УНИВЕРЗИТЕТ „ГОЦЕ ДЕЛЧЕВ” - ШТИП
ФАКУЛТЕТ ЗА ИНФОРМАТИКА

ISSN 1857- 8691

ГОДИШЕН ЗБОРНИК
2012
YEARBOOK
2012

ГОДИНА 1 **VOLUME I**

GOCE DELCEV UNIVERSITY - STIP
FACULTY OF COMPUTER SCIENCE

УНИВЕРЗИТЕТ „ГОЦЕ ДЕЛЧЕВ“ – ШТИП
ФАКУЛТЕТ ЗА ИНФОРМАТИКА

ГОДИШЕН ЗБОРНИК
2012
YEARBOOK
2012

ГОДИНА 1 МАРТ, 2013 VOLUME I

GOCE DELCEV UNIVERSITY – STIP
FACULTY OF COMPUTER SCIENCE

ГОДИШЕН ЗБОРНИК
ФАКУЛТЕТ ЗА ИНФОРМАТИКА
YEARBOOK
FACULTY OF COMPUTER SCIENCE

За издавачот:
Проф д-р Владо Гичев

Издавачки совет / **Editorial board**
Проф. д-р Саша Митрев — Prof. Saša Mitrev, Ph.D
Проф. д-р Лилјана Колева - Гудева — Prof. Liljana Koleva - Gudeva, Ph.D.
Проф. д-р Владо Гичев — Prof. Vlado Gicev, Ph.D.
Проф. д-р Цвета Мартиновска — Prof. Cveta Martinovska, Ph.D.
Проф. д-р Татајана Атанасова - Пачемска — Prof. Tatjana Atanasova - Pacemska, Ph.D.
Доц. д-р Зоран Здравев — Ass. Prof. Zoran Zdravev, Ph.D.
Доц. д-р Александра Милева — Ass. Prof. Aleksandra Mileva, Ph.D.
Доц. д-р Сашо Коцески — Ass. Prof. Saso Koceski, Ph.D.
Доц. д-р Наташа Коцеска — Ass. Prof. Natasa Koceska, Ph.D.
Доц. д-р Зоран Утковски — Ass. Prof. Zoran Utkovski, Ph.D.
Доц. д-р Игор Стојановиќ — Ass. Prof. Igor Stojanovik, Ph.D.
Доц. д-р Благој Делипетров — Ass. Prof. Blagoj Delipetrov, Ph.D.

Редакциски одбор / **Editorial staff**
Проф. д-р Цвета Мартиновска — Prof. Cveta Martinovska, Ph.D.
Проф. д-р Татајана Атанасова - Пачемска — Prof. Tatjana Atanasova - Pacemska, Ph.D.
Доц. д-р Наташа Коцеска — Ass. Prof. Natasa Koceska, Ph.D.
Доц. д-р Зоран Утковски — Ass. Prof. Zoran Utkovski, Ph.D.
Доц. д-р Игор Стојановиќ — Ass. Prof. Igor Stojanovik, Ph.D.
Доц. д-р Александра Милева — Ass. Prof. Aleksandra Mileva, Ph.D.
Доц. д-р Зоран Здравев — Ass. Prof. Zoran Zdravev, Ph.D.

Главен и одговорен уредник / **Managing/ Editor in chief**
Доц. д-р Зоран Здравев — Ass. Prof. Zoran Zdravev, Ph.D.

