VOLUME 10 NUMBER 1 BANJA LUKA, JUNE 2020 (1-68) ISSN 2232-9625 (Print) ISSN 2233-0194 (Online) UDC 004

Journal of Information Technology and Applications

(BANJA LUKA)

JITA

Exchange of Information and Knowledge in Research

APEIRON

THE AIM AND SCOPE

The aim and scope of the Journal of Information Technology and Applications (JITA) is:

- to provide international dissemination of contributions in field of Information Technology,
- to promote exchange of information and knowledge in research work and
- to explore the new developments and inventions related to the use of Information Technology towards the structuring of an Information Society.

JITA provides a medium for exchanging research results and achievements accomplished by the scientific community from academia and industry.

By the decision of the Ministry of Education and Culture of the Republic of Srpska, no.: 07.030-053-160-4/10 from 3/3/2010, the journal „Journal of Information Technology and Applications“ Banja Luka is registered in the Registry of public organs under the number 591.
Printed by Markos, Banja Luka in 300 copies two times a year.

Indexed in: LICENSE AGREEMENT, 3.22.12. **EBSCO** Publishing Inc., Current Abstracts

ebscohost.com
citefactor.org/contact
scholar.google.com
doisrpska.nub.rs
crossref.org
road.issn.org
citefactor.org
cosmosimpactfactor.com

Printed on acid-free paper

Annual subscription is 30 EUR
Full-text available free of charge at http://www.jita-au.com

UDC 004

ISSN 2232-9625 (Print)

ISSN 2233-0194 (Online)

JITA

VOLUME 10 NUMBER 1 BANJA LUKA, JUNE 2020 (1-68)

Contents

ZORAN AVRAMOVIĆ, PhD
EDITOR-IN-CHIEF

GORDANA RADIĆ, PhD

DUŠAN STARČEVIĆ, PhD

Dear readers,

Welcome to the tenth year of issue of the Journal of Information Technology and Applications (JITA), published by Pan-European University APEIRON, Banja Luka.

JITA continues to publish high-quality, original papers that contribute to the application of information technology (IT) methodology and case studies.

Our mission is to promote and enhance cooperation and dialogue in academia and institutions dedicated to the advancement of science and information technology, which involves the following:

- Promoting best practice standards in the field of academic IT education.
- Improvement of professional training of IT staff.
- Sharing the world's advanced IT achievements.
- Announcement of the latest and most advanced information on upcoming solutions.

JITA considers that such international cooperation can help students, researchers and scientists to learn from each other and encourages the dissemination of scientific and information knowledge.

JITA intends to achieve the mentioned goals and gain international visibility through the organization of the twelfth annual international conference ITeO (Information Technologies for E-education) and interaction with public and private universities from all parts of the world.

In this issue, JITA presents various articles.

The paper *Some Possibilities of Computer Linguistics on an Example of Analysis of Novels* shows some aspects of statistical analysis of three well-known novels. The goal of the analysis is to point to mutual similarities and differences of statistical data in these texts and to compare them with the up-to-date findings in that field.

In the second paper, authors have presented the possibility of correcting the positions of telecommunication facilities obtained using the GPS. Authors have shown that using permanent stations can achieve the required level of spatial position correction.

The third article focuses on the components and principles of expert systems.

Ambient intelligence will be presented in the fourth paper as a result of the artificial algorithm neural networks, through the following contexts: e-learning environment, identification, and security. The key role in raising students' achievements as well as competency levels belongs to modern information technology that works towards creating ambient intelligence. It is also executed through the concept of e-learning onto one of the convenient learning management platforms.

The fifth paper presents the models used for creating intelligent systems based on artificial non-chromic networks that indicate to the teachers which educational as well as teaching activities should be corrected. Results regarding data processing in artificial neural networks specifically indicate a specific activity that needs to be maintained, promoted, or changed in order to improve students' abilities and achievements.

