

Št. naloge: 01524/2008

Datum: 15.10.2008

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **TOMAŽ KUZMA**

Naslov: **HEVRISTIČNO SESTAVLJANJE KRIŽANK**
HEURISTIC CROSSWORD PUZZLE GENERATION

Vrsta naloge: Diplomsko delo univerzitetnega študija

Tematika naloge:

Sestavljanje križank je NP-poln problem, ki se ga lotevamo s hevrističnim preiskovanjem. Za praktično uporaben program, ki bi sestavljal križanke namesto človeka, je potrebno upoštevati jezikovne posebnosti, primeren nabor besed za ciljno publiko in zagotoviti, da se besede ne pojavljajo prepogosto.

V nalogi pripravite pregled hevrističnih preiskovalnih metod za sestavljanje križank in prosto dostopnih rešitev v angleškem jeziku. Na tej podlagi predlagajte prilagoditve potrebne za slovenščino, načrtujte in analizirajte podatkovne strukture in preiskovalni algoritem, zasnujete ustrezno bazo podatkov in izdelajte prototip aplikacije, ki bo omogočala dolgoročno praktično uporabo tako na spletu kot v tiskanih medijih.

Mentor:

doc. dr. Marko Robnik Šikonja

Dekan:

prof. dr. Franc Solina



Št. naloge: 01524/2008

Datum: 15.10.2008

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **TOMAŽ KUZMA**

Naslov: **HEVRISTIČNO SESTAVLJANJE KRIŽANK**
HEURISTIC CROSSWORD PUZZLE GENERATION

Vrsta naloge: Diplomsko delo univerzitetnega študija

Tematika naloge:

Sestavljanje križank je NP-poln problem, ki se ga lotevamo s hevrističnim preiskovanjem. Za praktično uporaben program, ki bi sestavljal križanke namesto človeka, je potrebno upoštevati jezikovne posebnosti, primeren nabor besed za ciljno publiko in zagotoviti, da se besede ne pojavljajo prepogosto.

V nalogi pripravite pregled hevrističnih preiskovalnih metod za sestavljanje križank in prosto dostopnih rešitev v angleškem jeziku. Na tej podlagi predlagajte prilagoditve potrebne za slovenščino, načrtujte in analizirajte podatkovne strukture in preiskovalni algoritem, zasnujete ustrezno bazo podatkov in izdelajte prototip aplikacije, ki bo omogočala dolgoročno praktično uporabo tako na spletu kot v tiskanih medijih.

Mentor:

doc. dr. Marko Robnik Šikonja

Dekan:

prof. dr. Franc Solina

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Tomaž Kuzma

**HEVRISTIČNO SESTAVLJANJE
KRIŽANK**

DIPLOMSKO DELO
NA UNIVERZITETNEM ŠTUDIJU

Mentor: doc. dr. Marko Robnik Šikonja

Ljubljana, 2009

Rezultati diplomskega dela so intelektualna lastnina Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje Fakultete za računalništvo in informatiko ter mentorja.

Namesto te strani **vstavite** original izdane teme diplomskega dela s podpisom mentorja in dekana ter žigom fakultete, ki ga diplomant dvigne v študentskem referatu, preden odda izdelek v vezavo!

Zahvala

Zahvaljujem se svojemu mentorju doc. dr. Marku Robniku Šikonji za pomoč in nasvete pri izdelavi diplomske naloge, staršem in dekletu pa za vso podporo v času študija.

Kazalo

| | |
|--|-----------|
| Povzetek | 1 |
| 1 Uvod | 4 |
| 1.1 Predstavitev problema | 4 |
| 1.2 Motivacija | 4 |
| 1.3 Različni pristopi pri sestavljanju križank | 5 |
| 1.4 Moj pristop | 6 |
| 1.5 Napoved vsebine | 7 |
| 2 Splošni opis križank | 8 |
| 2.1 Tipične vrste križank | 10 |
| 2.1.1 Križanka iz sredine in skrita križanka | 10 |
| 2.1.2 Italijanska in koordinatna križanka | 11 |
| 2.1.3 Slikovna in skandinavsko križanka | 12 |
| 2.1.4 Verižna križanka | 13 |
| 2.1.5 Kombinacijska križanka | 14 |
| 2.1.6 Zlogovna križanka | 15 |
| 2.2 Opis programa za stavljenje križank - Sfinga | 16 |
| 3 Opis problema | 18 |
| 3.1 Hevristično preiskovanje | 18 |
| 3.2 Splošne zahteve za slikovne križanke | 19 |
| 3.3 Možne dodatne zahteve | 20 |
| 4 Baza gesel | 21 |
| 4.1 Pridobitev baze | 21 |
| 4.2 Analiza baze gesel | 22 |
| 4.3 Indeksiranje baze gesel | 22 |
| 4.3.1 Računalniški pomnilnik ali SQL Server | 22 |
| 4.3.2 Indeksiranje | 23 |

| | | |
|----------|---|-----------|
| 4.3.3 | Iskanje besed | 24 |
| 4.4 | Velikost baze gesel | 25 |
| 5 | Potek generiranja križank | 30 |
| 5.1 | Predpriprava | 30 |
| 5.2 | Povezanost besed | 31 |
| 5.3 | Korak generiranja - vstavljanje besede | 31 |
| 5.3.1 | Nemogoči opisni kvadrati | 32 |
| 5.3.2 | Izbira položaja za vstavljanje besede | 32 |
| 5.3.3 | Izbira besede | 34 |
| 5.4 | Menjava vstavljene besede - vračanje en korak nazaj | 34 |
| 5.5 | Večnivojsko vračanje | 35 |
| 6 | Analiza rezultatov | 37 |
| 6.1 | Testiranje sestavljanja križank | 37 |
| 6.2 | Obrazložitev rezultatov | 41 |
| 7 | Sklep | 43 |
| 7.1 | Povzetek | 43 |
| 7.2 | Možnosti izboljšav | 43 |
| | Dodatki | 45 |
| | A Predloge križank za testiranje algoritma | 45 |
| | Seznam slik | 50 |
| | Seznam slik | 50 |
| | Seznam tabel | 51 |
| | Literatura | 52 |

Povzetek

V diplomski nalogi sem predstavil različne pristope za sestavljanje križank in podrobno opisal svoj pristop. Na splošno sem opisal križanke in predstavil nekaj tipičnih vrst križank, ki jih je mogoče sestaviti z mojim algoritmom. Opisal sem optimiziran način indeksiranja baze gesel glede na iskanja po geslih, ki so potrebna pri sestavljanju križank. Pri algoritmu za sestavljanje križank sem uporabil nov pristop dodajanja črnih polj med generiranjem križanke, in ne na začetku dodajanje vseh polj kot so uporabili ostali raziskovalci. Uporabil sem hevristično preiskovanje prostora. Za izbiro položaja vstavljene besede sem uporabil kriterij minimalnega števila ustreznih besed. Uporabil sem tudi večnivojsko vračanje, ki je nujno za rešitev problema sestavljanja križank, saj omogoča hitro zamenjavo besede, ki je bila vstavljena več nivojev nazaj. Na koncu sem svoj algoritem preizkusil na nekaj tipičnih primerih križank in rezultate analiziral.

Ključne besede:

križanka, hevristično preiskovanje, baza gesel, slikovna križanka, večnivojsko vračanje

Abstract

I present various approaches for generation crosswords and my own approach. I describe the crosswords puzzles and a few types of crosswords that can be generated with my algorithm. I optimize indexing database in order that the searches, which are necessary in construction crosswords, are as fast as possible. In the generation I used a novel approach, where I automatically adding black boxes during construction crossword, while existing approaches require their definition in advance. I use heuristic search of space. For the position in selection of the inserted words positions and I use a criterion of the minimum number of relevant words. I also use multi level back-tracking, which is necessary because it allows quick replacement of the word, which was inserted several levels back. I test the algorithm on some typical examples and analyze the results.

Key words:

crosswords, heuristic search, word database, multilevel back-tracking

Poglavje 1

Uvod

1.1 Predstavitev problema

Sestavljanje križank je NP-poln problem, ki ga rešujemo s hevrističnim preiskovanjem. NP-polnost pomeni, da za tak algoritem obstaja nedeterminističen algoritem, ki reši problem v polinomskem času, medtem ko determinističnega algoritma najverjetneje ni.

Pri sestavljanju križank je potrebno v mrežo polj določenih dimenzij vnesti besede. Besede potekajo navpično (od zgoraj navzdol) in vodoravno (od leve proti desni). Črke vodoravnih in navpičnih besed, ki se križajo, se morajo ujemati. Pri večini križank je na koncu besede še tako imenovan črn kvadrateg, ki ločuje dve besedi in v njem ni črke. Včasih je zaradi lažje sestave križanke dovoljeno še omejeno število dodatnih črnih kvadratkov, ki niso na začetku oz. koncu ene izmed besed v križanki.

Za sestavljanje kvalitetnih križank je potrebno zagotoviti dovolj veliko bazo besed, ki se v križankah imenujejo gesla. Brez dovolj velike baze oz. če lahko uporabimo le gesla iz nekega področja, je potrebno omiliti zahteve pri sestavljanju križanke. Kot zelo primerne besede pri sestavljanju križank se izkažejo besede iz ugankarskih slovarjev, ki vsebujejo veliko število raznolikih besed krajših dolžin.

1.2 Motivacija

V podjetju, v katerem delam, smo se lotili izdelave programa za stavljenje križank - Sfinga, ki omogoča izdelavo poljubnih križank z vsemi pripadajočimi elementi (slike, liki, razne črte, ...). Omogoča tudi izvoz v različne formate

za tisk in objavo na spletu. Program že omogoča avtomatizacijo nekaterih procesov pri izdelavi križanke, zato smo se v sodelovanju z naročnikom odločili še za avtomatizacijo najbolj zahtevnega dela pri izdelavi križanke, vnosa gesel v križanko. To je še posebej pomembno pri izdelavi križank za splet, kjer je potrebno veliko križank in kratek čas generiranja.

Vse slike križank v tej diplomski sem izdelal z omenjenim programom.

1.3 Različni pristopi pri sestavljanju križank

Na temo sestavljanja križank je v strokovni literaturi relativno malo napisanega, kljub temu, da je bila prva križanka narejena že pred skoraj 100 leti (1913). Obstajata dva pristopa:

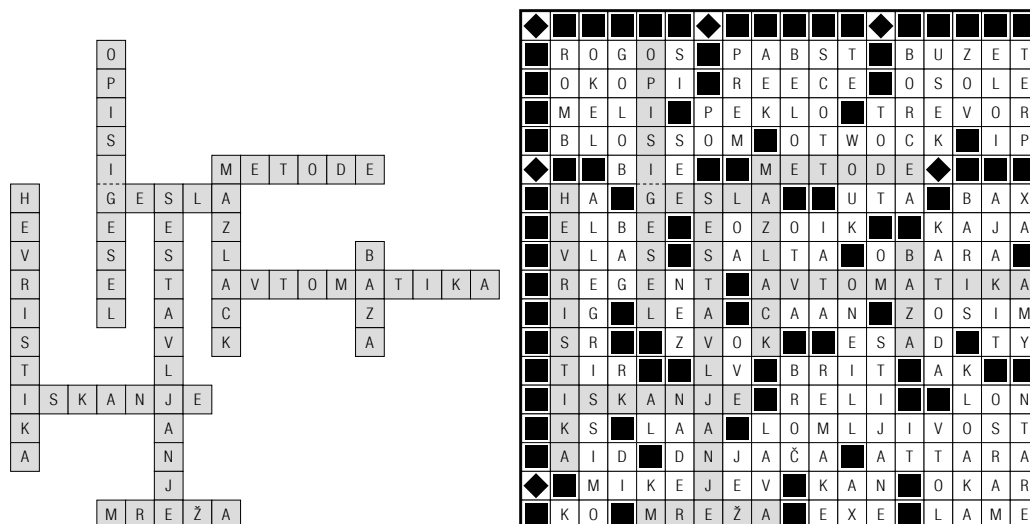
- vstavljanje črke za črko
- vstavljanje beseda za besedo

Prvi pristop je v sedemdesetih letih prejšnjega stoletja uporabljal Mazlack [1]. Problem drugega pristopa je bila premajhna količina računalniškega pomnilnika, kar pa danes ni več težava, saj je lahko celotna baza besed učinkovito indeksirana v spominu. Izpeljanke iz drugega pristopa pa uporablja večina novejših raziskovalcev. Ginsberg [6] uporablja pristop k sestavljanju križank na podlagi deklarativnih povezav. Pri fiksni mreži črnih polj lahko križanko prevedemo na množico deklarativnih stavkov in uporabimo metode za reševanje deklarativnih povezav. Aherne in Vogel [9] sta v svojem članku opisala problem sestavljanja križanke na določeno temo.

Večji del generatorjev križank na internetu je namenjenih izdelavi tako imenovanih mrež. To so križanke, kjer je večina kvadratkov v križanki povezanih samo z geslom v eni smeri. Takšne križanke so primerne za izobraževalne namene [10], ker določiš nekaj gesel, ki so tesno povezana z zelenim področjem in so vsa uporabljena, ni pa potrebna uporaba mašil¹. Sam se bom ukvarjal s sestavljanjem pravih križank, ne mrež. Razlika med obema križankama je predstavljena na sliki 1.1.

Najbolj znan program za sestavo križank je Crossword compiler [7]. Zastonj je možno dobiti preizkusno različico programa. Omogoča sestavljanje

¹Mašila so takšna gesla, ki si jih želimo v križanki čim manj, so pa nujna za popolnitev križanke.



