

IB113/03 – Domácí úkol 05

Zpracování digitálních dat

Deadline – 13.1.2021, 12.00

Pokyny

- Šestý domácí úkol odevzdejte do [odevzdáárny](#) naší seminární skupiny nejpozději do 13. ledna do pravého poledne (12.00).
- Svá řešení pište přímo do šablony, kterou naleznete ve studijních materiálech. Odevzdáváte jediný soubor pod názvem **du06.py**. V šabloně neměňte názvy funkcí ani jejich argumenty.
- Úkoly vypracovávejte samostatně. Dotazy k zadání pište veřejně do Discordového vlákna, nebo se na stejné platformě obračejte přímo na cvičícího předmětu.
- Za celý úkol můžete získat až **40 bodů**.

Rady

1. K vypracování úkolu si potřebujete ze studijní materiálů stáhnout soubory: *temperature_history.csv*, *diktat.txt*, *cameraman.png* a *cameraman_filtered.png*
2. Třetí úkol se zabývá zpracováním obrazu. To budeme procvičovat 7. ledna na 12. cvičení ještě před deadline úkolu.

1 Statistika teplot(10 bodů)

Funkce `weather_statistic` spočítá statistiku nad historickými daty o průměrné teplotě v měsíci a vytiskne letopočty s nejvyšší a nejnižší roční průměrnou teplotou ve správném pořadí.

Průměrné teploty jsou zaznamenány v souboru *temperature_history.csv*. Koncovka `.csv` (comma separated values) značí textový soubor s daty, kdy na každém řádku je jeden datový záznam, v něm jsou hodnoty odděleny symbolem `,`. Datový formát `.csv` se v Pythonu typicky zpracovává pokročilými balíčky pro zpracování dat (`pandas`), pro naši potřebu budeme data zpracovávat

po řádcích jako textový soubor. Data obsahují informaci o měsíční průměrné teplotě pro každý rok a každý měsíc v letech 1920 - 2019. Teplota byla měřena v pražském Klementinu a je zaokrouhlena na jedno desetinné místo.

Při řešení nejprve zpracujte vstupní soubor a vhodná data si uložte do slovníku. Data ze slovníku poté vyhodnoťte a vytiskněte roky, ve kterých byla naměřena průměrná nevyšší a nejnižší teplota. Tyto roky budou seřazeny vždy od extrému k průměru.

BONUS: Funkce může zpracovat i další statistiky (např. nejteplejší měsíce roku, teplotní rekordy, ...)

2 Diktát (15 bodů)

Cvičení v <https://www.umimecesky.cz/diktaty> používá jako interní reprezentaci úloh vlastní formát popsany níže, který umožňuje snadnou zpracovatelnost a kontrolu řešení. Napište program, který podle textu v tomto formátu vypíše správné řešení diktátu, tedy dosadí z možností vždy tu správnou variantu.

Formátovaný text načtete ze souboru *diktat.txt*. Místo, kde se rozhoduje o gramatice je označeno hranatými závorkami []. Před symbolem | jsou dvě možnosti k doplnění do slova, oddělené dopředným lomítkem /. Za symbolem | je klíč, který označuje, která možnost je správná. Klíč 10 značí že první, klíč 01 označuje druhou možnost jako správnou. K řešení můžete, ale nemusíte využít knihovnu pro regulární výrazy *re*. Příkladem formátovaného textu jsou následující dvě věty:

```
Zb[i/y|01]něk se rád zab[i/ý|01]vá hrou na b[i/y|10]cí.  
Cítím se jako ob[ě/j|10]ť ranního vstávání
```

Jejich správný přepis zní:

```
Zbyněk se rád zabývá hrou na bicí.  
Cítím se jako oběť ranního vstávání.
```

3 Mediánový filtr(15 bodů)

Funkce `median_filter` aplikuje na zadaný obrazek mediánový filtr velikosti 3x3 a výsledek uloží jako *xxx_filtered.png*, kde *xxx.png* je název originálního souboru. Mediánový filtr se běžně používá pro odstraňování šumu v obraze. Je to operace, která po každý pixel vezme jeho hodnotu a hodnoty okolních osmi pixelů (celkem devět čísel) a přiřadí mu hodnotu, která je mediánem této devítice, tzn. pátou nejvyšší hodnotou. Pixely při okraji obrazu nemají všech 8 sousedů. Proto je buď nevyhodnocujte, nebo zjišťujte medián pouze z existujících sousedů. Filter lze testovat na obrázku *cameraman.png*.



cameraman.png Zašuměný obraz, vstup mediánového filtru.



cameraman_filtered.png Vyčištěný obraz, očekávaný výstup mediánového filtru.