUEST COMPLETE.
    
```

```

ENTER COMMAND
..7NAME "MR. SAITO QUERY ANSWER"
REQUEST ACCEPTED.
    
```

```

ENTER COMMAND
..7LIMIT KEY,INDEX,TAG
REQUEST ACCEPTED.
    
```

```

CAPACITY EXCEEDED.
OVERFLOWED BY: TAG
REQUEST COMPLETE.
    
```

```

ENTER COMMAND
..2BROWSE KEY = "ELECT"
REQUEST ACCEPTED.

** A LIST OF SEVERAL VALUES OF A ENTRY WITHIN THE INCLUSIVE RANGE
BROWSED IS DISPLAYED. **
    
```

(d)

VN	VALUE	FREQ
\$01	EARTH RESOURCES TECHNOLOGY SATELLITE	3
\$02	EIGENVECTORS	1
\$03	ELASTIC COEFFICIENT	1
\$04	ELASTIC PROPERTIES	7
\$05	ELASTIC SCATTERING	1
\$06	ELECTRIC CURRENT	1
\$06	ELECTRIC DOUBLE LAYER	1
\$07	ELECTRIC EQUIPMENT	11
\$08	ELECTRIC FIELDS	1
\$09	ELECTRIC POWER GENERATION	3
\$10	ELECTRIC RELAYS	1

```

ENTER COMMAND
..7FIND #07
REQUEST ACCEPTED.

FOUND IN DATABASE.
FREQ OF VALUE: 11
ASSIGNED NAME: #01
REQUEST COMPLETE.
    
```

```

ENTER COMMAND
..7BROWSE #10
REQUEST ACCEPTED.

** A LIST OF SEVERAL VALUES OF A ENTRY WITHIN THE INCLUSIVE RANGE
BROWSED IS DISPLAYED. **
    
```

(e)

VN	VALUE	FREQ
\$01	ELECTRIC CURRENT	1
\$02	ELECTRIC DOUBLE LAYER	1
\$03	ELECTRIC EQUIPMENT	11
\$04	ELECTRIC FIELDS	1
\$05	ELECTRIC POWER GENERATION	3
\$06	ELECTRIC RELAYS	1
\$06	ELECTRICAL CONDUCTIVITY	1
\$07	ELECTRICAL RESISTANCE	13
\$08	ELECTROHYDRAULIC SERVO DRIVES	1
\$09	ELECTROMAGNETIC RADIATION	3
\$10	ELECTROMECHANICAL RELAYS	1

```

ENTER COMMAND
..7FIND #07
REQUEST ACCEPTED.

FOUND IN DATABASE.
FREQ OF VALUE: 13
ASSIGNED NAME: #02
REQUEST COMPLETE.
    
```

(f)

```

ENTER COMMAND
..7COMBINE #01 OR #02
REQUEST ACCEPTED.

CONDITION QUALIFIED.
COUNT OF RETRIEVED RECORDS: 24
TOTAL OF STORED RECORDS: 2797
ASSIGNED NAME: #01
REQUEST COMPLETE.
    
```

```

ENTER COMMAND
..7SHOW #01,AUT1,TITLE,KEY,INDEX,BIB (1)
REQUEST ACCEPTED.

MR. SAITO QUERY ANSWER
    
```

```

** A LIST OF RETRIEVED RECORDS FROM DATABASE IS DISPLAYED. **
    
```

```

RECORD: 1
AUT1 : WOLFF ERNEST G
TITLE : SEMI-CONDUCTORS FOR HIGH TEMPERATURE THERMOMETRY FINAL R
      EPT 22 JUN 64-22 MAY 65.
KEY : ELECTRICAL RESISTANCE
INDEX : RESPONSES
BIB : AD-480 105 USGRDR 72 20

REQUEST COMPLETE.
    
```

```

ENTER COMMAND
..7MORE 2
REQUEST ACCEPTED.