Slika 1.1: Na levi sliki je mreža, na desni pa je mreža dopolnjena v običajno križanko.

velikega števila različnih tipov pravih križank. Program Custom Crossword Generator [8] je dostopen z izvorno kodo, nima izdelanega uporabniškega vmesnika, omogoča pa sestavljanje poljubnih križank po predlogi.

1.4 Moj pristop

Moja rešitev problema temelji na pristopu beseda za besedo, z razliko, da ne uporablja fiksne mreže črnih polj. Pri vseh rešitvah, ki sem jih zasledil v literaturi, so vsi črni in opisni kvadrati določeni že pred generiranjem križanke. Pri moji rešitvi pa lahko na začetku določimo le nekaj nujnih črnih in opisnih kvadratkov, ostali pa se avtomatsko določijo med sestavljanjem križanke, glede na besede in omejitve.

Prednost tega načina je, da posnema človekovo izdelavo križanke, ki je v večini primerov brez vnaprej določenih črnih kvadratkov in omogoča interakcijo programa s sestavljavcem. Program lahko na primer dokonča križanko, ki jo je človek že na pol sestavil. Interaktivno je možno uporabljati programa tako, da sestavljavec usmerja program. Program doda eno ali več besed, sestavljavec pa se odloči o njihovi sprejemljivosti, lahko predlaga tudi svoje. Program pomaga tudi tako, da predlaga možne besede in sproti preverja omejitve pri sestavi križanke.

Pri izdelavi algoritma sem upošteval interaktivnost in ga zasnoval tako, da je vsak postopek izbire in vstavljanja besede en korak algoritma, ki je neodvisen od preteklih korakov. Globalno je potrebno voditi seznam vnesenih besed, ki se uporablja za vračanje.

Grob potek enega koraka mojega algoritma je naslednji:

- v mreži križanke, glede na omejitve, določim seznam polj v katera lahko vstavim besede,
- po določenem kriteriju izberem polje v katerega se bo vstavila beseda,
- za izbrano polje izračunam seznam ustreznih besed glede na že vstavljene črke in omejitve glede dolžine besede,
- po določenem kriteriju izberem eno besedo in jo vstavim v križanko,
- če se v določenem koraku zgodi, da v nekem polju ni ustrezne besede, se vrnem na korak kjer je bila vstavljena zadnja izmed problematičnih besed.

1.5 Napoved vsebine

V naslednjem poglavju bom na splošno opisal križanke in prikazal nekaj primerov križank, ki so primerne za sestavljanje z mojim programom. Opisal bom tudi program za stavljenje križank - Sfinga. V tretjem poglavju bom podrobno predstavil hevrstično preiskovanje prostora in navedel specifične zahteve pri sestavljanju slikovnih križank ter izpostavil možne zahteve, ki se bi lahko pojavile pri uporabi programa v praksi. V četrtem poglavju bom opisal in delno analiziral bazo gesel. Podrobno bom tudi predstavil indeksiranje baze, ki je prilagojeno iskanjem po bazi pri sestavljanju križank. V petem poglavju bom natančno opisal svoj algoritem za sestavljanje križank, še posebej bom izpostavil pojma povezanost besed in večnivojsko vračanje. V šestem poglavju bom analiziral rezultate sestavljanja križank na nekaj tipičnih primerih iz prakse.

Poglavje 2

Splošni opis križank

Križanka je vrsta besedne uganke, ki je zapisana v mreži polj, vodoravna in navpična gesla pa se križajo. Obstaja tudi nekaj vrst ugank, ki so zelo podobne križankam, kjer so gesla zapisana vodoravno, rešitev uganke oz. nek dodatni opis pa ne poteka navpično, vendar to niso križanke. Naj omenim še zlogovne križanke (slika 2.7), kjer se v vsakem polju nahaja celoten zlog namesto posamezne črke.

Sestavni del križanke je razlaga gesla, ki se imenuje opis gesla. Slednji ni pomemben pri sestavljanju, je pa nujen pri reševanju križank.

Ker bo moj algoritem optimiziran za sestavljanje skandinavskih in slikovnih križank (podpoglavje 2.1.3), pri katerih so opisi gesel vneseni v polje v križanki, bom dodatno ločeval med črnimi in opisnimi kvadrati. Opisni kvadrater je tisti iz katerega bo sledilo geslo (desno ali navzdol). Črn kvadrater je tisti, ki ima levo oz. spodaj še en črn ali pa opisni kvadrater in iz njega ni možno potegniti gesla. Razlika je prikazana tudi na sliki 2.1.

| | | | | | | |
|---|------------------------------|------------------------------|------------------------------|---|------------------------------|--|
|  | TO JE OPISNI KVADRATER | TO JE OPISNI KVADRATER | TO JE OPISNI KVADRATER |  | TO JE OPISNI KVADRATER | TO JE OPISNI KVADRATER |
| TO JE OPISNI KVADRATER | | | | TO JE OPISNI KVADRATER | | |
| TO JE OPISNI KVADRATER | | | | | | |
| TO JE OPISNI KVADRATER | | | TO JE OPISNI KVADRATER | | |  |

Slika 2.1: Razlika med črnimi in opisnimi kvadrati.

S stališča sestavljanja križank lahko veliko večino križank spravimo na

skupni imenovalec:

- imamo mrežo polj določenih dimenzij,
- takoj na začetku lahko določena polja oz. skupine polj odstranimo (to so črna polja),
- mrežo polj, ki je še ostala, napolnimo z besedami in morebitnimi dodatnimi črnimi polji in se držimo dodatnih omejitev.

Naj navedem nekatere mogoče dodatne omejitve: število črnih polj, način razporeditve črnih polj, minimalna povprečna dolžna besede, . . . Pri nekaterih križankah ni črnih polj, to pomeni, da si besede sledijo brez črnih polj. Obstajajo tudi križanke, pri katerih morajo biti črna polja razporejena na določen način (npr: simetrično).

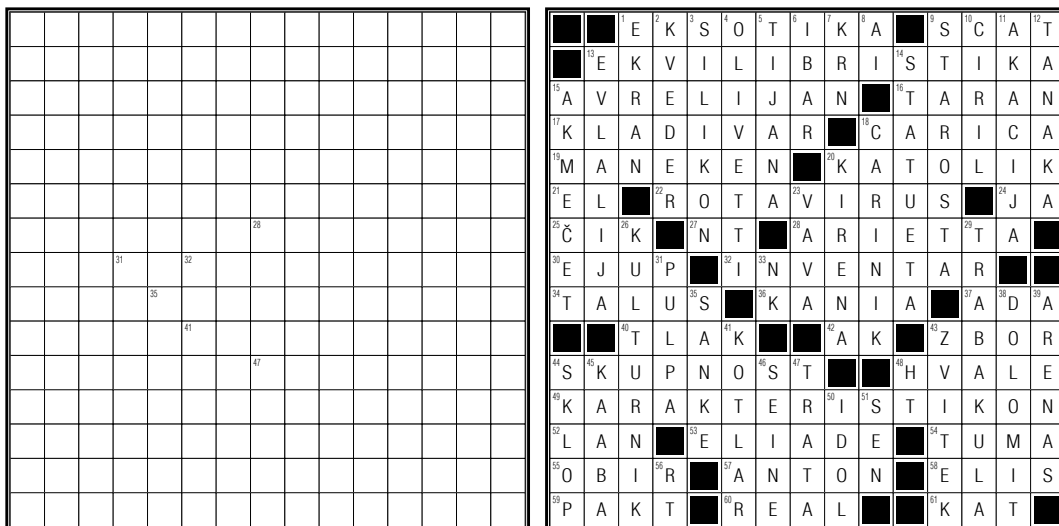
Ko pogledamo sestavljeno križanko s stališča reševanja, lahko isto križanko interpretiramo na več različnih načinov. Vrste križank bom opisal v naslednjem podpoglavju.

2.1 Tipične vrste križank

Uporabljeni primeri križank so iz revije Kih [2]. Opisi križank so delno povzeti iz revije Kih in s spletne strani Enigmatko [4].

2.1.1 Križanka iz sredine in skrita križanka

Pri križanki iz sredine vsebuje seznam oštevilčenih opisov za vodoravna gesla in seznam opisov za navpična gesla. V nerešeni križanki imamo le na sredini nekaj oštevilčenih kvadratkov, kjer se začenjajo besede. Začetke ostalih besed pa morajo reševalci določiti sami, zato se reševanja križanke lotijo iz sredine. Ta vrsta križanke vsebuje tudi črne kvadratke, ki jih reševalci prav tako določijo sami, v navodilih za reševanje je navedeno le skupno število črnih. Skrita križanka je zelo podobna križanki iz sredine, le da ne vsebuje oštevilčenih kvadratov na sredini. Prva beseda se pri skriti križanki nujno začne v zgornjem levem kvadratu, kar pa pri križanki iz sredine ne drži.



Slika 2.2: Nerešena in rešena križanka iz sredine.

2.1.2 Italijanska in koordinatna križanka

Italijanska križanka vsebuje seznam opisov in število črnih kvadratkov za vsako vrstico in vsak stolpec posebej. V prvi vrstici in stolpcu je le ena beseda, brez črnih kvadratkov. Koordinatna križanka pa ima za vsako vodoravno besedo koordinato kjer se beseda začne. Za navpične besede pa ima le seznam opisov, in mora reševalec sam določiti v kateri stolpec bo besedo vstavil. Omejitev glede ene besede v prvem stolpcu oz. vrstici tudi odpade. Koordinatna križanka vsebuje izbrano število črnih polj, ki jih mora reševalec sam razporediti.

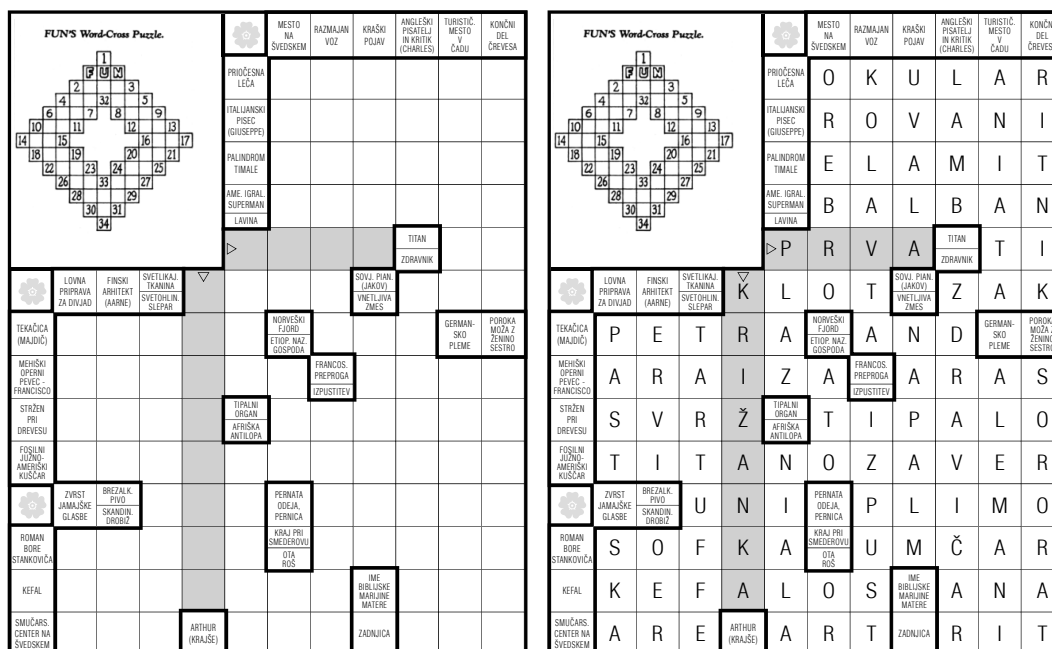
| ○ | A | B | C | Č | D | E | F | G | H | I | J |
|----|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |
| 4 | | | | | | | | | | | |
| 5 | | | | | | | | | | | |
| 6 | | | | | | | | | | | |
| 7 | | | | | | | | | | | |
| 8 | | | | | | | | | | | |
| 9 | | | | | | | | | | | |
| 10 | | | | | | | | | | | |

| ○ | A | B | C | Č | D | E | F | G | H | I | J |
|----|---|---|---|---|---|---|---|---|---|---|---|
| 1 | G | R | A | M | ■ | R | A | Š | I | C | A |
| 2 | R | A | D | I | K | A | L | I | Z | E | M |
| 3 | I | T | E | R | A | T | I | V | ■ | V | P |
| 4 | V | A | N | I | N | I | ■ | A | N | K | E |
| 5 | A | R | D | A | ■ | Š | B | ■ | A | A | R |
| 6 | ■ | ■ | ■ | M | A | Č | I | A | R | ■ | ■ |
| 7 | Š | E | R | E | M | E | T | J | E | V | O |
| 8 | E | R | A | T | O | ■ | O | V | Č | A | R |
| 9 | S | R | Š | E | N | ■ | L | A | J | T | A |
| 10 | T | I | A | R | A | ■ | A | R | E | A | L |

Slika 2.3: Nerešena in rešena koordinatna križanka.

