

# Kontrola pravopisu



Na vstupu programu bude (na samostatných řádcích)

- počet slov ve slovníku `d` (celé číslo od 0 do 10 000),
- `d` řádků, na každém jedno slovo (dlouhé od 1 do 100 znaků) ve slovníku,
- prázdný řádek,
- počet řádků textu `l` (celé číslo od 0 do 1 000),
- `l` řádků textu, na každém několik (od 0 do 100) slov (dlouhých od 1 do 100 znaků).

Slova ve slovníku budou obsahovat písmena anglické abecedy, z nichž jenom první může – ale nemusí – být velké. Slova textu budou moci kromě písmen obsahovat také číslice a budou moci končit jednou tečkou. V textu se neobjeví samotné tečky bez předcházejícího písmena nebo čísla. Před prvním ani za posledním slovem řádku nebudou žádné mezery.

V textu platí následující pravopisná pravidla:

pravidlo	hláška při porušení
Mohou být užita pouze slova uvedená ve slovníku.	unknown word
Mohou být užita libovolná čísla (slova složená pouze z číslic).	–
Když se slovo v textu nachází na začátku věty (tj. na počátku celého textu nebo po tečce), musí začínat velkým písmenem. Pozor, platí výjimka: je-li před nějakou tečkou číslo (slovo složené pouze z číslic), nemusí to znamenat konec věty a slovo za takovou tečkou může začínat libovolně – velkým i malým písmenem.	incorrect leading lowercase
Slovo, které ve slovníku začíná velkým písmenem, musí v textu také vždy začínat velkým, ať už je na začátku věty, anebo kdekoli jinde.	incorrect leading lowercase
Slovo, které ve slovníku začíná malým písmenem, může v textu začínat velkým jenom tehdy, když je na začátku věty.	incorrect leading uppercase

Slovo, které je ve slovníku dvakrát – jednou začínající malým a podruhé velkým písmenem – se může v textu objevit kdekoli v jakékoli z těchto variant. Stále pro něj však platí pravidlo o začátku věty.	-
V řadě za sebou nesmí stát více (než jedna) mezer.	duplicate space
Číslo o čtyřech a více cifrách musí být rozdělena mezerami po tisících, milionech atd.	incorrect number formatting

Slovo, které není obsaženo ve slovníku a zároveň porušuje pravidla o velikosti písmen, bude označeno pouze hláškou „unknown word“ – vizte poslední slovo textu v příkladu.

Úkolem je vypsát zadaný text s označenými pravopisnými chybami. Chyby se označují následovně: Chybná část je podtržena vlnovkami ( ~ ). Na první znak podtržení zároveň míří šipka. Ta se skládá ze stříšky ( ^ ), pod ní několika svislítek ( | ) a pěti spojovníků ( - ), za nimiž stojí mezera a příslušná chybová hláška. Je-li na jednom řádku zvýrazněno více chyb, šipky jsou vrstveny pod sebe, přičemž na první chybu na řádku míří nejnižší šipka, na druhou chybu druhá nejnižší šipka a tak dále. Mezi vodorovnými čarami šipek je vždy jeden prázdný řádek. Pod nejnižší šipkou daného řádku je jeden prázdný řádek navíc.

## Příklad

### Vstup

```
20
this
should
be
uppercase
is
a
misspelled
word
b
and
A
incorrect
but
correct
percent
wrong
John
too
many
spaces

5
this Should be uppercase. This is a mispeled word. B is 2. and A is 1.

1 000000 is incorrect but 1 000 000 is correct. 2 500 1000percent wrong.
3. Correct 16. john. This
is too
many spaces    End.
```

## Výstup

```
this Should be uppercase. This is a mispeled word. B is 2. and A is 1.
```

```
~~~~~  
^      ^  
|      |  
|      |  
|      |  
|      |  
----- incorrect leading uppercase  
|  
----- incorrect leading lowercase
```

```
1 000000 is incorrect but 1 000 000 is correct. 2 500 100percent wrong.
```

```
~~~~~  
^  
|  
|  
|  
----- incorrect number formatting
```

```
3. Correct 16. john. This
```

```
    ^  
    |  
    ----- incorrect leading lowercase
```

```
is too  
many spaces   End.
```

```
~~~~~  
^  ^  
|  |  
|  ----- unknown word  
|  
----- duplicate space
```

Veškeré výukové materiály se nachází na <https://github.com/delta-cs/seminar>.

V případě, že vám řešení nebude uznáno a nebudete vědět proč, navštivte nejdříve [stránku s technickými pravidly](#) a ujistěte se, že program všechna splňuje.