(g) MR. SAITO QUERY ANSWER
    
```

```

** A LIST OF RETRIEVED RECORDS FROM DATABASE IS DISPLAYED. **
    
```

```

RECORD: 1
AUT1 : KOZHARIN A S
TITLE : BEARING ASSEMBLY FOR AN ELECTRICAL MACHINE, BEARING ASSEM
      BLY FOR AN ELECTRICAL MACHINE--TRANSLATION
KEY : ELECTRIC EQUIPMENT
INDEX : *PLASTIC SEALS, BALL BEARINGS
BIB : AD-747 446 USGRDR 72 20

REQUEST COMPLETE.
    
```

```

RECORD: 2
AUT1 : LATYSHEV V N
TITLE : STUDY OF ELECTRICAL RESISTANCE OF LUBRICATING FILM IN THE
      FRICTION OF VARIOUS METALS, STUDY OF ELECTRICAL RESISTANC
      E OF LUBRICATING FILM IN THE FRICTION OF VARIOUS METALS--T
      RANSLATION
KEY : ELECTRICAL RESISTANCE
INDEX : *LUBRICANTS, RESISTANCE(ELECTRICAL
      BIB : AD-747 451 USGRDR 72 20

REQUEST COMPLETE.
    
```

第9図 QCL の対話の実例

5. 内容の閲覧

第9図(f)で保存されたサブセット #01 の内容から、例えば著者(AUTI)、標題(TITLE)、キーワード(KEY)などの内容の出力を求めるものである。

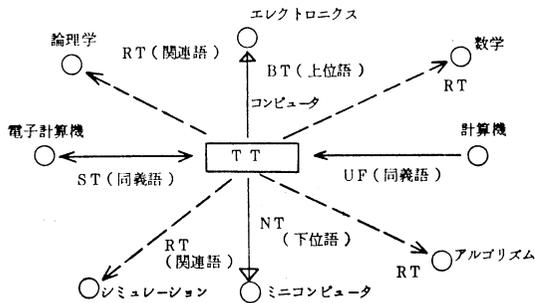
B. シソーラス学習における対話

1. 学習の効果

シソーラス学習における対話の目的から、次の効果が期待できる。

- (1) 検索すべき主題の糸口を与え、発見的探索を助ける。
- (2) 主題を記述するキーワード語彙の学習ができる。
- (3) 概念からアプローチし、それを表記するキーワードや組み合わせのヒントを得られる。
- (4) キーワードの階層的、統計的關係と関連といった文脈中の位置を把握することができる。
- (5) データベースの主題分析の傾向を知ることができる。
- (6) シソーラスエントリをイモヅル状に引き出すことから連想的主題の把握ができる。
- (7) シソーラス自身のキーワードとその関係付けが妥当であるかといった品質の検定ができる。

ここで示した対話型式でシソーラスを利用する意図は、直接検索効率を向上させようとするのではなく、単に学習という補助的用途のためである。



TT	: コンピュータ
UF	: 計算機
ST	: 電子計算機
BT	: エレクトロニクス
NT	: ミニコンピュータ
RT	: 論理学
RT	: 数学
RT	: シミュレーション
RT	: アルゴリズム

第10図 シソーラスエントリの構成

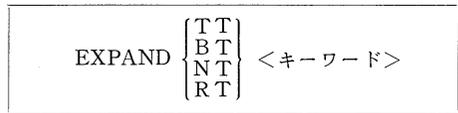
2. エントリの構成

シソーラスはエントリと呼ぶキーワードの用途の定義を、他のキーワードと関連子により位置づけたレコードの集合である。

エントリは第10図に示すようなターゲット語(TT)を中心に、それより上位の概念を示すキーワード(BT)と、下位の概念を示すキーワード(NT)と、同一レベル上にある同義語(SF)や、それに関連するキーワード(RT)などを表にしたものである。

3. 学習のコマンド

シソーラス学習のコマンドはEXPAND<sup>11)</sup>と呼ぶ。EXPAND文の一般形式を次に示す。



たとえば、EXPAND BT “IR system” とすると、上位語(BT)として “IR system” というキーワードを含む全てのエントリを拡張せよ、といった意味である。

4. 学習の実際

ここで実際のEXPANDコマンドを使った対話の実際を第11図に示そう。

①では “information retrieval” を関連語(RT)とするエントリの要求をし、②に “subject indexing” をターゲット語とする6件のエントリの中の1件のエントリが出力された。次々と④の “indexes”, さらに⑥の “documentation”, ⑧の “data retrieval” が出力された。新しく “data retrieval” を関連語(RT)として拡張すると、該当したエントリが2件あり、その1件の⑩の “information retrieval” を出力し、⑫の “data transmission” を追加した。そこで⑭では、⑭のエントリ中に関連語(RT)として含まれる “information theory” を拡張すると、“information” や “data processing” が出力されている。この学習の対話は、特定の主題に対して目的意識を持って行ったものでないが、図解すると第12図に示すような関連地図<sup>12)</sup>が描ける。

5. 検索への応用

QCLのコマンドと併用し、本格的なシソーラス活用の文献検索のための対話は、第13図のようになろう。

6. ファイルの構成

シソーラスファイルはTVFと呼び、その構造は、

文献検索とシソーラス学習のための相互作用データベースシステムの開発