2.1.3 Slikovna in skandinavska križanka

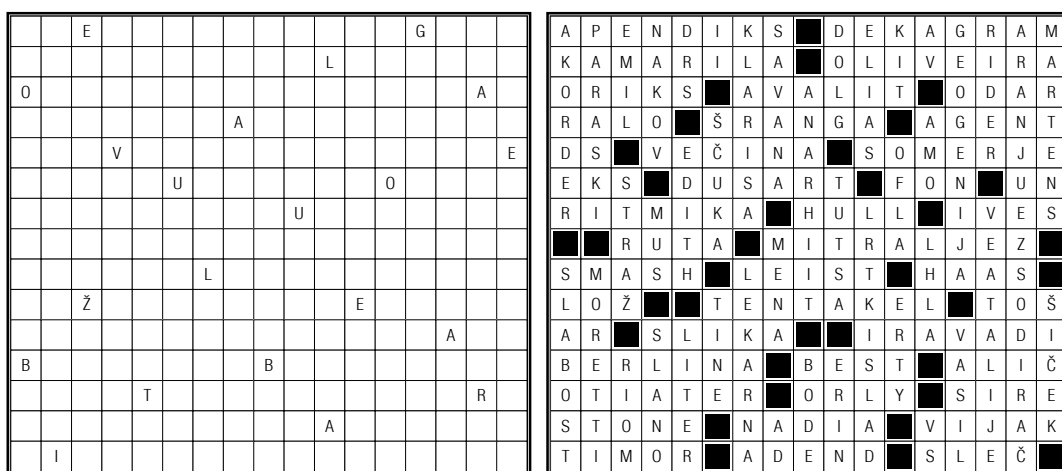
Pri slikovni in skandinavski križanki so opisi gesel napisani v opisnih kvadratih v sami križanki. Če sta pri katerem kvadratu možni dve smeri za geslo, je opisni kvadrat deljen na zgornji in spodnji opis. Pri slikovni križanki so lahko skupine črnih kvadratov predstavljene kot slike. Ravno prisotnost slike v križanki pa je razlika med slikovno in skandinavsko križanko.



Slika 2.4: Nerešena in rešena slikovna križanka.

2.1.4 Verižna križanka

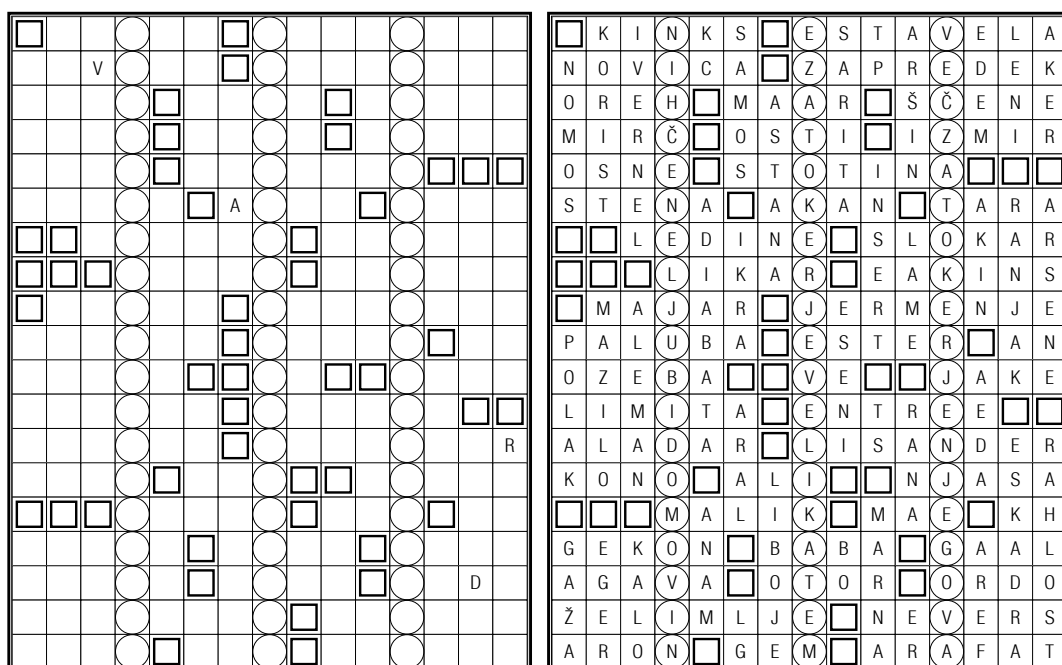
Pri verižni križanki se opisi verižijo. Vodoravni se verižijo po vrsticah od zgoraj navzdol, navpični pa po stolpcih od leve proti desni (navadno si navpični opisi sledijo po vrsticah od leve proti desni, ne po stolpcih). V mrežo križanke je potrebno razporediti tudi določeno število črnih kvadratkov. Za pomoč je običajno v križanko že vpisanih nekaj črk.



Slika 2.5: Nerešena in rešena verižna križanka.

2.1.5 Kombinacijska križanka

Kombinacijska križanka nima opisov, vsebuje le seznam gesel razporejen po številu črk za lažje kombiniranje. Vsa črna polja so že označena. Gesla je potrebno s pravilnim kombiniranjem razvrstiti v lik križanke. Za pomoč je včasih v križanko že vpisanih nekaj črk. Običajno se v posebno označenih poljih pojavlja zanimiva misel.



Slika 2.6: Nerešena in rešena kombinacijska križanka.

2.1.6 Zlogovna križanka

Za zlogovno križanko je značilno, da polja zapolnijo zlogi namesto črk. Vsa črna polja so določena vnaprej. Podobno kot pri mreži ni nujno, da potekata skozi vsako polje dve besedi. To vrsto križanke sem navedel kot primer križanke, ki je ne bom sestavljal z mojim programom, ker ne obstaja enostavna prilagoditev programa za eno-črkovne križanke na zlogovno križanko.

| | | | | | | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|----|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | 1 | 2 | <input type="checkbox"/> | 3 | 4 | <input type="checkbox"/> | 5 | 6 | <input type="checkbox"/> | | |
| 7 | | | <input type="checkbox"/> | 8 | | <input type="checkbox"/> | 9 | | | <input type="checkbox"/> | 10 |
| | <input type="checkbox"/> | 11 | 12 | | <input type="checkbox"/> | 13 | 14 | | <input type="checkbox"/> | 15 | 16 |
| 17 | 18 | <input type="checkbox"/> | 19 | | 20 | <input type="checkbox"/> | 21 | | 22 | <input type="checkbox"/> | 23 |
| <input type="checkbox"/> | 24 | 25 | | <input type="checkbox"/> | 26 | 27 | | <input type="checkbox"/> | 28 | 29 | <input type="checkbox"/> |
| 30 | | | <input type="checkbox"/> | 31 | | <input type="checkbox"/> | 32 | | | <input type="checkbox"/> | 33 |
| | <input type="checkbox"/> | 34 | 35 | | <input type="checkbox"/> | 36 | 37 | | <input type="checkbox"/> | 38 | 39 |
| 40 | 41 | <input type="checkbox"/> | 42 | | 43 | <input type="checkbox"/> | 44 | | 45 | <input type="checkbox"/> | 46 |
| <input type="checkbox"/> | 47 | | | <input type="checkbox"/> | 48 | | | <input type="checkbox"/> | 49 | | <input type="checkbox"/> |

| | | | | | | | | | | | | | | | | | | | |
|--------------------------|-----|--------------------------|-----|--------------------------|--------------------------|--------------------------|-----|--------------------------|--------------------------|--------------------------|------|--------------------------|------|--------------------------|--------------------------|--------------------------|------|--------------------------|-----|
| <input type="checkbox"/> | 1 | PRE | 2 | BEŽ | NIK | <input type="checkbox"/> | 3 | GRA | 4 | VID | NOST | <input type="checkbox"/> | 5 | PO | 6 | GRI | NJEK | <input type="checkbox"/> | |
| 7 | KAN | DIN | SKI | <input type="checkbox"/> | 8 | AL | HAM | BRA | <input type="checkbox"/> | 9 | SA | LA | MA | <input type="checkbox"/> | 10 | PE | | | |
| | TAV | <input type="checkbox"/> | 11 | DI | 12 | O | LEN | <input type="checkbox"/> | 13 | TO | 14 | SKA | NA | <input type="checkbox"/> | 15 | SA | 16 | BO | TER |
| 17 | TOR | 18 | KA | <input type="checkbox"/> | 19 | PRE | DE | 20 | LEK | <input type="checkbox"/> | 21 | LI | DER | 22 | KA | <input type="checkbox"/> | 23 | ŠKA | LE |
| <input type="checkbox"/> | 24 | O | 25 | RA | KELJ | <input type="checkbox"/> | 26 | SI | 27 | VI | NA | <input type="checkbox"/> | 28 | TE | 29 | ME | RIN | <input type="checkbox"/> | |
| 30 | BE | LIN | DA | <input type="checkbox"/> | 31 | KLJU | KA | NJE | <input type="checkbox"/> | 32 | KON | DRA | ŠIN | <input type="checkbox"/> | 33 | KRI | | | |
| | NET | <input type="checkbox"/> | 34 | MES | 35 | SI | NA | <input type="checkbox"/> | 36 | TA | 37 | TAR | STAN | <input type="checkbox"/> | 38 | KA | 39 | LU | ŽA |
| 40 | KE | 41 | FIR | <input type="checkbox"/> | 42 | PAL | ČI | 43 | CA | <input type="checkbox"/> | 44 | TA | TA | 45 | MI | <input type="checkbox"/> | 46 | PI | VEC |
| <input type="checkbox"/> | 47 | MA | LE | ČNIK | <input type="checkbox"/> | 48 | PIN | DA | ROS | <input type="checkbox"/> | 49 | STER | ŽI | NAR | <input type="checkbox"/> | | | | |

Slika 2.7: Nerešena in rešena zlogovna križanka.

Za vse navedene vrste eno-črkovnih križank velja, da so za sestavljanje podobne. Pri sestavljanju je razlika le pri razporeditvi in številu črnih oz. opisnih kvadratkov.

2.2 Opis programa za stavljenje križank - Sfinga

Kot sem omenil že v uvodu, sem v podjetju izdelal program za stavljenje križank. Njegov izgled je prikazan na sliki 2.8. Program omogoča izdelavo in shranjevanje vseh vrst križank s pripadajočimi parametri, sliko križanke (v različnih formatih) in seznamom opisov in gesel. Omogočeno je tudi ročno vnašanje gesel in opisov gesel, ki se shranijo v bazo in je možno že vnesena gesla uporabiti v naslednji križanki. Gesla v bazi se ločijo tudi po tem koliko vrstic vsebujejo. To je pomembno pri slikovni in skandinavski križanki, kjer so opisi gesel vneseni v polja križanke in morajo biti ustrezno krajši. Ostale križanke imajo gesla navedena posebej na seznamu in imajo lahko tudi daljše opise. Sfinga vsebuje tudi modul za urejanje gesel in določanje parametrov gesel (težavnost, kvalifikatorje), kar je uporabno tudi pri avtomatskem sestavljanju križank.

Program uporablja bazo slik sistema Atlas¹. Omogočeno je iskanje slik po vsebini, tako da zlahka najdemo ustrezno sliko za iskano geslo.

Pri ponavljajočih se križankah je v pomoč tudi uporaba predlog. Križanko lahko v vsakem koraku shraniš kot predlogo in jo kasneje spet uporabiš.

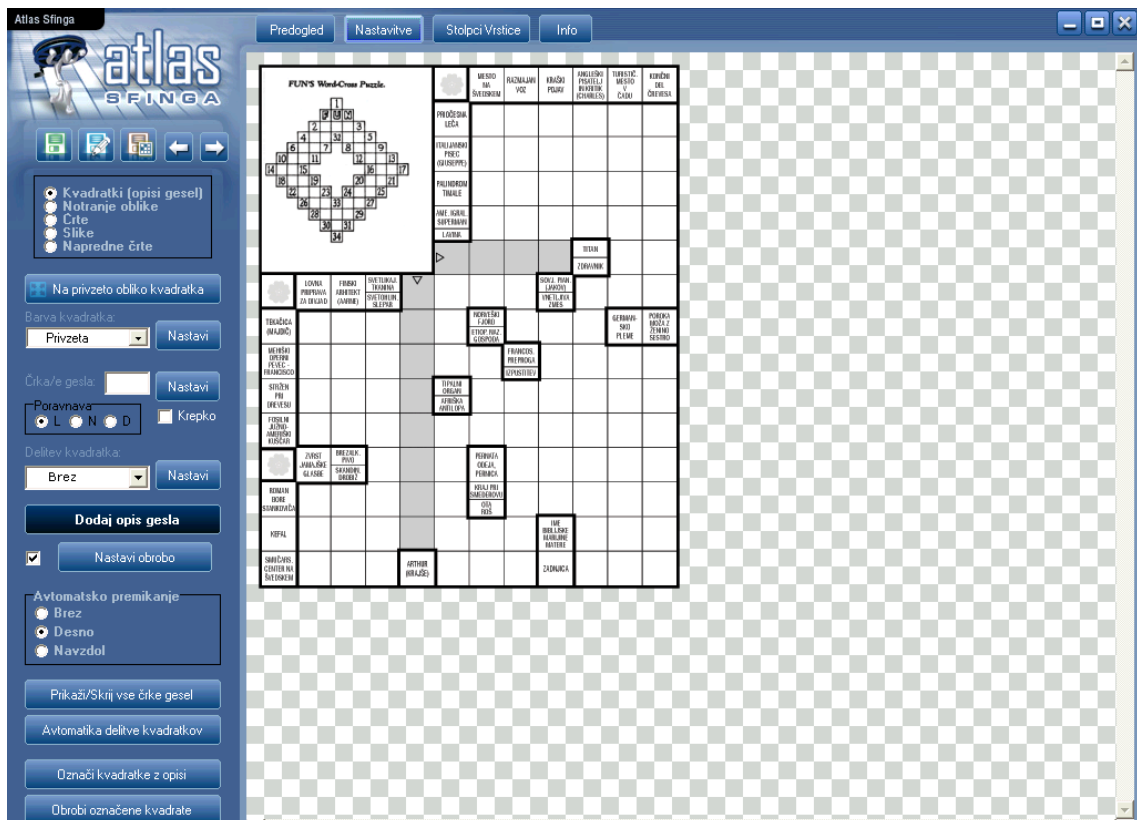
Za standardne križanke je vgrajena napredna avtomatika za:

- obrobo križanke, slik in opisnih kvadratkov,
- detekcijo in označevanje začetkov gesel,
- določanje opisnih kvadratkov, če so določene samo črke.

