

МАТЕМАТИЧКА ГИМНАЗИЈА



МАТУРСКИ РАД

ИЗ РАЧУНАРСТВА И ИНФОРМАТИКЕ

Решавање игре *Речко*

Аутор:

Јанко Поповић, 4д

Ментор:

проф. Милош Арсић

Београд, јун 2025.

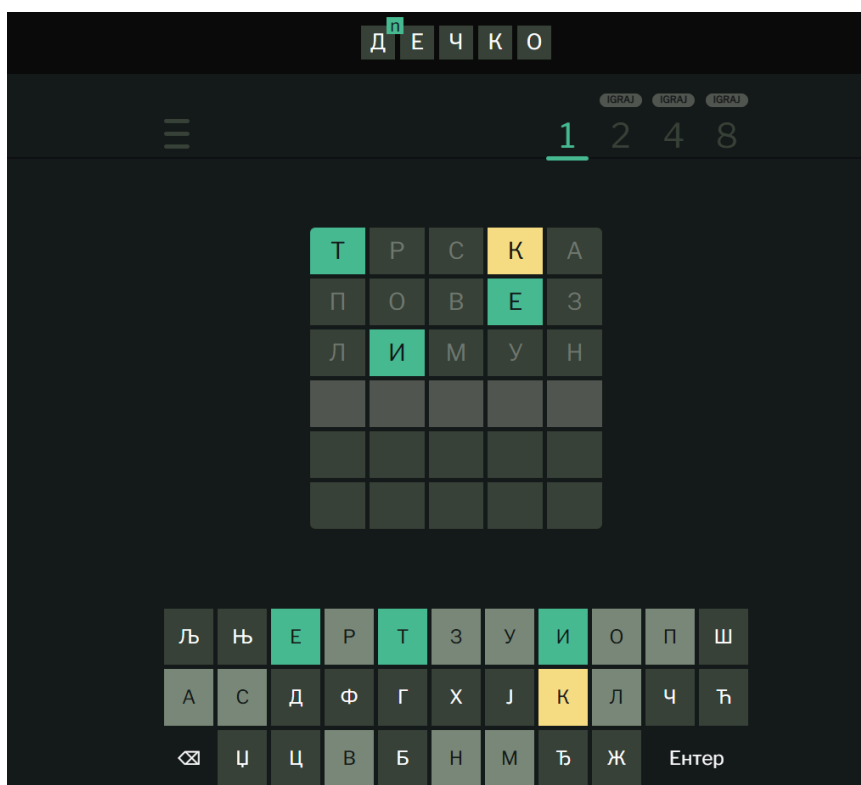
0 Садржај

0 Садржај.....	1
1 Увод.....	2
1.1 О игри Речко.....	2
1.2 Правила игре.....	2
2 Теорија информација.....	4
3 Технички детаљи.....	5
3.1 Формирање речника.....	5
3.2 Избор сајта.....	5
3.3 Решавање Речка.....	6
4 Резултати.....	7
4.1 Статистика.....	7
4.2 Најбоље речи.....	11
4.3 Најбоље комбинације.....	12
5 Закључак.....	14
6 Литература.....	15

1 Увод

1.1 О игри *Речко*

Речко је популарна игра која представља српску верзију познате светске игре *Wordle*. Циљ игре је погодити тачну реч од пет слова у ограниченом броју покушаја. У овом раду посебно ћу се бавити процесом решавања *Речка*, анализирајући стратегије и методе које помажу у успешној игри односно проналажењу решења у што мањем броју корака.



Слика 1.1: Интерфејс *Речка*

1.2 Правила игре

Циљ игре је погодити непознату реч у одређеном, унапред дефинисаном броју покушаја. Обе постојеће верзије *Речка* подржавају само петословне речи, па ће само

тај случај бити разматран у оквиру овог рада. Сваки покушај мора бити валидна реч, што значи да треба да припада неком унапред дефинисаном, али играчу не експлицитно познатом списку. Списак би требало да одговара речнику одговарајућег језика, у овом случају српског.

Након што играч укуца реч и притисне *Enter* свако слово обојиће се у једну од три боје - зелену (■), жуту (■) или сиву (■). Зелена боја означава да се одговарајуће слово налази на исправном месту, жута да слово постоји у речи али не на испробаном месту, а сива значи да тог слова нема у речи.

