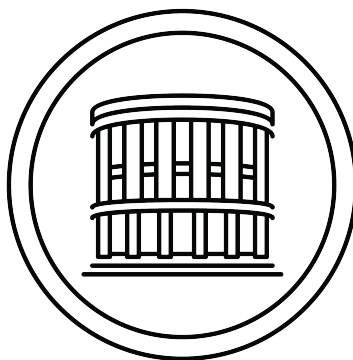


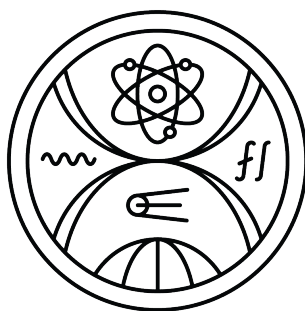
UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



VÁHOVANÁ SLOVNÁ SIETĚ PRE ZDRAVÝCH
ĽUDÍ A ĽUDÍ S PROBLÉMOM REČI

Diplomová práca

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



VÁHOVANÁ SLOVNÁ SIETĚ PRE ZDRAVÝCH ĽUDÍ A ĽUDÍ S PROBLÉMOM REČI

Diplomová práca

Študijný program: Aplikovaná informatika
Študijný odbor: Informatika
Pracovisko: FMFI.KAI - Katedra aplikovanej informatiky
Školiteľ: doc. RNDr. Mária Markošová, PhD.



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Lenka Čakurdová
Študijný program: aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Váňovaná slovná sieť pre zdravých ľudí a ľudí s problémom reči
Weighted word webs for healthy people and people with speech problems

Anotácia: Študent na báze textov zdravých ľudí aj ľudí s afáziou vytvorí váňovanú slovnú sieť. Zanalyzuje štatistické vlastnosti týchto sietí a porovná ich s cieľom zistiť signifikantné rozdiely. Zanalyzuje aj lingvistické vlastnosti týchto sietí.

Cieľ: Zistiť signifikantné rozdiely v slovných sieťach zdravých a chorých ľudí pre potenciálne diagnostikovanie afázie.

Literatúra: Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. Proceedings of the National Academy of Sciences 101 (11), 3747-3752. arXiv:cond-mat/0311416

Vedúci: doc. RNDr. Mária Markošová, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. RNDr. Tatiana Jajcayová, PhD.

Spôsob sprístupnenia elektronickej verzie práce:
prípustná pre vlastnú VŠ

Dátum zadania: 05.10.2023

Dátum schválenia: 11.11.2023

prof. RNDr. Roman Ďurikovič, PhD.
garant študijného programu

.....
študent

.....
vedúci práce

Čestne prehlasujem, že túto diplomovú prácu som vypracovala samostatne len s použitím uvedenej literatúry a za pomoci konzultácií u môjho školiteľa.

Bratislava, 2024

.....
Bc. Lenka Čakurdová

Pod'akovanie

Abstrakt

Kľúčové slová: grafy, váhovaná slovná sieť, afázia, poruchy reči

Abstract

Keywords: graphs, weighted word webs, aphasia, speech disorders

Obsah

Úvod	1
Terminológia	2
1 Východiská práce	3
1.1 Prehľad problematiky	3
1.1.1 Grafy	3
1.1.2 Váňované siete	6
1.1.3 Slovné siete	8
1.1.4 Distribúcia interpunkčných znamienok	10
1.1.5 Poruchy reči	11
1.2 Záverečné práce s podobnou tematikou	15
1.2.1 Analýza a tvorba jazykových grafov ľudí pred a po afázií na báze angličtiny	15
1.2.2 Hľadanie príznakov afázie v textoch na báze slovenčiny	16
1.3 Existujúce aplikácie	16
1.3.1 Gephi	17
1.3.2 Cytoscape	17
2 Motivácia a cieľ práce	20
3 Návrh riešenia	21
4 Implementácia	22
5 Výskum	23
5.1 Grafová analýza textov	23
5.1.1 Analýza textu Malý princ	23
5.1.2 Analýza textov amerického autora	24
Záver	25
Literatúra	26

Zoznam obrázkov

1.1	Vizualizácia neorientovaného grafu s vrcholmi $V = \{1, 2, 3, 4, 5, 6, 7\}$ a hranami $E = \{\{1, 2\}, \{1, 5\}, \{2, 5\}, \{3, 4\}, \{5, 7\}\}$. [3]	4
1.2	Príklad reprezentácie váhovanej siete [12]	6
1.3	Príklad reprezentácie multigrafu [12]	7
1.4	Vizualizácia slovnej siete	8
1.5	Vizualizácia grafu ľudského jazyka [5]	9
1.6	Vizualizácia pridania nového slova do slovnej siete jazyka [4]	9
1.7	Rozdelenie pravdepodobnosti výskytu slov a interpunkčných znamienok pre dva texty v poľskom a českom jazyku. [8]	11
1.8	Oblasti mozgu zodpovedné za produkciu a spracovanie reči.	12
1.9	Používateľské rozhranie aplikácie [7]	16
1.10	Používateľské rozhranie aplikácie [13]	17
1.11	Používateľské rozhranie aplikácie Gephi	18
1.12	Používateľské rozhranie aplikácie Cytoscape	18

Úvod

Terminológia

Pojmy

- Pojem 1
- Pojem 2

Skratky

- Skratka 1
- Skratka 2

1. Východiská práce

Kapitola východiská zahŕňa prehľad teórie, ktorá je relevantná pri riešení diplomovej práce. Definuje grafy, váhované grafy, slovné siete a tiež opisuje dôsledky afázie - poruchy reči v dôsledku poškodenia mozgu. Následne analyzuje podobné záverečné práce a existujúce aplikácie.

1.1 Prehľad problematiky

1.1.1 Grafy

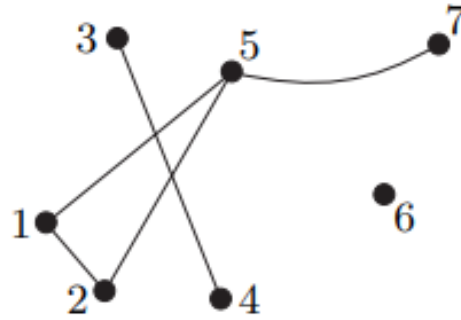
Mnoho systémov môžeme veľmi užitočne reprezentovať ako siete - grafy obsahujúce množinu vrcholov, ktoré sú prepojené hranami reprezentujúcimi rôzne vzťahy medzi danými vrcholmi. Medzi príklady patrí internet, citačné siete, sociálne siete a biologické a biochemické siete rôzneho druhu. Mnohé skúmané siete sú reprezentované binárne - to znamená, že buď sa hrana medzi dvoma vyskytuje, alebo nie. [12]

Graf môžeme definovať ako matematickú štruktúru, ktorá modeluje vzťahy medzi dvojicami objektov. Skladá sa z množiny vrcholov (uzlov), ktoré sú vo dvojiciach prepojené hranami. Poznáme orientovaný, neorientovaný alebo zmiešaný graf.

