

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

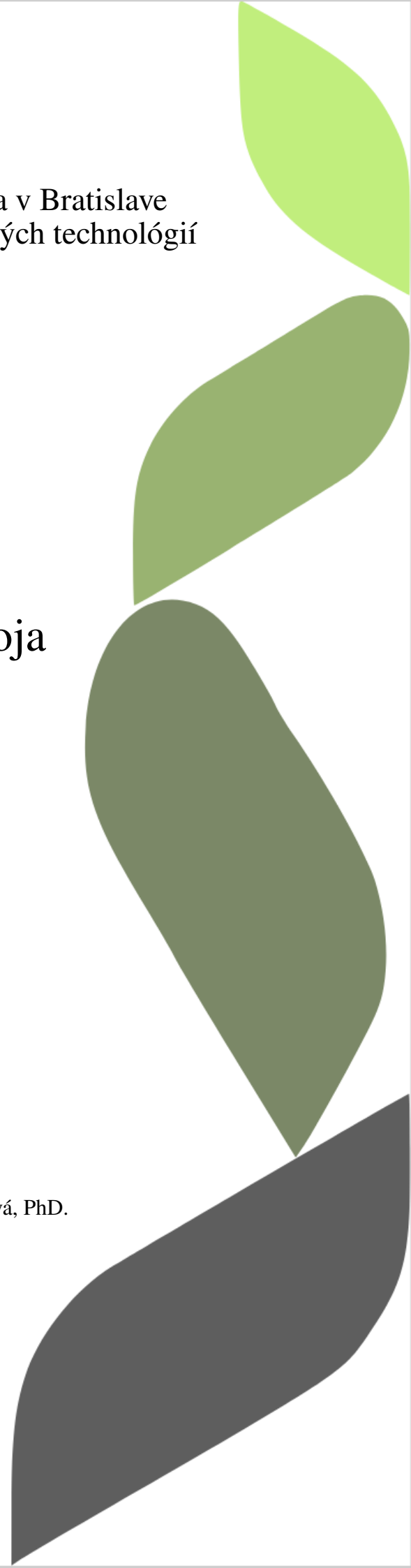
Tím č.5

Metodika vývoja v Pythone

Vedúci tímu: Mgr. Martin Sabo, PhD., Ing. Marta Prnová, PhD.

Vypracovala: Alena Valová

Kontakt: mikasa.fiit@gmail.com



1 Manipulácia s projektom

Pre zefektívnenie a zachovanie rovnorodosti práce na projekte sú v tomto dokumente definované pravidlá manipulácie s projektom.

Odporúčaným IDE pre písanie programov v jazyku Python je Pycharm. Zvoleným jazykom pre vytváranie pomenovaní je anglický jazyk.

Názvoslovie pre súbory a adresáre

Všetky priečinky budú pomenované:

- v anglickom jazyku
- s veľkým začiatočným písmenom

Súbory s koncovkou *.py* sa budú pomenovávať v závislosti od ich obsahu:

- **Triedy** - V prípade, že obsahom súboru bude definícia triedy, súbory sa pomenujú cez *camelCase*, s výnimkou veľkého začiatočného písmena. Podobne ako súbory. Napríklad trieda, ktorá definuje príspevky pre používateľov sa bude volať *UserPost*.
- **Scripty** - Ak obsahom súboru nie sú triedy, súbory budú pomenované cez *lower_case*, t.j. všetky slová v názve budú začínať malým začiatočným písmenom a budú oddelené cez podčiarkovník. Napríklad, pre funkcie upravujúce reťazce, by bol názov súboru *string_modifier*.

Adresáre v projekte Clasisfication

Pre zefektívnenie, sprehľadnenie a zjednodušenie práce na projekte budú v hlavnom projekte existovať nasledujúce adresáre:

- **Predprocessing** - všetky súbory z predspracovania.
- **FeatureSelection** - funkcie pre vytváranie črt.
- **Neurons** - neurónové siete.
- **StandardClasification** - štandardné klasifikátory.

