8. CVIČENÍ Z DATOVÝCH STRUKTUR 1, ZS23/24

Hešování univerzální, nezávislé, tabulační a možná i perfektní

1. *Univerzální nezávislost.*

a) Dokažte, že $(2, c)$-nezávislost implikuje c-univerzalitu.
b) Dokažte, že (k, c)-nezávislost implikuje $(k-1, c)$-nezávislost.

2. *Varianty lineární funkcí modulo p.* Na přednášce byl důkaz 2-nezávislosti systému $\mathcal{H}_{\text{lin}} := \{h_{a,b} \mid a, b \in [p]\}$ pro prvočíslo p, kde $h_{a,b}(x) := ((ax + b) \mod p) \mod m$.

a) Zdůvodněte, že $\mathcal{H}_{\text{lin}}$ není 3-nezávislá.
b) Co kdybychom měli vždy $b = 0$, tedy volili náhodně jen a? Bude výsledná rodina c-univerzální? Bude 2-nezávislá?

3. *Tabelační (tabulkové) hešování.*

a) Ukažte, že tabelační hešování je 2-nezávislé.
b) Ukažte, že tabelační hešování *není* 4-nezávislé, pokud používáme alespoň dvě tabulky.
c) Bonus: ukažte 3-nezávislost.

4. *Perfektní hešování dle Fredmana, Komlóse a Szemerédiho (FKS).* Máme dánu n-prvkovou množinu S jako podmnožinu nějakého (obrovského) univerza $\mathcal{U}$, např. 64-bitové integery. Cílem je navrhnout pro S datovou strukturu velikosti $O(n)$, která zvládne pro zadaný dotaz x zjistit, jestli x náleží v S, v konstantním čase vždy. (V čem klasická hešovací tabulka velikosti $O(n)$ nesplňuje požadavky?)

a) Připomeňte si narozeninový paradox, tedy nejmenší počet lidí s rovnoměrně náhodnými narozeninami takový, aby s pravděpodobností alespoň 50% měli dva stejné narozeniny.
b) Hešování s úplně náhodnou (nebo c-univerzální) hešovací funkcí funguje podobně. Jak zhruba musíme mít velkou tabulku, aby nastala kolize s pravděpodobností méně jak 50%?
c) Perfektní hešování vybudujeme takto: Pořídíme si hešovací funkci $h : \mathcal{U} \to [m]$ pro $m = O(n)$, kterou vybereme náhodně z c-univerzální rodiny. Pro každou přihrádku $i \in [m]$ zavedeme tabulku druhé úrovně, která bude dost velká, aby tam nenastala kolize s pravděpodobností alespoň 50% (pokud by kolize nastala, změníme hešovací funkci).
d) Teď už stačí jen omezit celkovou velikost tabulek druhé úrovně ve střední hodnotě.