Graf $G = (V, E)$ je definovaný pomocou množiny vrcholov V a množiny hrán E – neusporiadaných dvojíc $e = \{u, v\}$ vrcholov z V . Grafy sú zvyčajne reprezentované bodkou pre každý vrchol z množiny V a čiarou znázorňujúcou prepojenie dvoch z týchto bodiek, ktoré tvoria hranu - dvojica z množiny E . Graf je neorientovaný, pretože nezáleží na poradí vrcholov v týchto dvojiciach. [3]

Orientovaný graf $G = (V, E)$ je definovaný pomocou množiny vrcholov V a množiny E hrán – usporiadaných dvojíc $e = (u, v)$ vrcholov z V .

V prípade, že sú vrcholy prepojené viac ako 1 hranu, sú prepojené *násobnými hranami*. Ak je vrchol prepojený hranou sám so sebou, hrana sa nazýva *slučka*. Ak graf neobsahuje žiadne slučky, nazýva sa *jednoduchý*.



Obr. 1.1: Vizualizácia neorientovaného grafu s vrcholmi $V = \{1, 2, 3, 4, 5, 6, 7\}$ a hranami $E = \{\{1, 2\}, \{1, 5\}, \{2, 5\}, \{3, 4\}, \{5, 7\}\}$. [3]

Podgraf grafu $G = (V, E)$ je graf $G' = (V', E')$, kde $V' \subseteq V$ a $E' \subseteq E$.

Dva vrcholy x, y grafu G sú *susedné*, ak xy je hrana grafu G . Ak sú všetky vrcholy grafu G navzájom prepojené, potom je graf G *kompletný*. [3]

Grafy je možné reprezentovať maticou susedností A , v ktorej riadky a stĺpce reprezentujú vrcholy, a hodnota A_{ij} nadobúda hodnotu 1, ak vrcholy i, j sú prepojené a hodnotu 0, ak prepojené nie sú.

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

V prípade príkladu matice A môžeme pozorovať reprezentáciu neorientovaného grafu, v ktorej sa vyskytuje vlastnosť symetrie okolo diagonály. Keďže matica obsahuje na diagonále len nuly, graf neobsahuje žiadne slučky.

Ďalšou možnou reprezentáciou je incidenčná matica B , kde riadky reprezentujú vrcholy a stĺpce hrany. V tejto matici prvok B_{ij} nadobúda hodnotu 1, ak je hrana incidentná s daným uzlom, teda či z neho vychádza alebo do neho vchádza. Príklad incidenčnej matice :

$$B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Následne si vymenujeme dôležité štatistické parametre užitočné pri pozorovaní rôznych vlastností a analýze grafov. Medzi tieto parametre patrí najmä v komplexných sieťach *distribúcia stupňov uzlov*, *priemerná dĺžka cesty* a *koefficient zhukovania*.

Štruktúra grafu s N vrcholmi (uzlami) je reprezentovaná binárnou maticou $N \times N$, $A = a_{ij}$, známou ako matica susednosti, ktorej prvok a_{ij} sa rovná 1, keď existuje hrana medzi vrcholom i a vrcholom j a 0 v opačnom prípade. ($i, j = 1, 2, \dots, N$). [6]

Stupeň uzla

Stupeň k_i uzla i je definovaný ako počet jeho susedov, to znamená počet prepojení, ktoré sú viazané na daný uzol i .

$$k_i = \sum_{j \in \Pi(i)} a_{ij}, \quad (1.1)$$

kde a_{ij} sú prvky matice susedností A a $\Pi(i)$ sú susednosti uzla i .

Distribúcia, alebo rozdelenie stupňov je pravdepodobnosť, že nejaký uzol má k spojení s inými uzlami.[6]

Priemerná dĺžka cesty

Charakteristická dĺžka cesty siete je definovaná ako priemer najkratšej cesty medzi ľubovoľnými dvoma uzlami:

$$L = \frac{2}{N(N-1)} \sum_{\{i,j\}} d_{ij}, \quad (1.2)$$

kde d_{ij} je najkratšia cesta medzi uzlami i a j , definovaná ako minimálny počet prejdenných spojení pri ceste z uzla i do uzla j . [6]

Koeficient zhlukovania

V mnohých sieťach platí, že ak existuje prepojenie medzi uzlami i a j , a prepojenie medzi uzlami i a k , zvyšuje sa pravdepodobnosť prepojenia aj medzi uzlami j a k . Táto tendencia spájania susedov ľubovoľného uzla i sa nazýva *zhlukovanie* alebo *klasterizácia*. Je vyjadrená koeficientom c_i , ktorý reprezentuje podiel trojuholníkov, v ktorých sa uzol i vyskytuje a maximálny možný počet takýchto trojuholníkov.

$$c_i = \frac{n_i}{k_i(k_i - 1)} = \frac{\sum_{j,k} a_{ij}a_{jk}a_{ki}}{k_i(k_i - 1)}, k_i \neq 0, 1 \quad (1.3)$$

kde $(1/2)n_i$ je skutočný počet trojuholníkov, v ktorých je uzol i , t.j. skutočný počet prepojení medzi susedmi uzla i , a $k_i(k_i - 1)/2$ je maximálny možný počet spojení, keď je podgraf susedov uzla i úplne prepojený.

Koeficient zhlukovania, alebo klasterizačný koeficient c_i nadobúda hodnotu 1, keď uzol i je centrom úplne prepojeného klastra a hodnotu 0, ak susedia uzla i nie sú navzájom prepojení.[6]

Ďalšie informácie by sme mohli získať preskúmaním klasterizačného koeficientu $C(k)$, ktorý sa zameriava len na triedy uzlov s konkrétnym stupňom k .

$$C(k) = \frac{1}{NP(k)} \sum_{i/k_i=k} c_i \quad (1.4)$$

V mnohých sieťach je koeficient zhlukovania závislý od stupňa $C(k)$ klesajúcou funkciou, čo ukazuje že uzly s nízkym stupňom všeobecne patria do dobre prepojenej komunity, zatiaľ čo úseky s vysokým stupňom sú spojené s mnohými uzlami, ktoré môžu patriť do rôznych skupín, ktoré nie sú priamo prepojené. [1]

1.1.2 Váhované siete

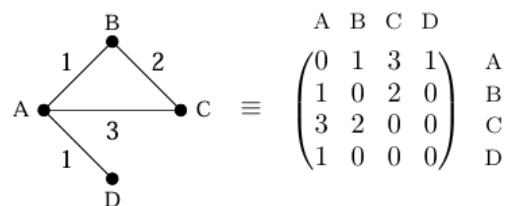
Binárna reprezentácia prepojení entít v sieťach však v mnohých prípadoch nie je postačujúca. Spolu s komplexnou topologickou štruktúrou, siete vykazujú veľkú heterogenitu v kapacite a intenzite prepojení - váhy prepojení. Pre ilustráciu uvedieme príklad: počet voľných miest v lietadle na konkrétnej linke, rôzne silné sociálne väzby medzi ľuďmi v sociálnej sieti, atď.

