

Question answering je disciplínou informatiky v oblasti zpracování přirozeného jazyka a získávání informací. Cílem je vytvořit systém, který automaticky najde odpověď na určitou otázku v textu. V dnešní době existuje spousta modelů trénovaných na obrovských tréninkových datových souborech v angličtině. Tato práce se zaměřuje na budování podobných modelů v češtině bez českých tréninkových datasetů. Při této práci jsme použili SQuAD 1.1 a přeložili jej do češtiny pro vytvoření trénovacích a testovacích datových souborů. Pak jsme trénovali a testovali modely BiDirectional Attention Flow a BERT. Nejlepší získaný výsledek na českém datasetu je z modelu BERT trénovaného na češtině s přesnou shodou 60,48% a skóre F1 73,46%. Kromě toho jsme také natrénovali model BERT na anglickém datasetu a vyhodnotili jsme ho na českém testovacím datasetu bez překladu. Dosáhli jsme přesné shody 63,71% a skóre F1 74,78%, což je mimořádně dobré navzdory tomu, že model dosud neviděl žádné české question answering data. Takový model je velmi univerzální a poskytuje systém odpovědí na otázky v jakémkoli jazyce, pro který máme dostatek monolingválních textů.