Јазично уредување / **Language editor**
Даница Гавриловаска - Атанасовска (македонски јазик) — Danica Gavrilovska-Atanasovska (macedonian language)
Павлинка Павлова-Митева (англиски јазик) — Pavlinka Pavlova-Miteva (english language)

Техничко уредување / **Technical editor**
Славе Димитров — Slave Dimitrov
Благој Михов — Blagoj Mihov

Редакција и администрација
Универзитет „Гоце Делчев“-Штип
Факултет за информатика
ул. „Крсте Мисирков“ 10-А
п. фах 201, 2000 Штип
Р. Македонија

Address of the editorial office
Goce Delcev University – Stip
Faculty of Computer Science
Krste Misirkov 10-A
PO box 201, 2000 Štip,
R. of Macedonia

СОДРЖИНА
CONTENT

DEVELOPING CLOUD COMPUTING'S NOVEL COMPUTATIONAL METHODS FOR IMPROVING LONG-TERM WEATHER GLOBAL FORECAST
Zubov Dmytro 7

PERVASIVE ALERT SYSTEM FOR FALL DETECTION BASED ON MOBILE PHONES
Kire Serafimov, Natasa Koceska 17

ESTABLISHEMENT OF A HEALTHCARE INFORMATION SYSTEM
Alexandar Kostadinovski, Drasko Atanasoski 26

TIME COMPLEXITY IMPROVEMENT OF THE FIRST PROCESSING STAGE OF THE INTELLIGENT CLUSTERING
Done Stojanov, Cveta Martinovska 36

MOODLE AS A TEACHING TOOLS IN MATHEMATICS-CASE STUDY IN UNIVERSITY "GOCE DELCEV" STIP
Tatjana Atanasova-Pacemska, Sanja Pacemska, Biljana Zlatanovska 45

TOURISM RECOMMENDATION SYSTEMS: ANALYTICAL APPROACH
Biljana Petrevska, Marija Pupinoska-Gogova, Zoran Stamenov 57

CLOUD COMPUTING APPLICATION FOR WATER RESOURCES MODELING AND OPTIMIZATION
Blagoj Delipetrev 66

IMPROVING THE SECURITY OF CLOUD-BASED ERP SYSTEMS
Gjorgji Gicev, Ivana Atanasova, Jovan Pehcevski 77

USING OF THE MOORE-PENROSE INVERSE MATRIX IN IMAGE RESTORATION
Igor Stojanovic, Predrag Stanimirovic, Marko Miladinovic 88

THE INFLUENCE OF THE BUSINESS INTELLIGENCE ON THE BUSINESS PERFORMANCE MANAGEMENT
Ljupco Davcev, Ana Ljubotenska 99

LINQ TO OBJECTS SUPPORTED JOINING DATA
Mariana Goranova 109

GLOBALIZATION, INFORMATION TECHNOLOGY AND NEW DIGITAL ECONOMIC LANDSCAPE
Riste Temjanovski 120

WEB БАЗИРАН СОФТВЕР ЗА SCADA АПЛИКАЦИИ INTEGRAXOR
Марјан Стоилов, Василија Шарац 130

SECURITY IN COMPUTER NETWORKS FROM THE PERSPECTIVE OF ACCESS CONTROL
Saso Gelev, Jasminka Sukarovska-Kostadinovska 139

FREQUENCY DISTRIBUTION OF LETTERS, BIGRAMS AND TRIGRAMS IN THE MACEDONIAN LANGUAGE
Aleksandra Mileva, Stojanče Panov, Vesna Dimitrova 149

TOWARDS A GENERIC METADATA MODELING
Pavel Saratchev 161

ECONOMIC VALUE OF INFORMATION SYSTEMS IN PRODUCTION PROCESSES
Aleksandar Krstev, Zoran Zdravev 175

TUNING PID CONTROLLING PARAMETERS FOR DC MOTOR SPEED REGULATION
Done Stojanov 185

COMPARISON OF THE PERFORMANCE OF THE ARTIFICIAL BOUNDARIES P3 AND P4 OF STACEY
Zoran Zlatev, Vasko Kokalanov, Aleksandra Risteska 192

CORRESPONDENCE BETWEEN ONE-PARAMETER GROUP OF LINEAR TRANSFORMATIONS AND LINEAR DIFFERENTIAL EQUATIONS THAT DESCRIBE DYNAMICAL SYSTEMS
Marija Miteva, Limonka Lazarova 200

