

Autocomplete II

Termín odevzdání:	08.12.2019 23:59:59
Pozdní odevzdání s penalizací:	06.01.2020 23:59:59 (Penále za pozdní odevzdání: 100.0000 %)
Hodnocení:	5.0000
Max. hodnocení:	5.0000 (bez bonusů)
Odevzdaná řešení:	2 / 20 Volné pokusy + 10 Penalizované pokusy (-10 % penalizace za každé odevzdání)
Nápovědy:	0 / 2 Volné nápovědy + 2 Penalizované nápovědy (-10 % penalizace za každou nápovědu)

Úkolem je vytvořit program, který bude zobrazovat nápovědu při zadání vyhledávaného výrazu.

Chceme realizovat program, který bude implementovat část funkcionality našeptávání - autocomplete. Program si pamatuje seznam často zadávaných frází. Pro každou frázi si pamatujeme její četnost. Chceme realizovat program, který načte seznam často zadávaných frází a následně bude schopen v tomto seznamu vyhledávat pravděpodobné fráze na základě zadání jejich částí.

Vstupem programu je seznam často zadávaných frází. Tyto fráze jsou zadané v podobě:

číslo:text fráze

kde číslo je četnost dotazu (desetinné číslo) a text fráze je řetězec. Často hledaných frází je dopředu neznámý počet, jejich zadávání končí zadáním prázdného řádku. Po zadání často vyhledávaných frází následuje vlastní vyhodnocování našeptávání. Na řádce vstupu je zadán text dotazu, tento text tvoří frázi nebo její část. Texty dotazu jsou zadávané na jednotlivých řádkách, jejich zpracování skončí po dosažení konce vstupu (EOF).

Výstupem programu je počet frází, které vyhoví zadanému textu dotazu. Text dotazu se může ve frází **vyskytovat kdekoliv** (nemusí být na začátku, v tomto se zadání liší od jednodušší varianty), při porovnávání nerozlišujeme malá a velká písmena. Za výpisem počtu vyhovujících frází následuje jejich výpis v pořadí klesající četností. Zobrazeno bude nejvýše 50 frází s nejvyšší četností. Tento výstup bude zobrazen pro každý text dotazu na vstupu.

Pokud je vstup neplatný, program to musí detekovat a zobrazit chybové hlášení. Chybové hlášení zobrazujte na standardní výstup (ne na chybový výstup). Za chybu považujte:

- u zadávané fráze není uvedena četnost nebo četnost není desetinné číslo,
- chybí dvojtečka oddělující četnost a frázi,
- byl zadán nulový počet frází.

Váš program bude spouštěn v omezeném testovacím prostředí. Je omezen dobou běhu (limit je vidět v logu referenčního řešení). V povinných testech jsou zadávané rozumné dotazy (frází je málo). Pro zvládnutí povinných testů postačuje rozumná implementace naivního algoritmu. Úloha nabízí bonusový test, kde je objem zpracovávaných dat velký (fráze jsou dlouhé, je jich mnoho). Pro zvládnutí bonusového testu je potřeba použít lepší algoritmu, který dokáže rychle eliminovat neperspektivní fráze.

Ukázka práce programu:

```
Casto hledane fraze:
80:Progtest random test failed
70:Segmentation fault
40:Progtest homework #2
80:Invalid input data
50.5:Program has stopped working
15:Validator result
20:Lid is open
60:Test in progress
```

```
Hledani:
test
Nalezeno: 3
> Progtest random test failed
> Test in progress
> Progtest homework #2
test fail
Nalezeno: 1
> Progtest random test failed
lid
Nalezeno: 3
> Invalid input data
> Lid is open
> Validator result
prog
Nalezeno: 4
> Progtest random test failed
> Test in progress
> Program has stopped working
> Progtest homework #2
data
Nalezeno: 1
> Invalid input data
ISO
Nalezeno: 0
exception
Nalezeno: 0
```

```
Casto hledane fraze:
80 foo
Nespravny vstup.
```

```
Casto hledane fraze:
Progtest. Progtest never changes.
Nespravny vstup.
```