Preto zavedieme pojem **váhované siete**, kde každej hrane z uzla i a uzla j je priradená príslušná váha w_{ij} . [12]

Váhovaná sieť môže byť reprezentovaná maticou susedností W s prvkami, ktoré nie sú jednoducho 0 alebo 1, ale namiesto toho sú rovné váham na hranách:

$$W_{ij} = \text{váha prepojenia uzla } i \text{ a uzla } j.$$

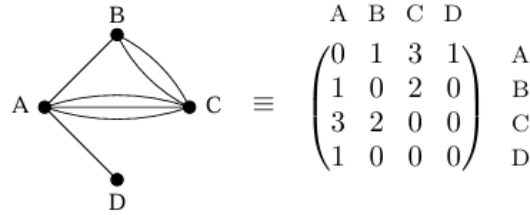
Príklad váhovanej siete :



Obr. 1.2: Príklad reprezentácie váhovanej siete [12]

No matica s kladnými číselnými hodnotami sa môže vyskytovať aj v inom prípade, pri reprezentácii multigrafov - sietí, alebo grafov, ktoré obsahujú viacnásobné hrany medzi vrcholmi.

Na obrázkoch 1.2 a 1.3 môžeme pozorovať, že ich matica susedností týchto sietí je rovnaká. Rovnaké bude aj ich správanie. Z toho môžeme usúdiť, že pri analýze váhovaných sietí by sme si mohli uľahčiť prácu premapovaním danej váhovanej siete na



Obr. 1.3: Príklad reprezentácie multigrafu [12]

nevážený multigraf. To znamená, že by sme každú hranu s váhou w nahradili w paralelnými hranami s váhou 1, ktoré by spájali rovnaké vrcholy ako predošlá váhovaná hrana. Matica susedností siete ostane bez zmeny a následne by sme mohli použiť techniky na analýzu nevážených grafov, ktoré sa uplatňujú aj na multigrafy. [12]

Na charakterizáciu váhovaných sietí je potrebné zaviesť nasledovné pojmy:

Stupeň uzla

Stupeň uzla vo váhovaných sieťach rovnako, ako v neváhovaných:

Stupeň k_i uzla i je definovaný ako počet jeho susedov, to znamená počet prepojení, ktoré sú viazané na daný uzol i .

$$k_i = \sum_{j \in \Pi(i)} a_{ij}, \quad (1.5)$$

kde a_{ij} sú prvky matice susedností A a $\Pi(i)$ sú susednosti uzla i . Tento pojem je však priamo rozšírený pojmom sila uzla alebo váhovaný stupeň uzla.

Váhovaný stupeň uzla

Predstavuje súčet váh všetkých hrán vychádzajúcich z daného uzla i .

$$s_i = \sum_{j \in \Pi(i)} w_{ij} \quad (1.6)$$

Sila uzla berie do úvahy konektivitu daného uzla a aj váhy (silu) prepojení. [1]

Váhovaný klasterizačný koeficient

Váhovaný klasterizačný koeficient uzla i je definovaný ako :

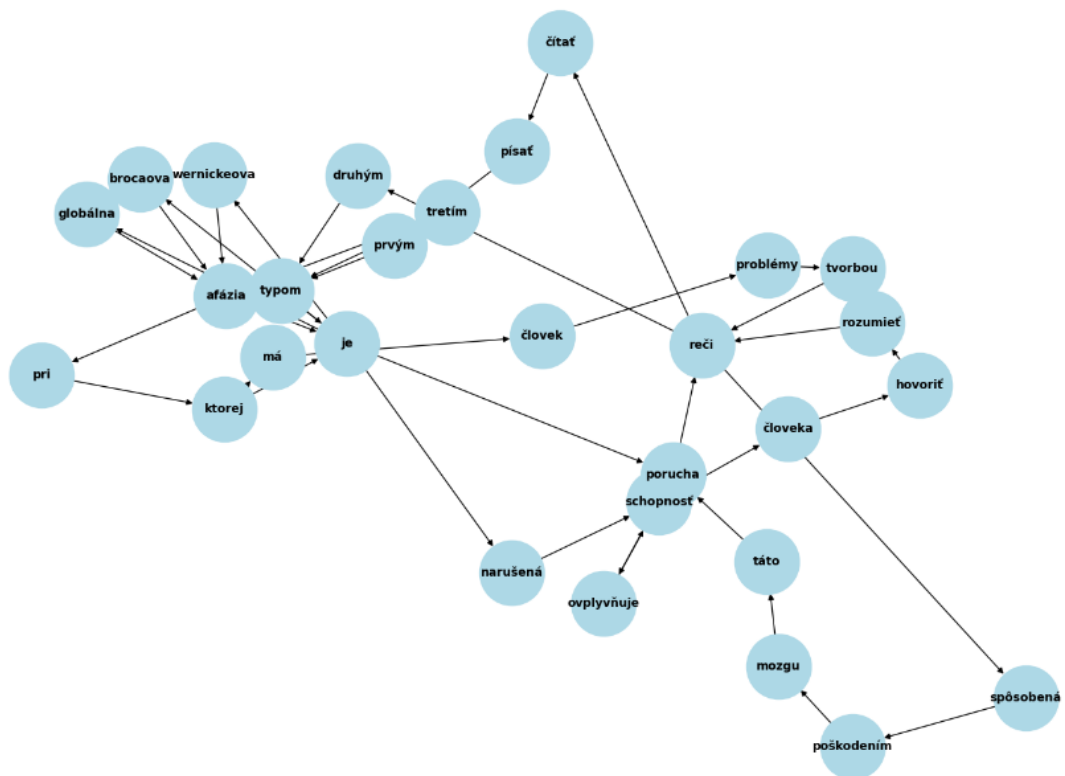
$$c_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh} \quad (1.7)$$

1.1.3 Slovné siete

Ľudský jazyk je komplexný systém, v ktorom medzi sebou interagujú rôzne slová, čo umožňuje tvorbu nekonečných variácií viet z obmedzenej uzavretej množiny slov. Spájanie konkrétnych slov medzi sebou nie je náhodné a vytvára organizáciu jazyka, ktorú by sme mohli reprezentovať ako slovný graf alebo slovnú sieť.

