

氏名	西原清一 にし はら せい いち
学位の種類	工学博士
学位記番号	論工博第1084号
学位授与の日付	昭和53年7月24日
学位授与の要件	学位規則第5条第2項該当
学位論文題目	計算機システムにおけるハッシュ法に関する研究

論文調査委員 (主査) 教授 萩原 宏 教授 大野 豊 教授 堂下修司

論文内容の要旨

本論文は計算機システムにおいて、データの識別名、すなわちキーをもとにして、表内のデータの格納場所を決定するために用いられるハッシュ法について考察したものである。ハッシュ法は、データ・アクセスの処理時間が表内のデータの総数に直接関係しない点にその特徴があるが、従来の方法ではその効率が必ずしも良くなかったので、処理時間を従来の方法より短縮するための新しい方法を提案し、そのアルゴリズムの評価を行い、さらに、連想的データ処理へのハッシュ法の適用可能性などについて検討を加えたものであり、本文6章および付録から成っている。

第1章は序論であって、計算機システムにおけるハッシュ法の問題の位置づけを行なっている。計算機で種々の問題を処理する場合、プログラムとならんでデータ構造が重要であることを指摘し、著者の試作したフローチャート・プログラミング・システムおよび計算機動画作成システムの2システムにおいて用いられたデータ構造について論じている。データ構造の方式にはポインタを基礎にする方法と直接アクセスによる方法に大別される。このうちで、ポインタ方式は構造検索に適しており、直接アクセス方式は内容検索に向いていることを明らかにし、ハッシュ法は後者の典型的な技法であることを指摘している。

第2章では、ハッシュ法の中心的な課題である走査法について考察している。まず、オープン・ハッシュ法の検索効率の低下の原因になっているクラスタリング現象を第1種、第2種に分け、つぎに、線形走査法、2次走査法、ランダム走査法について述べ、これらの方法におけるクラスタリング発生の回避について論じ、問題点を明らかにした。ついで、クラスタリング現象を取り除く新しい方法を2つ提案し、そのアルゴリズムを与え、解析的な証明ののち、計算機実験によって確認している。また、連鎖法については、改良間接連鎖法を提案し、その効率を解析的に評価した。とくに表占有率をパラメータとした効率比較においては、癒合を含む連鎖法よりは平均探索回数が少なくなることを示している。さらに、オープン・ハッシュ法、連鎖法に加えて、新たに予測子法を提案し、そのアルゴリズムを与え、その効率を解析的に評価し、計算機実験によって確認している。また、連鎖法、予測子法においてポインタを多重化した場合の効率についても、その下限値を求め、実験により確認している。

第3章では、前章に引続き、走査法に関連した考察を行ない、検索棄却、データの削除、ハッシュ表の動的再配置、および記憶効率の4つの問題について検討を加えている。まず、検索棄却については、格納時の衝突回数を解析し、検索用に最適配置する方法について考察し、予測子法に適用した場合、その有効性が顕著であることを実験的に明らかにしている。つぎに、データの格納、削除が混在して生起するような動的な動作環境におけるデータの削除やハッシュ表のサイズの動的変更、再配置について検討し、二三のアルゴリズムを与えている。さらに、オープン・ハッシュ法と予測子法、連鎖法と予測子法の2つの組について記憶効率の比較を行っている。

第4章では、ハッシュ法についての別の基礎技術であるハッシュ関数、すなわちキー・アドレス変換の問題を検討している。まず、良いキー・アドレス変換とはどういうものかという点に注目し、その良さの基準を確率的な観点から定義し、分散係数によって定量化した。さらに、キー・アドレス変換を実用にする場合の基準として、可用変換を定義し、また、キーの構成ディジットの統計的性質を用いるディジット解析による変換法を提案している。

第5章では、ハッシュ法の内容検索指向の性質に注目し、連想的なデータ検索に応用することを試みている。連想的なデータ検索とは、検索操作などの処理の対象となるデータについて、それらの一部の特徴のみを規定することによって取り出すことであるが、連想という用語があいまいであるので、まず、連想検索という言葉を明確にし、ついで、複数キーから成るデータの集りであるファイルに対する連想的検索方法を提案し、ハッシュ技術を応用した効率の良いアルゴリズムを与えている。続いて、複数のファイルにわたって関係するデータ処理の基礎技術としての集合演算について考察し、ハッシュ表を用いて与えられた集合式を直接処理することにより、集合演算をその全要素の線形時間で実行するような手法を提案している。また、これらの手法のデータベース・システムへの適用についても検討している。

第6章は結論であって、上述の結果をまとめたものである。また、付録として FORTRAN による効率評価用シミュレーション・プログラムが付けられている。

論文審査の結果の要旨

計算機システムにおけるハッシュ法は、データの識別名、すなわちキーをもとにして、表内のデータの格納場所を決定する技法であり、データ・アクセスの処理時間が表内のデータの総数に直接関係しない点に特徴がある。本論文ではこのハッシュ法について考察を行い、従来の方法にくらべて処理の効率の良い種々の新しい方法を提案し、そのアルゴリズムの評価を行い、さらに、連想的データ処理に対する応用などについて研究を行った結果をまとめたもので、得られた主な成果はつぎの通りである。

1. オープン・ハッシュ法について検討を加え、その備えるべき条件を明らかにし、これらを満足する方法として、改良2次走査法、改良線形走査法を提案し、そのアルゴリズムを与え、解析的に証明を行い、計算機実験によって効率の確認を行った。

2. キーの格納の際の衝突の処理を行うためにポインタを用いる連鎖法について検討し、平均探索回数を減少させる方法として改良間接連鎖法を提案し、解析的な評価を行い、計算機実験によって確認し、また理論上の下限値を求めた。

3. オープン・ハッシュ法に対して平均探索回数を小さくするために予測子を用いる方法を提案し、そのアルゴリズムを与え、効率を解析的に評価し、計算機実験によってその効果を確認した。

4. 走査法に関して検討を行い、検索棄却については衝突回数を解析して検索用に最適配置する方法を考案し、予測子法に適用した場合特に有効であることを示し、また、データの削除、ハッシュ表の再配置に関する有効なアルゴリズムを与えた。

5. キー・アドレス変換法について検討を行い、キーの構成ディジットの性質を利用した変換法を提案し、京都大学大型計算機センターのシステムに適用して有効性を確認している。

6. ハッシュ法の連想的なデータ処理への応用について検討し、処理効率の良い連想的検索のアルゴリズムを与え、また、集合演算を効率良く行う方法を提案した。

以上を要するに、ハッシュ法について、その問題点を明らかにし、従来の方法の改良および新しい方法の提案を行い、その効果を明らかにし、またその応用に関する新しい知見を得て、計算機によるデータ処理の有効な手段を提供したものであり、学術上、実際上貢献するところが少なくない。

よって、本論文は工学博士の学位論文として価値あるものと認める。