

連載講座「キー検索技法」の編集にあたって

青江 順一† 榊 博 史††

コンピュータ利用における一つの大きな分野としてその高速な探索能力を用いる分野がある。われわれがこの能力の恩恵を受けている身近な例としては、銀行のオンラインシステムでのデータベースからの検索や日本語ワードプロセッサのかな漢字変換辞書の検索などがある。

キー検索は、キー(key)を“見出し”としてその内容(record)を探すことであり、表探索法であるハッシュ法、探索木法である2分探索木法、B木法などがよく知られている。キー検索の技法に関してはキーとレコードの更新が行われるか否か、2次記憶を使うか否かなどの条件に応じて非常に多くの変異があり、また種々の改善と拡張が行われてきた。最近十数年間のこの分野の論文数は千件以上にのぼる。これらのほとんどがACM, IEEE, SIAMなどの外国の論文誌に掲載されたものであり、国内での最近の研究成果は数的には少ない。本学会誌では1980年(第21巻第1, 7, 9号)と1983年(第24巻第4号)でハッシュ法と探索木法に関する解説記事が掲載されたにとどまっている。個別のシステムの設定に当たっては、これらの多くの方法の中から最適なもの、もしくはその組合せを選ぶことになる。

本連載講座では、キー検索技法をその特徴により分類し、その基本概念、最近の動向、応用分野を平易に説明するものであり、初心者の方にも十分に理解できることを目指して具体的な例も可能なかぎり多く導入する予定である。

本講座は下に示す4編からなり、このうち第3編の探索木法は、青江順一(徳島大)、佐藤隆士(大阪教育大)の共同執筆であるが他の3編は青江の単独執筆である。この講座の取り扱い分野は青江が先に編集した論文集¹⁾とはほぼ同一であるのでこれも参考にさせていただきたい。以下に各編の概要を示す。

(1) 静的ハッシュとその応用

連載の最初の講座として、まずキー検索の概観を行い、以下の講座で紹介される検索技法の相互関係について述べる。次に、ハッシュ法の基本的原理をハッシュ表での衝突現象とともに説明した後、衝突の解決法として開番地法と連鎖法を取り上げ、その基本的技術と応用分野を説明する。最後に衝突の無い種々の完全ハッシュ法を例により平易に説明し、その応用分野を紹介する。

(2) 動的ハッシュとその応用

近年注目されている2次記憶を利用する動的ハッシュ法の基本原理と種々の動的ハッシュ法の特徴を説明する。また、動的ハッシュ法は、2次記憶上のファイルの再編成技法が重要となるので、空間効率をよくするための動的ファイル管理法についても触れる。

(3) 探索木法とその応用

探索木法の原理と説明を最初に行い、続いて2分探索木法の基礎概念と、木の平衡(バランス)を実現するための多種、多彩な方法の説明を行う。次に、多分木法として代表的なB木の検索手法を概説した後、B木の多彩な拡張形式(B⁺, B*法など)の原理と特徴を整理する。

(4) トライ法とその応用

一文字単位の走査に基づくデジタル探索木法(トライ)の基本原理と特徴を説明した後、代表的データ構造に対する検索と記憶効率に関する評価を示す。またトライの応用としてかな漢字変換、形態素解析、スペルチェック辞書などを説明し、トライに関する最近の新しい改善法についても触れる。

以上が各編の概略であるが各講座には、初心者のために基礎的な著書と論文、またより深く調査を進めるための新しい論文を参考文献として示すので、今回の連載で不足する部分をおぎなっていたら幸いである。

1) Aoe, J.: Computer Algorithms—Key Search Strategies—, IEEE Computer Society Press(1991).

(平成4年6月3日)

† 徳島大学工学部
†† KDD 研究所