```

ENTER COMMAND
?EXPAND RT "INFORMATION RETRIEVAL"
REQUEST ACCEPTED.

FOUND IN THESAURUS.
COUNT OF ENTRIES: 6

** A LIST OF ENTRY IN THESAURUS IS DISPLAYED. **

② TT: SUBJECT INDEXING          $C0
  UF: CLASSIFICATION            $C1
  UF: SUBJECT CLASSIFYING      $C2
  NT: COORDINATE INDEXING      $C3
  RT: AUTOMATIC INDEXING      $C4
  RT: CATALOGING              $C5
  RT: CLASSIFICATIONS          $C6
  RT: DESCRIBING               $C7
  RT: DESCRIPTIVE CATALOGING  $C8
  RT: DOCUMENT ANNOUNCEMENT   $C9
  RT: DOCUMENTATION            $10
  RT: IDENTIFYING              $11
  RT: INDEXES                  $12
  RT: INFORMATION RETRIEVAL    $13
  RT: INFORMATION RETRIEVAL EFFECTIVENESS $14
  RT: LIBRARY SCIENCE         $15

  * CATEGORY : C5C2
  * ENTRY ID : KWCCCC9

REQUEST COMPLETE.

DO YOU WANT MORE ENTRIES?
?YES

** A LIST OF ENTRY IN THESAURUS IS DISPLAYED. **

④ TT: INDEXES                   $C0
  UF: CATALOGS                  $C1
  RT: CLASSIFICATIONS           $C2
  RT: DESCRIPTIVE CATALOGING   $C3
  RT: DIRECTORIES              $C4
  RT: DOCUMENTS                 $C5
  RT: INFORMATION RETRIEVAL    $C6
  RT: LIBRARY SCIENCE          $C7
  RT: SUBJECT INDEXING         $C8

  * CATEGORY : C5C2
  * ENTRY ID : KWCCCC10

REQUEST COMPLETE.

DO YOU WANT MORE ENTRIES?
?YES

** A LIST OF ENTRY IN THESAURUS IS DISPLAYED. **

⑥ TT: DOCUMENTATION            $C0
  RT: ABSTRACTING              $C1
  RT: DESCRIPTIVE CATALOGING  $C2
  RT: DOCUMENT ACQUISITION    $C3
  RT: DOCUMENT ANNOUNCEMENT   $C4
  RT: DOCUMENT CIRCULATION     $C5
  RT: DOCUMENT STORAGE        $C6
  RT: DOCUMENTALISTS          $C7
  RT: DOCUMENTS                $C8
  RT: INFORMATION RETRIEVAL    $C9
  RT: INFORMATION SYSTEMS     $10
  RT: LIBRARY SCIENCE         $11
  RT: REPORTING                $12
  RT: SEARCH STRUCTURING      $13
  RT: SUBJECT INDEXING        $14
  RT: TECHNICAL WRITING       $15
  RT: TRANSLATING             $16

  * CATEGORY : C5C2
  * ENTRY ID : KWCCCC12

REQUEST COMPLETE.

