

# パソコンの関係データベースシステム桐の評価

5Q-6

曹 南珠 西村 恕彦  
東京農工大学数理情報工学科

## 1. まえがき

パーソナルコンピュータの普及につれ、最近は、設計の段階を知らなくても使用できるような一般利用者向けのデータベースシステムが増えている。ここでは、その中でも関係方式のデータベースシステムの評価と図書データベースシステムの設計を行った。

## 2. 関係データベースシステムの評価

### 関係演算の評価

関係演算の中でも関係代数の演算である選択および制約、射影、結合を評価の対象とした。実験で使用された図書データベースの中で、一番検索される率の高い項目である著者名と書名を条件項目として、データの格納率別に演算の実行時間を測定して、その結果を用いて関係演算の評価を行った。検索の条件を決めるためには、書名と著者名の分布を調べて、一般性の高い格納密度部分を選んだ。

### 検索時間の評価

検索時間を測定して同じ条件で検索を行ったときを推定した。桐の言語を用いて測定作業に必要なプログラムを作成したので、データを探すだけに消費された純粋な検索実行時間を算出するために、二つの方法で検索を行った。検索実行時間の中で、命令の解釈時間と命令の実行時間を計算するために、文献データベースからランダムに抽出した1レコードだけのテスト用文献ファイルを作成して、そのファイルから一件のレコードを検索する時間をベース時間とした。約1300の母集団から100の大きさのサンプルを抽出して、そのサンプルについて実験を行った。サンプルデータの抽出は乱数を発生させた。

検索方法において、一件ずつ検索して、各検索時間からベース時間を引く場合と、100件を連続して検索した時間から合計ベース時間を引く場合とを比較してみると後者の方がより正確であることが分かった。

た。格納密度と検索時間の分析は格納密度別(30%, 50%, 70%, 80%, 90%)について検索時間を算出して、その結果を用いて検索時間の推定を行う。

図 1.

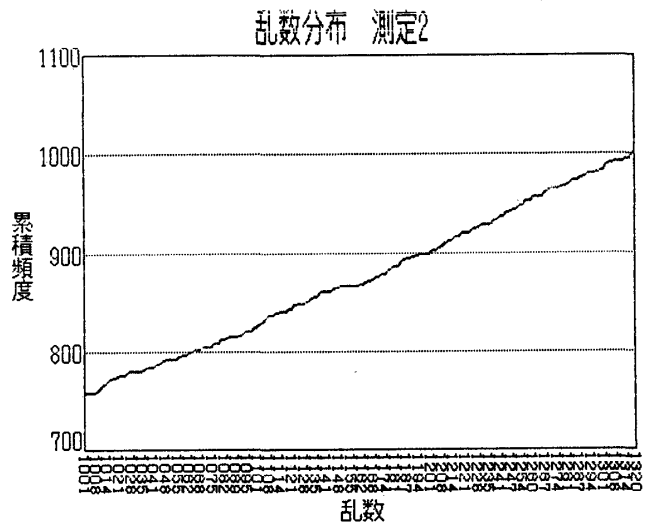


表 1. 文献検索時間測定結果

検索方法	1 回目	...
BASE TIME	00分02秒	...
索引 無 - BASE TIME	01分27秒	
索引 無 W/O BASE	01分28秒	
索引 有 - BASE TIME	00分15秒	
索引 有 W/O BASE	00分15秒	
方式B BASE TIME	00分04秒	
方式B 索引 無	01分13秒	
方式B 索引 有	00分12秒	

サンプルの大きさを100件のデータに固定しておいて、索引のある場合と索引のない場合と別々に検索時間を測定した。測定の結果から索引使用率を計算

して索引使用の評価を行った。索引使用率は100件につき10\*10回測定した結果から有/無\*100%を計算した。図書データベースシステムの検索に対する索引の項目は、図書番号、書名著者名、分類番号、発行年であるが、検索時間測定に使用される索引は書名と著者名である。索引のある、ない場合の検索時間を測定した結果は索引がある場合がない場合に比べ約3倍速いことがわかった。

#### 整列時間の評価

整列する条件を辞書順に固定しておいて、サンプルの大きさを100、300、500、700、800、900、1000、1300にわけて、それぞれのサンプルに対して文字列と整数との場合について検索時間を測定する。もう一つの評価方法として、整列の条件とサンプルの大きさを固定しておいて、項目の数を変えながら検索時間の測定を行う。

#### 物理的面からの評価

測定対象となった文献ファイルはフロッピーに収録されているので、領域不足のために表ファイルのバックアップをとってない。バックアップがないので、測定が行われたつど、表ファイルの物理的位置が変わることはないし、検索時間に影響を及ぼすことはないと思われる。ポインタのサイズは文献データベースの場合はデータベース全体の約9.41%にあたるということがわかった。索引レコードのサイズは整列項目値の長さの合計に14を加算したバイト数である。

## 2. 対話式図書データベースシステムの設計

パーソナルコンピュータにおけるデータベースシステムがもつべき機能としては、データの入力、データの蓄積、データの検索、データの加工、繰り返して実行する機能などを上げられる。本システムではデータ入力と検索の部分を中心に設計した。特にパーソナルコンピュータでデータベースシステムを使用する場合は、データベースの設計者と利用者が同一であるケースが多いので、全作業を対話形式にして、さらにメニュー画面、指示などは日本語にしてより使いやすさをはかった。

#### メニューでの検索項目

1. 分類番号
2. 発行年
3. 図書番号
4. 書名
5. 著者名

## 6. 出版社名

指定方法は、分類番号は3桁の数字、発行年は西暦で4桁の数字、図書番号は6桁、書名はローマ字で49字、かな漢字で24文字まで、著者名はローマ字で30文字、かな漢字で15文字まで、出版社名はローマ字で35文字まで指定できる。

データの入力と更新の場合も、検索の場合のように、入力項目指定画面と更新項目指定画面によって指定できるようになっている。

文献検索システムの全作業は、メニュー画面表示することによって対話式にした。さらに、補助機能としてHELPキーを準備してキーを押すことで各項目の指定の仕方を知ることができるようにした。さらに、検索命令が実行されている実行状態、検索件数、検索レコードの平均長さ、最大と最小の長さ、標準偏差なども知ることができる。実行状態は全体データに対する実行比率として表示されるようになっている。

## 3. まとめ

本実験で使用されたデータは、東京農工大学数理情報工学科図書室の1987年までの1321冊の図書に基づいたものなので、データベースの規模としては小さい方である。検索時間評価と領域問題について、もっと考えるべきである。最近、関係方式のデータベースシステムが利用されるようになっているが、使用される環境と条件などによる分析が必要だと思う。

#### 参考文献

- 1) 植村：データベースシステムの基礎、オーム社(1979)
- 2) 平尾：関係データベースシステム、近代科学社(1986)
- 3) 芝田：図書データベースの評価と設計、東京農工大学卒論(1984)
- 4) エン：データベースの評価と設計、東京農工大学修士論文(1983)
- 5) 酒井：データベースシステムの実際、サイエンス社(1987)