

ZÁPADOČESKÁ UNIVERZITA V PLZNI

Fakulta aplikovaných věd
Akademický rok: 2020/2021

ZADÁNÍ BAKALÁŘSKÉ PRÁCE (projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Matěj ZEMAN**
Osobní číslo: **A18B0360P**
Studijní program: **B3902 Inženýrská informatika**
Studijní obor: **Informatika**
Téma práce: **Automatické rozpoznávání dialogových aktů v tištěných textech**
Zadávající katedra: **Katedra informatiky a výpočetní techniky**

Zásady pro vypracování

1. Seznamte se s dodanými datovými kolekcemi pro automatické rozpoznávání dialogových aktů (DA) v textové podobě.
2. Navrhněte a implementujte nástroj, který z dodaných textových kolekcí vytvoří analogicky anotované datové sady složené z množiny obrazových dokumentů.
3. Seznamte se se systémem optického rozpoznávání znaků (OCR) Tesseract.
4. Prostudujte relevantní metody automatického rozpoznávání dialogových aktů založené na neuronových sítích.
5. Na základě předchozí studie navrhněte a implementujte prototyp systému pro automatické rozpoznávání dialogových aktů z obrazových dokumentů. U vytvořeného prototypu použijte k převodu obrazu na text systém Tesseract.
6. Funkčnost prototypu otestujte na vytvořených datových kolekcích.
7. Výsledky diskutujte a navrhněte další možná rozšíření.

Rozsah bakalářské práce: **doporuč. 30 s. původního textu**
Rozsah grafických prací: **dle potřeby**
Forma zpracování bakalářské práce: **tištěná**

Seznam doporučené literatury:

Dodá vedoucí bakalářské práce.

Vedoucí bakalářské práce: **Doc. Ing. Pavel Král, Ph.D.**
Katedra informatiky a výpočetní techniky

Datum zadání bakalářské práce: **5. října 2020**
Termín odevzdání bakalářské práce: **6. května 2021**

L.S.

Doc. Dr. Ing. Vlasta Radová
děkanka

Doc. Ing. Přemysl Brada, MSc., Ph.D.
vedoucí katedry