

SLOVA, KTERÁ NEPOMINOU



Petr Herian je český podnikatel a manažer, který založil firmu Newton, specializující se na rozpoznávání hlasu a přepis řeči na text. V tomto oboru překročil horizont lokální firmy a dostal se na světovou špičku. Dokáže konkurovat dokonce i takovým obrům, jako jsou Microsoft a Google.

Milý Petře, v čem podniká tvoje firma Newton?

Velmi zjednodušeně je to tak, že Newton se dělí na dvě skupiny. První je Newton Media, která se zaměřuje na monitoring médií a mediální analýzy. Druhá je Newton Technologies a ta se zabývá rozpoznáváním řeči. My jsme dlouhou dobu pro ten monitoring médií měli 20 písáček, které každý večer přepisovaly televizní a rozhlasové zprávy. Říkali jsme si, že na to přece musí existovat nějaký systém, a tak jsem šel tehdy, zhruba v roce 2003, do IBM, kde jsem věděl, že pracují na rozpoznávání řeči. Oni se mně vysmáli, že systém na přepis češtiny by byl hrozně složitý a těžký. A navíc byl pro ně český trh malý a nezajímavý. Takhle jsme obešli ještě pár firem a zjistili jsme, že to nejde koupit. Zkusili jsme jít i za vysokými školami a domluvili jsme se na Technické univerzitě v Liberci, že my dodáme data a oni to zkusí naprogramovat. A ukázalo se, že ta technologie, na tu dobu, docela něco přepisovala. V té době ministerstvo spravedlnosti hledalo systém na diktování do počítače a už byli domluveni s jednou firmou. Ale my jsme za tři měsíce vymysleli program na diktování do počítače, který byl výrazně lepší než ten konkurenční, a tak si nakonec vybrali nás.

Můžeš nám vysvětlit, jak přepis řeči na text funguje?

Dřív to fungovalo tak, že se vytvořil slovník slov a k tomu se udělal akustický model.

Zvuk lidského hlasu se rozdělil na jednotlivé hlásky, tzv. fonémy, a z těch hlásek se skládala slova. Slovo, které bylo ve slovníku nejpodobnější, se pak napsalo. Ještě do loňského roku to byl u nás hlavní princip přepisu řeči na text.

S rozvojem neuronových sítí a umělé inteligence (AI) se to teď dělá tak, že se vezme velké množství nahrávek spolu s doslovnými přepisy a ten systém se na tom vytrénuje. Sám si zpracuje a vytvoří model, podle kterého pak rozpoznává mluvenou řeč. Ale je to hlavně o tom, kolik a jak kvalitní jsou data na trénování pro daný jazyk. My našťestí máme obrovské množství dat, na kterých trénujeme zejména slovanské jazyky. Troufám si tvrdit, že ve většině



Co všechno přepisujete na text?

Zhruba ještě v době před deseti lety naši zákazníci chtěli všechna data, která se o nich objevila v novinách, rozhlase i televizi. Dneska je těch dat násobně více, takže zákazníci chtějí jenom to nejdůležitější. Začínali jsme s novinami a časopisy. Pak přišly přepisy rozhlasu a televize, pak internetových stránek a do toho se ještě přidala data ze sociálních sítí. Technologie rozpoznávání řeči se používá nejen na zpracování televize a rozhlasu, ale i různých podcastů, YouTube kanálů atd. Když jsme začínali, zpracovávali jsme data druhý den do 11 hodin. Když jsme v roce 1997 vyhráli zakázku na přepis řeči pro Senát České republiky, tak nás donutili, aby to bylo zpracované do osmi ráno. Dnes je potřeba data dodávat do pěti minut od jejich objevení. Když se v médiích o našem klientovi objeví nějaká negativní nebo klíčová zpráva, chtějí ji mít co nejdříve přepsanou a zpracovanou.

Státní instituce, politické strany či velké společnosti chtějí vědět, co se o nich píše v novinách, co se píše o jejich konkurenci, případně o tématech, která je zajímají. Sleduje se citovanost politiků, trendů, značek, jmen lidí a tak podobně.

slovanských jazyků jsme nejlepší na světě a ve světových jazycích se kvalitou můžeme poměřovat s Microsoftem i Googlem.

Může to využít i běžný uživatel?

Vyvíjíme systém, který se jmenuje Beey a slouží k přepisu jakékoliv řeči. Hodně je používán novináři pro přepis rozhovorů, protože je to nástroj, který nejenom umí přepsat řeč ve vysoké kvalitě, ale má velmi chytré nástroje k rychlé editaci. (Tento rozhovor je také přepsán pomocí aplikace Beey.io.) To znamená, že je možné text opravit jedním tlačítkem, přeložit ho do cizího jazyka nebo z něho udělat titulky. A to je směr, kterým teď jdeme. Dokážeme přepsat jakoukoliv nahrávku, ať se jedná o televizi, přednášku, rozhovor, jednání zastupitelstva nebo parlamentu. Máme za sebou i pilotní nasazení, kdy jsme náš systém použili na konferenci, kde mluvili řečníci v angličtině a vedle s nějakým minimálním zpožděním běžely titulky v češtině.

