

Hlučet dovoleno. Umělá inteligence zvládne přesně zapisovat řeč i v rušné místnosti

6.3.2023 - Adam Pluhař | Technická univerzita v Liberci

Za aplikací Beey, jež se úspěšně etablovala i v zahraničí a má uživatele v 50 zemích, stojí vývojáři z pražské firmy Newton Technologies a tým vědců z Laboratoře počítačového zpracování řeči na Fakultě mechatroniky, informatiky a mezioborových studií TUL (FM). Základem programu je jeden z nejkvalitnějších nástrojů pro automatické rozpoznání řeči, který je v současnosti na trhu. Poradí si nejen s češtinou, ale i s dalšími devatenácti světovými jazyky. Mezi všemi dvaceti jazyky umí aplikace Beey také překládat. Nyní už není překážkou ani hluk v místnosti.

„Nové architektury neuronových sítí umožňují trénovat systémy rozpoznávání řeči pouze na základě zvukových nahrávek a odpovídajících textových přepisů. Na rozdíl od minulé generace modelů tak není nutné vytvářet slovníky obsahující pro každé slovo různé výslovnostní varianty. A zejména novou generaci modelů není vůbec nutné učit, jak se slova v daném jazyce vyslovují. To výrazně zjednodušuje proces přípravy dat pro učení, kterých je nyní možné použít řádově více než v minulosti,“ říká Petr Červa, vedoucí týmu vědců z FM, který pracuje na počítačových modelech na rozpoznávání řeči.

Přesnost čtených zpráv je téměř stoprocentní

U čtené nebo předem připravené řeči je už dnes přesnost přepisu téměř sto procent. Pro záznamy obsahující promluvu více mluvčích najednou nebo velkého ruchu na pozadí se pohybuje od 80 procent výše. *„Prostor pro zlepšení ještě stále máme. Obecně ke zvyšování přesnosti přispívá kromě stále lepší architektury neuronových sítí také vzrůstající množství dat pro trénování, které máme k dispozici. Souvisí to s tím, že dané technologie využívá stále více lidí,“* dodává docent Červa.

Převod mluvené řeči na psaný text umožňuje získávat rychlé a levné přepisy podcastů, rozhovorů, schůzí nebo přednášek. Umělá inteligence v Beey tituluje videa na internetu a s využitím automatických překladů zpřístupňuje také audiovizuální obsah z celého světa. Technologie nachází využití i v průmyslu, zdravotnictví, médiích nebo státní správě.

„Naším cílem je, aby výsledky posledních výzkumů v oblasti umělé inteligence mohl využívat každý, ať je to student, lékař, novinář nebo třeba státní úředník. Proto si náš program může vyzkoušet zdarma každý, kdo má počítač nebo chytrý mobil a přístup k internetu,“ přibližuje základní filozofii společnosti firmy Newton Technologies její ředitel Petr Herian.

Program píše titulky, monitoruje média nebo pomáhá na soudech

Současný zvýšený zájem o novinky v oblasti umělé inteligence podle Petra Heriana pomáhá tyto technologie dále zlepšovat a hledat i jejich další uplatnění. *„Lidé si pomalu zvykají na to, že se hlasové technologie stávají součástí jejich práce i běžného života. A nejde už zdaleka jen o diktování zpráv do mobilu. Náš program tituluje oblíbené seriály, pomáhá neslyšícím nebo přepisuje jednání na soudech. Usnadňuje práci novinářům nebo výuku ve školách. Nových příležitostí je celá řada a já jsem hrdý na to, že jsou u toho i čeští vývojáři a vědci,“* říká Petr Herian.

V době pandemie koronaviru začala platforma Beey titulkovat vybrané, zejména zpravodajské, pořady pro sluchově postižené. „Kromě titulkování lze ale Beey obecně využít pro přepis a následnou opravu jakéhokoli zvukového záznamu. Může jít o soudní jednání, záznam rozhovoru z diktafonu nebo třeba automatický přepis televizního či rozhlasové pořadu pro účely monitoringu médií. Právě pro poslední zmíněnou aplikaci se Beey využívá asi nejvíce. Je skvělé, že zpřístupňuje hlasové technologie stále širšímu okruhu osob,“ uzavírá docent Červa.

Počáteční impulz k využívání hlasové technologie pro účely zautomatizování monitoringu televizního a rozhlasového vysílání dali před lety firmě Newton Media vědci z Laboratoře počítačového zpracování řeči na FM. Společnost vsadila na spolupráci s TUL a časem založila firmu Newton Technologies. Ta kromě aplikace Beey stojí například i za softwarem Newton Dictate, který existuje v několika jazykových variantách (na T-UNI jsme psali například o ovládnutí mluvené norštiny). Používá se zejména pro diktování medicínských nálezů a textů v oblasti justice a práva. Právě Newton Dictate byl ve prvním výsledkem spolupráce FM a společnosti Newton.

<https://tuni.tul.cz/a/hlucet-dovoleno-umela-inteligence-zvladne-presne-zapisovat-rec-i-v-rusne-mistnosti-146328.html>