

Název práce: Normalizace čísel pro výslovnost syntézou řeči

Autor: Jakub Růžička

Ústav: Ústav formální a aplikované lingvistiky

Vedoucí bakalářské práce: Mgr. et Mgr. Ondřej Dušek, Ph.D., Ústav formální a aplikované lingvistiky

**Abstrakt:** Nezbytnou součástí každého systému pro syntézu řeči je normalizace slov obsahujících číslice. Přesnost tohoto procesu může významně ovlivnit kvalitu výsledné promluvy. Hlavním cílem této práce je návrh a implementace modulu pro normalizaci číslic v češtině. Slova obsahující číslice jsou nejprve zařazena do jedné z předem definovaných kategorií. Na jejímž základě jsou následně generovány možné rozepsané podoby. Pro výběr kontextově správné varianty je využit existující jazykový model. Systém je distribuován jako balíček v jazyce Python. Může běžet na systému Linux nebo v Docker kontejneru, jehož konfigurace je součástí projektu. Současně byla navržena a napsána specializovaná aplikace pro anotaci dat, která slouží k vytváření datových sad pro normalizaci textu. Pomocí aplikace byly vytvořeny dvě datové sady s 1.882 větami a 3.185 slovy vyžadujícími normalizaci. Systém dosáhl na obou získaných datasetech přesnosti přes 80 % (měřeno na úrovni věty). Na základě analýzy chyb popisujeme možná zlepšení.

**Klíčová slova:** normalizace českého textu, normalizace čísel, systém pro syntézu řeči, vážený konečný převodník