Посебно, ако се у испробаној речи неко слово јавља више пута, биће обојено жуто или зелено онолико пута колико се јавља у непознатој речи. Наиме, ако је скривена реч *књига*, а покушај *ратар*, резултат ће бити ■■■■■, а не ■■■■■, зато што се слово *а* појављује само једном у речи *књига*. Увек ће бити обојено оно слово које се у покушају јавља раније (односно на позицији са мањим радним бројем). Да је покушај био *ватра*, резултат би био ■■■■■, а не ■■■■■. Обратите пажњу на то да је у овом случају обојено касније слово *а*. Зелена слова имају предност како би играчу било дато више информација.

2 Теорија информација

За потребе овог рада није неопходно подробно познавање теорије информација. Биће коришћена само формула за ентропију H (облици (1) или (2)), при чему наравно важи услов (3), да је збир вероватноћа свих исхода једнак 1.

$$H(p_1, \dots, p_n) = \sum_{i=1}^n p_i \log_2 \left(\frac{1}{p_i} \right) \quad (1)$$

$$H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (2)$$

$$\sum_{i=1}^n p_i = 1 \quad (3)$$

$$S = k_B \ln \Omega \quad (4)$$

Занимљиво је приметити сличност формуле (2) са формулом за термодинамичку ентропију (4), која је и инспирисала назив ентропија за овај појам из теорије информација.

Информациона ентропија се изражава у *битовима*.

3 Технички детаљи

3.1 Формирање речника

За решавање *Wordle* најлакше и најсмисленије је користити јавно доступну листу која броји 12000 речи које се могу играти. Од тога постоји и посебна листа 1200 пробраних речи које могу бити решење.

У случају *Речка* не постоји овакав луксуз. Почетна идеја била је коришћење *Речника* Матице српске [5], као званичног извора речи. Како је он доступан у *pdf* формату, теоретски би било могуће извући листу из овог речника. У покушају да се то уради, дошло је до низа проблема од којих је главни то што су, због формата речника, речи записане специјалним карактерима па више речи може дати исти запис, па се самим тим оне не могу реконструисати.

Зато је, уместо [5], коришћена јавно доступна листа речи [3]. Ова листа је иницијално садржала преко 2 милиона речи. Значајан део њих се не може сматрати стандардним српским језиком или представља речи које нису у основном облику. Листа је профилирана и добијено је око 60 хиљада петословних речи, од којих је, укуцавањем сваке од њих у *Речка* преостало 18.047 речи које се могу играти. С обзиром на опширност почетног речника, извесно је да је овај број близак укупном броју могућих речи. Ако је аутор *Речка* користио исти извор [3] или неки његов подскуп, што је врло вероватан случај, онда пронађена листа одговара потпуном списку. У тренутку писања овог рада није пронађена ни једна реч која се може уписати у *Речка* а да није део листе.

3.2 Избор сајта

За *Речко* постоје два сајта - [1] и [2]. На оба је игра у основи иста - имамо 6 покушаја да погодимо једну реч. Сајт [1] такође има могућност играња дуплог *Речка*, тзв. *Дечка*, као и четвороструког и осмоструког. Додатно правило дуплог речка је то да две скривене речи морају имати неку међусобну везу нпр. *бајка* и *принц*. Још једна предност сајта [1] је што се може утврдити да ли је реч валидна само укуцавањем

исте, без притискања *Enter*, што је значајно убрзало процес проверавања свих 60 хиљада речи.

Због свих наведених погодности, [1] је изабран као сајт на коме ће бити решен *Речко*.

3.3 Решавање *Речка*

Идеалан решавач би био онај који гарантује најмањи број корака за сваку почетну реч. Како би тако нешто захтевало претраживање огромног броја опција, тиме се у овом раду нећемо бавити. Уместо тога, урађене су три ствари које је аутор након свог великог искуства са *Речком* сматрао најкориснијим за његово решавање.

Први програм за дату реч и одговор на њу (комбинација ,  и ) враћа број и листу могућих речи после сваког корака.

Други програм рачуна ентропију сваке речи тј. колико свака реч доприноси решавању.