DO YOU WANT MORE ENTRIES?
?YES

** A LIST OF ENTRY IN THESAURUS IS DISPLAYED. **

⑧ TT: DATA RETRIEVAL          $C0
  RT: DATA PROCESSING         $C1
  RT: DATA ACQUISITION        $C2
  RT: DATA STORAGE            $C3
  RT: DATA TRANSMISSION       $C4
  RT: GRAPH THEORY             $C5
  RT: INFORMATION RETRIEVAL    $C6

  * CATEGORY : C5C2
  * ENTRY ID : KWCCCC20

REQUEST COMPLETE.

DO YOU WANT MORE ENTRIES?
?NO

REQUEST COMPLETE.

ENTER COMMAND
?EXPAND RT "DATA RETRIEVAL"
REQUEST ACCEPTED.

FOUND IN THESAURUS.
COUNT OF ENTRIES: 2

** A LIST OF ENTRY IN THESAURUS IS DISPLAYED. **

⑩ TT: INFORMATION RETRIEVAL    $C0
  UF: DOCUMENT RETRIEVAL      $C1
  UF: RECORDS RETRIEVAL       $C2
  RT: DATA RETRIEVAL         $C3
  RT: DESCRIPTIVE CATALOGING $C4
  RT: DOCUMENTATION           $C5

```

```

RT: INDEXES                   $C0
RT: INFORMATION RETRIEVAL EFFECTIVENESS $C7
RT: INFORMATION SYSTEMS       $C8
RT: SUBJECT INDEXING          $C9

  * CATEGORY : C5C2
  * ENTRY ID : KWCCCC8

REQUEST COMPLETE.

DO YOU WANT MORE ENTRIES?
?YES

** A LIST OF ENTRY IN THESAURUS IS DISPLAYED. **

⑬ TT: DATA TRANSMISSION       $C0
  UF: DIGITAL DATA TRANSMISSION $C1
  RT: TRANSMISSION             $C2
  RT: CODING                    $C3
  RT: DATA ACQUISITION        $C4
  RT: DATA LINKS              $C5
  RT: DATA PROCESSING         $C6
  RT: DATA RETRIEVAL          $C7
  RT: ELECTROMAGNETIC WAVE TRANSMISSION $C8
  RT: INFORMATION THEORY       $C9
  RT: MULTIPLEXING             $10
  RT: RADIO TELEMETRY         $11
  RT: READING                   $12
  RT: STOCHASTIC PROCESSING    $13
  RT: TELECOMMUNICATION        $14
  RT: TELEMETRY                $15
  RT: WRITING                   $16

  * CATEGORY : C9C4
  * ENTRY ID : KWCCCC27

REQUEST COMPLETE.

DO YOU WANT MORE ENTRIES?
?NO

REQUEST COMPLETE.

ENTER COMMAND
?EXPAND TT "INFORMATION THEORY"
REQUEST ACCEPTED.

NOT FOUND IN THESAURUS.
PLEASE TRY AGAIN.

⑮ ?EXPAND RT "INFORMATION THEORY"
REQUEST ACCEPTED.

FOUND IN THESAURUS.
COUNT OF ENTRIES: 4

** A LIST OF ENTRY IN THESAURUS IS DISPLAYED. **

⑰ TT: INFORMATION              $C0
  UF: RECORDED INFORMATION      $C1
  NT: BATHYTHERMOGRAPH DATA   $C2
  NT: EXPERIMENTAL DATA       $C3
  NT: METEOROLOGICAL DATA     $C4
  NT: OCEANOGRAPHIC DATA      $C5
  NT: STATISTICAL DATA        $C6
  NT: STEAM TABLES            $C7
  NT: TABLES                   $C8
  NT: TELEMETERING DATA       $C9
  RT: INFORMATION THEORY        $10

  * CATEGORY : 14C7
  * ENTRY ID : KWCCCC7

REQUEST COMPLETE.