Vytváranie slovných sietí z jazyka je užitočné, pretože umožňuje vizualizovať vzťahy medzi slovami na základe ich významu, gramatických funkcií alebo používania. Tieto siete pomáhajú lepšie pochopiť štruktúru jazyka, identifikovať kľúčové pojmy a skúmať jazykové vzory. Môžeme na nich analyzovať frekvenciu slov, či synonymické a antonymické vzťahy čo je cenné pre lingvistický výskum, strojové učenie a podobne.

Graf, ktorý reprezentuje slovnú sieť obsahuje vrcholy, alebo uzly reprezentujúce jednotlivé slová a hrany medzi uzlami reprezentujú prepojenia týchto slov.

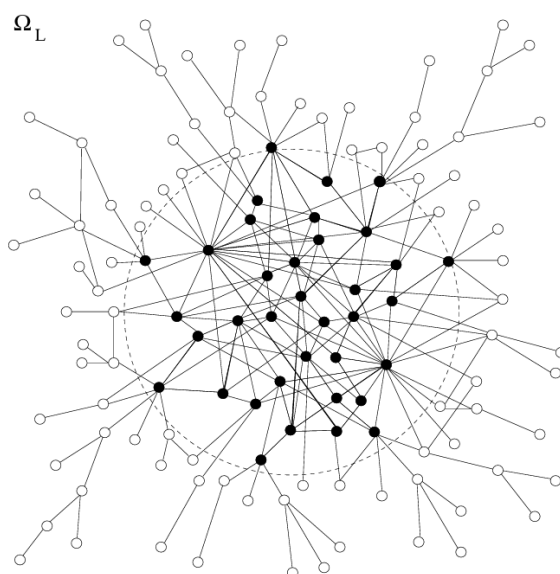


Obr. 1.4: Vizualizácia slovnéj siete

Ľudský mozog v sebe uchováva tisíce slov. Slovníky jednotlivcov sa od seba líšia najmä v závislosti od veku, vzdelania, či povolania. Dôležité je však identifikovať tzv. jadro slovníka, ktoré obsahuje najfrekventovanejšie slová, nevyhnutné pre bežnú komunikáciu naprieč vekovými kategóriami, či sociálnym postavením.

Predstavme si graf ľudského jazyka Ω_L , ktorý je definovaný ako $\Omega_L = (W_L, E_L)$, pričom $W_L = \{w_i\}$, $i = 1, \dots, N_L$ je množina slov jazyka a $E_L = \{\{w_i, w_j\}\}$ je množina všetkých hrán, ktoré dané slová spájajú. [5]

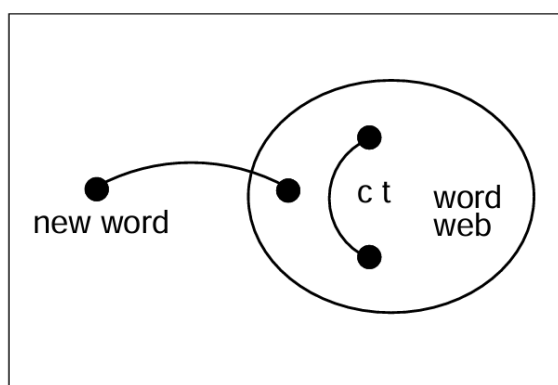
Na obrázku 1.5 môžeme pozorovať reprezentáciu slovnej siete jazyka, kde čierne kruhy predstavujú najpoužívanejšie slová (jadro jazyka) veľmi husto prepojené a svetlé kruhy predstavujú menej využívané slová, špecifické pre rôzne skupiny ľudí.



Obr. 1.5: Vizualizácia grafu ľudského jazyka [5]

Evolúcia jazyka

Jazyk je neustále sa rozvíjajúci a meniaci sa systém, v ktorom niektoré slová zanikajú, iné môžu meniť svoj význam, alebo vznikajú úplne nové slová. Súčasná štruktúra jazyka je podmienená jeho predošlým vývojom. [4]



Obr. 1.6: Vizualizácia pridania nového slova do slovnej siete jazyka [4]

Tento komplexný systém sa sám organizuje počas svojho rastu. Jednotlivé slová sa navzájom ovplyvňujú a môžu sa objavovať nové interakcie medzi novými a starými slovami, kedy pribúdajú do slovnej siete nové hrany.

Po pridaní nového slova (uzla) do slovnej siete, sa nové slovo pripojí k niekoľkým starým slovám (uzlom). Podľa prirodzenej verzie preferenčného spájania sa nové slovo prepojí s nejakým starým uzlom i v závislosti od jeho stupňa k_i (Barabási-Albert model). Okrem toho v každom časovom prírastku vzniká medzi starými slovami ct nových hrán, kde c je konštantný koeficient, ktorý charakterizuje konkrétnu sieť. Tieto nové hrany vznikajú medzi starými slovami i a j s pravdepodobnosťou úmernou súčinu ich stupňov $k_i \times k_j$. [4]

1.1.4 Distribúcia interpunkčných znamienok

Prirodzený jazyk je jeden z najkomplexnejších systémov. Tvoria ho slová, ktoré odkazujú na všetko čo ľudia vedia opísať, či pomenovať. No nie len slová sú základom jazykovej štruktúry. Veľmi dôležitým elementom sú vety. Na tejto úrovni vieme skúmať sémantiku jazyka nielen pomocou slov ale aj vetných členov a interpunkcie. [8]

Písanie, na rozdiel od hovoreného slova vyžaduje používanie interpunkcie, najmä na oddelenie logických častí textu, alebo na vylúčenie nejednoznačnosti. V texte teda interpunkcia hrá dôležitú úlohu. Zo štatistického hľadiska interpunkčné znamienka vykazujú vlastnosti, ktoré sú kvalitatívne podobné vlastnostiam najfrekvencovanejších slov, ako sú podstatné mená, spojky, zámená a predložky.

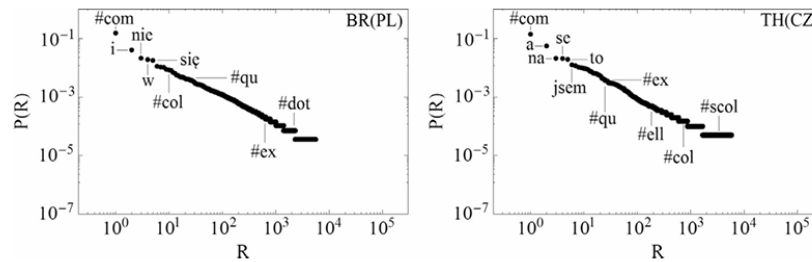
Štatistická analýza vzoriek jazyka odhalila pravidlá, ktoré jazyk dokážu popísať kvantitatívne. Štúdia sa zaoberá distribúciou frekvencie slov, ktorá sa porovnáva s modelom mocninového zákona známym ako Zipfov zákon, a tiež pozoruje závislosť dĺžky textu a počtu unikátnych slov, ktoré sa v texte nachádzajú.