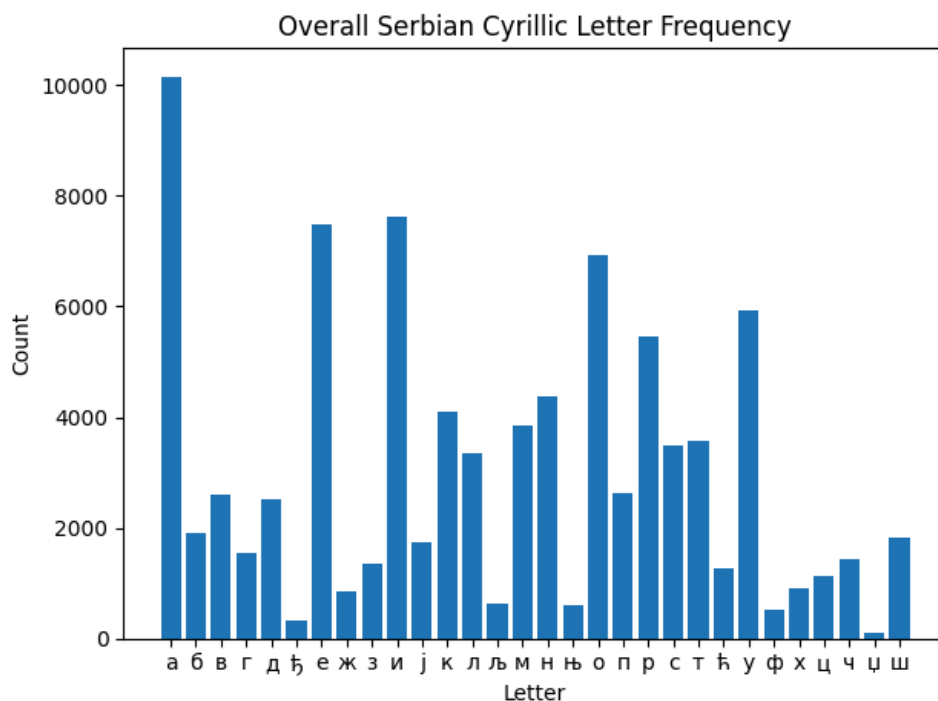
Трећи за дати комбинацију речи рачуна њихову збирну ентропију (која није збир ентропија!).

