

Spracovanie jazyka pri inteligentnom hlasovom vyhľadávaní*

Martin Kvietok

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií
xkvietokm@stuba.sk

30. november 2023

Abstrakt

V súčasnosti zohráva spracovanie jazyka (NLP) veľmi dôležitú úlohu pri efektívnom vyhľadávaní informácií na internete. Bežným postupom ako niečo vyhľadať je napísanie výrazu na klávesnici. No s dnešnými modernými technológiami vieme vyhľadať potrebný výraz ešte ľahšie a dokonca bez písania. Takéto riešenie výrazne urýchľuje proces práce s informáciami a ušetrí používateľom veľa času. V tejto práci predstším použiteľnosť tohto riešenia, inovatívne vyhľadávanie hlasom pomocou Smart Voice Search Engine (SVSE), ktorého základ tvorí spracovanie prirodzeného jazyka. NLP umožňuje počítačom porozumieť ľudskému jazyku, čo je kľúčovým prvkom aktuálnej problematiky. V tejto práci sa ďalej budem snažiť detailnejšie vysvetliť princíp fungovania NLP a jeho implementáciu v rámci SVSE, taktiež funkciu rozpoznávania hlasu používateľa a praktickú využiteľnosť v budúcich projektoch.

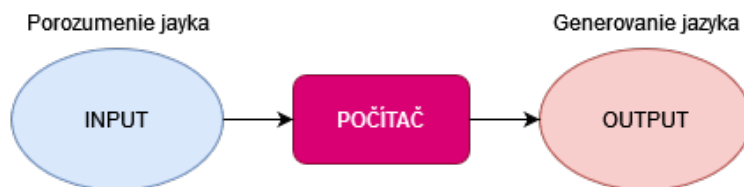
1 Úvod

V dnešnej dobe je prístup k informáciám a službám na internete takmer samozrejмый, pričom hovorené slovo sa stáva prirodzeným nástrojom na vyhľadávanie a pomaly nahrádza nepraktické písanie na klávesnici. V uplynulých desaťročiach sme boli svedkami prudkého vývoja v oblasti rozpoznávania prirodzeného jazyka. To malo za následok, že už v roku 2007 dosiahla firma Google prelomové výsledky, keď rozšírila aplikáciu Google Maps o funkciu hlasového vyhľadávania. Tento vývoj následne viedol k vzniku personalizovaných hlasových asistentov, ktorí sú schopní okamžite vyhľadávať informácie na základe hovorených otázok a efektívne transformovať reč na užitočný obsah. Tento článok sa zaoberá problematikou rozpoznávania jednotlivých používateľov a ich konkrétnej reči. Taktiež uvádza komplexný model s označením SVSE, ktorý je navrhnutý s cieľom zvýšiť dostupnosť a efektivitu hlasového vyhľadávania na internete.

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2023/24, vedenie: Ing. Mohammad Yusuf Momand, MSc.

2 Spracovanie prirodzeného jazyka (NLP)

Spracovanie prirodzeného jazyka (NLP) je odvetvie informatiky, ktoré sa zaoberá vytváraním počítačových systémov schopných pracovať s ľudským jazykom. Začiatky spracovania prirodzeného jazyka (NLP) sa datujú približne od roku 1950 [12], no od tej doby prechádza NLP neustálym vývojom, ktorý pokračuje až po súčasnosť. Každý rok so sebou prináša stále nové a fascinujúce možnosti. Jeden z najvýraznejších momentov bol dosiahnutý modelom GPT-3 (Generative Pre-trained Transformer 3) od spoločnosti OpenAI už v roku 2022. GPT-3 je najväčší a najvýkonnejší jazykový model schopný produkovať pôsobivý a gramaticky správny text, odpovedať na otázky, a dokonca vytvárať kreatívne textové obsahy. Celý princíp fungovania je založený na procesoch porozumenia a generovania prirodzeného jazyka.



Obr. 1: Diagram 1: Princíp NLP
[8]

”Porozumenie jazyka (NLU) je akýmsi jazykovým detektívom, ktorý zisťuje význam slov v texte, zatiaľ čo Generovanie jazyka (NLG) je kováčom, ktorý formuje slová zrozumiteľné pre počítače.”

Vstupný text vždy prechádza 5 analýzami, počas ktorých sa kontroluje.

