



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV AUTOMATIZACE A INFORMATIKY

INSTITUTE OF AUTOMATION AND COMPUTER SCIENCE

INTELIGENTNÍ PŘEDPOVĚĎ POČASÍ

INTELLIGENT WEATHER FORECAST

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Juraj Cigáň

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Daniel Zuth, Ph. D.

BRNO 2019

Zadání bakalářské práce

Ústav:	Ústav automatizace a informatiky
Student:	Juraj Cigáň
Studijní program:	Strojírenství
Studijní obor:	Aplikovaná informatika a řízení
Vedoucí práce:	Ing. Daniel Zuth, Ph.D.
Akademický rok:	2018/19

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma bakalářské práce:

Inteligentní předpověď počasí

Stručná charakteristika problematiky úkolu:

Práce se bude zabývat algoritmy pro počítačem generované texty, využitelné pro inteligentní předpověď počasí a následným automatickým publikováním na síti Twitter. Mohou být použité nástroje OpenWeatherMap a Thingspeak.

Cíle bakalářské práce:

- Rešerše v oblasti mikrokontrolérů.
- Rešerše v oblasti počítačem generovaného textu.
- Popsat API navrhovaných řešení.
- Vytvořit SW řešení.

Seznam doporučené literatury:

Current weather and forecast - OpenWeatherMap [online], [cit. 2018-10-25]. Dostupné z: <https://openweathermap.org/>

Internet of Things - MATLAB & Simulink [online], 1994. MathWorks [cit. 2018-10-25]. Dostupné z: <https://www.mathworks.com/solutions/internet-of-things.html>

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2018/19

V Brně, dne

L. S.

doc. Ing. Radomil Matoušek, Ph.D.
ředitel ústavu

doc. Ing. Jaroslav Katolický, Ph.D.
děkan fakulty

ABSTRAKT

Táto bakalárska práca sa zaoberá tvorbou algoritmu automaticky generovaných textov a ich konkrétnym využitím v počítačových predpovediach počasia. Je zameraná na zber poznatkov v spomenutých oblastiach a tiež rieši problematiku softvéru, ktorý bol použitý na dosiahnutie vytýčených cieľov. Prvá časť pojednáva o histórii vývoja digitálneho generovania a jeho využití v súčasnosti. Vymedzuje hlavné pravidlá a metódy potrebné pre zostrojenie podobného programu. Taktiež sa venuje zhromaždeniu potrebných informácií pre zostrojenie vlastnej meteorologickej stanice a opisuje dôležité vlastnosti softvéru, ktorý môže byť pri takomto procese použitý. Nakoniec, s ohľadom na rešerše vytvorené v predchádzajúcich kapitolách, opisuje postup tvorby vlastného programu, ktorý dokáže získať informácie o predpovedi počasia, spracovať ich a následne odoslať a zdieľať na sociálnej sieti Twitter.

ABSTRACT

This bachelor thesis focuses on the creation of an algorithm of automatically generated texts and their particular application in computer weather forecasts. It is aimed at gathering information from the fields mentioned above as well as about in the sphere of the software used to reach the set goals. The first part regards the history of development of digital generating and its present-day use. It specifies main rules and methods necessary to design such a programme. It also deals with the accumulation of required information for the designing of its own meteorological station and describes important features of the software, which can be used in such a process. Lastly, it also describes the creation of a programme which is able to gain and process information about weather forecasts and subsequently share them on social platform Twitter, all in consideration with the researches in previous chapters.

KLÚČOVÉ SLOVÁ

automatické generovanie textu, počítačová predpoveď počasia, zdieľanie na Twitteri, domáca meteostanica, OpenWeatherMap API, ThingSpeak

KEYWORDS

natural language generation, computer weather forecast, sharing on Twitter, home meteorostation, OpenWeatherMap API, ThingSpeak

BIBLIOGRAFICKÁ CITÁCIA

CIGÁŇ, Juraj. *Inteligentní předpověď počasí*. Brno, 2019. Dostupné také z: <https://www.vutbr.cz/studenti/zav-prace/detail/117085>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta strojního inženýrství, Ústav automatizace a informatiky. Vedoucí práce Daniel Zuth.

POĎAKOVANIE

Touto cestou by som chcel poďakovať rodičom za podporu počas štúdia, priateľke Linde za neutíchajúce povzbudenie a pomoc, kamarátovi Ondrejovi za konštruktívne postrehy a mnohým ďalším, bez ktorých by tieto riadky ťažko vznikali. Rovnako by som sa chcel poďakovať Ing. Danielovi Zuthovi Ph. D., za jeho odborné vedenie pri tejto bakalárskej práci, cenné rady a pripomienky.

ČESTNÉ PREHLÁSENIE

Prehlasujem, že táto práca je mojím pôvodným dielom, spracoval som ju samostatne pod vedením Ing. Daniela Zutha Ph. D. a s použitím literatúry uvedenej v zozname literatúry.

V Brne dne 1. 3. 2019

.....

Juraj Cigán

OBSAH

1	ÚVOD.....	15
2	DIGITÁLNE GENEROVANIE TEXTU.....	17
2.1	História a vývoj	17
2.2	Spôsoby generovania textu	18
2.3	Kombinačné generovanie textov	18
2.4	Automatické (prirodzené) generovanie textov	19
2.4.1	Spôsob fungovania algoritmu	19
2.5	Analýza textu pre automatické generovanie.....	19
2.6	Šablóny v metóde automaticky generovaných textov	22
2.7	Obmedzenia a upozornenia	23
3	POČÍTAČOVÁ PREDPOVEĎ POČASIA	25
3.1	Textová interpretácia dát	25
3.2	Výber slov na použitie v predpovedi	26
4	MIKROKONTROLÉRY V PREPOVEDIACH.....	27
4.1	Typy mikrokontrolérov.....	27
4.2	Programovateľnosť	28
4.3	Modul pre získavanie dát počasia.....	28
5	SOFTVÉR A PROSTREDIA	31
5.1	MatLab.....	31
5.1.1	Vlastnosti MatLabu	31
5.1.2	Použitelnosť MatLabu.....	31
5.1.3	Thing Speak	32
5.2	OpenWeatherMap.....	32
5.2.1	API – application program interface	33
5.3	Twitter	34
5.3.1	Formy zdieľania na Twitteri	34
6	REALIZÁCIA PREDPOVEDI A ZDIEĽANIA.....	37
6.1	Analýza textu.....	37
6.1.1	Plánovanie textu	38
6.1.2	Mikro-plánovanie	39
6.1.3	Povrchová realizácia.....	44
6.2	Tvorba algoritmu	45
6.2.1	Získanie a odovzdanie informácií.....	46
6.2.2	Automatické generovanie textu	47
6.3	Výsledok generovania	48
7	ZÁVER	49
8	ZOZNAM POUŽITEJ LITERATÚRY.....	51
9	ZOZNAM SKRATIEK, SYMBOLOV, OBRÁZKOV A TABULIEK.....	53
10	ZOZNAM PRÍLOH.....	55

1 ÚVOD

Nové storočie prinieslo so sebou množstvo nových možností práce s informáciami. Táto revolúcia spôsobila zásadné zmeny v chápaní úlohy digitálnych technológií v súkromnom živote a v profesionálnom živote. Počítače prenikli doslova na každý krok a ich neustály vývoj nahradzuje človeka v čoraz zložitejších úlohách. Inak tomu nie je ani v oblasti spracovania a analýzy informácií. Už niekoľko desiatok rokov sú skúmané možnosti zapojenia počítačových programov do tvorby súvislých a zmysluplných textov, pričom tento cieľ je čoraz viac dosiahnuteľný a dosahovaný. Množstvo voľne dostupných algoritmov a postupov sa už dnes v určitých oblastiach snaží vyrovnat' ľudskej inteligencii. Rôzne postupy a metódy dokážu „naučiť“ počítač vytvoriť diela, ktoré by sa len ťažko odlišovali od tých, ktoré pochádzajú od človeka.

Takisto však môžeme pozorovať zvýšený nárast informácií, ktorý sa k nám dostane počas bežného dňa. Sú to dáta, ktoré často nepotrebujeme alebo ich preberieme vo forme, ktorá vyžaduje ďalšie spracovanie. To znamená vynaloženie ďalšej energie a času na ich interpretáciu. Podobné procesy sa však v určitej miere dajú nahradiť vhodným využitím umelej inteligencie a odbremenit' sa tak od niektorých nadbytočných úkonov.

Zvážením týchto dvoch bodov sa nám vytvára spoločenská objednávka premýšľať nad zhotovovaním systémov, ktoré budú mimoriadne uľahčovať život svojimi funkciami. Proces spracovania dát môže byť užitočný v rôznych smeroch. Či už je to zjednodušenie každodenného života alebo špecializovaná pomoc pre ľudí, ktorí to nevyhnutne potrebujú.

Táto bakalárska práca pojednáva o spomenutých možnostiach a úskaliach. Jej cieľom je načrtnúť efektívne využitie digitálnych technológií v súčasných podmienkach. Pojednáva o zjednodušení databázových dát predpovedí počasia algoritmom, vytvoreným konkrétne pre túto úlohu. Informácie potom podáva v príjemnej forme, ktorá svojou zložitou pripomína bežný priateľský rozhovor. Následne ukazuje, že je možné prakticky využiť sociálnu sieť na zdieľanie aj takejto formy obsahu. Tieto méty sú pritom dosahované službami, z ktorých väčšina je dostupná pre používateľov zdarma. Radíme sem OpenWeatherMap s možnosťou získavania počasia a ThingSpeak, ktorý pochádza od výrobcu MatLabu, s jednotlivými jeho funkciami, ktoré dokážu rôzne formy dát spracovať.

2 DIGITÁLNE GENEROVANIE TEXTU

Výborným ukazovateľom schopnosti počítača čeliť problémom a prichádzať s riešeniami na úrovni mysle človeka je Turingov test. Prvýkrát sa skúška „ľudskosti“ počítača uskutočnila v roku 1951 a fungovala na báze Imitačnej hry (*Imitation game*) a je využitelná dodnes. Skúšajúci dostane dvoch respondentov. Jeden z nich je človek a druhý naprogramovaná umelá inteligencia. Medzi skúšanými a skúšajúcim prebieha komunikácia, ktorá je výlučne textová. Jej úseky tvoria časové intervaly, ktoré majú maximálne 5 minút a opakujú sa niekoľkokrát. Rozmedzie tém určených na konverzáciu je neobmedzené. Test sa pokladá za úspešný, pokiaľ skúšajúci alebo hodnotiaca komisia nevie rozlíšiť, ktorý z odpovedajúcich je robot, a ktorý ľudská bytosť. Dôležitá hranica sú traja hodnotitelia z desiatich, čiže aspoň 30 % musí odpoveď hodnotiť ako nerozlišiteľnú.

Prvý úspešný riešiteľ tohto testu je počítačový program nazvaný Eugene Goostman, vytvorený v Petrohrade programátorom Vladimirom Veselovom. Program simuloval komunikačné zdatnosti trinásťročného chlapca a jeho úspešnosť bola 33% (1 z 3 ľudí komisie hodnotil odpovede ako nerozlišiteľné). Nie však celá vedecká obec má rovnaký názor na kvalitu Turingovho testu. Je dôležité poznamenať, že Turingov test nedokazuje reálnu inteligenciu, ale schopnosť ju imitovať, čo v mnohých oblastiach generovania textu postačuje. [1, 2]

2.1 História a vývoj

Korene generovaného textu siahajú do dávnej minulosti. Nachádzame tu rôzne príklady modelov systémov rôznej úrovne zložitosti. Prvé pokusy, založené na náhodnom výbere častí viet alebo slov, sa datujú do obdobia okolo roku 1200-1300, kedy arabskí astrológovia používali modely pohybujúcich sa stĺpcov na predpovedanie budúcnosti.

V 16. storočí experimentoval nemecký básnik George Philipp Harsdörffer s hracími kockami a kúskami dreva, čím sledoval tvorbu anagramov. Tým vytvoril predpoklad pre svoj zdokonalený systém náhodného generovania slov, ktorý pomáhal pri inšpirácii jeho literárnej činnosti. Spočíval v zostrojení piatich sústredných kruhov s písmenami a fragmentami slov napísanými na nich. Na prvom kruhu z vnútra sa nachádzali predpony, na druhom začiatkové písmená, tretí obsahoval prostredné písmená, štvrtý koncové a na koniec pridal prípony. Zatočením všetkých alebo len častí kruhov dostal rôzne slová alebo ich skomoleniny. Podobných pokusov možno nájsť v minulosti viacero. [3]

Skutočný rozmach na inej úrovni zložitosti, či už po obsahovej, tak i po technickej stránke, začal v päťdesiatych rokoch 20. storočia. Spočiatku generovanie jazyka nevyžadovalo príliš abstrakcie, zamerali sa skôr na preklad jednotlivých slov alebo slovných spojení medzi dvoma odlišnými jazykmi. Sedemdesiate a osemdesiate roky priniesli prvé akademické články, pojednávajúce o problematike generovania

prirodzeného jazyka. Boli zdefinované prvé prekážky a problémy, ktoré bolo potrebné preskúmať v tomto odbore. Medzinárodný workshop, ktorý sa konal v roku 1983, sa stal dôležitým míľnikom v histórii automaticky generovaných textov. Položil základy výskumu v odbore a zároveň vzbudil záujem, ktorý sa v deväťdesiatych rokoch ešte viac prehĺbil. Rozšírila sa podpora workshopov, projektov a prác v rôznych oblastiach informatiky a počet dostupných informácií začal exponenciálne rásť.

Každý párny rok sa koná konferencia NLG (International Natural Language Generation Conference), ktorá sa venuje práve tejto problematike. Organizujú sa tu diskusie, šíria a archivujú sa výsledky danej oblasti. Stará sa o to skupina SIGGEN (Special Interest Group on Generation) patriaca do asociácie ACL (Association for Computational Linguistics). [4]

2.2 Spôsoby generovania textu

Náhrada cieľavedomej práce ľudskej inteligencie v oblastiach umenia a kreativity je jedným z najväčších problémov výskumu digitálnych technológií, a to z veľkej časti preto, že popísať ich princíp pomocou inštrukcií určených pre počítač, je veľmi namáhavé.

Rovnako je to aj pri generovaní textových diel. Tie totižto musia obsahovať viac ako iba náhodne poskladané vety do súvislého celku. Dôležitá je zápleтка, práca s pointou, zmysluplné konanie a koniec koncov musí dávať výsledok zmysel a nevyzerat' umelo. Pre dosiahnutie požadovaného výsledku je, podobne ako v iných oblastiach informatiky, možné využiť viacero typov metód. Medzi dve najhlavnejšie radíme kombinačné, založené na šablóne (Combinatory Text Generation) a prirodzené (Natural Language Generation) generovanie textov.