DO YOU WANT MORE ENTRIES?
?YES

** A LIST OF ENTRY IN THESAURUS IS DISPLAYED. **

⑲ TT: DATA PROCESSING         $C0
  UF: AUTOMATIC DATA PROCESSING $C1
  UF: COMPUTER TECHNOLOGY       $C2
  UF: DATA ANALYSIS            $C3
  UF: DATA HANDLING            $C4
  UF: ELECTRONIC DATA PROCESSING $C5
  UF: INTEGRATED DATA PROCESSING $C6
  NT: DATA REDUCTION           $C7
  NT: DATA RETRIEVAL           $C8
  NT: DATA SMOOTHING          $C9
  NT: DATA STORAGE             $10
  RT: AUXILIARY EQUIPMENT       $11
  RT: CODING THEORY             $12
  RT: COMPUTATION               $13
  RT: COMPUTER PROGRAMMING      $14
  RT: COMPUTER STORAGE DEVICES  $15
  RT: COMPUTER SYSTEMS HARDWARE $16
  RT: COMPUTER SYSTEMS PROGRAMS $17
  RT: COMPUTERS                  $18
  RT: DATA                      $19
  RT: DATA PROCESSING EQUIPMENT $20
  RT: DATA RECORDING           $21
  RT: DATA SAMPLING            $22
  RT: DATA TRANSMISSION        $23
  RT: INFORMATION THEORY        $24
  RT: PROGRAMMING LANGUAGES     $25
  RT: RECORDS MANAGEMENT       $26
  RT: SYSTEMS ENGINEERING       $27
  RT: TABULATION PROCESSES      $28

  * CATEGORY : C9C2
  * ENTRY ID : KWCCCC22

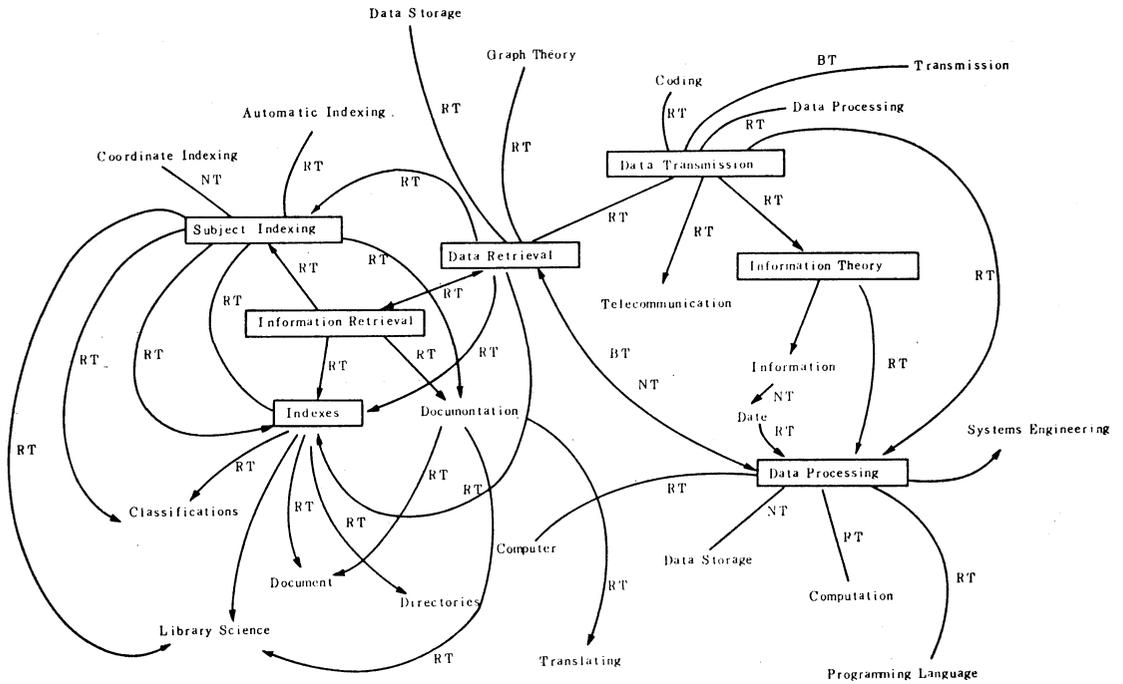
REQUEST COMPLETE.