Poznámky:

- Ukázkové běhy zachycují očekávané výpisy Vašeho programu (tučné písmo) a vstupy zadané uživatelem (základní písmo). Zvýraznění tučným písmem je použité pouze zde na stránce zadání, aby byl výpis lépe čitelný. Váš program má za úkol pouze zobrazit text bez zvýrazňování (bez HTML markupu).
- Znak odřádkování (`\n`) je i za poslední řádkou výstupu (i za případným chybovým hlášením).
- Maximální délka vstupu není určena. V této úloze je potřeba paměť alokovat dynamicky.
- Maximální délka řádky není omezena. Řetězce nelze rozumně realizovat pomocí polí fixní délky.
- Pro načítání řádek textu se hodí funkce `getline` případně `fgets`. Nepoužívejte funkci `gets`.
- Při implementaci používejte prostředky jazyka C. C++ knihovna STL je pro tuto úlohu úmyslně vyřazena.
- Pro řazení se hodí knihovní funkce `qsort`.
- Při optimalizaci v bonusové části můžete předpokládat, že dotazů je mnoho, tedy vyplatí se předzpracování načtených frází. Předpočítané struktury zaberou nějakou paměť. Nemusíte mít obavy z nedostatku paměti, k dispozici je nejméně 20x větší prostor než zabírají vlastní vstupní data.
- Slovní popis struktury platných vstupních dat není zcela exaktní. Proto na výtky některých studentů přikládáme i popis vstupního jazyka v EBNF:

```
input      ::= phrase { phrase } '\n' { search }
phrase     ::= decimal ':' text '\n'
search     ::= text '\n'
text       ::= ' ' | '!' | '"' | '#' | '$' | ... | '{' | '|' | '}' | '~'
decimal    ::= [ '+' | '-' ] number [ '.' number ] [ ( 'e' | 'E' ) [ '+' | '-' ] number ] |
               [ '+' | '-' ] '.' number [ ( 'e' | 'E' ) [ '+' | '-' ] number ]
number     ::= digit { digit }
digit      ::= '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9'
```

Pozn.: text je tvořen libovolnými ASCII znaky kromě znaků řídicích.

Vzorová data:

Download

Referenční řešení

- Hodnotitel: automat**
  - Program zkompilován
  - Test 'Základní test s parametry podle ukázky': Úspěch
    - Dosaženo: 100.00 %, požadováno: 100.00 %
    - Max doba běhu: 0.006 s (limit: 2.000 s)
    - Čelková doba běhu: 0.016 s
    - Úspěch v závazném testu, hodnocení: 100.00 %
  - Test 'Test mezních hodnot': Úspěch
    - Dosaženo: 100.00 %, požadováno: 50.00 %
    - Max doba běhu: 0.195 s (limit: 5.000 s)
    - Čelková doba běhu: 0.240 s
    - Úspěch v nepovinném testu, hodnocení: 100.00 %
  - Test 'Test ošetření nesprávných vstupů': Úspěch
    - Dosaženo: 100.00 %, požadováno: 50.00 %
    - Max doba běhu: 0.006 s (limit: 1.000 s)
    - Čelková doba běhu: 0.026 s
    - Úspěch v nepovinném testu, hodnocení: 100.00 %
  - Test 'Test náhodnými daty': Úspěch
    - Dosaženo: 100.00 %, požadováno: 50.00 %
    - Max doba běhu: 0.017 s (limit: 2.000 s)
    - Čelková doba běhu: 0.111 s
    - Úspěch v nepovinném testu, hodnocení: 100.00 %
  - Test 'Test náhodnými daty + mem debugger': Úspěch
    - Dosaženo: 100.00 %, požadováno: 50.00 %
    - Max doba běhu: 0.118 s (limit: 5.000 s)
    - Čelková doba běhu: 0.283 s
    - Úspěch v nepovinném testu, hodnocení: 100.00 %
  - Test 'Bonusový test (rychlost)': Úspěch
    - Dosaženo: 100.00 %, požadováno: 100.00 %
    - Max doba běhu: 0.697 s (limit: 2.000 s)
    - Čelková doba běhu: 1.254 s
    - Úspěch v bonusovém testu, hodnocení: 125.00 %
    - Čelkové hodnocení: 125.00 % (= 1.00 \* 1.00 \* 1.00 \* 1.00 \* 1.00 \* 1.25)
  - Čelkové procentní hodnocení: 125.00 %
  - Bonus za včasné odevzdání: 0.50
  - Celkem bodů: 1.25 \* ( 5.00 + 0.50 ) = 6.88