1. Rozdelenie textu na slová a vety
2. Gramatické štruktúra vety
3. Význam slova z kontextu
4. Vzájomná závislosť viet
5. Správna interpretácia

[1]

Pri identifikácii jazyka zohráva kľúčovú úlohu predovšetkým strojové učenie (Machine Learning). V minulosti sa identifikácia jazyka zakladala na množstve pravidiel a výnimiek (if-else), čo robilo tento proces časovo neefektívnym a nepraktickým. Strojové učenie je založené na automatickom získavaní a používaní informácií prostredníctvom analýzy príkladov z reálneho sveta. Tieto algoritmy využívajú ako vstupný zdroj veľké množstvo dát a to v praxi znamená - čím viac dát je k dispozícii, tým má daný algoritmus menšiu mieru chybovosti. Systémy založené na algoritmoch strojového učenia majú mnohé výhody oproti ručne vytvoreným pravidlám. [3]

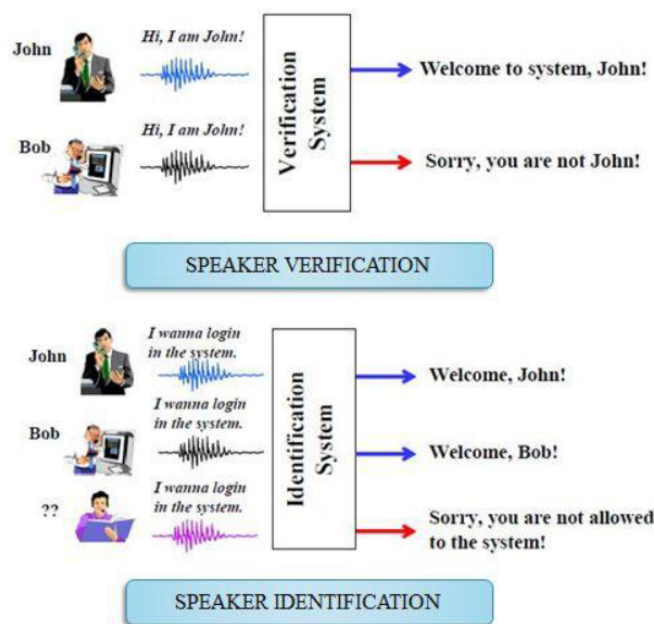
3 Koncept SVSE

Prvé využitie vyhľadávania hlasom bolo uvedené na trh už v roku 2008 v aplikácii Google Mobile App pre iPhone. Táto funkcia rozšírila možnosti hlasového vyhľadávania zo skôr obmedzeného vyhľadávania podnikov na mapách na neobmedzené vyhľadávanie po celom internete. Vyhľadanie na Google hlasom musí byť schopné spracovať všetko, čo dokážu aj bežné internetové prehliadače. To z neho robí pomerne náročný rozpoznávací systém, pretože slovná zásoba a komplexnosť otázok sú veľmi veľké. [9]

3.1 Princíp Rozpoznania Hlasu Používateľa

Proces rozpoznávania hlasu užívateľa sa skladá z 3 dôležitých aspektov.

1. **Identifikácia** - je proces určenia, od ktorého z registrovaných rečníkov pochádza daný prejav
2. **Overenie** - je proces prijatia alebo odmietnutia identifikácie
3. **Diarizácia** - je proces odelenia reči v prípade rozprávania viacerých užívateľov súčasne



Obr. 2: Identifikácia a overenie používateľa
[6]

3.2 Model SVSE

Základné parametre s ktorými SVSE pracuje.

- **Word Error Rate** (miera chýb, UX vs SVSE)
- **WebScore** (optimalizácia chýb UX)
- **Perplexity** (predpoveď vstupu UX)
- **Out-of-Vocabulary Rate** (rozpoznanie jazyka UX)
- **Latency** (celkový čas vyhľadávania)

[9]

3.3 Princíp vyhľadávania

Pred samotným vyhľadávaním je nevyhnutné digitalizovať zvuk pomocou analógovo-digitálneho konvertora. Tým sa akustický zvuk z mikrofónu prevedie do digitálnej podoby. Tento digitálny formát sa ďalej analyzuje a vizualizuje vo forme spektrogramu, ktorý podáva detailné informácie o frekvencii, intenzite a časovej štruktúre zvukového záznamu.