Zipfov zákon tvrdí, že pravdepodobnosť $P(R)$ výskytu R -tého najčastejšieho slova sa mení podľa $P(R) \sim R^{-\alpha}$, kde $\alpha \approx 1$. Zipfova distribúcia funguje pre väčšinu hodnôt, no pri najfrekvencovanejších prvkoch sa odchyľuje od očakávaného vzoru. Namiesto rovného mocninového poklesu sú tieto prvky menej časté, ako by mali byť. Preto sa na presnejšie modelovanie používa Zipfov-Mandelbrotov zákon, ktorý tento posun zohľadňuje:

$$P(R) \sim (R + c)^{-\alpha} \quad (1.8)$$

kde c je parameter zodpovedný za vyššie uvedenú výchylku. [8]

V štúdií [8] boli pre účely výskumu rozloženia interpunkčných znamienok použité prózy v rôznych jazykoch. Z nich sa vybrali najfrekventovanejšie slová, ktoré mali dôležitú gramatickú funkciu vo vetách a boli na prvých priečkach v Zipfovej distribúcií. V tomto výskume boli interpunkčné znamienka zahrnuté do analýzy rovnako, ako keby boli slová. Na obrázku 1.7 je vizualizácia distribúcie týchto znamienok. Je veľmi zaujímavé že rôzne znamienka interpunkcie sa správajú v distribúcií rovnako, ako niektoré vetné členy a sú zaradené do rozličných priečok Zipfoho rozloženia podľa hodnoty frekvencie.



Obr. 1.7: Rozdelenie pravdepodobnosti výskytu slov a interpunkčných znamienok pre dva texty v poľskom a českom jazyku. [8]

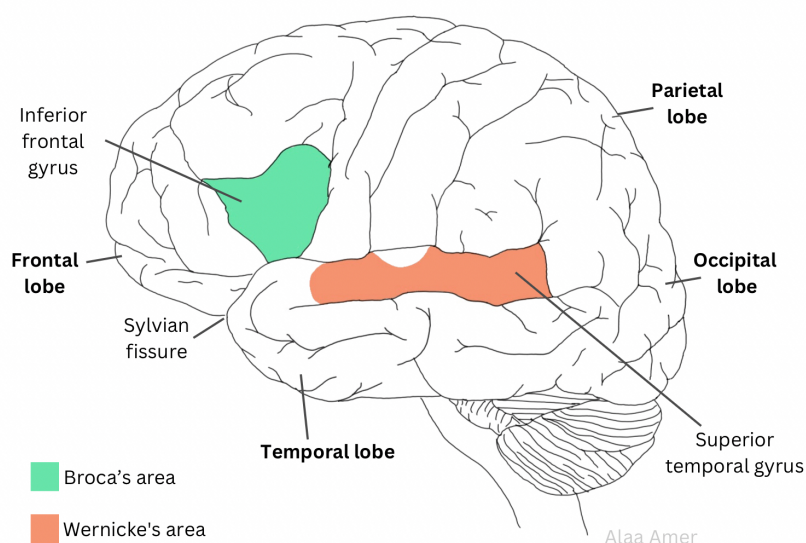
Interpunkčné znamienka sú takmer nerozoznatelné od ostatných najčastejších a stredne častých slov (funkčných aj lexikálnych) pri skúmaní ich štatistických vlastností. Keďže interpunkčné znamienka majú nezanedbateľný význam, je užitočné ich zahrnúť do akéhokoľvek typu analýzy textov. Zahrnutie interpunkčných znamienok do analýzy rozširuje jej dimenzionalitu, a preto otvára väčší priestor pre možné prejavenie niektorých predtým nepozorovaných efektov.

1.1.5 Poruchy reči

V bežnom rozhovore často berieme produkciu jazyka a spracovanie či porozumenie reči za samozrejmosť. Vedíme konverzáciu, vyjadrujeme sa a vnímame a rozumieme bez problémov tomu, čo počujeme. Neuvedomujeme si, aké komplexné procesy prebiehajú v našom mozgu, od fonologického spracovania až po lexikálnu a syntaktickú analýzu. Pozornosť tento proces získa najmä v prípade traumy, či poškodenia mozgu. Vedecké výskumy jazykovej architektúry a spracovania sa opierajú najmä o skúmanie porúch jazyka/reči u ľudí s nejakým typom nervového poškodenia. Prejavy sa líšia v závislosti od oblasti mozgu, ktorá utrpela traumu.

Mnohé z toho, čo sme teraz o tom, ako mozog spracúva jazyk, vzniklo v polovici až na konci 19. storočia v publikovaných štúdiách o skupinách a jednotlivcoch, ktorí si osvojili jazyk po mozgovej príhode. Táto jazyková porucha sa dnes označuje ako *afázia*. [9]

Afázia je porucha reči spôsobená poškodením mozgu. Poškodenia sú rôznej závažnosti a môžu ovplyvniť sluchové vnímanie reči, tvorbu reči, čítanie a/alebo písanie. Väčšina afázií sa v minulosti zamerala na osoby, ktoré utrpeli mozgovú príhodu (narušenie krvného obehu v mozgu), čo malo za následok afáziu, ale afázia môže byť dôsledkom takmer akéhokoľvek typu poškodenia mozgu, vrátane traumatického poškodenia mozgu, nádoru, chirurgického odstránenia mozgového tkaniva alebo infekcie. Afázia môže byť aj dôsledkom neurodegeneratívnych ochorení, ako je frontotemporálna demencia, najmä jeden z podtypov často označovaný ako primárna progresívna afázia. [10]



Obr. 1.8: Oblasti mozgu zodpovedné za produkciu a spracovanie reči.

Posudzovanie pacienta a identifikácia miery poškodenia môže byť zdĺhavý proces. Ak sa jedná o akútne poranenie mozgu, je možné zistiť poškodenie pomerne rýchlo, no u niektorých ľudí to vyžaduje rozsiahle testy a dlhodobú liečbu u logopéda a ďalších odborníkov. Dôležitým aspektom pri posudzovaní afázie je skúmanie miery poškodenia vo viacerých dimenziách jazyka vrátane sluchového porozumenia (slov, viet), spontánnej reči, opakovania reči, pomenovania, čítania a písania. [10]

Pojem **parafázia** označuje rečovú poruchu, pri ktorej človek neúmyselne nahrádza správne slová nesprávnymi, čo narúša plynulosť a zrozumiteľnosť jeho prejavu. Tento jav sa často vyskytuje pri neurologických poruchách.