The topic of the sixth paper is banking. Banks are known as monetary management institutions because they deal with money. Instead, with the participating and implementation of new Virtual Assistant powered with AI and Machine Learning technology in the banking sector, the institutions are again using the legacy system or may be bound to use the legacy system.

Finally, the subject of the seventh research is CNC machine programming.

Acknowledgements

On behalf of the Editorial Board, we would like to thank the authors for their high quality contributions, as well as the reviewers for the effort and time invested into the preparation of this issue of Journal of Information Technology and Applications. However, any errors or shortcomings remain our full responsibility.

Conflicts of Interest

The authors declare no conflict of interest.

Editor-in-Chief, Zoran Avramović
zoran.z.avramovic@apeiron-edu.eu
Editors, Gordana Radić, Dušan Starčević

Some Possibilities of Computer Linguistics on an Example of Analysis of Novels

Nedim Smailović, Zoran Ž. Avramović

Pan-European University APEIRON, Banja Luka, Republic of Srpska, Bosnia and Herzegovina

nedim.i.smailovic@apeiron-edu.eu, zoran.z.avramovic@apeiron-edu.eu

Contribution to the State of the Art

https://doi.org/10.7251/JIT2001005S UDC: 004.82:81`322, 821.163.4`322:004.9

Abstract: This paper shows some aspects of statistical analysis of well-known novels: Death and the Dervish by Meša Selimović (1966), Autobiography by Branislav Nušić (1924) and In the Registrar's Office by Ante Kovačić (1888). The goal of the analysis is to point to mutual similarities and differences of statistical data in those texts and to compare them with the up to date findings in that field. A part of the analysis relates to comparison of languages of these writers with today's language, used by column authors in electronic media. These kinds of researches belong to linguistics, as a science on language, but the results may be used in the contemporary development of artificial intelligence.

Keywords: computer linguistics, language, text analysis, visualization of data.

Introduction

Linguistics deals with language (French *linguistique,* after the Latin *lingua:* language) and it may be divided into: phonetics, phonology, morphology, syntax, semantics, stylistics and pragmatics. Linguistics is also multidisciplinary, so its specialized branches emerged in the 20th century, such as: mathematical linguistics, computer linguistics, psycholinguistics, sociolinguistics, neurolinguistics, etc.

A language may be described as a system of signs that serve for understanding among men. According to Jahic, "It is a form of human expression as a thinking being, disclosing its essence and its differentiation from other living beings." [18]. Different sources also offer other definitions of a language. Merriam Webster dictionary gives: "A language is a systemic means of communicating ideas or feelings through use of conventionalized signs, sounds, movements or markings in order to comprehend a meaning." [14]

In everyday communication we utter language automatically, without much thinking about the correctness of its use. Language is acquired and learned from the birth throughout a lifetime, and we could say it becomes an integral part of and a significant characteristic of every individual. Nevertheless, the world of science pays due attention to studying speech and systems of writing. The research area is practically endless as relevant sources confirm that we live in a world in which over 7.7 billion people [17] communicate, speak and write in over 7.100 languages. The number is not easy to ascertain due to intertwining among languages and dialects. A language can also die out, so in the 20th century alone a total of 110 languages were proclaimed extinct. In its reports, UNESCO assesses that, unless something is done about it, by the end of this century half of languages spoken today might disappear. With disappearance of uncodified and undocumented languages, mankind would lose not only cultural wealth, but also important knowledge about ancestors that is installed, in particular, into autochthonic languages. [11]

In a contemporary world of large and swift changes in all aspects of life, language is changing as well. Some words vanish, some get new meanings,

and some completely new emerge. In such an ambient language survives, adjusts and remains the most powerful means of human communication.

STATISTICAL ANALYSIS OF TEXT IN SELECTED NOVELS

Language is alive and is changing, but there are foundations of which those changes stand. An answer to the question: *Are the foundations susceptible to changes?* may be given by modern information technology with its powerful tools and statistical analyses. Results of one such analysis are presented in this paper. The analysis encloses the following novels: *Death and the Dervish* by Meša Selimović (1910 – 1982), *Autobiography* by Branislav Nušić (1864 – 1938) and *In the Registrar's Office* by Ante Kovačić (1854 – 1899).

