

### 3. CVIČENÍ Z ADS 2, PÁTEK 12:20

Vyhledávání více jehel v kupce sena

1. *Automat Aho-Corasickové aneb Ahomat:* Sestrojte vyhledávací automat pro slova dar, radar, adam, a.
2. *Víc jak lineárně mnoho výskytů:* Nalezněte příklad jehel a sena, v němž je asymptoticky více než lineární počet výskytů. Přesněji řečeno ukažte, že pro každé  $n$  existuje vstup, v němž je součet délek jehel a sena  $\Theta(n)$  a počet výskytů není  $\mathcal{O}(n)$ .
3. *Naivní skákání bez zkratk:* Uvažujme zjednodušený algoritmus AC, který nepoužívá zkratkové hrany a vždy projde po zpětných hranách až do kořene. Ukažte vhodnými příklady vstupů, že tento algoritmus je asymptoticky pomalejší.
4. *Frekvence výskytů:* Popište algoritmus, který v lineárním čase pro každou jehlu spočítá, kolikrát se v seně vyskytuje. Časová složitost by neměla záviset na počtu výskytů – ten, jak už víme, může být superlineární.
5. *Předpočítané množiny:* Jednoduchý způsob, jak si poradit s hlášením výskytů, je předpočítat si pro každý stav  $s$  množinu  $\text{out}(s)$  slov (jehel) k ohlášení. Dokažte, že tyto množiny není možné sestavit v lineárním čase s velikostí slovníku, protože součet jejich velikostí může být pro některé vstupy superlineární. (Poznámka: Množiny  $\text{out}(s)$  obsahují jen *indexy* jehel ve slovníku, nikoliv celá slova.)
6. *Nejčastější výskyt:* Zjistěte jaké slovo délky  $k$  se vyskytuje nejčastěji jako podslovo slova  $\sigma$ ?  
(U tohoto příkladu je chyták: časová složitost. Pokud chcete, můžete se ponořit do detailů.)

7. *Dynamické vyhledávání*: Navrhněte datovou strukturu pro dynamické vyhledávání v textu. Jehla je pevná, v seně lze průběžně měnit jednotlivé znaky a struktura odpovídá, zda se v seně právě vyskytuje jehla.

8. *Nejdelší opakovaný podřetězec*: Najděte nejdelší podslovo  $\sigma$ , které se v  $\sigma$  vyskytuje alespoň dvakrát.

9. *Největší překryv*: Mějme slovník (pro jednoduchost nad malou abecedou). Najděte v něm dvojici různých slov  $s$  a  $t$ , které se co nejvíc překrývají. Jinými slovy pro dvojici  $(s, t)$  existuje řetězec  $p$  (překryv), který je sufixem  $s$  a prefixem  $t$ , a mezi všemi dvojicemi  $(s, t)$  je  $p$  nejdelší. Pokud je řešení více, najděte libovolné.