2.3 Kombinačné generovanie textov

Tento typ generovania textu využíva lingvistické štruktúry existujúcich viet. Algoritmus spočíva v postupnej obmene ich súčastí za vzniku nových, odlišných viet s iným významom, ale rovnakým vetným rámcom.

Využívame tu metódy tvorby permutácií, kombinácií a dopĺňovanie do voľného políčka. Poznáme tri hlavné metódy kombinačného generovania textov. Pokiaľ konkrétne elementy zamieňame za nové alebo využívame variácie, hovoríme o permutačných metódach. Kombinačná metóda spočíva vo vytvorení špeciálneho zoznamu slov, z ktorého neskôr vyberáme v kontrolovaných alebo náhodných kombináciách jednotlivé požadované slová. Posledná metóda je tiež kombinačná, s rozdielom, že využíva dopĺňovanie slov do syntaktických šablón, čo spôsobuje, že výsledný text je gramaticky a syntakticky bezchybný. [5, 8]

Takéto procedúry možno použiť k tvorbe krátkych textov ako sú básne alebo príslovia, pretože program ovplyvňuje iba skladbu vety, nepracuje však s jej skutočným

významom. Príkladom využitia pravidiel v praxi je ‘*phrase á trous*‘ algoritmus, ktorý použil Oulipo. Pracuje na princípe vymieňania podstatných mien vety inými podstatnými menami získanými zo synonymického slovníka. Vyžitie tohto typu generovania textu môžeme nájsť aj v práci autora Thea Lutza. Vytvoril dielo zvané „Stochastické texty“ (nemecky *Stochastische Texte*, ktoré pozostávalo iba zo slov a fráz z novely Franza Kafku Hrad (*The Castle*).

Táto metóda bola najpoužívanejšou v oblasti digitálnej poézie v osemdesiatych rokoch 20. storočia. [6]

2.4 Automatické (prirodzené) generovanie textov

Porozumenie prirodzenému jazyku je, dá sa povedať, recipročný proces k jeho generovaniu. Zmysluplné čítanie textu zväčša znamená snahu porozumieť tomu, čo chcel autor svojim dielom povedať. Túto snahu teda sprevádza myslenie, ktoré funguje na princípe hľadania abstraktných pojmov a vytvárania prepojení medzi nimi. Slová sú v myšlienkach reprezentované všeobecnými pojmami a ich tvar (v mnohých jazykoch ohýbaný a skloňovaný) špecifikuje charakter pojmov.

Generovanie textu funguje opačne. Program je vytvorený tak, aby dosiahol svoj komunikačný cieľ, čo predstavuje predanie nejakej informácie pomocou sformulovaných slov alebo vyšších celkov. Používa abstraktné štruktúry, ktoré prebiehajú procesom spresnenia, usporiadania a konečného formulovania. Oba procesy, generovanie i porozumenie, sú hlavným nástrojom komunikácie, preto je ich cieľom efektívne zdieľanie informácií. [4]

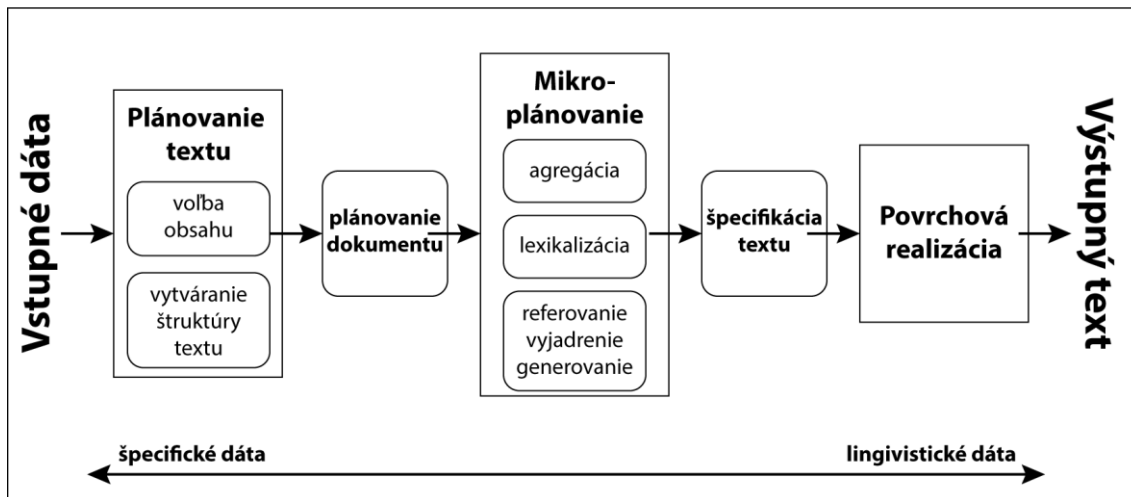
2.4.1 Spôsob fungovania algoritmu

Najbežnejšou cestou k uspokojivému výsledku prirodzeného generovania textu je určiť si cieľ výstupu, čiže predstavu, ako by mala požadovaná veta vyzeráť po obsahovej a formálnej stránke. Následne plánujúci sektor algoritmu vytvorí špecifikácie vety použitím získaných dát opisujúcich výstup. Rozhodne o výbere vhodných obsahových a lexikálnych jednotiek, ktoré ovplyvňujú konečný spôsob výstupu. Tento výber je poskytnutý poslednému prvku procesu tvorby vety, ktorý vezme vytvorený základ a s využitím gramatických pravidiel prirodzeného jazyka vetu dokončí. Obširnejšie preberáme postup tvorby algoritmu a spôsob jeho fungovania v nasledujúcej kapitole. [9]

2.5 Analýza textu pre automatické generovanie

Jeden z hlavných objektov záujmu pri vytváraní algoritmu automatického generovania je textový korpus, ktorý sa v tejto oblasti považuje za zdroj hlavných informácií pre celý proces. Existuje však niekoľko spôsobov ako získať požadované výsledky. Jedným z nich je postup, ktorý sa skladá z analýzy vhodného korpusu a následnej práce so získanými informáciami.

Tvorba automaticky generovaných textov z nelingvistických vstupných dát pozostáva z troch hlavných fáz: plánovanie textu, mikro-plánovanie a povrchová realizácia. Najjednoduchšia architektúra realizácie tohto modelu je tzv. zreťazené spracovanie, kde každý z hlavných procesov spotrebúva alebo/a vytvára prechodnú reprezentáciu: z netextového vstupu prostredníctvom dokumentového plánu a textovej špecifikácie, až na výstupný text (obr. 1). Dôležitým argumentom pre zváženie použitia týchto troch fáz (namiesto iba plánovania a iba realizácie) je typ znalostí, aký je použitý v každej z nich. [10]



Obr. 1: Hlavná architektúra automaticky generovaných textov. [vlastné spracovanie podľa 10]

Plánovanie textu. Na úplnom začiatku je potrebné si vytvoriť štruktúru výsledného textu, čo predstavuje dôslednú voľbu obsahu a jeho následné zasadenie do logických súvislostí. Všetko je v tomto bode zatiaľ vytvárané bez ohľadu na konkrétny tvar viet a gramatické detaily. V určitom zmysle je táto fáza najdôležitejšia, pretože určuje, aké informácie sa budú vyskytovať vo výslednom texte. Pre cieľavedomý výber dát je nutné zhodnotiť oblasť, odkiaľ bude algoritmus čerpať, a tiež cieľových užívateľov, pre ktorých bude text určený - teda celkový účel textu. Ide o veľmi špecifický proces zameraný na prehodnotenie využitia vygenerovaného výsledku. Vstup pre fázu plánovania textu sú všeobecne štyri veci:

1. Zdroj informácií tvorený databázou. Predstavuje najdôležitejšiu zložku, z ktorej algoritmus čerpá.
2. Komunikačný cieľ, ktorý má byť dosiahnutý. Ohraničuje účel generovaného výsledku a úroveň jeho splnenia je v určitých prípadoch merateľná.
3. Užívateľský model, určujúci koncovú skupinu ľudí, pre ktorých je výsledok určený.
4. História, čo znamená, že disponujeme súborom predošlých úspešne vytvorených textov, ktoré sa týkali podobnej témy a je žiadúce, aby sa im text podobal.

Výstupom prvého kroku je **plán dokumentu**. Jeho štruktúra je špeciálne navrhnutá s osobitým prístupom pre konkrétny prípad a pre konkrétnu technológiu, ktorá bola v našom prípade využitá. Väčšinou má vzhľad stromovej štruktúry. Rôzne metódy sa líšia úrovňou abstrakcie a tiež usporiadaním stromu v jeho vnútornej štruktúre. Niekedy môžu byť časti vyjadrovanej informácie iba listy stromu a uzly tu definujú ich vlastnosti, kdežto v iných prípadoch nesú podobný typ informácií aj uzly, aj listy. Jednotlivé vetvy stromu potom reprezentujú správy, ktoré majú byť odovzdané koncovému používateľovi s najvyššou mierou abstrakcie, teda bez ohľadu na ich konkrétne formulácie. Táto štruktúra nie je závislá na jazyku, v ktorom bude výsledok vytvorený.

Nasledujúcou fázou procesu je **mikro-plánovanie**. Spočíva vo vytváraní abstraktného stromu, ktorý prepája jednotlivé lexikálne časti so syntaktickými štruktúrami. Správy, popísané v pláne dokumentu, sú následne prevedené do sekvencie špecifikovaných viet alebo fráz, a tento prevod je úplne závislý od jazyka interpretácie.

Správy, ktoré vzniknú, môžu byť individuálne opísané vlastnými vetami alebo ich možno spojiť do jednej vety či súvetia. Takýto proces sa nazýva **agregácia**. Jej zmyslom je vyhodnotiť všetky možné metódy realizácie a na základe toho vybrať takú, ktorá bude s ohľadom na vzájomné súvislosti správ najpriateľnejšia. Rozhodovanie je ovplyvnené konvenciami a pravidlami pojednávaneho žánru a taktiež vychádza z už hotového plánu dokumentu – napríklad sa agregujú iba tie vety, ktoré majú v stromovej štruktúre rovnakého najbližšieho rodiča.

Výber konkrétnych slov, ktoré sú vhodné na použitie pri vytváraní a zdieľaní obsahu, sa nazýva **lexikalizácia**. [4] Inak povedané, je to proces voľby vhodných jazykových prostriedkov pre realizovanie predikačnej časti správy. [10] Toto rozhodovanie môže prebehnúť prostredníctvom šablóny, pomocou výberových stromov alebo metódami prepisovania grafov. Je potrebné brať ohľad na znalosti a preferencie adresáta správy a na fungovanie slov vo vzťahu s ich použitím s ostatnými slovami textu. Taktiež treba zväziť využitie rovnakých výrazových prostriedkov v predošlých častiach textu. Dôležitý je dohľad nad úrovňou zrozumiteľnosti a presnosti celej správy, a to v súvislosti so zamedzením zdieľania nepráv a polopráv.

Jav, ktorý sa bežne využíva v ľudskej reči a vyžaduje vlastnú časť programu na spracovanie, sa nazýva **referovanie**. Spočíva v používaní zámen alebo iných odkazov na objekt, o ktorom momentálne text pojednáva. K tomuto dochádza, keď veta hovorí o entite, ktorá sa už v texte nachádzala napríklad v predošlej vete a nie je možné ju zameniť s iným objektom. Vtedy je na mieste nahradiť pomenovanie objektu zámenom (kráľ Ondrej – on), skrátenou formou jeho mena (Ondro), prípadne nájsť inú formu vyjadrenia (panovník). [4]

Ďalším princípom v tejto fáze tvorby automatického textu je **generovanie**. Tento proces možno prirovnať k fungovaniu komunikácie medzi ľuďmi. Každý človek predtým ako vytvorí vetu, musí skombinovať niekoľko správ dokopy. Realizácia vety je podmienená poznaním, aká správa bude vytvorená a s akými syntaktickými konštrukciami bude pracovať. [10]

Výstupom celej fázy mikro-plánovania je špecifikácia textu, čiže strom slov v základnom tvare s reprezentáciami súvislostí medzi nimi. V tomto bode je text skoro hotový, ostáva už len gramatická úprava a formátovanie viet.

Povrchová realizácia. Rozdiel medzi touto a predošlými dvoma fázami spočíva v tom, že povrchová realizácia je vždy oddelená od zbytku programu. Jej tvorba prebieha nezávisle, a to preto, aby nebola ovplyvnená gramatickými detailmi konkrétneho používaného jazyka. Ako vstup slúži strom syntaktických štruktúr, ktoré možno previesť na sekvenciu slov v základnom tvare. Výstupom je potom niekoľko hotových, gramaticky správnych viet. Táto fáza teda spočíva v lingvistickej realizácii, ktorá má za úlohu vyberať a následne vkladať funkčné slová do viet, zaistiť ich správne skloňovanie, usporiadanie a dodržanie pravopisu. Medzi hlavné techniky riešenia radíme obojstrannú špecifikáciu gramatiky a špecifikáciu gramatiky prispôbenú pre generovanie založené na šablónach.

Obojstranná špecifikácia je opísaná ako súbor logických prepojení medzi syntaktickými (vetnými) a sémantickými (významovými) súčasťami textu. Tento spôsob je teoreticky veľmi elegantný a frekventovane sa využíva v programoch zabezpečujúcich strojový preklad. Na generovanie prirodzenej reči sa príliš nehodí, pretože vstup povrchovej realizácie je už sčasti spracovaný a nemá takú formu, ktorá by obsahovala iba sémantický význam textu.

Prispôbená gramatika pre vytváranie konkrétneho typu textu obsahuje rad pravidiel, ktoré boli vymyslené pre adresné riešenia a je teda neprenosná. To znamená, že jej použitie sa obmedzuje iba na jeden určitý prípad. Zvážením typu a obsahu vstupov sa rozhoduje o najefektívnejších riešeniach pre realizáciu. Často nie je potrebné vytvárať sofistikované algoritmy, ktoré by riešili všetky syntaktické pravidlá konkrétneho jazyka, ale generovanie si vystačí aj so šablónou alebo vytvorením niekoľkých nekomplikovaných princípov. [4]

2.6 Šablóny v metóde automaticky generovaných textov

Niektoré systémy na tvorbu automaticky generovaného textu teda nepracujú na princípe syntaktických väzieb, namiesto toho sa voľba obsahu sústreďuje na špecifikáciu textu pomocou šablóny. Takýto prístup spôsobuje pár odlišností v analýze textu:

- Voľba obsahu a plánovanie dokumentu pokračuje rovnako, ako bolo spomenuté vyššie, avšak výstupom je textový plán, ktorého listy (stromová štruktúra) sú štruktúrované do predom určenej formy.
- Pri agregácii viet nie je možné použiť niektoré vetné štylistické metódy, ako napríklad elipsu (výpustka). Dôvodom takejto zmeny je neschopnosť šablón špecifikovať vetné súvislosti na dostatočne abstraktnej úrovni tak, aby bola dovolená takáto manipulácia.
- Lexikalizácia väčšinou neprebieha, namiesto nej však systém voľby obsahu pomáha zvoliť šablónu, ktorá obsahuje vhodné slová pre

zachovanie rôznorodosti textu. V niektorých prípadoch je potrebné zvoliť viacero predpripravených foriem automatického generovania pre tú istú správu a vytvoriť mechanizmy, ktoré budú medzi nimi rozhodovať.

