

## NAKI-II-USTR-UKONCENE - Task #3889

### Vybrat triedy dokumentov pre ucely klasifikacie

29.04.2016 14:38 - Hrúz Marek

<b>Status:</b>	Closed	<b>Start date:</b>	29.04.2016
<b>Priority:</b>	Urgent	<b>Due date:</b>	
<b>Assignee:</b>	Hrúz Marek	<b>% Done:</b>	50%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Description</b>			
Pre ucely projektu bude vhodne klasifikovat dokumenty, aby sme vedeli ako ich dalej spracovat.			
Triedy, do ktorych sa bude klasifikovat nie su zatiaľ urcene a vramci tejto ulohy budu dane triedy identifikovane. Dokumenty sa musia prezriet manualne.			

#### History

##### #1 - 03.10.2016 16:11 - Zajíc Zbyněk

udělat supervised shlukování

- potřeba supervised množina dokumentů (zkusit natrénovat v prvním nástřelu unsupervised třídy a zkouknout je ručně) aby poté bylo možné dokumnty zpracovávat OCR s nějakou přidanou znalostí o dokumentu

##### #2 - 02.02.2017 14:06 - Hrúz Marek

- File document\_types.txt added

- % Done changed from 0 to 90

Vybrana množina tried pre klasifikáciu (viz. document\_types.txt)

V budúcnosti sa možno rozšíri.

##### #3 - 06.02.2017 18:00 - Zajíc Zbyněk

- Subject changed from Vybrat triedy dokumentov pre ucely klasifikacie to Vybrat triedy dokumentov pre ucely klasifikacie - natrénovat supervised klasifikátor

- % Done changed from 90 to 50

Ze schůzky [[<https://wiky.zcu.cz/redmine/projects/naki-ii-ustr/wiki/Schuzka16-10-03>]]:

- LBU unsupervised klasifikace do 3 tříd (předpoklad: hand/printed/photo) - Kmeans z příznaků z poslední vrstvy CNN (Alex net) - nefunguje
- co funguje - hledání nejbližšího dokumentu ke zvolenému
- MH určit ručně analýzou dokumentů 11 tříd
- **Dodělat** -> najít reprezentativní obrázky k daným 11 třídám a jimi natrénovat supervised klasifikátor

##### #4 - 22.10.2018 16:08 - Zajíc Zbyněk

- Subject changed from Vybrat triedy dokumentov pre ucely klasifikacie - natrénovat supervised klasifikátor to Vybrat triedy dokumentov pre ucely klasifikacie

- Priority changed from Normal to Urgent

##### #5 - 22.10.2018 16:20 - Zajíc Zbyněk

Připravit návod na anotaci

##### #6 - 18.11.2019 09:08 - Zajíc Zbyněk

- Status changed from Assigned to Closed

#### Files

document_types.txt	2.86 KB	02.02.2017	Hrúz Marek
--------------------	---------	------------	------------