

NAKI-II-USTR-UKONCENE - Task #4482

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

Task # 3803 (Closed): OCR - jazykový model v Tesseractu

Balík OCR

22.10.2018 16:17 - Zajíc Zbyněk

| | | | |
|---|---------------|------------------------|------------|
| Status: | Closed | Start date: | 22.10.2018 |
| Priority: | High | Due date: | 05.11.2018 |
| Assignee: | Neduchal Petr | % Done: | 80% |
| Category: | | Estimated time: | 0.00 hour |
| Target version: | | | |
| Description | | | |
| Přípravit (v Pythonu?) balík zpracování obrazu s předzpracováním a vyhodnocením zatím jen pomocí Tesseractu (bez Kaldi), 1Best hypoteza + BoundingBox | | | |

History

#1 - 05.11.2018 13:50 - Neduchal Petr

- % Done changed from 0 to 80

Připravený balík obsahující cpp aplikaci využívající api Tesseractu + python balíček předán Honzovi Švecovi.

Zjištění: Tesseract 4 má lepší výsledky než předchozí verze, ovšem v současné fázi vývoje z nějakého důvodu vrací pouze one_best nikoliv alternativy i přesto že to přes API přímo požadujeme.

Předzpracování se musí ještě upravit z toho co měl Dan tak aby se spouštělo přímo v pythonu a ne z shellu --> Dodělám v tomto týdnu

#2 - 18.11.2019 09:08 - Zajíc Zbyněk

- Status changed from Assigned to Closed