Druhy parafázie

- **fonemická** - nahradenie jednej fonémy inou v rámci jedného slova (namiesto „mačka“ osoba povie „pačka“)

- **naologistická** - záměna hlásek, které foneticky a sémanticky nesúvisia s výsledným slovom (osoba si vymyslí nezmyselné slovo napr. „tranlák“)
- **sémantická** - nahradenie jedného slova druhým, ale nie vždy sémanticky súvisiace (namiesto „stolička“ osoba povie „stôl“)

Parafázie sa vyskytujú najmä pri typoch afázie s plynulou rečou, kde sa vyskytuje problém s nájdením správneho slova, ktoré sa nahradí nezmyselným. [9]

Parafázia narúša efektívnu komunikáciu a často sa rieši prostredníctvom logopedickej terapie.

Typy afázie

Afázia sa prejavuje rôznorodými príznakmi v závislosti od miesta a rozsahu poškodenia mozgu. Na základe charakteristických jazykových porúch sa rozlišuje niekoľko typov afázie, pričom každý z nich ovplyvňuje rôzne aspekty rečových schopností. Porozumenie týchto typov je kľúčové pre správnu diagnostiku a výber vhodnej terapie. Na obrázku 1.8 sú znázornené časti mozgu, ktorých poškodenie vedie k rôznym prejavom afázie.

Brocova afázia

Brocova afázia je pravdepodobne najviac skúmaným typom afázie. Charakteristická je pre ňu neplynulá prerušovaná reč s výskytom gramatických chýb, či chybné opakovanie reči. V skladbe viet zväčša chýbajú gramaticky funkčné slová a tiež skloňovanie či časovanie slov. Produkcia jazyka je teda agramatická ("bez gramatiky"). Agramatický komponent tohto typu afázie tiež ovplyvňuje porozumenie reči. Porozumenie jednotlivým slovám a jednoduchým vetám sa (väčšinou) javí ako relatívne neporušené. V súčasnosti vieme, že jazyková porucha sa vzťahuje na porozumenie zložitejších vetných konštrukcií obsahujúcich zvrtné podstatné mená.[9] V rečovej produkcii pacientov s Brocovou afáziou sú často prítomné fonematické aj verbálne parafázie a pacienti si zvyčajne uvedomujú svoje deficity a pokúšajú sa o nápravu chýb.

Typ afázie aj oblasť mozgu sú pomenované podľa *Paula Broca*, francúzskeho vedca a lekára z polovice 19. storočia, ktorý bol jedným z prvých, ktorý spájal poruchy reči s poškodením ľavého predného laloku.[10]

Wernickeova afázia

Wernickeova afázia je asi jednou z najznámejších typov afázie. Charakteristická je pre ňu tvorba plynulej reči, ale narušené porozumenie reči vrátane porozumenia jednotlivým slovám. Reč jedincov s Wernickeovou afáziou zvyčajne obsahuje sémantické aj fonemické parafázie. Títo jedinci si zvyčajne neuvedomujú svoje deficity, a preto sú

pokusy o nápravu chýb minimálne. [10] Reč produkovaná pacientom s týmto typom afázie môže obsahovať aj parafázie, ktoré znejú ako jazyk hovoriaceho, no nie sú reálne slová, len možno spĺňajú fonologické pravidlá daného jazyka. Wernickeovu afáziu zväčša spôsobuje poškodenie ľavého temporálneho a parietálneho laloku s centrom v zadnom temporálnom laloku.

Konduktívna afázia

Konduktívnu afáziu charakterizuje plynulá produkcia reči s občasnými parafáziami, no chybné opakovanie vnímanej reči. Tento typ afázie sa tradične považuje za „syndróm odpojenia“, ktorý je spôsobený poškodením arcuate fasciculus, veľkej 'dráhy' bielej hmoty, ktorá spája zadný horný temporálny lalok a Brocovu oblasť. Postihnutí pacienti môžu správne reprodukovať jednoduché slová, či frázy, no majú problémy pri viacslabičných slovách, či abstraktných vetách z dôvodu komplexnejšieho fonologického spracovania. Deficity pri tomto afázii vznikajú najmä pri zapamätávaní a opakovaní vnímanej/vypočutej reči. Najčastejšie sa vyskytuje pri poškodení zadných temporálnych oblastí mozgu. [10]

Medzi ďalšie typy afázie patrí *Transkortikálna motorická afázia*, ktorá sa vyznačuje podobne ako Brocova afázia neplynulou agramatickou produkciou reči. Rozdiel spočíva v takmer bezchybnom opakovaní vnímanej reči. Poškodenie mozgu u tejto skupiny býva v oblastiach okolo Brocovej oblasti, v prednom hornom frontálnom laloku.

Transkortikálna senzorická afázia je podobná Wernickeovej afázii. Pacienti s týmto typom sú charakteristickí tvorbou plynulej, no prázdnej nezmyselnej reči a vymyslených nereálnych slov. Poškodené je porozumenie reči a narozdiel od Wernickeovej afázie jedinci dokážu opakovať vnímané slová a vety. Príčinou tohto typu afázie je poškodenie mozgu v blízkosti Wernickeovej oblasti. [9]

Najzávažnejším typom afázie je *globálna afázia*, kedy je poškodená produkcia aj porozumenie reči. Lézie na mozgu sa zvyčajne vyskytujú na prednom a temporálnom laloku.

Anomická afázia je naopak najmenej závažným typom. Prejavuje sa najmä neschopnosťou spomenúť si na mená alebo niektoré slová. Porozumenie, opakovanie, či produkcia reči ostáva pri tomto type takmer nepoškodená.

Klasifikácia typov afázie

Tabuľka 1.1 klasifikuje typy afázie a rozdeľuje ich na základe toho, či sa zaraďujú do kategórie plynulej alebo neplynulej, prerušovanej reči. Popisuje symptómy spojené

s konkrétnym typom afázie, ktoré sa týkajú produkcie reči, pochopenia vnímanej reči, pomenovávaní slov a tiež opakovania reči.