Omogočen je izvoz v naslednje formate:

- pdf (za tisk),
- InDesign (za nadaljnjo obdelavo),
- eps (za črno-beli tisk),
- xml-flash (interaktivna križanka za splet),
- jpg (predogled).

¹Atlas [11] je redakcijski sistem za avtomatizirano stavljenje vsebin v založništvu



Slika 2.8: Program Sfinga.

Poglavje 3

Opis problema

3.1 Hevristično preiskovanje

Križanke bomo sestavljali s preiskovanjem prostora stan, zato si najprej pogledamo definicijo in kratko formalizacijo problema.

V splošnem za preiskovanje prostora možnih rešitev potrebujemo začetno stanje S_z in množico operatorjev, ki transformirajo trenutno stanje v množico naslednikov: $\{S' \mid S \rightarrow S'\}$ Z množico operatorjev in začetnim stanjem je definiran prostor stanj:

$$P = \{S \mid S_z \rightarrow^* S\} \quad (3.1)$$

Poleg tega potrebujemo še oceno kvalitete stanja q . Naloga algoritma je, poiskati stanje $S_0 \in P$, ki maksimizira oceno kvalitete:

$$S_0 = \arg \max_{S \in P} q(S) \quad (3.2)$$

V mojem primeru je začetno stanje prazna oz. delno predizpolnjena križanka, operator pa je dodajanje ene besede na dovoljeno mesto. Ocene za kvaliteto stanja pa nisem uporabil eksplicitno, pač pa preko urejanja dovoljenih besed in vrstnega reda naslednjih besed.

Algoritmi pri preiskovanju prostora možnih stanj uporabljajo različne strategije. Osnovne strategije so [3]:

- deli in vladaj
- strategije iskanja optimalne rešitve
- strategije iskanja približne rešitve

- stohastični preiskovalni algoritmi

Za dodajanje novih besed bom uporabljam iskanje v globino. Ko pride algoritem v slepo vejo, kjer ni mogoče več nadaljevati, uporabljam večnivojsko vračanje, kar pa je približna rešitev. Pri izbiri položaja za besedo in potem tudi pri izbiri besede pa lokalno izberem najbolj ugodno besedo.

Izbira položaja oz. mesta besede teoretično ne vpliva na rešljivost križanke, ker je na vsa mesta potrebno dodati besede, njihov vrstni red pa za rešljivost ni pomemben. Z izbiro vrstnega reda besed pa lahko v veliki meri vplivamo na hitrost algoritma. Dobro je kot naslednje mesto izbrati takšno, ki mu ustreza manjše število besed, saj s tem omejimo preiskovalni prostor. Če izberemo mesto, kjer je število besed veliko, imamo tudi več možnosti, da izberemo takšno, ki bo povzročila, da na nekem mestu kjer je bilo prej malo ustreznih besed, ne bo nobene ustrezne besede več, kar sproži (nepotrebno in časovno drago) vračanje.

Pri izbiri besede pa napačna besede lahko pomeni tudi nerešljivost križanke.

Večnivojsko vračanje celotno zadevo zaplete. Če sklepamo po prej napisanem, bi bilo najboljšo vedno izbrati mesto kjer je najmanj možnih besed, vendar temu ni tako. Zaradi vračanja je zaželeno, da so po vrsti vstavljene besede razporejene precej blizu skupaj, kar pa pri minimalnih mestih ne drži vedno. Včasih je boljše, da vzamemo ne-minimalno mesto po številu ustreznih, ki je povezano s prejšnjo besedo, ne odstopa pa preveč od minimalnega.

3.2 Splošne zahteve za slikovne križanke

- povprečna dolžina gesel mora biti vsaj 4,7,
- maksimalno 3 črni kvadrati (dodatni črni kvadrati so v obliki slik),
- dve gesli dolžine 2 ne smeta biti vzporedni,
- geslo se lahko ponovi v križanki samo enkrat,
- minimalna dolžina gesla je 2.

Minimalna povprečna dolžina gesel je prvi kriterij težavnosti križank. Pri večji povprečni dolžini gesel je manj dvo in tri-črkovnih gesel, ki so v povprečju lažja za reševanje. Omejitve so različne glede na vrsto križanke, 4,7 je omejitev za slikovno križanko. V križanki so trije črni kvadrati navadno namenjeni navedbi avtorja, ilustratorja in številke križanke. Morebitni dodatni črni

kvadratki se pri slikovni križanki zaradi lepšega izgleda nadomestijo s slikami. Omejitev vzporednosti dvo-črkovnih gesel je vpeljana zato, da sestavljalci ne bi križali dveh daljših vzporednih besed z večjim številom enostavnih dvo-črkovnih gesel, kar bi imelo za posledico enostavnost križanke. Zadnji dve omejitvi spadata med splošna pravila sestavljanja križank.

3.3 Možne dodatne zahteve

- povprečna dolžina gesel (tipa vsaj, največ, približno 4,5),
- možnost ročnega dodajanja rešitve križanke (oz. delno izpolnjena križanka),
- možnost izbire nekaj 'sponzorskih' gesel, ki se bodo nujno pojavila v križanki,
- glede na kvalifikatorje¹
 - izbira gesel, pri katerih je zaželeno, da se pojavijo v križanki (če je možno, se izbere takšno geslo),
 - izbira gesel, ki se lahko pojavijo v križanki (to geslo izbere, če ni prvih gesel),
 - izbira gesel, ki se v nobenem primeru ne pojavijo v križanki.
- glede na zgodovino uporabe (določim dve časovni meji npr: pred enim mesecem in pred dvema mesecema)
 - gesla, ki so se uporabljala pred manj kot mesecem se ne uporabijo,
 - gesla, ki so se uporabljala pred več kot enim mesecem in manj kot dvema mesecema, se uporabijo manj verjetno,
 - ostala gesla se uporabijo enako verjetno.
- upoštevajo se omejitve glede števila vrstic v opisu gesla pri sestavljanju križanke (deljeni kvadratki pri slikovni križanki),
- lahko se določi minimalna dolžina gesla večja od 2.

Pri natančnem opisu rešitve bom posamezne točke, ki sem jih realiziral še podrobno opisal za ostale pa nakazal rešitve.

¹kvalifikatorji so največkrat vezani na neko področje (šport, geografija, zvezdniki,...), lahko pa so tudi drugačni (čudna gesla - mašila, splošno znano, iz SSKJ,...)

Poglavje 4

Baza gesel

4.1 Pridobitev baze

Bazo gesel sem prenesel s spletne strani slovarji.com [5]. Tam je baza razporejena po začetnih črkah in dolžinah gesel. Vsebuje skoraj vse besede slovenskega pravopisa, slovarja tujk, mitologije sveta, najpomembnejše pisce in glasbenike do leta 1980, veliko število imen starejših filmskih igralcev in drugega (imena mest, rek, geografskih pojmov, ...). Baza gesel se je nahajala v ločenih html datotekah z vključenimi opisi gesel. Najprej sem pridobil vsa gesla z opisi vred. Nato sem ločil gesla od opisov. Ohranil sem samo gesla, ki vsebujejo črke:

ABCČDEFGHIJKLMNOPRSŠTUVZŽ XYQW (vse slovenske in angleške črke)

Ker sta za vsako črko v križanki potrebni dve besedi, ena vodoravna beseda in ena navpična beseda, sem se odločil, da besed, ki vsebujejo črke, ki jih ni na zgornjem seznamu ne bom uporabil oz. bom problematične črke besed ustrezno prilagodil. Tudi sicer je z redko uporabljenimi črkami problem, kadar se takšna beseda pojavi v križanki. Možno je, da se pojavi na začetku, kasneje pa je malo verjetno, ker se težko najde ustrezna beseda, ki bi se križala s takšno besedo. Pri večbesednih geslih so dovoljeni tudi presledki, ki pa sem jih odstranil, ko sem besedo shranil v bazo. Presledek ni pomemben za samo sestavljanje križank, ampak le pri končni izdelavi križanke, da se na nek način označi mejo med kvadratoma oz. črkama. V bazi tudi ni gesel daljših od 40 črk, ker je maksimalna dimenzija križanke, ki jo stavimo 40x40 in ni potrebe po takšnih geslih.

Skupno sem v bazo shranil približno 120.000 različnih gesel (veliko gesel ima več različnih možnih opisov gesel).

4.2 Analiza baze gesel

Iz tabele 4.1 lahko razberem, da se samoglasniki (razen U) v besedah pojavljajo najpogosteje in tudi v isti besedi se pogosteje ponavljajo, kot ostale črke. Angleške črke XYQW so precej redke.

Tabela 4.2 nam pove koliko besed pride največ v seznamu, ko indeksiram besede. To je pomembno pri odločanju ali je potrebno več-črkovno indeksiranje. Podrobno bom to opisal v naslednjem podpoglavju.

Tabele 4.3 kaže zanimivo ugotovitev: 19 različnih zaporedij (2,2%) dveh črk se pojavlja v 25% vseh zaporedij dveh črk iz baze. Iz rezultatov, ki jih ni v tabeli pa sledi: 50 različnih zaporedij (6%) dveh črk se pojavlja v 50% vseh zaporedij dveh črk iz baze.

Vsaka beseda vsebuje za ena manj, kot je njena dolžina, dvo-črkovnih zaporedij. Pri 120.000 besedah v bazi je to približno 850.000 dvo-črkovnih zaporedij. Različnih dvo-črkovnih zaporedij pri 29 različnih črkah pa je 841.

Po mojem mnenju pri zaporedjih več črk tudi obstajajo bolj verjetna zaporedja, kar nam pride prav pri izbiri besed, da izberemo takšne, ki jih generirajo.

4.3 Indeksiranje baze gesel

4.3.1 Računalniški pomnilnik ali SQL Server

Potrebno se je bilo odločiti ali naj med sestavljanjem križank berem gesla iz baze SQL ali pa naj gesla predhodno shranim v računalniški pomnilnik. Za lažje odločanje sem v SQL bazi naredil novo tabelo in v njo vstavil en zapis. Primerjal sem čas, ki je potreben, da se prebere zapis iz tabele in čas da se preveri nek pogoj (recimo ali je tretja črka I) na vseh geslih v pomnilniku. Vsako od omenjenih operacij sem 1000x ponovil in ugotovil, da je SQL komunikacija 10-20 krat počasnejša. Če bi bil SQL strežnik v mreži in ne na lokalnem računalniku in če bi bilo potrebno opraviti še selekcijo med 120.000 gesli, bi to še dodatno upočasnilo postopek.

Gesla so zaradi komunikacije z ostalimi programi, še vedno shranjena v bazi, pred sestavljanjem križanke pa jih shranim v pomnilnik. Poleg gesel shranim še attribute kvalifikatorjev in zgodovino uporabe.

4.3.2 Indeksiranje

Poizvedbe za gesla pri generiranju križanke so lahko na primer naslednjih tipov (prikazano na sliki 4.1):

- gesla, ki so dolga najmanj 5 in največ 9 znakov (7 ni možno ¹)
- gesla, ki imajo tretjo črko A in so dolga največ 7 znakov (5 ni možno ¹)
- gesla, ki imajo peto črko C in so dolga 6 ali 9 znakov
- gesla, ki imajo drugo črko N, peto črko E in so dolga 3,5 ali 6 znakov

Seznam gesel je torej potrebno urediti na način, da bodo zgornje poizvedbe čim hitreje. Najprej sem seznam uredil po dolžini gesel od najkrajših do najdaljših. To je globalni seznam gesel. Za hiter dostop do gesel določenih dolžin sem dodal še seznam začetkov besed določene dolžine.

Globalni seznam je namenjen za poizvedbe prvega tipa. Število besed dolžine najmanj 3 do največ 7 dobiš kar tako, da od indeksa začetka besed dolžine $7+1$ odšteješ indeks začetka besed dolžine 3.

Gesla sem indeksiral po določni črki na določenem mestu. Vse besede, ki imajo A,B,C, ... Ž na 1.,2., ... 40. mestu. Nastali dvodimenzionalni seznam vsebuje seznam indeksov ustreznih besed in še seznam začetkov besed določene dolžine v tem ustreznem seznamu. To indeksiranje omogoča poizvedbe tipa dva in tri.

Za poizvedbe tipa 4 uporabljam pravkar navedeno indeksiranje, le da se moram naknadno sprehoditi po seznamu in pregledati, če izbrane besede vsebujejo 2.,3., ... črko na ustreznem mestu.

