

Návod na použitie

---

Konverzia Jupyter Notebooku pomocou nb2tex

---

Vygenerované z Jupyter Notebooku

pf

6. októbra 2019

# Obsah

<b>1</b>	<b>Export Jupyter Notebooku do LaTeX/u</b>	<b>2</b>
1.1	Úvod . . . . .	2
1.2	Rýchlonávod . . . . .	2
1.3	Konverzia pomocou skriptu nb2tex . . . . .	3
1.3.1	Konverzie obrázkov . . . . .	3
1.3.2	Potlačenie nekonvertovateľných častí notebooku . . . . .	4
1.3.3	Riadenie vertikálneho toku textu . . . . .	4
1.4	Inštalácia . . . . .	4
1.4.1	nb2tex . . . . .	4
1.4.2	LaTeX . . . . .	4
1.4.3	Python a Jupyter-Notebook . . . . .	4
1.5	Použitie skriptu . . . . .	4
1.5.1	Štruktúra prostredia notebooku . . . . .	4
1.5.2	Parametre skriptu, konfigurácia a spustenie . . . . .	5
1.6	Prostredie LaTeX-u . . . . .	5
1.7	Čo (zatiaľ) nefunguje ... a doporučená . . . . .	6

# Kapitola 1

## Export Jupyter Notebooku do LaTeX/u

### 1.1 Úvod

Štandardný export Jupyter-Notebooku do formátu prostredia LaTeX využíva knižnicu *pandoc*, ktorá konvertuje len základné konštrukcie formátovania textu. Rozšírenie Jupyteru o plugin LaTeX *latex.envs* umožnilo doplnenie notebookov o ďalšie konštrukty (číslovanie rovníc, lokálne odkazy a pod.), zároveň ale konvertovaný výstup v LaTeX-e vyžaduje ručné editovanie textu, navyše nie je prispôbené na konverziu dokumentov pozostávajúcich z viacerých notebookov tvoriacich kapitoly budúceho dokumentu. Z tohoto dôvodu vznikol skript *nb2tex*, ktorý čiastočne eliminuje vyššie uvedené nedostatky.

### 1.2 Rýchlonávod

- Predpokladajme, že máte vytvorené jupyter notebooky (napr. *notebook\_1.ipynb* a *notebook\_2.ipynb*) v nejakom adresári napr. **Notebooky**
- Rozbalte archív **jp2tex.zip** do nejakého adresára, napr. **Konvertor**
- Vytvorte si adresár, v ktorom sa bude vytvárať dokument, napr. **Export**
- Najskôr vygenerujeme konfiguračné súbory pre budúcu konverziu, v adresári **Export** spustíte

```
python3 $HOME/Konvertor/jp2tex.py -c
```

- V adresári sa vytvorilo niekoľko súborov s príponou **.py** a **.tex**, pomocou ktorých viete definovať titulnú stránku dokumentu, formátovanie zdrojových textov a pod. Súbory **chapters.path.tex** a **graphics.path.tex** sa prepisujú pri každej konverzii, ostatné konfiguračné súbory sa nemenia. V súbore *config.py* zadefinujeme štruktúru budúceho dokumentu, t.j. ktoré notebooky chceme exportovať a v akom poradí budú usporiadané

```
source = 'Notebooky'
```

```
chapters = [  
    ['notebook_1', True, True],  
    ['notebook_2', True, True]  
]
```

- Vlastnú konverziu spustíme v adresári **\$HOME/Export** skriptom

```
python3 $HOME/Konvertor/jp2tex.py
```

- Ak notebooky neobsahujú chyby a neznáme konštrukcie, výsledkom skriptu je vygenerovaný koreňový súbor **main.tex** a pomocné inkludované súbory, ktoré obsahujú cesty k obrázkom a jednotlivým kapitolám.

## 1.3 Konverzia pomocou skriptu nb2tex

Pretože Jupyter-Notebook sa priebežne vyvíja a aj autori rozšírení vedomí si nedostatkov priebežne svoje diela vylepšujú a dopĺňajú, nie je možné pokladať nižšie popisovaný skript za finálny a konečný produkt. Originálny export je riadený template, ktorého popis je na stránke Customizing nbconvert  
Dokument obsahuje

- Popis funkcie konverzného skriptu
- Popis usporiadania adresárovej štruktúry notebookov
- Návod na použitie a konfiguráciu konverzného skriptu
- Popis a usporiadanie prostredia LaTeX-u pre vytváranie finálnych dokumentov
- Demonštračná kapitola s použitím rozšírení LaTeX\_envs

Hlavným zámerom na vytvorenie skriptu bolo vygenerovanie takého prostredia, v ktorom sa primárny dokument edituje **len** v prostredí Jupyter-Notebooku, bez potreby ďalších alebo len minimálnych úprav v prostredí LaTeX-u. Pôvodný notebook môže obsahovať aj interaktívne časti, ktoré nie je možné principiálne konvertovať do statického dokumentu (animácie, videá a pod.)

