



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

**BAKALÁŘSKÁ PRÁCE**

Daniela Jurášová

**Automatické generovanie hrebeňoviek**

Ústav formální a aplikované lingvistiky

Vedoucí bakalářské práce: Mgr. Rudolf Rosa, Ph.D.

Studijní program: Informatika

Studijní obor: Obecná informatika

Praha 2021

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Ďakujem vedúcemu tejto práce, Mgr. Rudolfovi Rosovi, Phd., za mimoriadne pozitívny prístup a motiváciu. Ďakujem mamke a mojej rodine, za nekonečnú podporu počas štúdia. V neposlednom rade ďakujem spolužiakom, za spoločné roky bakalárskeho štúdia plné radostí aj starostí, a za vzájomnú pomoc, vďaka ktorej mohla táto práca vzniknúť.

Název práce: Automatické generovanie hrebeňoviek

Autor: Daniela Jurášová

Ústav: Ústav formální a aplikované lingvistiky

Vedoucí bakalářské práce: Mgr. Rudolf Rosa, Ph.D., Ústav formální a aplikované lingvistiky

Abstrakt: Cieľom tejto práce je predstavenie programu pre automatické generovanie tematických hrebeňoviek. Hrebeňovka je typ krížovky, v ktorej niekoľko slov umiestnených v stĺpcoch bezprostredne vedľa seba, vytvorí v určenom riadku tajničku. Výsledný program má formu počítačovej aplikácie s jednoduchým grafickým prostredím. Užívateľ si na začiatok zvolí tému krížovky z dostupného zoznamu. Program následne použitím voľne prístupnej databázy WordNet vygeneruje krížovku obsahujúcu slová týkajúce sa danej témy. Zároveň vytvorí legendu pre riešenie krížovky. Súčasný návrh je oproti pôvodnému zadaniu rozšírený o možnosť definovať si vlastnú tému. Databáza slov je rozšírená o slovenský WordNet, takže užívateľ si môže vybrať medzi dvomi jazykmi slov dopĺňaných do krížovky.

Klíčová slova: krížovky téma WordNet

Title: Automatic generation of crosswords

Author: Daniela Jurášová

Institute: Institute of Formal and Applied Linguistics

Supervisor: Mgr. Rudolf Rosa, Ph.D., Institute of Formal and Applied Linguistics

Abstract: The aim of this work is to present a program for automatic generation of thematic crosswords. The program generates a specific type of crossword puzzle, where a few words in columns placed right next to each other create a secret word in a specified row.

The resulting program is in the form of a computer application with a simple graphical environment. The user initially chooses a crossword puzzle topic from the available list. The program then uses a freely accessible WordNet database to generate a crossword puzzle containing words related to the topic. At the same time, it will create a legend for solving the crossword puzzle.

Compared to the original assignment, the current proposal is extended by the possibility to define your own topic. The word database is extended by Slovak WordNet, therefore the user can choose between two languages of words filled in the crossword puzzle.

Keywords: crosswords topic WordNet

# Obsah

<b>1</b>	<b>Úvod</b>	<b>3</b>
<b>2</b>	<b>Motivácia</b>	<b>4</b>
2.1	Druhy krížoviek . . . . .	4
2.2	Krátka história slovných krížoviek . . . . .	5
<b>3</b>	<b>Známe prístupy k riešeniu úlohy</b>	<b>7</b>
3.1	Techniky spracovania prirodzených jazykov . . . . .	7
3.2	Programovanie s obmedzujúcimi podmienkami . . . . .	8
3.3	Použitie voľne prístupných databáz . . . . .	8
<b>4</b>	<b>WordNet</b>	<b>10</b>
4.1	Anglický WordNet . . . . .	10
4.1.1	Štruktúra . . . . .	11
4.2	Slovenský WordNet . . . . .	13
4.2.1	Štruktúra . . . . .	13
<b>5</b>	<b>Implementácia</b>	<b>15</b>
5.1	Serializácia a deserializácia . . . . .	15
5.1.1	Formát JSON . . . . .	15
5.2	Preprocessing . . . . .	16
5.2.1	Techniky mapovania slovenských a anglických synsetov . . . . .	18
5.2.2	WordNet API . . . . .	20
5.3	Zostavenie tematického slovníka . . . . .	21
5.3.1	Vymedzenie slov týkajúcich sa témy . . . . .	22
5.3.2	Zvažované varianty algoritmu . . . . .	22
5.4	Tvorba krížoviek . . . . .	23
5.5	Grafické prostredie . . . . .	23
<b>6</b>	<b>Vývojová dokumentácia</b>	<b>25</b>
6.1	Word Source Maker . . . . .	25
6.1.1	Potrebné súbory . . . . .	25
6.1.2	Trieda Record . . . . .	25
6.1.3	Trieda Program . . . . .	26
6.1.4	Výstup . . . . .	30
6.2	Thematic Crosswords . . . . .	30
6.2.1	Stavy aplikácie . . . . .	30
6.2.2	Obsluha interaktívnych komponent . . . . .	33
6.2.3	Prehľadávanie databázy . . . . .	36
6.2.4	Tvorba krížoviek . . . . .	37
6.3	Užitočné zdroje . . . . .	39
<b>7</b>	<b>Užívateľská dokumentácia</b>	<b>41</b>
7.1	Úvod . . . . .	41
7.2	Voľba jazyka . . . . .	41
7.3	Voľba témy . . . . .	42

7.4	Čakanie . . . . .	44
7.5	Pravidlá . . . . .	45
7.6	Krížovka . . . . .	46
7.7	Tlačidlá . . . . .	47
<b>8</b>	<b>Odozva užívateľov</b>	<b>50</b>
8.1	Číselné vyhodnotenie . . . . .	50
8.2	Podnety a návrhy . . . . .	52
8.3	Možné vylepšenia . . . . .	53
	<b>Záver</b>	<b>54</b>
	<b>Zoznam použitej literatúry</b>	<b>55</b>
<b>A</b>	<b>Prílohy</b>	<b>56</b>
A.1	WordSourceMaker . . . . .	56
A.2	ThematicCrosswords . . . . .	56
A.3	TCApplcation . . . . .	57
A.4	Licencie . . . . .	58

# 1. Úvod

Hrebeňovka je typ krížovky, v ktorej niekoľko slov v stĺpcoch umiestnených bezprostredne vedľa seba vytvorí v určenom riadku tajničku. Hrebeňovky a iné podobne tvarom jednoduché krížovky nachádzajú uplatnenie predovšetkým vo vzdelávaní, kde sa využívajú pri výučbe cudzích jazykov. Súčasnú učebnicu cudzích jazykov sú zvyčajne zostavené podľa tematických celkov. Ku každej téme má študent k dispozícii súbor súvisiacich slov. Tie môže používať v rôznych cvičeniach, medzi ktoré patria krížovky. Väčšina učebníc a pracovných zošitov má však k dispozícii len jednu krížovku na tému, ktorá navyše nepokrýva všetky výrazy. Študent tak po vyplnení krížovky nemá iný priestor na precvičenie slovíčok.

Cieľom tejto práce je vytvoriť jednoduchú počítačovú aplikáciu, ktorá automaticky generuje tematické hrebeňovky. Užívateľ na úvod zadá požadovanú tému. Následne sa automaticky vygeneruje krížovka so slovami týkajúcimi sa tejto témy. Aplikácia umožňuje generovať krížovky na ľubovoľnú tému alebo opakovane generovať rôzne zadania krížoviek na rovnakú tému.

Cieľovou skupinou sú študenti, ktorí používaním aplikácie získajú prístup k neobmedzenému množstvu krížoviek na precvičovanie slovné zásoby. Aplikáciu môžu využívať nielen študenti cudzích jazykov, ale aj maturanti alebo študenti vedných disciplín, ktorí sa potrebujú zamerať na konkrétny súbor odborných slov. Uplatnenie nájde aj u vyučujúcich, ktorí pomocou aplikácie môžu pripravovať materiály pre svojich študentov.

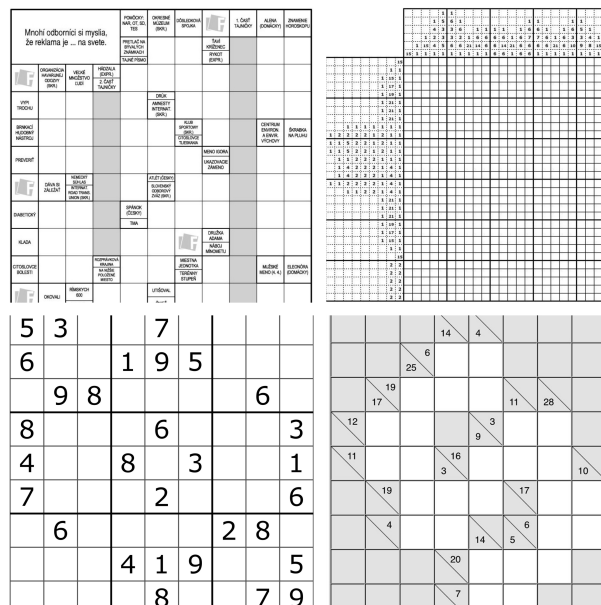
## 2. Motivácia

Križovky už po viac ako sto rokov zábavným spôsobom rozširujú vedomosti riešiteľov. Definujeme ich ako logické hry založené na križovaní výrazov vpisovaných do obrazca. Cieľom riešenia križovky je správne vyplniť obrazec a vylúštiť tak tajničku. V tejto kapitole si predstavíme základné druhy križoviek a ich históriu.

### 2.1 Druhy križoviek

Podľa typu vpisovaného (alebo hľadaného) výrazu rozlišujeme tri hlavné druhy križoviek:

- slovné križovky, kde výraz pozostáva z písmen
  - klasické a švédske križovky
  - osemsmierovky
- obrazové križovky, kde výrazom rozumieme postupnosť štvorčiekov, z ktorých riešiteľ niektoré vymaľuje
  - maľované križovky (nonogramy)
- číselné križovky, kde výraz pozostáva z číslíc zo špecifického intervalu
  - sudoku
  - kakuro



Obr. 2.1: Druhy križoviek. Zľava doprava švédska križovka, nonogram, sudoku, kakuro.



Krížovky môžeme nájsť v rôznych formátoch. Najčastejšie sa vyskytujú ako špeciálne stránky v novinách alebo ako pravidelne vychádzajúce mesačníky a týždenníky. Nemenej populárne sú krížovkárske knihy a aplikácie, ktoré prevyšujú ostatné formáty v kvantite a rozmanitosti krížoviek.

Pre účely tejto práce definujeme krížovku ako slovný hlavolam, pozostávajúci z buniek na mriežke. V krížovke sa vyskytujú 3 druhy buniek :

- prázdne bunky, kam je potrebné dopĺňať písmená, a ktoré ďalej môžeme deliť na obyčajné bunky a bunky tajničky
- bunky s vysvetlivkou, umiestnené na začiatku riadkov alebo stĺpcov (legenda môže byť umiestnená mimo mriežky)
- neplatné bunky, slúžiace na vyplnenie nepotrebného miesta na mriežke

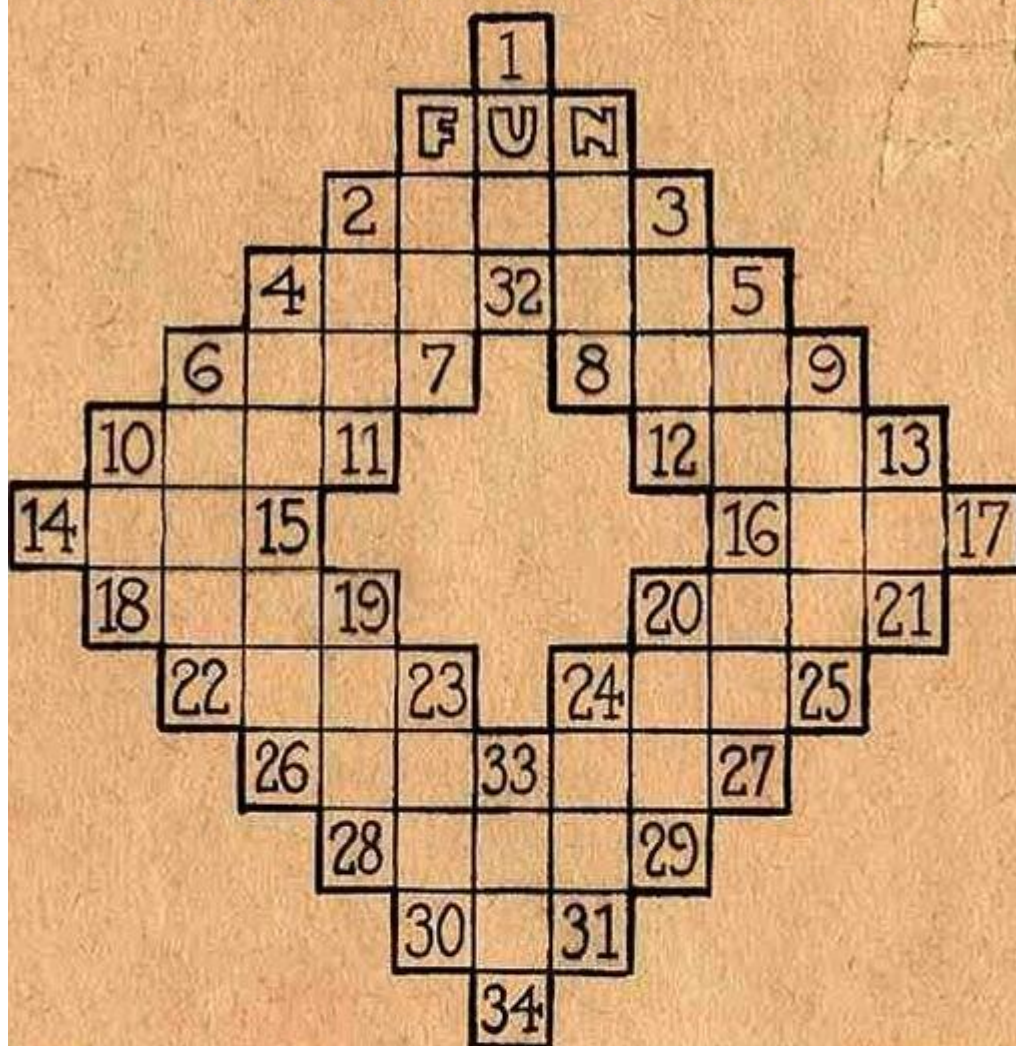
## 2.2 Krátka história slovných krížoviek

Prvá krížovka bola uverejnená v roku 1913 v tzv. zábavnej sekcii denníka New York World. Mala tvar diamantu a jej tvorca, Arthur Wynne, ju pomenoval „Word-Cross“ (obrázok 2.2). Wynne sa považuje za otca krížoviek, keďže svojím výtvorom vzbudil záujem širokej verejnosti. Napriek narastajúcej popularite krížoviek vznikli univerzálne štandardy pre ich konštrukciu až v štyridsiatych rokoch dvadsiateho storočia. Neskôr, v roku 1942 sa krížovky stali pravidelnou súčasťou dennej tlače.

V úvode vzniku krížoviek si ich riešenie vyžadovalo znalosti súboru faktov a myšlienok. Prvé krížovky sa týkali geografie, poézie, ale aj latinčiny, preto boli sprvu určené pre vzdelanú spoločnosť. Neskôr sa predmet krížoviek prispôboval moderným témam, alebo celkovému zameraniu periodika, v ktorom vychádzali. (Tausig, 2013)

V tejto kapitole sme si v krátkosti predstavili základné druhy krížoviek a ich históriu. V súčasnosti si môžeme vyberať zo širokej škály typov krížoviek. Tematika krížoviek je zameraná na tajničky, kde sú obľúbené najmä tajničky tvoriace komplexné kriminálne príbehy, alebo dopĺňajúce ľudové múdrosti, vtipy a recepty. V tejto práci sa budeme venovať krížovkám, v ktorých sa okrem tajničky viažu k určitej téme aj dopĺňané slová. Ako formu krížoviek sme si zvolili špeciálny typ - hrebeňovky.

## FUN'S Word-Cross Puzzle.



**F**ILL in the small squares with words which agree with the following definitions:

2-3. What bargain hunters enjoy.

4-5. A written acknowledgment.

6-7. Such and nothing more.

10-11. A bird.

14-15. Opposed to less.

18-19. What this puzzle is.

22-23. An animal of prey.

26-27. The close of a day.

28-29. To elude.

30-31. The plural of is.

8-9. To cultivate.

12-13. A bar of wood or iron.

16-17. What artists learn to do.

20-21. Fastened.

24-25. Found on the seashore.

10-18. The fibre of the gomuti palm.

6-22. What we all should be.

4-26. A day dream.

2-11. A talon.

19-28. A pigeon.

F-7. Part of your head.

23-30. A river in Russia.

1-32. To govern.

32-34. An aromatic plant.

N-8. A fist.

24-31. To agree with.

3-12. Part of a ship.

20-29. One.

5-27. Exchanging.

9-25. To sink in mud.

13-31. A boy.

Obr. 2.2: Prvá krížovka, Wynne (1913), v denníku New York World.

# 3. Známe prístupy k riešeniu úlohy

Pred tvorbou samotnej aplikácie sme si museli zodpovedať dve kľúčové otázky:

- Čo bude zdrojom dát ?
- Akým spôsobom budeme tvoriť krížovku ?

V nasledujúcich sekciách prezentujeme možné odpovede na tieto otázky.

## 3.1 Techniky spracovania prirodzených jazykov

Rigutini, Diligenti, Maggini a Gori (2008) predstavili vo svojej práci softvér pre plne automatické generovanie krížoviek. Softvér pozostáva z dvoch modulov, z ktorých jeden extrahuje definície z internetových stránok a druhý generuje schémy krížovky. Modul generovania schémy krížovky popíšeme v ďalšej podsekcii.

Modul pre extrakciu definícií sťahuje stránky z dostupných internetových zdrojov. Čistý text z týchto stránok posúva ďalej na analýzu. Analýza textu prebieha použitím techník spracovania prirodzených jazykov a metód strojového učenia, a to v niekoľkých vrstvách. Vybrané techniky sa na text aplikujú postupne. To znamená, že každá vrstva analýzy textu aplikuje určenú techniku na výstup predchádzajúcej vrstvy. Prvou z používaných techník je *Part-of-speech tagging*, čo je proces pridelovania značiek k slovám vo vete. Značka obsahuje gramatické vlastnosti slova, konkrétne k akému slovnému druhu patrí. Druhá technika zoskupuje výrazy a identifikuje základné zložky vety. V teórii gramatiky rozlišujeme tri hlavné zložky vety, a to mennú frázu, slovesnú frázu a predložkovú frázu. Definované skupiny výrazov ďalej spracováva tretia technika, *syntaktická analýza*. Každá skupina priradí jednu syntaktickú značku spomedzi: *podmet*, *určovací sklad<sup>1</sup> medzi prísudkom a predmetom alebo príslovkovým určením*, *prisudzovací sklad<sup>2</sup>*, *iné*. Podľa zistení výpočtovej skupiny Univerzity v Siene, 80% definícií pozostáva z podmetu nasledovaného popísaným určovacím skladom a doplnkami. Posledná technika, *extrakcia definícií*, skúma výsledky analýzy predchádzajúcich vrstiev. Na základe týchto výsledkov a spomenutého zistenia sa rozhoduje, či je daná fráza definíciou. Ak je fráza klasifikovaná ako definícia, modul ju uloží do výslednej databázy.

Modul pre extrakciu definícií bol experimentálne spustený na stránkach talianskej Wikipédie. Vo výsledkoch práce sa uvádza, že percento nevhodných alebo nesprávnych definícií bolo vďaka spracovaniu a analýze textu veľmi nízke. Problémom však boli definície, ktoré síce boli správne, ale neboli vhodné pre krížovky. Boli príliš dlhé alebo používali výrazy týkajúce sa riešenia danej nápovedy, prípadne boli nejednoznačné. Porovnaním 1000 náhodných definícií vygenerovaných modulom a 1000 definícií z existujúcich krížoviek zistili, že je modul schopný generovať realistické krížovky. Preto bol v konečnom dôsledku modul označený za úspešný.

---

<sup>1</sup>vzťah medzi nadradeným a podradeným vetným členom

<sup>2</sup>vzťah podmetu a prísudku (podmet môže byť zamlčaný)

## 3.2 Programovanie s obmedzujúcimi podmienkami

V module generovania schémy krížovky využívajú Rigutini a kol. (2008) postupy programovania s obmedzujúcimi podmienkami. Vytvorenie krížovky z predom daného slovníka sa považuje za časovo náročný vyhľadávací problém určený pre umelú inteligenciu. Pre vytvorenie väčších krížoviek je nutné používať heuristiky<sup>3</sup> na prehľadávanie slovníka. Prehľadávanie hrubou silou môže byť použité len k vytvoreniu menších schém (zhruba 5x5).