Če se izkaže, da je gesel z določeno črko na določenem mestu preveč, se lahko izvede nadaljnje indeksiranje po vseh položajih črke večjih od trenutne črke. To podindeksiranje je smiselno po potrebi uporabiti le za nekaj najbolj problematičnih črk, ker zavzame precej prostora. V prejšnjem poglavju sem ugotovil, da se lahko na seznamu pojavi tudi 22.000 besed za vse dolžine besed. V realnih primerih se velikokrat zgodi, da je možnih le nekaj različnih dolžin besed, kar močno zmanjša seznam besed, ki jih je potrebno pregledati.

Za lažjo predstavo sem dodal sliko 4.2, ki prikazuje kako se besede indeksirajo.

¹Pri minimalni dolžini gesla 2 vedno ni mogoča dolžina gesla, ki je enaka za dva zmanjšani maksimalni možni dolžini gesla. Če bi tam postavili črn kvadrateg, bi imel za posledico eno-črkovno geslo.

| | | | | | | | | | |
|---|---|----------------|---|---|----------------|---|---|---|----------------|
| ■ | □ | □ | □ | □ | ■ | ■ | ■ | □ | 2 |
| □ | ■ | ■ | ■ | ■ | □ | □ | □ | ■ | 1 |
| 3 | 1 | 2 | 3 | 4 | 5 C | 6 | 7 | 8 | 9 2 |
| □ | □ | □ | □ | □ | ■ | ■ | ■ | ■ | A ³ |
| □ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | 4 |
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 5 |
| □ | □ | □ | □ | ■ | ■ | ■ | ■ | ■ | 6 |
| □ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | 7 |
| 4 | 1 | N ₂ | 3 | 4 | E ₅ | 6 | □ | □ | ■ |
| □ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| □ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |

Slika 4.1: Prikaz tipov poizvedb za gesla. Potemnjeni kvadratici ne morejo postati črni oz. opisni kvadratici.

4.3.3 Iskanje besed

Na primeru gesel, ki imajo 5. črko N, 8. črko A in 10. črko C ter so dolga največ 15 znakov bom pokazal, katere skupine gesel ustrezajo temu pogoju in katere tipe iskanj iz prejšnjega podpoglavja potrebujemo.

Gesla, ki ustrezajo pogoju (-, -, -, -, N, -, -, A, -, C, -, -, -, -, -):

- vsa gesla dolžine 4 ali manj (iskanje tipa 1),
- vsa gesla dolžine 7 ali manj, ki imajo 5. črko N (iskanje tipa 2),
- vsa gesla dolžine 9 ali manj, ki imajo 5. črko N in 8. črko A (iskanje tipa 4),

- vsa gesla dolžine 15 ali manj, ki imajo 5. črko N, 8. črko A in 10. črko C (iskanje tipa 4).

Pri križankah je bolj ustrezno, če od navedenih dolžin gesel še odštejemo 1, ker ne more biti geslo recimo dolžine 4, če je na naslednjem mestu že črka N (potreben je vmesni opisni kvadrateg).

4.4 Velikost baze gesel

Poskus generiranja križanke (20x20) na bazi, ki je imela 850.000 besed, je pokazal, da sem precej hitreje sestavil križanko z večjo omejitvijo (minimalna dolžina gesel 3). Poskusil sem tudi zmanjšati bazo gesel z odstranjevanjem le teh. Ugotovil sem, da pri velikosti baze manj kot 30.000 gesel, križanke praktično ne morem več sestaviti v 10 minutah.

Večja kot je baza, večje omejitve lahko postavim za križanko in manjša mora biti zahtevnost algoritma za izbiro besede.

| črka | % besed s črko | povprečno ponovitev iste črke |
|------|-------------------|-------------------------------------|
| A | 68,2 | 1,46 |
| E | 55,5 | 1,38 |
| I | 52,3 | 1,27 |
| N | 49,2 | 1,19 |
| O | 49,2 | 1,31 |
| R | 48,7 | 1,15 |
| T | 36,8 | 1,18 |
| L | 36,2 | 1,11 |
| S | 32,7 | 1,15 |
| K | 30,6 | 1,11 |
| J | 24,5 | 1,1 |
| V | 23,9 | 1,09 |
| P | 21,5 | 1,08 |
| D | 19,7 | 1,06 |
| M | 19,5 | 1,06 |
| C | 19 | 1,05 |
| U | 16,2 | 1,06 |
| B | 14,5 | 1,05 |
| G | 13,7 | 1,05 |
| Z | 12,8 | 1,05 |
| Č | 10,1 | 1,04 |
| H | 8,8 | 1,04 |
| Š | 7,5 | 1,02 |
| F | 5,5 | 1,06 |
| Ž | 4,7 | 1,02 |
| Y | 1,7 | 1,02 |
| W | 1,3 | 1,03 |
| X | 0,3 | 1,01 |
| Q | 0,2 | 1,02 |

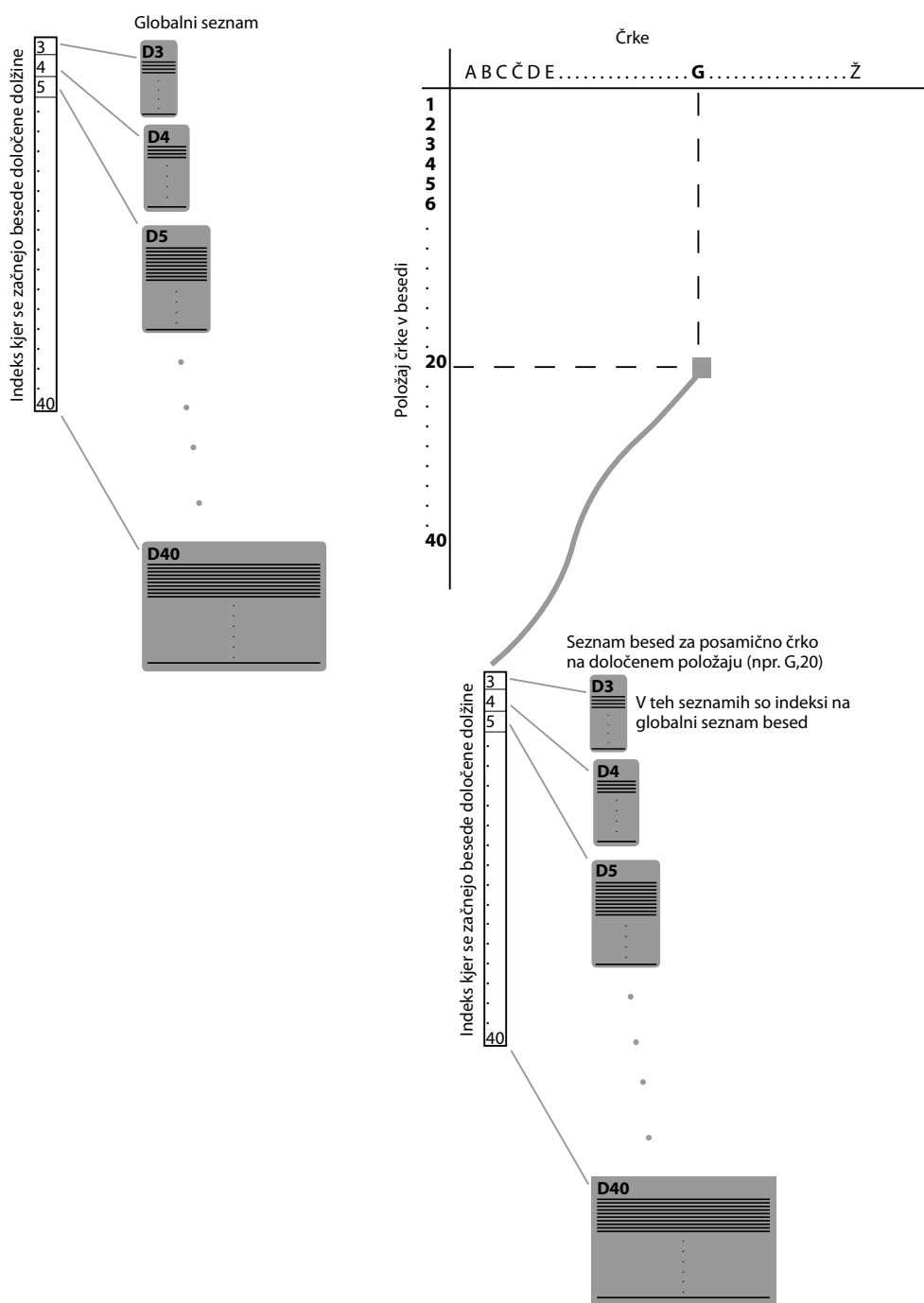
Tabela 4.1: Pogostnost posamezne črke v besedi in pogostnost ponavljanja posamezne črk v besedi - upoštevane so samo tiste besede v katerih se črka pojavi vsaj enkrat.

| črka | položaj v besedi | št. besed | % besed |
|------|------------------|-----------|---------|
| A | 2 | 22615 | 19,1 |
| O | 2 | 18201 | 15,4 |
| E | 2 | 16255 | 13,7 |
| R | 2 | 14545 | 12,3 |
| P | 1 | 14087 | 11,9 |
| A | 5 | 13120 | 11,1 |
| A | 4 | 12765 | 10,8 |
| A | 6 | 12333 | 10,4 |
| R | 3 | 12223 | 10,3 |
| E | 4 | 11513 | 9,7 |
| E | 5 | 11325 | 9,6 |
| O | 4 | 11100 | 9,4 |
| I | 4 | 11077 | 9,3 |
| A | 7 | 11041 | 9,3 |
| I | 5 | 10757 | 9,1 |
| N | 6 | 10740 | 9,1 |
| S | 1 | 10416 | 8,8 |
| I | 2 | 9997 | 8,4 |

Tabela 4.2: Število besed, ki ima na določenem mestu v besedi določeno črko.

| dve črki | št. pojavitev zaporedja | % besed | kumulativni % najbolj pogostih kombinacij |
|----------|-------------------------|---------|---|
| A N | 17918 | 15,1 | 2,1 |
| S T | 15543 | 13,1 | 3,9 |
| R A | 13234 | 11,2 | 5,5 |
| J E | 13137 | 11,1 | 7,1 |
| N I | 11846 | 10 | 8,5 |
| E N | 11625 | 9,8 | 9,8 |
| A R | 11510 | 9,7 | 11,2 |
| K A | 11435 | 9,7 | 12,5 |
| J A | 10849 | 9,2 | 13,8 |
| I N | 10654 | 9 | 15,1 |
| O S | 10406 | 8,8 | 16,3 |
| E R | 10353 | 8,7 | 17,5 |
| R E | 9868 | 8,3 | 18,7 |
| N J | 9713 | 8,2 | 19,8 |
| A L | 9570 | 8,1 | 21 |
| R I | 9353 | 7,9 | 22,1 |
| N A | 9000 | 7,6 | 23,1 |
| N O | 8475 | 7,2 | 24,1 |
| L A | 8067 | 6,8 | 25,1 |

Tabela 4.3: Pogostnost zaporedja dveh črk in kumulativni procent pogostnosti.



Slika 4.2: Prikaz globalnega indeksiranja in indeksiranja za posamezne črke na določenih položajih.

Poglavje 5

Potek generiranja križank

5.1 Predpriprava

Na začetku je potrebno določiti dimenzijo križanke in vsaj levo in zgoraj vse opisne kvadratke (Slika 5.1 - levo). Če je križanka večjih dimenzij je priporočljivo dodati še na vsakih 5-8 kvadratkov zgoraj in levo še en opisni kvadratek, da ni potrebno vstavljati zelo dolgih besed. Zaradi hitrejše sestave križanke je priporočljivo dodati še en opsijski opisni kvadratek. Podobno lahko dodamo kvadratke še desno in spodaj (Slika 5.1 - desno). To veliko pripomore k hitrosti sestavljanja križanke.

| OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |

| OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK | OPISNI KVADRATEK |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|---------------------------|------------------|---------------------------|
| OPISNI KVADRATEK | | | | | | | OPISNI KVADRATEK | | |
| OPISNI KVADRATEK | | | | | | | OPCIJSKO OPISNI KVADRATEK | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | OPCIJSKO OPISNI KVADRATEK |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | | | |
| OPISNI KVADRATEK | | | | | | | OPCIJSKO OPISNI KVADRATEK | | |

Slika 5.1: Začetna predpriprava križanke.

Če imamo na začetku v križanki že nekaj besed, po potrebi dodamo opisni kvadrater pred geslo in na primeren način označimo, da je beseda v izbrano smer že vnesena, nato pa dodamo opisni kvadrater na koncu besede, ki bo omogočal nadaljevanje sestavljanja križanke.

V podpoglavju 3.3 sem omenil sponzorske besede. Za dodajanje teh besed še nimam implementiranega algoritma in jih za testiranje dodajam ročno.

5.2 Povezanost besed

V tem podpoglavju bom analiziral vpliv vstavljene besede na ostale še ne vnesene prostore za besede oz. začetne opisne kvadratke le teh. Primer na sliki 5.2 je za vodoravno besedo, pri navpični potrebno zamenjati ustrezne smeri.

Kvadratke na katere vnesena beseda vpliva, bom ločil s pojmom: močno povezani kvadratki (SC) in šibko povezani kvadratki (LC).

Močno povezani so tisti kvadratki na katere vplivajo črke vnesene besede in posledično zmanjšajo število besed. Nastala omejitev je obvezna črka na določenem mestu.