- **Data** - adresár obsahuje všetky dáta.
 - **Test** - adresár obsahuje testovaciu vzorku dát.
 - * **Raw** - predspracované dáta.
 - * **Features** - dáta s vypočítanými črtami.
 - **Train** - adresár obsahuje trénovaciu vzorku dát a menšie vzorky
 - * **Raw** - predspracované dáta.
 - * **Features** - dáta s vypočítanými črtami, adresáre s obrázkami a diffmapami a súbormi šablóna a maska pre diff.
- **Statistics** - adresár obsahujúci štatistické údaje, grafy a iné.
 - **Charts** - adresár obsahujúci obrázky výsledkov modelov a matíc zámen.
- **UnitTests** - adresár obsahujúci unit testy.
- **Models** - adresár obsahujúci klasifikačné modely.

2 Písanie programov

Hlavička dokumentu - Importy a globálne premenné

Všetky globálne importy sa budú nachádzať na začiatku súboru. Prvý import sa umiestni na riadok č.1. V prípade že sa za importom globálne premenné:

- nachádzajú - Nasledujú za importom 1 prázdny riadok.
- nenachádzajú - Nasledujú za importom 2 prázdne riadky.

Následne budú umiestnené deklarácie globálnych premenných. Názvy globálnych premenných budú napísané veľkými písmenami, oddelených podčiarkovníkom. Napríklad globálna premenná pre maximálnu hodnotu bude označená ako *MAX_VALUE*. Príklad je uvedený na obrázku č. 1

```
import csv

MAX_VALUE = 1000000

def function:
    # some code to add
```

Obr. 1: Príklad deklarácie hlavičky súboru

Telo dokumentu

Funkcie

Medzi funkciami budú 2 voľné riadky.

V prípade, ak sa jedná o spustiteľný súbor, je povinná funkcia main v tvare:

```
if __name__ == "__main__":
```

Podmienky

Uprednostnenie *or* a *and* pred `||` a `&&`.

Pre podmienky *if*, *elif*, *else* bude blok kódu začínať vždy na novom riadku.

Cykly

Blok kódu začínať vždy na novom riadku.

Polia s kľúčovou hodnotou

Pri deklarácii polí s kľúčovou hodnotou sa odporúča nepridávať čiarku na koniec posledného riadku. Ukážka je uvedený na obrázku č. 2

```
thisdict = {  
    "brand": "Ford",  
    "model": "Mustang",  
    "year": 1964  
}
```

Obr. 2: Príklad deklarácie polí s kľúčovou hodnotou

Reťazce

Deklarácie reťazcov budú primárne začínať a končiť dvojitoú úvodzovkou ”.

Tvorba pomenovaní

Črty

Názvy črt sa skladajú z dvoch častí, ktoré sú oddelené podčiarkovníkom. Prvá časť je názov funkcie, ktorá danú črtu ráta. Druhú časť tvorí stĺpec alebo viacero stĺpcov (riadkov), nad ktorými sa daná črta ráta.

Príklad črty priemer všetkých hodnôt zo stĺpca v_1:

mean_v_1

Ak je v názve viac ako jeden stĺpec (riadok) oddeľujú sa pomocou pomlčky. V prípade názvov, kde sa nachádzajú viaceré stĺpce v kombinácii s riadkami píše sa iba číslo daného stĺpca (riadku). Ak názov funkcie tvorí viac slov spájajú sa pomocou pomlčky.

Príklad výpočtu maxima z bodu (v-s-n):

max-point_1-7-10

V prípade, že chceme pri nejakej funkcii vyrátať črtu nad všetkými stĺpcami použijeme ako názov stĺpca slovo „ALL“.

Príklad výpočtu minima nad všetkými stĺpcami:

min_ALL

Verziovane modelov

Názvy modelov konkrétnych klasifikátorov sa budú písať cez lower_case, oddelené od časovej pečiatky pomocou pomlčky. Každý model končí príponou *.joblib*.