Následne pre lepšie pochopenie a spracovanie textu sa využijú pokročilé metódy, ako je Hidden Markov Model (HMM) alebo Neurónové Siete, ktoré sú zamerané na prácu s fonémami (základnými jednotkami reči).

Nakoniec je potrebné prejsť procesom normalizácie textu, ktorý transformuje písaný text do jeho štandardizovanej formy. Pri vytváraní jazykových modelov, napríklad pre vyhľadávanie pomocou hlasu, je dôležité dosiahnuť konzistentnú a zrozumiteľnú reprezentáciu dopytov nezávisle na ich pôvodnej podobe. Uvediem príklad na pôvodnú podobu dopytu:

”Koľko je teraz hodín v Bratislave?”

Podobne ako v predchádzajúcom príklade, konzistentná a zrozumiteľná reprezentácia dopytu by zahŕňala aj iné možné formulácie ako napríklad:

I. **”Ďakujem, aký je aktuálny čas v Bratislave?”**

II. **”Povedz mi prosím, koľko je hodín v Bratislave?”**

III. **”V Bratislave je teraz koľko hodín?”**

Aby jazykový model rozumel týmto formuláciám otázok a bol schopný na ne adekvátne reagovať, je potrebné ich všetky zaradiť do reprezentácie pre dopyt ohľadom aktuálneho času v Bratislave. Hlavným cieľom je zjednodušiť text pre ďalšiu analýzu alebo spracovanie. Medzi metódy normalizácie patrí odstránenie diakritiky, tokenizácia (rozdelenie textu na jednotlivé slová alebo vety), odstránenie interpunkcie, konverzia slov do ich základných tvarov a rozvoj skratiek. Významnou časťou normalizácie sú procesy ako stemming a lemmatizácia.

Proces tokenizácie textu na príklade:

[**'Kolko', 'je', 'teraz', 'hodín', 'v', 'Bratislave'**]

Stemming sa snaží redukovať slová na ich koreňový tvar, často nezávisle na tom, či výsledný tvar je platné slovo v slovníku.

Ukážka stemmingu na príklade:

[**'kolko', 'je', 'teraz', 'hodín', 'v', 'bratislav'**]

Naopak, lemmatizácia sa snaží slová redukovať na základný tvar, ktorý je platným slovom v danej jazykovej sústave.

Ukážka lemmatizácie na príklade:

[**'Kolko', 'je', 'teraz', 'hodín', 'v', 'Bratislava'**]

Tieto techniky normalizácie majú za cieľ zjednodušiť text a urobiť ho vhodným pre ďalšiu analýzu a spracovanie v rámci aplikácií spracovania prirodzeného jazyka (NLP) a podobných oblastí. [4]

Vo finálnom kroku bude postupne skúmaná sekvencia slov, ktorá bola identifikovaná. Tieto slová budú použité na vyhľadanie informácií na Google. Výsledky vyhľadávania budú overované, aby sa zistilo, či presne zodpovedajú tomu, čo používateľ hľadá. Ak nie, systém sa pozrie na inú možnosť sekvencie slov a pokračuje v tomto procese, kým nenájde tú správnu. Potom systém zaznie signál, aby používateľa informoval o výsledkoch. [9]

4 Efektívnosť

Rozhovor(min)	Počet hovoriacich	Pohlavie	Presnosť rozpoznania
Do 10	2	Muž/Žena/Mix	65%/62%/71%
Do 10	5	Muž/Žena/Mix	55%/51%/65%
Nad 10	2	Muž/Žena/Mix	69%/67%/74%
Nad 10	5	Muž/Žena/Mix	60%/55%/69%

Tabuľka 1: Tabuľka testov SVSE
[8]

Efektívnosť SVSE spočíva predovšetkým v presnom rozpoznávaní používateľa a správnom porozumení hľadaného textu. Počas testov zameraných na rozpoznávanie používateľov sa systému podarilo dosiahnuť presnosť vyšiu ako 50% u každého rečníka. Najlepšie výsledky boli zaznamenané pri dlhších rozhovoroch medzi mužom a ženou.

Tento úspech môže byť pripisovaný implementácii strojového učenia. Systém sa dokázal prispôbiť rôznorodej slovnej zásobe a výrazom používateľov. Kľúčovým faktorom bolo, že systém je schopný lepšie porozumieť a rozpoznať rečníka, keď má k dispozícii viac vzorových údajov.