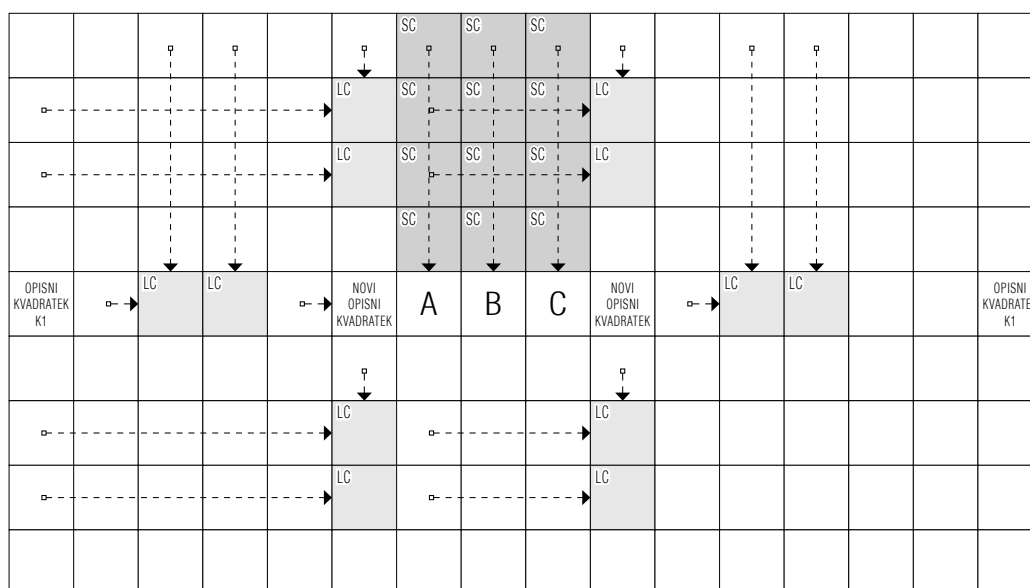
Med šibko povezane kvadratke pa sodijo tisti na katere vplivajo opisni kvadratki, ki so posledica vnosa opisnega kvadratka in manj zmanjšajo število besed. Nastala omejitev je, da določena dolžina besede ni možna.

Večina šibko povezanih kvadratkov je posledica omejitve minimalne dolžine besede na 2 oz. v primeru slike 5.2 na 3 kvadratke. Na sliki so kvadratki, ki zaradi prejšnje omejitve ne smejo biti črni oz. opisni, označeni z LC, v njih pa, iz besede na katere vplivajo, prihajajo puščice. Puščica tudi označuje smer v katero poteka omejena beseda (samo na to smer ima vpliv).

V nadaljevanju bom pri opisovanju algoritma večkrat uporabil pojma močno in šibko povezani kvadratki.

5.3 Korak generiranja - vstavljanje besede

Pri vsakem koraku generiranja križanke, pri katerem se beseda vnese na novo oz. zamenja, moramo opraviti korake, ki jih bom opisal v naslednjih podpoglavjih. Postopek pa ne velja za korake pri katerih je potrebno večnivojsko vračanje, kar bom opisal pozneje.



Slika 5.2: Omejitve po vstavitvi besede.

5.3.1 Nemogoči opisni kvadrati

Glede na omejitve pri križanki določim nekatere kvadratke za nemogoče opisne kvadratke.

Pri minimalni dolžini gesla N so takšni vsi, ki so od nekega opisnega kvadratka v vodoravni oz. navpični oddaljeni od 2 do N kvadratkov.

Takšni so tudi spodnji sosedi kvadratkov, ki imajo za desnega soseda že opisni kvadrater oz. obratno. To bi imelo za posledico dodatni črni kvadrater, ki pa ga ne bom dovolil, ker bom vse morebitne črne kvadratke določil že na začetku.

Glede na splošne zahteve pri slikovni križanki so takšni tudi tisti kvadrati, ki bi povzročili vzporednost, dveh dvo-črkovnih gesel.

5.3.2 Izbira položaja za vstavljanje besede

Na začetku sem iskal ustrezen položaj za besede samo po trenutnih opisnih kvadratih in nastal je problem pri daljših prostorih za besede. Pri slednjih je veliko možnih besed od kratkih do dolgih. Problem nastane, če se pri določeni besedi, zaradi vstavitve neke druge besede, zapolnijo črke, ki so bolj proti koncu besede. Kljub tem črkam, ki so lahko popolnoma neustrezne za kakršnokoli besedo, je na začetku še dovolj krajših besed in se to mesto ne detektira kot

problematično. Ko pa se začetne črke zapolnijo, pa z ostankom črk na koncu, ni mogoče sestavit besede.

Ugotovitev me je pripeljala do sklepa, da je potrebno za možne položaje uporabiti vsa opisna in vsa prazna polja, ki niso med nemogočimi opisnimi polji. To je rešilo prejšnji problem, pojavil pa se je nov. Zaradi malo možnih besed pri krajših prostorih za besede, so se večinoma izbirale kratke besede, kar je imelo za posledico zelo nizko povprečno dolžino besed pri križanki.

Uvedel sem postopek podaljševanja gesel. Pri tri-črkovnih oz. štiri-črkovnih prostorih sem poskušal podaljšati prostor na pet-črkovni, če veččrkovni prostor vsebuje več ali enako število ustreznih besed, kot manjčrkovni prostor, kar je prikazano na sliki 5.3. Na sliki je pod prostorom napisno tudi število ustreznih besed določene dolžine. Podaljševanje besed je povprečno dolžino besed v križanki v povprečju povečalo za pol črke.

| | | | | | |
|-----|----|----|--|---|---|
| | | | | | |
| 5 | 4 | 3 | | A | N |
| 205 | 55 | 17 | | | |

Slika 5.3: Daljšanje besed (podaljša se lahko iz dolžine 3 na 5).

Med vsemi ustreznim položaji sem položaj za vnos besede izbiral na 3 načine:

- po minimalnem številu ustreznih besed,
- po minimalnem številu ustreznih besed in sosednosti (kombinacija),
- po sosednosti (če ni sosednosti po prvem načinu).

Pri prvem načinu sem za vse možne položaje besed izračunal število ustreznih besed glede na že vnesene črke in izbral tisti položaj, ki je imel minimalno besed.

Pri drugem načinu sem prav tako izračunal število ustreznih besed za vse položaje in primerjal število besed pri minimalnem izmed vseh in minimalnem izmed sosednjih. Če število besed pri minimalnem sosedu ni preveč odstopalo od absolutnega minimuma sem izbral soseda. Sosednji kvadrati so tisti, ki so močno povezani (Podpoglavje 5.2) s prejšnjo besedo.

Pri zadnjem načinu sem uporabil minimalnega soseda. Absolutni minimum sem uporabil le, ko se je določen podprostor zapolnil in vstavljena beseda ni več imela prostega soseda. Ta izbira se ni izkazala za uspešno.

Poskusil sem tudi možnost, da na 10% najboljših položajev za besede (po določenem kriteriju) opravi dodatno računanje in tako izberem najboljši položaj. Za vsak položaj sem določil katere črke so možne na določenem mestu glede na njegove močno povezane sosede. V bistvu sem opravil izračun za en korak naprej na določenih mestih. Ta postopek se ni izkazal za preveč uspešenega, saj je bil časovno precej potraten.

5.3.3 Izbira besede

Pri izbiri besede iz seznama ustreznih besed sem favoriziral daljše besede, nato pa med določenim deležem najdaljših eno izbral naključno. Nisem upošteval značilnosti besed glede na razporeditev oz. pojavljanje določenih črk. Tukaj je še veliko možnosti za izboljšave algoritma. Izbrano besedo vstavim in ponovim korak algoritma.

Če v določenem koraku na nekem mestu ni bilo možno vstaviti besede, sem uporabil vračanje, ki ga bom opisal v naslednjih poglavjih.

5.4 Menjava vstavljene besede - vračanje en korak nazaj

Ko pride v nekem koraku do problema, da v en ali več položajev v križanki ni mogoče vstaviti besede, je potrebno zadnjo vstavljeno besedo odstraniti in predlagati novo. Potrebno je odkriti vzrok nastale težave in odstraniti vse besede, ki bi lahko pripeljale do enakih težav.

Opisal bom naslednje možnosti za reševanje težav:

- prostoru, ki je močno povezan z zadnjo vstavljeno besedo sledi odstranitev vseh besed, ki imajo na problematičnem mestu enako črko kot vstavljena beseda,
- prostoru, ki je šibko povezan z zadnjo besedo in je posledica opisnega kvadratika na koncu besede sledi odstranitev vseh besed, ki imajo enako dolžino kot trenutna beseda,

- prostoru, ki je šibko povezan z zadnjo besedo in je posledica opisnega kvadratka na začetku besede sledi odstranitev vseh besed na trenutnem položaju in označitev začetnega kvadratka za nemogoči opisni kvadrater.

5.5 Večnivojsko vračanje

Pri večnivojskem vračanju se pojavijo tri vprašanja:

- Kdaj se sploh vračati več nivojev nazaj?
- Na katere nivoje se je ustrezno vrniti?
- Kateri nivo izmed ustreznih izbrati?

Večnivojsko vračanje nazaj je smiselno uporabiti le pri položajih za besede, kjer je bilo na začetku kar nekaj besed, ki pa so se zaradi različnih problemov odstranile. Pri tej ugotovitvi sem sklepal tako: če sem na trenutnem položaju preizkusil veliko besed, pa nobena ni bila ustrezna, potem verjetno ni problem v tem položaju ampak v nekem prejšnjem. Če sem preizkusil zelo malo besed, potem tega ne morem trditi.

Potrebno se je vrniti na nivoje, kjer so besede na nek način povezane s problematično besedo. Povezavo bom razložil s pomočjo slike 5.4. Pri geslu za opis 6 pride do problemov. Vse besede, ki so bile ustrezne na tem mestu, so bile že preizkušene zadnja je SABLJAČ, ki tudi ni ustrezna, zato je potrebno vračanje. Na levi strani so potemnjeni kvadrati označeni s PR. To so vsi kvadrati, ki so bili problematični med vstavljanjem besed za opis 6. Potrebno se je vrniti v nivo, kjer je bila vstavljena beseda povezana z enim izmed teh problematičnih kvadratkov. V navedenem primeru je to na opis 2. Pri implementaciji v mojo rešitev sem uporabil poenostavitev. Izračunal sem močno povezane sosedbe opisa 6 in 2 nato pa preveril, če imata kašnega skupnega ter se v tem primeru vrnil na nivo 2. Ta poenostavitev sem lahko izvedel, ker se pri večjem številu preverjenih gesel redko zgodi situacija na sliki, da v tretji in četrti vrstici ni nobenih problematičnih kvadratkov. V navedenem primeru bi me moja poenostavitev sicer lahko pripeljala do vračanja na opis 4.

Ob predpostavki, da na sliki 5.4 kvadrater označen z P* upoštevamo tudi kot problematičnega, imamo dva možna položaja za vrnitev. Potrebno se je odločiti na kateri položaj naj se vrnemo.

Možni so trije pristopi:

- vračamo se čim manj nivojev nazaj,

| | | | | | | | | | |
|----|-----------|-----------|-----------|----|----|-----------|---|-----------|-----------|
| | | | OPIS 2 | | | OPIS 4 | | OPIS 6 | |
| | OPIS 1 | F | I | N | A | L | I | S | T |
| | | | D | P* | | I | | A | |
| | | | E | | | K | | B | |
| | PR | PR | A | | PR | | | L | |
| | PR | PR | L | | | PR | | J | |
| PR | PR | PR | I | | | PR | | A | OPIS 5 |
| | | OPIS 3 | S | E | Č | I | Š | Č | E |
| | | | T | | | | | | B |
| | | | | | | | | | E |
| | | | | | | | | | N |

Slika 5.4: Večnivojsko vračanje.

- vračamo se čim več nivojev nazaj,
- primerjamo vse možne položaje glede na omejevanje, ki ga povzročijo pri ostalih besedah in se odločimo za menjavo tistega, ki najbolj omejuje.

Drugi pristop lahko povzorči preveč vračanja še posebno, če je velika razlika med nivojema za prvi in zadnji položaj. Možno je, da zadnji pristop boljši, vendar ga zaradi zapletenosti nisem izvedel. Odločil sem se za prvi pristop.

Ostaja še odprto vprašanje, katere besede odstraniti iz seznama ustreznih besed pri večnivojskem vračanju. Na podlagi problematičnih kvadratkov iz slike 5.4 bi sicer lahko določili katere črke so problematične na dotičnem nivoju, vendar se lahko tudi ob menjavi neke druge črke zgodi, da se potek sestavljanja popolnoma spremeni glede na trenutni potek in so problematične popolnoma druge črke. Takšnim besedam sem le zmanjšal verjetnost izbire v naslednjem poskusu, ne pa dokončno odstranil.

Poglavje 6

Analiza rezultatov

6.1 Testiranje sestavljanja križank

Testiranje sestavljanja sem izvedel na šestih različnih mrežah. Pri mreži 5x5 sem uporabil dva pristopa glede vstavljanja črnih kvadratkov: s fiksnimi črnimi kvadratki in sprotno vstavljanje črnih kvadratkov. Pri slikovni križanki sem uporabil dve zelo podobni mreži. Pri prvi je bil eden dodatni črni kvadrateg na robovih mreže pri drugi pa dva.

Za vsako mrežo sem opravil vse kombinacije generiranj glede na minimalno dolžino besede (2 ali 3) in velikost baze (celotna ali polovična). Za vsako kombinacijo pa sem nato sestavljal križanko z privzetimi parametri (večnivojsko vračanje, izberi mesto z minimalno besed in izbiraj daljše besede iz seznama ustreznih) in še tako, da sem vsakega izmed njih dal na neprivzeto vrednost (brez večnivojskega vračanja, izberi mesto s kombinacijo minimalnega števila besed in povezanih besed, izbiraj popolnoma naključno iz seznama ustreznih besed) ostala dva pa pustil na privzeti vrednosti.