Takže studenti ve škole mohou mít v mobilu spuštěnou vaši službu a rovnou jim to přepíše přednášku?

Přesně tak. Samozřejmě bychom se museli zamyslet, jak to těm studentům finančně zpřístupnit. Ale když si člověk změří čas strávený přepisem přednášky, naším systémem ušetří obrovské množství času, a tím i peněz.



AI vtrhla do našich životů silou povodně. Máš představu, jakým způsobem by mohla změnit společnost?

To je hrozně těžká otázka. Já jsem ještě před rokem nebo dvěma byl na jedné konferenci, kde mluvil Tomáš Mikolov, jeden z našich nejlepších AI vědců, který pracoval v Googlu a Facebooku. Když se ho ptali na AI, říkal, že je to jenom statistika. A já si to také stále ještě myslím. Je to o tom, že AI má velké množství dat a naučí se je rozpoznávat. To znamená, že když jí dáme miliony obrázků zvířat, ona je schopna je pak s velkou přesností rozpoznat. Dneska už ty systémy mají tolik dat a neuronové sítě fungují tak dobře, že jsou schopny jakoby něco vytvořit nebo namalovat. Nejnovější je teď ChatGPT, což je jazykový model. Když po něm budete chtít, aby vám napsal v jakémkoliv jazyce obchodní dopis, tak vám ho napíše naprosto super, protože má spoustu dat, na kterých se natrénoval.



Jednou jsme zkoušeli na chlapeckém spolku, jestli nám napíše vtip a všechny nás výsledkem překvapil.

Je zatím obrovské množství lidské práce. Co já vím, společnost OpenAI do tréninku AI vložila asi stovky tisíc hodin lidské práce. Vzhledem k tomu, co bylo před rokem, to funguje překvapivě dobře. I v našem systému rozpoznávání řeči si musíme dát obrovský pozor, abychom ho netrénovali na špatných datech. Jakmile tam jsou špatná data, a to byly případy prvních chatbotů, systém sám nepozná, že ta data jsou špatná. Kdybych ho například při vyslovení slova sklenička naučil, ať napíše slovo křeslo, on to tak udělá. Výběr dat, ze kterých se trénuje, musí být velmi pečlivý.

S těmi vtipy je to tak, že ChatGPT měl nejspíš v trénovacích datech milióny vtipů v různých jazycích a tím, že překladové systémy fungují velmi dobře, se dá model natrénovat na datech v jednom jazyce a použít v jiném.

Jak vidíš budoucnost AI?

Jak bude AI napojena a aktuální data z internetu, bude schopna nám velmi uleh-

čit jakoukoliv činnost. Na druhou stranu vidím nebezpečí v tom, že si ty systémy často vymyslí, uživatelé si na to postupně zvyknou a nebudou posuzovat kvalitu. My jsme si dělali testy jazykových modelů a nejčastější chybou byla změna času minulého za budoucí a „kreativní“ doplnění informací, které nebyly pravdivé, ale pravděpodobně byly v trénovacích datech.

Tím, že tyto systémy vyvíjí společnosti jako je Microsoft a Google, které mají spoustu dat a obrovsky výkonné počítače, je otázkou, kam se to časem posune. Stále ale systémy zpracovávají jen jednu úlohu. Jazykový model umí něco napsat, ale už nerozpozná obrázek. Ale jsem moc zvědav, kam se to celé dostane.

Trochu se bojím, že technologie nebudou dokonalé, ale protože nám budou šetřit spoustu času, odpustíme jim různé chyby. Postupně všechnu kancelářskou činnost začneme ovládat hlasem a místo toho, abychom přemýšleli, jak napsat dopis, jak spočítat matematický příklad, budeme říkat, ať nám to napíše a vypočítá. Obávám se, že z toho trochu zlenivíme.

**Děkujeme za rozhovor
Otázky kladl František Jakubec ml.**



ING. PETR HERIAN,

ženatý, otec pěti dětí.

Zakladatel a výkonný ředitel (CEO) firem Newton Media a Newton Technologies.

Oblíbený citát: Nikdy neposuzuj druhého, pokud jsi neušel aspoň míli v jeho topánkách.

Oblíbený film: Mám velice rád film Cesta zhyčkaného dítěte, který pro mne v mládí byl takovou školou podnikání.