Typ afázie	Produkcia	Pochopenie	Opakovanie	Pomenovanie
<i>neplynulá reč</i>				
Brocova	prerušovaná	zhoršené	slabé	slabé
Transkortikálno-motorická	prerušovaná	zhoršené	nepoškodené	slabé
Globálna	poškodená	poškodená	slabé	slabé
<i>plynulá reč</i>				
Anomická	chybná	nepoškodené	nepoškodené	mierne ťažké
Wernickeova	parafázie	slabé	slabé	slabé
Transkortikálno-senzorická	parafázie	slabé	nepoškodené	slabé
Kodnuktívna	zjednodušená	nepoškodené	slabé	nepoškodené

Tabuľka 1.1: Typy afázie a súvisiace symptómy

1.2 Záverečné práce s podobnou tematikou

Táto časť práce opisuje záverečné práce, ktoré sa zaoberajú spracovaním textu pomocou grafov, pričom analyzujú zistené výsledky a porovnávajú rôzne štatistické parametre za účelom skúmania zmien v textoch u ľudí s poškodením mozgu a poruchami reči.

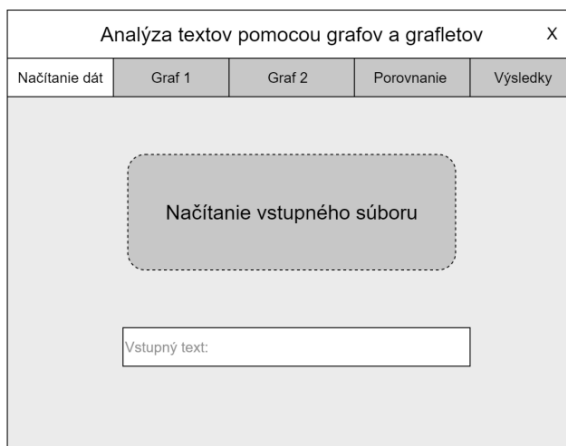
1.2.1 Analýza a tvorba jazykových grafov ľudí pred a po afázií na báze angličtiny

Záverečná práca Bc. Jozefa Kubíka *Analýza a tvorba jazykových grafov ľudí pred a po afázií na báze angličtiny* [7] sa zaoberá spracovaním textu autora, ktorý utrpel poškodenie mozgu. Práca analyzuje autorove texty pred a po afázií a skúma dopady rečovej poruchy na jeho schopnosť písať a vyjadrovať sa.

Výsledok práce zahŕňa aplikáciu, ktorá umožňuje používateľovi vložiť dva texty v anglickom jazyku a následne ich spracovať a analyzovať. Aplikácia spracuje texty, upraví ich do požadovanej podoby a zanalyzuje ich viacerými metódami z teórie grafov. Dostupná je základná analýza a taktiež grafletová analýza vstupného textu. Následne výsledky a štatistiky zobrazí a umožní porovnanie hodnôt. Zistené rozdiely by mali pomôcť identifikácii a skúmaniu mozgovej poruchy a jej vplyvu na vyjadrovanie, či slovnú zásobu pacienta. V aplikácii je tiež umožnené porovnávať vložený text s náhodným grafom.

Grafické rozhranie aplikácie je jednoduché a prehľadné, s príjemným dizajnom. Je určené aj pre používateľov s minimálnymi počítačovými zručnosťami a obsahuje aj vysvetlivky a obrázky pre lepšie pochopenie teórie grafov a uľahčenie analýzy výsledkov.

Po spracovaní textu aplikácia vypočíta viaceré štatistické parametre vytvorených grafov a následne ich zobrazí v užívateľskom rozhraní. Používateľ si môže sám určiť, pomocou akých typov grafov sa mu výsledky zobrazia a taktiež môže určiť, ktoré štatistiky chce zobraziť.



Obr. 1.9: Používateľské rozhranie aplikácie [7]

1.2.2 Hľadanie príznakov afázie v textoch na báze slovenčiny

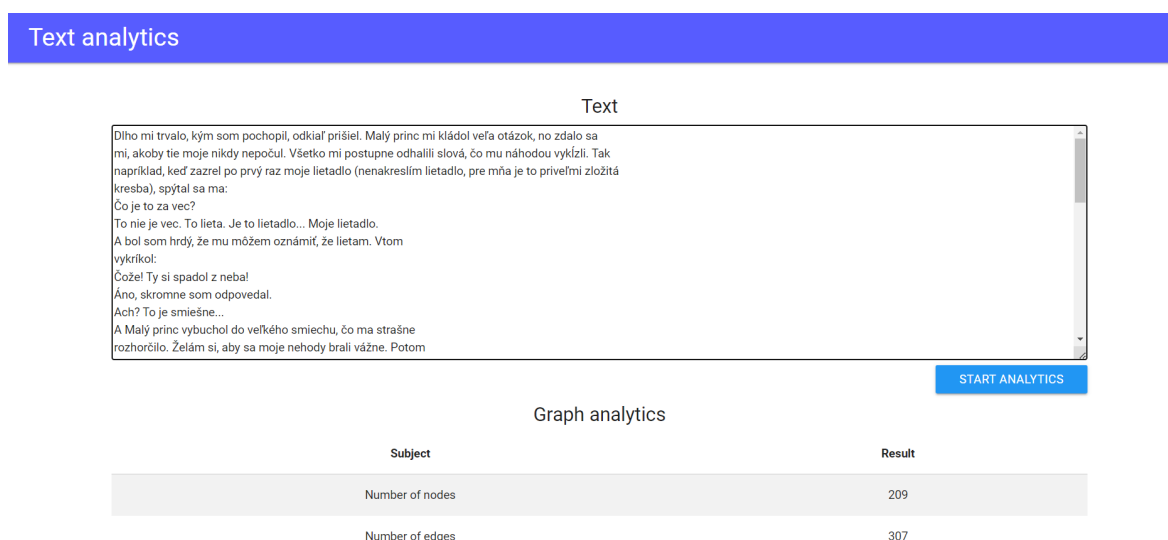
Hľadanie príznakov afázie v textoch na báze slovenčiny [13] je záverečná práca Bc. Daniela Petriľaka, ktorý navrhol aplikáciu na spracovanie a grafovú analýzu textov v slovečine.

Výsledná aplikácia poskytuje grafovú a grafovú analýzu. Používateľské rozhranie umožňuje používateľovi vložiť text určený na analýzu. Aplikácia následne text spracuje, vykoná grafovú a tiež jazykovú analýzu. Následné výsledky a štatistiky zobrazí nižšie v aplikácii v podobe tabuliek a grafov.

Aplikácia umožňuje používateľovi prihlásiť sa a tiež uložiť skúmané dáta do databázy.

1.3 Existujúce aplikácie

Následujúca časť práce predstavuje existujúce aplikácie určené na spracovanie dát a analýzu grafov. Popisuje ich funkcionality, výhody a prípadné nevýhody.