Za vsako mrežo sem tako opravil 16 različnih testov in vsakega ponovil 100-krat. Zelo podobne rezultate sem smiselno združil v tabeli.

Pri rezultatih me je zanimala povprečna dolžina gesla, čas sestavljanja in odstotek neuspešnih sestavljanj.

Bolj natančno (v takšem vrstnem redu kot v tabeli):

- odstotek sestavljanj, ki so po času za več kot trikrat odstopala od povprečnega časa uspešno sestavljenih križank,
- odstotek sestavljanj, ki niso bila uspešno dokončana,

- povprečna dolžina gesla pri uspešno sestavljenih križankah,
- povprečni čas sestave križanke pri uspešno sestavljenih križankah,
- povprečni čas sestave križanke pri vseh križankah (časovna omejitev sestavljanja je bila 400 sekund).

| min dol. bes. | cel. baza | večn. vrač. | min. besed | dalj. besed | % napak | % velik. odstop. | povp. dolž. | povp. čas (ok) | povp. čas (vse) |
|---------------|-----------|-------------|------------|-------------|---------|------------------|-------------|----------------|-----------------|
| 2 | da | da | da | da | 0 | 1 | 4,41 | 1,1 | 1,1 |
| 2 | da | da | da | ne | 0 | 0 | 4,3 | 0,98 | 0,98 |
| 2 | ne | da | da | da | 0 | 0 | 4,38 | 1,23 | 1,23 |
| 2 | ne | da | da | ne | 1 | 3 | 4,3 | 2,34 | 8,95 |
| 3 | da | da | da | da | 26 | 31 | 5,15 | 29,76 | 129,67 |
| 3 | da | da | da | ne | 14 | 20 | 5,1 | 15,07 | 70,4 |
| 3 | ne | da | da | da | 20 | 28 | 5,11 | 25,72 | 104,85 |
| 3 | ne | da | da | ne | 16 | 21 | 5,1 | 21,97 | 85,91 |
| 2,3 | da,ne | da | ne | da | 100 | 100 | -1 | -1 | -1 |
| 2 | da | ne | da | da | 95 | 95 | 4,36 | 0,47 | 143,99 |
| 2 | ne | ne | da | da | 97 | 97 | 4,29 | 0,46 | 146,5 |
| 3 | da,ne | ne | da | da | 100 | 100 | -1 | -1 | -1 |

Tabela 6.1: Rezultati sestavljanja slikovne križanke (slika A.1) z vključenimi OP* črnimi kvadrati.

| min dol. bes. | cel. baza | večn. vrač. | min. besed | dalj. besed | % napak | % velik. odstop. | povp. dolž. | povp. čas (ok) | povp. čas (vse) |
|---------------|-----------|-------------|------------|-------------|---------|------------------|-------------|----------------|-----------------|
| 2 | da | da | da | da | 0 | 2 | 4,49 | 5,18 | 8,52 |
| 2 | da | da | da | ne | 0 | 0 | 4,4 | 1,13 | 1,13 |
| 2 | ne | da | da | da | 3 | 4 | 4,47 | 1,43 | 14,82 |
| 2 | ne | da | da | ne | 0 | 1 | 4,4 | 1,2 | 1,2 |
| 3 | da | da | da | da | 41 | 47 | 5,16 | 65,31 | 207,13 |
| 3 | da | da | da | ne | 42 | 48 | 5,2 | 52,99 | 202,69 |
| 3 | ne | da | da | da | 40 | 46 | 5,16 | 71,43 | 207,89 |
| 3 | ne | da | da | ne | 35 | 40 | 5,2 | 45,2 | 171,39 |
| 2,3 | da,ne | da | ne | da | 100 | 100 | -1 | -1 | -1 |
| 2,3 | da,ne | ne | da | da | 100 | 100 | -1 | -1 | -1 |

Tabela 6.2: Rezultati sestavljanja slikovne (slika A.1) križanke brez OP* črnih kvadratkov.

| min dol. bes. | cel. baza | večn. vrač. | min. besed | dalj. besed | % napak | % velik. odstop. | povp. dolž. | povp. čas (ok) | povp. čas (vse) |
|---------------|-----------|-------------|------------|-------------|---------|------------------|-------------|----------------|-----------------|
| 2 | da | da | da | da | 0 | 0 | 4,42 | 1,45 | 1,45 |
| 2 | ne | da | da | da | 1 | 2 | 4,43 | 1,17 | 7,91 |
| 3 | da | da | da | da | 21 | 29 | 5,01 | 32,24 | 113,19 |
| 3 | ne | da | da | da | 24 | 29 | 5 | 23,76 | 115,94 |
| 2 | da | da | da | ne | 0 | 0 | 4,4 | 0,76 | 4,14 |
| 2 | ne | da | da | ne | 0 | 0 | 4,4 | 0,89 | 0,89 |
| 3 | da | da | da | ne | 20 | 27 | 5 | 26,22 | 105,04 |
| 3 | ne | da | da | ne | 22 | 29 | 5 | 21,14 | 107,6 |
| 3 | da,ne | ne | da | da | 100 | 100 | -1 | -1 | -1 |
| 2 | da,ne | ne | da | da | 97 | 97 | 4,35 | 0,4 | 146,48 |
| 2,3 | da,ne | da | ne | da | 100 | 100 | -1 | -1 | -1 |

Tabela 6.3: Rezultati sestavljanja skandinavske križanke (slika A.2).

| min dol. bes. | cel. baza | večn. vrač. | min. besed | dalj. besed | % napak | % velik. odstop. | povp. dolž. | povp. čas (ok) | povp. čas (vse) |
|---------------|-----------|-------------|------------|-------------|---------|------------------|-------------|----------------|-----------------|
| 2 | da | da | da | da | 4 | 5 | 4,48 | 1,75 | 18,53 |
| 2 | da | da | da | ne | 4 | 5 | 4,4 | 1,47 | 18,34 |
| 2 | ne | da | da | da | 4 | 8 | 4,47 | 6,34 | 22,91 |
| 2 | ne | da | da | ne | 5 | 6 | 4,4 | 2,27 | 22,38 |
| 3 | da | da | da | da | 45 | 51 | 5,21 | 39,39 | 207,15 |
| 3 | da | da | da | ne | 37 | 41 | 5,2 | 24,14 | 167,03 |
| 3 | ne | da | da | da | 40 | 44 | 5,17 | 42,6 | 187,55 |
| 3 | ne | da | da | ne | 45 | 49 | 5,2 | 29,71 | 198,55 |
| 2,3 | da,ne | da | ne | da | 100 | 100 | -1 | -1 | -1 |
| 2,3 | da,ne | ne | da | da | 100 | 100 | -1 | -1 | -1 |

Tabela 6.4: Rezultati sestavljanja sponzorske križanke (slika A.3).

| min dol. bes. | cel. baza | večn. vrač. | min. besed | dalj. besed | % napak | % velik. odstop. | povp. dolž. | povp. čas (ok) | povp. čas (vse) |
|---------------|-----------|-------------|------------|-------------|---------|------------------|-------------|----------------|-----------------|
| 2 | da,ne | da | da | da,ne | 0 | 0 | 4,48 | 0,1 | 0,1 |
| 3 | da,ne | da | da | da | 0 | 0 | 5 | 0,19 | 0,19 |
| 3 | da,ne | da | da | ne | 0 | 0 | 5 | 0,14 | 0,14 |
| 2 | da,ne | da | ne | da | 76 | 76 | 4,35 | 0,07 | 115,09 |
| 3 | da,ne | da | ne | da | 94 | 94 | 5 | 0,08 | 138,84 |
| 2 | da,ne | ne | da | da | 65 | 65 | 4,41 | 0,09 | 100,11 |
| 3 | da,ne | ne | da | da | 88 | 88 | 5 | 0,08 | 134,01 |

Tabela 6.5: Rezultati sestavljanja križanke 5x5 (slika A.4) brez fiksnih črnih kvadratkov.

| min dol. bes. | cel. baza | večn. vrač. | min. besed | dalj. besed | % napak | % velik. odstop. | povp. dolž. | povp. čas (ok) | povp. čas (vse) |
|---------------|-----------|-------------|------------|-------------|---------|------------------|-------------|----------------|-----------------|
| 2,3 | da,ne | da | da | da,ne | 0 | 0 | 5 | 0,16 | 0,16 |
| 2,3 | da,ne | da | ne | da | 95 | 95 | 5 | 0,08 | 142,09 |
| 2,3 | da,ne | ne | da | da | 88 | 88 | 5 | 0,09 | 132,14 |

Tabela 6.6: Rezultati sestavljanja križanke 5x5 (slika A.4) s fiksnimi črnimi kvadratki.

| min dol. bes. | cel. baza | večn. vrač. | min. besed | dalj. besed | % napak | % velik. odstop. | povp. dolž. | povp. čas (ok) | povp. čas (vse) |
|---------------|-----------|-------------|------------|-------------|---------|------------------|-------------|----------------|-----------------|
| 2,3 | da,ne | da | da | da,ne | 0 | 0 | 4,6 | 0,26 | 0,26 |
| 2,3 | da,ne | da | ne | da | 100 | 100 | -1 | -1 | -1 |
| 2,3 | da,ne | ne | da | da | 87 | 87 | 4,6 | 0,17 | 131,22 |

Tabela 6.7: Rezultati sestavljanja velike križanke (slika A.5) s fiksnimi črnimi kvadratki.

6.2 Obrazložitev rezultatov

Najbolj nepričakovano je, da izbira položaja za vstavitve besede ne deluje dobro, če položaj z minimalnim številom besed v določenih primerih zamenjam s položaji, ki so povezani s prejšnjo besedo. Pričakoval bi, da bo takšna izbira malo boljša od izbire minimalnega mesta, je pa v vseh primerih zelo slaba.

Poskus brez večnivojskega vračanja se je pričakovano zelo slabo obnesel. Večnivojsko vračanje je ključnega pomena pri mojem algoritmu sestavljanja križank, saj edino omogoča odstranitev neoptimalnih gesel, ki so bila dodana več kot 10 korakov pred trenutnim.

Zmanjšanje velikost baze iz 120.000 gesel na 60.000 gesel nima opaznega vpliva na sestavljanje križanke. V nekaterih primerih je sestavljanje križanke celo hitrejše z manjšo bazo.

Pri sestavljanju večje slikovne (sl. A.1), skandinavske (sl. A.2) in sponzorske (sl. A.3) križanke se je pojavil pričakovani problem z gesli, ki potekajo ob robu (označeno s črtkano črto na slikovni križanki) čez celotno dolžino križanke. Taka gesla so zelo dolga, posledično jih je zelo malo in se v križanko

vstavijo na začetku, zato je nujno potrebno večnivojsko vračanje, da se jih po potrebi lahko zamenja. Če je križanka brez dodatnega črnega kvadratka (slikovna križanka), je pri sestavljanju še večji problem, kar je razvidno iz tabele 6.2, ker sta pri minimalni dolžini gesla 3, zelo problematični kar dve vrstici.

Na primeru križanke 5x5 (sl. A.4), kjer sem primerjal načina sestavljanja s fiksnimi črnimi kvadrati in z dodajanjem črnih kvadratkov med potekom sestavljanja, pričakovno ni bilo večjih razlik. Edina razlika je bilo malo več neuspešnih poskusov pri izbiri brez večnivojskega vračanja in brez minimalnega ustreznega mesta, kar pa je posledica povečanja povprečne dolžine besed. To lahko to opazimo v tabeli 6.6 in 6.5, le pri minimalni dolžini gesla 2, pri minimalni dolžini gesla 3 je povprečna dolžina enaka, zato razlike ni.

Pri problematičnih izbirah (brez večnivojskega vračanja oz. neizbira kvadratka z minimalnim številom ustreznih besed) so rezultati velike križanke s fiksnimi črnimi kvadrati (tabela 6.7) primerljivi s skandinavsko križanko (tabela 6.3), pri lažjih izbirah pa je križanka s fiksnimi črnimi kvadrati uspešno sestavljena v večjem številu primerov, predvsem zaradi problemov na začetku sestavljanja, ki sem jih že navdel.

Na splošno gledano je križanka zelo hitro sestavljena ali pa sploh ni sestavljena. Pri omejitvi minimalne dolžine gesla na 2 je povprečni čas sestavljanja nekaj sekund, pri minimalni dolžini 3 pa nekaj deset sekund. To velja za križanke dimenzij med 15x15 in 20x20.

Omejitev minimalne povprečne dolžine gesla 4,7 lahko dosežem le pri minimalni dolžini gesla 3. V posameznih primerih se takšna povprečna dolžina gesla doseže pri minimalni dolžini gesla 2, v povprečju pa ne.

Poglavje 7

Sklep

7.1 Povzetek

V diplomskem delu mi je uspelo doseči vse pomembne cilje, ki sem si jih zastavil.

Seznanil sem se s pristopi drugih raziskovalcev, z malo zastarelim pristopom vstavljanja posamezne črke in novejšim pristopom vstavljanja posamezne besede, na katerem temelji tudi moje raziskovanje. Osvojil sem teoretično ozadje preiskovanja prostora s hevrističnimi metodami. Uspel sem pridobiti dovolj veliko bazo slovenskih gesel in vpeljati indeksiranje za hitro iskanje po bazi gesel. Razvil sem algoritem za sestavljanje križank, ki je uspel sestaviti križanko tako, da je zadostila omejitvam iz podpoglavja 3.2. Podrobno sem preučil področje sestavljanja križank, in tako prišel do nekaterih idej za izboljšave pri sestavljanju križank, ki jih bom na kratko opisal v naslednjem podpoglavju.