Skript *nb2tex* (aktuálna verzia 1.3/190420) napísaný v Pythone ver.3.x konvertuje Jupyter-Notebook v nasledujúcich krokoch:

- Skopíruje jeden alebo viacero notebookov do samostatnej zvolenej adresárovej štruktúry. Do pôvodných zdrojových textov skript nezasahuje. Voľba súborov je definovaná v textovom konfiguračnom súbore.
- Pre skopírované notebooky spustí v neinteraktívnom móde obnovenie obsahu notebooku, skriptov, generovanie obrázkov skriptami (*matplotlib*) a iné. Po obnovení sú v lokálnej adresárovej štruktúre vygenerované obrázky, ktoré sú štandardnou súčasťou notebooku.
- V texte notebooku **\*.ipynb** sú pomocou regulérnych výrazov upravené a nahradené niektoré problematické konštrukcie, ako sú obrázky, vyradenie častí notebooku, ktoré nie je možné konvertovať do LaTeX-u (videá, animácie ..) a úprava vertikálneho formátovania.
- Konverzia textu notebooku na LaTeX je pomocou štandardných konverzných nástrojov *nbconvert* a *pandoc*
- Z konvertovaného textu v LaTeX-e sú odstránené časti, ktoré by bránili poskladaniu viacerých dokumentov do jedného celku. Úvodná stránka a členenie dokumentu je definované v titulnom súbore LaTeX-u.

### 1.3.1 Konverzie obrázkov

Štandardný notebook umožňuje vloženie obrázkov bez popisu (pomocou HTML tagu `<img>`), pri ktorých je možné nastaviť šírku v px, táto informácia nie je ale konvertovaná do LaTeX-u. Na druhej strane rozšírenie LaTeX-envs umožňuje vkladanie obrázkov s popisom a veľkosťou pomocou prostredia LaTeX-u `figure{}`, informáciu o veľkosti obrázku, ale interpreter notebooku toto ignoruje.

Pre riešenie tohoto problému skript nahradí pred konverziou HTML tag konštrukciou LaTeX-u, z tohoto dôvodu je tag rozšírený o neštandardný (ignorovaný) textový parameter *scale* a HTML tag s definíciou obrázka má potom tvar (dva apostrofy v príklade reprezentujú uvozovky)

```
<img src='../img/map01.png' width=500px alt="Skalárne pole" scale="0.4"/>
```

je konvertovaný na LaTeX-ový popis obrázku nahradzujúci pôvodné vloženie obrázku

```
\begin{figure}[h!][h!]
```

```
\centerline{\includegraphics[scale=0.4]{../img/map01.png}}
```

```
\caption{Skalárne pole}
```

### 1.3.2 Potlačenie nekonvertovateľných častí notebooku

Pre zabránenie chybám pri konverzii vložených animácií, videí a pod. umožňuje skript vyradiť z konverzie riadok, pred ktorým je vložený komentár s príkazom REMOVE

```
<!-- REMOVE -->
```

Odstráni nasledujúci riadok po príkaze REMOVE

### 1.3.3 Riadenie vertikálneho toku textu

Pomocou príkazov vložených v komentároch je možné upraviť vertikálne formátovanie textu. Príkazy sú nahradené štandardnými makrami LaTeX-u.

```
\smallskip  
\medskip  
\bigskip  
\newpage
```

## 1.4 Inštalácia

### 1.4.1 nb2tex

Pre konverziu je použitý skript *nb2tex* pre Python ver. 3.x. Skript sa neinštaluje, používa sa ako akýkoľvek iný skript v OS Linux.

### 1.4.2 LaTeX

Pre správnu konverziu notebooku je potrebné mať nainštalované základné balíky LaTeX-u, fonty a podporu slovenčiny

```
sudo apt-get install texlive texlive-generic-recommended texlive-latex-extra  
sudo apt-get install texlive-lang-czechslovak
```

tieto obsahujú len základnú podmnožinu z veľkého počtu balíkov dostupných v LaTeX-e. Pre prípadné úpravy finálneho textu je vhodné mať nainštalované aj vhodné prostredie, napr. TexMaker.

### 1.4.3 Python a Jupyter-Notebook

Okrem štandardného Jupyter-Notebooku je potrebné nainštalovať aj rozšírenie *LaTeX\_envs* a pomocné konverzné utility, knižnicu *pandoc* pre konverziu textových formátov, *nbconvert* a *jupyter\_latex\_envs* pre export notebooku

```
sudo apt-get install pandoc  
sudo apt-get install python-setuptools  
  
pip install wheel  
pip install nbconvert  
pip install jupyter_contrib_nbextensions  
  
jupyter contrib nbextension install --user
```

Podrobnejšie je inštalácia rozšírení a ich aktivovanie v prostredí Jupyter popísané v Kapitole 2. V prípade, že používate ako default verziu Python3, použite miesto *pip* alternatívu *pip3*.