Pre účely tvorby schémy krížovky si Rigutini a kol. (2008) definovali „čierny box“, teda pole, ktoré nesmie byť nikdy prepísané, a „biely box“, teda pole, ktoré môže obsahovať hlásku. Horizontálna alebo vertikálna postupnosť *bielych boxov* definuje slot, kam môže byť vložené slovo. Umiestnenie slova je potom obmedzené hranicami schémy krížovky alebo *čiernym boxom*. Slová, ktoré už majú určené svoje umiestnenie v schéme, vytvárajú dodatočné podmienky pre umiestnenie ďalších slov. Obmedzenia týkajúce sa tvaru a veľkosti schémy krížovky sú nemenné, zatiaľ čo obmedzenia týkajúce sa dopĺňaných slov sú vytvárané postupne. Jadro systému pre tvorenie schémy krížovky pozostáva z prioritného frontu<sup>4</sup>, ktorý skladuje množinu čiastočných riešení. V každom kroku tvorenia schémy je z frontu vybraté jedno čiastočné riešenie. Pre zvolený voľný slot v danom riešení sa zo slovníka získa zoznam vhodných kandidátov. Náhodným výberom sa zo zoznamu zvolí slovo, ktoré je doplnené do schémy. V tomto kroku nastupuje heuristika, ktorá po každom doplnení slova priradí schéme skóre na základe toho, ako blízko je k úplnému zaplneniu slovami. Zohľadňuje aj pravdepodobnosť javu, že schéma môže byť naplnená. Z dostupných čiastočných riešení sa potom vyberie to s najlepším skóre.

## 3.3 Použitie voľne prístupných databáz

Okrem automatického získavania dát z internetových stránok sa nám ponúka možnosť využitia databáz. Ideálne by bolo použitie výkladových slovníkov. Oficiálne výkladové slovníky jednotlivých jazykov zvyčajne nie sú voľne prístupné, rovnako tak ich forma nie je strojovo dobre spracovateľná. Potrebujeme voľne prístupnú databázu, ktorá by obsahom predstavovala výkladový slovník a mala formu vhodnú pre automatické spracovanie. Jednou z možností je WordNet.

WordNet je lexikálna databáza, v ktorej sú slová prepojené pomocou sémantických vzťahov, a z ktorej je možné extrahovať výrazy spolu s definíciami. Bližším preskúmaním tejto databázy sme zistili, že WordNet je stále aktuálny projekt, ktorý nachádza množstvo uplatnení. Na medzinárodnej konferencii k WordNetu (Sojka a kol., 2004) boli predstavené desiatky projektov, ktoré používajú WordNet ako zdroj dát, alebo sa snažia WordNet skúmať a pridávať nové funkcie. Medzi zaujímavé projekty patrí napríklad tvorba lexikálnych databáz inšpirovaných WordNetom v cudzích jazykoch (španielčina, nemčina, taliančina), budovanie te-

<sup>3</sup>metóda riešenia problémov, ktorá nie je zaručene optimálna, ale poskytuje dobré riešenie

<sup>4</sup>dátová štruktúra typu „prvý dnu, prvý von“ - čo prvé sa do štruktúry vloží, to sa ako prvé použije, respektíve odstráni

zaurov<sup>5</sup> za pomoci WordNetu, alebo skúmanie štatistických vlastností WordNetu, ktorých výsledky neskôr využijeme v našej práci.

Okrem najznámejšej verzie WordNetu od americkej univerzity Princeton existuje aj slovenská verzia WordNetu. O plánovanom vytvorení slovenskej databázy WordNet pojednáva Garabík (2010). Databáza má slúžiť ako základ pre ďalší projekt vývoja medzinárodného slovníka. Neskôr bol cieľ vytvorenia databázy naplnený a Šimková a Garabík (2014) opisujú v diele o rozvoji jazykových technológií na Slovensku početné zaradenie slovenského WordNetu v celoeurópskych ale aj medzinárodných lingvistických projektoch.

V tejto kapitole sme si predstavili možný spôsob získavania dát do krížoviek pomocou strojového učenia, alebo použitím voľne prístupných databáz. Za stavebný prvok aplikácie sme si zvolili databázu WordNet, ktorá slúži ako primárny zdroj dát pre krížovky. Predstavili sme si aj spôsob tvorby schém krížoviek. Klasické švédske krížovky, tvorené týmto spôsobom, sú pre účely našej aplikácie zbytočne zložité. Keďže krížovky môžu byť lúštené aj mladšími užívateľmi, zamerali sme sa na špeciálny, jednoduchší tvar, hrebeňovky.

---

<sup>5</sup>súbor termínov usporiadaných do hierarchickej štruktúry s definovanými vzťahmi medzi termínmi

## 4. WordNet

WordNet je lexikálna databáza, tzv. pojmová sieť, organizovaná podľa významu slov. Slová sú prepojené pomocou sémantických vzťahov ako synonymia, hyponymia a meronymia. Prvý WordNet bol vytvorený v anglickom jazyku na Univerzite Princeton v osemdesiatych rokoch dvadsiateho storočia. V súčasnosti je k dispozícii viac ako 200 jazykových variánt. (Wikipedia, 2021b) WordNet sa využíva najmä v lingvistike a pri spracovaní prirodzeného jazyka.

WordNet je stavebným prvkom našej aplikácie. V tejto práci používame dve jazykové verzie, anglickú a slovenskú. V nasledujúcich sekciách si predstavíme oba používané varianty WordNetu, ich štruktúru a sémantické vzťahy, ktoré reprezentujú.

### 4.1 Anglický WordNet

Princeton WordNet je lexikálna databáza anglického jazyka. Obsahuje podstatné mená, prídavné mená, príslovky a slovesá. Hlavným vzťahom medzi slovami vo WordNete je synonymia. Synonymá sú slová, ktoré majú rôznu formu ale podobný alebo rovnaký význam, a je možné ich v rámci kontextu vymieňať. Synonymá sú zoskupené v neusporiadaných množinách, takzvaných synsetoch (obrázok 4.1). Teda synset pre učité slovo je množina jemu významovo podobných slov. Slovo sa nachádza v toľkých synsetoch, koľko rôznych významov reprezentuje. Preto každá dvojica slova a významu je jednoznačná. WordNet obsahuje viac ako 117 000 synsetov, ktoré sú prepojené pomocou sémantických a lexikálnych vzťahov a je možné sa medzi nimi pohybovať.

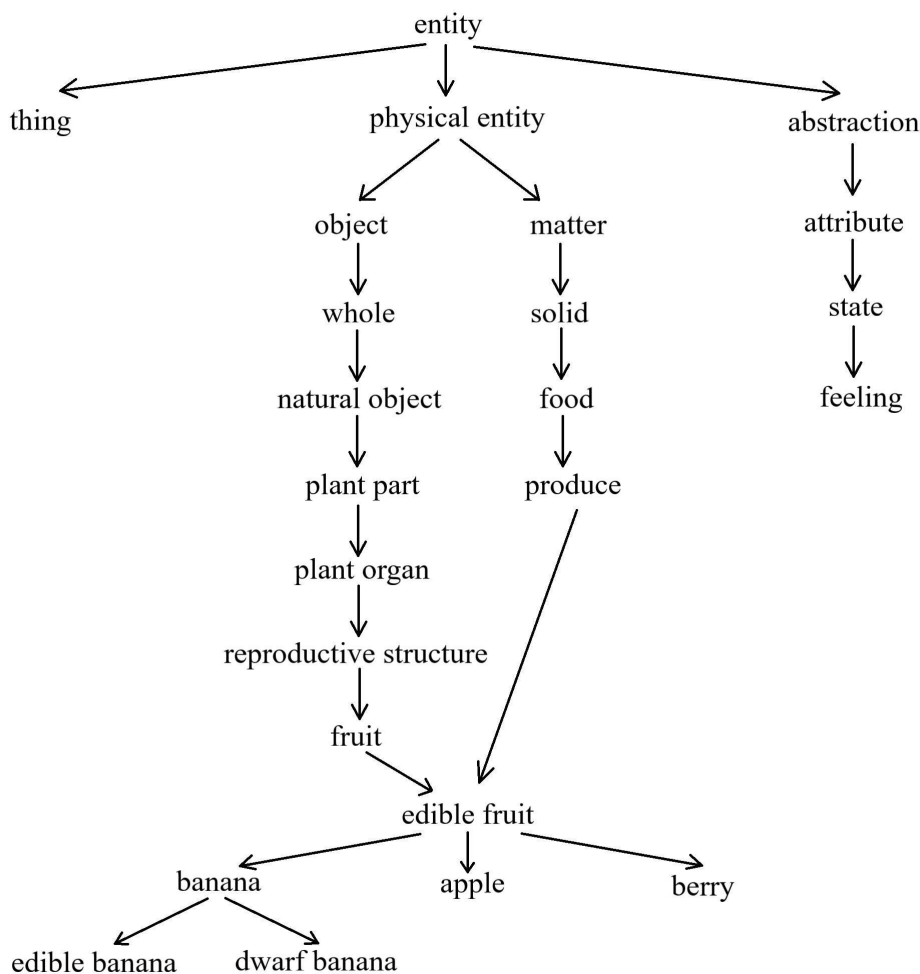
Obr. 4.1: Vizualizácia synsetu pre slovo „entity“

*entity* → {*physical entity, abstraction, thing*}

Dôležitým vzťahom medzi synsetmi je významová podradenosť (hyponymia) a významová nadradenosť (hyperonymia). Hyperonymum je všeobecnejšie označenie hyponyma. Naopak hyponymum je konkretizáciou hyperonyma. Napríklad slovo *kvet* je hyperonymom slova *tulipán*, pretože tulipán je špecifický druh kvetu. Slovo *stôl* je hyponymom slova *nábytok*, pretože nábytok súhrnne označuje interiérové zariadenie, medzi ktoré patrí aj stôl. Podobným vzťahom je meronymia predstavujúca časť celku. Slovo *palec* je meronymom slova *ruka*.

Všetky dostupné slová sú usporiadané v hierachii podľa príslušných vzťahov, a je možné po tejto hierarchii postupovať. (Miller a kol., 1990) Príklad hierarchickej štruktúry môžeme vidieť na obrázku 4.2.

Obr. 4.2: Príklad hierarchickej štruktúry slov z WordNetu



### 4.1.1 Štruktúra

Princeton WordNet je podľa potreby možné používať tromi spôsobmi, ako:

- online nástroj na vyhľadávanie slov
- počítačová aplikácia na vyhľadávanie slov
- databázové súbory pre použitie pri vývoji aplikácií

Pre potreby tejto práce sme si zvolili použitie prostredníctvom databázových súborov, ktoré sme si stiahli v komprimovanom súbore z oficiálnej stránky WordNetu. (Princeton University, 2021) Hlavný obsah databázy WordNet je reprezentovaný súbormi `index.pos` a `data.pos`, kde `pos` označuje prítomné slovné druhy (podstatné mená, prídavné mená, príslovky, slovesá). Okrem toho WordNet ponúka niekoľko pomocných, lexikografických súborov, ktoré sú používané aplikáciami na prehľadávanie databázy. Lexikografické súbory nie sú pre našu aplikáciu potrebné.

Vo výslednej aplikácii používame tzv. dátové súbory označené predponou `data`. Každý zo štyroch dostupných dátových súborov začína niekoľkými riadkami obsahujúcimi licenciu, upozornenie o autorských právach a číslo aktuálnej verzie. Tieto riadky začínajú dvomi medzerami. Všetky zvyšné riadky majú jednotný formát, ktorý pozostáva z nasledujúcich prvkov oddelených vždy jednou medzerou (obrázok 4.3):

- osemciferné desiatkové celé číslo reprezentujúce aktuálny bajtový ofset<sup>1</sup> v súbore
- dvojciferné desiatkové číslo, ktoré reprezentuje pomocný lexikografický súbor obsahujúci daný synset
- písmeno, určujúce slovný druh synsetu
  - **n** = podstatné mená
  - **a** = prídavné mená
  - **s** = „adjective satellite“ - koncept, ktorý je významovo podobný niektorému konceptu prídavného mena
  - **r** = príslovky
  - **v** = slovesá
- dvojciferné hexadecimálne číslo, stanovujúce počet slov v danom synsete
- slovo vo formáte ASCII<sup>2</sup> - medzery sú nahradené podčiarkovníkom, v súbore `data.adj` sa bezprostredne za slovom môže nachádzať syntaktická značka v zátvorkách
- jednociferné hexadecimálne číslo, používané pre jednoznačné určenie významu slova v pomocných lexikografických súboroch
- trojmiestne desiatkové celé číslo, označujúce počet ukazovateľov z daného synsetu do iných
- nula alebo viac ukazovateľov do iných synsetov tvaru
  - symbol ukazovateľa, popisujúci sémantický alebo lexikálny vzťah medzi aktuálnym synsetom a cieľovým synsetom, na ktorý ukazovateľ ukazuje
  - ofset cieľového synsetu v súbore, ktorý odpovedá slovnému druhu
  - slovný druh (**n**, **a**, **r**, **v**)
  - štvorbajtové pole pozostávajúce z dvoch dvojciferných hexadecimálnych čísel, rozlišujúce sémantické a lexikálne ukazovatele
- zoznam čísel zodpovedajúci generickým rámcom<sup>3</sup>, vyskytujúci sa len v súbore `data.verb`

---

<sup>1</sup>celé číslo označujúce vzdialenosť medzi začiatkom objektu a prvkom, ktorý sa v danom objekte nachádza

<sup>2</sup>kódovací systém znakov anglickej abecedy, číslic, špeciálnych znakov a riadiacich kódov

<sup>3</sup>rámec ilustruje typy jednoduchých viet, v ktorých je možné slovesá z daného synsetu používať



- glosa, teda definícia slova, ktorá je reprezentovaná ako zvislá čiara (|) nasledovaná textom, môže obsahovať jednu alebo viacero vzorových viet, ktoré ilustrujú použitie slova, a sú ohraničené úvodzovkami (“”)

Uniformná štruktúra riadkov WordNetu vytvára priestor pre automatické metódy spracovania tejto databázy a uľahčuje jej používanie. (Princeton University, 2021)

Obr. 4.3: Ukážky z dátových súborov WordNetu.

```
data.noun
08204378 14 n 01 theater_company 0 002 @ 08203951 n 0000 ~ 08204487
n 0000 | a company that produces plays
```

```
data.adj
00028294 00 s 01 declarable 0 001 & 00027360 a 0000 | that must be declared;
"declarable income"
```

```
data.verb
00031174 29 v 01 cramp 1 001 @ 00019441 v 0000 01 + 10 00 | affect with
or as if with a cramp
```

```
data.adv
00023852 02 r 01 conformably 0 001 \ 01617433 a 0101 | in a conformable
manner
```

## 4.2 Slovenský WordNet

Podobne ako anglická verzia je slovenský WordNet lexikálna databáza slov usporiadaných podľa sémantických vzťahov. Obsahuje najfrekvencovanejšie podstatné mená, prídavné mená, príslovky a slovesá. Každý záznam je spojený s ekvivalentným záznamom anglického WordNetu.

### 4.2.1 Štruktúra

Slovenský WordNet je voľne dostupný na stiahnutie ako jediný textový súbor. Každý riadok súboru zodpovedá jednému synsetu a je rozdelený na dva záznamy, slovenský a anglický (obrázok 4.4). Slovenský synset sa v súbore nachádza viackrát, ak mu zodpovedá viacero anglických synsetov.

Prvý zo záznamov je slovenský. Obsahuje štyri položky oddelené tabulátorom:

- číslo synsetu - jednoznačný identifikátor
- slovný druh - jedno z písmen spomedzi :
  - n = podstatné mená
  - a = prídavné mená
  - r = príslovky

- v = slovesá
- slovo - literál, prípadne ďalšie varianty slova oddelené bodkočiarkou, môže obsahovať:
  - zátvorku s upresnením významu slova
  - znak + pre označenie hlavného slova spomedzi synonym
  - znak - pre označenie nepresného sémantického vzťahu medzi anglickým a slovenským synsetom
  - znak ? pre označenie nejasného ekvivalentu
- glosa - definícia, resp. vysvetlenie slova (slovenská glosa je väčšinou prázdna)

Nasleduje znak s kódovým označením **U+241E** nazývaný **record separator** (oddeľovač záznamov, na obrázku 4.4 označený ako **RS**) a príslušný anglický ekvivalent záznamu. Štruktúra anglického záznamu sa dokonale zhoduje so štruktúrou záznamov Princeton WordNetu. Podobne ako pri anglickej verzii ponúka uniformná štruktúra riadkov slovenského WordNetu možnosť automaticky spracovávať túto databázu. (Slovenský Národný Korpus, 2021)

Obr. 4.4: Ukážka dát zo slovenského WordNetu.

```
716 n spoločenská hra RS 00458800 04 n 01 party__game 0 001 @ 00430606
n 0000 | a game to amuse guests at a party

2189 a rýchlonohý; rýchly RS 01031602 00 s 02 swift-footed 0 fast-footed 0
001 & 01031232 a 0000 | having rapidly moving feet

2842 v ničiť; pohlcovať RS 01565360 35 v 01 devour 0 001 @ 01564144 v
0000 01 + 11 00 | destroy completely; "Fire had devoured our home"

346 r priemerne RS 00169659 02 r 02 on_the_average 0 on_average 0 000
| typically; "on average he watches three movies a week"
```

V tejto kapitole sme si v krátkosti predstavili anglickú aj slovenskú databázu WordNet, ich štruktúru a sémantické vzťahy, ktoré zobrazujú.

# 5. Implementácia

Tvorbu aplikácie sme si rozdelili na niekoľko myšlienkových častí:

- preprocessing<sup>1</sup> dát a vytvorenie základného dátového zdroja
- logika prehľadávania dát a tvorby krížovky
- vytvorenie grafického prostredia

V predchádzajúcej kapitole sme si predstavili databázu WordNet, ktorá slúži ako hlavný dátový zdroj pre našu aplikáciu. Samotnej tvorbe krížoviek predchádza niekoľko krokov. Prvým je spracovanie databáz, kde odpovedajúce záznamy slovenského a anglického variantu spojíme v jeden a výslednú dátovú štruktúru uložíme pomocou serializácie. Po zadaní témy prechádzame dátovú štruktúru a prehľadávame slová postupovaním skrz hyponymá a hyperonymá. Vzťah významovej nadradenosti a podradenosti nám pomôže vytvoriť slovník slov súvisiacich so zadanou témou. Z tohto slovníka potom náhodne vyberieme slová tvoriace krížovku. Posledným krokom je vloženie získaných údajov do grafického prostredia. V tejto kapitole si detailne popíšeme jednotlivé kroky programu, algoritmus prehľadávania dát a tvorenia krížoviek, ako aj bližšie informácie o použitých technických prostriedkoch a ťažkostiach pri implementácii.

## 5.1 Serializácia a deserializácia

Serializácia je proces premeny objektu na postupnosť bajtov<sup>2</sup>, ktorá sa ďalej ukladá do pamäte, databázy alebo súboru. Hlavným cieľom serializácie je zachovať pôvodný stav objektu. Opačný proces sa nazýva deserializácia, a ide o znovuvytvorenie objektu v pôvodnom stave. V našej aplikácii využívame serializáciu na uloženie predom spracovaných dát do súboru vo formáte JSON (obrázok 5.1). Načítanie dát pomocou deserializácie z tohto súboru významne urýchlilo chod aplikácie.

### 5.1.1 Formát JSON

JSON je verejný štandard formátu súborov. Používa sa na ukladanie a prenos dátových objektov. Tieto objekty môžu pozostávať z párov atribút<sup>3</sup>-hodnota, alebo dátových polí a rôznych kolekcii. (Wikipedia, 2021a) Ide o často používaný formát dát s rozsiahlou škálou využití.