4 Резултати

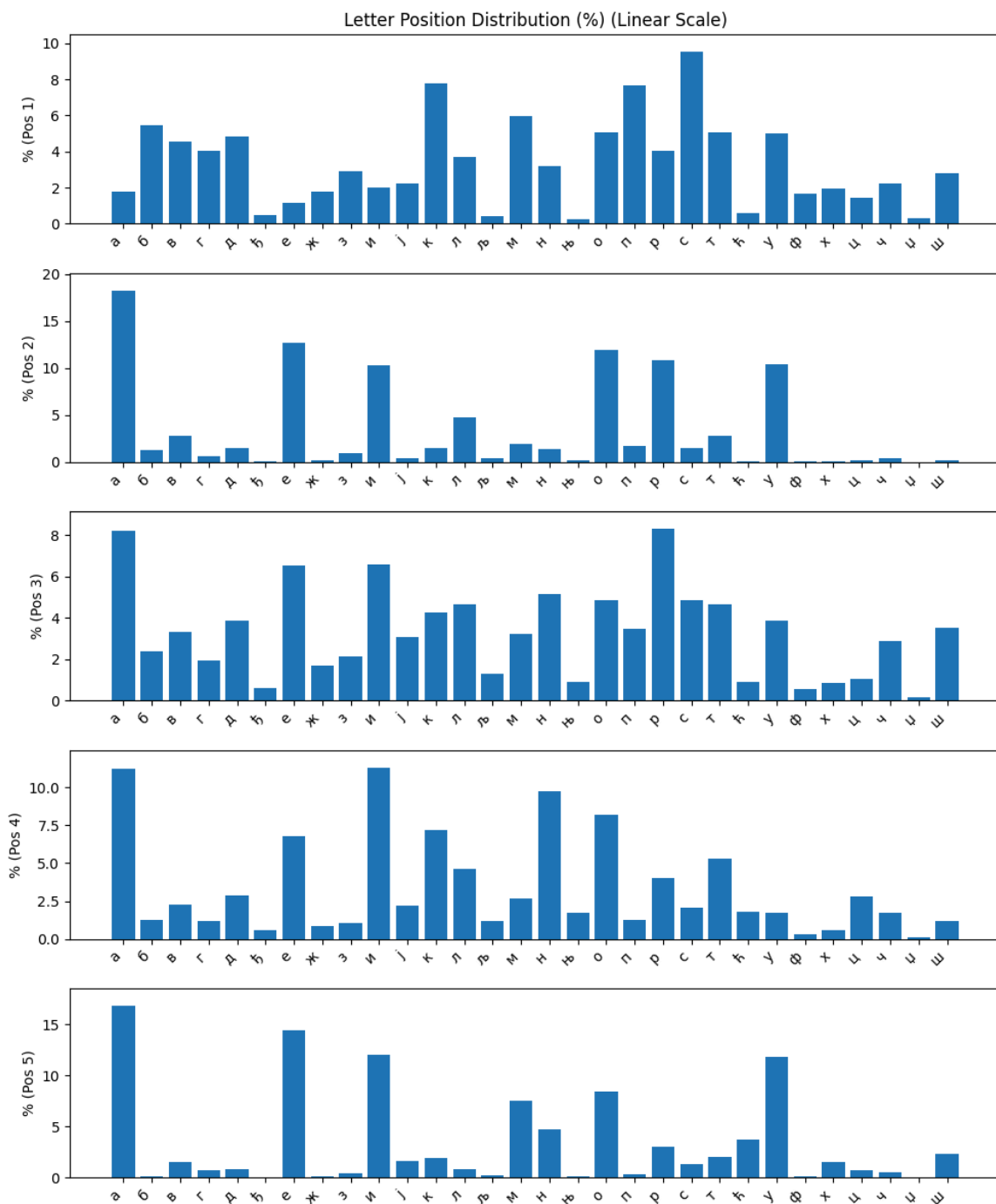
4.1 Статистика

На Слици 4.1 је приказана укупна расподела. Може се приметити да су самогласници најчешћи редом: *а, и, е, о, у*. Међу сугласницима су најзаступљенији *р, н, к, м, т*.

На Слици 4.2 расподела учесталости по свакој позицији у речи. Може се уочити да позиције 2 и 5 и позиције 1 и 3 имају квалитативно сличне расподеле. Такође, занимљиво је да је слово *с* најчешће почетно слово, и да је то једина позиција на којој самогласници нису доминантни.

