

1.

Při vyhledávání vzoru o délce m v textu o délce n ($n > m$) je možno vhodnou volbou algoritmu docílit v nejlepším případě až asymptotické složitosti

- a) $\Theta(m)$
- b) $O(1)$
- c) $\Theta(n)$
- d) $\Theta(n \cdot m)$
- e) $\Theta(n-m)$
- f) $\Theta(n/m)$

Boyer–Moore algoritmus po neshodě posledního znaku vzoru s odpovídajícím znakem v textu posune vzorek o m znaků dopředu. Po n/m neúspěšných porovnáních tak dosáhne konce textu. Platí tedy možnost f).

2.

Při vyhledávání vzorku AABA v textu nad abecedou $\{A,B,C,D\}$ zkonstruuje Boyer-Mooreův algoritmus (Horspoolova varianta) tabulku posunů danou vpravo. Hodnoty x y z w jsou tedy (v tomto pořadí):

- a) 1 2 3 4
- b) 2 1 4 4
- c) 1 1 2 1
- d) 4 2 1 1
- e) 1 2 3 1

A	B	C	D
x	y	z	w

V daném algoritmu se do tabulky zapisuje pro každý znak abecedy jeho vzdálenost od konce vzorku. Přitom se uvažuje jeho první výskyt v tomto pořadí a neuvažuje se 0, která by odpovídala poslednímu znaku vzorku. Všem znakům, které ve vzorku nejsou obsaženy, se přiřadí číslo rovné délce vzorku. Protože se znaky C a D ve vzorku nevyskytují, musí být v tabulce $z = 4$ a $w = 4$. Této podmínce ovšem odpovídá toliko varianta b).

3.

Při vyhledávání vzorku v textu nad abecedou $\{A,B,C,D\}$ zkonstruuje Boyer-Mooreův algoritmus (Horspoolova varianta) tabulku posunů danou vpravo. Vzorek je tedy

- a) BCAD
- b) DACB
- c) ACBA
- d) BCBC
- e) ABCD

A	B	C	D
3	1	2	4

V daném algoritmu se do tabulky zapisuje pro každý znak abecedy jeho vzdálenost od konce vzorku. Přitom se uvažuje jeho první výskyt v tomto pořadí a neuvažuje se 0, která by odpovídala poslednímu znaku vzorku. Všem znakům, které ve vzorku nejsou obsaženy, se přiřadí číslo rovné délce vzorku. Tento poslední fakt vyřazuje ze soutěže variantu a) b) a e), protože je v nich obsažen znak D, jemuž je tabulkou přiřazen posun 4, čímž se říká, že D není obsažen ve vzorku (Všechny navrhované vzorky mají totiž délku právě 4). Dále nám tabulka prozrazuje, že znak A je od konce vzorku vzdálen o 3 znaky, takže ve vzorku přítomen je, čímž odpadá také varianta d). Zbývá jen varianta c).

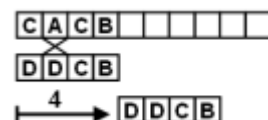
4.

Při vyhledávání vzorku v textu nad abecedou {A,B,C,D} zkonstruuje Boyer-Mooreův algoritmus (Horspoolova varianta) tabulku posunů danou níže. Vzorek je tedy

- a) BCAD
- b) DACB
- c) DDCB
- d) BCBC
- e) ABCD

A	B	C	D
4	4	1	2

Poslední prvek tabulky říká, že vzdálenost znaku D od konce vzorku je 2. Tomu odpovídá jediná z předložených variant, jmenovitě varianta c). Zbytek tabulky to potvrzuje. Vzdálenost znaku C od konce vzorku je 1. Vzorek má délku 4, takže když tabulka tvrdí, že posun pro znak A a B je 4, znamená to, že ani A ani B se nemohou vyskytovat ve vzorku na první až předposlední pozici. Na poslední pozici se např. znak B podle tabulky vyskytovat může, neboť v i takovém případě je stejně posun vzorku při neshodě roven 4 jak ukazuje obrázek.



5.

Při vyhledávání vzorku ADDB v textu nad abecedou {A,B,C,D} zkonstruuje Boyer-Mooreův algoritmus (Horspoolova varianta) tabulku posunů danou níže. Hodnoty x y z w jsou tedy (v tomto pořadí):

- a) 1 2 3 4
- b) 2 1 4 4
- c) 1 1 2 1
- d) 4 2 1 1
- e) 3 4 4 1

A	B	C	D
x	y	z	w

Ve vzorku ADDB je vzdálenost znaku A od konce vzorku rovna 3. Takže $x = 3$. To zachycuje pouze varianta e). B je na posledním místě ve vzorku, takže po jakékoli neshodě znaku B v textu a některého znaku vzorku se musí vzorek posunout o celou svou délku tj. o 4 znaky, takže v tabulce bude $y = 4$. Znak C ve vzorku není, takže bude triviálně $z = 4$ a konečně vzdálenost znaku D od konce vzorku je 1, tudíž $w = 1$.

6a.

Boyer-Mooreův algoritmus hledá v textu všechny výskyty vzorku ACBAC. Jaké hodnoty obsahuje tabulka posunů (Bad Character Shift)?

- a) 1 2 3
- b) 3 2 1
- c) 4 5 6
- d) 2 2 5
- e) 5 4 2

A	B	C

Tabulka BSC určuje jak daleko ve vzorku je ten který znak od konce vzorku, aby při neshodě (která se zjišťuje od konce vzorku) mohl být shodný znak vzorku nastaven pod aktuálně kolidující znak v textu. Tímto je již řečeno, že u A bude 1 a u B 2. u C je pak 3, protože nula

by znamenala žádný posun a tudíž žádný pokrok v hledání, trojka odkazuje na C = čtvrtý znak od konce vzorku. Platí varianta a).

6b.

Boyer Mooreův algoritmus hledá v textu všechny výskyty vzorku MLKLM. Jaké hodnoty obsahuje tabulka posunů (Bad Character Shift)?

- a) 1 3 2
- b) 5 2 1
- c) 4 5 6
- d) 2 1 4**
- e) 5 4 2

K	L	M

Úloha je analogická předchozí úloze. Tabulka BSC určuje jak daleko ve vzorku je ten který znak od konce vzorku, aby při neshodě (která se zjišťuje od konce vzorku) mohl být shodný znak vzorku nastaven pod aktuálně kolidující znak v textu. Tímto je již řečeno, že u K bude 2 a u L 1. u M je pak 4, protože nula by znamenala žádný posun a tudíž žádný pokrok v hledání, čtyřka odkazuje na M = pátý znak od konce vzorku. Platí varianta d).