ソーティングにおける高速化の一手法（Ⅰ）

1P-7

－ タグトーナメント技法 －

前島泰*，米城範正**，大熊和明**，小端則夫**
* ㈱富士通静岡エンジニアリング
** 富士通株式会社

1．はじめに

近年，データ量は著しく増加しており，これを扱う外部ソートも一層の高速化が必要となっている．

外部ソートは，入力したレコードから複数の昇順（降順）に並んだレコード列を作るストリング生成部と，これら複数のストリングを一つにまとめるストリング併合部の，２段階を経る．本稿では，ストリング生成部を高速化する一手法について報告する．

2．トーナメント技法

従来，ストリング生成部では，総合処理性能が良いことから，トーナメントソート技法を用いている．トーナメントソート技法は，一連のレコード群について，トーナメント（勝ち抜き戦）方式で比較し，最も強いレコードの選出と，作業ファイルへの出力を，繰り返し行うソート技法である．”強いレコード”とは，昇順に並べる場合は小さいキーを持つレコードのことである．

図１にトーナメントソート技法の処理手順を示す．

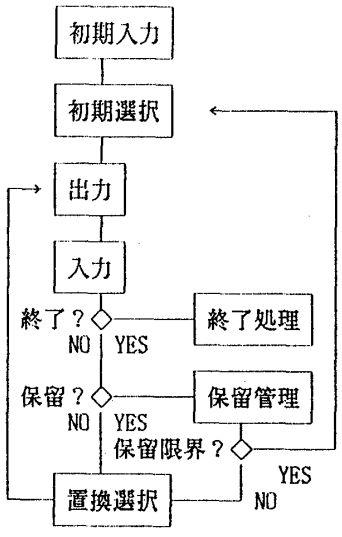

図１　トーナメントソート技法

初期入力で，トーナメントに参加させるレコードを読み込む．初期選択では，初回のトーナメントを行い，優勝レコードを選出し出力する．優勝レコード出力後，次のレコードを入力し優勝レコードと比較する．入力レコードのキーが優勝レコードのキーより弱い場合，入力レコードと残ったレコードでトーナメントを行い，新たに優勝レコードを選出する．入力レコードのキーが優勝レコードのキーと同値の場合，優勝レコードと同様に出力する．入力レコードのキーが優勝レコードのキーより強い場合，トーナメントに参加させずに保留しておく（保留レコード）．なぜなら，優勝レコードのキーより強いキーを持つレコードを，トーナメントに参加させると，ストリング内の並びが壊れるためである．

保留レコードが保留限界数（トーナメントするレコード数）になった場合，初期選択処理を含めたトーナメントを行う．

3．問題点

従来のトーナメントソート技法では，保留レコードについて以下の問題点がある．

１）保留レコードのトーナメントを行うごとに，初期選択が必要である．

２）保留レコードの管理が必要である．

レコード数が増加するにつれて，保留レコード発生率が高くなり，ストリング生成部の処理時間に，大きな影響を及ぼしている．

4．タグトーナメント技法

これらの問題点を解決すべく，各レコード単位にタグ情報を付加することにより，非保留レコードと保留レコード

A Method for Acceleration of Sort Processing (1)
Yasushi MAEJIMA, Norimasa YONEKI, Kazuaki OHKUMA, Norio OBATA
FUJITSU SHIZUOKA Engineering Ltd.
FUJITSU Ltd.

の識別を可能にし，保留レコードもトーナメントに参加できるように考案した．これをタグトーナメント技法と名付けた．タグトーナメントにおける対戦相手と勝敗は，次のように決定する．

- 非保留レコード対非保留レコード
  → キーの強弱により勝敗を決定する．
- 非保留レコード対保留レコード
  → 非保留レコードを勝ちとする．
- 保留レコード対保留レコード
  → キーの強弱により勝敗を決定する．

図２にタグトーナメント技法の処理フローを示す．従来と同様，初期入力，初期選択を行い，優勝レコードの選出と次のレコードの入力を行う．このとき，レコードに付加するタグ情報は，すべて非保留として初期化しておく．入力レコードのキーが，優勝レコードキーよりも強い場合には，“保留”を示すタグ情報を付加し，入力レコードをトーナメントに参加させて，再び優勝レコードを選出する．

非保留レコード同士，または保留レコード同士の対戦では，キーの値によって強弱が決定するため，出力するストリングは，通常の強弱順にレコードが並ぶことになる．これに対し，非保留レコードと保留レコードとの対戦では，必ず非保留レコードを勝たせ，保留レコードの値によってストリング内の，レコード順序が逆転することを防ぐ．

以上のように，保留レコードと非保留レコードを共にトーナメントに参加させることにより，保留レコード発生時の初期選択，および保留レコードの管理が不要になる．

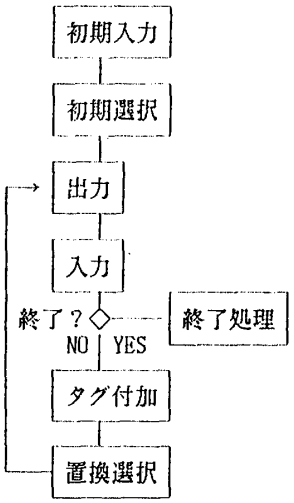

図２　タグトーナメント技法

５．効果

従来の技法では，保留レコードが発生するたびに保留管理を行い，保留レコードが限界値に達した場合，初期選択処理を繰り返す必要があった．しかし，タグトーナメント技法では，ソート開始時点の１回だけである．また，保留レコードもトーナメントに参加できるため，保留レコードの管理が不要となる．さらに副次効果として，対戦相手を決定するとき，従来のトーナメントソート技法では，対戦相手がいるかどうかの，判定が必要であったが，タグトーナメント技法では，常に相手が存在するためその必要がない．図３に，タグトーナメント技法の効果を，ＣＰＵ時間の差で表す．

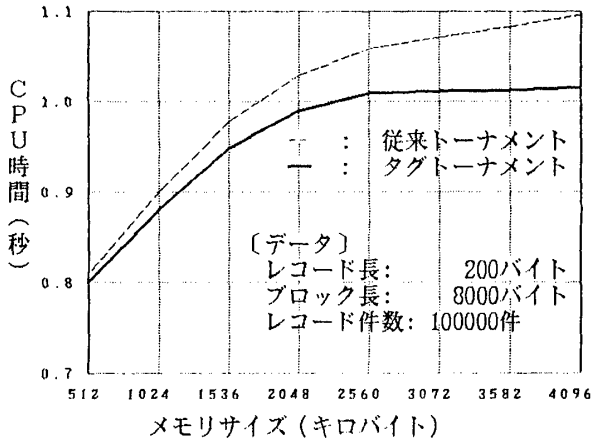

図３　トーナメントソート性能評価結果

タグトーナメントは従来のトーナメントソートに比べ，メモリサイズが多くなるほど性能差が開いている．これは，トーナメント規模が大きくなるにしたがって，一つの優勝レコードを決定するまでの比較回数が増加するため，保留処理時間や，対戦相手の存在を調べる処理時間の差が，より明確に現れてくるためである．

６．おわりに

以上ソート処理時間の性能を改善する手法について述べた．レコードにタグ情報を付加することによって，効率の良いトーナメントソートを行うことができる．