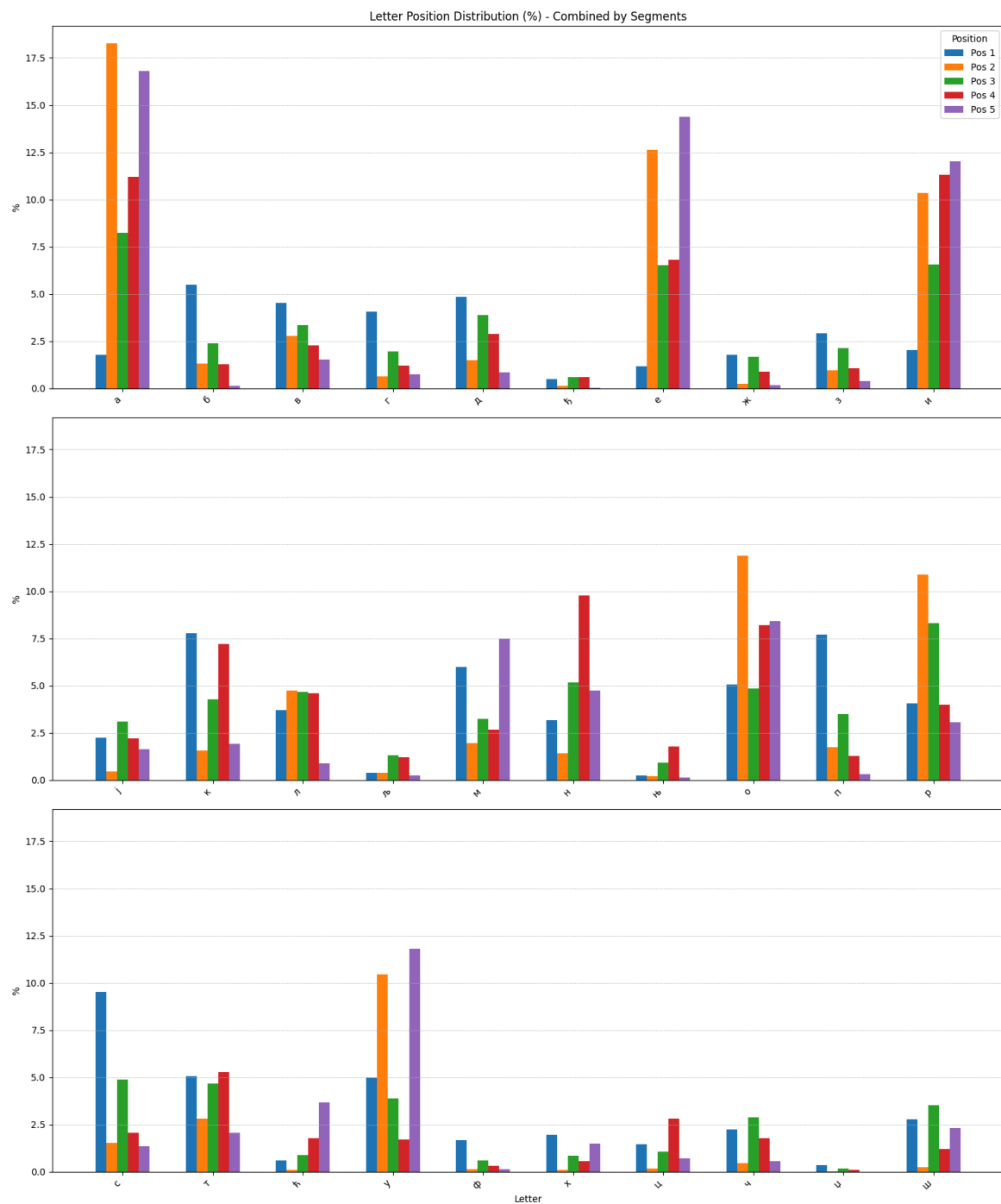


Слика 4.1: Расподела учесталости слова

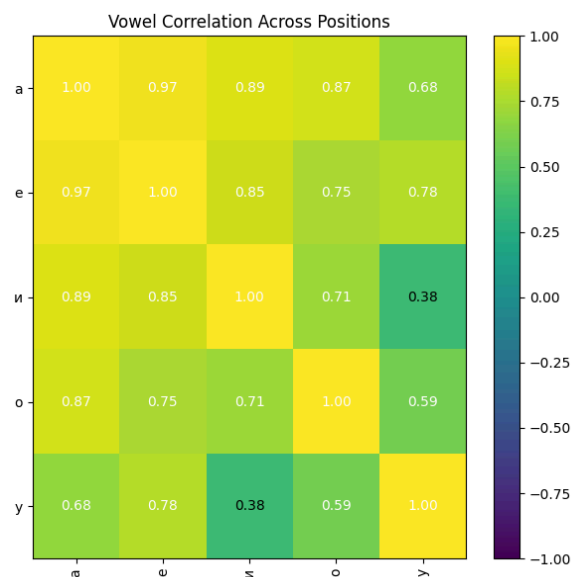


Слика 4.2: Расподела учесталости слова на свакој од позиција у речи

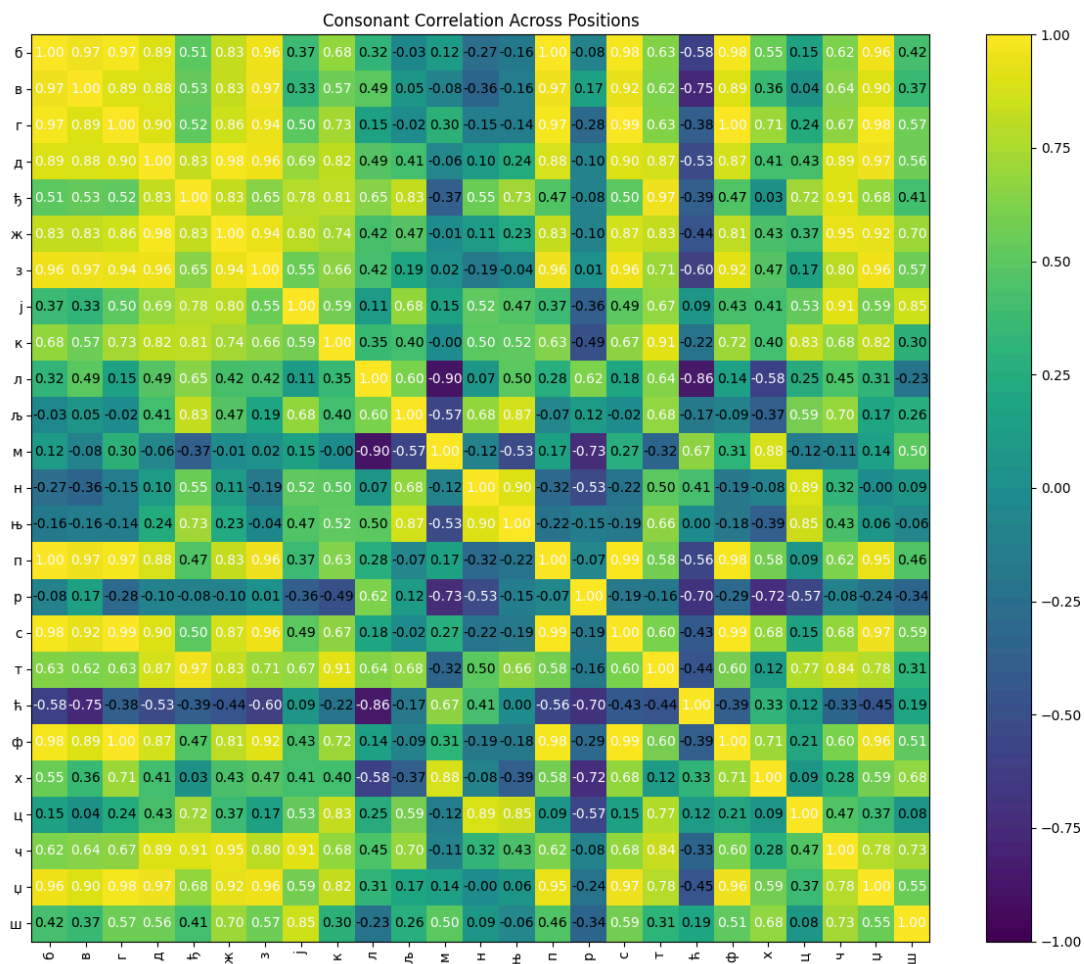
Такође је занимљиво погледати графике корелације међу сугласницима и међу самогласницима, где је свако слово представљено петочланим вектором његове дистрибуције. Слика 4.3 приказује расподелу по позицијама за свако слово, а слике 4.4 и 4.5 корелације међу самогласницима и сугласницима, редом.



