

# Sekáček

## Informace pro uživatele

### Co je Sekáček?

Program Sekáček slouží k dělení slov českého textu na slabiky. Slova dělí tak, jak slabikuje člověk, ne tak, jak se dělí slova, která se při psaní celá nevejdou na jeden řádek. Rozdíl v tom je malý, ale je. Například není přípustné dělit na dva řádky *mo/drý*, ale vyslabikovat *mo/drý* lze.

Rozdělit text strojově na slabiky se může hodit například při výrobě zpěvníku, kde je třeba přiřadit ke každé notě jednu slabiku. Sekáček se dá využít také při skládání básniček, když chceme, aby byl ve verši určitý počet slabik. Aby se daly spočítat, musí se nejdřív vědět, kde která slabika je. A to vám řekne Sekáček.

Sekáček předpokládá, že mu uživatel zadá český text. S cizojazyčným si neporadí správně, nemůžete čekat, že ho rozdělí podle toho, jak se slabikuje například v angličtině nebo němčině. Slovo *language* rozdělí na *lan/gu/a/ge*, jako kdybyste ho vyslovili česky. (Zato ale *lengvidž* rozdělí na *len/gvidž*. :)

Pokud chcete Sekáčku říct, že Beethoven se dělí jako *Bee/tho/ven* a ne *Be/e/tho/ven*, jak by si jinak myslel, máte tu možnost. Sekáček totiž hledí na kombinace písmen a ty rozdělí podle nějakých pravidel, která zná. Ta pravidla ale nefungují na všechno. Existují slova jako bezolovnatý nebo podobojí, která se podle nich rozdělí na *be/zo/lov/na/tý* a *po/do/bo/jí*. Je třeba znát nějaké výjimky. Některé už zná, můžete mu je změnit nebo nějaké přidat sami. Více se dočtete dále v tomto textu.

### Požadavky pro spuštění

Autor zaručuje správné fungování pouze na počítači s unixovým operačním systémem, na kterém je nainstalován interpret Pythonu 3. Ten je standartně na většině distribucí. Minimální systémové požadavky pro počítač se tedy odvíjí od požadavků Pythonu 3.

### Ovládání

Sekáček komunikuje s uživatelem pouze přes příkazový řádek.

#### **Spuštění, vstup a výstup**

Spouští se příkazem `sekacek.py`. Je-li zadán bez parametrů, čeká na text k nasekání na standartním vstupu. V tom případě napište Váš text do terminálu, konce řádků odděluje Enterem, na konci stiskněte Enter a Ctrl+D. Použijte kódování UTF-8.

Výstup vypíše Sekáček opět na standartní výstup.

Za názvem programu lze zadat přepínače a nakonec název souboru nebo souborů, které Sekáček otevře a naseká jejich obsah. Potom je vypíše na standartní výstup, nemění je na místě.

#### **Přepínače**

`--help`

Vypíše nápovědu a skončí.

`-o-`, `-o -`, `--oddělovač=-`

Oddělovač slabik. Mezi slabiky uprostřed slov vkládá sekáček „/“, není-li zadáno jinak. Přepínačem `-o-` můžete jeho chování změnit, pak bude mezi každou slabiku vkládat řetězec „-“. Dočkáte se tím například výstupu `a-hoj`.

Místo `-` můžete napsat libovolný znak nebo i více znaků.

`-s+`, `-s +`, `--spojovník=+`

Spojovač slov, která tvoří jednu slabiku, například `k+le/su`. Implicitně vkládá „~“, vlnku. Podobně jako u oddělovače lze zadat i více znaků.

`-v SOUBOR [SOUBOR]...`, `--výjimky=SOUBOR`

Soubor výjimek. Otevře soubor se zadaným názvem, z něj načte slova, která má dělit jinak než pravidelně. Při volbě přepínače `-v` lze zadat i více souborů. V souboru Sekáček očekává přesný formát popsáný níže. Pokud soubor neexistuje, skončí chybou. Pokud v něm není požadovaný obsah, vypíše to na standartní chybový výstup.

Sekáček vždy otvírá soubor výjimek `.sekacek`, který se má nacházet v aktuálním adresáři.

`-t`, `--tiše`

Tichý mód. Sekáček nevypisuje žádné chybové hlášky na standartní chybový výstup. U závažných chyb ale skončí.

`-i`

Ignoruje všechna varování, která jinak způsobují konec programu, a běží dál. Chybové hlášky ale vypisuje (není-li zadáno `-t`).

## Soubory výjimek

Na každém řádku souboru výjimek Sekáček očekává dva regulární výrazy oddělené neprázdnou posloupností mezer či tabulátorů. Regulární výrazy nesmí obsahovat mezery.

Kotvy `^` a `$` znamenají začátek a konec slova, slovo je ohraničeno znaky jinými než písmena.

První výraz znamená, co se má dělit jinak, druhý výraz znamená, jak se to má dělit. Mezi hranice slabik uprostřed slov ve druhém výrazu vložte „/“. Před i za výraz Sekáček vkládá konec slabiky.

Do souborů je možno vkládat komentáře. Ty začínají znakem `%`, `#`, `"` nebo dvěma lomítky `//` a pokračují až do konce řádku.

Prázdné řádky se přeskakují.

Sekáček zpracovává soubory výjimek postupně od začátku, na již změněná slova aplikuje další změny ze souboru výjimek. Pokud napíšete do souboru výjimek například `abc def\ndef ghi\n` a na vstup dáte `abc`, dostanete `ghi`. První `abc` změní vstup na `def` a další řádek ho změní na `ghi`.

Příklad:

`odopero od/o/pe/ro`

Rozdělí odoperovat na `od/o/pe/ro/vat`, ale odoperuj tato výjimka neošetří, to rozdělí standartně `o/do/pe/ruj`

`sedm(de|náct) se/dm/\1`

Slovo sedmdesát rozdělí na `se/dm/de/sát`, sedmnáct na `se/dm/náct`, ale pozor: hledá se i uprostřed slov a za konec výjimky se vloží konec slabiky, takže ze sedmnáctý bude `se/dm/náct/ý`.

## O autorovi

Sekáček vznikl jako zápočtový program předmětu Programování I na Matematicko-fyzikální fakultě UK v Praze v zimním semestru 2013/14.

Autorem je Dominik Macháček. Kontakt na autora: [gldkslfmsd@gmail.com](mailto:gldkslfmsd@gmail.com).

Sekáček je přístupný pod licencí Creative Commons.