In quality electronic versions of these novels [6], [9], [4] text processor statistics was as follows:

Table 1. *General statistics of a text processor*

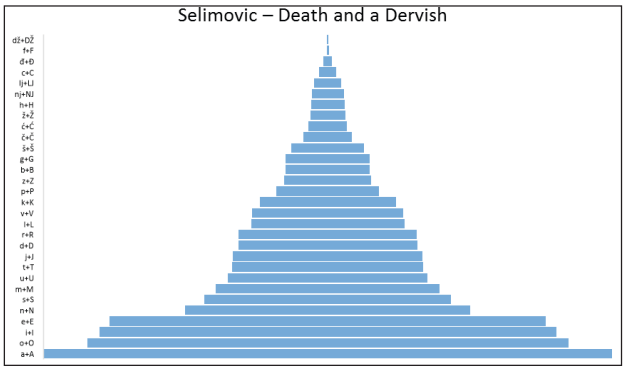
	Selimović	Nušić	Kovačić
Number of pages	248	126	210
Number of words	119,045	63,898	147,323
Number of characters (no blanks)	541,162	297,749	685,059
Number of characters (with blanks)	662,322	363,026	832,585
Number of paragraphs	3,465	1,511	2,752
Number of lines	9,435	4,924	10,521

Considering that the analyzed texts are of different length, for some analyses we took percentage of participation of certain elements, while in some other examples we were satisfied with absolute number.

Linguistics is a science that studies internal order among language units. Table 2 shows results of analysis of total number of particular letters (both capital and small) in the analyzed novels, with graphical interpretation of obtained results on Pictures 1-3. Below each graph, we have given an order of letters in that language in a descending order, left to right.

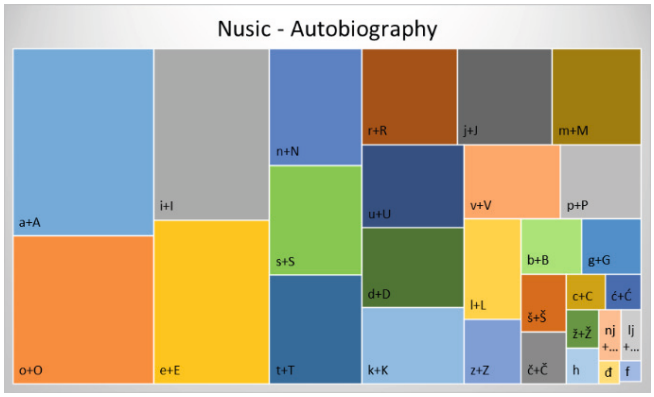
Table 2. *Order of the total number of letters in novels*

Selimović		Nušić		Kovačić	
Letter	Total	Letter	Total	Letter	Total
a+A	62,258	a+A	36,018	a+A	80,282
o+O	52,725	o+O	28,568	i+I	63,222
i+I	50,102	i+I	27,193	o+O	60,699
e+E	47,781	e+E	25,942	e+E	55,143
n+N	31,335	n+N	14,681	n+N	33,407
s+S	27,107	s+S	13,830	j+J	30,215
m+M	24,555	t+T	13,643	s+S	30,182
u+U	21,939	r+R	12,618	u+U	30,161
t+T	21,012	j+J	12,434	t+T	29,903
j+J	20,850	m+M	11,772	r+R	27,488
d+D	19,662	u+U	11,617	m+M	24,669
r+R	19,536	d+D	11,161	d+D	24,125
l+L	16,831	k+K	10,691	k+K	23,843
v+V	16,619	v+V	9,786	v+V	22,891
k+K	14,989	p+P	8,175	p+P	18,059
p+P	11,319	l+L	7,858	l+L	17,780
z+Z	9,577	z+Z	5,014	g+G	12,165
b+B	9,284	b+B	4,576	š+Š	10,654
g+G	9,240	g+G	4,471	z+Z	10,088
š+Š	8,061	š+Š	3,578	b+B	9,542
č+Č	5,407	č+Č	3,202	č+Č	6,264
ć+Ć	4,329	c+C	1,926	c+C	6,213
ž+Ž	3,900	ć+Ć	1,795	h+H	5,496
h+H	3,740	ž+Ž	1,666	ć+Ć	4,962
nj+NJ	3,580	h+H	1,578	ž+Ž	4,072
lj+LJ	3,089	nj+NJ	1,542	lj+LJ	3,882
c+C	1,901	lj+LJ	1,430	nj+NJ	3,138
đ+Đ	1,038	đ+Đ	654	đ+Đ	1,266
f+F	328	f+F	629	f+F	436
dž+DŽ	237	dž+DŽ	59	dž+DŽ	59