DO YOU WANT MORE ENTRIES?
?NO

REQUEST COMPLETE.

```

第11図 シソーラス学習の対話例



第 12 図 対話から生成される関連地図

IDS (Integrate Data Store)<sup>13)</sup> の得意とするチェーン構造を採用しており、第 14 図のように関連子とエントリのチェーンになっている。この構造の特色はただ一つのキーワードを一回登録することによりエントリ間に重複することなく、しかも記憶スペースの大幅な節約ができることである。

おわりに

この種のオンライン IR システムの例に、古くは INTREX<sup>14)</sup>、DIALOG<sup>15)</sup>、そして STAIRS<sup>16)</sup> があるし、また ORBIT<sup>17)</sup> とその大規模な実用化システム MEDLINE<sup>18)</sup> が有名である。そして日本においても JOLDOR<sup>19)</sup> と DOOR<sup>20)</sup> と呼ぶ試作システムがあった。これらを網羅的に展望した最近の名著に Lancaster に

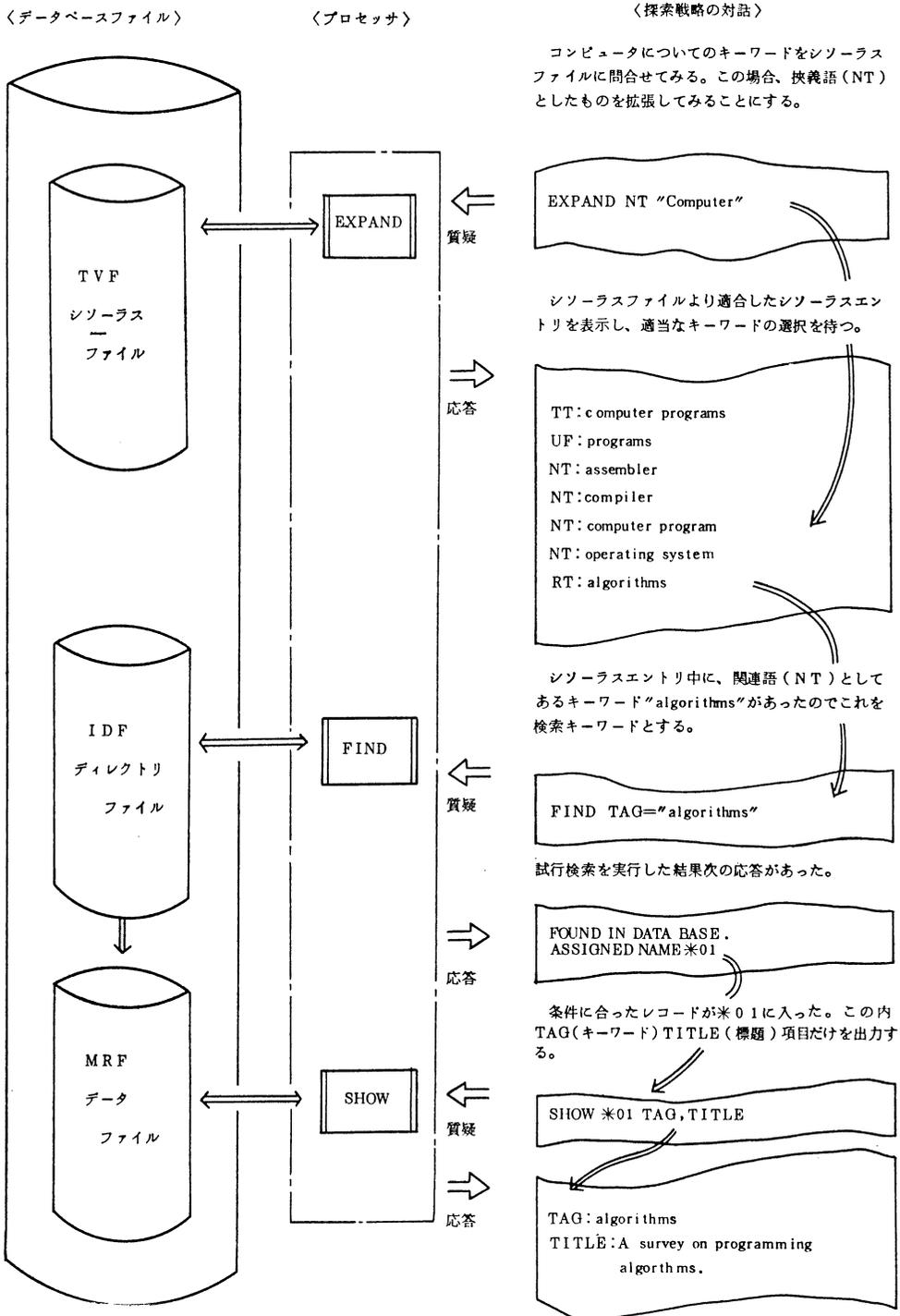
よる *Information retrieval on-line*<sup>21)</sup> がある。

オンライン IR システムの価値は、対話と呼ぶ相互作用ができることから、利用者が直接自分自身のベースで探索戦略をたてることのできる点にあった。しかし、これには無駄な対話による冗長性の増大による経済性の低下などの欠点もある。

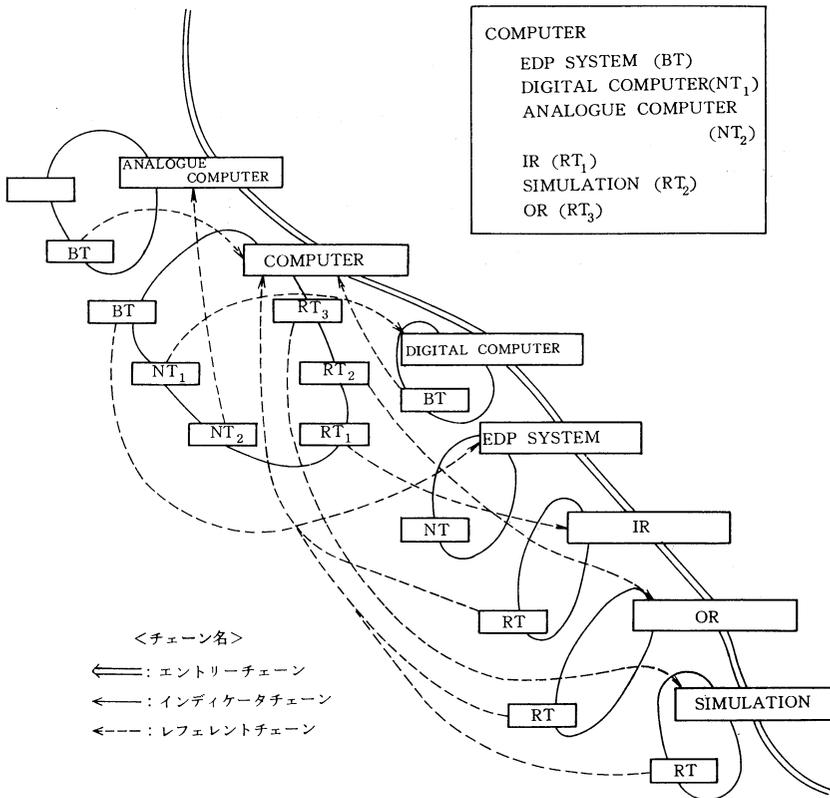
また、データベースの準備と管理などのバッチ処理が、そのタイミングと操作に思わぬ障害となる。このことは情報の頻繁な更新を必要とする SDI サービスでは大問題になる。

さらに、指標のないデータベースにおいては依然としてシーケンシャルファイルを対象にする自然語検索が有効であろう。

文献検索とシソーラス学習のための相互作用データベースシステムの開発



第 13 図 シソーラスと文献の対話



第14図 チェーン構造展開のシソーラスエントリ

