

Propojování entit je úloha, ve které jsou zmínky z textu propojovány s příslušnými entitami. Stejně jako v mnoha jiných oblastech zpracování přirozeného jazyka se i v propojování entit výrazně projevily vliv hlubokého učení, což vedlo k významnému zlepšení výkonu. V současnosti se ale stávající modely trénují pomalu a spoléhají na nejednotné zdroje dat, což ve výsledku komplikuje reprodukovatelnost. V této práci vyvíjíme několik systémů, které se učí rychle, čímž ukazujeme, že konkurenceschopných výsledků lze dosáhnout i bez velkého GPU clusteru. Zároveň trénujeme na konkrétním veřejně dostupném datasetu. Naše výsledky jsou tak snadno reprodukovatelné. Modely vyhodnocujeme na devíti jazycích, což nám poskytuje kvalitní přehled o jejich silných stránkách. Mimo to také podrobně analyzujeme nastavení značného množství hyperparametrů bi-enkóderů — populárního přístupu pro propojování entit — čímž zjednodušujeme rozhodování navazujícím pracem. Náš výzkum ukazuje, že lze vytvářet silné mnohojazyčné systémy na propojování entit i za použití pouze omezených výpočetních zdrojů. Tím činíme celou úlohu přístupnější.