- Referovanie a jeho výrazy môžu byť pri takomto type generovania nevhodné, pretože algoritmus nevie vyhodnotiť, aké entity boli spomenuté v konzervovanej časti šablóny.
- Pri povrchovej realizácii treba dať pozor na skloňovanie a ostatné aspekty morfológie a ortografie. Neprebíha tu však žiadne syntaktické spracovávanie textu.

2.7 Obmedzenia a upozornenia

Skupina výskumníkov v danej oblasti - Reiter a kolektív, v roku 2000 popísali niekoľko techník, ktoré zabezpečovali získanie potrebných informácií pre použitie v systémoch automatického generovania textu. Jedna z nich je práve metóda spočívajúca v analýze korpusu textu napísaného človekom, ktorou sa táto kapitola zaoberá. Upozorňujú však aj na nástrahy tejto techniky. Varujú pred:

- zlou interpretáciou dát získaných z korpusu,
- výlučným spoliehaním sa na získavanie informácií iba z korpusu,
- výberom vhodných kritérií pri výbere správneho korpusu. Je vhodné použiť metódu výberu, kde kvalitatívne vlastnosti majú väčšiu váhu ako tie kvantitatívne, čo znamená, že úroveň primeranosti zdroja informácií nie je vhodné súdiť podľa jeho rozsiahlosti, ale podľa obsahu.

3 POČÍTAČOVÁ PREDPOVEĎ POČASIA

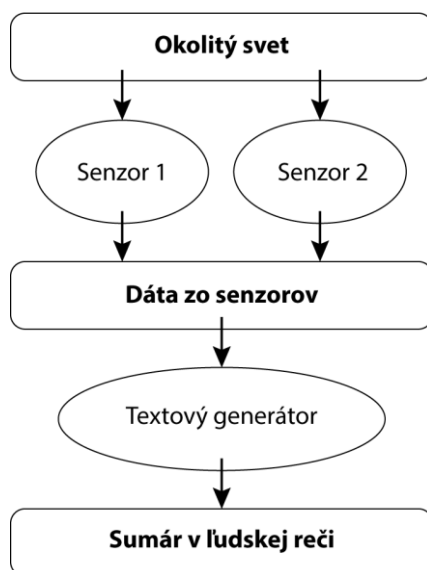
Moderné predpovede počasia sú stavané na numerických predpovediach, čo predstavuje masívne simulácie atmosféry spúšťané na superpočítačoch. Výstupy takýchto modelov sú súbory meteorologických parametrov ako rýchlosť vetra, teplota, zrážky a podobne, a to pre určité lokácie v určitý čas.

Organizácie zaberajúce sa predpoveďami počasia následne prijímajú tieto dáta a modifikujú si ich vzhľadom na svoje znalosti a skúsenosti, čo znamená na základe predošlých nepresností spôsobených podcenením určitých údajov superpočítačom. Výstupom sú informácie zahŕňajúce numerické údaje pre určité oblasti, ktoré zaujímajú konečných adresátov. Tí ich potom vo forme, v akej sú im poskytované, používajú pri rozhodovaní v rôznych oblastiach života.

Prvý krok pri manipulácii s hotovými dátami je cieľavedomý výber tých, ktoré sú svojou dôležitosťou vhodné pre adresáta. Napríklad pilot, ktorý pristáva, musí byť oboznámený s aktuálnou viditeľnosťou a rýchlosťou vetra na území koncového letiska. Druhý krok je vytvoriť sumár týchto dát, ktorý môže byť textový, tabuľkový, grafický alebo iný vzhľadom na zákazníkove požiadavky.

3.1 Textová interpretácia dát

Žijeme v modernom svete, ktorý je denne zaplavovaný obrovskými množstvami informácií a porozumieť každej z nich nebýva ľahké. Často požívaný systém, ktorý získava dáta o okolitom svete pomocou senzorov (obr. 2), je možné využiť napríklad na sledovanie plynovej turbíny. Tá vyprodukuje 200 MB dát za deň, ktoré zahŕňajú jednu hodnotu za sekundu z 250 senzorov, ktoré súčasne merajú teplotu, vibrácie, prietok paliva, hodnoty síl, atď.



Obr. 2: Prepojenie jazyka so svetom prostredníctvom senzorov. [vlastné spracovanie podľa 11]

Inžinier údržby má, povedzme, jednu hodinu (denne) na porozumenie tomuto množstvu údajov. Takisto doktor na oddelení intenzívnej starostlivosti je zahlcovaný megabajtami údajov o svojich pacientoch (tep, tlak srdca, atď.), keď potrebuje určiť diagnózu a od nej sa odvíjajúci typ liečby. Spracovať všetko, čo je potrebné (a to často pod časovým tlakom) môže byť veľmi vyčerpávajúcou úlohou. Informácie sú všadeprítomné a ovplyvňujú veľa dôležitých rozhodnutí, preto sa vytvára dopyt po efektívnom spôsobe ich spracovania a prezentácie.

Tu sa dnes najčastejšie používajú numerické (tabuľkové) alebo grafické (vizualizačné) metódy. Pokiaľ sa zamyslíme nad teoretickou podstatou tvorby textových a grafických prezentácií informácií, zistíme, že druhá spomenutá má obrovskú výhodu, pretože môže byť generovaná automaticky (v konečnom dôsledku lacno), kdežto textová forma musí byť napísaná človekom (čo znamená draho). Preto sa v súčasnej dobe stále viac hľadajú možnosti, ako tento spôsob tvorby textových dát zautomatizovať, a tým zatriktívniť jeho používanie.

3.2 Výber slov na použitie v predpovedi

Kľúčovým aspektom prepojenia jazyka s okolitým svetom (obr. 2) je vyberanie správnych slov, ktoré vyjadria lingvistické dáta. To je hlavná časť problému, ktorá spočíva v určení, ako majú lingvistické konštrukcie súvisieť s nelingvistickými informáciami.

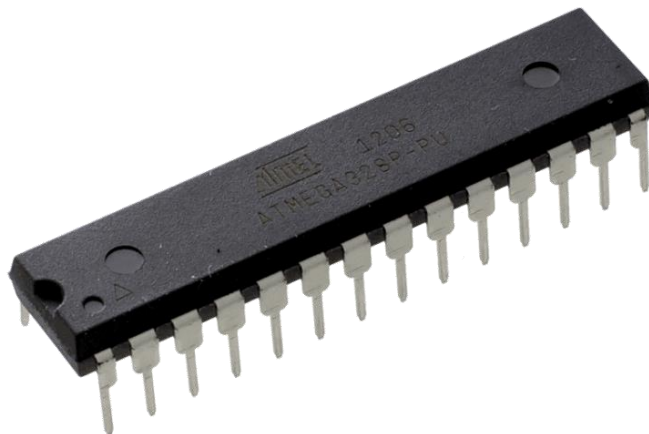
Nedávny výskum na tému využitia jazyka ukázal, že väčšina ľudí má zafixované podobné asociácie slov a ich významov, tie však v určitých konkrétnych prípadoch nemusia nevyhnutne platiť. Napríklad jeden človek si môže predstaviť ružovú farbu inak ako druhý. To znamená, že predmet opísaný rovnakou farbou, nemusí v mysliach dvoch ľudí vytvoriť rovnakú predstavu, a to aj za rovnakých svetelných podmienok. Takéto odlišnosti sú dôkazom individuálneho chápania jazyka a sú problémom pre softvér generujúci prirodzenú formu jazyka. Otázkou teda zostáva, ako generovať texty, ktorých kontext bude interpretovaný rovnako aj za použitia iných výrazových prostriedkov.

Generovanie prirodzeného textu preto vyžaduje niekoľko dôležitých rozhodnutí v súvislosti s uspokojiteľným porozumením obsahu predpovede. Toto môže pozitívne ovplyvniť výber správnych slov pre vyjadrovanie. Väčšinou sa naskytá možnosť vyberať z niekoľkých podobných označení. Napríklad spôsob zapísania:

- smeru vetra - môže byť zapísaný ako Z alebo západný;
- rýchlosti vetra - 8 km/h môže byť zapísaných ako 8 alebo 08;
- slovesa – vietor, ktorý prestal fúkať, možno opísať slovami ustať, ustúpiť;
- časovej frázy – používať do večera alebo do neskorého večera; vyjadriť polnoc slovom alebo časom 00:00.

4 MIKROKONTROLÉRY V PREPOVEDIACH

Mikrokontrolér (skratka μC , uC alebo MCU) je v podstate zmenšenina počítača, ktorého funkcie sú integrované do jedného malého čipu (obr. 3), z čoho vyplýva aj jeho názov: jednočipový mikropočítač. Tak ako obyčajný osobný počítač, aj mikrokontrolér disponuje pamäťou, vstupno-výstupnými a komunikačnými obvodmi. [13]



Obr. 3: Znárodnenie frekventovaného výzoru samotného mikrokontroléru. [22]

Často ho nájdeme v riadiacich zariadeniach, kde väčšinou plní špeciálnu funkciu, pre ktorú bol naprojektovaný a vyrobený (na rozdiel od osobného počítača, ktorého využitie je univerzálne). Mikrokontroléry môžu obsahovať niekoľko periférnych obvodov, ako napríklad AD prevodník, USB, PWM a podobne. Je možné ich zapájať z rôznych prenosných zdrojov napätia ako je batéria, akumulátor alebo fotočlánky. Sú veľmi rozšírené, nájdeme ich v množstve zariadení okolo nás. V podstate sa nachádzajú v každom aspoň sčasti inteligentnom prístroji, čo znamená, že pre svoju činnosť nevyžadujú nevyhnutný zásah človeka. Môžu riadiť ohrievanie vody v bazénoch, ich zakrývanie v prípade zlých poveternostných podmienok, polievanie záhrad a trávnikov alebo vytvoriť internetový server. Nájdeme ich v hračkách, USB kľúčoch, prehrávačoch, medicínskych prístrojoch, práčkach, kuchynských spotrebičoch a podobne.

4.1 Typy mikrokontrolérov

Poznáme niekoľko typov a skupín mikrokontrolérov. Najznámejší a najstarší je mikrokontrolér 8051 od dlhoročného výrobcu čipov Intel, ktorý bol uvedený na trh už v roku 1980. V licencií ho okrem Intelu vyrábali aj ďalšie známe firmy ako Phillips, Dallas Semiconductors, Atmel, Winbond a iné. Známu skupinou mikrokontrolérov sú aj modely od firmy Microchip. Veľmi známy je typ PIC16F84, ktorý našiel uplatnenie hlavne medzi študentmi stredných a vysokých technických škôl. Tretiu skupinu tvoria čipy od firmy Atmel, ktorá koncom deväťdesiatych rokov vytvorila rad AVR. Devízou

tohto typu bolo ľahké programovanie v jazyku C. Veľkú popularitu nadobudli hlavne vďaka pomerne vysokému výkonu, prijateľnej cene a veľkému počtu variantov s množstvom zabudovaných periférií. Uplatnenie našli medzi konštruktérmi elektroniky, priemyselných systémov a v neposlednom rade medzi domácimi technikmi. Samozrejme, poznáme aj ďalšie typy týchto zariadení od firiem ako Freescale (predtým Motorola), Acrom, Zilog a iné.

Na trhu sú momentálne predávané samostatne alebo ako kity, ktoré obsahujú aj iné súčiastky potrebné na prácu pri riešení konkrétnych problémov. Takýto kit je napríklad Arduino UNO od firmy Atmel. [13]

4.2 Programovateľnosť

Ruka v ruke s napredovaním miery dômyselnosti mikrokontrolérov sa vyvíjali aj rôzne nástroje na ich programovanie. Spočiatku prebiehalo programovanie (a ešte stále v určitých prípadoch prebieha) v jazyku *assembler*. Existujú však aj efektívnejšie nástroje, ako napríklad PIC Basic Compiler pre mikrokontroléry typu PIC.

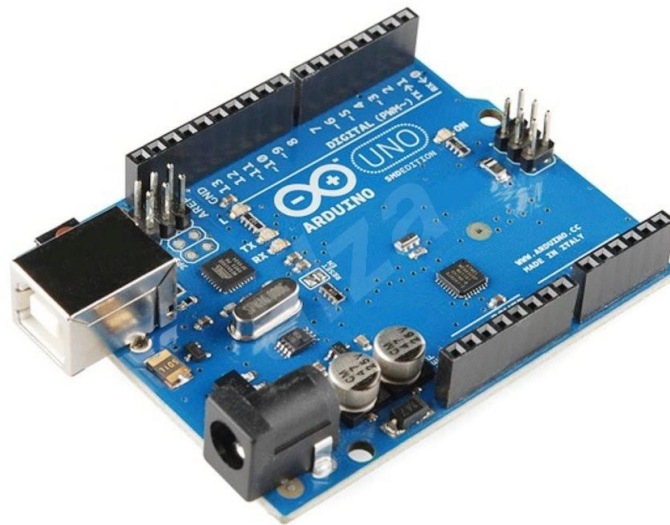
Nevýhodou tohto nástroja je jeho komerčnosť, a z toho vyplývajúca cena nákupu. V poslednom čase sa rozšírilo programovanie v jazykoch C a C++, nielen v prostrediach mikrokontrolérov, ale aj na osobných počítačoch. Ide o obľúbený programovací jazyk, ktorý jeho výrobca sprístupnil pre všetkých zadarmo. Možnosť programovania ponúkajú aj vyššie programovacie jazyky, ako Pascal, tie však nemajú takú mohutnú podporu užívateľov. [13]

4.3 Modul pre získavanie dát počasia

Stanice, ktoré sa zaoberajú zberom holých dát počasia sú veľké, senzitívne a spoľahlivé systémy. Kľúčovým dôvodom dopytu po takýchto zariadeniach je hlavne už spomenutý problém plánovania. Väčšina letísk, leteckých spoločností a iných transportných sústav potrebuje vopred pripravovať rozvrh trás, ktoré úzko súvisia s klimatickými podmienkami. Na druhú stranu, takéto systémy nie sú lacné, a to hlavne v prípadoch, kedy je potrebné periodické skenovanie a dohľad nad veľkými mestami a ich počasím.