Proces serializácie a deserializácie dát vo formáte JSON sa zvyčajne vykonáva pomocou tried patriacich do priestoru názvov `System.Text.Json`. Balíček JSON bol však v tomto priestore názvov označený ako zastaralý a neodporúča sa jeho používanie. Po preskúmaní dostupných variantov sme zvolili použitie `Json.NET` od spoločnosti `Newtonsoft`. (Newton-King, 2021) `Newtonsoft.Json` je rovnako

---

<sup>1</sup>spracovanie dát pred ich použitím, môže zahŕňať redukciu, úpravu, selekciu požadovaných častí

<sup>2</sup>základná jednotka digitálnych informácií, 1 bajt = 8 bitov

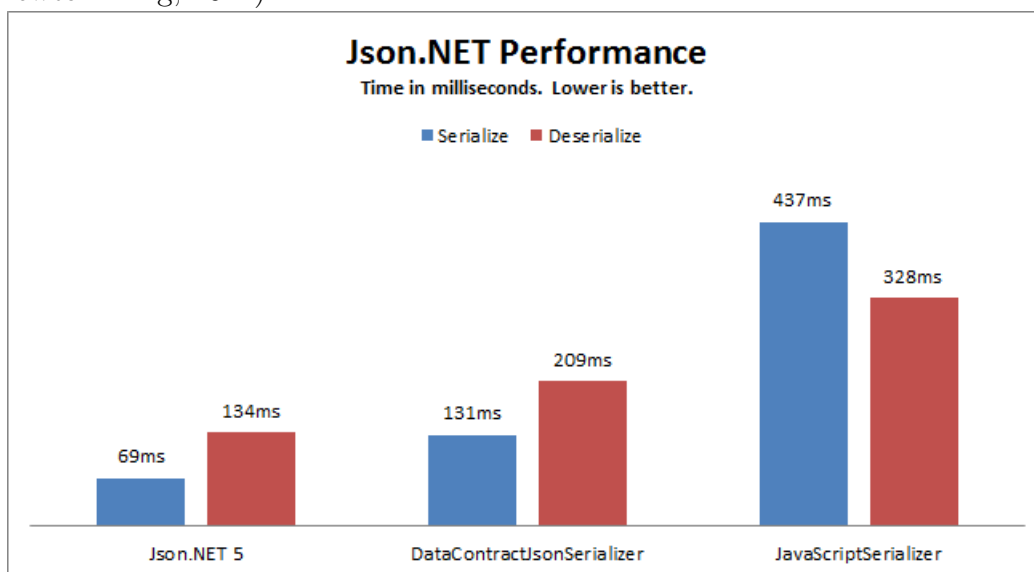
<sup>3</sup>vlastnosť, črta

priestor názvov poskytujúci triedy, ktoré sa používajú na implementáciu základných funkcií štruktúry JSON. `Json.NET` je v súčasnosti populárny a niekoľkonásobne výkonnejší ako konkurenčné varianty (obrázok 5.2).

Obr. 5.1: Príklad synsetu z databázy WordNet uloženého vo formáte JSON:

```
{ "offset": "00001740",  
  "type": "n",  
  "word_EN": "entity",  
  "word_SK": "entita",  
  "hyponyms": { "00001930": "n", "00002137": "n", "04431553": "n" },  
  "hypernyms": {},  
  "gloss": "that which is perceived or known or inferred to have its own distinct  
existence (living or nonliving)" }
```

Obr. 5.2: Vizualizácia porovnania výkonnosti `Json.NET` a konkurenčných variant: (Newton-King, 2021)



## 5.2 Preprocessing

Za dátový zdroj považujeme vo všeobecnosti akúkoľvek databázu, ktorá nám poskytuje informácie. V prípade aplikácií, ktoré pracujú s jazykovými dátami je potrebné, aby bola databáza dostatočne veľká, ľahko použiteľná, ale aj rozumne náročná na pamäť a čas spracovania. S ohľadom na uvedené vlastnosti sme si za hlavný zdroj dát zvolili databázu WordNet.

V aplikácii využívame dve jazykové varianty databázy WordNet, anglickú a slovenskú. Princeton WordNet obsahuje viac ako 177 000 synsetov, zatiaľ čo slovenský WordNet obsahuje len 28 800 synsetov. Keďže chceme získať čo najrozsiahljšiu databázu slov využívame slovenský WordNet ako doplnok k Princeton WordNetu. Obe databázy obsahujú množstvo informácií, ktoré sú pre chod aplikácie nadbytočné. Preto databázy najprv spracujeme a extrahujeme užitočné informácie. Slovenské synsety potom mapujeme na synsety anglického WordNetu, čím obe databázy spojíme a vytvoríme dvojjazyčnú databázu slov, ktorá slúži ako konečný zdroj dát pre výslednú aplikáciu.

Z dostupných databázových súborov Princeton WordNetu používame štyri súbory pomenované `data.pos`, kde `pos` reprezentuje slovné druhy (podstatné mená, prídavné mená, príslovky, slovesá). Tieto súbory sú postačujúce pre obsiahnutie celej databázy a získanie súboru vyše 117 000 slov vrátane ich definícií. Naproti tomu je slovenský WordNet obsiahnutý v jedinom textovom súbore.

Pre načítanie oboch databáz WordNetu do pamäti a ich spracovanie sme si vytvorili samostatný program. Prostredníctvom nami napísaných pomocných programov na prehľadávanie databázových súborov sme definovali dva regulárne výrazy (obrázok 5.3), anglický a slovenský, ktoré odpovedajú štruktúre riadkov v spracovávaných súboroch. Viac informácií o štruktúre riadkov uvádzame v sekciiach *Štruktúra* 4.1.1 a 4.2.1. Podrobný popis jednotlivých častí regulárnych výrazov vzhľadom k štruktúre riadkov uvádzame v sekcii 6.1.3 *Program* a konkrétne príklady mapovania výrazov na riadky WordNetu môžeme vidieť na obrázkoch 6.3 a 6.4.

Obr. 5.3: Regulárne výrazy používané v programe:

Regulárny výraz pre anglický WordNet

```
@"(\\d{8})\\s\\d{2}\\s(\\w)\\s[0-9a-f]{2}\\s((()-_'.\\/\\w\\d+))\\s[0-9a-f]{1}\\s((()-_'.\\/\\w\\d)*\\s[0-9a-f]{1}\\s)*\\d{3}\\s((\\^\\d\\s)+\\s\\d{8}-\\s\\w\\s[0-9a-f]{4}\\s)*\\[\\]^*\\|\\s(.+)"
```

Regulárny výraz pre slovenský WordNet

```
@"\\d+\\t(\\w)\\t(\\[\\];\\t+);\\t.+"
```

Ako prvý krok načítame a spracujeme slovenský Wordnet. Keďže riadky tejto databázy sú zložené zo slovenského aj anglického záznamu, využívame počas parsovania<sup>4</sup> oba definované regulárne výrazy. Postupne čítame riadky slovenského WordNetu a delíme ich pomocou špeciálneho znaku `record separator` na slovenskú a anglickú časť. Zo slovenskej časti si uložíme slovenský preklad slova, z anglickej časti zasa anglický termín a glosu. Tieto údaje sú nevyhnutné pre úspešné spojenie záznamov slovenského a Princeton WordNetu. Zvyšné položky nevyužijeme.

Po kompletom spracovaní slovenského WordNetu, postupne čítame riadky databázových súborov Princeton WordNetu. Z načítaných riadkov extrahujeme

<sup>4</sup>proces analýzy textu a určenia jeho gramatickej štruktúry

pomocou regulárneho výrazu ofset, slovný druh, anglický termín, glosu a ukazovateľa na ďalšie záznamy. Všetky tieto prvky uložíme ako atribúty jediného záznamu. Ofset a ukazovateľa využijeme pri prehľadávaní dát počas tvorby zoznamu tematických slov. Pri prehľadávaní budeme využívať dva druhy ukazovateľov. Hyponymá, označené symbolmi  $\sim$  a  $\sim_i$ , a hyperonymá označené symbolmi  $\@$  a  $\@_i$ . Dostupné ukazovateľa preto potrebujeme roztriediť a uložiť spomínané typy. Postupne prechádzame reťazec reprezentujúci množinu ukazovateľov. Akonáhle narazíme na niektorý zo symbolov  $\@$  a  $\sim$ , uložíme offset cieľového ukazovateľa a slovný druh, na ktorý ukazuje, do dátovej štruktúry pre hyponymá, respektíve hyperonymá. Následne sa pokúsime podľa anglického termínu a glosy nájsť odpovedajúci slovenský záznam a pridať tak k už získaným atribútom záznamu slovenský preklad. Po spracovaní súborov jednotlivých slovných druhov uložíme všetky získané informácie do výslednej dátovej štruktúry. Spojenie anglického a slovenského WordNetu je potom reprezentované ako množina záznamov so spomínanými atribútmi.

Popísaný proces načítania, spracovania a párovania všetkých dát zo slovenského a anglického WordNetu a ich uloženie do výslednej dátovej štruktúry trvá viac ako 56 sekúnd. Ak by sme pri každom spustení aplikácie získavali dáta týmto spôsobom, bolo by to príliš zdĺhavé a spokojnosť užívateľov by sa mohla znížiť. Jedným z našich cieľov bolo získať komplexný súbor dát, ktorého načítanie zaberie zanedbateľný čas. Preto sme použili proces serializácie. Výslednú dátovú štruktúru obsahujúcu všetky údaje sme serializovali a uložili do súboru formátu JSON. Tento súbor slúži ako konečný zdroj dát pre aplikáciu na generovanie krížoviek, ktorá získava dáta deserializáciou. Čas načítania dát sme pomocou deserializácie výrazne znížili, a to zo spomínaných 56 sekúnd na jednu sekundu. Pozorujeme 98% zrýchlenie.

### 5.2.1 Techniky mapovania slovenských a anglických synsetov

Počas spájania slovenského a anglického WordNetu sme narazili na niekoľko komplikácií. Jednoznačným identifikátorom záznamu v anglickom WordNete je jeho ofset. Každý ofset sa v súbore reprezentujúcom jeden slovný druh vyskytuje práve jedenkrát. Rovnaký ofset sa môže vyskytovať len medzi slovami rôznych slovných druhov. Príkladom je nasledujúca štvorica záznamov s rovnakým ofsetom ale rôznymi slovnými druhmi.

<code>data.noun:</code>	<code>00001740 03 n 01 entity</code>
<code>data.verb:</code>	<code>00001740 29 v 04 breathe</code>
<code>data.adj:</code>	<code>00001740 00 a 01 able</code>
<code>data.adv:</code>	<code>00001740 02 r 02 a_cappella</code>

Na základe tejto skutočnosti sme sa rozhodli záznamy zo slovenského aj anglického WordNetu načítať do samostatných kolekcí reprezentujúcich jednotlivé slovné druhy. Ako prvý sme spracovali slovenský WordNet. Výhodou je jeho štruktúra, pretože pozostáva zo slovenskej a anglickej časti oddelenej špecifickým separátorom. Navyše je anglická časť svojou štruktúrou identická záznamom Princeton WordNetu.

Postupne sme načítali riadky slovenského WordNetu a rozdelili ich pomocou spomínaného oddeľovača. Zo slovenskej časti sme extrahovali slovenský preklad slova. Z anglickej časti sme vyňali výraz, pomocou ktorého sme chceli dané slovenské slovo spojiť s anglickým záznamom. Tieto dva prvky sme si ukladali do kolekcie `dictionary`<sup>5</sup> s výrazom pre párovanie ako kľúč a slovenským prekladom ako hodnota. Po načítaní slovenského WordNetu a uložení potrebných dát do kolekcií sme postupne načítali súbory anglického WordNetu a pokúšali sa na jednotlivé záznamy napárovať slovenské preklady. V otázke „Čo má obsahovať výraz?“ sme vyskúšali niekoľko prístupov.

Prvou taktikou bolo párovať slovenské a anglické záznamy na základe ofsetu. Tu sme narazili na problém zastaralých ofsetov v slovenskom WordNete, ktoré ilustrujú nasledujúce príklady. Prvý príklad ukazuje rozdiely v ofsetoch pri rovnakých slovách. Druhý príklad ukazuje rovnaké ofsety pri rozdielnych slovách, ktoré dokonca nepatria k rovnakému slovnému druhu.

```
Princeton: 03355726 fishery
SK:      03350880 fishery
```

```
Princeton: 00441881 n cross-country_skiing
SK:      00441881 v diminish
```

Druhou taktikou bolo párovať slová na základe glosy. Počas testovania sme však zistili, že existuje niekoľko rozličných slov, ktoré majú rovnakú glosu. Išlo najmä o mestá alebo prírodné úkazy v rovnakej oblasti, prípadne rôzne jazyky, ktorými sa hovorí na rovnakom území.

```
Princeton: 08761524 Leon | a city in central Mexico
           08763240 San_Luis_Potosi | a city in central Mexico
SK: 08743691 Leon | a city in central Mexico
     08745407 San_Luis_Potosi | a city in central Mexico
```

Tretou uvažovanou taktikou bolo párovanie na základe anglického výrazu. Tu sa okamžite vynára problém synonymie, ktorý týmto spôsobom nebol vyriešený. Ako príklad uvádzame anglické slovo *date* a príslušné významy. V prípade párovania podľa anglického slova by sa slovo *date* vo význame *datľa* preložilo nesprávne ako *dátum*.

```
Princeton:
date | the specified day of the month
date | the particular day, month, or year that an event occurred
date | a particular day specified as the time something happens
date | sweet edible fruit of the date palm with a long woody seed
date | a meeting arranged in advance
date | a participant in a date
date | the present
```

```
SK:
dátum | the specified day of the month
```

<sup>5</sup>dátová kolekcia pozostávajúca z párov kľúč a hodnota

dátum   the particular day, month, or year that an event occurred
dátum   a particular day specified as the time something happens

Po týchto neúspešných krokoch sme usúdili, že záznamy je potrebné párovať pomocou viac ako jedného výrazu, medzi ktoré nepatrí ofset. Zvolili sme si glosu a anglický výraz. Slovo a definícia jeho významu jednoznačne určujú záznam.

Vo výslednej zvolenej taktike párovania sa taktiež prejavuje zastaranosť dát slovenského WordNetu. Keďže používame najnovšiu verziu anglického WordNetu, niektoré glosy sú zmenené, a teda párovanie určitých záznamov neuspeje. Konkrétne ide o 192 prípadov. Pri bližšom skúmaní sme zistili, že mnohé z glos v slovenskom WordNete, ktoré neuspajú pri párovaní, majú vo svojom reťazci nadbytočný alebo nesprávny znak (medzera, čiarka namiesto bodkočiarky). Takže aj keď sú významovo identické a správne, záznamy sa nespárujú.

Mohli by sme použiť rovnakú verziu anglického WordNetu, podľa ktorej bol tvorený aj slovenský WordNet. Tým by sme však prišli o viac ako 140 000 synsetov, navyše sa používanie starých verzií Princeton WordNetu neodporúča kvôli ich chybovosti.

Princeton: beauteous   (poetic) beautiful, especially to the sight
SK: beauteous   (poetic )beautiful, especially to the sight
Princeton: recite   render verbally; "recite a poem"; "retell a story"
SK: recite   render verbally, "recite a poem"; "retell a story"

Okrem zastaraných dát v súvislosti s mapovaním na Princeton Wordnet sa v databáze slovenského WordNetu nachádza viac ako 3500 identických, opakujúcich sa záznamov. Pre 213 záznamov nie je uvedený slovenský preklad (obrázok 5.4). Vo väčšine týchto prípadov neexistuje jednoduchý slovenský ekvivalent slova, preto daný záznam obsahuje len zdĺhavé vysvetlenie významu. V konečnom dôsledku je viac ako 13% záznamov nepoužiteľných.

Všeobecný popis programu pre spracovanie WordNetu uvádzame v sekcii 5.2 *Preprocessing*.

Obr. 5.4: Príklady záznamov slovenského WordNetu, kde nie je uvedený slovenský ekvivalent anglického termínu:

- [14028 n ? RS 04057846 06 n 01 reactor 1 002 @ 03269401 n 0000 ~ 03065708 n 0000 | an electrical device used to introduce reactance into a circuit](#)
- [14102 n - anatomická štruktúra oddeľujúca jednotlivé časti v organizme RS 05222790 08 n 01 partition 0 003 @ 05220461 n 0000 ;c 06057539 n 0000 ~ 05222940 n 0000 | \(anatomy\) a structure that separates areas in an organism](#)

## 5.2.2 WordNet API

Aplikáciu pre spracovanie WordNetu sme tvorili v jazyku C#. Aby sme proces spracovania databázy zjednodušili, hľadali sme verejne prístupné API<sup>6</sup> pre použí-

<sup>6</sup>application programming interface - rozhranie pre programovanie aplikácií



vanie WordNetu v tomto jazyku. Oficiálna stránka Princeton WordNetu uvádza v sekcii súvisiacich projektov niekoľko API, vytvorených a ponúkaných používateľmi databázy. (Princeton University, 2021) Napriek tejto skutočnosti, je väčšina z uverejnených užívateľských projektov nedostupná, odstránená alebo zastaraná. Preto sme sa rozhodli vytvoriť vlastné funkcie pre spracovanie databázových súborov, ktorých hlavnou zložkou je použitie regulárnych výrazov.

## 5.3 Zostavenie tematického slovníka

Užívateľ si po spustení aplikácie zvolí tému. Našou úlohou je nájsť všetky slová, ktoré sa danej témy týkajú. Pri skúmaní vlastností a štruktúry WordNetu sa nám ako najvhodnejšie javilo prehľadávanie dát postupovaním po hyperonymách a hyponymách. Hyperonymia je vzťah významovej nadradenosti, kde hyperonymom k určitému slovu rozumieme zovšeobecnenie. Naopak hyponymia je vzťah významovej podradenosti. Hyponymom k určitému slovu rozumieme konkretizáciu. Hierarchickú štruktúru, ktorá zobrazuje vzťah hyponymie a hyperonymie môžeme pozorovať na obrázku 4.2.

Navrhovaný algoritmus pozostáva z troch hlavných krokov.

1. Najprv nájdeme všetky slová, ktoré sa priamo týkajú zvolenej témy. To znamená, že reťazec reprezentujúci tému je prefixom týchto slov, alebo je im rovný. Získame prvotnú množinu slov. Všetky slová z prvotnej množiny pridáme do výsledného slovníka.
2. Pre každé slovo z prvotnej množiny nájdeme všetky synsety predstavujúce hyponymá, a pridáme ich do množiny hyponým. Rovnako tak pre každé slovo z prvotnej množiny nájdeme všetky synsety predstavujúce hyperonymá, a pridáme ich do množiny hyperoným.
3. Všetky slová reprezentované synsetmi zo získaných množín pridáme do výsledného slovníka. V prehľadávaní pokračujeme tzv. „rovnakým smerom“. Pre každý záznam z množiny hyponým nájdeme všetky synsety predstavujúce hyponymá, a pridáme ich do novej množiny hyponým. Pre každý záznam z množiny hyperoným nájdeme všetky synsety predstavujúce hyperonymá, a pridáme ich do novej množiny hyperoným. Znovu opakujeme krok 3.

Otázkou zostávalo, ako dlho je treba prehľadávať hyponymá a hyperonymá, aby bol výsledok optimálny. Odpoveď nájdeme v článku o metrikách WordNetu napísanom Ann Devitt a Carlom Vogelom (Devitt a Vogel, 2004). Označme si synsety, ktoré majú ukazovatele na podradené pojmy ako vnútorné vrcholy, a ostatné ako listy. Potom môžeme databázu WordNet chápať ako acyklický orientovaný graf. Aby sme preskúmali všetky slová súvisiace s určitým synsetom, chceli by sme prechádzať grafom až k listom, prípadne aj koreňom. Výsledky spomínanej práce ukazujú, že medzi ľubovoľným vnútorným vrcholom a listom sa nachádza v priemere 2 až 5 vrcholov. Z toho vyplýva, že optimálny počet iterácií prehľadávania hyperoným a hyponým pre dosiahnutie listových synsetov je 5.

Ďalším zaujímavým výsledkom práce je zoznam najčastejších koreňov. Koreňom rozumieme synset, ktorý nie je podradený žiadnemu inému synsetu, a teda

vyjadruje najvšeobecnejší popis určitej kategórie. Devitt v práci uvádza 9 najvšeobecnejších koreňov (zoradené zostupne podľa množstva podradených synsetov): *entita, abstrakcia, skupina, činnosť, psychologická vlastnosť, stav, fenomén, udalosť, vlastníctvo*. Toto zistenie sa odzrkadľuje aj na výsledkoch našej práce. Vo vytvorenom slovníku slov súvisiacich s témou vždy nachádzame niektorý alebo viaceré z týchto koreňov.

*Použitie výsledky práce Devitt a Vogel (2004) sa vzťahujú k WordNetu verzie 1.1.7, kedy obsahoval 74488 synsetov pre podstatné mená.*

### 5.3.1 Vymedzenie slov týkajúcich sa témy

Proces hľadania slov týkajúcich sa témy je kľúčový pre správne fungovanie aplikácie. Po zvolení témy užívateľom prehľadávame reťazce anglických slov. Najprv sme hľadali slová, ktoré obsahujú celý reťazec predstavujúci tému, použitím metódy `contains`. Testovaním sme narazili na problematiku podreťazcov, ktoré sa vyskytovali v slovách nesúvisiacich so zadanou témou. Napríklad pri vstupnom slove *sport* sme vo výslednom zozname vhodných slov našli dopravné prostriedky. Príčinou bolo obsiahnutie slova *sport* vo výraze *transport*, teda doprava. Zmenili sme spôsob hľadania, a to použitím metódy `startsWith`. Teda slovo považujeme za patriace k danej téme, ak sa začína reťazcom témy. Tým sme vyššie spomenutý problém vyriešili, vynára sa však problém vhodnosti zadanej témy. Táto metóda funguje dobre ak si užívateľ vyberie zo zoznamu dostupných tém, ktoré boli vopred otestované. Ak však užívateľ zadá vlastnú tému, toto slovo nemusí byť v základnom tvare a je potrebné ošetriť možné modifikácie reťazca, ako množné číslo alebo skloňovanie.

