

## Dvanácté cvičení ADS 1

**Příklad 1:** Je dán text rozdělený na slova. Chceme vypsat frekvenční slovník, tedy tabulku všechslov seřazených podle počtu výskytů.

**Příklad 2:** Navrhněte algoritmus, který pro daný řetězec a hodnotu  $m$  najde všechny opakované výskyty slov délky  $m$ .

**Příklad 3:** V implementaci Bucketsortu (přihrátkového třídění) může být nešikovně ukládat přihrádky jako spojové seznamy, protože spotřebujeme hodně paměti na ukazatele. Upravte Bucketsort, aby si vystačil se vstupním polem, výstupním polem a jedním neboněkolicar-prvkovými poli

**Příklad 4:** Jak reprezentovat slovník, abyste uměli rychle vyhledávat všechny přesmyčky zadaného slova?

**Příklad 5:** *Bloomův filtr* je datová struktura pro přibližnou reprezentaci množiny. Skládá se z pole bitů  $B[1..m]$  a hešovací funkce  $h$ , která prvkům univerza přiřazuje indexy  $v$  poli.  $\text{Insert}(x)$  nastaví  $B[h(x)] = 1$ .  $\text{Member}(x)$  otestuje, zda  $B[h(x)] = 1$ . Vložme nyní do filtru nějakou  $n$ -prvkovou množinu  $M$ . Pokud  $x \in M$ ,  $\text{Member}(x)$  vždy odpoví správně. Pokud se ale zeptáme na  $x \notin M$ , může se stát, že  $h(x) = h(y)$  pro nějaké  $y \in M$ , a  $\text{Member}(x)$  odpoví špatně. Spočítejte, s jakou pravděpodobností se to pro dané  $m$  a  $n$  stane.

Využijte toho, že  $1 + \alpha \leq e^\alpha$  pro každé  $\alpha \in \mathbb{R}$ .

**Příklad 6:** Spolehlivost Bloomova filtru můžeme zvýšit tak, že si pořídíme  $k$  filtrů s různými hešovacími funkcemi.  $\text{Insert}$  ude vkládat do všech,  $\text{Member}$  se zeptá všech a odpoví ano pouze tehdy, když se na tom všechny filtry shodnou. Je-li pravděpodobnost chyby jednoho filtru  $p$ , pak kombinace  $k$  filtrů chybuje s pravděpodobností pouhých  $p^k$ . Vymyslete, jak nastavit  $m$  a  $k$  pro případ, kdy chceme ukládat  $10^6$  prvků s pravděpodobností chyby nejvýše  $10^{-9}$ . Minimalizujte spotřebu paměti.

**Domácí úkol 6:** Navrhněte datovou strukturu pro básníky, která si bude pamatovat slovník a bude umět hledat rýmy. Tedy pro libovolné zadané slovo najde jiné slovo ve slovníku, kterému se zadaným co nejdélší společný suffix.

Vyřešené příklady odevzdávejte buď e-mailem se subjektem obsahujícím **ADS** na adresu [hubicka@kam.mff.cuni.cz](mailto:hubicka@kam.mff.cuni.cz) nebo na papíře na některém cvičení.