Jednoduchým riešením meteorologického problému môže byť využitie mikrokontrolérov na zber a čiastočné spracovanie dát. Rýchlou prosperitou a prevahou na trhu hral dôležitú úlohu v tomto odvetví mikrokontroler Arduino UNO (obr. 3), ktorý je možné vylepšiť mnohými doplnkovými zariadeniami, a to s cieľom rozšíriť možnosti jeho využitia. Vybavený DHT senzorom dokáže merať teplotu a vlhkosť ovzdušia a senzor pre meranie rýchlosti vetra vie merať rovnomennú veličinu. LDR senzor je používaný pri získavaní dát o intenzite svetla dopadajúceho na zariadenie. Postup pri práci s týmito veličinami je nasledovný: Arduino, pripojené k týmto senzorom najskôr získa požadované informácie, ktoré dokáže priamo zobrazovať na LCD displeji alebo ich ukladať do súboru MS Excelu. Následne sú dáta analyzované a ukladané do tabuliek

prostřednictvím R-jazyka. Výsledkem takéhoto procesu je přehledná databáze s uspořádanými hodnotami jednotlivých veličin, kterou možno použít při různých spracovaniach dát, vizualizáciách, správach a podobne. [14]



Obr. 4: Zobrazenie hardvéru vývojovej dosky Arduino UNO. [12]

5 SOFTVÉR A PROSTREDIA

5.1 MatLab

MatLab (vychádza z názvu MATrix LABoratory, v preklade „maticové laboratórium“ - jeho hlavnou dátovou štruktúrou sú matice) je interaktívna platforma programovacieho jazyku štvrtej generácie, pôvodne určená pre UNIX, dnes používaná na operačných systémoch Linux, Windows a Mac OS X. Bol vytvorený s cieľom uľahčiť používateľom prácu s matematickými výpočtami, ale postupom času a po výraznom dopyte trhu, boli pridané ďalšie mnohé funkcie a rozšírenia. Dnes pokrýva veľmi široký záber aplikácií a programov, určených hlavne pre vedecko-technických pracovníkov pri vývoji a výskume, pre študentov a učiteľov na vysokých školách. K dopytu po ňom dochádza aj v treťom sektore služieb. Jeho schopnosti sú široké:

- počítania matic,
- vykresľovania 2D a 3D grafov funkcií,
- implementácie algoritmov,
- modelovania a simulácií,
- rozboru predmetu a prezentácie získaných informácií,
- tvorby aplikácií (aj po grafickej stránke),
- inžinierskej grafiky, atď.

Na univerzitách je požívaný ako pomôcka pri výučbe matematiky a technických predmetov. V priemysle pomáha ako veľmi efektívny nástroj pri analýze dát.

5.1.1 Vlastnosti MatLabu

Základnou vlastnosťou MatLabu je, že všetky objekty sú považované za prvky pol'a alebo matice. Môžu to byť nielen čísla a premenné, ale aj zložitejšie štruktúry, ako napríklad obrázky. Jeho všestrannú funkčnosť a výkonnosť zabezpečuje nadväzujúci softvér, ktorý tvoria programy tzv. „*toolboxy*“, zväčša určené pre konkrétne spektrum problémov. K nadväzujúcim rozšíreniam môžeme zaradiť aj používateľom vytvorené programy (*m-files*, v preklade „*m-súbory*“). Majú koncovku *.m a obsahujú zoznam funkcií, skriptov alebo tried. Je možné ich neobmedzene vytvárať, ukladať a následne načítať a použiť podľa potreby. [15, 16]

5.1.2 Použitelnosť MatLabu

Problém, ktorý spočíva v otázke, či je vhodné použiť MatLab na náš konkrétny problém, nie je vždy úplne jednoznačne riešiteľný, a je potrebné pristupovať ku každému jednému osobitne. Záleží na niekoľkých faktoroch, ktoré vyplývajú z požiadaviek a podmienok užívateľa. Pokiaľ potrebujeme rozsiahle výpočty, spracovávať rozmerné dátové súbory, pracovať s obsiahlymi maticami a všeobecne v prípadoch, kedy sa dá riešenie problému previesť na vektorové a maticové operácie, je MatLab obzvlášť vhodný. Vzhľadom k jeho možnostiam programovania je použiteľný aj pokiaľ riešime rozvetvené, prípadne

iteračné algoritmy. Má však, samozrejme, aj svoje slabé stránky. Neodporúča sa ho používať, pokiaľ potrebujeme vykonávať stále tie isté výpočty dokola alebo pracovať s veľkými objemami dát. V skratke vtedy, keď je dôležitým faktorom celková rýchlosť procesu výpočtu a nie vývoj algoritmu. V takom prípade volíme skôr hotový algoritmus napísaný v nejakom inom programovacom jazyku, ktorý má schopnosť vytvoriť spustiteľný kód a nepoužiť pri písaní MatLab, ktorý by bol v tomto prípade len interpret.

MatLab je výkonný a efektívny prostriedok k dosiahnutiu požadovaných výsledkov, avšak je vhodné si dopredu určiť:

1. Čo potrebujeme – disponovať jasnou a pochopiteľnou formuláciou úlohy.
2. Ako to chceme dosiahnuť – poznať možné spôsoby riešenia problému.
3. Ako overíme výsledok – mať na konci procesu možnosť otestovať správnosť celého priebehu výpočtu a konečného výsledku. [16]

5.1.3 Thing Speak

ThingSpeak je rozšírenie MatLabu, patriace do skupiny programov s licenciou *open source*, čiže voľne šíriteľného softvéru. Je aplikáciou IOT - Internet Of Things, čo v informatike znamená označenie spojenia internetu a vstavaných zariadení.

Možno ho použiť na uchovávanie a znovu získanie dát použitím protokolu HTTP cez Internet lebo lokálne cez *Local Area Network*. Taktiež sa prostredníctvom neho dajú vytvoriť aplikácie na zaznamenávanie informácií zo snímačov, lokalizovanie polohy alebo siete s aktualizáciou statusu. Okrem uchovávania alfanumerických dát, má ThingSpeak mnoho ďalších funkcií, ako napríklad časové škálovanie, priemerovanie, zaokrúhľovanie, vytváranie súm alebo mediánu. Každý ThingSpeak Channel (v preklade kanál) podporuje vstupy dát až do ôsmich dátových polí, zemepisnú šírku a dĺžku, nadmorskú výšku a status. Je možná integrácia JSON, SML a CSV súborov do aplikácií. Taktiež je dostupná funkcia manažmentu časových zón, čítania a zapisovania API kľúča a implementovanie diagramov založených na programovacom jazyku JavaScript z Highslide softvéru. [17, 18]

5.2 OpenWeatherMap

OpenWeatherMap je online služba, ktorá sprostredkúva dáta počasia, zahŕňajúce jeho aktuálny a minulý stav a predpoveď. Tieto informácie poskytuje výrobcem webových a mobilných aplikácií. Ako zdroje dát používa meteorologické vysielacie služby, ktoré získavajú čisté nespracované údaje z letísk, radarových staníc a iných oficiálnych miest zaoberajúcich sa počasím. Následne tieto údaje spracováva a poskytuje čo najpresnejšie predpovede a ich zaznačenia do máp, ktoré môžu mať rôzne zameranie. Napríklad mapy oblačnosti, zrážok, tlaku, teploty, vetra. Je ich však o mnoho viac. Filozofiou tejto služby je fungovať podobne ako OpenStreetMap (túto službu využíva na zobrazenie máp počasia) alebo Wikipedia, čiže vytvárať informácie zadarmo (pri splnení určitých

podmienok) a prístupné pre každého. Je možné dokúpiť si balíčky (obr. 5), ktoré zahŕňajú rozšírené možnosti práce s celým rozhraním.

	Free	Startup	Developer	Professional	Enterprise
Price per month Price is fixed, no other hidden costs (VAT is not included)	Free	40 USD / month	180 USD / month	470 USD / month	2,000 USD / month
Subscribe	Get API key and Start	Subscribe	Subscribe	Subscribe	Subscribe
Calls per minute (no more than)	60	600	3,000	30,000	200,000
Current weather API	✓	✓	✓	✓	✓
4 days/hourly forecast API ^{NEW}	-	-	✓	✓	✓
5 days/3 hour forecast API	✓	✓	✓	✓	✓
16 days/daily forecast API	-	✓	✓	✓	✓
Weather maps 2.0: Current, Forecast, Historical ^{NEW}	-	-	✓	✓	✓
Relief maps ^{NEW}	-	-	✓	✓	✓
Weather maps 1.0	✓	✓	✓	✓	✓
Bulk download	-	-	-	✓	✓
UV index	✓	✓	✓	✓	✓
Weather alerts	✓	✓	✓	✓	✓
Service					
Availability	95.0%	95.0%	99.5%	99.5%	99.9%
SLA	-	-	-	-	✓
Weather API data update	< 2 hours	< 2 hours	< 1 hour	< 10 min	< 10 min
Weather maps data update	< 3 hours	< 3 hours	< 3 hours	< 3 hours	< 3 hours
API lifetime support	Current version	Current version	Current version	All versions	All versions
License for maps, APIs, and other products	CC BY-SA 4.0	CC BY-SA 4.0	CC BY-SA 4.0	CC BY-SA 4.0	CC BY-SA 4.0 or custom
License for data and database	ODbL	ODbL	ODbL	ODbL	ODbL or custom
Tech support	Helpdesk	Helpdesk	Helpdesk	Direct	Direct 24x7

Obr. 5: Ceny balíčkov a ich obsah s výhodami a nevýhodami. Ponuka pochádza priamo zo stránok OpenWeatherMap. [20]

Licencia, pod ktorou celá služba funguje sa nazýva Creative Commons Attribution - ShareAlike. Jej konkurenciou je napríklad Weather Underground alebo Yahoo Weather.

5.2.1 API – application program interface

OpenWeatherMap poskytuje API s JSON, XML a HTML formami výstupov dát a ich limitované použitie zadarmo. Viac ako 60 požiadaviek za minútu vyžaduje platbu vo výške začínajúcej na 40 USD za mesiac. Prístup k informáciám z minulosti stojí od 150 USD za mesiac.

Aktuálny stav počasia sa obnovuje každých 10 minút a miesto, ktorého počasie nás zaujíma, hľadáme podľa geografických súradníc, podľa jeho názvu alebo štátu,

v ktorom sa nachádza. Služba ponúka hodinovú predpoveď fungujúcu v intervale do štyroch dní, trojhodinovú predpoveď počasia až na päť dní vopred a dennú predpoveď až na šesť dní dopredu. Taktiež v balíčkoch nájdeme možnosti ako upozornenia pri výkyvoch počasia alebo UV index (obr. 5). [19, 20]

5.3 Twitter

Twitter bol založený v roku 2006 a odvtedy sa stal jednou z najpoužívanejších sociálnych sietí na svete. Miera možností práce s touto platformou je veľmi rozsiahla, čo znamená, že je možné ho využiť napríklad ako marketingový nástroj, prostriedok na zdieľanie noviniek alebo sledovanie myšlienok a statusov obľúbenej celebrity a priateľov. Človek má vďaka nemu k dispozícii obrovské množstvo informácií, pretože dokáže spojiť veľké množstvo ľudí po celej Zemi. Noví používatelia môžu byť zo začiatku zmätení z celkovej formy fungovania na sieti, pretože si vytvorila vlastný ekosystém virtuálneho života, možno povedať - vlastnú kultúru.

5.3.1 Formy zdieľania na Twitteri

Tweets alebo *tweety* sú príspevky (ktorých rozsah nemôže prekročiť hodnotu 280 znakov) používateľa, ktoré sa zobrazujú na nástenkách jeho sledovateľov. „Sledovať“ niekoho znamená, že si pýtame povolenie na odoberanie obsahu prispievateľa. Pokiaľ o to požiadame, príde mu upozornenie, ktoré musí potvrdiť. Je tu aj možnosť verejného profilu. V tomto prípade môže vidieť príspevky zdieľajúceho ktokoľvek. Nie je pravidlom, aby sa nasledovanie opätovoľovalo, čiže pokiaľ nasleduje používateľ A používateľa B, neznamená to, že B bude nasledovať A.

Twitter Newsfeed je prakticky nekončiaca nástenka príspevkov v chronologickom poradí. Pozostáva z príspevkov ľudí, ktorých sledujeme a čím viac je takých ľudí, tým viac je naša nástenka zaujímavá a diverzifikovaná. Redukuje tiež šancu, že uvidíme rovnaké príspevky viackrát. Pokiaľ sa nám na nástenke zobrazí príspevok od niekoho, o kom sme nikdy nepočuli, je pravdepodobné, že náš priateľ zdieľal tweet od jeho známeho. Tomu vravíme retweet.

Retweet čiže pripnúť cudzí príspevok na svoj profil, môže mať niekoľko dôvodov. Človek, ktorý takto zdieľa, súhlasí s názorom obsiahnutým v správe *tweetu* a chce ho šíriť aj medzi okruh svojich známych, pretože ho pokladá za dôležitý a hodnotný.

Twitter Threads je ďalšia súčasť slovníka Twitteru. Je to súbor *tweetov* jednej osoby zaoberajúcich sa jedným druhom témy. Prispievateľ vytvorí svojimi poznatkami pripnutými na nástenke vlákno a neustále ho oživuje novými. Tento typ komunikácie spravidla rieši dlhšie a hlbšie problémy.

Quote tweets požívame, pokiaľ chceme opísať, prečo sme použili funkciu retweet. To znamená, že zdieľame cudzí príspevok na svoj profil a pridáme k nemu ešte vlastný dodatok vo forme komentáru. Pridanie vlastnej myšlienky môže mať vysvetľujúci charakter, ako sme informáciu pochopili.

Twitter trending je funkciou Twitteru, ktorá pridáva príspevky na našu nástenku. Automaticky vyberá také, ktoré by sa nám mohli páčiť. Takýto príspevok vlastne vytvorí ľudia početným tweetovaním a pridávaním hashtagov o určitej idei alebo udalosti. Často sa to stáva napríklad pri predávaní Oskarov alebo športovej akcii Superbowl. Kliknutím na hashtag vidíme, čo si myslia ľudia o téme.

Hashtag je slovo alebo slovné spojenie so symbolom mriežky „#“ pred ním. Slúži na rýchle označenie toho, o čom náš príspevok bude. Pomáha kategorizovať jednotlivé myšlienky používateľov. To znamená, že pokiaľ k *tweetu* pridáme fotku svojho domáceho miláčika, môžeme taktiež napísať *hashtag* #domacimilacik. Každý, kto sa bude zaujímať o túto tému, klikne na *hashtag* a zobrazí sa mu história príspevkov s rovnakým označením.

Direct message je formou priamej komunikácie medzi ľuďmi prihlásenými na Twitteri. Je veľmi podobná osobným správam na ostatných sociálnych sieťach. Používateľ môže mať otvorený *inbox* (doručenú poštu), čiže mu môže napísať ktokoľvek, alebo je k dispozícii nastavenie, pri ktorom musí adresát poštu najskôr schváliť. Dôvodom takýchto obmedzení je zabrániť obťažovaniu a diskriminácii. [21, 23]

6 REALIZÁCIA PREDPOVEDI A ZDIEĽANIA

Inteligentná predpoveď počasia sleduje cieľ vytvoriť automaticky fungujúcu, verejnú predpoveď z dostupných informácií. Práca bola rozdelená na dve časti.