Pri definovaní témy užívateľom preto najprv skontrolujeme, či sa dané slovo vyskytuje v databáze WordNet. Ak áno, vytvoríme slovník tematických slov a pokúsime sa z neho vytvoriť krížovku. Naopak ak sa užívateľom zadané slovo nevyskytuje vo WordNete, upozorníme ho na túto skutočnosť a požiadame ho o kontrolu tvaru zadaného slova, prípadne o výber novej témy. V oboch prípadoch kontrolujeme, či sme schopní vytvoriť krížovku z dostupného slovníka tematických slov.

### 5.3.2 Zvažované varianty algoritmu

Pôvodný návrh algoritmu pozostával zo získania prvotnej množiny záznamov obsahujúcich tému. Následne sme pre každý záznam z prvotnej množiny našli jeho hyponymá aj hyperonymá. Tým vznikla nová množina záznamov a prehľadávanie pokračovalo rovnakým spôsobom, teda pre každý záznam obsiahnutý v množine sme hľadali všetky jeho hyperonymá aj hyponymá.

Tento spôsob prehľadávania záznamov sa osvedčil pri zadaní príliš konkrétnej témy. Ak užívateľ zadal slovo, ktorého záznam obsahoval veľmi málo ukazovateľov alebo neobsahoval žiadne ukazovatele na hyponymá alebo hyperonymá, program zvládol vytvoriť dostatočne veľký slovník slov pre generovanie krížovky. Avšak ak téma, ktorú si užívateľ zvolil alebo zadal bola dostatočne všeobecná, algoritmus vrátil príliš rozsiahly slovník, ktorý obsahoval väčšinu slov netýkajúcich sa danej témy.

Rozhodli sme sa, že je lepšie v niektorých prípadoch nevytvoriť krížovku, ako

by sme v drvivej väčšine prípadov vytvárali krížovky s tematicky nesúvisiacimi slovami. Preto sme algoritmus pozmenili na prehľadávanie len priamo po hyponymách a po hyperonymách, ako je popísané v sekcii 5.3 *Zostavenie tematického slovníka*.

## 5.4 Tvorba krížoviek

V tejto práci sa venujeme generovaniu špeciálneho typu krížoviek - hrebeňovkám. Hrebeňovka je typ krížovky, v ktorej niekoľko stĺpcov umiestnených bezprostredne vedľa seba vytvorí v určenom riadku tajničku. Program najprv vytvorí slovník slov týkajúcich sa zvolenej témy. Pre vytvorenie krížovky si zo slovníka vyberieme jedno slovo, ktoré bude slúžiť ako tajnička. Výber opakujeme, ak je slovo príliš krátke alebo obsahuje číslicu. Po úspešnom výbere slova do tajničky iterujeme skrz znaky tohto slova a zo slovníka vyberáme náhodné slová, ktoré obsahujú daný znak. Výber opakujeme, ak slovo obsahuje číslicu alebo neobsahuje želaný znak. Slovo so spoločným znakom hľadáme maximálne dvadsaťkrát. Ak neuspějeme viac ako dvadsaťkrát, ohlásime neúspech a vraciame sa k výberu slova do tajničky. Pokiaľ aj tu presiahneme počtom dvadsať opakovaní, vyhlásime, že nie je možné vytvoriť krížovku a požiadame užívateľa o výber novej témy, prípadne si užívateľ môže zvoliť generovanie náhodnej krížovky. Týmto spôsobom jednoducho ošetríme situácie, kedy je slovník tematických slov príliš krátky alebo nie je možné z tohto slovníka vytvoriť krížovku. V praxi táto situácia nastáva ojedinele. Ak si užívateľ vyberie z nami definovaných tém, krížovka sa zakaždým úspešne vygeneruje. Ak užívateľ definuje vlastnú, príliš konkrétnu tému, môže sa stať, že z daného slovníka tematických slov nebude možné krížovku vytvoriť.

## 5.5 Grafické prostredie

Pre vytvorenie prívetivého grafického prostredia používame nástroje Windows formulárov. Nasledujúci zoznam predstavuje nástroje kľúčové pre chod našej aplikácie.

- **DataGridView** - prispôsobiteľná mriežka, ktorá slúži na zobrazovanie dát alebo prijíma vpisované údaje od užívateľa, jednotlivé bunky mriežky je možné formátovať a upravovať podľa toho či je bunka aktívna, v našej aplikácii mriežka reprezentuje schému krížovky a používame ju aj pre reprezentáciu legendy
- **SplitContainer** - ovládací prvok, ktorý pomocou pohyblivej lišty rozdeľuje zobrazovacie okno na dva panely vo vertikálnom alebo horizontálnom smere, lišta môže byť pohyblivá alebo znehybnená, v našej aplikácii využívame tento prvok na rozdelenie okna na časť pre zobrazenie schémy krížovky a časť pre zobrazenie legendy
- **ComboBox** - padacie menu, ktoré umožňuje užívateľovi vybrať si niektorý z ponúkaných prvkov, a pomocou ktorého reprezentujeme množinu ponúkaných tém

- `TextBox` - okno, ktoré umožňuje užívateľovi zadať text do aplikácie, v našej aplikácii slúži pre definovanie vlastnej témy užívateľom
- `MessageBox` - vyskakovacie okno (známe ako dialógový box), ktoré zobrazuje správu pre užívateľa a blokuje ďalšie akcie v aplikácii, pokým nie je zatvorené, môže obsahovať tlačidlá a text alebo obrázky, vyskakovacie okná využívame pri hlásení chýb užívateľovi a tiež pri obsluhu tlačidiel

Okrem vyššie spomenutých, menej známych nástrojov využívame aj tlačidlá (`Button`), tabuľky (`TableLayoutPanel`) a ovládacie prvky pre obrázky (`PictureBox`) a textové označenia (`Label`). Všetky nástroje, s ktorými užívateľ interaguje disponujú metódami na ich obsluhu. Bližší popis programového prevedenia a implementácie metód pre obsluhu jednotlivých nástrojov predstavíme v kapitole 6 *Vývojová dokumentácia*.

V tejto kapitole sme si detailne popísali základné myšlienky programu. Predstavili sme jednotlivé kroky programu, predspracovanie dát, prehľadávanie dát a spôsob tvorby krížoviek. Odôvodnili sme výber algoritmu prehľadávania dát a popísali sme technické prostriedky použité počas implementácie. V krátkosti sme popísali grafické nástroje kľúčové pre chod aplikácie. V ďalšej kapitole (6 *Vývojová dokumentácia*) sa budeme venovať technickému popisu programového prevedenia.

## 6. Vývojová dokumentácia

Aplikáciu sme písali v jazyku `C#` v prostredí `Visual Studio`. Celkovo sme vytvorili dva programy. `WordSourceMaker` slúži pre spracovanie slovenskej a anglickej databázy `WordNet`. `ThematicCrosswords` je výsledná aplikácia pre generovanie tematických hrebeňoviek. Oba programy využívajú metódy balíčka `Newtonsoft.Json`. Tento balíček sme nainštalovali prostredníctvom tzv. manažéra balíčkov `NuGet Package Manager`, ktorý ponúka prostredie `Visual Studio`.

V tejto kapitole si detailne predstavíme technické prevedenie týchto programov, jednotlivé metódy, z ktorých pozostávajú, aké súbory používajú a čo je ich výstupom.

Dokumentáciu odporúčame čítať spolu s kódom programov pre názornejšie ukážky metód a používaných vlastností komponent.

### 6.1 Word Source Maker

Program `WordSourceMaker` slúži pre spracovanie slovenskej a anglickej databázy `WordNet`. Načíta a zjednotí záznamy oboch databáz a výsledok zjednotenia uloží pomocou serializácie do súboru `WordNet` vo formáte `JSON`.

#### 6.1.1 Potrebné súbory

Program čerpá údaje zo štyroch databázových súborov `Princeton WordNetu`: `data.adj`, `data.adv`, `data.noun`, `data.verb`. Tieto súbory sú voľne prístupné a stiahnuteľné v komprimovanom priečinku z oficiálnej stránky `Princeton WordNetu` (`Princeton University`, 2021). Ďalej program čerpá údaje zo slovenského `WordNetu`, ktorý je reprezentovaný jediným textovým súborom pod názvom `sk-wn-2013-01-23`, a ktorý je voľne dostupný na stránke `Slovenského Národného Korpusu` (`Slovenský Národný Korpus`, 2021).

#### 6.1.2 Trieda `Record`

Program pozostáva zo základnej triedy `Program` a triedy `Record`. Instancia<sup>1</sup> triedy `Record` predstavuje jeden záznam databázy `WordNet`. Obsahuje atribúty:

- `offset` = ofset typu `string` (napriek tomu, že ofset je číslo nemôžeme použiť číselný typ `int`, pretože ofset môže začínať nulou)
- `type` = slovný druh typu `string`
- `word_EN` = slovo v anglickom jazyku typu `string`
- `word_SK` = slovo v slovenskom jazyku typu `string`
- `hyponyms` = zoznam ukazovateľov, hyponým, typu `Dictionary<string, char>`, kde ukladáme ofset a slovný druh

---

<sup>1</sup>v objektovo orientovanom programovaní je instancia objekt patriaci do danej triedy

- `hypernoms` = zoznam ukazovateľov, hyperoným, typu `Dictionary<string, char>`, kde ukladáme ofset a slovný druh
- `gloss` = glosa / definícia slova typu `string`

Trieda `Record` obsahuje konštruktor, ktorý vytvára nový záznam a prijíma jednotlivé atribúty záznamu ako parametre. Parametrom pre hyponymá a hyperonymá je reťazec predstavujúci všetky ukazovatele záznamu. Tento reťazec je potrebné spracovať. Konštruktor preto zavola metódu `extractSynsets`, raz pre extrahovanie ukazovateľov na hyponymá a raz pre hyperonymá.

Metóda `extractSynsets` prijíma ako parametre reťazec reprezentujúci množinu ukazovateľov a znak (`char`), ktorý reprezentuje typ ukazovateľov, ktoré chceme extrahovať. Znakom môže byť `~` pre hyponymá alebo `@` pre hyperonymá. Metóda prechádza reťazec znak po znaku. Akonáhle narazí na znak zhodný so želaným typom ukazovateľa, uloží ofset ukazovateľa a slovný druh, na ktorý ukazuje do výstupnej dátovej štruktúry typu slovník (`Dictionary<string, char>`). Konštruktor následne uloží všetky atribúty danej instance záznamu.

### 6.1.3 Trieda Program

Trieda `Program` obsahuje metódy pre spracovanie databázových súborov, ktoré využívajú regulárne výrazy. V kapitolách 4.1.1 a 4.2.1 *Štruktúra* sme popísali štruktúru riadkov anglického a slovenského WordNetu. Na základe tohto popisu sme definovali dva regulárne výrazy, ktoré dokonale zodpovedajú štruktúre riadkov databáz, a pomocou ktorých môžeme extrahovať komponenty riadkov potrebné pre našu aplikáciu. V nasledujúcich odsekoch detailne vysvetlíme používané regulárne výrazy.

#### Regulárny výraz pre anglický jazyk

Obr. 6.1: Regulárny výraz pre anglický WordNet

```
@"(\\d{8})\\s\\d{2}\\s(\\w)\\s[0-9a-f]{2}\\s((()-_'.\\./w\\d+)\\s[0-9a-f]{1}\\s((()-_'.\\./w\\d)*\\s[0-9a-f]{1}\\s)*\\d{3}\\s((\\^\\d\\s|+\\s\\d{8})\\s\\w\\s[0-9a-f]{4}\\s)*\\[\\]*\\s(\\s.+)"
```

Anglický regulárny výraz pozostáva z nasledujúcich častí, vždy oddelených medzerou reprezentovanou znakom `\s`: (*poznámka: číslo v zložených zátvorkách vyjadruje počet opakovaní daného znaku, znaky alebo intervaly znakov v hranatých zátvorkách vyjadrujú platné/prijímané znaky, + vyjadruje, že daný znak sa musí v reťazci vyskytovať aspoň raz, \* značí ľubovoľný počet opakovaní, teda nula alebo viac*)

- ofset - osemciferné desiatkové celé číslo `\\d{8}`
- lexikografický súbor - dvojciferné desiatkové číslo `\\d{2}`
- slovný druh - jeden znak `\\w`
- počet slov - dvojciferné hexadecimálne číslo `[0-9a-f]{2}`

- slovo - reťazec  $([()\_'\./\w\d]+)$ 
  - slovo môže obsahovať gultaté zátvorky (), pomlčky (-), podčiarkovník (\_), apostrof ('), bodky (\.), písmená (\w) a číslice (\d)
  - v riadku sa musí vyskytovať aspoň jedno slovo, ale môže ich byť viac (vyjadrené pomocou reťazca  $([()\_'\./\w\d]^*\s[0-9a-f]{1}\s)^*$ )
- jednociferné hexadecimálne číslo  $[0-9a-f]{1}$
- počet ukazovateľov - trojmiestne desiatkové celé číslo  $\d{3}$
- ukazovatele  $([\d\s]+\s\d{8}\s\w\s[0-9a-f]{4}\s)^*$ , ktoré sú tvaru
  - symbol ukazovateľa  $[\d\s]^+$  = čokoľvek okrem písmen a číslíc aspoň raz
  - ofset - osemciferné desiatkové celé číslo  $\d{8}$
  - slovný druh - jeden znak  $\w$
  - štvorbajtové pole pozostávajúce z dvoch dvojčiferných hexadecimálnych čísel  $[0-9a-f]{4}$
- generické rámce, nula alebo viackrát, ošetrujeme pomocou výrazu  $[^]*$  = čokoľvek okrem zvislej čiary (|) ľubovoľne veľakrát
- glosa - reťazec slov nasledujúci po zvislej čiare (|)  $|\s(.+)$  = ľubovoľné znaky po zvislej čiare

## Regulárny výraz pre slovenský jazyk

Riadky slovenského WordNetu pozostávajú zo slovenského a anglického záznamu. Anglický záznam sa štruktúrou dokonale zhoduje so záznamami anglického WordNetu, preto pre prácu s ním využívame rovnaký regulárny výraz, ktorý sme definovali vyššie. Pre prácu so slovenskou časťou záznamu používame nasledovný regulárny výraz.

Obr. 6.2: Regulárny výraz pre slovenský WordNet

`@”\d+\t(\w)\t([\d\s]+\s\d{8}\s\w\s[0-9a-f]{4}\s)^*\s|\s(.+)`

Všetky časti slovenského záznamu sú vzájomne oddelené tabulátorom ( $\t$ ). Záznam pozostáva z nasledujúcich častí:

- číslo synsetu - desiatkové celé číslo  $\d+$
- slovný druh - jeden znak  $\w$
- slovo - reťazec  $[\d\s]^+$  = vezmeme všetko až po bodkočiarku (prvú variantu slova), alebo tabulátor
- glosa - reťazec (väčšinou prázdny)  $[\d\s]^+\s|\s(.+)$ , teda čokoľvek po poslednom tabulátore alebo oddeľujúcej bodkočiarku až do konca riadka

Na obrázkoch 6.3 a 6.4 môžeme vidieť reálne záznamy z databáz WordNet a ich rozdelenie na jednotlivé časti pomocou regulárnych výrazov.

	noun	adj	adv	verb
<code>\d{8}</code>	05643469	00188413	00016322	00004492
<code>\d{2}</code>	09	00	02	29
<code>\w</code>	n	s	r	v
<code>[0-9a-f]{2}</code>	01	02	01	01
<code>(( _ '\w\d)+</code>	fortification	astir(p)	prosperously	hold
<code>[0-9a-f]{1}</code>	0	0	1	4
<code>(( _ '\w\d)*\s[0-9a-f]{1}\s)*</code>		up(p) 0		
<code>\d{3}</code>	003	001	001	001
<code>((^\d\s]+\s\d{8}\s\w\s[0-9a-f]{4}\s)*</code>	@ 05646832 n 0000 + 01089608 v 0103 + 00221295 v 0103	& 00188104 a 0000	\ 02029807 a 0103	@ 00233707 v 0000
<code>[^]*</code>				01 + 08 00
<code>\\s(.+)</code>	the art or science of strengthening defenses	out of bed; "are they astir yet?"; "up by seven each morning"	in the manner of prosperous people	keep from exhaling or expelling; "hold your breath"

Obr. 6.3: Príklad parsovania riadkov databázových súborov Princeton WordNetu podľa definovaného regulárneho výrazu.

	slovenský záznam
<code>\d+</code>	1999
<code>\w</code>	a
<code>[^;\t]+</code>	abnormálny
<code>[:\t].+</code>	; nadmerný; priveľký RS 01533535 00 s 01 abnormal 0 001 & 01533120 a 0000   much greater than the normal; "abnormal profits"; "abnormal ambition"

Obr. 6.4: Príklad parsovania slovenského záznamu z riadka slovenského WordNetu podľa definovaného regulárneho výrazu.

## Metódy

V triede `Program` sme si definovali tri metódy:

### 1. `loadTranslations`

Metóda si na začiatok vytvorí prázdny zoznam, ktorý bude reprezentovať záznamy slovenského WordNetu. Postupne načíta riadky slovenského WordNetu zo súboru `sk-wn-2013-01-23`. Každý riadok rozdelí v mieste znaku `record separator \U+241E` na anglickú a slovenskú časť záznamu. Zo slovenskej časti extrahuje reťazec reprezentujúci slovenské slovo. Z anglickej časti extrahuje reťazec reprezentujúci anglické slovo a reťazec reprezentujúci glosu. Pre extrakciu používame definované regulárne výrazy a metódu `Regex.Matches`, ktorá dokáže vyňať časti regulárneho výrazu uzavreté v guľatých zátvorkách. Slovenské a anglické slovo môže obsahovať špeciálne znaky ako zátvorky, pomlčky alebo podčiarkovníky. Špeciálne znaky odstránime pomocou metódy `Regex.Replace`. Podobným spôsobom je treba



upraviť glosu. Potrebujeme odstrániť vzorové vety, preto použitím metódy `Regex.Replace` odstránime všetky znaky nachádzajúce sa za úvodzovkami (”), ktorými vzorové vety začínajú. Metóda uloží všetky získané časti riadku ako atribúty nového záznamu a tento záznam pridá do výsledneho zoznamu. Zoznam záznamov, teda štruktúru `List<Record>`, vracia metóda ako svoj výstup.

## 2. loadWordNet

Metóda si na začiatok vytvorí prázdny výstupný zoznam. Definujeme koncovky databázových súborov, z ktorých budeme čerpať, a to `noun`, `adj`, `adv`, `verb`. Pre každý zo súborov vykonáme rovnaký proces načítania a spracovania.

Metóda postupne načíta riadky súboru. Z každého riadku extrahuje reťazce reprezentujúce offset, slovný druh, anglické slovo, ukazovatele a glosu. Pre extrakciu používame definovaný regulárny výraz a metódu `Regex.Matches`, ktorá dokáže vyňať časti regulárneho výrazu uzavreté v guľatých zátvorkách. Rovnako ako v metóde `loadTranslations` odstránime špeciálne znaky z reťazcov reprezentujúcich anglické slovo a glosu. Následne sa metóda pokúsi nájsť slovenský záznam so zhodným anglickým termínom a glosou. K hľadaniu využíva metódu `Find`, ktorá vráti prvok spĺňajúci stanovený predikát. Ak taký záznam neexistuje, slovenské slovo bude prázdny reťazcom. Metóda potom uloží všetky získané časti riadku ako atribúty nového záznamu a tento záznam pridá do výsledneho zoznamu. Zoznam záznamov, teda štruktúru `List<Record>`, vracia metóda ako svoj výstup.

## 3. Serializer

Metóda dostáva ako vstupný parameter zoznam záznamov (`List<Record>`), ktorý chceme serializovať. Inicializuje prvok `serializer`, ktorého metódy vykonávajú proces serializácie a deserializácie a definuje cestu k cieľovému súboru pre uloženie serializovaného objektu. Pomocou metódy `Serialize` uloží daný objekt vo formáte JSON do súboru podľa definovanej cesty.