7.2 Možnosti izboljšav

Možne izboljšave bom ločil na tri vrste:

- vsebinske (izbira dobrih gesel),
- glede na večjo povprečno dolžino gesel,
- glede na hitrost sestavljanja križanke.

Pri vsebinskih izboljšavah je potrebno zagotoviti, da se uporabi čim več gesel, ki so glede na različne kvalifikatorje najbolj zaželena. Dobro je tudi,

da število manj zelenih gesel iz določenega področja ne izstopa po številu. Izogibati se je potrebno križanju dveh težjih gesel.

Če želimo izbrati geslo glede na vsebinski kriterij, pridemo včasih v navzkrižje z kriterijem povprečne dolžine gesel. Potrebno je ustrezno obtežiti izbiro po vsakem izmed njiju in na ta način določiti geslo.

Pri povprečni dolžini besede je vedno potrebno pogledati celotno situacijo, ki je med dvema obstoječima črnima kvadratkoma, ne samo stanje pri enem kvadratu. V tem primeru je potrebno presoditi ali je smiselno vmesni prostor razdeliti na dva ali več delov. Pomembno je tudi na katerem mestu opravimo delitev, da bo ta prinesla čim manj delitev v prihodnje. Dobra možnost je tudi, da najprej poskusimo z gesli, na več različnih položajih, ki ne povzročijo dodatne delitve, šele nato pa ostale besede iz minimalnega položaja.

Izbira daljših besed v negativnem smislu vpliva na hitrost, ker je več križanj in tako več možnosti za nemogoče kombinacije.

Kot sem že ugotovil, je za hitrost sestavljanja križank ključnega pomena večnivojsko vračanje, predvsem v katerem primeru in kam se vračati.

Lahko si pomagamo z vodenjem seznama preizkušenih besed. Pri preiskovanju se včasih pojavi zaporedje določenih besed, ki se ponavljajo na različnih nivojih sestavljanja križanke. Na več različnih načinov lahko pridemo v nek podprostor, kjer ni mogoče rešiti križanke. Potrebno je detektirati omenjeno situacijo in se vrniti na nivo, kjer se prvič pojavi takšno ponavljanje.

Dodatek A

Predloge križank za testiranje algoritma

| | | | | | | | | | | | | | | |
|----|-------|-------|----|----|----|-----|-------|-----|-------|----|----|-----|-------|--|
| | ■ | OP | OP | OP | OP | OP | OP | ■ | OP | OP | OP | OP | OP | |
| | OP | | | | | | | OP | | | | | | |
| | OP | ----- | | | | | | OP* | ----- | | | | | |
| | OP | | | | | | | | | | | | | |
| | OP | | | | | | | | | | | | | |
| | OP | | | | | | | | | | | | | |
| | OP | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | |
| ■ | OP | OP | OP | OP | OP | | | | | | | OP* | OP | |
| OP | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | |
| OP | ----- | | | | | OP* | ----- | | | | | OP* | ----- | |
| OP | | | | | | OP | | | | | OP | | | |

Slika A.1: Predloga za sestavljanje slikovne križanke.

| | | | | | | | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| ■ | OP | OP | OP | OP | OP | OP | OP | OP | ■ | OP | OP | OP | OP | OP | OP | OP | ■ | OP | OP | OP | OP |
| OP | | | | | | | | | OP | | | | | | | | OP | | | | |
| OP | | | | | | | | | OP | | | | | | | | OP | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| ■ | OP | OP | | | | | | | | | | | | | | | | | | OP | OP |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | | |
| ■ | OP | OP | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | OP | OP |
| OP | | | | | | | | | | | | | | | | | | | | | |
| OP | | | | | | | | | OP | | | | | | | | OP | | | | |
| OP | | | | | | | | | OP | | | | | | | | OP | | | | |

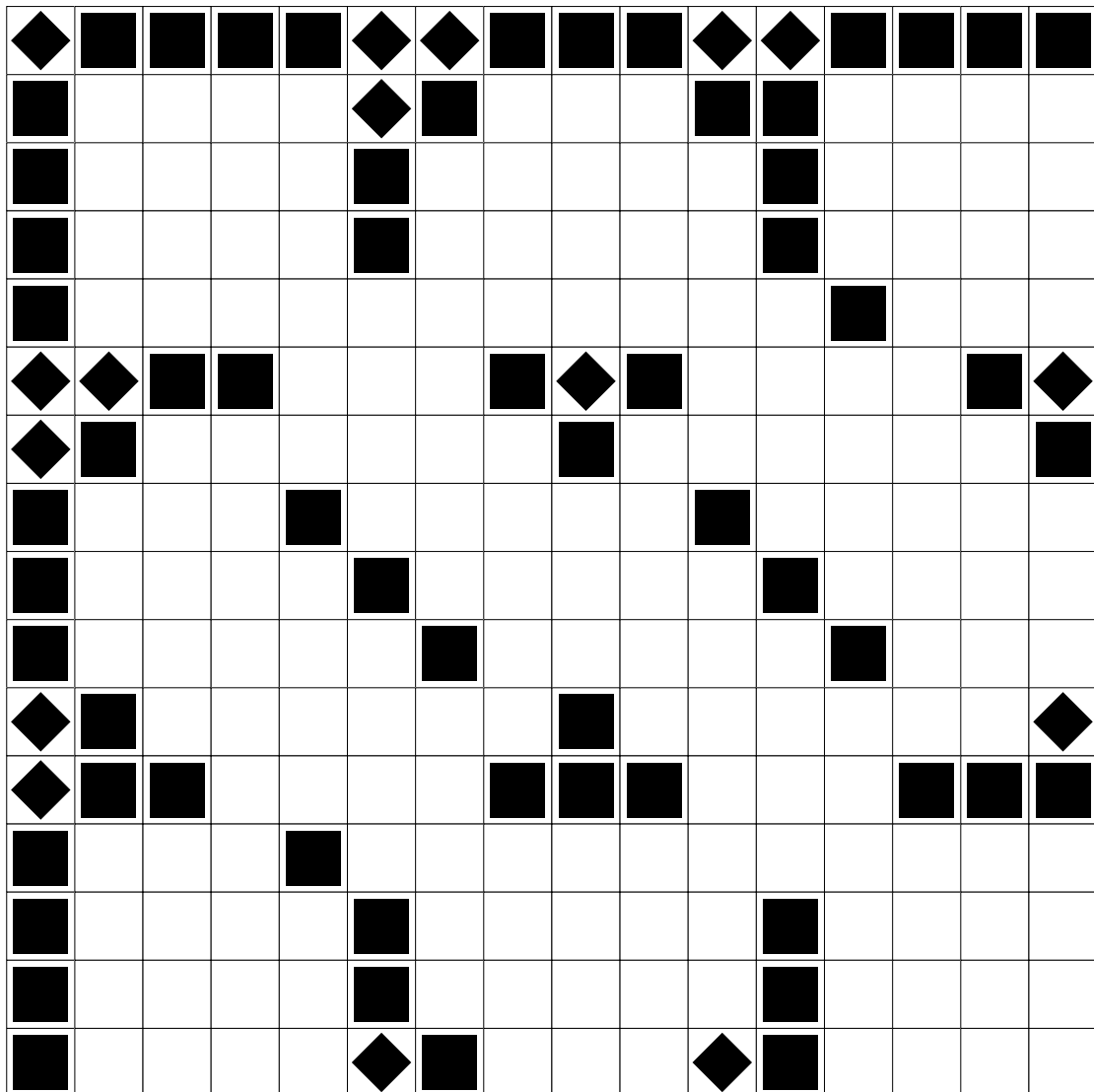
Slika A.2: Predloga za sestavljanje skandinavske križanke.

| | | | | | | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| ■ | OP | OP | OP | OP | OP | OP | OP | OP | ■ | OP | OP | OP | OP | OP | ■ | OP | OP | OP | OP | OP |
| OP | | | | | | | | | OP | | | | | | OP | | | | | |
| OP | | | | | | | | | OP | | | | | | | | | | | |
| OP | | | | | | | | | OP | | | | | | | | | | | |
| OP | | | | | | | | | | | | | | | | | | | | |
| OP | Z | E | L | O | OP | | | | | | | | | | | | | | | |
| OP | | | | | Z | | | | | | | | | | | | | | | |
| OP | | | | | A | | | | | | | | | | | | | | | |
| OP | | | | | P | | | | | | | | | | | | OP | OP | OP | |
| ■ | OP | OP | OP | | L | | | | | | | | | | | | | | | |
| OP | | | | | E | | | | | | | | | | | | | | | |
| OP | | | | | T | | | | | | | | | | | | | | | |
| OP | | | | | E | | | | | | | | | | | | | | | |
| OP | | | | | N | | | | | | | | | | | | | | | |
| OP | | | | | A | OP | K | R | I | Ž | A | N | K | A | OP | | | | | OP |
| OP | | | | | OP | | | | | | | | OP | | | | | | | |
| OP | | | | | OP | | | | | | | | OP | | | | | | | |
| OP | | | | | OP | | | | | | | | OP | | | | | | | |

Slika A.3: Predloga za sestavljanje sponzorske križanke.

| | | | | | |
|----|----|----|----|----|----|
| ■ | OP | OP | OP | OP | OP |
| OP | | | | | |
| OP | | | | | |
| OP | | | | | |
| OP | | | | | |
| OP | | | | | |

Slika A.4: Predloga za sestavljanje križanke 5x5.



Slika A.5: Predloga za sestavljanje velike križanke s fiksnimi črnimi kvadrati.

Slike

| | | |
|-----|--|----|
| 1.1 | Mreža in običajna križanka. | 6 |
| 2.1 | Opisni in črni kvadrati. | 8 |
| 2.2 | Križanka iz sredine. | 10 |
| 2.3 | Koordinatna križanka. | 11 |
| 2.4 | Slikovna križanka. | 12 |
| 2.5 | Verižna križanka. | 13 |
| 2.6 | Kombinacijska križanka. | 14 |
| 2.7 | Zlogovna križanka. | 15 |
| 2.8 | Programa Sfinga. | 17 |
| 4.1 | Poizvedbe za gesla. | 24 |
| 4.2 | Indeksiranje gesel. | 29 |
| 5.1 | Začetna predpriprava križanke. | 30 |
| 5.2 | Omejitve po vstavitvi besede. | 32 |
| 5.3 | Daljšanje besed. | 33 |
| 5.4 | Večnivojsko vračnje. | 36 |
| A.1 | Predloga za sestavljanje slikovne križanke. | 46 |
| A.2 | Predloga za sestavljanje skandinavske križanke. | 47 |
| A.3 | Predloga za sestavljanje sponzorske križanke. | 48 |
| A.4 | Predloga za sestavljanje križanke 5x5. | 48 |
| A.5 | Predloga za sestavljanje velike križanke s fiksnimi črnimi kvadrati. | 49 |

Tabele

| | | |
|-----|---|----|
| 4.1 | Pogostnost in ponavljanje posameznih črk. | 26 |
| 4.2 | Število besed, ki ima na določenem mestu določeno črko. | 27 |
| 4.3 | Pogostnost zaporedja dveh črk. | 28 |
| 6.1 | Sestavljanje slikovne križanke 1. | 38 |
| 6.2 | Sestavljanje slikovne križanke 2. | 39 |
| 6.3 | Sestavljanje skandinavske križanke. | 39 |
| 6.4 | Sestavljanje sponzorske križanke. | 40 |
| 6.5 | Sestavljanje križanke 5x5 1. | 40 |
| 6.6 | Sestavljanje križanke 5x5 2. | 41 |
| 6.7 | Sestavljanje križanke s fiksnimi črnimi kvadrati. | 41 |

Literatura

- [1] L. J. Mazlack, Computer construction of crossword puzzles using precedence relationships, *Artificial Intelligence*, št. 7, zv. 1, str. 1-19, 1976.
- [2] Revija Kih, *Delo revije*, št. 890, oktober 2008.
- [3] I. Kononenko, Načrtovnje podatkovnih struktur in algoritmov, *Fakulteta za računalništvo in informatiko, Ljubljana*, 2. popravljena in doplonjena izdaja, 1999.
- [4] Enigmatko.com, Dostopno na:
<http://www.enigmatko.com/tekmovanje/pravila.html>
- [5] Slovarji.com, Dostopno na:
<http://www.slovarji.com/>
- [6] M. L. Ginsberg, Search lessons learned from crossword puzzles, *In Proceedings of the Eighth National Conference on Artificial Intelligence*, 1990
- [7] Crossword compiler, Dostopno na:
<http://www.crossword-compiler.com/>
- [8] Custom Crossword Generator, Dostopno na:
<http://pdos.csail.mit.edu/cgi-bin/theme-cword>
- [9] A. Aherne, C. Vogel, Wordnet Enhanced Automatic Crossword Generation, *University of Ireland, Dublin*, Computational Linguistics Group
- [10] E. A. Gomez, J. M. Scher, Design Strategies for the Pedagogical Use of Crossword Puzzle Generation Software, In Individual and Collaborative Design Modes *New Jersey Institute of Technology* Department of Information Systems, College of Computing Sciences
- [11] Citadela - Atlas, Dostopno na:
<http://www.citadela.si/content/view/4/37/lang,si/>