Prvá sa zaoberá analýzou a inventarizáciou informácií a výrazových prostriedkov. Druhá časť spočíva vo vytvorení algoritmu, ktorý preniesie všetky údaje od vstupu k výstupu a medzitým poskladá tieto informácie metódou automaticky generovaných textov do zmysluplného celku.

6.1 Analýza textu

Na samom začiatku pri vyváraní generovaného textu je potrebné rozhodnúť, akým spôsobom budeme pri zostavovaní algoritmu postupovať, aby mohla byť využitá určitá štruktúra metód, s prípadnými malými obmenami, podľa potreby požadovaného výsledku. Výstup nášho programu bude *tweet* na sociálnej sieti Twitter. Mal by byť stručný, jasný, ale zároveň obsahovať všetky elementárne dáta o predpovedi počasia.

Nad formou výsledného statusu alebo správy chceme mať dokonalú kontrolu, pretože preklepy a nejasné vety by mohli zniesť nedôveryhodne a mohli adresáta zavádzať. Preto bol navrhnutý postup – zvoliť metódu automaticky generovaného textu s prvkami kombinačnej techniky, čiže využiť pri práci šablóny. To zabezpečí, že výsledný text bude vždy zodpovedať našim predpisom.

Pokiaľ máme zvolenú metodiku riešenia zadania, možno sa orientovať podľa postupu zhromaždeného v teoretickej časti (kapitola 2.4). Tá odporúča zvoliť si vhodný textový korpus, ktorý bude zdrojom väčšiny informácií. Vzhľadom na fakt, že sa pohybujeme v oblasti predpovedí počasia, náš cieľový používateľ je človek, ktorý môže predstavovať akúkoľvek vekovú kategóriu, sociálnu skupinu, pohlavie. Samozrejme, nezohľadňujeme extrémny ako príliš mladí alebo ľudia nerozumejúci výslednému jazyku.

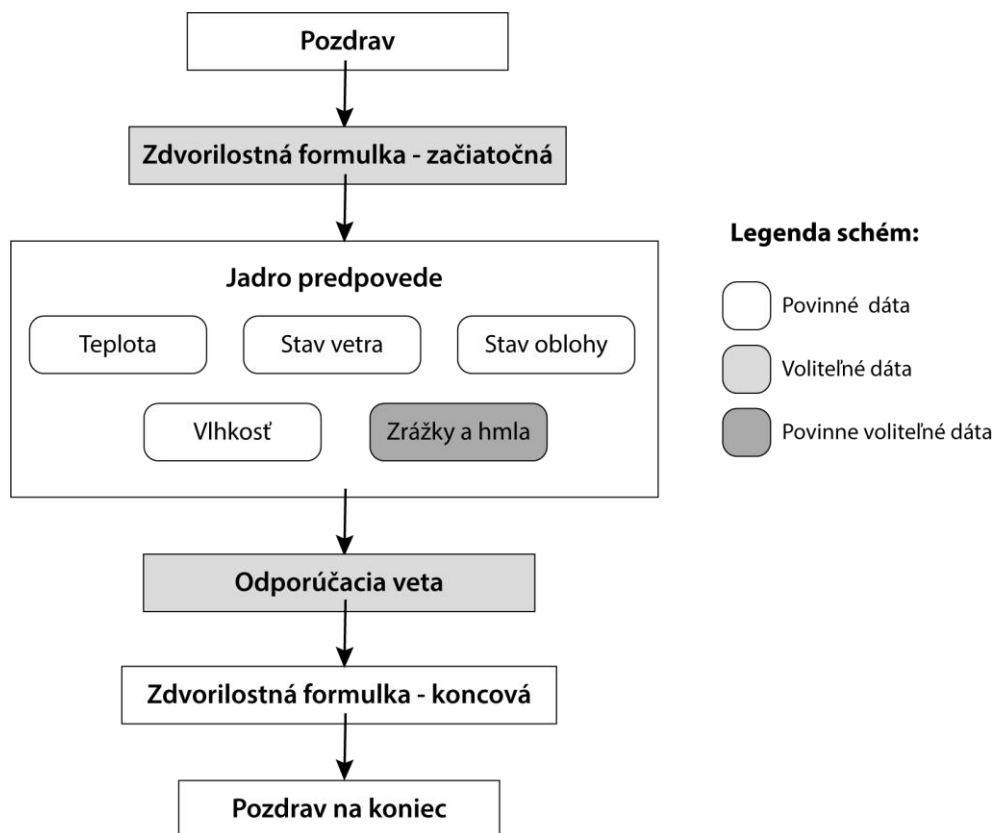
Z týchto dôvodov volíme ako vhodný korpus (príloha 1) televízne relácie, zaoberajúce sa predpoveďami pre rovnako univerzálneho adresáta, fungujúce niekoľko rokov a overené mnohými sledovateľmi. Vyberáme texty z viacerých mediálnych zdrojov pre zabezpečenie rôznorodosti vstupu. Pokúsime sa zabezpečiť vzor predpovedí z dátumov každého ročného obdobia, kvôli dostatočnému obsiahnutiu potrebných informácií. Niektoré texty televíznych relácií sú príliš dlhé a pre nás zbytočné. Tieto prejavy sa zväčša venujú predpovedi na niekoľko dní, zatiaľ čo my sledujeme iba najbližších niekoľko hodín. Z tohto dôvodu sa snažíme vyberať úryvky, ktoré obsahujú dennú predpoveď alebo v nich nájdeme frázy, ktoré sa neopakujú a opisujú prvky použiteľné v našom prípade. Vhodné je pridať aj textové predpovede z internetových portálov priamo zameraných na informovanie o počasí.

6.1.1 Plánovanie textu

Prvá fáza procesu tvorby automaticky generovanej predpovedi počasia po vytvorení vhodného korpusu, by mala podľa kapitoly 2.5 začať voľbou obsahu a tvorbou štruktúry textu. Ako už bolo spomenuté, náš výsledný text bude stručná a vecná predpoveď, podaná milou cestou ku koncovému používateľovi. Vstup bude predstavovať:

1. Zdroj znalostí – tabuľkový model získaných informácií o stave počasia. Máme na výber, buď použijeme dáta zo služby OpenWeatherMap alebo z domácej meteostanice. V tomto prípade bude prvá možnosť vhodnejšia, a to z dôvodu rýchlej dostupnosti a dôveryhodnosti dát. Rolu zohráva aj cenová dostupnosť.
2. Komunikačný cieľ – informovať bežného používateľa Twitteru o počasi daného dňa.
3. Užívateľský model – používatelia Twitteru, univerzálny typ koncového adresáta.
4. História – fungujúce systémy generovania počasia ako SumTime alebo CBR-Meteo.

Čo sa týka štruktúry, správa bude obsahovať niekoľko častí (obr. 6), ktoré si rozdelíme na povinné (správa ich bude obsahovať vždy), povinne voliteľné (budú vložené



Obr. 6: Schéma štruktúry plánovaného obsahu.

do vety na základe zisteného výskytu konkrétneho stavu počasia, napr. zrážky) a voliteľné (zdvorilostné formulky a nie úplne nevyhnutné informácie, ktoré však dodávajú efekt plnosti a nasýtenosti textu).

Konkrétna forma, z ktorej bude *tweet* pozostávať, bude teda pozdrav, resp. zaželenie pekného dňa a prípadné pridanie zdvorilého privítania. Ďalej budú zobrazené časti správy vo forme osobitých viet alebo súvetí, ktoré budú obsahovať informácie o aktuálnej predpovedi:

- teploty,
- stavu vetra,
- zrážok,
- stavu oblohy,
- vlhkosti.

Po nich bude nasledovať voliteľná zdvorilostná veta pojednávajúca o oblečení alebo inom prispôsobení sa na očakávané poveternostné podmienky. Na konci správy bude forma zaželenia pekného zbytku dňa alebo pozdrav.

6.1.2 Mikro-plánovanie

Fáza, ktorá nasleduje, sa podľa teoretického základu nazýva mikro-plánovanie. Skladá sa z podprocesov: lexikalizácia, agregácia, referovanie, vyjadrenie a generácia. V šablónovej metóde sa tieto procesy jemne líšia, preto budeme brať ohľad na rozdiely, popísané v kapitole 2.6.

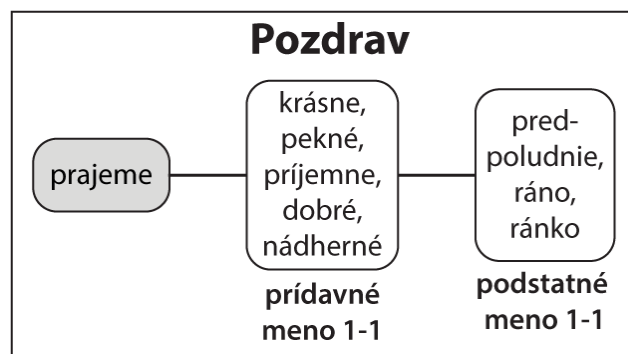
Na začiatku je dôležité vybrať vhodné výrazové prostriedky, a to na základe plánu dokumentu a zvoleného korpusu. Tiež možno použiť vhodné slová, ktoré budeme čerpať zo skúseností z bežného života.

Hľadáme také výrazové prostriedky, ktoré sa budú čo najviac zhodovať s predstavou, ako má výsledok vyzeráť. Zvažujeme pri tom ich vecný význam a dopad na konečnú správu (kapitola 3.2). Po procese hľadania a výberu zobrazíme do stromovej štruktúry vyskloňované slová, pretože sa nám potom bude s nimi ľahšie pracovať pri povrchovej realizácii a pri písaní algoritmu pre ňu.

Schéma štruktúry plánovaného obsahu predstavuje koreň riešenia problému, z ktorého budeme čerpať v nasledujúcich krokoch. Je základným zdrojom informácií pri zostavovaní stromových štruktúr jednotlivých súčastí obsahu. Každá by mala obsahovať použiteľné výrazové prostriedky a ich vzájomné syntaktické prepojenia.

Niektoré moduly schém obsahujú synonymické obmeny, ktoré zabraňujú monotónnosti výsledného textu. Tie časti viet, ktoré sú schopné takýchto modifikácií, špeciálne označujeme názvami, vyjadrujúcimi ich slovný druh alebo inú charakteristiku. Pokiaľ ide o súbor väčšieho množstva slov, používame označenie „fráza“. Jednotlivé slová bez synonym v schéme neoznačujeme názvom.

Pozdrav. Ako už názov napovedá, jedná sa o úvod správy, vyjadrený pomocou slušného pozdravenia adresáta. Jeho výsledná forma bude obsahovať maximálne 3 slová (obr. 7).



Obr. 7: Súčasť - Pozdrav.

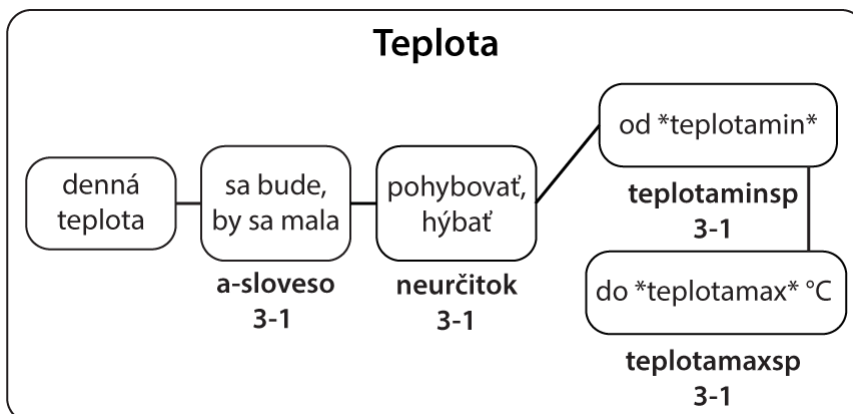
Ďalšou súčasťou je **Zdvorilostná formulka – začiatočná** (obr. 8), ktorá sa nebude vyskytovať vo výsledku vždy, pretože ju radíme do skupiny voliteľná. Ide o modul, ktorý milým spôsobom uvíta čitateľa a charakterizuje, čo sa v celej správe deje. Obsahuje dve vetvy, pretože skloňovanie v použítom slovenskom jazyku neumožňuje skladať slová v jednom type ohýbania slov.



Obr. 8: Súčasť – Zdvorilostná formulka - začiatočná.

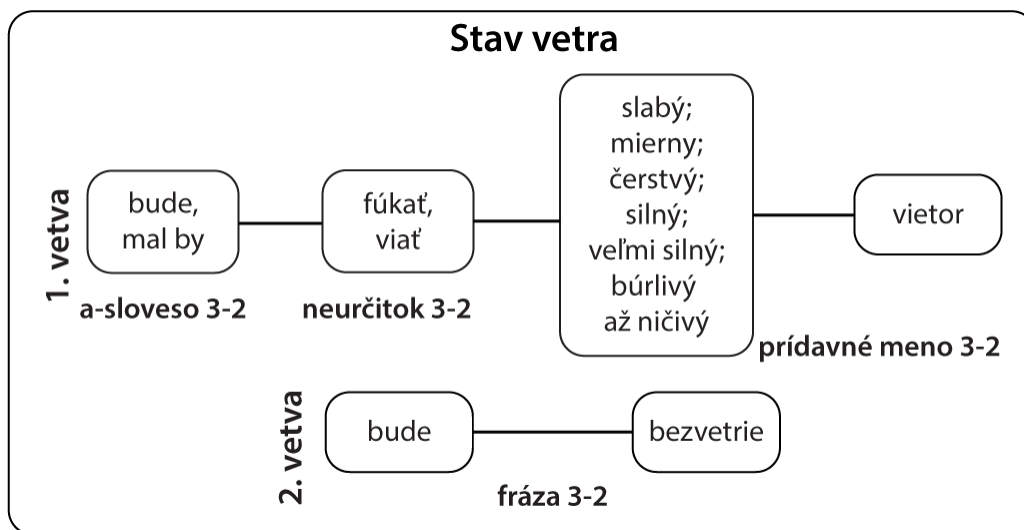
Súčasť **Jadro predpovede** je zložitejšia, pretože sa rozvetvuje na niekoľko sekcií. Sú to , Stav vetra, Zrážky a hmla, Stav oblohy a Vlhkosť. Jednotlivé zložky spolu nesúvisia a ich poradie vo vete neurčuje ich dôležitosť, ktorú im každý čitateľ môže priradiť inak. Preto sa budeme snažiť ich zoradiť k univerzálnej spokojnosti.