V dnešnej dobe je teda stále kľúčové investovať do zbierania a analýzy vzorových údajov a ich následné využitie pri vylepšovaní schopností hlasových

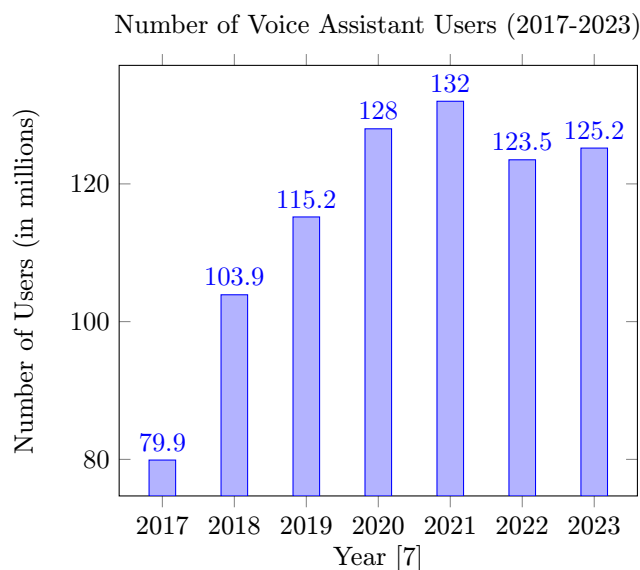
vyhľadávačov. Týmto spôsobom môžu tieto systémy lepšie zodpovedať potrebám používateľov a ponúkať im personalizovaný a efektívny zážitok pri interakcii s technológiou. [5]

5 Potencionálne nedostatky

- (a) **Neznámy jazykový vstup** - Algoritmy rozpoznávania hlasu môžu rozumieť iba naučeným jazykom, dialektom a prízvukom.
- (b) **Ochrana súkromia** - Tieto zariadenia zbierajú a uchovávajú informácie o tom, čo používatelia vyhľadávali a ukladajú ich hlasové záznamy. Na zmiernenie niektorých obáv je dnes hlasový údaj klasifikovaný ako biometrická informácia a chránený všetkými hlavnými zákonmi o súkromí. Firmy, ktoré chcú využívať túto technológiu, musia dodržiavať striktné predpisy. [2]
- (c) **SEO** - Prvé zobrazované výsledky vyhľadávania majú vždy najvyššiu prioritu. Algoritmus prehliadne nižšie hodnotené webové stránky, čo znamená, že potenciálni zákazníci sa o nich nikdy nedozvedia. [10]

6 Budúcnosť hlasových vyhľadávačov

Graf používania hlasových vyhľadávačov jasne ukazuje, že sa hlasovo-ovládne technológie rýchlo stávajú neoddeliteľnou súčasťou každodenného života ľudí. Táto zmena v správaní a preferenciách používateľov bezpochyby súvisí s výhodami, ktoré hlasové vyhľadávanie prináša. Pohodlie a rýchlosť, charakteristické pre túto formu interakcie, vytvárajú prostredie, v ktorom je hlasové vyhľadávanie vnímané ako efektívny a prirodzený spôsob získavania informácií. [11]



Historické súvislosti. Aj v oblasti informatiky nájdeme plno fascinujúcich osobností, ktoré zanechali za sebou nezmazateľnú stopu v histórii ľudstva a ktorých výsledky prác používame dodnes. Existovali by vôbec počítače keby nebolo 7 statočných inžinierov? Gottfried Wilhelm Leibnitz, ktorý priniesol ideu zápisu čísel do binárnej sústavy. George Boole zakladateľ Booleovskej algebry a logiky. Gottlog Frege s jeho prvým symbolického jazykom založeného na matematike a logike. Taktiež Cantor, Hilbert a Gödel, ktorí svojimi príspevkami položili základy informatiky prostredníctvom pokrokov v teórii množín, matematickej logike a pochopení výpočtových limitov. A nakoniec Alan Turing vynálezca Univerzálneho stroja, dnes označovaný ako počítač (stroj používaný na viac nezávislých činností). Nanešťastie, žiadny z týchto pánov už dnes nie je medzi nami, avšak výsledky ich práce budú ovplyvňovať naše každodenné životy ešte dlhú dobu.