### Metóda Main

Hlavná metóda pozostáva z troch jednoduchých krokov. Do dátovej štruktúry `slovak` načítame pomocou metódy `loadTranslations` záznamy slovenského WordNetu. Následne pomocou metódy `loadWordNet` načítame záznamy anglického WordNetu a zjednotíme ich so slovenskými záznamami. Výstup metódy uložíme do štruktúry `english`. Nakoniec objekt `english` serializujeme pomocou metódy `Serialize` do výsledného súboru, ktorý predstavuje dvojjazyčnú verziu WordNetu.

```
slovak = loadTranslations();
english = loadWordNet();
Serialize(english);
```

## 6.1.4 Výstup

Výstupom programu `WordSourceMaker` je súbor pod názvom `WordNet` vo formáte `JSON`. Predstavuje dvojjazyčnú verziu databázy `WordNet`. Ide o zjednotenie anglickej a slovenskej verzie tejto databázy, ktorá slúži ako hlavný zdroj dát pre výslednú aplikáciu na generovanie tematických hrebeňoviek. Formát súboru umožňuje jeho jednoduché spracovanie a mimoriadne rýchle načítanie dát.

## 6.2 Thematic Crosswords

`ThematicCrosswords` je výsledná aplikácia pre generovanie tematických hrebeňoviek s jednoduchým grafickým prostredím. Pomocou deserializácie načíta údaje zo súboru `WordNet.json`, vytvoreného programom `WordSourceMaker`. Tento súbor je jediným zdrojom dát pre aplikáciu. Na základe témy zvolenej užívateľom prehľadáva načítané dáta a vytvorí slovník tematických slov. Z tohto slovníka automaticky vygeneruje hrebeňovku, obsahujúcu slová týkajúce sa zvolenej témy.

Ide o takzvanú `Windows Forms` aplikáciu, ktorá využíva vizuálne prvky jazyka `C#` a prostredia `Visual Studio`, pre vytvorenie graficky prívetivého užívateľského prostredia. Program používa rovnakú triedu `Record`, akú sme popísali v sekcii *6.1.2 Trieda Record*. Metóda `extractSynsets` ani konštruktor nie sú potrebné, definovali sme si novú metódu `dictForm`. Metóda vracia reťazec s vybranými atribútmi záznamu oddelenými dvojbodkou (:), ktorý využívame pri ukladaní tematických záznamov, a ktorý zjednodušuje extrakciu jednotlivých atribútov záznamu.

Pri spustení aplikácie sa ako prvá použije hlavná metóda `Form1`. Inicializuje zobrazovacie okno, nastaví jeho parametre a presunie program do stavu `START`. V nasledujúcich sekciách popíšeme jednotlivé stavy programu, metódy pre obsluhu tlačidiel a prehľadávanie databázy, a tvorbu krížoviek.

### 6.2.1 Stavy aplikácie

Aplikácia sa presúva medzi šiestimi stavmi : `START`, `LANG`, `TOPIC`, `PROCESS`, `RULES`, `PUZZLE`. V nasledujúcom zozname detailne popíšeme metódy používané v jednotlivých stavoch aplikácie:

#### 1. `SetState`

Metóda popisuje jednotlivé stavy aplikácie. Podľa stavu zadaného ako parameter metódy určuje, čo sa v danom stave zobrazuje a aké procesy sa vykonávajú. V stave `START` sa do štruktúry `WordNet` pomocou metódy `Deserialize` načíta databáza `WordNet` a volá sa metóda `Introduction`. V stave `LANG` sa inicializujú štruktúry pre slovník tematických slov `TopicWords` a pre slová patriace do krížovky `Crosswords` a volá sa metóda `Language` (nová inicializácia využitá pri opätovnom tvorení krížovky). V ostatných stavoch sa volajú metódy, ktoré zodpovedajú názvu stavu.

#### 2. `Introduction`

Metóda predstavuje úvodnú obrazovku, na ktorej sa nachádza logo, uvítací text a tlačidlo pre presun do stavu `LANG` - výber jazyka. Jednotlivé kom-

ponenty sú vytvárané a pridávané do okna prostredníctvom kódu. Komponenty majú rôzne vlastnosti (funkcia väčšiny vlastností je jasná podľa názvu), medzi ktoré patrí aj `Dock` - určuje ako je daný prvok ukotvený v okne a ako sa mení jeho veľkosť so zmenou veľkosti okna. Používame `DockStyle.Top` teda ukotvenie k vrchnej hranici okna. Prvky sú pridávané do okna pomocou `Controls.Add` v poradí od najspodnejšieho po najvrchnejší prvok (napriek tomu, že dokumentácie uvádzajú, že prvky, ktoré majú vlastnosť `dock` sú v okne zoradené podľa poradia v akom ich vkladáme, je ich poradie v skutočnosti opačné). Nakoniec sa pridá metóda pre obsluhu tlačidla `EventHandler`.

### 3. Language

Metóda predstavuje obrazovku pre výber jazyka, ktorá zobrazuje informatívny text a dve tlačidlá. Jedno pre slovenský, jedno pre anglický jazyk. Všetky prvky majú nastavenú vlastnosť `DockStyle.Top`. Komponenty sú pridávané do okna pomocou `Controls.Add` v poradí od najspodnejšieho po najvrchnejší prvok a na záver sú pridané metódy pre obsluhu tlačidiel.

### 4. Topic

Metóda predstavuje obrazovku pre výber témy, ktorá obsahuje všetky komponenty z metódy `Language`. Saturácia farby týchto komponent je znížená, čím je indikovaná ich nefunkčnosť a tlačidlám sú odobrané metódy pre ich obsluhu. Navyše sa na obrazovke nachádza informačný text k výberu témy a prvok `ComboBox`, ktorý predstavuje padacie menu. Podľa zvoleného jazyka metóda pridá do padacieho menu prvky z poľa dostupných tém (`topics_en` alebo `topics_sk`). Medzi prvky padacieho menu je pridaný prvok *Click to select topic*, ktorý sa zobrazuje v padacom menu až pokiaľ užívateľ nerozbalí ponuku tém. Všetky komponenty obrazovky sú pridané do okna pomocou `Controls.Add` v poradí od najspodnejšieho po najvrchnejší prvok, na záver sú pridané metódy pre obsluhu tlačidiel.

### 5. Confirm

Metóda sa volá po zvolení témy z padacieho menu. Obsahuje všetky komponenty z metódy `Topic`, navyše obsahuje text s informáciou o aktuálnej zvolenej téme a tlačidlo `Confirm`, ktorým užívateľ definitívne potvrdí zvolenú tému. Pokiaľ nie je stlačené tlačidlo `Confirm`, užívateľ môže ľubovoľne meniť voľbu témy. Všetky komponenty obrazovky sú pridané do okna pomocou `Controls.Add` v poradí od najspodnejšieho po najvrchnejší prvok. Na záver sú pridané metódy pre obsluhu tlačidiel.

### 6. DefineTopic

Metóda sa používa v prípade, kedy si užívateľ spomedzi prvkov padacieho menu zvolil *define my own topic*. Obrazovka obsahuje všetky komponenty z metódy `Topic`, navyše je pridaný informačný text s inštrukciami k zadávaniu témy a prvok `TextBox` - okno, do ktorého je možné vpisovať text. Všetky komponenty obrazovky sú pridané do okna pomocou `Controls.Add` v poradí od najspodnejšieho po najvrchnejší prvok, na záver sú pridané metódy pre obsluhu textového okna.

## 7. Processing

Metóda predstavuje obrazovku čakania, kým sa na pozadí vytvára krížovka. Metóda je zavolaná akonáhle užívateľ potvrdil výber svojej témy. Obrazovka obsahuje informatívny text a pohyblivý obrázok presýpacích hodín. Všetky komponenty obrazovky sú pridané do okna pomocou `Controls.Add` v poradí od najspodnejšieho po najvrchnejší prvok. Na konci tejto metódy je inicializovaný tzv. `BackgroundWorker` a sú pridané metódy pre jeho obsluhu. `BackgroundWorker` je trieda, ktorá umožňuje vykonávať operácie a procesy v samostatnom, oddelenom vlákne.

Metóda `backgroundWorker_DoWork` predstavuje procesy vykonávané týmto prvkom na pozadí. Vytvára slovník tematických slov a následne sa pokúsi vytvoriť krížovku. Ak je zvoleným jazykom slovenčina, pomocou metódy `extractTopicRelated` získa slovník tematických slov. Extrahuje slovenské výrazy pomocou metódy `extractTranslations` a následne sa pokúsi vytvoriť krížovku pomocou metódy `createCrosswordList`. Výsledok operácie ukladá ako vlastnosť `e.Result`. Pre anglický jazyk postupuje rovnako s výnimkou použitia metódy `extractTranslations`.

Metóda `backgroundWorker_RunWorkerCompleted` potom kontroluje úspešnosť vytvorenia krížovky. Ak je vlastnosť `e.Result` `true`, aplikácia sa presunie do stavu `RULES` - pravidlá lúštenia. Ak je `e.Result` `false` a nepodarilo sa vytvoriť krížovku, volá sa metóda `noPuzzle`, ktorá upozorní užívateľa na neúspech a dá mu možnosť zvoliť si novú tému alebo pokračovať s náhodnou krížovkou.

## 8. Rules

Metóda predstavuje obrazovku s pravidlami lúštenia. Obsahuje text s popisom tvaru krížovky, spôsobom pohybu po krížovke, inštrukciami pre zadávanie písmen a popisom dostupných tlačidiel. Text je modifikovaný podľa zvoleného jazyka. Pre slovenčinu je pridané upozornenie o používaní diakritiky. Obrazovka obsahuje aj tlačidlá `Start` a `Return` pre zahájenie lúštenia, respektíve návrat na voľbu jazyka. Všetky komponenty obrazovky sú pridané do okna pomocou `Controls.Add` v poradí od najspodnejšieho po najvrchnejší prvok. Na záver sú pridané metódy pre obsluhu tlačidiel.

## 9. Puzzle

Metóda predstavuje obrazovku s krížovkou. Obsahuje mriežku krížovky, legendu, nápovedu k legende a pomocné tlačidlá. Vo vrchnej časti obrazovky sa nachádza prvok `TableLayoutPanel` - tabuľka obsahujúca 4 stĺpce, v ktorých sú vložené pomocné tlačidlá `Help`, `Show`, `Restart`, `End`. Tieto tlačidlá majú oproti iným komponentám nastavenú vlastnosť `DockStyle Fill`, pretože chceme, aby vyplnili celú bunku prvku `TableLayoutPanel`. Tlačidlá sú pridávané do konkrétneho stĺpca tabuľky podľa čísla stĺpca. Následne sú pridané metódy pre obsluhu tlačidiel.

Krížovka, legenda a nápoveda k legende sú reprezentované prvkom `DataGridView` - ide o tabuľku, do ktorej buniek je možné vpisovať, a ktorej bunky je možné ľubovoľne formátovať. Pre rozmiestnenie komponent na obrazovke používa metóda prvok `SplitContainer` - pomocou posuvnej lišty rozdeľuje

obrazovku na dva panely. Do ľavého z panelov metóda pridá mriežku krížovky, do pravého z panelov pridá mriežku legendy spolu s nápodou k legende. Mriežky majú nastavené rôzne vlastnosti. Patrí medzi ne aj definovanie základného štýlu buniek mriežky, ktorý je rozličný pre všetky spomínané mriežky. Líšia sa vo farbe a možnosti vpisovania údajov. Metóda následne vyplňa riadky legendy a riadky nápodu k legende. Do legendy sú vkladané definície slov zo zoznamu slov tvoriacich krížovku **Crossword**. Pridaná je aj metóda pre obsluhu vpisovania znakov do krížovky. Pomocou metódy `formCrossword` sa vytvorí mriežka krížovky. Zvyšné bunky, ktoré nepatria do krížovky nesmú mať povolené vpisovanie - to je zaručené pomocou vlastnosti `ReadOnly`.

#### 10. `noPuzzle`

Metóda sa používa ak sa aplikácii nepodarilo vytvoriť krížovku. Inicializuje okno druhého formulára `Form2`, ktorý slúži ako prispôsobiteľné vyskakovacie okno pre chybové hlásenia. Zobrazuje informáciu o neúspešnom vytvorení krížovky a dáva užívateľovi možnosť znovu si vybrať tému alebo pokračovať s náhodnou krížovkou. Pre výber ďalšieho postupu slúžia tlačidlá s textom `Choose topic` a `Random puzzle`. Na záver sú pridané metódy pre obsluhu tlačidiel.

## 6.2.2 Obsluha interaktívnych komponent

V aplikácii používame množstvo interaktívnych komponent, najmä tlačidiel, ktorými sa užívateľ presúva medzi stavmi aplikácie, alebo ktorými vyvoláva rozličné procesy. Tento zoznam detailne popisuje funkcie jednotlivých komponent:

#### 1. `introB_Click`

Tlačidlo `introB` sa zobrazuje na úvodnej obrazovke. Po kliknutí na tlačidlo sa obrazovka vyčistí a aplikácia sa presunie do stavu `LANG` - výber jazyka.

#### 2. `lang_SK_Click` / `lang_EN_Click`

Obe tlačidlá `lang_SK` a `lang_EN` sa zobrazujú na obrazovke pre výber jazyka. Po kliknutí na niektoré z tlačidiel sa uloží zvolený jazyk do premennej `language`, obrazovka sa vyčistí a aplikácia sa presunie do stavu `TOPIC` - výber témy.

#### 3. `ta_SelectedIndexChanged`

`ta` v názve metódy symbolizuje `ComboBox topics_avail`. Ide o padacie menu, ktoré obsahuje dostupné témy, z ktorých si užívateľ môže vyberať kliknutím myšou, a ktoré sa zobrazuje na obrazovke pre výber témy. Po kliknutí na položku z padacieho menu sa spustí metóda `SelectedIndexChanged`, ktorá vo vlastnosti `SelectedIndex` uchováva index zvolenej položky. Podľa zvoleného jazyka v premennej `language` získa metóda pole tém (`topics_en` alebo `topics_sk`). Ak je zvolenou témou posledná položka poľa (`define my own topic`) zavolá sa metóda `DefineTopic` - definícia vlastnej témy. Ak je zvolená náhodná krížovka (`random`) nastaví sa premenná `topic` predstavujúca tému na prázdny reťazec. Inak sa do premennej `topic` uloží

téma z pola tém na pozícii `SelectedIndex`, vyčistí sa obrazovka a zavolá sa metóda `Confirm` - potvrdenie zvolenej témy.

#### 4. `ta_DropDown`

`ta` v názve metódy znovu symbolizuje `ComboBox topics_avail`. Metóda `DropDown` je zavolaná, ak je padacie menu rozkliknuté. Odstráni z dostupných tém položku *Click to select topic*, ktorá sa zobrazuje užívateľovi až do chvíle, kedy rozbalí padacie menu.

#### 5. `continueB_Click`

Tlačidlo `continueB` sa zobrazuje po zvolení témy, a slúži na potvrdenie zvolenej témy. Po kliknutí sa vyčistí obrazovka a aplikácia sa presunie do stavu `PROCESS` - prehľadávanie dát a vytvorenie krížovky.

#### 6. `topic_Entered`

Metóda sa používa počas procesu definovania vlastnej témy. Je vyvolaná stlačením tlačidla `Enter`. Metóda potlačí zvukovú signalizáciu po stlačení tlačidla `Enter` a pridá na obrazovku nové tlačidlo `Confirm` pre potvrdenie definovanej témy.

#### 7. `define_Clicked`

Metóda je vyvolaná kliknutím na textové okno `custom_Topic`, kam užívateľ vpisuje tému. Toto okno v úvode obsahuje informatívny text *Enter text here....* Po kliknutí sa textové okno vyčistí a užívateľ doň môže priamo písať.

#### 8. `confirmB_Click`

Tlačidlo `confirmB` slúži na potvrdenie témy definovanej užívateľom. Po kliknutí sa metóda pokúsi v databáze slov nájsť zadané slovo. Ak dané slovo nájde, vyčistí obrazovku a aplikácia sa presunie do stavu `PROCESS` - prehľadávanie dát a vytvorenie krížovky. Ak sa dané slovo nenachádza v databáze, zobrazí užívateľovi vyskakovacie okno (`MessageBox`) s upozornením a výzvou na skontrolovanie správnosti tvaru tohto slova, prípadne výber novej témy.

#### 9. `StartB_Click`

Tlačidlo `StartB` sa zobrazuje na obrazovke s pravidlami a slúži na zahájenie lúštenia krížovky. Po kliknutí sa vyčistí obrazovka a aplikácia sa presunie do stavu `PUZZLE` - krížovka.

#### 10. `ReturnB_Clicked`

Tlačidlo `ReturnB` sa takisto zobrazuje na obrazovke s pravidlami a slúži na návrat do stavu pre výber jazyka. Po kliknutí sa vyčistí obrazovka a aplikácia sa presunie do stavu `LANG` - výber jazyka.

#### 11. `helpB_Click`

Tlačidlo `helpB` sa zobrazuje počas lúštenia krížovky. Slúži na doplnenie písmena na náhodnú pozíciu v krížovke, ktorá nie je tajničkou. Pomocou

metódy `ExistsEmpty` zistí, či sa v krížovke nachádza prázdna bunka, ktorá nie je tajničkou. Ak áno, na náhodnú prázdnu bunku krížovky, ktorá nie je tajničkou, doplní správne písmeno (uložené ako vlastnosť bunky `Tag`). Ak prázdna bunka neexistuje, zobrazí užívateľovi vyskakovacie okno (`MessageBox`) s upozornením a výzvou na kompletizáciu tajničky alebo ukončenie lúštenia.

#### 12. `showB_Click`

Tlačidlo `showB` sa zobrazuje počas lúštenia krížovky. Slúži na vyplnenie všetkých buniek krížovky. Pomocou metódy `isCompleted` zistí, či je krížovka kompletne vyplnená. Ak áno, zobrazí užívateľovi vyskakovacie okno (`MessageBox`) s výzvou na ukončenie lúštenia. Ak nie je vyplnená, na každú prázdnu bunku krížovky doplní správne písmeno (uložené ako vlastnosť bunky `Tag`).

#### 13. `restartB_Click`

Tlačidlo `restartB` sa zobrazuje počas lúštenia krížovky. Po kliknutí vymaže obsah všetkých buniek krížovky.

#### 14. `endB_Click`

Tlačidlo `endB` sa zobrazuje počas lúštenia krížovky. Po kliknutí inicializuje druhý formulár `Form2`, okno ukončenia aplikácie, kde informuje užívateľa o uzatváraní aplikácie a pýta sa, či chce začať novú krížovku. Zobrazuje dve tlačidlá, `Yes` a `No`.

#### 15. `yesButton_Click`

Tlačidlo `yesButton` sa zobrazuje v okne pre ukončenie aplikácie. Po kliknutí zavrie vyskakovacie okno, vyčistí obrazovku a presunie aplikáciu do stavu `LANG` - výber jazyka.

#### 16. `noButton_Click`

Tlačidlo `noButton` sa zobrazuje v okne pre ukončenie aplikácie. Po kliknutí sa aplikácia ukončí.

#### 17. `topicButton_Click`

Tlačidlo `topicButton` sa zobrazuje vo vyskakovacom okne, ktoré oznamuje, že sa aplikácii nepodarilo vytvoriť krížovku. Slúži pre návrat na výber témy. Po kliknutí vyčistí obrazovku a presunie aplikáciu do stavu `LANG` - výber jazyka, následne výber témy.

#### 18. `randomButton_Click`

Tlačidlo `randomButton` sa zobrazuje vo vyskakovacom okne, ktoré oznamuje, že sa aplikácii nepodarilo vytvoriť krížovku. Po kliknutí vyčistí obrazovku, nastaví premennú `topic` na prázdny reťazec a presunie aplikáciu do stavu `PROCESS` - prehľadávanie dát a vytvorenie krížovky.

### 6.2.3 Prehľadávanie databázy

Algoritmus prehľadávania, detailne popísaný v sekcii 5.3 *Zostavenie tematického slovníka*, pozostáva z troch hlavných krokov.

- nájdenie prvotnej množiny slov týkajúcich sa témy
- nájdenie množiny hyponým a hyperoným pre každé slovo z prvotnej množiny
- prehľadávanie hyponým ďalej po hyponymách a prehľadávanie hyperoným ďalej po hyperonymách v piatich iteráciách

Nasledujúci zoznam podrobne popisuje všetky metódy využívané pri prehľadávaní dát:

#### 1. Deserialize

Metóda slúži pre načítanie databázy WordNet zo súboru `WordNet.json` do dátovej štruktúry `WordNet`. Pre načítanie využíva metódu `Deserialize`, ktorú ponúka balíček `Newtonsoft.json`.

#### 2. extractTranslations

Metóda sa používa pri zvolení jazyka slovenčina a slúži pre extrakciu slovenských tematických slov. Najprv vytvorí nový prázdny zoznam slov. Vezme slovník tematických slov uložený v dátovej štruktúre `TopicWords` a každé slovo v tejto štruktúre, ktoré má slovenský preklad pridá do zoznamu. Potom tento zoznam uloží ako štruktúru `TopicWords`.

