

Softwarové programovanie pomocou automatickeho rozpoznávania reči.*

Roman Kuruc

Slovenská technická univerzita v Bratislave

Fakulta informatiky a informačných technológií

xkurucr [at] stuba.sk

18. september 2021

Abstrakt

Hlasové programovanie umožňuje inžinierom zmierniť ich fyzické nepohodlie, prípadne. Má to potenciál byť rýchlejšie ako normálne písanie na klávesnici a môže to pomôcť ľuďom s rôznymi postihnutiami. Zatiaľ čo existujú mnohé riešenia na hlasové programovanie, modelovo riadené inžinierstvo musí ešte len využiť tento netradičný, ale vysoko potenciálny prístup do softwarového vývoja. V mojej práci by som sa chcel zamerať na to, čo je to automatické rozpoznávanie reči, ako sa vytvára, aký to má potenciál a akú to má budúcnosť v softwarovom vývoji.

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2021/22, vedenie: Ing. Fedor Lehocki, PhD.

1 Úvod

V tomto semestrálnom projekte sa Vám pokúsim vysvetliť históriau, vývoj programov a takisto aj programy, ktoré v súčasnosti využívajú funkciu automatického rozpoznávania reči. Budem sa to snažiť vysvetliť tak, aby aj človek, ktorý nikdy nepočul o tejto technológii tomu chápal. Aj keď v dnešnej dobe rozpoznávanie reči je uplatnené v mnohých smeroch, stále sa snažia vývojarí nájsť nové spôsoby vývoja programov.

1.1 Prečo som si vybral tuto tému?

Potom čo bola zverejnená rámcová téma na semestrálny projekt z predmetu informačné vzdelávanie, tak nemal som vôbec žiadne tušenie, o čom by som mal písať. Dlho som hľadal tému, ktorá by ma zaujala na toľko, aby som si ju vybral. Po nejakom čase som prišiel k článku – softwarové modelovanie pomocou rozpoznávania reči. Po prečítaní abstraktu a obsahu článku, ktorý bol publikovaný na I EEE xplore, ma zaujala téma natoľko, že, som sa rozhodol vyberať túto tému na moju semestrálnu prácu.

2 Čo je to rozpoznávanie reči?

Rozpoznávanie reči nie je žiadna nová vec, ktorá bol vyvinutá v 21. storočí. Prvé systémy s touto technológiou boli vytvorené v 50. rokoch minulého storočia. Odvtedy boli mnohé pokusy o vytvorenie iných systémoch ovládanými hlasom, ktoré by asistovali so softwarovým inžinierom vo vytváraní kódov pre programy. Výhodou takýchto systémov je rýchlosť a presnosť. V priemere človek píše približne 60-70 slov za minútu, zatiaľ čo programy s týmito funkciami dokážu písanie v priemere 107 slov za minútu. Tradičný spôsob klávesnica a myš je oveľa pomalší. Tak isto je oveľa väčší počet chýb u užívateľov, ktorí píšu na klávesnici. Často krát písanie na klávesnici vyžaduje pohyb rúk z jednej strany na druhu, aby sme boli schopní zadať špeciálne znaky, čo značne spomaľuje naše písanie. Ako som už spomienul, písanie pomocou rozpoznávania reči má výhodu v rýchlosti, ale aj v tom, že dokáže pomôcť ľuďom s fyzickým postihnutím. Tak isto dokáže zmierniť bolest pre ľudí, ktorí trpia fyzickou bolestou spôsobenou repetitívou činnosťou, ako je napríklad syndróm karpálneho tunela. [2]