THE BLACK-SCHOLES MODEL AND VALUATION OF THE EUROPEAN CALL OPTION
Limonka Lazarova, Marija Miteva, Natasa Stojkovik 209

BITCOIN SCHEMES- INOVATION OR A THREAT TO FINANCIAL STABILITY?
Violeta Madzova 221

JAVA IDEs FOR EASILY LEARNING AND UNDERSTANDING OBJECT ORIENTED PROGRAMMING
Aleksandra Stojanova, Natasha Stojkovic, Dusan Bikov 232

STUDENTS' KNOWLEDGE TEST CONTROL – METHODS AND RESULTS' INTERPRETATION
Ludmila Stoyanova, Daniela Minkovska 241

WEB SERVICE FOR AMBIGUOUS TRANSLITERATION OF FULL SENTENCES FROM LATIN TO CYRILLIC ALPHABET
Stojance Spasov, Zoran Zdravev 252
ON THE APPLICATION OF KEEDWELL CROSS INVERSE QUASIGROUP TO CRYPTOGRAPHY
Jaíyéọlá Tèmítọpé Gbọláhàn 264

FREQUENCY DISTRIBUTION OF LETTERS, BIGRAMS AND TRIGRAMS IN THE MACEDONIAN LANGUAGE

Aleksandra Mileva[1]*, Stojanče Panov[1], Vesna Dimitrova[2]

1Faculty of Computer Science, "Goce Delčev" University in Štip, Republic of Macedonia
e-mail: aleksandra.mileva@ugd.edu.mk, stojance.10139@student.ugd.edu.mk

2Faculty of Computer Science and Engineering, "Ss. Cyril and Methodius" University in Skopje, Republic of Macedonia
e-mail: vesna.dimitrova@finki.ukim.mk

Abstract: Frequency analysis in cryptanalysis is based on the fact that, in any given piece of written text, certain letters and combinations of two or three letters occur with varying frequencies. In this paper we present average frequency distribution of letters, bigrams and trigrams in the Macedonian language. Letter frequency of the most common first letter and last letter in words is also given. Our results are based on approximately 15000 pages of written text from the following subjects: poetry, prose, drama, natural sciences, social sciences, law, different laws, economy, and computer science. Obtained letter frequency sequence is "А О И Е Т Н Р С В Д К Л П М У З Ј Г Б Ч Ш Ц Ж Њ Ф Ќ Х Ѓ Џ Љ S", the most common letter pairs are "НА АТ ТА НИ ТЕ РА ОТ СТ ТО КО" and the most common trigrams are "ИТЕ АТА УВА ИЈА АЊЕ СТА ОСТ ВАЊ ПРО ПРЕ".

Keywords: frequency distribution of letters, bigrams, trigrams, Macedonian language.

1 Introduction

Frequency analysis is a cryptanalysis's tool for breaking classical ciphers. It studies frequencies of letters and group of letters in a given ciphertext. For example, if you have got a message encrypted using the simple substitution cipher that you want to break, you can use frequency analysis. Each letter of the plaintext is replaced with another, and any particular letter in the plaintext will always be transformed into the same letter in the ciphertext. You can still recognise the original letter, because the frequency characteristics of the original letter will be passed on the new letters. Usually, cryptanalyst may need to try several combinations of mappings between ciphertext and plaintext letters, before he/she discovers the original plaintext.

Letter frequencies are also important for design of some keyboard layouts like Dvorak Simplified Keyboard, for several games like Scrable, for data-compression techniques such as Huffman coding, etc.

Letter frequencies for the English can be found in several sources, like [1, 2], for German in [3], for Spanish in [4], for Italian in [5], etc. The absence of similar results for the Macedonian language was our basic motivation for this study.

2 Letter frequency distribution

In our experiments, we use approximately 15000 pages of written text, or more accurate: 2000 pages of prose and poetry, 1700 pages of drama, 2000 pages of natural sciences (geology, mining, geography, etc), 1400 pages of social sciences (history, pedagogy, etc), 1400 pages of law, 3000 pages of different laws, 1600 pages of economy and 2000 pages of computer science.

