

V tejto bakalárskej práci skúmame možnosti generovania textu v prirodzenom jazyku podmieneného štruktúrovanými vstupnými dátami. Konkrétne trénujeme hlboké neurónové siete na datasete RotoWire, ktorý obsahuje štatistické údaje o basketbalovom zápase spárované s ich súhrnmi. Najprv analyzujeme dataset a navrhujeme niekoľko metód preprocessingu (napríklad Byte Pair Encoding). Následne trénujeme základný model založený na architektúre Encoder-Decoder na pripravenom datasete. Popíšeme niekoľko problémov základného modelu a skúmame niekoľko pokročilých neurónových architektúr, ktoré majú tieto problémy riešiť (Copy Attention, Content Selection, Content Planning). Naša analýza naznačuje, že modely nedokážu pochopiť štruktúru vstupných dát. Navrhujeme preto metódu na zjednodušenie jej zložitosti. Náš najlepší model vytrénovaný na zjednodušených dátach prekonáva základný model o viac ako 5 bodov BLEU.