

# Hashovanie

Martin Babka

22. septembra 2011

- ① Slovníky
- ② Hashovanie
- ③ Výsledky a pokračovanie

## Statické

- Napríklad perfektné hashovanie, typicky FKS schéma.
- Garantovaný čas na vyhľadávanie.
- Možná dynamizácia spolu so zachovaním  $O(1)$  času na vyhľadanie prvku.

## Dynamické

- Napríklad stromy (binárne, vyvažované, van Emde Boasove stromy), ale zaoberáme sa ...
- ... hashovaním – veľa schém, hlavne univerzálne hashovanie.
- Najlepšie metódy dávajú worst-case garanciu na vyhľadanie, problémom je update či zložitý univerzálny systém.

## Cuckoo hashing

- Použijeme dve tabuľky a dve funkcie – prvok  $x$  je na pozícii  $T_1[h_1(x)]$  alebo  $T_2[h_2(x)]$ .
- Teda  $O(1)$  vyhľadanie a mazanie. Vkladanie môže zlyhať, ale je to nepravdepodobné.
- Vyžadujú veľmi kvalitný uniformný hashovací systém –  $O(1)$  teoreticky ale pri použití už spôsobuje mierny problém.

## Two choices

- S vysokou pravdepodobnosťou dosahuje  $O(\log \log n)$  čas na vyhľadanie.
- Používa aspoň dve hashovacie funkcie a dá sa kombinovať aj s jednoduchými univerzálnymi systémami.

- *Polynómy a ich modifikácie* – jednoduché a rýchle, ale nefungujú s two-choices, otvoreným hashovaním (ak použijeme malý stupeň).
- *Lineárne funkcie* – experimenty s two-choices sú ok, ale nemáme dôkaz.
- *Tabulačné hashovanie* – kľúč rozdelíme na podreťazce a tie hashujeme nezávisle, t.j. jednoduché a rýchle. Majú výhodné teoretické vlastnosti, fungujú s two-choices, s linear probing, ale poskytujú len insertion-only kukučkové hashovanie.
- *Uniformné systémy* – výhodné teoretické vlastnosti, ale pomalé, použité s kukučkovým hashovaním.

- *Limit function* je maximálna dĺžka reťazca, ktorú povolíme.
- Ak po vložení prvku nie je dodržaná, tabuľku prehashujeme.
- Dá sa použiť pri separovaných reťazcoch, ako aj s otvoreným hashovaním.
- Extrakcia z rôznych univerzálnych systémov:
  - uniformné systémy –  $\log n / \log \log n$ , implementačne nevýhodné,
  - lineárne funkcie –  $\log n \log \log n$ , jednoduchý, ale nie úplne výhodný systém,
  - two choices –  $\log \log n$ , funguje aj s otvoreným hashovaním.

- univerzálne hashovanie a nové limit funkcie pre linear probing
- prakticky rýchle hashovanie založené na linear probing schéme alebo rýchle kukučkové hashovanie
- tabulačné hashovanie + kukučka + updaty – máme len statickú verziu
- prečo sú lineárne zobrazenia kvalitné spolu s two-choices
- ďalšie vlastnosti či vylepšenie už známych výsledkov o lineárnych funkciách