Obtained results for Macedonian letter frequency distribution are given in Table 1. Obtained relative frequencies of letters, ordered by Macedonian alphabet and by frequencies, are shown on Figure 1 and 2, respectively.

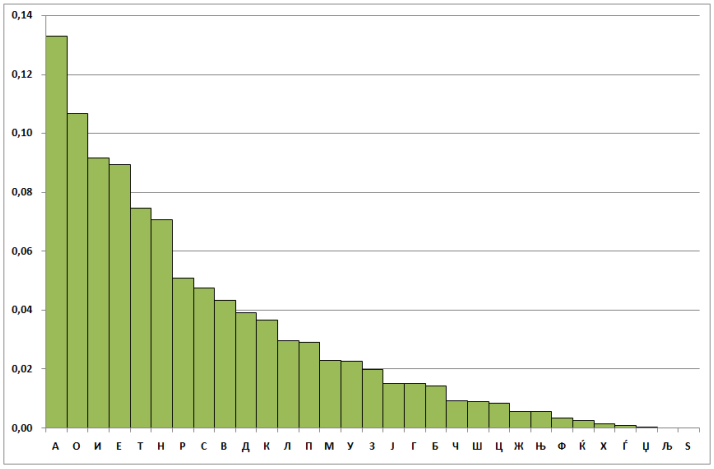

Figure 1 Relative frequencies of letters in text

Table 1 Macedonian letter frequency

Letter	Count	Frequency
А	3.819.909	13,293%
О	3.068.901	10,679%
И	2.632.308	9,160%
Е	2.570.604	8,945%
Т	2.144.354	7,462%
Н	2.033.735	7,077%
Р	1.462.718	5,090%
С	1.365.208	4,751%
В	1.247.904	4,343%
Д	1.123.847	3,911%
К	1.055.096	3,672%
Л	850.394	2,959%
П	840.787	2,926%
М	663.552	2,309%
У	649.953	2,262%
З	568.707	1,979%
Ј	434.240	1,511%
Г	432.547	1,505%
Б	412.019	1,434%
Ч	265.775	0,925%
Ш	263.335	0,916%
Ц	246.244	0,857%
Ж	163.334	0,568%

Њ	159.379	0,555%
Ф	99.687	0,347%
Ќ	75.905	0,264%
Х	41.978	0,146%
Ѓ	30.340	0,106%
Џ	7.086	0,025%
Љ	4.252	0,015%
S	2.708	0,009%

Note that, these frequencies are averages, that means that letter A will not always constitute 13,293% of all the letters in a text, and may not even be the most common letter. If you analyse just one sentence, its letter frequency distribution probably will not match the bar chart above. But if you pick a longer Macedonian text, the match should be surprisingly good.

By analysing Table 1, one can see that the “top twelve” letters comprise about 81,3% of the total usage and the “top eight” letters comprise about 66,5% of the total usage.

Recent analysis show that letter frequencies, tend to vary by subject. In the Table 2 and on the Figure 3 we give the letter frequency distribution by examined subjects.

