

拡張関係データベースエンジン

X R D B の性能評価

6Q-7

成田 実香\*\*, 山根 康男\*, 小櫻 文彦\*\*

\* ㈱富士通研究所, \*\* 富士通株式会社

1. はじめに

拡張関係データベースエンジンX R D Bは、UNIX上で動くシステムである。シングルユーザ環境で稼働するプロトタイプの完成にあたり性能評価を行ったので、結果を報告する。

評価には2通りの方法を用いた。まず、ウィスコンシンベンチマークテスト〔1〕をもとに、性能測定および他システムとの比較を行った。次に、辞書アクセスのために専用で作成されていたルーチン(今後、辞書専用ルーチンと呼ぶ)との形態素解析に関する性能比較を行った。以下、X R D Bの性能評価の方法について説明し、考察を述べる。

2. ウィスコンシンベンチマーク

性能測定は東芝ワークステーションAS3260(SUN-3/260のOEM)上で行った。主記憶は16MBで、CPUはMC68020(25MHz)である。

文献〔1〕に従って選択・結合・射影・集合関数・更新の5項目について測定を行ったが、ここでは選択・結合演算の測定結果の一部を示すにとどめる。

選択演算では、10000タプルのシーケンシャルリレーションから10%を選択する。結合演算では、10000タプルのシーケンシャルリレーションから10%の選択を行い、その結果と10000タプルのリレーションとを結合する。

X R D Bの処理時間における入出力の占める割合を調べるために、データがバッファ上に全てのっている状態から測定をはじめて入出力処理が全く起こらないようにした場合と、データがバッファ上に全くのっていない状態から測定を行った通常の場合との2通りの測定を行った。ここでは、前者を主記憶測定、後者を二次記憶測定と呼ぶことにする。

表 2.1, 2.2 より、経過時間に占める入出力処理時間の割合は、選択演算で 9%, 結合演算で12% となっていることがわかる。これは、X R D Bではその処理の大部

表2.1 選択演算

	経過時間	CPU 時間	入出力回数
二次記憶測定	4.7 sec	1.8 sec	241 回
主記憶測定	0.4 sec	0.4 sec	0 回

表2.2 結合演算

	経過時間	CPU 時間	入出力回数
二次記憶測定	10.8 sec	4.2 sec	637 回
主記憶測定	1.3 sec	1.3 sec	0 回

分を入出力処理が占めていることを示している。

このようにCPU処理の割合が少なくなっているのは、X R D Bで採られた高速化の手法の効果があらわれているものと考えられる。

X R D Bでは、処理の高速化の手法として、結合処理にハッシュを用いている。従来の結合処理ではソートが用いられていたが、X R D Bで測定したところ、10000タプルのソートを行うのに26秒の時間を要することがわかった。このことは、ソートによる結合処理は少なくとも26秒かかるということの意味している。一方ハッシュを用いた結合の場合には、表2.2 から、処理に 10.8 秒かかることがわかる。結合処理にハッシュを用いたことにより、X R D Bの高速化がはかられたといえる。

表2.3 は、上記の選択・結合演算に関する他システムとの性能比較を、経過時間について示したものである。表中で、I N G R E Sの測定値はVAX11/750上のデータを用いた。D B C / 1 0 1 2, G a m m a に関しては文

献〔2〕の測定値を用いた。ただしDBC/1012, Gammaではデータのタプル長は208バイトになっている。

表2.3 他システムとの比較(経過時間)

	選択演算	結合演算
XRDB	4.7 sec	10.8 sec
INGRES	53.9 sec	108 sec
DBC/1012	16.0 sec	35.6 sec
Gamma	2.1 sec	5.1 sec

表2.3においてDBC/1012やGammaが並列処理を行っていることを考慮すると、XRDBの高速性が示せたものと思う。

### 3.形態素解析

性能測定はSONYワークステーションNEWS上で行った。主記憶は8MBで、CPUはMC68020(16.67MHz)である。

辞書専用ルーチンの単語辞書は約20.5Mバイトで、タプル数は約116,000である。単語辞書をXRDBで使用する形式にコンバージョンした結果、生成された辞書の大きさは約23Mバイトであった。このXRDBの辞書は、辞書の昇順にソートされたシーケンシャルなリレーションにB-treeによるインデックスをつけた構成になっている。

XRDBと辞書専用ルーチンを用いて形態素解析を行った。形態素解析の中でもとくに、辞書から単語を検索してくる部分の時間と入出力回数について、測定を行った。したがって、辞書検索を除いた形態素解析処理と辞書のオープン・クローズ処理とにかかる実行時間と入出力回数は、測定結果に含まれていない。

辞書専用ルーチンを1としたときのXRDBの相対性能を表3.1に示す。

入力文として、文(a), (b)を用いた。

(a) 私は風邪です。

(b) これまで日本の広告の中では、他社の製品やブランドをそのまま使って広告するという事はあまり見られなかった。

表3.1の結果より、形態素解析専用の辞書専用ルーチンと比較してもXRDBを用いた検索は遅色がないことがわかる。

表3.1 XRDBの相対性能  
(辞書専用ルーチンを1としたとき)

	経過時間	CPU時間	入出力回数
文(a)	0.88倍	1.00倍	1.50倍
文(b)	0.97倍	0.95倍	0.98倍

### 4.おわりに

XRDBのプロトタイプにおける性能評価について報告した。XRDBと他システムとの比較を2通りの評価方法に関して行ったが、いずれもXRDBの高速性を示す良好な結果を得た。

今後はマルチユーザ環境への移行・並列処理等の課題を解決し、システムの機能を拡張していく予定である。

### 謝辞

この研究は通産省工業技術院大型プロジェクト「電子計算機相互運用データベースシステムの研究開発」の一環として行った。

### 参考文献

- 〔1〕 Bitton, D. D. J. DeWitt et al.: Benchmarking Database systems, A Systematic Approach Computer Sciences Department University of Wisconsin-Madison, 1983
- 〔2〕 D. J. DeWitt et al.: A Single User Evaluation of The GAMMA Database Machine, Proc. of the 5th Int. Workshop on Database Machine, 1987