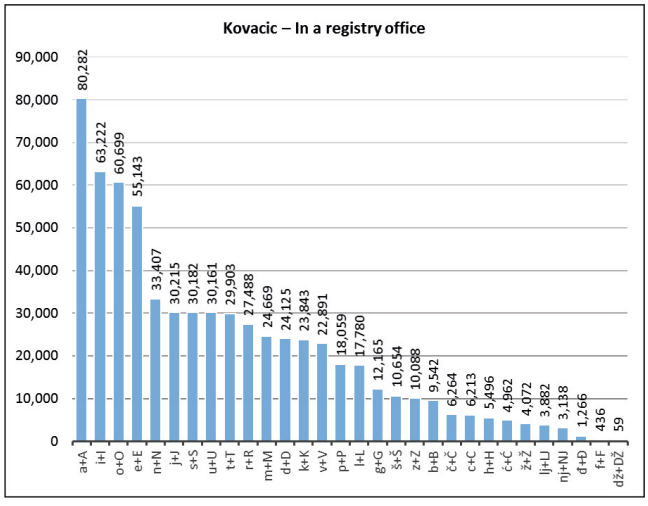
Picture 1. *Number of particular letters: Selimović – Death and the Dervish*

a+A, o+O, i+I, e+E, n+N, s+S, m+M, u+U, t+T, j+J, d+D, r+R, l+L, v+V, k+K, p+P, z+Z, b+B, g+G, š+Š, č+Č, ć+Ć, ž+Ž, h+H, nj+NJ, lj+LJ, c+C, đ+Đ, f+F, dž+DŽ.



Picture 2. *Number of particular letters: Nušić - Autobiography*

a+A, o+O, i+I, e+E, n+N, s+S, t+T, r+R, j+J, m+M, u+U, d+D, k+K, v+V, p+P, l+L, z+Z, b+B, g+G, š+Š, č+Č, c+C, ć+Ć, ž+Ž, h+H, nj+NJ, lj+LJ, đ+Đ, f+F, dž+DŽ.



Picture 3. *Number of particular letters: Kovačić – In the Registrar's Office*

a+A, i+I, o+O, e+E, n+N, j+J, s+S, u+U, t+T, r+R, m+M, d+D, k+K, v+V, p+P, l+L, g+G, š+Š, z+Z, b+B, č+Č, c+C, h+H, ć+Ć, ž+Ž, lj+LJ, nj+NJ, đ+Đ, f+F, dž+DŽ.

In the paper [19], on sample of contemporary texts [6], [13], [5], [16] this analysis gave very similar results. It was noticed that the order of letter presence with particular authors was as in Table 3.

Table 3. *Order of letter presence with particular authors*

Author	Letter order
Filipović	aioenstrjudmkvlpgzbščchćnjžljfđdž
Jergović	aioentsrjukvmdlpgzbčšchnjžćljfđdž
Apostolovski	aioenrsutkdvljmpgzbcčšćnjhžljđfdž
Lekić	aieonrstujkmvdlpgzbcčnjšhljžćfđdž

Table 4 shows unified statistical data from this research and paper [19]. Starting from the highest to lowest frequency of letter appearance, the order is as follows:

a i o e n s r t u j m k d v l p g z b š č c h ć nj ž lj đ f dž

The column *Average* shows average percentage presence of particular letters in observed texts of all authors altogether. In other columns we gave absolute amounts of discrepancy for each author in relation to the average (average of an author minus summary average for all). Discrepancies are very small. The biggest positive discrepancy is with Lekić and a letter R (the difference is 1.04%), while the largest negative discrepancy is with Selimović and letter R (difference is -0.84%).