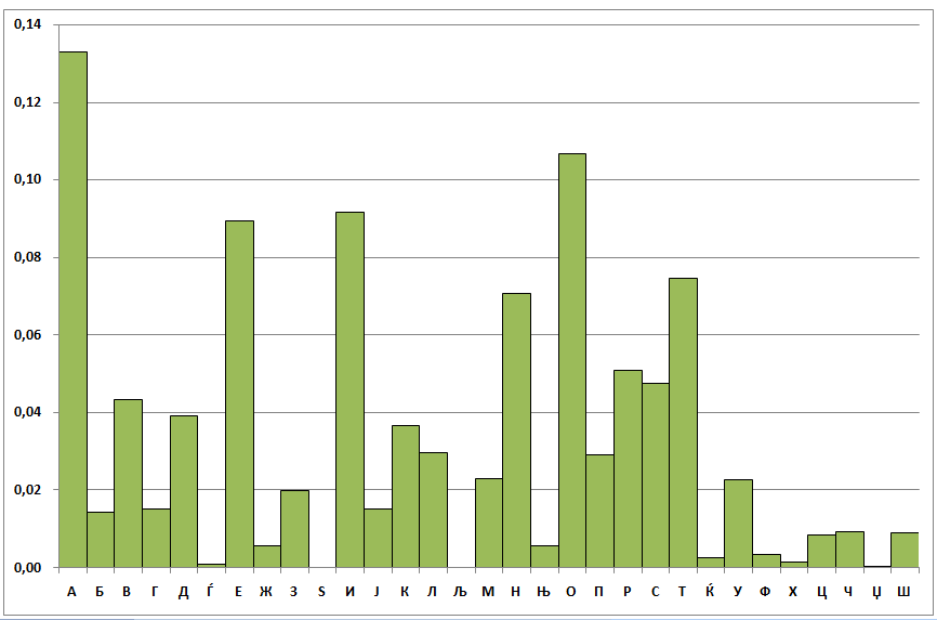

Figure 2 Relative frequencies ordered by frequency

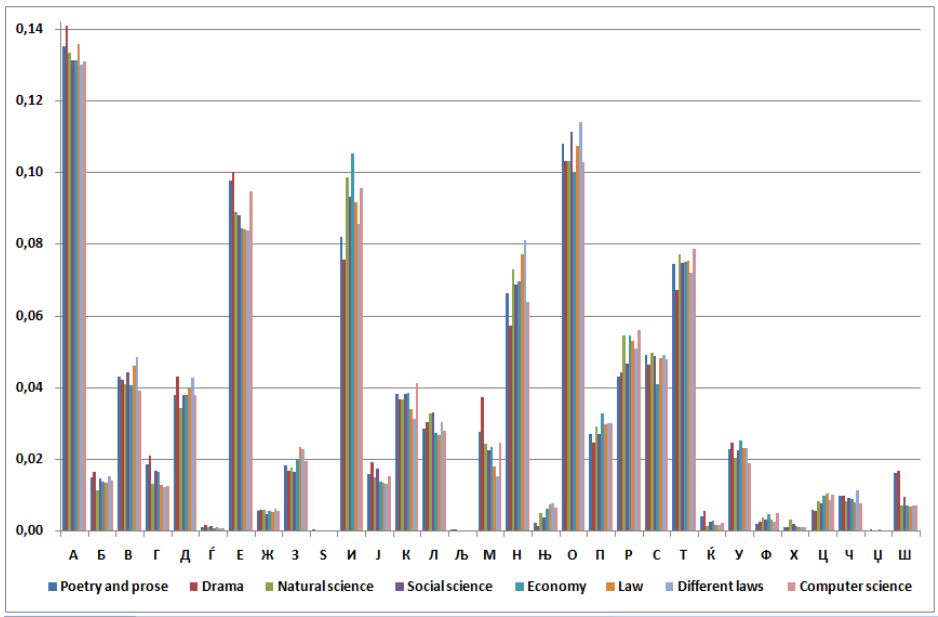

Figure 3 Relative frequencies of letters in text by subjects

Table 2 Macedonian letter frequency by subjects

L.	Frequency for poetry and prose	Frequency for drama	Frequency for natural science	Frequency for social science	Frequency for economy	Frequency for law	Frequency for different laws	Frequency for computer science
А	13,507%	14,081%	13,341%	13,129%	13,127%	13,591%	13,000%	13,099%
Б	1,508 %	1,647 %	1,142 %	1,480 %	1,387 %	1,363 %	1,519 %	1,408 %
В	4,296 %	4,225 %	4,098 %	4,436 %	4,070 %	4,618 %	4,844 %	3,914 %
Г	1,869 %	2,102 %	1,314 %	1,688 %	1,650 %	1,290 %	1,232 %	1,274 %
Д	3,805 %	4,298 %	3,435 %	3,783 %	3,789 %	3,988 %	4,272 %	3,786 %
Ѓ	0,126 %	0,169 %	0,122 %	0,135 %	0,095 %	0,118 %	0,068 %	0,068 %
Е	9,785 %	10,007%	8,912 %	8,819 %	8,451 %	8,418 %	8,386 %	9,480 %
Ж	0,574 %	0,604 %	0,592 %	0,483 %	0,561 %	0,532 %	0,622 %	0,569 %
З	1,821 %	1,672 %	1,762 %	1,656 %	2,016 %	2,338 %	2,281 %	1,970 %
S	0,038 %	0,017 %	0,006 %	0,012 %	0,008 %	0,000 %	0,002 %	0,003 %
И	8,215 %	7,579 %	9,863 %	9,311 %	10,545%	9,168 %	8,580 %	9,578 %
J	1,582 %	1,926 %	1,514 %	1,755 %	1,380 %	1,337 %	1,326 %	1,521 %
К	3,839 %	3,682 %	3,666 %	3,832 %	3,860 %	3,406 %	3,129 %	4,144 %
Л	2,866 %	3,051 %	3,281 %	3,302 %	2,746 %	2,677 %	3,046 %	2,786 %
Љ	0,045 %	0,047 %	0,013 %	0,016 %	0,007 %	0,002 %	0,004 %	0,006 %