- 1) DBTG report to the CODASYL. New York, Programming Language Committee, ACM, 1971.
- 2) Hewitt, C. *PLANNER: A language for manipulating models and proving theorems in a robot.* Cambridge, MIT, 1970. (AI memo 168).
- 3) Codd, E. F. "A relational model of data for large shared data banks," *Communications of the ACM*, vol. 13, June 1970, p. 377-87.
- 4) 斎藤 孝. "時分割処理による対話型文献検索システム IDEAS/1," *情報管理*, vol. 16, 1974. 2, p. 874-81.
- 5) 斎藤 孝. "TOSBAC-5600 オンライン時分割データベース検索システム IDEAS," *パルス*, vol. 4, no. 2, 1974, p. 3-7.
- 6) 斎藤 孝. "対話型データベースシステムの展望," *事務管理*, vol. 3, no. 7, 1974, p. 23-30.
- 7) 斎藤 孝 et al. 対話型質疑応答システムのための指令順任意と指令順固定方式記述言語プロセッサの開発と評価. <日本科学技術情報センター. 第10回情報科学技術研究会発表論文集. 東京, 1973> p. 171-84.
- 8) IBM. *Interactive query facility for IMS/360: Version 2.* General information manual GH20-0765-2. New York, 1972.
- 9) 斎藤 孝 et al. "時分割処理による対話形データベース検索システム IDEAS/1," *東芝レビュー*, vol. 29, 1974. 1, p. 73-8.
- 10) 東京芝浦電気株式会社. TOSBAC-5600 対話型データベース検索システム IDEAS システム概説書. (56AP418A). 東京, 1973.
- 11) 東京芝浦電気株式会社. TOSBAC-5600 シソーラスを用いた IR システム EXPAND 解説書 (56AP 422A). 東京, 1973.
