

Zadanie č.1

15 bodov

Termín odovzdania: do 21.03.2021 do 23:59 hod.

Uloha:

Mame subor dictionary.txt. Subor obsahuje slova z nejakého anglického textu spolu s frekvenciou ich výskytu. Slova obsahujú iba malé písmena anglickej abecedy, t.j. ASCII znaky 97 až 122. Jeden riadok suboru obsahuje frekvenciu výskytu slova a samotné slovo. Frekvencia a slovo sú oddelené medzerou. Slova sú v subore usporiadané podľa frekvencie výskytu: slovo s najvyššou frekvenciou je v prvom riadku, slovo s najnižšou frekvenciou výskytu je v poslednom riadku.

Vasou úlohou je zostrojiť **optimalný** binárny vyhľadávaci strom pre vyhľadávanie slov s frekvenciou výskytu ostro väčšou ako 50 000. Ďalej budem používať termíny a notáciu z kapitoly 15.5 z knihy Introduction to Algorithms od autorov Cormen, Leiserson, Rivest a Stein . Pri vytváraní stromu postupujte nasledovne:

- kľúče budú slova s frekvenciou výskytu ostro väčšou ako 50 000.
- na slovach uvažujte lexikografické usporiadanie
- pravdepodobnosť p_i , že vyhľadávame kľúč k_i , vypočítajte ako podiel frekvencie výskytu slova k_i a súčtu frekvencií výskytu všetkých slov v dokumente dictionary.txt.
- Uvažujeme, že budeme vyhľadávať iba slova z dokumentu dictionary.txt. Pravdepodobnosť q_i , že vyhľadávame slovo, ktoré je v lexikografickom usporiadaní medzi k_i a k_{i+1} , preto vypočítajte ako podiel súčtu frekvencií výskytu tých slov z dictionary.txt, ktoré sú v lexikografickom usporiadaní medzi k_i a k_{i+1} , a súčtu frekvencií výskytu všetkých slov v dictionary.txt. Analogicky vypočítajte aj pravdepodobnosti q_0 a q_n .

Okrem toho, vytvorte funkciu **pocet_porovnaní()**. Vstupom do funkcie bude reťazec. Funkcia vráti počet porovnaní, ktoré sa vykonajú počas hľadania vstupného reťazca v zostrojenom optimálnom binárnom vyhľadávacom strome.

Odovzdávanie:

Do vytvoreného miesta odovzdania odovzdajte zdrojové súbory.

Hodnotia sa len zadania odovzdané do AISu !!!

Pre získanie bodov zo zadania je potrebné riešenie odprezentovať v termíne po dohode s cvičiacimi !!!

Hodnotenie:

15 bodov - správne vytvorený optimálny binárny vyhľadávaci strom a správne fungujúca funkcia pocet_porovnaní(). Student musí vedieť podrobne popísať postup, ktorý použil pri vytváraní stromu a funkcie.

V prípade, že študent nevie vysvetliť fungovanie svojho riešenia, riešenie sa hodnotí 0 bodmi !!!

Je možné používať implementácie z internetových zdrojov, treba však uviesť ich zdroj a rozumieť princípom, ktorý daná implementácia používa.