М	2,773 %	3,742 %	2,453 %	2,271 %	2,353 %	1,790 %	1,543 %	2,481 %

Н	6,644 %	5,744 %	7,306 %	6,865 %	6,958 %	7,736 %	8,114 %	6,407 %
Њ	0,239 %	0,130 %	0,512 %	0,398 %	0,620 %	0,749 %	0,792 %	0,657 %
О	10,811%	10,315%	10,323%	11,135%	9,985 %	10,742%	11,401%	10,300%
П	2,722 %	2,455 %	2,912 %	2,703 %	3,289 %	2,982 %	3,024 %	3,024 %
Р	4,303 %	4,436 %	5,448 %	4,679 %	5,464 %	5,311 %	5,095 %	5,601 %
С	4,902 %	4,653 %	4,971 %	4,888 %	4,084 %	4,814 %	4,910 %	4,794 %
Т	7,456 %	6,726 %	7,732 %	7,494 %	7,507 %	7,543 %	7,220 %	7,866 %
Ќ	0,410 %	0,556 %	0,146 %	0,272 %	0,280 %	0,187 %	0,166 %	0,242 %
У	2,292 %	2,469 %	2,039 %	2,263 %	2,527 %	2,328 %	2,317 %	1,904 %
Ф	0,214 %	0,270 %	0,380 %	0,320 %	0,463 %	0,311 %	0,263 %	0,513 %
Х	0,119 %	0,118 %	0,324 %	0,200 %	0,131 %	0,114 %	0,114 %	0,099 %
Ц	0,585 %	0,562 %	0,827 %	0,777 %	0,989 %	1,043 %	0,869 %	1,015 %
Ч	0,993 %	0,983 %	0,849 %	0,919 %	0,894 %	0,797 %	1,128 %	0,770 %
Џ	0,032 %	0,063 %	0,007 %	0,026 %	0,053 %	0,013 %	0,010 %	0,011 %
Ш	1,629 %	1,672 %	0,708 %	0,952 %	0,710 %	0,693 %	0,724 %	0,710 %

3. Bigram and trigram frequency distribution

For cryptanalysis, more complex use of statistics can be achieved made, such as considering frequency of pairs of letters (digrams or bigrams or digraphs), triplets (trigrams), and so on. This provide more information to the cryptanalyst, for instance, Њ and Е nearly always occur together in that order in Macedonian language, even though Њ itself is rare.