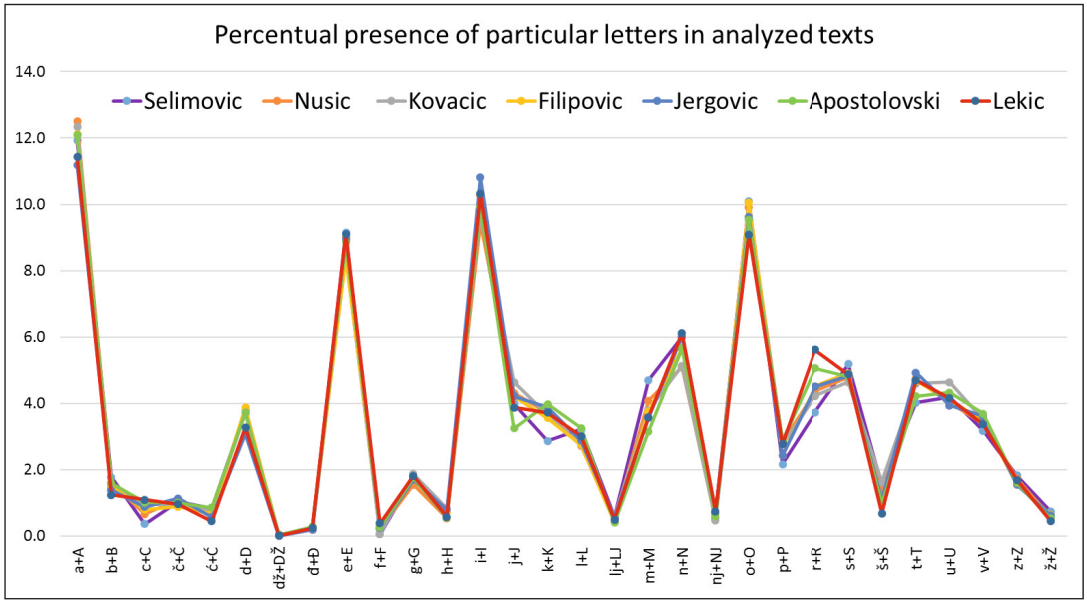
Table 4. *Percentage of presence of every letter in all the texts altogether and individual discrepancies for each of the authors*