		Celkem	Průměr	Maximum	Jméno funkce
SW metriky:	Funkce:	12	--	--	--
	Řádek kódu:	232	19.33 ± 8.22	33	speedSearch
	Cyklomatická složitost:	50	4.17 ± 1.72	7	readOnePhrase

223.11.2019 02:01:10

Download

Stav odevzdání: Ohodnoceno

Hodnocení: 5.5000

- Hodnotitel: automat**
  - Program zkompilován
  - Test 'Základní test s parametry podle ukázky': Úspěch
    - Dosaženo: 100.00 %, požadováno: 100.00 %
    - Max doba běhu: 0.006 s (limit: 2.000 s)
    - Čelková doba běhu: 0.017 s
    - Úspěch v závazném testu, hodnocení: 100.00 %
  - Test 'Test mezních hodnot': Úspěch
    - Dosaženo: 100.00 %, požadováno: 50.00 %
    - Max doba běhu: 0.011 s (limit: 5.000 s)
    - Čelková doba běhu: 0.037 s
    - Úspěch v nepovinném testu, hodnocení: 100.00 %
  - Test 'Test ošetření nesprávných vstupů': Úspěch
    - Dosaženo: 100.00 %, požadováno: 50.00 %
    - Max doba běhu: 0.004 s (limit: 1.000 s)
    - Čelková doba běhu: 0.022 s
    - Úspěch v nepovinném testu, hodnocení: 100.00 %
  - Test 'Test náhodnými daty': Úspěch
    - Dosaženo: 100.00 %, požadováno: 50.00 %
    - Max doba běhu: 0.045 s (limit: 2.000 s)
    - Čelková doba běhu: 0.166 s
    - Úspěch v nepovinném testu, hodnocení: 100.00 %
  - Test 'Test náhodnými daty + mem debugger': Úspěch
    - Dosaženo: 100.00 %, požadováno: 50.00 %
    - Max doba běhu: 0.034 s (limit: 5.000 s)
    - Čelková doba běhu: 0.212 s
    - Úspěch v nepovinném testu, hodnocení: 100.00 %
  - Test 'Bonusový test (rychlost)': Program překročil přidělenou maximální dobu běhu
    - Program násilně ukončen po: 2.001 s (limit: 2.000 s)
    - Neúspěch v bonusovém testu, hodnocení: Bonus nebude udělen
    - ☐ Překročení doby běhu
    - Čelkové hodnocení: 100.00 % (= 1.00 \* 1.00 \* 1.00 \* 1.00 \* 1.00)
  - Čelkové procentní hodnocení: 100.00 %
  - Bonus za včasné odevzdání: 0.50
  - Celkem bodů: 1.00 \* ( 5.00 + 0.50 ) = 5.50

		Celkem	Průměr	Maximum	Jméno funkce
SW metriky:	Funkce:	5	--	--	--
	Řádek kódu:	98	19.60 ± 15.63	48	loadSentences
	Cyklomatická složitost:	30	6.00 ± 4.15	13	loadSentences