Časová pečiatka je v tvare "RRRR-MM-DD HH-MM-SS". Rok mesiac a deň sú navzájom oddelené pomlčkou. Rok mesiac a deň sú oddelené od hodín minút a sekúnd medzerou. Hodiny minúty a sekundy sú taktiež navzájom oddelené pomlčkou.

Príklady

random_forest-2020-11-25 07-22-36.joblib

svm-2020-11-25 22-02-06.joblib

Premenné

Názvy premenných v programe sa budú písať cez lower_case. Názvy premenných by nemali obsahovať slovesá, preferované sú podstatné mená. Odporúča sa používať jednotné číslo, s malými výnimkami. Napríklad názov premennej, ktorá uchováva príspevky používateľov sa bude volať *user_posts*.

Triedy

Triedy budú pomenované cez camelCase s veľkým začiatočným písmenom. Napríklad trieda uchováajúca príspevky používateľov sa bude volať *UserPosts*.

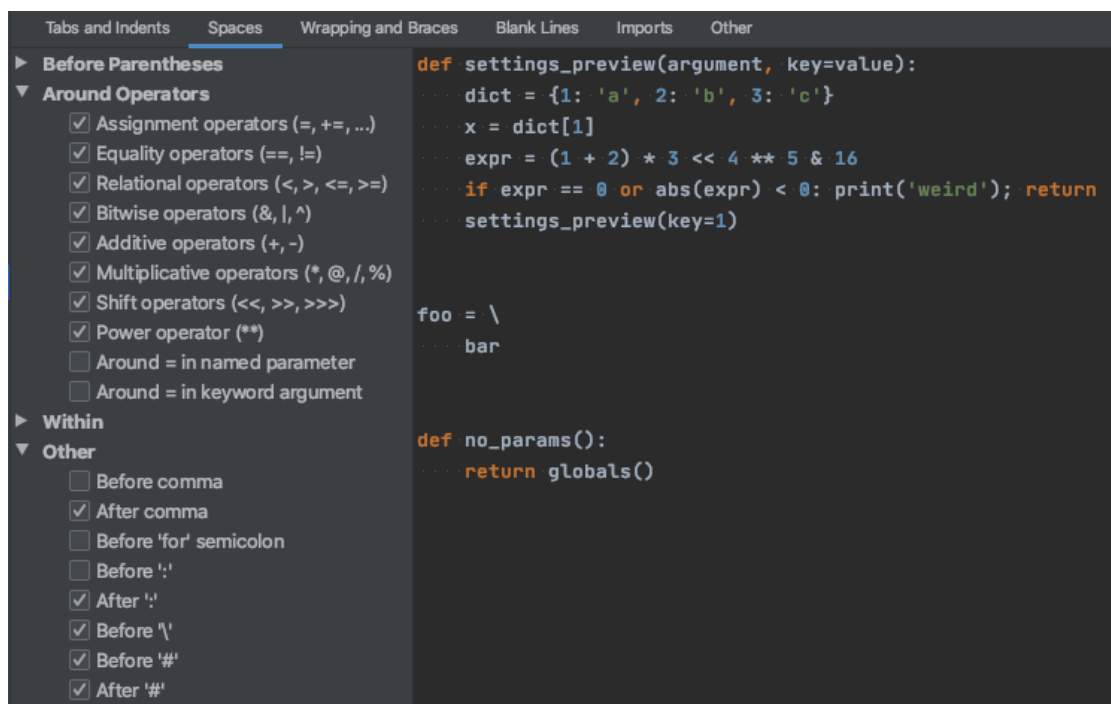
V prípade nedefinovaných pravidiel sa budeme spoliehať na štandardy PEP-8 a PEP-257. Pravidlá definované v tomto dokumente sú nadradené pravidlám týchto štandardov.

