

## NAKI-II-USTR-UKONCENE - Task #3957

Task # 3633 (Closed): Etapa 01 - Příprava dat a datových struktur, testy existujících metod

### OCR - čtení hlavičky/nadpisu dokumentu

27.06.2016 14:50 - Zajíc Zbyněk

<b>Status:</b>	Closed	<b>Start date:</b>	27.06.2016
<b>Priority:</b>	Normal	<b>Due date:</b>	31.10.2016
<b>Assignee:</b>	Neduchal Petr	<b>% Done:</b>	100%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Description</b>			
k určení tématu/klasifikace dokumentu			

#### History

##### #1 - 29.06.2016 13:46 - Neduchal Petr

- Due date changed from 30.09.2016 to 31.08.2016
- Assignee changed from Hrúz Marek to Neduchal Petr

analýza lokace, velikosti textového bloku...  
čtení a zjištění co v tom je.

##### #2 - 03.10.2016 15:48 - Zajíc Zbyněk

- Due date changed from 31.08.2016 to 31.10.2016

##### #3 - 12.10.2016 16:16 - Neduchal Petr

- % Done changed from 0 to 60

Provedena řada pokusů založených výhradně na vlastnostech obrazu (velikost textového bloku, počet znaků v textovém bloku, druh znaků, atp). Při kvalitním rozpoznání většinou dobré výsledky. Problém nastává při nerozpoznání bloku (spojení s jiným blokem) či špatném rozpoznání textu. To se dá odstranit lepším rozpoznáváním, lepším předzpracováním (také řešíme).

Pravděpodobně bude také potřeba projít dokumenty a vytvořit seznam typů dokumentů. Na to zkusím vymyslet metodiku tak aby se do toho mohlo pustit více lidí a nedocházelo k zavádění různých tagů pro jeden druh dokumentu.

##### #4 - 19.11.2018 07:55 - Neduchal Petr

- Status changed from Assigned to Resolved
- % Done changed from 60 to 100

Resolved --> Task je díky velkému rozptylu ve vzhledu dokumentů těžko vyřešitelný. Možná by v něm šly využít anotace tagů, které se před nedávnem dělaly. Stejně by ale zůstal problém velkých rozdílů ve vzhledu dokumentů

##### #5 - 19.11.2018 08:52 - Zajíc Zbyněk

- Status changed from Resolved to Closed