

KSI 2013/2014

Úloha 3-2: Hledání mediánu

Jan Horáček

Gymnázium, Brno, Vídeňská 47; jan.horacek@seznam.cz

4. ledna 2014

1 Teoretický popis řešení

K řešení problému nalezení mediánu využijeme 2 haldy, resp. jejich typické vlastnosti - po relativně rychlém setřepání haldy jsme schopni z nich jednoduše vyjmout největší, resp. nejmenší prvek.

Prvky budeme postupně přerozdělovat mezi 2-ma haldami tak, aby se každý prvek nacházel právě v jedné haldě podle následujících pravidel: zjednodušeně lze říct, že první halda (říkejme ji min-heap) bude obsahovat prvky momentálně menší, než medián a druhá halda (max-heap) bude obsahovat prvky momentálně větší, než medián. Zjednodušeně proto, že medián se může vlastně na "koncích" obou hald nacházet, protože může mezi být mezi prvky haldy přítomen z toho důvodu, že se numericky rovná jednomu, nebo více těmto prvkům.

Obě dvě haldy musí obsahovat stejné množství prvku s tolerancí 1 prvek. Tzn., že min-heap může obsahovat maximálně o jeden prvek více, nebo méně, než max-heap. Pokud obsahuje min-heap o jeden prvek více, než max-heap, je právě tento největší prvek z min-heap mediánem (celkový počet prvků je v tomto případě logicky lichý). Pokud obsahuje max-heap o jeden prvek více, než min-heap, je nejmenší prvek z max-heap mediánem (celkový počet prvků je opět lichý). Pokud obsahují obě dvě haldy stejný počet prvků, je medián průměrem největšího prvku min-heap a nejmenšího prvku max-heap.

2 Praktická realizace řešení

Celý algoritmus se tedy bude skládat ze dvou hald, u kterých budou implementované funkce *INSERT*, u min-heap pak funkce *GET_MAX* a u max-heap funkce *GET_MIN*, jejichž význam snad není potřeba vysvětlovat. Samotné třídění pak bude fungovat na základě algoritmu Heap sort (viz [1]).

Při vkládání prvků bude program ověřovat vyrovnanost obou hald a pokud bude zapotřebí přerozdělit prvky mezi nimi, tak to udělá. Po vložení prvku provede výpočet mediánu výše zmíněným postupem.

3 Závěr

Výběr minima a maxima z každé haldy má časovou složitost $O(N * \log N)$, celková časová složitost algoritmu je tedy $O(2N * \log N)$. Další kroky algoritmu lze, vzhledem k jejich poměrně malé časové náročnosti, zanedbat.

Na závěr bych rád zmínil, že jsem prakticky celé řešení převzal z [2].

Reference

- [1] Wikipedia
Heapsort
<http://en.wikipedia.org/wiki/Heapsort>
- [2] Stack Overflow
How to implement a Median-heap
<http://stackoverflow.com/questions/15319561/how-to-implement-a-median-heap>