Obr. 1.10: Používateľské rozhranie aplikácie [13]

1.3.1 Gephi

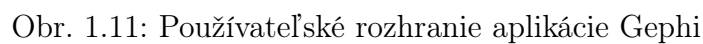
Gephi je open-source aplikácia určená na analýzu a vizualizáciu grafov a sietí. Tento nástroj je obzvlášť populárny v oblastiach, ako je analýza sociálnych sietí, biológia, politológia a ďalšie disciplíny, ktoré sa zameriavajú na analýzu komplexných dátových štruktúr a interakcií medzi rôznymi prvkami. Gephi umožňuje používateľom vizualizovať grafy, analyzovať ich štruktúru a vykonávať rôzne analytické úlohy.

Výhodou aplikácie je najmä dostupnosť: open-source. Poskytuje tiež výkonné nástroje na vizualizáciu grafov, užitočné pre pochopenie komplexných vzorcov v dátach. Tiež podporuje analýzu váhovaných grafov, interaktívnu manipuláciu a podporu rôznych typov algoritmov.

Nevýhodou je pomalé spracovanie komplexných sietí. Ďalšou nevýhodou je tiež obmedzené spracovanie textových dát. Gephi tiež ponúka množstvo funkcií, no môže byť zložitá zorientovať sa v aplikácii a naplno využiť jej funkcionality z dôvodu nie veľmi prívetivého používateľského rozhrania.

1.3.2 Cytoscape

Cytoscape je open-source aplikácia určená na analýzu, vizualizáciu a interpretáciu komplexných sietí a biologických dát. Hoci pôvodne vznikla pre potreby bioinformatiky, najmä pre analýzu biologických sietí, dnes sa používa aj v mnohých iných oblastiach, ako sú sociálne siete, analýza textov, či chemické interakcie.



Výhodou aplikácie Cytoscope je podpora pre váhované a dynamické grafy, čo je užitočné pre analýzu sietí, kde sú interakcie rôzneho typu a intenzity. Podporuje tiež dynamické grafy, kde sa môže štruktúra siete meniť v čase. Ďalšou výhodou je pokročilá vizualizácia: Cytoscape ponúka pokročilé nástroje pre vizualizáciu grafov, vrátane možnosti prispôsobenia farieb, tvarov uzlov, hrán a rozloženia siete. To pomáha zvýrazniť dôležité vzory a štruktúry v dátach.

Nevýhodou aplikácie je obmedzené možnosti pre analýzu textov. Aplikácia Cytoscape nie je špecificky navrhnutá na analýzu textu, takže jeho možnosti v oblasti textového spracovania, ako je analýza sentimentu alebo klasifikácia textu sú obmedzené.

2. Motivácia a cieľ práce

3. Návrh riešenia

4. Implementácia

5. Výskum

Kapitola Výskum obsahuje výsledky práce. Zahŕňa grafy, porovnania štatistík a slovné zhodnotenie. Analýza bola vykonávaná na viacerých textoch, no najmä na dielach amerického autora, ktorý utrpel poškodenie mozgu. Prvý vstupný text je dielo *The Rat man of Paris*, ktoré napísal pre afáziou a následne druhé dielo *The shadow factory*, ktoré napísal po afázií.

5.1 Grafová analýza textov

5.1.1 Analýza textu Malý princ

V tejto časti popisujem výsledky grafovej analýzy vstupných textov. Analýza bola uskutočnená na slovnej sieti a následne na váhovanej slovnej sieti, vytvorenej z príslušného textu.

Tabuľka 5.1 obsahuje výsledky analýzy slovnej siete a následne váhovanej slovnej siete zo vstupného textu *Malý princ* v slovenčine.

vlastnosť	neváhovaná sieť	váhovaná sieť
počet uzlov	209	209
počet hrán	307	313
maximálny stupeň uzla	26	13
minimálny stupeň uzla	1	1
priemerný stupeň uzla	2.937799043062201	1.6057692307692308
klasterizačný koeficient	0.04249291878337754	0.0268
priemerná najkratšia cesta	4.512881854987119	7.2799

Tabuľka 5.1: Analýza slovnej siete - Malý princ

V tejto časti si môžeme porovnať výsledky a identifikovať rozdiely pri analýze rovnakého textu na neváhovanej a váhovanej slovnej sieti.

5.1.2 Analýza textov amerického autora

Tabuľka 5.2 obsahuje výsledky analýzy váhovanej slovnej siete zo vstupného textu *The Rat man of Paris* a textu *The shadow factory* v anglickom jazyku.

vlastnosť	<i>The Rat man of Paris</i>	<i>The shadow factory</i>
počet uzlov	8392	6557
počet hrán	36805	26510
maximálny stupeň uzla	2969	2045
minimálny stupeň uzla	1	1
priemerný stupeň uzla	7.033484270734032	6.347414976361141
klasterizačný koeficient	0.2671	0.2559
priemerná najkratšia cesta	2.92499	2.89893

Tabuľka 5.2: Analýza váhovanej slovnej siete textu *The Rat man of Paris* a *The shadow factory*

Záver

Literatúra

- [1] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Modeling the evolution of weighted networks. *Physical Review E*, 70(6), December 2004.
- [2] Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. Characterization and modeling of weighted networks. *Physica A: Statistical Mechanics and its Applications*, 346(1–2):34–43, February 2005.
- [3] Reinhard Diestel. Graph theory, volume 173 of. *Graduate texts in mathematics*, 7, 2012.
- [4] S. N. Dorogovtsev and J. F. F. Mendes. Language as an evolving word web, 2001.
- [5] Ramon Ferrer-i Cancho and Ricard Sole. The small world of human language. *Proceedings. Biological sciences / The Royal Society*, 268:2261–5, 12 2001.
- [6] Antoniou Ioannis and Tsompa Eleni. Statistical analysis of weighted networks, 2007.
- [7] Jozef Kubík. Analýza a tvorba jazykových grafov ľudí pred a po afázií na báze angličtiny. Master’s thesis, Universita Komenského v Bratislave, Bratislava, 2021.
- [8] Andrzej Kulig, Jarosław Kwapień, Tomasz Stanisław, and Stanisław Drożdż. In narrative texts punctuation marks obey the same statistics as words. *Information Sciences*, 375:98–113, 2017.
- [9] Tracy Love and Kathleen Brumm. *Language processing disorders*, pages 202–226. 12 2012.
- [10] William Matchin and Corianne Rogalsky. *Aphasia and Syntax*, pages 593–634. 09 2023.
- [11] Adilson E Motter, Alessandro PS De Moura, Ying-Cheng Lai, and Partha Dasgupta. Topology of the conceptual network of language. *Physical Review E*, 65(6):065102, 2002.

- [12] M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 70(5), November 2004.
- [13] Daniel Petriľak. Hľadanie príznakov afázie v textoch na báze slovenčiny. Master's thesis, Universita Komenského v Bratislave, Bratislava, 2023.

Prílohy