Teplota (obr. 9) popisuje rozmedzie denných teplôt, ktoré sa v daný dátum predpokladajú. Nachádza sa tu prepojenie s ďalšou sekciou – Stavom vetra, a to v module *a-sloveso 3-1*. Tieto slovesá majú rovnaký vetný základ, preto chceme dosiahnuť ich rozdielny výber v dvoch potenciálne susedných vetných konštrukciách.



Obr. 9: Sekcia Teplota.

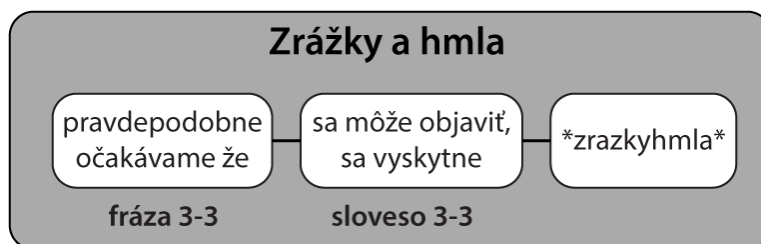
Stav vetra (obr. 10) patrí taktiež medzi hlavné predpovedné veličiny. Vyjadrená bude jeho sila na základe získaných dát. Obsahuje aj druhú vetvu, ktorá pokrýva možnosť, že bude bezvetrie.



Obr. 10: Sekcia Stav vetra.

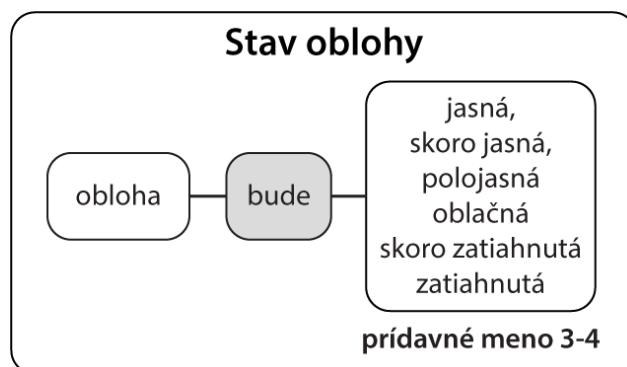
Zrážky a hmla (obr. 11) obsahujú výstrahu pred rôznymi typmi zrážok a zároveň hmlu, ktorá je dôležitým faktorom v predpovedi počasia. Túto časť textu radíme do skupiny povinne voliteľných, čiže o výslednom výskyte v správe rozhoduje jej samotná existencia. Slovo ohraňované v hviezdičkách (*zrazkyhmla*) vyjadruje hodnotu, ktorá sa bude plniť v programovej časti. Jej určenie je rovnaké ako aj pri ostatných moduloch,

avšak nečerpá z nášho korpusu, ale z informácií získaných z webovej aplikácie OpenWeatherMap, ktorá tieto dáta ponúka priamo vo forme slov v slovenskom jazyku.



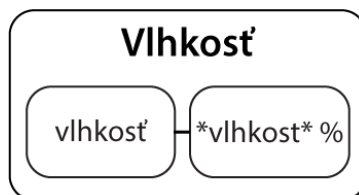
Obr. 11: Sekcia Zrážky a hmla.

Stav oblohy (obr. 12) má jednoduchú stavbu. Opisuje oblačnosť prídavnými menami, ktoré rozvíjajú slovo „obloha“. Ich zdrojom sú okrem nami zvoleného korpusu oficiálne informácie o meteorologickej terminológii [24].



Obr. 12: Sekcia Stav oblohy.

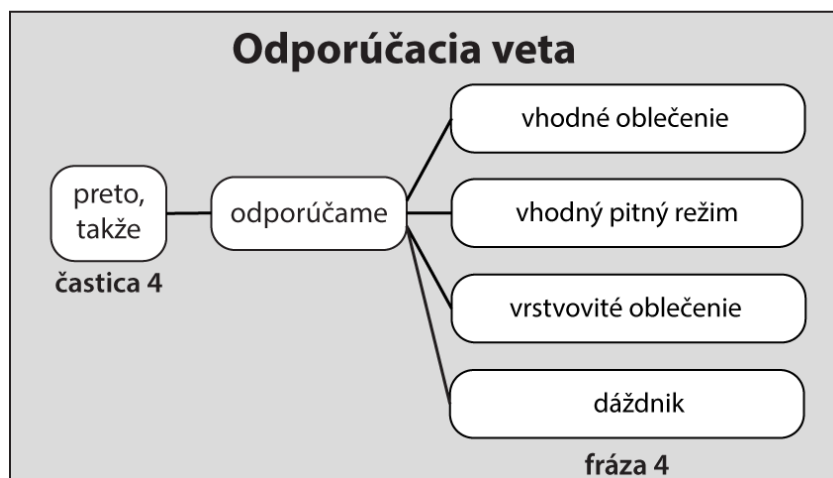
Poslednou časťou Jadra predpovede je **Vlhkosť** (obr. 13). Obsahuje tak, ako aj Zrážky a hmla, hodnotu v hviezdikách (*vlhkost*), kde počítač dosadí číselný údaj zo získaných informácií. Výsledok bude v percentách.



Obr. 13: Sekcia Vlhkosť.

V hlavnej štruktúre plánu dokumentu ďalej pokračuje **Odporúčacia veta** (obr. 14), ktorá bude pôsobiť v podstate ako varovanie pred extrémami a bude pripomínať potrebu vhodne sa na ne pripraviť. Vždy bude nasledovať po vete, ktorej následkom sú opísané opatrenia. V našom prípade pri zrážkach alebo pri extrémne vysokej a nízkej

teplote vzhľadom na našu zemepisnú šírku. Použiť ju však bude možné iba pri jednom zo spomenutých javov, na čo je potrebné dbať pri povrchovej realizácii.



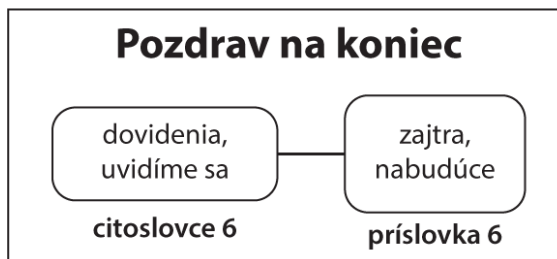
Obr. 14: Súčasť – Odporúčacia veta.

Zdvorilostná formulka – koncová (obr. 15) je vytvorená z rovnakého dôvodu ako začiatková, líši sa však v tom, že namiesto uvádzania do deja ho hodnotí. Význam ich obsahu je však rovnaký. 1. vetva je vybratá, pokiaľ vo vete nepoužijeme Zdvorilostnú formulku – koncovú, 2. vetva funguje v opačnom prípade.



Obr. 15: Súčasť – Zdvorilostná formulka – koncová.

Pozdrav na koniec. Skladá sa z formulácie pozdravu a frázy, ktorá okrem toho, že signalizuje koniec správy, podnecuje k prečítaniu *tweetu* aj nabudúce.

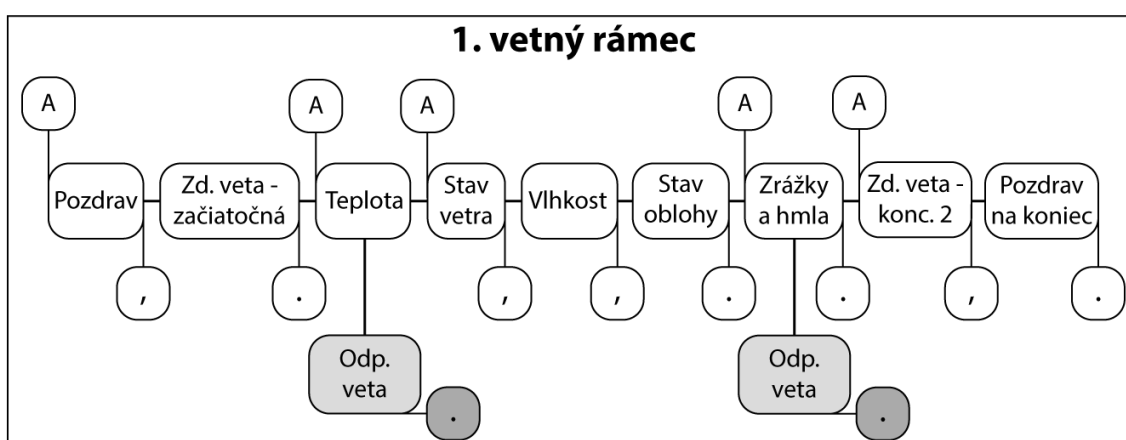


Obr. 16: Súčasť – Pozdrav na koniec.

6.1.3 Povrchová realizácia

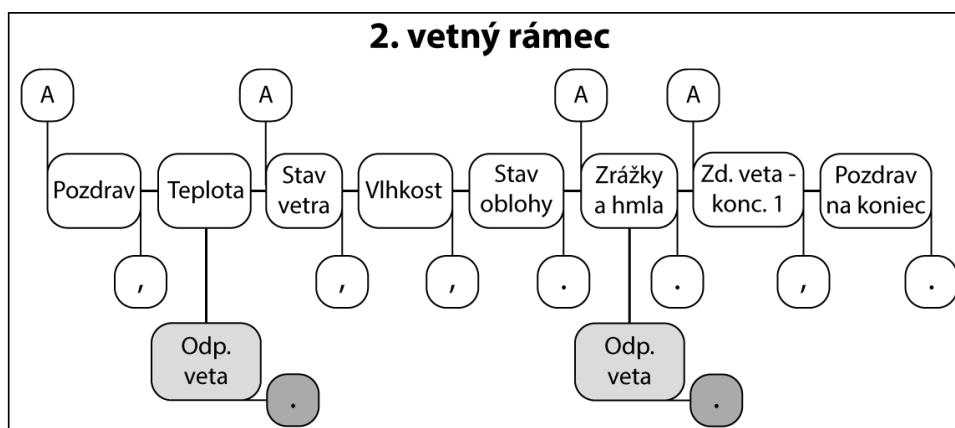
V poslednej časti procesu tvorby automatického generovania je našou úlohou vytvoriť konečné štruktúry viet, ktoré budú gramaticky a syntakticky správne. Budeme vychádzať z predošlých šablón, vytvorených na mieru nášmu problému. Kvôli zachovaniu diverzity výsledku, vytvoríme 4 rámce, ktoré budú podobne ako v prechádzajúcich kapitolách charakterizovať výstup. V tomto prípade už konečný.

1. vetný rámec (obr. 17) bude obsahovať všetky súčasti spomenuté v pláne dokumentu a využijeme ho aj pri tvorbe ostatných. Budú sa líšiť hlavne usporiadaním súčastí. Tvorba pevných šablón viet je namieste, pretože maximálny počet slov na výstupe je 280 (obmedzenie na Twitter [23]) a pokiaľ by bol spôsob výberu jednotlivých súčastí a sekcií v nich voľnejší, mohlo by dochádzať k častému prekračovaniu spomenutej podmienky.



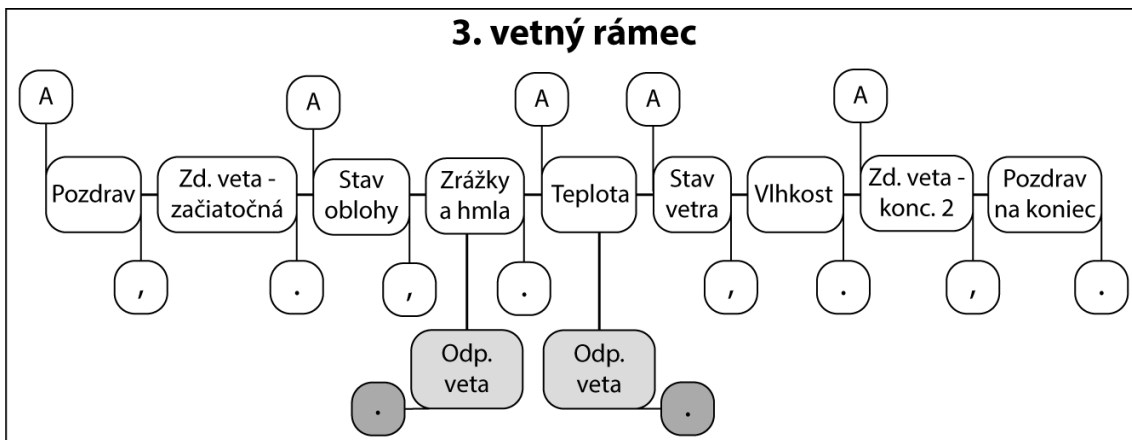
Obr. 17: 1. vetný rámec.

2. vetný rámec (obr. 18) sa od prvého líši použitím Zdvorilostných viet, a to tak, že v prvom prípade použijeme začiatočnú a 2. vetvu koncovú. V druhom vetnom rámci použijeme iba koncovú, konkrétne jej prvú vetvu.

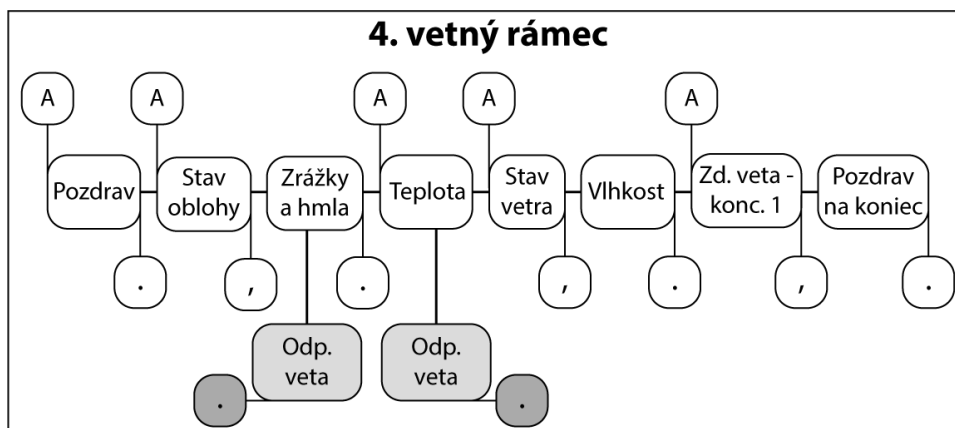


Obr. 18: 2. vetný rámec.

3. a 4. vetný rámeč (obr. 19 a obr. 20) sa líšia od predošlých dvoch vnútorným usporiadaním hodnôt, ktoré referujú počasie. Rozdiely, ktoré majú navzájom, sú rovnaké ako medzi 1. a 2.



Obr. 19: 3. vetný rámeč.



Obr. 20: 4. vetný rámeč.

Čo sa týka legendy znakov v schémach – symbol A znamená veľké písmeno na začiatku vety, ku ktorej sa vzťahuje. Zbytok grafických vyjadrení je jasný z ich vizuálnej podoby, znamenajú to, čo vyjadrujú.