- 12) 斎藤 孝 et al. シソーラスを対話型に活用するプロセッサの開発. 情報処理学会第14回大会において発表. 1974.
- 13) 斎藤 孝 et al. シソーラスファイルのジェネレーションとリアルタイムによる問合せ処理のプロセッサの設計. <日本科学技術情報センター. 第9回情報科学技術研究会発表論文集. 東京, 1972> p. 235-42.
- 14) Reintjes, J. F. "System characteristics of IN-

文献検索とソース学習のための相互作用データベースシステムの開発

- TREX," *AFIPS conference proceedings*, Spring joint computer conference. vol. 34, 1969, p. 457-9.
- 15) Summit, R. K. "DIALOG: an operational on-line reference retrieval system," *Proceedings of the ACM national meeting*, 1967, p. 51-6.
- 16) IBM. *Storage and information retrieval system (STAIRS)*: General information manual (QH12-5107). New York, 1973.
- 17) Seiden, H. R. *ORBIT system information*. Santa Monica, System Development Corporation, 1971.
- 18) National Library of Medicine. *The principles of MEDLARS*. Bethesda, 1970.
- 19) 小川義久 *et al.* "オンライン文献検索システム JORDOR." 第10回プログラミングシンポジウム報告集. 1969. p. 24-46.
- 20) 林 省三, 福島芳直. "オンライン検索," *情報管理* vol. 15, 1972. 5, p. 125-40.
- 21) Lancaster, F. W. *Information retrieval on-line*. Los Angeles, Melville, 1973.