Technológia a ľudia. Projekt = premietnutie idey do reality. Pred začatím každého projektu je potrebné všetko detailne premyslieť a zdokumentovať. Následne by sme mali konať v iteráciách v poradí 1.Špecifikácia 2.Návrh 3.Implementácia. Dôležité je dodržiavať stanovený plán a venovať sa prioritám. Efektívna metóda pre prácu v tíme je využitie SCRUMu. Scrum je jednoduchý spôsob riadenia projektov, kde tímy pracujú spoločne na malých kúskoch produktu, stretávajú sa denne na krátkom stretnutí na prehodnocujú svoje ciele. Týmto spôsobom poskytujeme tímu jasný rámec na plánovanie a sledovanie práce.

Udržateľnosť a etika. Rozvoj = postupný rast, rozširovanie, zdokonaľovanie. Pre spoločnosť vplýva pozitívne a je potrebný pri každej práci, vrátane inžinierov. Prináša nám nové objavy a cesty do budúcnosti, avšak čelíme aj mnohým výzvam ako vyčerpávanie prírodných zdrojov alebo

nadmerná produkcia odpadu. Tieto problémy si vyžadujú inovatívne riešenia v podobe tzv. udržateľného rozvoju (rozvoj, ktorý berie ohľad na budúcnosť, nie iba na aktuálny stav) Etika = vo vzťahu ku kultúre a globálnej spoločnosti, v práci a podnikaní rovnako v informačných technológiách. Etika má aj svoje regulácie, ktoré bývajú najčastejšie spísané v kódexoch pre určité zamerania, ku príkladu medici majú Hippokratova prisaha, IEEE je pre profesijné združenia a Software Engineering Code of Ethics využívame v informačných technológiách.

7 Záver

Spracovanie jazyka je nevyhnutnou súčasťou inteligentného hlasového vyhľadávania, ktoré sa stáva čoraz populárnejším prostriedkom interakcie medzi človekom a technológiou. Táto technológia umožňuje užívateľom komunikovať so zariadeniami a aplikáciami prirodzenejším spôsobom prostredníctvom hlasu. Efektívnosť vyhľadávačov sa každým rokom zlepšuje aj vďaka postupnému nasadzovaniu algoritmov strojového a hĺbkového učenia. No predsa tu môžeme nájsť pár nedostatkov, ktoré je potrebné do budúcnosti vyriešiť. Celkový proces získavania informácií z hlasového vstupu a následné porovnávanie pomocou vyhľadávača predstavuje súbor komplexných procesov, ktorý som sa snažil v tejto práci zjednodušiť spracovať. Celkovo je však jasné, že spracovanie jazyka hrá kľúčovú úlohu v rozvoji inteligentného hlasového vyhľadávania a prispieva k vytváraniu užívateľsky príjemnejších a efektívnejších interakcií s technológiou prostredníctvom hlasu.

Literatúra

- [1] An introduction to natural language processing (nlp).
- [2] Luigi's Box. How to optimize for voice searches.
- [3] Abhimanyu Chopra, Abhinav Prashar, and Chandresh Sain. Natural language processing. *International Journal of Technology Enhancements and Emerging Engineering Research*, 1(4):131, 2013.
- [4] Diego Lopez Yse. Text normalization for natural language processing (nlp): Stemming and lemmatization with python. *Towards Data Science*, February 2021.
- [5] Master of Code. 9 key predictions for the future of voice assistants.
- [6] Ashish Kumar Panda and Amit Kumar Sahoo. Study of speaker recognition systems, 2011. Submitted in partial fulfillment of the requirements for Bachelor in Technology in Electronics & Communication.
- [7] Demand Sage. Voice search statistics, 2023.
- [8] Shahenda Sarhan. Smart voice search engine. *International Journal of Computer Applications*, 90(3):40, March 2014.

- [9] Johan Schalkwyk, Doug Beeferman, Françoise Beaufays, Bill Byrne, Ciprian Chelba, Mike Cohen, Maryam Kamvar, and Brian Strope. *“Your Word is my Command”: Google Search by Voice: A Case Study*, chapter Chapter. Publisher.
- [10] Victorious SEO. The impact of voice search on seo, 2023.
- [11] John A. Smith and Mary B. Doe. Examining user perception and usage of voice search. *Journal of Science*, 10(2):123–145, 2022.
- [12] TechTarget. Natural language processing (nlp), 2023.