

Abstrakt

Cílem práce bylo použití hlubokého učení k rozpoznávání učebnic a jejich klasifikaci do vyučovacích předmětů a úrovní, a to na základě textových údajů, jako je název knihy, autor, nakladatel a stručný popis obsahu. Jako součást práce formulujeme vlastní definici učebnice, vytváříme dataset pomocí extrakce údajů ze zdrojů dostupných na internetu a ručně anotujeme trénovací a testovací množinu dat. Pro klasifikaci používáme jako baseline naivní bayesovský klasifikátor, z neuronových sítí pak konvoluční a rekurentní architekturu i jejich kombinace. Porovnáváme také různé způsoby reprezentace dat vektory (tzv. word embedding) a dosažené výsledky podrobně analyzujeme. Výsledný nejlepší model dosahuje na testovacích datech u všech tří úloh vysokých úspěšností a nabízí se tak možnost jeho uplatnění v praxi.