Bigram and trigram frequency distribution for Macedonian language are presented in Table 3. “Top twenty” bigrams in Macedonian language comprise about 31,2% of the total usage and “top ten” bigrams comprise about 18,3% of the total usage.

Table 3 “Top thirty” bigrams and trigrams in Macedonian language ordered by frequencies

Bigrams	Count	Frequency
НА	752.140	3,235%
АТ	432.434	1,860%
ТА	420.266	1,808%
НИ	410.148	1,764%
ТЕ	399.981	1,721%
РА	392.436	1,688%
ОТ	375.777	1,616%
СТ	375.112	1,614%
ТО	358.245	1,541%
КО	338.492	1,456%
ВА	338.320	1,455%
ВО	335.956	1,445%
ЕН	328.730	1,414%
ОД	319.684	1,375%
ПР	295.522	1,271%
РЕ	286.956	1,234%
КА	282.442	1,215%
ИТ	278.301	1,197%
ПО	276.883	1,191%
НО	269.854	1,161%
ЗА	261.579	1,125%
ДА	228.540	0,983%
ЈА	228.029	0,981%
СЕ	222.786	0,958%
УВ	209.417	0,901%
РИ	207.967	0,895%
ОВ	206.159	0,887%
ТИ	203.082	0,874%
ЛИ	199.693	0,859%
РО	195.123	0,839%

Trigrams	Count	Frequency
ИТЕ	227.851	1,261%
АТА	207.369	1,147%
УВА	203.097	1,124%
ИЈА	128.096	0,709%
АЊЕ	115.397	0,639%
СТА	110.629	0,612%
ОСТ	104.699	0,579%
ВАЊ	87.284	0,483%
ПРО	86.551	0,479%
ПРЕ	86.010	0,476%
ЕТО	77.521	0,429%
ИСТ	74.462	0,412%
РЕД	73.494	0,407%
АКО	73.165	0,405%
НАТ	71.059	0,393%
ИРА	66.287	0,367%
ШТО	65.471	0,362%
НОС	64.349	0,356%
ЕНИ	64.270	0,356%
АНИ	62.884	0,348%
ПРИ	61.401	0,340%
ИОТ	59.740	0,331%
НИТ	59.468	0,329%
ОТО	58.371	0,323%
ААТ	57.511	0,318%
ОВИ	55.883	0,309%
РАН	55.117	0,305%
ИНА	55.035	0,305%
СКИ	53.291	0,295%
МЕН	52.775	0,292%

“Top twenty” trigrams in Macedonian language comprise about 11,3% of the total usage and “top ten” bigrams comprise about 7,5% of the total usage.

4. Letter frequency distribution from the Macedonian dictionary

We examined also the Macedonian letter frequency distribution for more than 80000 Macedonian basic words from the Macedonian dictionary [6, 7, 8, 9]. Obtained results are given in Table 4.

Another interesting analysis about letters is the letter frequency of the most common first letter in words and most common last letter in words. We use the same sample from the dictionary for this analysis, and results are presented in Table 5. It is interesting that approximately 73% of all the examined words finished with a vocal. In 27,327% of examined words that vocal is А, and in 22,581% the vocal is *И*. The most frequent first letter in word with 23,880% is П.

Table 4 Macedonian letter frequency for words in the Macedonian dictionary.

Letter	Count	Frequency
А	92.952	12,996%
И	61.135	8,548%
Е	59.170	8,273%
О	52.885	7,394%
Р	47.300	6,613%
В	40.902	5,719%
Н	38.244	5,347%
У	34.068	4,763%
Т	32.226	4,506%
П	31.793	4,445%
С	31.193	4,361%
К	26.045	3,642%
Л	23.945	3,348%
Д	19.002	2,657%
З	17.290	2,417%
Ј	16.998	2,377%
М	14.103	1,972%
Ќ	11.233	1,571%
Њ	10.807	1,511%

Б	10.620	1,485%
Г	10.058	1,406%
Ч	9.690	1,355%
Ш	6.660	0,931%
Ц	6.415	0,897%
Ж	4.046	0,566%
Ф	3.018	0,422%
Х	1.281	0,179%
Џ	793	0,111%
Ѓ	545	0,076%
S	478	0,067%
Љ	327	0,046%

Table 5: Letter frequency of the most common first and last letter in words.

