

Automatická korekce textu (natural language correction) je obor zabývající se vytvářením systémů pro automatickou opravu lidmi psaných textů. Tento obor zahrnuje mimo jiné obecnou opravu gramatiky, opravu překlepů či doplnění diakritiky. V posledních letech jsme byli svědky velkého rozmachu tohoto oboru, objevily se nové modely pro korekci textu, datasety a také vyhodnocovací metriky. Tato disertace prezentuje formou souboru publikovaných prací naše příspěvky do této oblasti. Jelikož je čeština mateřským jazykem autora této práce, věnovali jsme velké úsilí zlepšování tohoto oboru v češtině. Mezi hlavní výstupy naší práce patří: (1) vytvoření velké datové sady pojmenované Grammar Error Correction Corpus for Czech, která obsahuje anotované texty psané různými typy uživatelů jako třeba eseje či příspěvky ve webových diskuzích, a zároveň natrénování a evaluaci automatických systémů založených na neuronových sítích a také provedení meta-evaluace automatických metrik, (2) vytvoření systémů pro automatickou korekci textu, které fungují dobře v situacích, kdy je k dispozici pouze malé množství anotovaných dat pro řízené učení a (3) vytvoření dvou systémů pro automatickou diakritizaci textu dosahujících nejlepších známých výsledků a také vytvoření velké datové sady pro učení a vyhodnocování systémů pro automatickou diakritizaci textu ve 12 jazycích.