Слика 4.3: Расподела учесталости за свако слово



Слика 4.4: Корелација самогласника



Слика 4.5: Корелација сугласника

4.2 Најбоље речи

Ентропија целог речника је бинарни логаритам од 18.047 односно 14,14. При оваквој дефиницији, смањење ентропије за један бит одговара половљењу могућих решења.

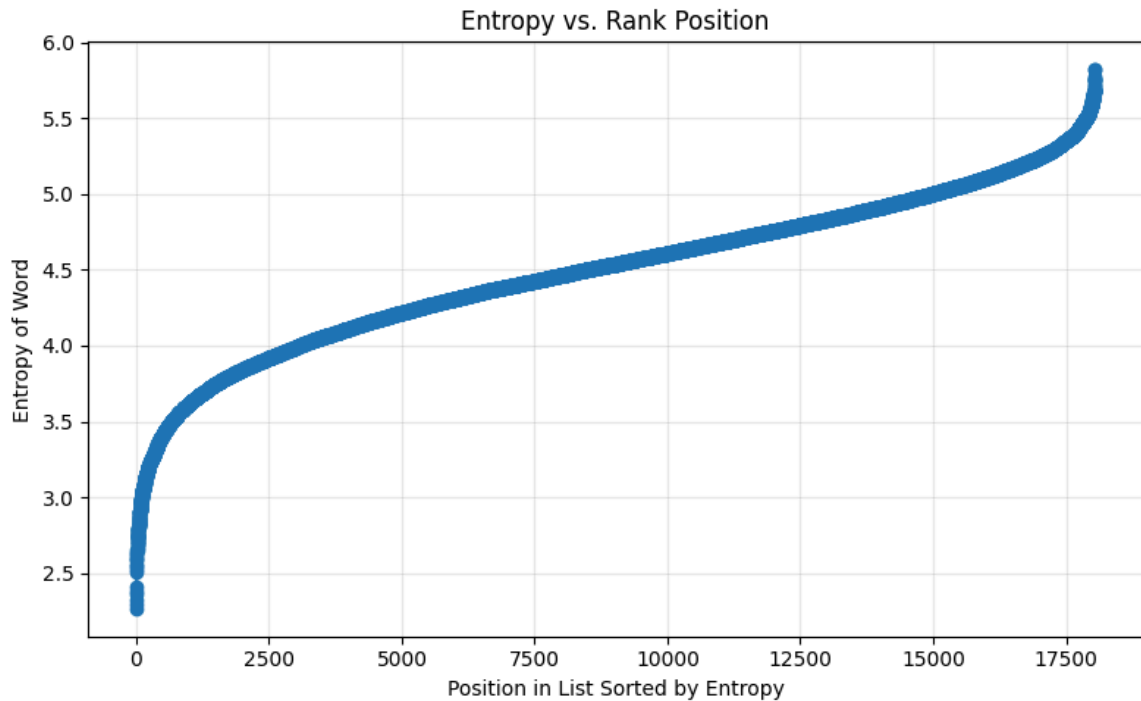
Следећа табела приказује најбоље и најгоре речи заједно са њиховим ентропијама. У последње две колоне су узете у обзир само смислене речи у основном облику - у загради је дата позиција у укупној листи.