First letter		Last letter	
Letter	**Frequency**	**Letter**	**Frequency**
П	23,880%	**А**	27,327%
С	8,769%	**И**	22,581%
Н	6,901%	**Е**	19,262%
З	6,220%	**Н**	7,562%
Р	5,814%	**Т**	6,023%
К	5,458%	**О**	3,733%
О	5,127%	**К**	2,648%
Д	4,294%	**Р**	2,520%
И	3,974%	**В**	1,746%
Б	3,607%	**М**	1,205%
В	3,446%	**Л**	0,894%
М	3,251%	**Ц**	0,828%
Т	2,742%	**Ч**	0,651%
Г	2,564%	**С**	0,615%
А	2,092%	**Д**	0,444%
У	1,612%	**Г**	0,380%
Л	1,513%	**Ј**	0,269%
Ш	1,489%	**П**	0,241%
Ф	1,256%	**Ш**	0,212%
Е	1,182%	**З**	0,209%

Ч	1,097%	**Ж**	0,155%
Ц	0,920%	**Ф**	0,150%
Х	0,724%	**У**	0,114%
Ж	0,723%	**Б**	0,097%
Ј	0,521%	**Х**	0,073%
Џ	0,274%	**Ќ**	0,023%
S	0,180%	**Љ**	0,018%
Ќ	0,175%	**Џ**	0,009%
Ѓ	0,130%	**Њ**	0,007%
Љ	0,057%	**Ѓ**	0,004%
Њ	0,009%	**S**	0,000%

5. Conclusion

In this paper we investigate the frequency distribution of letters, bigrams and trigrams in the Macedonian language and in the Macedonian dictionary, too. We investigate, letter frequency of the most common first letter and last letter in basic words from the Macedonian dictionary, also.

References

[1] H. S. Zim (1962): *Codes and secret writing*. Scholastic Book Services.

[2] H. Beker and F. Piper (1983): *Cipher Systems: The Protection of Communications*, John Wiley & Sons, pp. 397.

[3] A. Beutelsacher (2005): *Kryptologie. Eine Einführung in die Wissenschaft vom Verschlüsseln, Verbergen und Verheimlichen. Ohne alle Geheimniskrämerei, aber nicht ohne hinterli ... utzen und Ergötzen des allgemeinen* Publikums, Aufl., Wiesbaden: Vieweg Verlagsgesellschaft, pp.10.

[4] F. Pratt (1939): *Secret and Urgent: the Story of Codes and Ciphers*, Blue Ribbon Books, pp 254-255.

[5] S. Singh (1999): *Codici e Segreti*, RCS.

[6] A. Petrovski (2005): *About a Macedonian Computational Dictionary*, Proceedings of 2nd Balkan Conference in Informatics BCI 2005, Ohrid, 17-19 November, pp. 76-83.

[7] K. Zdravkova and A. Petrovski (2007): *Derivation of Macedonian Verbal Adjectives*, International Conference RANLP, Borovets, 27-29 September, Incoma, Ltd, pp. 661-665.

[8] K. Zdravkova (2007-2008): *Создавање компјутерски ресурси за македонскиот јазик*, Македонски јазик, Институт за македонски јазик „Крсте Мисирков“, година LVIII-LIX, pp. 153-174.

[9] T. Erjavec (2010): *MULTEXT-East Version 4: Multilingual Morphosyntactic Specifications, Lexicons and Corpora*, Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May, Valletta, Malta, pp. 2544-2547.