#### 3. extractTopicRelated

Metóda predstavuje celý algoritmus prehľadávania. Vytvára slovník tematických slov uložený ako `TopicWords`. Ako parameter prijíma premennú `topic` (zvolená téma). Ak je `topic` prázdny reťazec, užívateľ si zvolil náhodnú krížovku a metóda pridá do slovníka tematických slov `TopicWords` všetky slová z databázy `WordNet`. Ak `topic` nie je prázdny reťazec, do dátových štruktúr `oldHypos` a `oldHypers` uloží pomocou metódy `searchTopic` hyponymá a hyperonymá prvotnej množiny slov. Následne v piatich iteráciách pomocou metódy `searchPointers` prehľadáva štruktúry `oldHypos` a `oldHypers` a príslušné hyponymá a hyperonymá ukladá do štruktúr `newHypos` a `newHypers`. Aktualizuje štruktúry hyponým a hyperoným, teda „nové štruktúry (new-)“ uloží ako „staré štruktúry (old-)“ a pokračuje v ďalšej iterácii.

#### 4. searchTopic

Metóda slúži pre nájdenie prvotnej množiny slov, ich pridanie do výsledného slovníka a extrakciu hyponým a hyperoným tejto množiny slov. Metóda prijíma ako parametre reťazec reprezentujúci tému `topic`, a slovníky ukazovateľov pre hyponymá a hyperonymá (`Dictionary<string, char>`). Pomocou metódy `FindAll`, ktorá vracia kolekciu prvkov spĺňajúcich stanovený predikát, nájde prvotnú množinu záznamov (slová, ktoré sa začínajú reťazcom témy). Následne každý zo záznamov pridá do výsledného slovníka



**TopicWords**. Pomocou metódy `searchSynsets` nájde množinu hyponým záznamu, ktorú uloží do slovníka hyponým a rovnako množinu hyperoným záznamu, ktorú uloží do slovníka hyperoným.

#### 5. `searchSynsets`

Metóda slúži na extrakciu hyponým alebo hyperoným záznamu. Metóda prijíma ako parametre záznam, slovník ukazovateľov (`Dictionary<string,-char>`) a číslo predstavujúce typ ukazovateľov (0=hyponymá, 1=hyperonymá). Podľa typu ukazovateľa hľadá hyponymá alebo hyperonymá vstupného záznamu a pokiaľ sa daný ukazovateľ nevyskytuje v slovníku ukazovateľov, metóda ho tam pridá.

#### 6. `searchPointers`

Metóda slúži na nájdenie záznamu v databáze WordNet a extrakciu jeho hyponým alebo hyperoným. Ako parametre prijíma vstupný a výstupný slovník ukazovateľov (`Dictionary<string,char>`) a číslo predstavujúce typ ukazovateľov, ktoré hľadáme (0=hyponymá, 1=hyperonymá). Pre každý ukazovateľ vo vstupnom zozname nájde príslušný záznam v databáze WordNet. Pridá záznam do výsledného slovníka tematických slov **TopicWords** a pomocou metódy `searchSynsets` pridá do výstupného slovníka podľa hľadaného typu ukazovateľov hyponymá alebo hyperonymá záznamu.

### 6.2.4 Tvorba krížoviek

V nasledujúcom zozname detailne popíšeme metódy používané pre tvorbu krížoviek, či už logicky, teda skladaním slov, alebo graficky, teda formovaním mriežky krížovky a jej formátovaním:

#### 1. `createCrosswordList`

Metóda zo slovníka tematických slov **TopicWords** náhodne vyberie slová, ktoré budú tvoriť krížovku, a uloží ich do štruktúry **Crossword**. Ako parameter prijíma reťazec reprezentujúci zvolený jazyk `lang`. Záznamy v slovníku tematických slov sú tvaru *slovenské slovo : anglické slovo : definícia*, teda podľa zvoleného jazyka si metóda uloží index želannej časti záznamu (0 pre slovenské slovo, 1 pre anglické slovo). Náhodne zvolí záznam zo slovníka, ktorý reprezentuje tajničku (horizontálne slovo) a rozdelí ho podľa dvojbodky pomocou metódy `Split` na jednotlivé časti. Ak je slovo príliš krátke (menej ako 5 znakov) alebo obsahuje číslicu, opakuje výber. Po úspešnom výbere uloží slovo spolu s definíciou do výslednej štruktúry **Crossword** a pokračuje výberom vertikálnych slov. Pre každý znak horizontálneho slova náhodne vyberie záznam zo slovníka tematických slov a rozdelí ho na jednotlivé časti. Ak slovo neobsahuje hľadaný znak alebo obsahuje číslicu, prípadne sa už nachádza vo výslednej štruktúre **Crossword**, výber sa opakuje. Inak sa slovo spolu s definíciou pridá do štruktúry **Crossword** a iterácia pokračuje na ďalšie písmeno horizontálneho slova. Náhodný výber vertikálneho slova metóda opakuje maximálne 20-krát. Ak výber zlyhal viac ako 20-krát, vyprázdni štruktúru **Crossword** a opakuje výber horizontálneho

slova. Ak je výber horizontálneho slova opakovaný dvadsiaty krát, nepodarilo sa vytvoriť krížovku a metóda vracia pravdivostnú hodnotu `false`, inak vracia hodnotu `true`, teda úspech.

## 2. `formatCell`

Metóda slúži na formátovanie buniek patriacich do krížovky. Ako parametre prijíma x-ovú a y-ovú súradnicu bunky, znak, ktorý sa v danej bunke má správne vyskytovať a číslo, ktoré označuje, či je daná bunka tajničkou (0=nie je tajničkou, 1=je tajničkou). Metóda podľa čísla nastaví farbu bunky na danej pozícii (klasické bunky sú sivé, bunky tajničky sú zlaté). Nastaví farbu výberu bunky na svetlú červenú. Vlastnosť `Tag` nastaví na správne písmeno, a umožní vpisovanie do bunky pomocou vlastnosti `ReadOnly`.

## 3. `numberCell`

Metóda slúži na očíslovanie bunky. Ako parametre prijíma x-ovú a y-ovú súradnicu bunky a číslo. Bunku krížovky na danej pozícii zvýrazní svetlomodrou farbou, vloží do bunky číslo a znemožní užívateľovi vpisovanie do bunky pomocou vlastnosti `ReadOnly`.

## 4. `formCrossword`

Metóda slúži na vyformovanie mriežky krížovky. Najprv vypočíta maximálny obdĺžnik opisujúci krížovku. Počet stĺpcov je rovný dĺžke horizontálneho slova. Pre výpočet počtu riadkov postupne prechádza znaky horizontálneho slova. V križujúcom slove nájde pomocou metódy `IndexOf` priesečník a zistí, koľko znakov sa nachádza pred a koľko za týmto priesečníkom. Inými slovami, koľko buniek potrebuje nad tajničkou a koľko pod tajničkou. Vždy udržiava maximum týchto čísel. Do výslednej mriežky (`DataGridView crossword`) pridá počet stĺpcov a zistený počet riadkov + 2 (jeden pre tajničku, jeden pre číslo stĺpca). Metóda formuje krížovku od tajničky. Prvá bunka tajničky sa nachádza na pozícii maximálnej výšky + 1. Prechádza jednotlivé znaky tajničky. Pre každý znak nájde priesečník s príslušným vertikálnym slovom a zistí koľko buniek bude formovať smerom nahor a koľko smerom nadol. Pomocou metódy `formatCell` zvýrazňuje bunky smerom nahor od pôvodnej pozície tajničky. Pomocou metódy `NumberCell` očísľuje bunku na samom vrchu stĺpca slova v krížovke. Rovnakým spôsobom formátuje bunky smerom nadol od pôvodnej pozície tajničky. Po vyformovaní celej mriežky krížovky zvýrazní bunky tajničky.

## 5. `existsEmpty`

Metóda zisťuje, či v krížovke existuje prázdna bunka, ktorá nie je tajničkou. Výsledok vracia ako pravdivostnú hodnotu `true` alebo `false`.

## 6. `isCompleted`

Metóda zisťuje, či je krížovka vrátane tajničky vyplnená. Výsledok vracia ako pravdivostnú hodnotu `true` alebo `false`.

## 7. `crossWP_CellValueChanged`

Metóda je vyvolaná ak sa zmení hodnota bunky na mriežke krížovky. Ak je hodnota bunky prázdna (`null`), metóda to ignoruje (pri vymazaní obsahu buniek krížovky nechceme vyvolávať metódu). Ak bol vložený znak iný ako písmeno, zobrazí sa vyskakovacie okno `MessageBox` s upozornením, že je potrebné vkladať len písmená a hodnota bunky sa vymaže (toto upozornenie je potlačené pri číslovaní buniek počas tvorby mriežky krížovky). Inak metóda kontroluje dĺžku zadaného reťazca. Ak presiahol dĺžku 2, zobrazí sa vyskakovacie okno `MessageBox` s upozornením, že do jednej bunky krížovky je možné vložiť len jedno písmeno. Ak sa žiadny z týchto problémov nevyskytol, metóda kontroluje správnosť zadaného znaku porovnaním s vlastnosťou bunky `Tag`. Pri správnom písmene je jeho farba čierna, pri nesprávnom písmene sa jeho farba zmení na červenú (nastavením vlastnosti bunky `ForeColor`).

## 6.3 Užitočné zdroje

Počas vývoja aplikácie sme využívali množstvo návodov pre implementáciu technických riešení, ako aj dokumentácie popisujúce funkcie metód a grafických komponent. V nasledujúcom zozname uvádzame užitočné odkazy na dokumentácie a nápomocné stránky:

Regular Expression Language - Quick Reference

<https://docs.microsoft.com/en-us/dotnet/standard/base-types/regular-expression-language-quick-reference>

Grouping Constructs in Regular Expressions

<https://docs.microsoft.com/en-us/dotnet/standard/base-types/grouping-constructs-in-regular-expressions>

Regex.Matches Method

<https://docs.microsoft.com/en-us/dotnet/api/system.text.regularexpressions.regex.matches?view=net-5.0>

Regex.Replace Method

<https://docs.microsoft.com/en-us/dotnet/api/system.text.regularexpressions.regex.replace?view=net-5.0>

Jak serializovat a deserializovat (zařazování a zrušit zařazení) JSON v .NET

<https://docs.microsoft.com/cs-cz/dotnet/standard/serialization/system-text-json-how-to?pivots=dotnet-5-0>

Serialization (C#)

<https://docs.microsoft.com/en-us/dotnet/csharp/programming-guide/concepts/serialization/>

Serializing and Deserializing JSON

<https://www.newtonsoft.com/json/help/html/SerializingJSON.htm>

Install and manage packages in Visual Studio using the NuGet Package Manager

<https://docs.microsoft.com/sk-sk/nuget/consume-packages/install-all-use-packages-visual-studio>

DataGridView Class

<https://docs.microsoft.com/en-us/dotnet/api/system.windows.forms.datagridview?view=net-5.0>

How to: Set Default Cell Styles for the Windows Forms DataGridView Control

<https://docs.microsoft.com/en-us/dotnet/desktop/winforms/controls/how-to-set-default-cell-styles-for-the-windows-forms-datagridview-control?view=netframeworkdesktop-4.8>

Position and layout of controls (Windows Forms .NET)

<https://docs.microsoft.com/en-us/dotnet/desktop/winforms/controls/layout?view=netdesktop-5.0>

ComboBox Class

<https://docs.microsoft.com/en-us/dotnet/api/system.windows.forms.combobox?view=net-5.0>

SplitContainer Class

<https://docs.microsoft.com/en-us/dotnet/api/system.windows.forms.splitcontainer?view=net-5.0>

BackgroundWorker Class

<https://docs.microsoft.com/en-us/dotnet/api/system.componentmodel.backgroundworker?view=net-5.0>

TableLayoutPanel Class

<https://docs.microsoft.com/en-us/dotnet/api/system.windows.forms.tablelayoutpanel?view=net-5.0>

C# Disable Or Hide Close Button(X) in Windows Form

<http://csharpexamples.com/c-disable-hide-close-buttonx-windows-form/>

How to deploy a windows form application including database to another PC

<https://stackoverflow.com/questions/2173336/how-to-deploy-a-windows-form-application-including-database-to-another-pc>

V tejto kapitole sme si popísali programové prevedenie aplikácie a pomocného programu. Predstavili sme si dva programy - **WordSourceMaker** slúžiaci na spracovanie databáz WordNet a **ThematicCrosswords**, ktorý je výslednou aplikáciou pre generovanie tematických hrebeňoviek s grafickým prostredím. Podrobne sme popísali metódy používané v jednotlivých programoch, procesy, ktoré vykonávajú ale aj obsluhu interaktívnych komponent výslednej aplikácie.

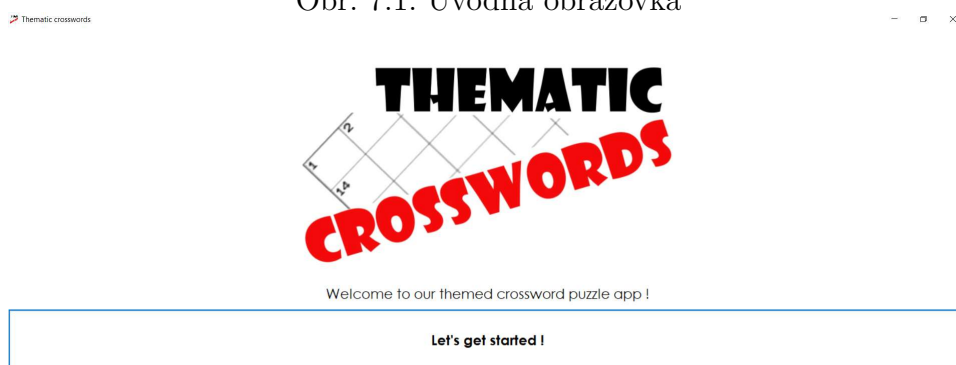
## 7. Uživatelská dokumentácia

Aplikácia sa podľa interakcie s užívateľom presúva medzi šiestimi základnými stavmi : *úvod*, *voľba jazyka*, *voľba témy*, *čakanie*, *pravidlá*, *krížovka*. V tejto kapitole sa na aplikáciu pozrieme z pohľadu užívateľa. Detailne si predstavíme jednotlivé stavy programu, nástroje, ktoré sa zobrazujú v okne, spôsob akým sa ovládajú, a procesy, ktoré spúšťajú.

### 7.1 Úvod

Priamo po spustení sa aplikácia nachádza v stave *úvod* a užívateľovi je zobrazená úvodná obrazovka. Obsahuje logo, uvítací text a tlačidlo pre pokračovanie.

Obr. 7.1: Úvodná obrazovka

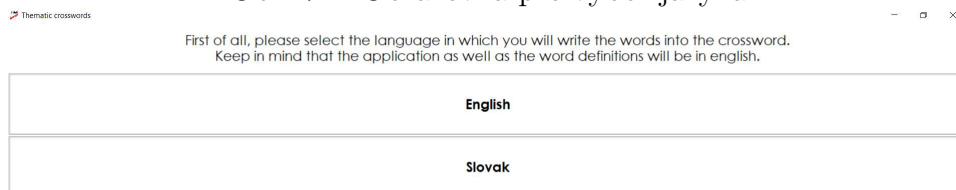


Stlačením tlačidla **Let's get started !** sa aplikácia presunie do stavu pre výber jazyka.

### 7.2 Voľba jazyka

V stave pre výber jazyka sa na obrazovke zobrazuje text s výzvou pre zvolenie jazyka a užívateľ má k dispozícii dve tlačidlá. **English** pre anglický jazyk a **Slovak** pre slovenský jazyk. Jazykom rozumieme jazyk slov dopĺňaných do krížovky, legenda je vždy v anglickom jazyku.

Obr. 7.2: Obrazovka pre výber jazyka

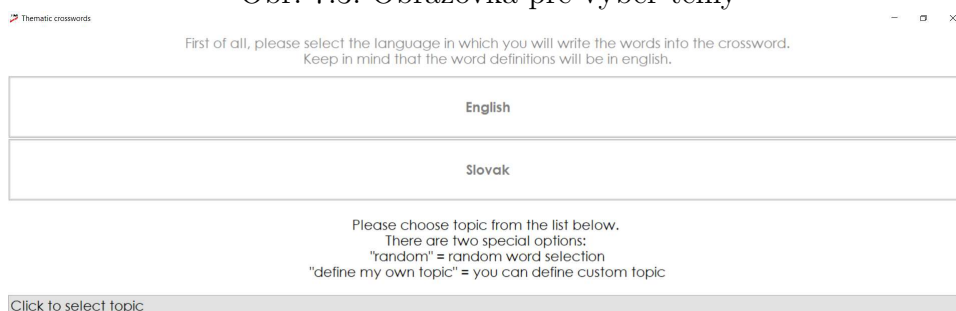


Po kliknutí na ľubovoľné z tlačidiel sa aplikácia presúva do stavu pre výber témy.

## 7.3 Voľba témy

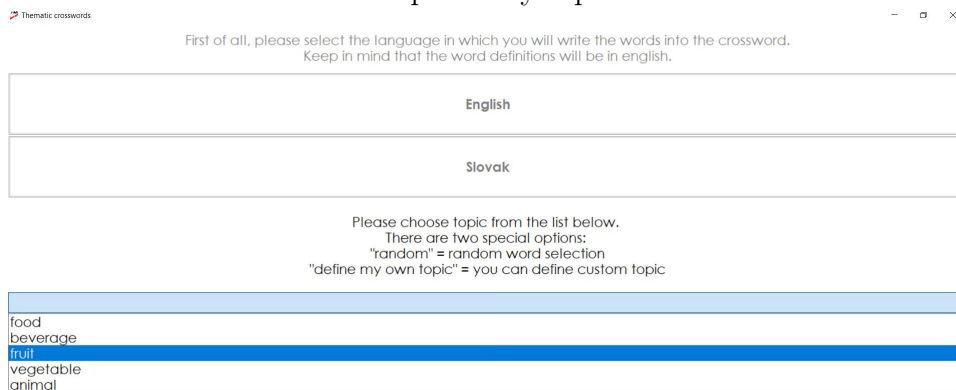
V stave pre výber témy sa na obrazovke vyskytujú všetky komponenty zo stavu pre výber jazyka. Saturácia farby je znížená a tlačidlá pre výber jazyka sa zdajú byť sivé, čo značí, že sa už nemajú používať. Kliknutie na tieto tlačidlá nevyvolá žiadnu akciu. Novými komponentami na obrazovke sú sprievodný text k výberu témy a padacie menu, ktoré obsahuje všetky dostupné témy.

Obr. 7.3: Obrazovka pre výber témy



Po kliknutí na padacie menu sa zobrazia dostupné témy. Užívateľ si ich môže prezerat prostredníctvom posuvníka a výber uskutočniť kliknutím. Výberom témy sa na obrazovke zobrazí oznam s aktuálne zvolenou témou a tlačidlo pre potvrdenie výberu. Stlačením tlačidla **Confirm** sa aplikácia presunie do stavu čakania. Pokiaľ nebolo tlačidlo **Confirm** stlačené je možné výber témy ľubovoľne meniť.

Obr. 7.4: Dostupné témy v padacom menu



Obr. 7.5: Aktuálne zvolená téma a tlačidlo potvrdenia

Thematic crosswords

First of all, please select the language in which you will write the words into the crossword.  
Keep in mind that the word definitions will be in english.

English

Slovak

Please choose topic from the list below.  
There are two special options:  
"random" = random word selection  
"define my own topic" = you can define custom topic

fruit

You chose topic: fruit

Confirm

Ak si užívateľ spomedzi dostupných tém zvolí **define my own topic**, zobrazí sa textové okno, do ktorého je možné písať a inštrukcie k definovaniu témy. Užívateľ musí tému vpísanú do textového okna potvrdiť stlačením klávesy **Enter**. Následne sa na obrazovke zobrazí tlačidlo pre potvrdenie definovanej témy.

Obr. 7.6: Textové okno pre zadanie témy

Thematic crosswords

First of all, please select the language in which you will write the words into the crossword.  
Keep in mind that the application as well as the word definitions will be in english.

English

Slovak

Please choose topic from the list below.  
There are two special options:  
"random" = random word selection  
"define my own topic" = you can define custom topic

define my own topic

Please write your desired topic as a single english noun in the basic form in lowercase. Then press enter.  
Remember that a word other than the basic form may cause a malfunction.

Enter text here...

Obr. 7.7: Potvrdenie témy zadanej užívateľom

Thematic crosswords

First of all, please select the language in which you will write the words into the crossword.  
Keep in mind that the application as well as the word definitions will be in english.