редни број	најбоље	најгоре	најбоље*	најгоре*
1	кореа 5,825	жежељ 2,262	орати (75) 5,579	уздуж (18) 2,668
2	марие 5,821	рсфср 2,290	круна (118) 5,532	мњење (48) 2,826
3	марио 5,770	изиђи 2,319	рента (141) 5,515	зулуф (55) 2,842

Табела 4.1: Најбоље и најгоре речи

Може се приметити да су смислене речи доста ретке односно лоше рангиране и на листи најбољих и на листи најгорих, али из различитих разлога. За најбоље речи је неопходан велик број самогласника што лакше постижу речи које нису у основном облику, а за најгору реч је потребан што већи број ретких слова што лакше постижу стране и чудне речи.

Такође је приложена Слика 4.6 - график зависности ентропије од редног броја речи.



Слика 4.6: График зависности ентропије од редног броја речи

4.3 Најбоље комбинације

То што је једна реч сама по себи најбоља не значи да ће она заједно са следећом речју највише смањити ентропију. У Табели 4.2 су поређене вредности ентропије за неке комбинације 2 до 5 речи састављених од честих, различитих слова.

р.бр.	реч 1	реч 2	реч 3	реч 4	реч 5	ентропија
1	насео	трули				9,7189
2	трска	повез				9,0383
3	трска	лимун				9,5458
4	лимун	повез				8,6044
5	амери	конту				9,6215
6	трска	повез	лимун			12,2265
7	авети	дрмну	склоп			12,2441
8	чајлд	губиш	првом	некст		13,4032
9	жуњић	хајнц	зглоб	кемпф	чврст	13,6629

Табела 4.2: Збирне ентропије

5 Закључак

Из ауторовог емпиријског искуства, овај *Решавач Речка*, ако се исправно користи, побољшава резултат за око један корак што је значајан бољитак.

Могуће унапређење резултата могло би се постићи прављењем листе могућих решења. Од 18 хиљада признатих речи мали је број оних које би могле да се нађу као решење. Ово побољшање би се могло постићи проналажењем неког званичног речника и повећавањем вероватноће свих речи које му припадају при рачунању ентропије.

Такође, минимално унапређење кода могло би бити да уместо да програм препоручи насумичну могућу реч, препоручује ону са највећом почетном ентропијом.

Хтео бих да напоменем да, иако је овај решавач нешто што сам јако дуго хтео да направим, игра постаје зачуђујуће досадна уз његову помоћ.

За крај бих волео да се захвалим свом ментору, професору Милошу Арсићу, на указаном бескрајном стрпљењу при изради овог рада.

6 Литература

[1] Дупли *Речко* - <https://www.hangman.rs/dechko/>

приступљено 30. маја 2025.

[2] *Речко* - <https://rechko.com/>

приступљено 30. маја 2025.

[3] Коришћен списак речи - <https://github.com/turanjanin/spisak-srpskih-reci>

приступљено 26. маја 2025.

[4] Решавање *Wordle* - <https://www.youtube.com/watch?v=v68zYyaEmEA>

приступљено 25. маја 2025.

[5] М. Вујанић et al, *Речник српскога језика*, Матица српска, Нови Сад 2011.