2.1 Programy využívajúce VDSM

V dnešnej sú rôzne programy, ktoré využívajú VDSM, prostredníctvom ktorého dokážeme vytvárať programové kódy pomocou programov ako je napríklad NatLink, SPEED, voicecode.io, Vocola alebo Aenae. Teraz vám vysvetlím, aká je funkcia každého programu. Natlink je hlasovo riadené pythonové prostredie vytvorené Joelom Gould. Toto je základných projektov, ktoré je rozšírením pre VoiceCode, Vocola a ďalších iných. Aenae je projekt, ktorý rozširuje NatLink, aby mohol fungovať ako prostredie pre klientovi server. Tieto prostredia sú častokrát využívané vývojármi, ktorí zoberú verbálne generované príkazy a premenia ich naprieč prepojeniami, ako keby boli písané na klávesnici. Jedným z ďalších projektov je SPEED, program vytvorený Andrewom Begel z Harmonia Research na University of California at Berkeley. Tento systém využíva verbálne rozhranie na vytvorenie hovoreného Java kódu, ktorý je podobný štandardnému Java jazyku a môže byť skompilovaný vo väčšine štandardných Java kompilátoroch. Dodnes je Vocola je jedna z najpopulárnejších pomôcok v VDSM. Má rovnaké vlastnosti ako SPEED a VoiceCode. Jedným z posledných programoch je VoiceCode, vytvorený Benom Meyerom. Voicecode dokáže zmapovať rôzne špecifické verbálne frázy do textových strún, makier a akcii, ktoré mu dovolili reťazenie a hniezdenie príkazov. Tieto preddefinované makrá dovoľujú explicitne integráciu s deviatimi rôznymi kódovými editormi a teoretický integrácia s akýmkoľvek inými editormi. Teda potom užívateľ je schopný vytvoriť a testovať vhodné makrá.

Posledným programom je ModelByVoice. Tento nástroj je určený špeciálne pre zrakovo postihnutým pri modelovacích aktivitách. Snaží sa nahradieť už existujúce softvérové modelovacie nástroje ako Simulink. Je určený hlavne na odstránenie všetkých grafických a textových požiadaviek z modelovacieho prostredia. ModelByVoice funguje tak, že využíva Google Cloud Speech-to-Text, ktorý prijme výstup od používateľa a vďaka FreeTTS vytvorí počúvateľný text ako výstup. [2]

2.2 Hidden Markov Model - HMM

Hidden Markov Model (HMM) je štatistická metodika pre automatické rozpoznávanie reči. Táto metóda bola testovaná vo veľkom spektri aplikácií. Na druhú stranu Prediction by Partial Matching (PPM) je technika štatistického modelovania s konečným kontextom. Táto technika dokáže predpovedať ďalšie znaky na základe kontextu a dokázala, že má skvelí potenciál vo vývoji nových riešení pre rôzne jazykové modeli v rozpoznávaní reči. Zatiaľ čo HMM sa sústredí na globálne štatistické funkcie v reči, PPM zdôrazňuje čiastočnú predikciu kontextu pre vylepšenie rozpoznávania. Metóda PPM môže byť ďalej vylepšená začlenením neurčitej zhody do procesu párovania. [3]

3 Využitie v leteckom priemysle

Jedným z možných využítí tejto technológie je v leteckom priemysle, presnejšie v ATC (Air Traffic Control) prostredí. Hlavným dôvodom je kvôli komunikácií s pilotmi pomocou hlasu. Aj keď nové technológie dovoľujú datalink komunikáciu, hlas je stále hlavným zdrojom komunikácie medzi zemou a vzduchom. Tým pádom automatické rozpoznávanie reči v komunikácii vedie k automatickej transkripcii komunikácie a následne s dobrým back-end systémom aj k automatickej detekcii kontrolných udalostí, ktoré dokáže kontrolór vykonávať pomocou hlasu. [1]

4 Záver

Cieľom tohto článku bolo objasnenie témy, ktorú som si vybral. To, či sa mi to podarilo, je už na osobnom zvážení každého čitateľa. Ako som už spomenul v úvode, dúfam, že táto technológia sa bude v budúcnosti ďalej vyvíjať a čoraz viac sa budeme s ňou stretávať v obyčajnom živote.

Literatúra

- [1] José Manuel Cordero, Manuel Dorado, and José Miguel de Pablo. Automated speech recognition in atc environment. In *Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems*, ATACCS '12, page 46–53, Toulouse, FRA, 2012. IRIT Press.
- [2] Yadong Li, Danlan Li, Wenqiang Cui, and Rui Zhang. Research based on osi model. In *2011 IEEE 3rd International Conference on Communication Software and Networks*, pages 554–557, 2011.
- [3] Jackson Zhang and Bruce Wang. A novel voice recognition model based on hmm and fuzzy ppm. In *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*, pages 637–640, 2010.