English

Slovak

Please choose topic from the list below.  
There are two special options:  
"random" = random word selection  
"define my own topic" = you can define custom topic

define my own topic

Please write your desired topic as a single english noun in the basic form in lowercase. Then press enter.  
Remember that a word other than the basic form may cause a malfunction.

geometry

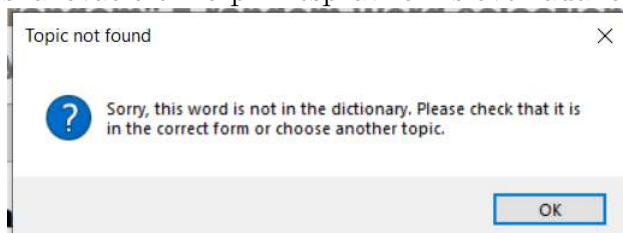
Confirm

Po stlačení tlačidla **Confirm** sa uloží definovaná téma a aplikácia skontroluje, či sa dané slovo nachádza v databáze WordNet. Ak áno, aplikácia sa presunie do

stavu čakania. Ak nie, zobrazí sa vyskakovacie okno. Informuje užívateľa, že sa dané slovo nevyskytuje v slovníku a vyzve ho k skontrolovaniu tvaru zadaného slova alebo k zvoleniu inej témy.

Stlačením tlačidla **Confirm** sa ukladá téma, ktorá je aktuálne napísaná v textovom okne. To využijeme v prípade, že si užívateľ rozmyslel voľbu témy, definoval novú tému ale zabudol stlačiť klávesu **Enter**.

Obr. 7.8: Vyskakovacie okno pri nesprávnom slove zadanom užívateľom

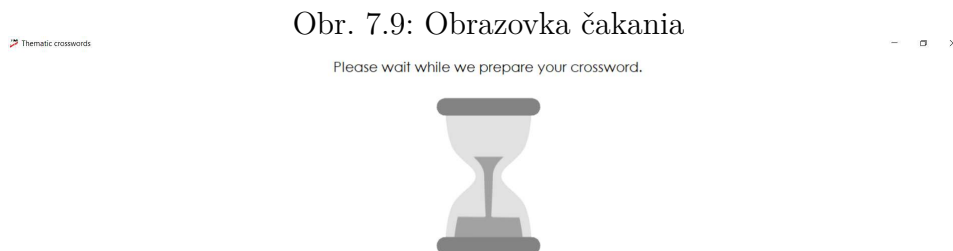


## 7.4 Čakanie

Generovanie krížovky pozostáva z dvoch základných krokov:

- vytvorenie slovníka tematických slov
- vytvorenie zoznamu slov tvoriacich výslednú krížovku

Tento proces môže trvať niekoľko sekúnd, preto je vykonávaný na pozadí. Počas pripravovania krížovky sa užívateľovi zobrazuje obrazovka čakania. Obsahuje informáciu o prebiehajúcom procese tvorby krížovky a obrázok presýpacích hodín.

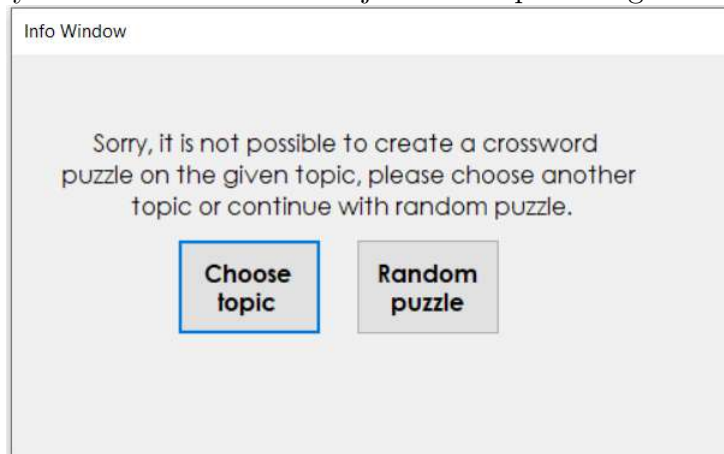


Obr. 7.9: Obrazovka čakania

Pokiaľ sa aplikácii nepodarilo vytvoriť krížovku, zobrazí sa vyskakovacie okno. Informuje užívateľa o zlyhaní a ponúka dve tlačidlá. **Choose topic** slúži pre návrat užívateľa do stavu výberu témy. **Random puzzle** presunie užívateľa znovu do stavu čakania a aplikácia vygeneruje krížovku obsahujúcu náhodné slová.



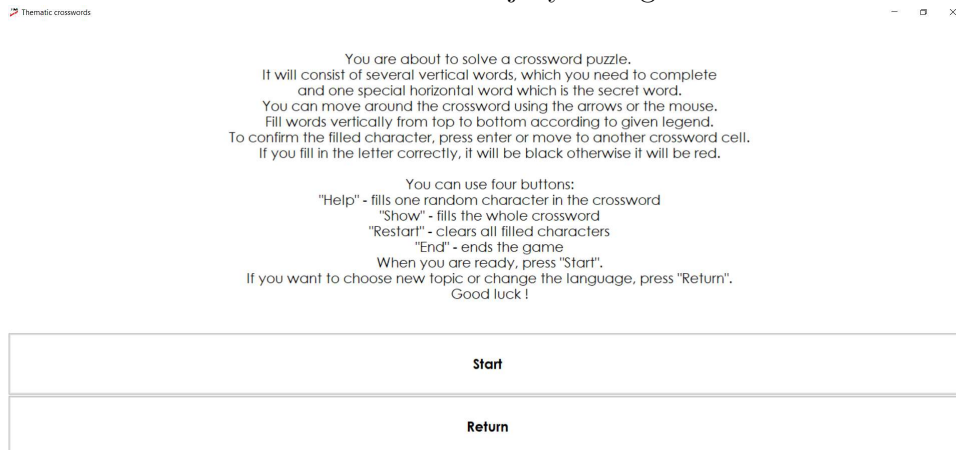
Obr. 7.10: Vyskakovacie okno informujúce o neúspešnom generovaní krížovky



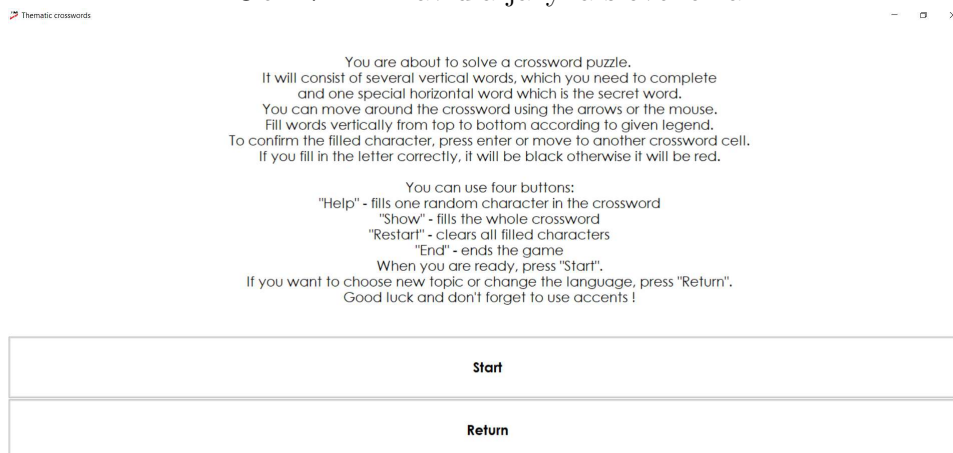
## 7.5 Pravidlá

V stave *pravidlá* sa užívateľovi zobrazí obrazovka so základnými inštrukciami pre lúštenie krížovky. Obsahuje popis tlačidiel, ktoré má užívateľ počas lúštenia k dispozícii a dve tlačidlá pomenované **Start** a **Return**. Tlačidlo **Start** presunie aplikáciu do stavu *krížovka* a zobrazí užívateľovi krížovku aj s legendou. Tlačidlo **Return** presunie aplikáciu do stavu pre výber jazyka a následne výber témy.

Obr. 7.11: Pravidlá jazyka angličtina



Obr. 7.12: Pravidlá jazyka slovenčina

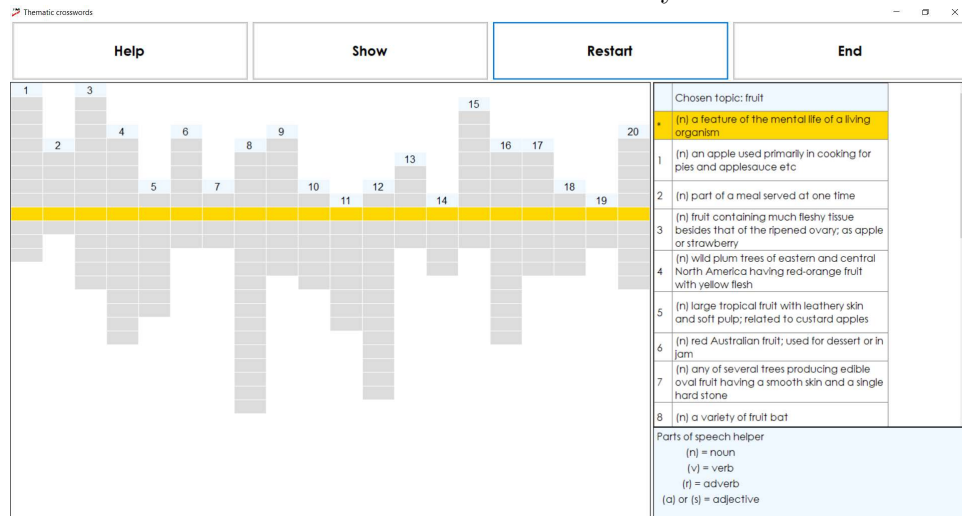


## 7.6 Krížovka

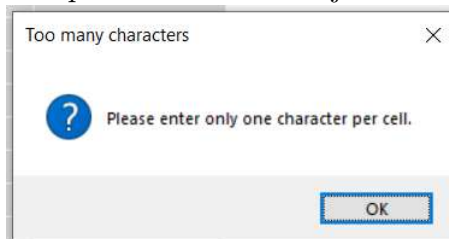
V stave *krížovka* je obrazovka celkovo rozdelená na tri časti. V hornej časti sa nachádza skupina pomocných tlačidiel, ktoré popíšeme v nasledujúcej sekcii *Tlačidlá*. Zvyšok obrazovky je rozdelený na dve časti, kde ľavá strana je vyplnená krížovkou a pravá strana je vyplnená legendou. Krížovka je vyobrazená pomocou zvýraznenia buniek sivou farbou na bielej mriežke. Bunky tajničky sú zvýraznené zlatou farbou. Stĺpce krížovky sú očíslované. Bunky, ktoré nie sú označené sivou farbou, a teda nepatria do krížovky sú neaktívne. Kliknutím na ne sa nespúšťa žiadna akcia a nie je možné do týchto buniek písať. Po kliknutí na bunku krížovky sa jej farba zmení na svetlo červenú, indikuje výber bunky. Užívateľ môže do týchto buniek vpisovať. Do každej bunky je možné vložiť len jeden znak, ktorý je písmenom. Ak je vložených znakov viac, zobrazí sa vyskakovacie okno s upozornením, že je treba vkladať len jeden znak na jednu bunku. Ak je vložený znak číslica, alebo iný špeciálny znak, ktorý nie je písmenom, zobrazí sa vyskakovacie okno s upozornením. Ak je vložený znak správny, jeho farba je čierna. Ak je nesprávny, zvýrazní sa červenou farbou.

Legenda v pravej časti obrazovky pozostáva z tabuľky, ktorá má dva stĺpce. Prvý stĺpec obsahuje poradové číslo, ktoré odpovedá stĺpcu na mriežke krížovky kam je potrebné slovo vložiť. Nasleduje definícia slova. Riadok reprezentujúci tajničku je zvýraznený zlatou farbou a namiesto poradového čísla je označený hviezdíčkom (\*). Na samom vrchu legendy je riadok zobrazujúci zvolenú tému zvýraznený svetlomodrou farbou. Pod legendou sa nachádza menšie pomocné okno, kde sú vysvetlené skratky slovných druhov nachádzajúce sa v definíciách.

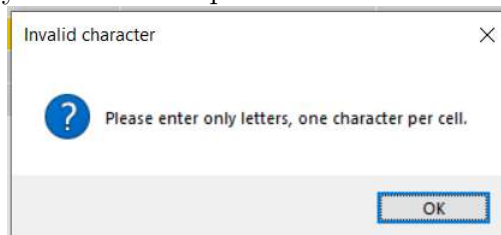
Obr. 7.13: Príklad krížovky



Obr. 7.14: Chybové hlásenie pri vložení viac ako jedného znaku do bunky krížovky



Obr. 7.15: Chybové hlásenie pri vložení znaku iného ako písmeno

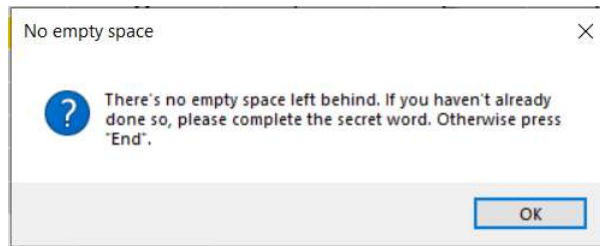


## 7.7 Tlačidlá

Užívateľ má počas lúštenia k dispozícii sadu štyroch tlačidiel :

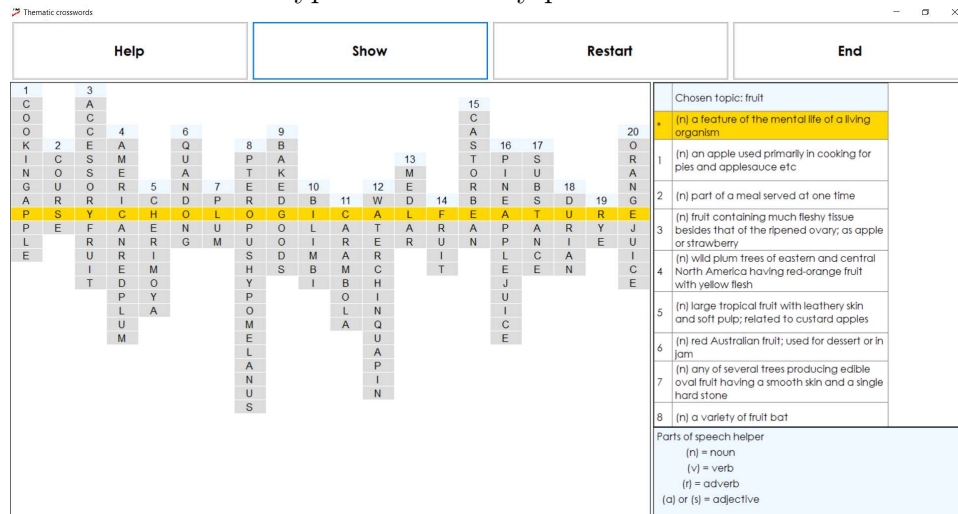
- **Help** - kliknutím sa na náhodnú voľnú pozíciu v krížovke doplní znak, voľnou pozíciou rozumieme ľubovoľnú bunku krížovky, ktorá nepatrí do tajničky, ak sa v krížovke nenachádza voľná bunka, zobrazí sa vyskakovacie okno s upozornením

Obr. 7.16: Stlačenie tlačidla Help v čase keď sa v krížovke nenachádza voľná bunka

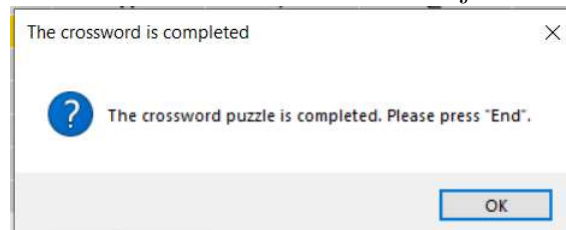


- **Show** - kliknutím sa vyplnia všetky bunky krížovky, vrátane buniek tajničky, ak je krížovka vyplnená a užívateľ znovu stlačí tlačidlo Show zobrazí sa vyskakovacie okno s upozornením

Obr. 7.17: Vyplnenie krížovky pomocou tlačidla Show

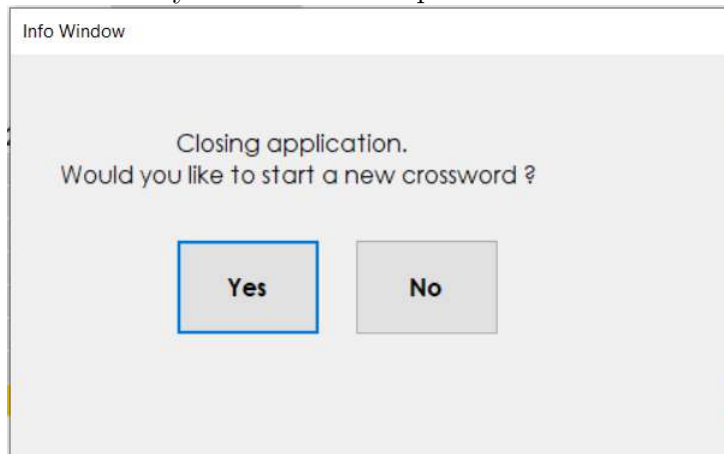


Obr. 7.18: Stlačenie tlačidla Show v čase keď je krížovka vyplnená



- **Restart** - kliknutím sa vyprázdni všetky bunky krížovky
- **End** - kliknutím sa vyvolá vyskakovacie okno, užívateľ má možnosť začať novú krížovku použitím tlačidla Yes, ktoré presunie aplikáciu do stavu výberu jazyka, alebo aplikáciu ukončiť pomocou tlačidla No

Obr. 7.19: Vyskakovacie okno pri stlačení tlačidla End



V tejto kapitole sme si detailne popísali aplikáciu z pohľadu užívateľa. Predstavili sme si jednotlivé stavy aplikácie, procesy, ktoré užívateľ spúšťa interakciou s aplikáciou a ukázali sme si funkcie tlačidiel ako aj ošetrovanie chybových hlásení.

Podľa tohto popisu vidíme, že sú všetky možné akcie užívateľa dobre ošetrené a nemôže sa stať, že by sa užívateľ nevedel pohnúť ďalej.

## 8. Odozva užívateľov

Prvú výslednú verziu aplikácie sme testovali na vzorke 22 užívateľov, z ktorých zhruba polovica sa venuje informatickým odborom a zhruba polovica je iného, nie informatického zamerania. Pripravili sme dotazník s uzavretými otázkami a jednou otvorenou otázkou, ktorý mali užívatelia po vyskúšaní aplikácie vyplniť. V tejto kapitole si predstavíme vyhodnotenie dotazníka, podnety a návrhy respondentov a celkovú spokojnosť užívateľov s aplikáciou.

### 8.1 Číselné vyhodnotenie

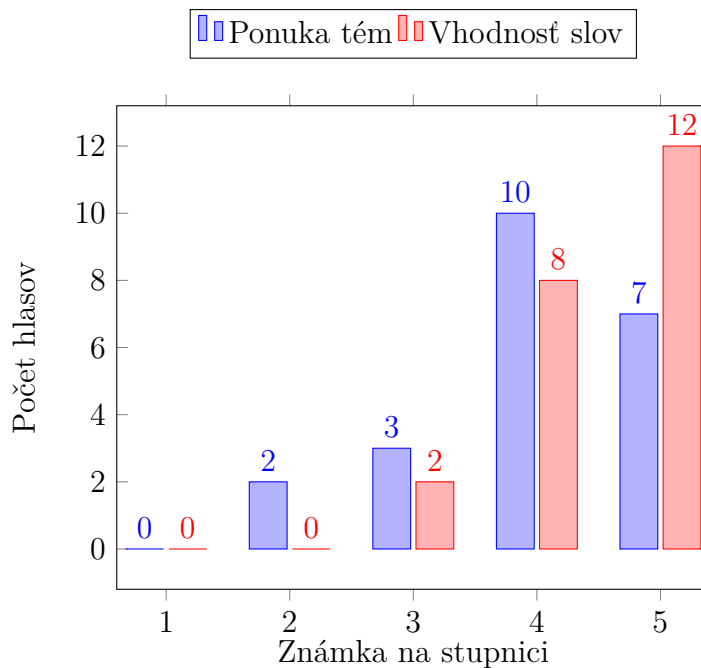
V dotazníku k aplikácii sme sa respondentov pýtali päť otázok:

1. Ako hodnotíte ponuku tém ?
2. Sú podľa Vás slová v krížovke primerané k zvolenej téme ?
3. Ako hodnotíte grafickú úpravu aplikácie ?
4. Bolo pre Vás ovládanie aplikácie náročné ?
5. Je podľa Vás aplikácia užitočná ? Použili by ste ju pri svojom štúdiu ?

