

Cuckoo Hashing com Perfect Rehash

Judismar Arpini Junior

Universidade Federal do Rio de Janeiro (UFRJ)
Av. Athos da Silveira Ramos, 274 - Edifício do Centro de Ciências Matemáticas e da Natureza,
Bloco D, Cidade Universitária, Rio de Janeiro - RJ
jj.arpini@gmail.com

Vinícius Gusmão Pereira de Sá

Universidade Federal do Rio de Janeiro (UFRJ)
Av. Athos da Silveira Ramos, 274 - Edifício do Centro de Ciências Matemáticas e da Natureza,
Bloco D, Cidade Universitária, Rio de Janeiro - RJ
vigusmao@dcc.ufrj.br

RESUMO

O tempo de pior caso para inserção de uma chave no esquema clássico de Cuckoo Hashing é uma variável aleatória que pode assumir valores arbitrariamente altos, dada a probabilidade estritamente positiva de se precisar fazer uma sequência interminável de realocações (*rehashes*). Apresentamos uma variante em que o tempo de pior caso da inserção é controlado. Para isso, aplicamos duas ideias: a primeira é o emprego de um método de Perfect Hashing sempre que for necessária uma realocação, fazendo com que no máximo uma realocação seja necessária por inserção; a segunda ideia é fazer com que o número de tabelas deixe de ser constante, como de praxe, e passe a ser função do número de chaves, para que a complexidade da realocação seja baixa. O preço que pagamos é piorar a complexidade de pior caso da operação de busca, que deixa de ser constante. Nosso objetivo, no entanto, é equilibrar as complexidades de pior caso da inserção e da busca, tornando ambas sub-lineares no número total de chaves.

Reportamos também a saída de alguns experimentos computacionais, em que comparamos os resultados obtidos por nosso algoritmo e por uma implementação do Cuckoo Hashing tradicional.

PALAVRAS CHAVE. cuckoo hashing, perfect hashing, estruturas de dados, complexidade
Outras aplicações em Pesquisa Operacional

ABSTRACT

The worst-case time for inserting a key in the classic Cuckoo Hashing scheme is a random variable that may assume arbitrarily big values, owing to the strictly positive probability that a neverending sequence of rehashes take place. We present a Cuckoo Hashing variation in which the worst-case insertion time is controlled. To accomplish this, we use two basic ideas: the first one is to employ a Perfect Hashing method whenever a rehash is called for, so the number of rehashes by insertion is at most one; the second idea is to make it so that the number of underlying hash tables is no longer constant, but rather a function of the number of keys, downgrading the time complexity of lookups. The price to pay is a higher lookup time, which is no longer constant. Our aim, though, is to balance the worst case complexities of the insertion and lookup operations, both of which become sub-linear in the number of keys.

We also report the outcome of some computational experiments, in which we compare the results obtained by our algorithm and by an implementation of the standard Cuckoo Hashing method.

KEYWORDS. cuckoo hashing, perfect hashing, data structures, complexity
Other applications in Operational Research