## 1.5 Použitie skriptu

### 1.5.1 Štruktúra prostredia notebooku

Adresárovú štruktúru, v ktorej vytvárame notebooky, je vhodné usporiadať v nasledujúcom tvare:

```

<Source>
|
|   +--- notebook_01.ipynb      subory notebookov, jednotlivé kapitoly a pod.
|   +--- notebook_02.ipynb
|   +--- ...
|   |
|   +--- <img>                  adresar so statickymi (negerovanymi)
|       |                      obrazkami pouzitymi v notebookoch
|       +--- *.png             pmocou prikazov <img src ... >

```

## 1.5.2 Parametre skriptu, konfigurácia a spustenie

Skript je potrebné spustiť s nasledujúcimi parametrami

usage: nb2tex.py [-h] [-c]

kde

-h optional, zobrazenie helpu  
-c optional, vygenerovanie konfiguracnych suborov

Konfiguračný súbor **config.py** definuje notebooky, ktoré budú zaradené do konverzie. Formát konfiguračného súboru:

```

source = 'cesta_k_notebookom'

chapters = [
    # notebook      COPY    UPDATE
    ['meno_notebooku', True, True],
    [ .... ],
]

# význam flagov
# COPY - ak True, notebook sa bude kopirovať, inak sa použije lokálna kópia
# UPDATE - ak True, vykona sa refresh celého notebooku (spustenie skriptov,
#          generovanie obrázkov z matplotlibu)
#          voľbu False je možné zvoliť len pri notebookoch, ktoré neobsahujú
#          interaktívne časti

```

## 1.6 Prostredie LaTeX-u

Výsledkom konverzie je sada súborov **\*tex** v adresároch pomenovaných podľa notebookov. Štruktúra prostredia pre generovanie finálnych dokumentov má tvar (doporučené)

```

<Moja_Kniha>
|
|   |
|   |   +--- <meno_notebooku_01>      generovaný adresar
|   |   |
|   |   +--- notebook_01.ipynb        kópia originálneho notebooku
|   |   +--- notebook_01.tex
|   |   |
|   |   +--- <meno_notebooku_01>
|   |   |
|   |   +--- *.png                    skriptom generované obrázky (UPDATE)
|   |
|   +--- <img>                        kópia adresaru so statickymi (negerovanymi)
|       |                            obrazkami pouzitymi v notebookoch

```

```

|      +-- *.png
|
+--- main.tex          konfiguracia dokumentu, usporiadanie a poradie kapitol
+--- listing.tex       format zobrazenia vypisov programov
+--- title_page.tex    titulna stranka dokumentu
+--- chapters_path.tex cesty ku kapitolam (needitovat !)
+--- graphics_path.tex cestu k obrazkom (needitovat !)

```

V každom vygenerovanom adresári je podadresár vytváraný pri UPDATE notebooku, ktorý obsahuje súbory s obrázkami vygenerovanými skriptami notebookov. Pre spracovanie v systéme LaTeX-u treba jednotlivé súbory usporiadať do finálneho dokumentu a nastaviť cesty ku generovaným a vkladným obrázkom, príslušné konfiguračné súbory vytvára skript automaticky.

```

% konfiguracia dokumentu, pouzite prostredia, slovenscina

\documentclass[10pt,a4paper]{report}
\usepackage[utf8]{inputenc}
\usepackage[slovak]{babel}
\usepackage[IL2]{fontenc}
...

%importovanie konfiguracneho suboru pre formatovanie vypisu programov
\include{listing}

% cesty k obrazkom
\include{graphics_path}

% struktura dokumentu
\begin{document}

% titulna stranka
\include{title_page}

% obsah
\newpage

% kapitoly dokumentu
\include{chapters_path}

\end{document}

```

Výsledný dokument vytvárame obvyklým spôsobom, z prostredia editora LaTeX-u alebo skriptom.

## 1.7 Čo (zatiaľ) nefunguje ... a doporučenia

V automatickom móde generovania LaTeXu zatiaľ nefunguje

- Matematické výrazy v popise obrázkov
- Tabulky - je v tom zatiaľ chaos ... asi vhodnejšie je tabulky vytvoriť vo vlastnom \*.tex súbore a vložiť ich do finálneho dokumentu cez *include*
- Odkazy medzi medzi kapitolami na vzorce a pod.
- Variabilné nastavovanie veľkosti generovaných obrázkov, nastavenie veľkosti pomocou *figsize=()* sa neprenáša do LaTeXu
- Použitie italic textu v nadpisoch generuje chybu

Parser Jupyter-u sa v niektorých prípadoch mierne odlišuje od LaTeX-u, prípadne je tolerantnejší k nedodržaniu pravidiel LaTeX-u. Pri tvorbe notebookov je vhodné dodržiavať niektoré odporúčenia

- Vkladané obrázky v \*.png formáte generovať v jednom rozlíšení (napr. 600dpi)
- Pri vkladaní matematických výrazov do textu nevkladať medzery za a pred znakom koncového znaku dolára.