6.2 Tvorba algoritmu

Programová realizácia je jednou z najdôležitejších fáz celej práce. Zhromažďuje všetky potrebné údaje, triedi ich a vytvára konečný výsledok, ktorý uvidí čitateľ. Algoritmus sme vytvárali vo vývojovom prostredí MatLab, a ten bol po dokončení prenesený do webového *cloudu* ThingSpeak. Tieto dve služby patria pod jedného výrobcu, čo znamená, že sú plne kompatibilné, čo sa týka spustiteľného kódu.

Proces je rozdelený na dve časti. Prvá spočíva vo vytvorení inštrukcií pre ThingSpeak, ktorý informácie získava z OpenWeatherMap prostredníctvom API kľúča a následne podobným spôsobom odovzdáva na sociálnu sieť Twitter. Druhý krok je samotné zlučovanie získaných informácií o počasí s výrazovými prostriedkami.

6.2.1 Získanie a odovzdanie informácií

Získanie informácií (nasledujúci kód) spočívalo vo vytvorení zoznamu príkazov pre aplikáciu MatLab Analysis v ThingSpeaku, ktorá je na základe nich schopná vyberať príslušné dáta z JSON štruktúry, dostupnej z platformy OpenWeatherMap. Prevzali sme údaje o teplote (max, min), vetre, oblohe, vlhkosti a zrážkach. Požitý bol na to cyklus *for* s 5 opakovaniami, pretože od siedmej hodiny, kedy kód spúšťame, potrebujeme hodnoty pre čas 9, 12, 15, 18 a 21 hodín (denné hodnoty sa merajú od 7:00 do 21:00 [24]). Všetky získavané veličiny sú zoradené do poľa, z ktorých následne v jednotlivých sekciách vytvárame priemer.

```
r = RequestMessage;
uri =
URI('https://api.openweathermap.org/data/2.5/forecast?q=Bratislava,sk&lang=sk&appid=430e926f7bc26ea583f86d01ffded141&units=metric');
resp = send(r,uri);
status = resp.StatusCode;

%o siedmej 5 hodnôt, hodiny: 9, 12, 15, 18, 21
pocet= 5;
for i=1:pocet
teplotamin(1,i)=resp.Body.Data.list{i, 1}.main.temp_min;
teplotamax(1,i)=resp.Body.Data.list{i, 1}.main.temp_max;
vietor(1,i)=resp.Body.Data.list{i, 1}.wind.speed;
obloha(1,i)=resp.Body.Data.list{i, 1}.clouds.all;
vlhkost(1,i)=resp.Body.Data.list{i, 1}.main.humidity;
all=resp.Body.Data.list{i, 1}.weather.description;
z_s_zrazkyhmla(1,i)=convertCharsToStrings(all)
z_c_zrazkyhmla(1,i)=resp.Body.Data.list{i, 1}.weather.id
end
```

Odovzdanie vytvorených informácií (kód nižšie) prebehne na základe nasledujúceho algoritmu. Funguje podobne ako prijímanie, a to pomocou API kľúča vytvoreného cez ThingTweet.

Vykonanie prijímacích aj odosielacích inštrukcií v predom určenej dobe (7:00) zabezpečuje aplikácia TimeControl, v ktorej je možné nastaviť interval opakovania, časovú zónu, *fuzzy time* a niekoľko ďalších hodnôt.

```
r1 = RequestMessage;
adresa='https://api.thingspeak.com/apps/thingtweet/1/statuses/update?api_key=4XXC5QXKD5H55UK2&status=';
tweet=urlencode(vetny_ramec1);
poziadavka=sprintf('%s%s',adresa,tweet);
uri = URI(poziadavka);
resp = send(r,uri);
```

6.2.2 Automatické generovanie textu

Pri písaní programu v MatLabe (príloha 2) sme sa riadili analýzou textu, podrobne popísanou v kapitole 6.1. Jednotlivé schémy nám pomohli pri zostavovaní celkového výstupu. Kód je rozdelený komentármi (znakom %) na jednotlivé časti plánu dokumentu. Každú súčasť sme skladali zvlášť, a to na základe variácií. Podľa nich sme potom vyberali vhodné slová, ktoré syntakticky a lexikálne sedeli so zbytkom správy.

Postupovali sme rovnako ako aj v časti plánovania dokumentu:

- Prvá bola riešená súčasť **Pozdrav**. Táto časť kódu nám potom poslúžila pri ostatných súčastiach ako vzor. Vytvorili sme si polia so skupinami slov, ktoré sme v predošlých kapitolách predpísali a z ktorých potom, na základe náhody, program vyberal. Po vykonaní týchto procesov algoritmus spojil slová do slovných spojení.
- Rovnako sme postupovali pri **Zdvorilostnej formulke – začiatovej**, kde nám však vznikla ešte druhá vetva možnej vety, ktorej vyberanie funguje rovnako ako voľba slov do viet – na základe náhodného čísla.
- **Teplota** obsahovala dve vstupné premenné, a to minimálna a maximálna teplota, kde sme z 5 hodnôt vytvorili priemer a potom ich zaradili do správy.
- Pri riešení **Stavu vetra** sme využili podmienku *if*, aby z číselnej hodnoty opisujúcej rýchlosť mohla byť zvolená slovná reprezentácia.
- **Zrážky a hmla** boli odlišné v tom, že zdroj informácií pozostával aj z cifry, aj slovného zo vyjadrenia. Tieto dva nositele informácií sme prepojili a lexikálna časť sa na základe číselnej objavuje vo výsledku. Vždy sme hľadali extrém konkrétnej udalosti, čo znamená, že búrka má prednosť pred silným dažďom, silné sneženie pred slabým snežením.
- **Stav oblohy** funguje veľmi podobne ako stav vetra, tu však pracujeme s hodnotou oblačnosti.
- **Vlhkosť** je jednoduché spojenie získaného čísla a slova.
- **Odporúčacia veta** bola tvorená na základe podmienky existencie určitých javov – zrážky dažďové a snehové a vysoká a nízka teplota. Dážď je zo spomenutých procesov najčastejší a dokáže najviac človeka prekvapiť, preto má upozornenie na jeho možnú prítomnosť prednosť.
- Kód **Zdvorilostnú formulku – koncovú** bol napísaný podobne ako pre začiatkovú, nemusel však obsahovať riešenie pre viacero vetiev.
- Písanie kódu pre mikro-plánovanie sme ukončili súčasťou **Pozdrav na koniec**, kde prebieha jednoduchý výber slov do fráz.

Algoritmus povrchovej realizácie najskôr, na základe náhody, vyberá konkrétny vetný rámec, ktorého premennú plní plným obsahom viet, zložených z vyššie spomenutých súčastí. Výsledok je v dátovom type *string*, pričom následným prevedením na *char* pomocou `convertStringsToChars` je umožnené programu merať dĺžku

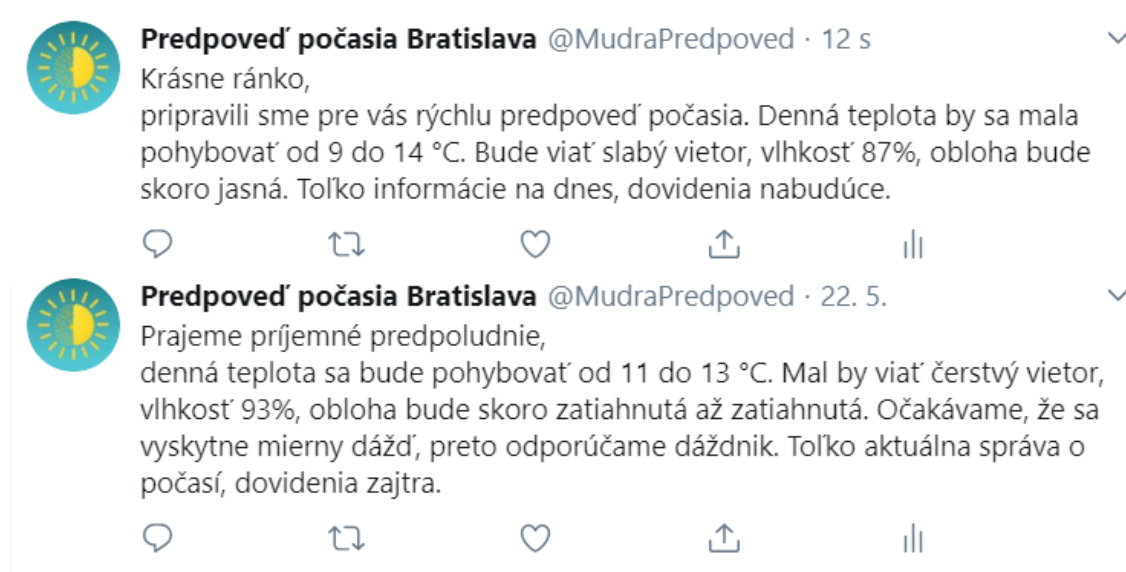
správy. To je nevyhnutné z dôvodu maximálnej dĺžky *tweetu*, ktorá bola v algoritme ošetrená cyklom *while*.

Na konci kódu je umiestnená vlastná funkcia, potrebná pri skladaní vetných rámcov. Jej úlohou bolo zväčšovať prvé písmená vo vete. Vstup aj výstup tvorí vetná konštrukcia, vo výsledku teda so zmeneným začiatkom.

6.3 Výsledok generovania

Funkčný kód pracuje na platforme ThingSpeaku, kde plní všetky funkcie, popísané v predošlých kapitolách. Úspešne vygenerované správy, ktorých tvorba prebieha každé ráno o 7:00 sú posielané na Twitter, kde sme si vytvorili účet s názvom Predpoveď počasia Bratislava. Jeho názov obsahuje popis činnosti a miesto, ktorého sa týka. To je výhodné, pretože lokalizáciu potom nie je potrebné spomínať v texte správy (názov účtu sa nachádza na nástenke vždy nad *tweetom*). Mesto bolo vyberané na základe jazyka, v ktorom je správa písaná (slovenský).

Ďalej boli nastavené profilové obrázky, ktorých tvorba prebehla konkrétne pre tento účel. Vniesli do prostredia vhodný vizuál a dôveryhodnosť. Snažili sme sa o minimalistický dizajn, a tým zaručiť, že sa budú adresáti radi vracat'. Príklady výslednej správy môžu vyzerat' napríklad takto (obr. 21).



Obr. 21: Príklady výsledného *tweetu* na sociálnej sieti.

7 ZÁVER

Riešenie problematiky, ktorej sa týkala táto bakalárska práca, zahŕňalo rozšírenie si teoretických znalostí v oblastiach automaticky generovaných textov, počítačovej predpovedi počasia a fungovania mikrokontrolérov. Získavali sme potrebné informácie aj vo sfére softvérového vybavenia a internetových služieb, kde boli popísané funkcie programu MatLab, jeho súčasti ThingSpeak, a tiež nás zaujímal Twitter a OpenWeatherMap. Inteligentná predpoveď počasia vyžadovala aj prehĺbenie vedomostí o meteorológii a jej termínoch.

Tieto informácie boli následne využité v realizačnej časti, ktorá spočívala v niekoľkých dôležitých bodoch. Začali sme výberom metódy, ktorá určila ďalšie kroky postupu. Výberom vhodného korpusu, jeho zhromaždením a spracovaním sme si zabezpečili zdroj potrebných dát pre vytvorenie programu generujúceho automatické texty. Nasledovala tvorba plánu dokumentu, ktorá spočívala vo zvolení charakteru obsahu, typu podania a tiež určovala, ako bude výstup vyzerat'.

Mikro-plánovaním sme sa dostali hlbšie k podstate riešeného problému. Bolo potrebné vytvoriť si niekoľko schém jednotlivých súčastí, ktoré obsahovali lexikálnu výbavu viet, a tiež inštrukcie na určenie ich správneho tvaru. Predstavovali vstup pre následnú fázu povrchovej realizácie, kde sme jednotlivé vety vybavili diakritikou, aby boli vo výsledku gramaticky a syntakticky správne. Tak ako v predošlej fáze, aj tu nám pomáhali na mieru vytvorené šablóny.

Celý proces bol dôležitý, pretože tvoril koreň riešenia programovej časti práce. Algoritmus bol tvorený pomocou MatLabu, pretože jeho užívateľské prostredie je prehľadné a plne kompatibilné s využitou službou ThingSpeak. Najskôr sme získali údaje z databázy OpenWeatherMap, následne spracovali, postavili do formy zmysluplného textu a odovzdali na Twitter, kde boli publikované na profile Predpoveď počasia Bratislava.

Výrazové prostriedky boli do viet vkladane na základe náhodných premenných alebo boli pevne vymedzené šablónou a poskladané s ohľadom na predom vytvorené vetné rámce. Tie boli využité, pretože obsah *tweetu* je limitovaný na 280 znakov. [23] Potrebovali sme preto pevnú kontrolu nad formou viet a ich maximálnou možnou dĺžkou.

Úplným výsledkom bakalárskej práce je každodenná správa o predpovedi počasia v daný deň. Má veľké množstvo možností vonkajšieho výzoru, preto súbor niekoľkých takýchto *tweetov* nepôsobí monotónne. Zamýšľané bolo aj pridávanie obrázkov k predpovediam, takúto funkciu ale použitý softvér neponúkal. Generovanie automatického textu bolo úspešné a výstup obsahuje všetky popísané náležitosti.