Pri každej otázke bola k dispozícii hodnotiaci číselná stupnica od jeden do päť, kde jednotka predstavovala najhoršie možné hodnotenie a päťka najlepšie možné hodnotenie. Okrem uzavretých otázok bola k dispozícii jedna otvorená otázka s dlhou odpoveďou, kde mohli respondenti vyjadriť svoje postrehy a návrhy k aplikácii.

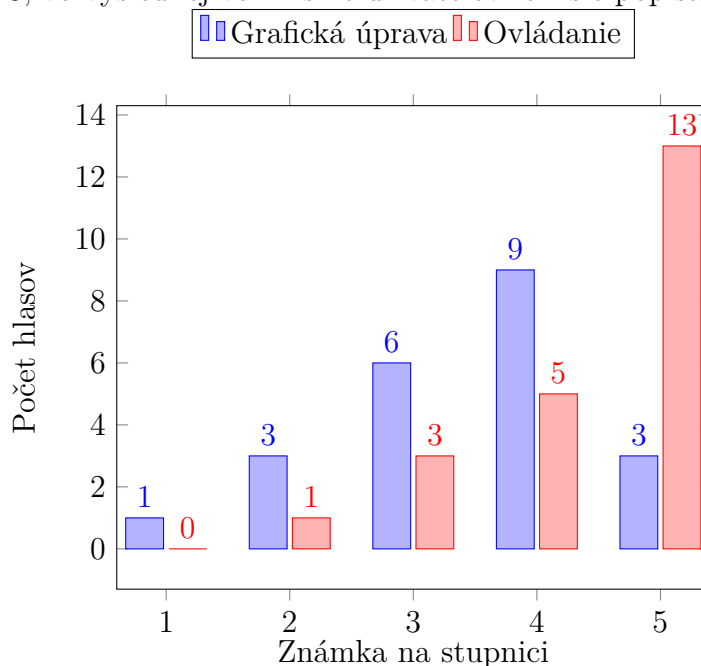
Respondenti hodnotili ponuku tém ako širokú až veľmi širokú. Priemerná a zároveň najpočetnejšia známka v prvej otázke bola 4 s celkovým počtom hlasov 10 (45,5%). Keďže v anglickom jazyku je k dispozícii 31 tém (plus náhodné slová a možnosť definovať si vlastnú tému), takýto výsledok bol očakávaný a považujeme ho za adekvátny.

Slová v krížovke boli hodnotené ako úplne vhodné vzhľadom k zvolenej téme, alebo vhodné. Priemerná známka v druhej otázke bola 4 s celkovým počtom hlasov 8 (36,4%) a najpočetnejšia známka bola 5 s celkovým počtom hlasov 12 (54,5%). Medzi slovami v krížovke sa často vyskytujú korene, teda všeobecné pomenovania základných kategórií slov, preto sme ako najpočetnejšiu známku očakávali 4. Výborné hodnotenie vhodnosti slov nás príjemne prekvapilo.



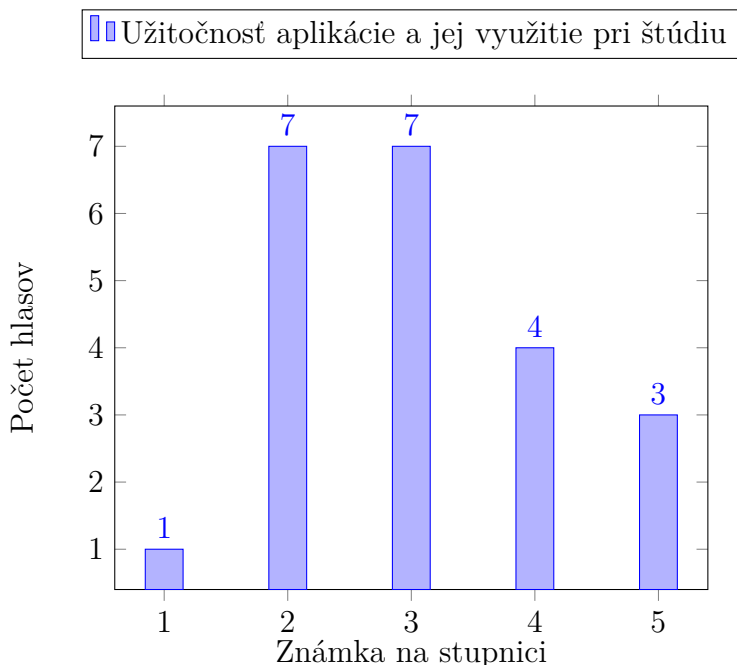
Grafická úprava bola hodnotená priemernou známkou 3 s celkovým počtom hlasov 6 (27,3%) a najpočetnejšou známkou bola 4 s celkovým počtom hlasov 9 (40,9%). Nízke známky sme dostávali najmä od študentov inžinierskych odborov, ktorí sa venujú počítačovej grafike a očakávali lepšiu a viac responzívnu grafickú úpravu. Respondenti, ktorí sa nevenujú inžinierskym odborom naopak hodnotili grafickú úpravu ako veľmi dobrú až výbornú. Napriek pomerne dobrému hodnoteniu sme vo výslednej aplikácii grafickú úpravu mierne vylepšili.

Ovládanie aplikácie dostalo výborné hodnotenie. Priemernou známkou bola 4 s celkovým počtom hlasov 5 (22,7%) ale najpočetnejšou známkou s celkovo 13 hlasmi (59,1%) bola známka 5. Ovládanie považujeme za intuitívne a veľmi jednoduché, takýto výsledok sme očakávali. Keďže sa objavili hlasy aj pri známkach 2 a 3, vo výslednej verzii sme užívateľovi bližšie popísali spôsob ovládania.



Poslednou uzavretou otázkou bola užitočnosť aplikácie, ktorá dostala prie-

mernú známku 3 s celkovým počtom hlasov 7 (31,8%). Priemerná známka je spolu so známkou 2 s rovnakým počtom hlasov najpočetnejšou odpoveďou. Tento výsledok nás zaskočil, nakoľko sme očakávali vyššie hodnotenia. Respondenti svoje nízke hodnotenie zdôvodnili v otvorenej otázke, a to tým, že by aplikáciu využili väčšmi na relax ako na štúdium.



Celkovo dostala aplikácia od užívateľov dobré až veľmi dobré hodnotenia. Testovanie aplikácie považujeme za úspešné. Nebolo potrebné vykonávať výrazné zmeny v aplikácii.

## 8.2 Podnety a návrhy

V otvorenej otázke mali respondenti možnosť vyjadriť ďalšie poznámky a požrehy k aplikácii, prípadne aké chyby v aplikácii našli.

Najviac pripomienok prichádzalo v súvislosti s jazykom slovenčina. Pri výbere jazyka bolo v informatívnom texte uvedené, že sa jedná o jazyk slov dopĺňaných do krížovky. Napriek tomu, boli užívatelia zmätení. Väčšina užívateľov si tento text neprečítala do konca, alebo význam textu zle pochopili. Viac ako 45% užívateľov očakávalo, že celá aplikácia bude od tohto kroku v slovenskom jazyku. To zahŕňa všetky informatívne texty, text tlačidiel ale aj nápovedy k slovám v krížovke. Vo výslednej verzii sme tento informatívny text upravili. Jasne stanovujeme, že si užívateľ vyberá jazyk, v ktorom bude dopĺňať slová do krížovky. Následne oznamujeme, že celá aplikácia vrátane legendy bude bez ohľadu na voľbu jazyka v angličtine.

Druhou najčastejšou pripomienkou bolo zobrazenie legendy. Niektoré definície sú dlhšie ako priestor, v ktorom sa zobrazujú, preto je nutné používať posuvník, pomocou ktorého si užívateľ postupne prečíta celú definíciu. 27% užívateľov opisovalo používanie posuvníka ako rušivé a otravné. Vo výslednej verzii text definícií zalamujeme tak, aby bola v priestore pre legendu zobrazená celá definícia.



## 8.3 Možné vylepšenia

V odpovediach na otvorenú otázku sa objavili dva ďalšie návrhy. Traja z opýtaných respondentov by prijali rôzne obtiažnosti krížoviek. Dvaja z respondentov by prijali možnosť napísať do prvej bunky stĺpca celé slovo, ktoré sa automaticky rozdelí do zvyšných buniek. Tieto návrhy boli ojedinelé, preto si ich nechávame ako možné vylepšenia pre ďalšie verzie.

Niektorí z opýtaných užívateľov si nemohli aplikáciu vyskúšať, pretože majú zariadenia od spoločnosti Apple s operačným systémom macOS, alebo používajú operačný systém Linux. V budúcnosti by sme chceli rozšíriť funkčnosť aplikácie aj na tieto platformy.

Keďže nadpolovičná väčšina pripomienok sa týkala slovenského jazyka, uvažovali sme nad použitím automatických prekladačov aby sme doplnili databázu slovenských slov a preložili príslušné definície. Po zvolení slovenského jazyka užívateľom by celá aplikácia vrátane všetkých súčastí mohla pokračovať v tomto jazyku. Existujú rôzne možnosti použitia prekladačov, avšak platené. Automatický preklad sme preto nevyužili.

V tejto kapitole sme vyhodnotili odozvu užívateľov na výslednú aplikáciu. Vyčíslili sme výsledky dotazníka a predstavili sme návrhy a podnety užívateľov. Výsledky sme vizualizovali pomocou grafov. Aplikáciu vnímali užívatelia pozitívne. Niektorí z respondentov sa nám spätne ozvali s otázkou, či si môžu aplikáciu ponechať a naďalej ju používať pre osobné účely. Podľa užívateľského hodnotenia považujeme aplikáciu za úspešnú, a veríme, že by si v spoločnosti našla množstvo priaznivcov.

# Záver

Táto práca rozsiahlym spôsobom popisuje vytvorenie aplikácie pre automatické generovanie tematických hrebeňoviek.

K vytvoreniu aplikácie nás motivoval najmä nedostatočný priestor na precvičovanie slovných zásob počas učenia sa cudzích jazykov. Spomedzi známych prístupov k riešeniu úlohy sme si zvolili získavanie slov do krížoviek z voľne prístupných databáz a formovanie špeciálneho druhu krížoviek, hrebeňoviek. Ako stavebný prvok aplikácie slúžia databázy slovenského a anglického WordNetu. Vyvinuli sme dva programy, **WordSourceMaker** pre spracovanie databáz a **ThematicCrosswords** ako výslednú aplikáciu s grafickým prostredím. V práci podrobne popisujeme logickú funkciu týchto programov a ich metód, ako aj výsledné programové prevedenie. V posledných kapitolách opisujeme aplikáciu z pohľadu užívateľov a vyhodnocujeme spätnú väzbu od malej vzorky reálnych užívateľov.

Cieľ práce považujeme za splnený. Vyvinuli sme plne funkčnú aplikáciu, ktorá prijíma od užívateľa tému krížovky a na základe tejto témy pomocou prehľadávania databázy skrz hyponymá a hyperonymá získa slovník tematických slov. Z tohto slovníka následne vytvorí krížovku, ktorá obsahuje slová týkajúce sa zadanej témy a vyformuje príslušnú legendu. Užívateľ vidí graficky príjemné prostredie, s ktorým interaguje pomocou tlačidiel, a ktoré mu umožňuje pohodlne vyplňať krížovku. Oproti pôvodnému zadaniu sme výslednú aplikáciu rozšírili o slovenský jazyk, teda užívateľ si môže vybrať spomedzi dvoch jazykov slov dopĺňaných do krížovky (anglický a slovenský jazyk). Užívateľovi navyše dávame možnosť definovať si vlastnú tému, pokiaľ je preňho ponuka tém nedostatočná.

Aplikácia bola počas testovania užívateľmi hodnotená pozitívne. Preto veríme, že by si pri uvedení do prevádzky pre širokú verejnosť našla množstvo priaznivcov.

Ako každý vyvíjaný projekt, ani naša aplikácia nie je dokonalá. V súčasnosti aplikácia funguje len na systémoch Windows. Jedným zo zásadných zlepšení, ktoré by sme chceli v budúcnosti implementovať je multiplatformovosť.

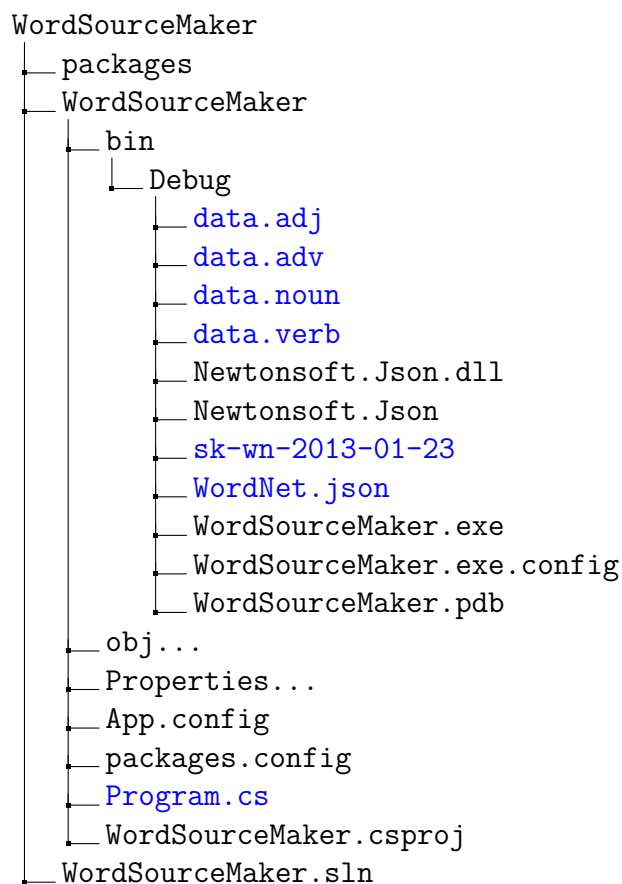
# Zoznam použitej literatúry

- DEVITT, A. a VOGEL, C. (2004). The Topology of WordNet: Some Metrics. *Computational Linguistics Group, Trinity College, Dublin, Ireland.*
- GARABÍK, R. (2010). Slovak national corpus tools and resources. In *Proceedings of the 5th Workshop on Intelligent and Knowledge oriented Technologies (WIKT 2010)*, pages 2–7.
- MILLER, G. A., BECKWITH, R., FELLBAUM, C., GROSS, D. a MILLER, K. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography.*
- NEWTON-KING, J. (2021). Newtonsoft Json.NET. *Popular high-performance JSON framework for .NET.* URL <https://www.newtonsoft.com/json>.
- PRINCETON UNIVERSITY (2021). WordNet. *A Lexical Database for English, Princeton University.* URL <https://wordnet.princeton.edu/>.
- RIGUTINI, L., DILIGENTI, M., MAGGINI, M. a GORI, M. (2008). A Fully Automatic Crossword Generator. In *Seventh International Conference on Machine Learning and Applications*, pages 326–367.
- ŠIMKOVÁ, M. a GARABÍK, R. (2014). Slovenský národný korpus (2002–2012): východiská, ciele a výsledky pre výskum a prax. *Jazykovedné štúdie XXXI. Rozvoj jazykových technológií a zdrojov na Slovensku a vo svete (10 rokov Slovenského národného korpusu).* Ed. Katarína Gajdošová—Adriána Žáková. Bratislava: VEDA, pages 35–64.
- SLOVENSKÝ NÁRODNÝ KORPUS (2021). Slovenský WordNet. *Jazykovedný ústav E. Štúra Slovenskej akadémie vied.* URL <https://korpus.sk/WordNet.html>.
- SOJKA, P., PALA, K., SMRŽ, P., FELLBAUM, C. a VOSSEN, P. (2004). Second International WordNet Conference. In *Brno, Czech Republic.*
- TAUSIG, B. (2013). *The Curious History of the Crossword: 100 Puzzles from Then and Now.* Race Point Publishing.
- WIKIPEDIA (2021a). JSON. *The Free Encyclopedia.* URL <https://en.wikipedia.org/wiki/JSON>.
- WIKIPEDIA (2021b). WordNet. *The Free Encyclopedia.* URL <https://en.wikipedia.org/wiki/WordNet>.
- WYNNE, A. (1913). FUN's Word-Cross Puzzle. *Prevzaté z webového článku.* URL <https://medium.com/fgd1-the-archive/crossword-puzzle-1913-2a1da960bbaa>.

# A. Prílohy

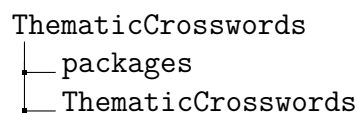
## A.1 WordSourceMaker

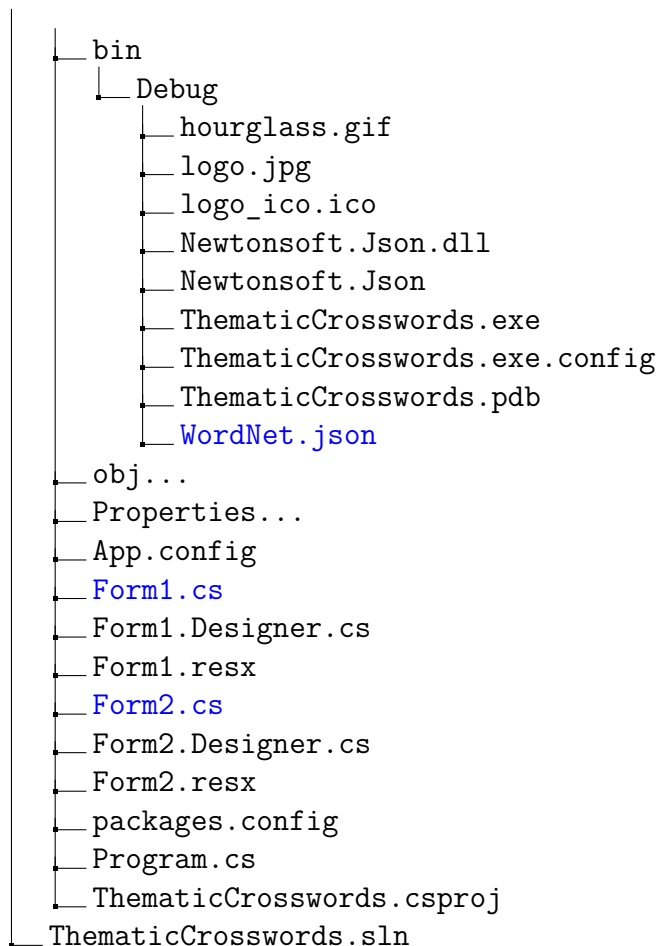
Priečink WordSourceMaker obsahuje program pre spracovanie a serializáciu slovenskej a anglickej databázy WordNet. Dôležitými súbormi sú zdrojový kód programu (`Program.cs`), zdrojové dátové súbory databáz WordNet (`data.adj`, `data.adv`, `data.noun`, `data.verb`, `sk-wn-2013-01-23`) a výstupný súbor formátu JSON reprezentujúci zjednotenie oboch databáz (`WordNet.json`). Nasledujúci graf reprezentuje štruktúru súborov priečinku, s cestou k spomínaným súborom.



## A.2 ThematicCrosswords

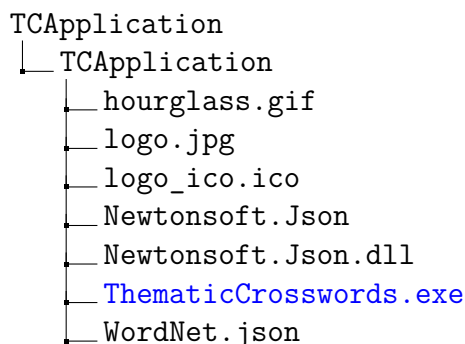
Priečink ThematicCrosswords obsahuje program pre automatické generovanie tematických hrebeňoviek s grafickým prostredím. Dôležitými súbormi sú zdrojový kód programu (`Form1.cs` a `Form2.cs`) a súbor, ktorý je hlavným zdrojom dát (`WordNet.json`). Nasledujúci graf reprezentuje štruktúru súborov priečinku, s cestou k spomínaným súborom.





### A.3 TCApplication

Komprimovaný priečinok TCApplication obsahuje výslednú spustiteľnú aplikáciu pre automatické generovanie tematických hrebeňoviek, a súbory, ktoré sú potrebné pre správne fungovanie aplikácie. V priečinku je najdôležitejším súborom ThematicCrosswords.exe, čo je samotná aplikácia. Pre spustenie aplikácie stačí dvojkliknúť na tento súbor. Nasledujúci graf reprezentuje štruktúru súborov priečinku, s cestou k súboru aplikácie.



## A.4 Licencie

Textový súbor obsahujúci licencie softvérových a dátových produktov použitých v práci (Princeton WordNet, Slovenský WordNet, Newtonsoft.json). Licencie povolujú použitie týchto produktov v rozsahu našej práce.