A Nastavenie IDE pre všetkých

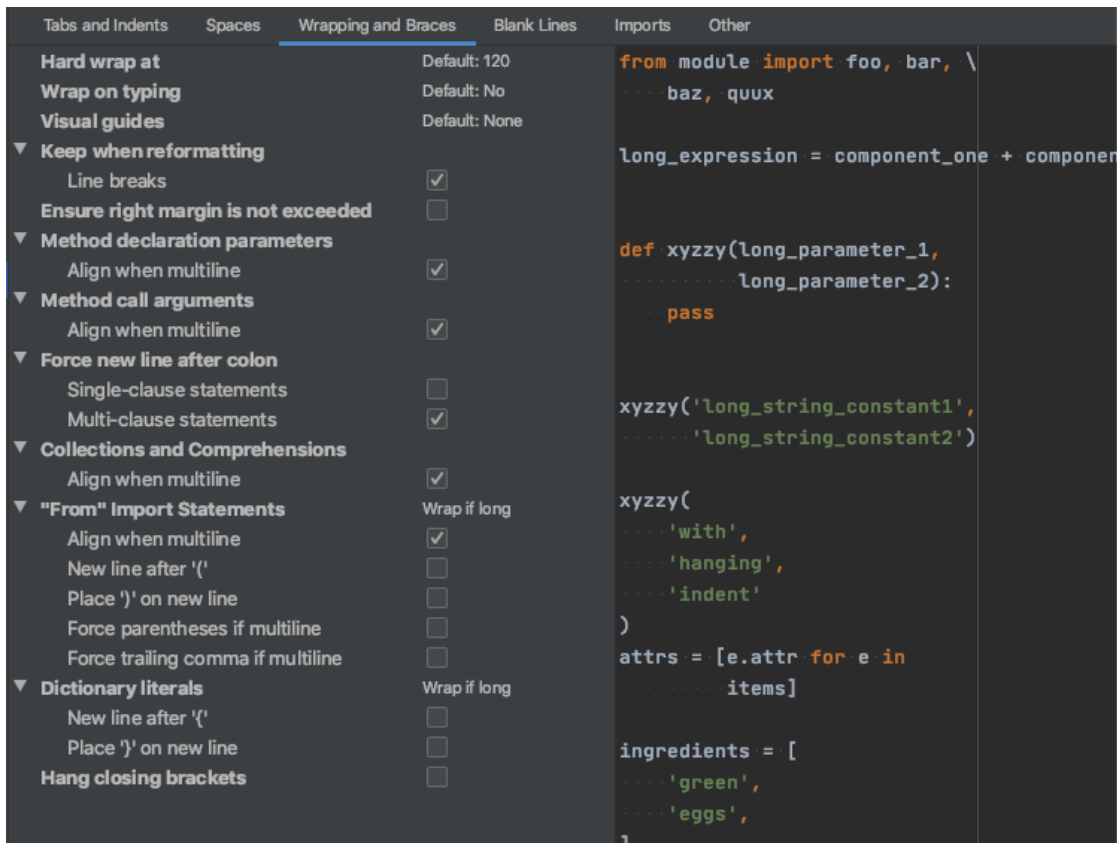
Setting/Editor/Code style/Python



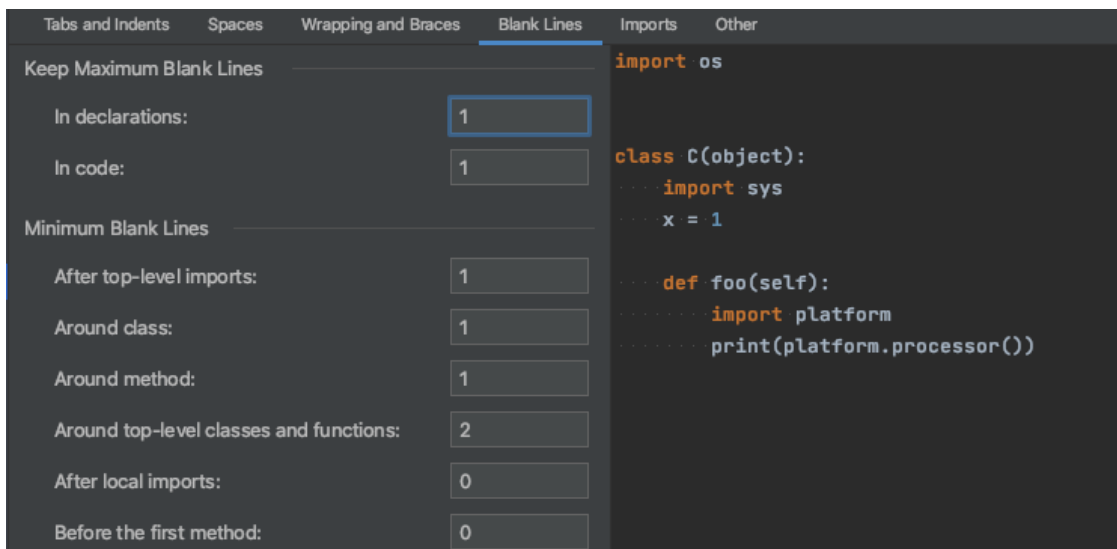
Obr. 3: Tabs and Indents



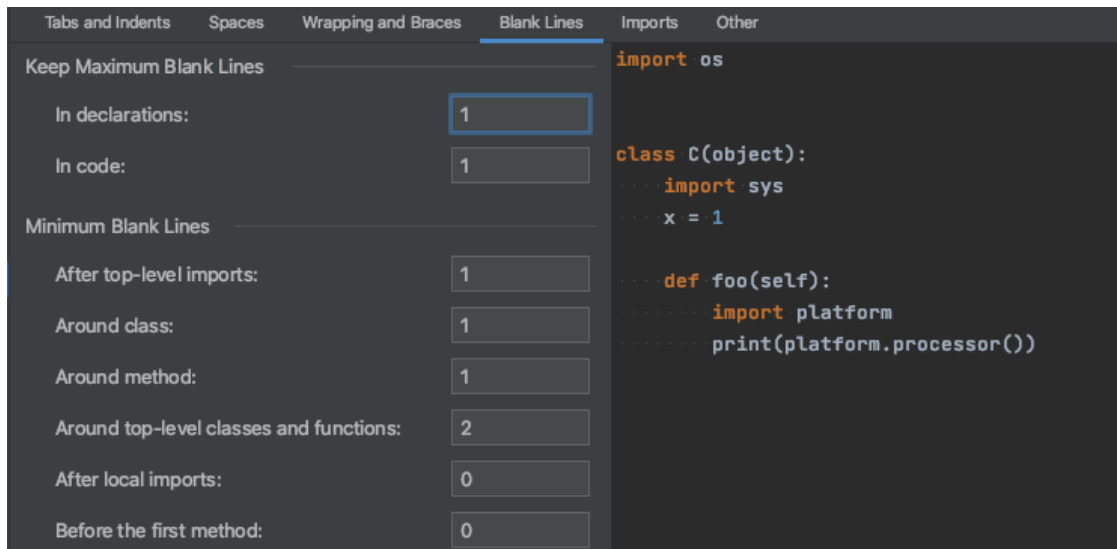
Obr. 4: Spaces



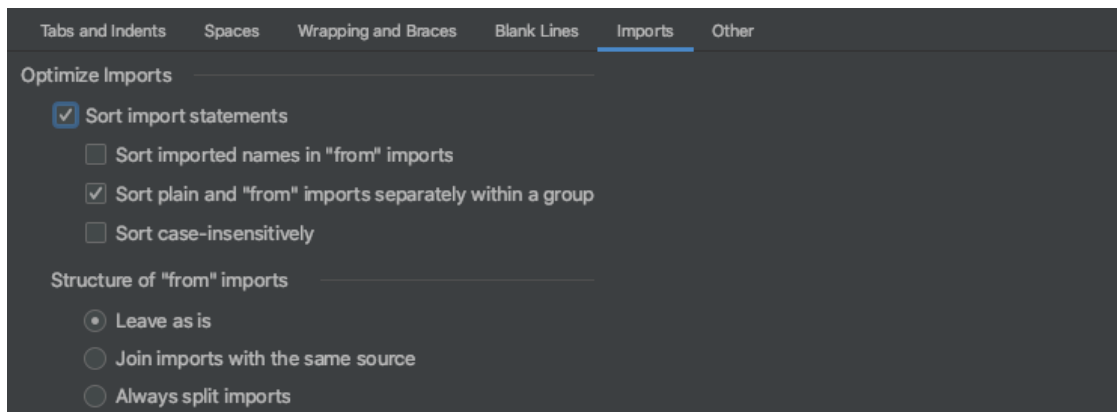
Obr. 5: Wrapping and Braces



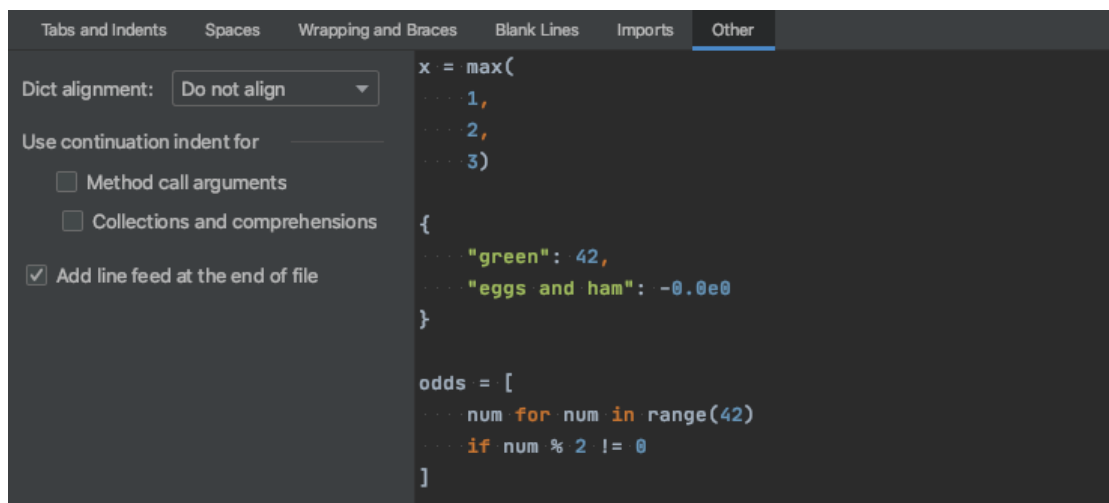
Obr. 6: Blank Lines



Obr. 7: Blank Lines



Obr. 8: Imports



Obr. 9: Other

B História

Meno a Priezvisko	Dátum	Opis zmeny
Alena Valová	17.10.2020	Vytvorenie metodiky
Mária Rajníková	24.10.2020	Pridanie adresára Data/Original
Simona Klučková	24.10.2020	Pridanie konvencie pre vytváranie názvov čít
Mária Rajníková	7.11.2020	Pridanie adresára s obrázkami a diffmapami do features
Simona Klučková	27.11.2020	Pridanie konvencie pre pomenovávanie a verziovanie modelov
Mária Rajníková	3.12.2020	Pridanie adresárov Train a Test
Jakub Kučečka	3.12.2020	Pridanie adresárov pre Aplikáciu
Zuzana Popovcová	4.2.2021	Odstránenie podpričinkov v pričinku Models, premenovanie pričinku Preprocessing → Predprocessing
Simona Klučková	18.2.2021	Odstránenie pričinkov Origins a pridanie pričinku Charts
...