8 ZOZNAM POUŽITEJ LITERATÚRY

- [1] REINGOLD, Eyal - Nightingale, Johnathan. *The Turing Test*. [online] 1999 [cit. 2019-05-04]. Dostupné z: <http://www.psych.utoronto.ca/users/reingold/courses/ai/>
- [2] ARON, Jacob. *Software tricks people into thinking it is human*. [online]. 2011 [cit. 2019-10-04]. Dostupné z: <https://www.newscientist.com/article/dn20865-software-tricks-people-into-thinking-it-is-human/#.VaEpOvmlVzk>
- [3] MARCAULT, Matheson. *The History of Text Generation*. [online] 2015 1999 [cit. 2019-11-04] Dostupné z: <http://mathesonmarcault.com/index.php/2015/12/15/randomly-generated-title-goes-here/>
- [4] PROCHÁZKOVÁ, Martina. *Generátor příběhů* [online]. Brno, 2016 [cit. 2019-12-04]. Dostupné z: <https://is.muni.cz/th/fy8lk/>. Diplomová práce. Masarykova univerzita, Fakulta informatiky. Vedoucí práce Zuzana Nevěřilová.
- [5] HORKÁ, Kateřina. *Knihy psané počítačem: trendy, možnosti, výzvy, úskalí* [online]. Brno, 2017 [cit. 2019-05-13]. Dostupné z: <https://is.muni.cz/th/bbask/>. Bachelor's thesis. Masaryk University, Faculty of Arts. Vedoucí práce Michal Černý.
- [6] RYAN, Marie-Laure – EMERSON, Lori – ROBERTSON, Benjamin J.. *The Johns Hopkins guide to digital media*. Baltimore: Johns Hopkins University Press, 2014. ISBN 1421412241.
- [7] REITER, Ehud – DALE, Robert (eds.). *Building natural language generation systems*. Cambridge: Cambridge University Press, c2000. Studies in natural language processing. ISBN 0-521- 62036-8.
- [8] DAZA, Angel - CALVO, Hiram - FIGUEROA-NAZUNO, Jesús. *Automatic Text Generation by Learning from Literary Structures* [online]. San Diego, California, USA, 2016 [cit. 2019-26-04]. Dostupné z: <https://www.aclweb.org/anthology/W16-0202>.
- [9] MATSKEVICH, Sergey – GORDON, Colin S. *Generating Comments From Source Code with CCGs* [online]. USA, 2018 [cit. 2019-26-04]. Dostupné z: <https://arxiv.org/pdf/1810.06599.pdf>
- [10] GELDOLF, Sabine. *Corpus-analysis for NLG*. [online]. Sydney, USA, 2003 [cit. 2019-30-04]. Dostupné z: <https://www.aclweb.org/anthology/W03-2305>
- [11] REITER Ehud - SRIPADA, Somayajulu – HUNTER, Jim – YU, Jin – DAVY, Ian. *Choosing words in computer-generated weather forecasts* [online]. United Kingdom, 2005 [cit. 2019-31-04]. Dostupné z: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.3627&rep=rep1&type=pdf>
- [12] ALZA. *Arduino UNO Rev3* [online]. [cit. 2019-01-05]. Dostupné z: <https://cdn.alza.cz/ImgW.ashx?fd=FotoAddOrig&cd=ARD204a-03&i=1.jpg>
- [13] ORAVEC, Miroslav. *ARDUINO - programujeme mikrokontroléry /1.část* [online]. 2018 [cit. 2019-01-05]. Dostupné z: <https://www.pcrevue.sk/a/ARDUINO---programujeme-mikrokontrolery-1-cast>
- [14] MAHMOOD, Sarmad Nozad – HASAN, Forat Falih. *Design of Weather Monitoring System Using Arduino Based Database Implementation* [online]. Iraq, 2017 [cit. 2019-01-05]. Dostupné z: https://www.researchgate.net/publication/316597515_Design_of_Weather_Monitoring_System_Using_Arduino_Based_Database_Implementation_-----_httpwwwjmestorgwp-contentupl
- [15] *MATLAB* [online], [cit. 2019-01-05]. Wikipedie. Dostupné z: <https://cs.wikipedia.org/wiki/MATLAB>

- [16] DUŠEK, František. *MATLAB a SIMULINK: úvod do používání*. Pardubice: Univerzita Pardubice, 2000. ISBN 80-7194-273-1. Skripta. Univerzita Pardubice.
- [17] MATHWORKS. *Documentation* [online]. 2019 [cit. 2019-04-05]. Dostupné z: <https://www.mathworks.com/help/thingspeak/>
- [18] VŠETEČKA, Roman. *Firmy bojují o ovládnutí internetu věcí. Uživatel může prodělat* [online]. 2015 [cit. 2019-01-05]. Dostupné z: https://www.idnes.cz/technet/pc-mac/souboj-internetu-veci-iot.A150316_161724_hardware_vse
- [19] *OpenWeatherMap* [online], [cit. 2019-05-05]. Wikipedie. Dostupné z: <https://en.wikipedia.org/wiki/OpenWeatherMap>
- [20] OPENWEATHERMAP. *WeatherApi* [online]. 2019, [cit. 2019-05-05]. Dostupné z: <https://openweathermap.org/api>
- [21] PINEGAR, Grace. *What is Twitter: A Beginner's Guide* [online]. 2018, [cit. 2019-05-05]. Dostupné z: <https://learn.g2crowd.com/what-is-twitter>
- [22] PNGMART. *Microcontroller PNG Clipart* [online]. [cit. 2019-05-05]. Dostupné z: <http://www.pngmart.com/image/90218>
- [23] TWITTER. *Tweet* [online]. [cit. 2019-05-16]. Dostupné z: <https://twitter.com>
- [24] CHMI. *Základní meteorologická terminologie* [online]. [cit. 2019-05-16]. Dostupné z: <http://portal.chmi.cz/informace-pro-vas/prezentace-a-vyuka/meteorologicka-terminologie>
- [25] RTV:. *Počasié* [online]. 20. 7. 2018, [cit. 2019-05-16]. Dostupné z : <https://www.rtvsk.sk/televizia/archiv/13984/161801#138>
- [26] RTV:. *Počasié* [online]. 19. 10. 2018, [cit. 2019-05-16]. Dostupné z : <https://www.rtvsk.sk/televizia/archiv/13984/169781#329>
- [27] RTV:. *Počasié* [online]. 26. 1. 2019, [cit. 2019-05-16]. Dostupné z : <https://www.rtvsk.sk/televizia/archiv/13984/177823#141>
- [28] ČT. *Předpověď počasí* [online]. 24. 1. 2019, [cit. 2019-05-16]. Dostupné z : <https://www.ceskatelevize.cz/porady/1178166999-predpoved-pocasi/219411000410124/>
- [29] TA3. *POČASIE: Predpoveď na dnes* [online]. 4. 5. 2019, [cit. 2019-05-16]. Dostupné z : <https://www.ta3.com/clanok/1154193/pocasio-predpoved-na-dnes.html>
- [30] RTV:. *Počasié* [online]. 6. 5. 2019, [cit. 2019-05-16]. Dostupné z : <https://www.rtvsk.sk/televizia/archiv/13984/186971#336>
- [31] MARKÍZA. *Počasié* [online]. 7. 5. 2019, [cit. 2019-05-16]. Dostupné z : <http://videoarchiv.markiza.sk/video/pocasio/epizoda/21309-pocasio>
- [32] SHMÚ. *Rozšířená predpoveď počasí* [online]. 14. 5. 2019, [cit. 2019-05-16]. Dostupné z : http://www.shmu.sk/sk/?page=1&id=meteo_tpredpoved_skcele

:

9 ZOZNAM SKRATIEK, SYMBOLOV, OBRÁZKOV A TABULIEK

Obr. 1: Hlavná architektúra automaticky generovaných textov. [vlastné spracovanie podľa 10]	20
Obr. 2: Prepojenie jazyka so svetom prostredníctvom senzorov. [vlastné spracovanie podľa 11]	25
Obr. 3: Znázornenie frekventovaného výzoru samotného mikrokontroléru. [22] .	27
Obr. 4: Zobrazenie hardvéru vývojovej dosky Arduino UNO. [12]	29
Obr. 5: Ceny balíčkov a ich obsah s výhodami a nevýhodami. Ponuka pochádza priamo zo stránok OpenWeatherMap. [20]	33
Obr. 6: Schéma štruktúry plánovaného obsahu.	38
Obr. 7: Súčasť - Pozdrav.....	40
Obr. 8: Súčasť – Zdvorilostná formulka - začiatočná.....	40
Obr. 9: Sekcia Teplota.....	41
Obr. 10: Sekcia Stav vetra.	41
Obr. 11: Sekcia Zrážky a hmla.....	42
Obr. 12: Sekcia Stav oblohy.....	42
Obr. 13: Sekcia Vlhkosť.....	42
Obr. 14: Súčasť – Odporúčacia veta.....	43
Obr. 15: Súčasť – Zdvorilostná formulka – koncová.....	43
Obr. 16: Súčasť – Pozdrav na koniec.....	43
Obr. 17: 1. vetný rámec.	44
Obr. 18: 2. vetný rámec.	44
Obr. 19: 3. vetný rámec.	45
Obr. 20: 4. vetný rámec.	45
Obr. 21: Príklady výsledného tweetu na sociálnej sieti.....	48

10 ZOZNAM PRÍLOH

Príloha 1: Korpus pre generovanie textu

Príloha 2: Dátový nosič (CD) s digitálnou verziou práce a kódom pre aplikáciu ThingSpeak

PRÍLOHY

Príloha 1: Korpus pre generovanie textu

20. 7. 2018

„Dobrý večer, pripravené máme aktuálne informácie o počasí. [...] V noci bude jasno až polooblačno, pri zväčšenej oblačnosti spočiatku ešte prehánky alebo slabý dážď. Do rána poklesne teplota na 17 až 12, miestami na 10 stupňov. Fúkať bude slabý vietor. [...] V južnej časti západného Slovenska predpokladáme zajtra spočiatku slnečný deň, popoludní pribudne oblačnosť a vo večerných hodinách sa ojedinele vyskytnú prehánky alebo búrky. Maximum teploty dosiahne 29 až 31 stupňov. [...] S touto informáciou sa s vami lúčim a prajem vám pekný večer a pokojný víkend.“ [25]

19. 10. 2018

„Dobrý večer. Vitajte pri predpovedi. [...] V noci bude polooblačno, ojedinele bude pršať. K ránu výnimočne sa vyskytne aj hmla a najnižšia teplota dosiahne 11 až 6 stupňov, ale v údoliach bude mestami aj o niečo chladnejšie. Viat' bude severný vietor s rýchlosťou do 20 kilometrov za hodinu. Na strednom Slovensku a v Nitrianskom kraji očakávame miestami len slabý a premenlivý vietor. Aj zajtra cez deň bude polooblačno až oblačno a ojedinele bude slabo pršať. Na severe bude prevládať oblačná až zamračená obloha, a tu sa zrážky vyskytnú miestami. Najvyššia teplota 16 až 20 v Žilinskom kraji a na Spiši miestami len 12 až 15 stupňov Celzia. [...] V každom prípade dúfam, že vám počasie náladu cez víkend nepokazí, a že naberiete nové sily. Aj keď v nedeľu už bude cez deň výrazne chladnejšie, zajtrajšok bude ešte relatívne príjemný a pršať bude len ojedinele. Želám vám ešte pekný zvyšok piatka.“ [26]

26. 1. 2019

„Dobrý večer. Vitajte pri aktuálnych informáciách o počasí. [...] V noci bude oblačno až zamračené a na viacerých miestach rátaťte so snežením. Na západe sa vyskytne aj dážď alebo dážď so snehom. Najnižšia teplota dosiahne mínus tri až mínus osem, na západe a na juhu stredného Slovenska plus 1 až mínus tri stupne. Fúkať bude slabý, na juhozápade spočiatku mierny západný vietor. Pocitová teplota bude blízka reálnej, na juhozápade o niečo nižšia. Zajtra bude oblačno až zamračené a hmlisto, miestami sa však oblačnosť zmenší. Ojedinele sa môže vyskytnúť slabé sneženie v nižších polohách aj mrholenie. [...] S touto informáciou tak s vami lúčim a prajem vám ešte pekný večer. Dovidenia.“ [27]

24. 1. 2019

„Dobrý večer pripravené pravidelná počasí, já vás vítám. [...] V sobotu a v nedeľi už budú teploty vyššie, takže odpoledne by se teplota měla dostávat na nulu jenom v sobotu. Na severovýchodě může být odpoledne ještě mrazivější, s teplotou kolem mínus jednoho stupně Celzia. Rána ale zůstanou mrazivá [...] V sobotu a v nedeľi se teplota v Česku

může dostat až na čtyři stupně Celsia a tady už je vidět, že převládat bude spíše zataženo až oblačno. O počasí poslední informace, přeju hezký večer.“ [28]

4. 5. 2019

„Oblačno až zamračené, postupne od juhozápadu na mnohých miestach dážď, prehánky, ojedinele aj búrky. Vo vysokých polohách sneženie. Najvyššia denná teplota 12 až 17, v Žilinskom kraji a na Spiši okolo 10 stupňov. Teplota na horách vo výške 1500 m okolo 3 stupňov. Slabý, na východe prevažne juhozápadný vietor 2 až 6 m/s (5 až 20 km/h). Predpokladané množstvo zrážok: do 10 mm, miestami, najmä na strednom Slovensku, 10 až 20 mm.“ [29]

6. 5. 2019

„Dobry večer. Vitajte pri predpovedi počasia. [...] Streda prinesie viac slnka a už spomínané studené ráno. Popoludní od západu opäť oblačnosť pribudne. Maximum teploty dosiahne 15 až 20 na severe miestami 13 stupňov. Štvrtok bude oblačný až zamračený a na viacerých miestach rátať s dažďom či prehánkami. Teplotne bude približne o 4 stupne chladnejšie ako v stredu. V piatok bude polooblačno až oblačno a miestami, najmä popoludní, sa vyskytnú prehánky, ojedinele aj búrky. Najnižšia teplota dosiahne 13 až 20 stupňov. [...] Počas víkendových dní bude spočiatku menej oblačnosti, postupne sa však zamračí a na mnohých miestach bude pršať. Zatiaľ čo v sobotu vystúpi teplota na juhu nad 20 stupňov, v nedeľu bude o niekoľko stupňov chladnejšie. [...] S touto informáciou sa dnes s vami lúčim a prajem vám ešte pekný večer.“ [30]

7. 5. 2019

„Vietor je dôležitým činiteľom, podľa ktorého môžeme určiť, aké počasie nás čaká. Označujeme ho podľa toho odkiaľ fúka. 7. mája v roku 1774 sa narodil syr Francis Beaufort. Bol to admirál britského námorníctva a autor stupnice pre odhad sily vetra, bez použitia prístrojov, teda podľa účinkov vetra na rôzne objekty. Má 12 stupňov a používame ju aj my v predpovedi počasia. Dobry večer. Tlakovú níž, ktorá so svojou oblačnosťou ovplyvňovala počasie aj u nás, vystriedal už v noci od západu výbežok tlakovej výše. Tlaková výš ovplyvňovala počasie u nás aj dnes a bude aj zajtra. Užite si preto sviatočný deň, lebo od západu sa nás chystá tento frontálny systém. Zrážky sa zajtra neočakávajú, ale zrána sa lokálne môžu tvoriť hmly. Teplotne sa bude naša krajina nachádzať v pásme od 15 do 20, [...] Tolko počasie. Dovidenia.“ [31]

14. 5. 2019

„Oblačno až zamračené. Na mnohých miestach dážď alebo prehánky, spočiatku od cca 800 m sneženie. Hranica sneženia bude od juhovýchodu postupne stúpať. Veterno a mimoriadne chladno. Najvyššia denná teplota 4 až 9, v južnej polovici stredného a východného Slovenska miestami 9 až 13 st. Teplota na horách vo výške 1500 m od -3 st. v Malej Fatre do +2 st. v Nízkych Tatrách, od juhovýchodu postupne otepľovanie. Prevažne severný vietor 4 až 11, na juhovýchode do 15 m/s (15 až 40, 55 km/h), v

nárazoch okolo 15 m/s, na juhovýchode až do 22 m/s (55, 80 km/h). Na strednom Slovensku miestami slabý vietor. Predpokladané množstvo zrážok: do 6 mm, na východe do 10 mm.“ [32]