Letter	Average	Filipovic	Jergovic	Apostolovski	Lekic	Selimovic	Nusic	Kovacic
a+A	11.94	0.17	-0.75	0.16	-0.51	-0.02	0.56	0.40
b+B	1.51	0.01	-0.13	0.09	-0.27	0.27	0.08	-0.04
c+C	0.83	-0.03	0.06	0.19	0.27	-0.46	-0.16	0.13
č+Č	1.02	-0.12	0.13	-0.01	-0.05	0.02	0.10	-0.05
ć+Ć	0.68	-0.01	-0.11	0.18	-0.23	0.15	-0.06	0.08
d+D	3.62	0.26	-0.53	0.10	-0.33	0.15	0.26	0.09
dž+DŽ	0.03	0.01	-0.01	0.01	-0.01	0.02	-0.01	-0.02
đ+Đ	0.23	0.03	-0.02	0.05	0.01	-0.03	0.00	-0.04
e+E	8.85	-0.46	0.09	0.01	0.27	0.30	0.16	-0.37
f+F	0.22	0.08	0.02	0.04	0.17	-0.16	0.00	-0.15
g+G	1.74	-0.03	-0.03	0.03	0.07	0.03	-0.19	0.13
h+H	0.68	0.04	0.12	-0.12	-0.10	0.03	-0.14	0.16
i+I	10.01	0.27	0.79	-0.09	0.31	-0.42	-0.57	-0.29
j+J	4.07	0.15	0.15	-0.83	-0.20	-0.08	0.24	0.57
k+K	3.63	-0.07	0.25	0.36	0.10	-0.76	0.08	0.04
l+L	2.94	-0.19	-0.06	0.32	0.07	0.28	-0.22	-0.21
lj+LJ	0.51	-0.08	0.06	-0.10	-0.01	0.08	-0.02	0.08
m+M	3.81	-0.05	-0.24	-0.65	-0.22	0.89	0.28	-0.01
n+N	5.67	-0.01	0.35	-0.01	0.45	0.33	-0.58	-0.53
nj+NJ	0.63	0.03	0.09	-0.04	0.12	0.05	-0.10	-0.15
o+O	9.67	0.39	-0.04	-0.13	-0.57	0.43	0.25	-0.33
p+P	2.62	-0.18	-0.18	0.26	0.17	-0.45	0.22	0.16
r+R	4.58	-0.06	-0.08	0.49	1.04	-0.84	-0.20	-0.35
s+S	4.87	0.04	-0.04	-0.05	0.01	0.32	-0.07	-0.22
š+Š	1.17	-0.17	-0.09	-0.18	-0.48	0.38	0.07	0.47
t+T	4.56	0.12	0.38	-0.34	0.15	-0.53	0.18	0.04
u+U	4.20	-0.11	-0.26	0.14	-0.04	0.00	-0.17	0.44
v+V	3.47	0.03	0.15	0.23	-0.09	-0.29	-0.07	0.05
z+Z	1.65	-0.08	-0.09	-0.04	0.04	0.18	0.09	-0.10
ž+Ž	0.60	0.04	0.03	-0.07	-0.14	0.15	-0.02	0.03

Based on performed analysis and data presented in the above table we can draw several conclusions, and one of the most important one is that in the overall text of 2.026.250 letters almost half 931.227, or 46.10%, are the five most frequently used: **a i o e n.**

Looking from another angle, 17 letters: **v l p g z b** š č **c h** ć **nj** ž **lj** đ **f dž** make just above one fifth of the text, 20.54%).

Picture 4 graphically presents previously described analysis results.



Picture 4. *Graphical presentation of percentual presence of particular letters in the analyzed texts*

The significance of five most frequent letters can be seen in the next example. If we take the first three sentences of the novel *Death and the Dervish*:

"*Počinjem ovu svoju priču, nizašto, bez koristi za sebe i za druge, iz potrebe koja je jača od koristi i razuma, da ostane zapis moj o meni, zapisana muka razgovora sa sobom, s dalekom nadom da će se naći neko rješenje kad bude račun sveden, ako bude, kad ostavim trag mastila na ovoj hartiji što čeka kao izazov. Ne znam šta će biti zabilježeno, ali će u kukama slova ostati nešto od onoga što je bivalo u meni, pa se više neće gubiti u kovitlacima magle, kao da nije ni bilo, ili da ne znam šta je bilo. Tako ću moći da vidim sebe kakav postajem, to čudo koje ne poznajem, a čini mi se da je čudo što uvijek nisam bio ono što sam sad.*"

And then we delete and leave blanks in the spots where letters **a i o e n** were:

"P č j m vu sv ju pr ču, z št , b z k r st z s b z drug , z p tr b k j j j č d k r st r zum , d st z p s m j m , z p s muk r zg v r s s b m, s d l k m d m d ć s ć k rj š j k d bud r ču sv d , k bud , k d st v m tr g m st l v j h rt j št č k k z z v. z m št ć b t z b lj ž , l ć u kuk m sl v st t št d g št j b v l u m , p s v š ć gub t u k v tl c m m gl , k d j b l , l d z m št j b l . T k ću m ć d v d m s b k k v p st j m, t čud k j p z j m, č m s d j čud št uv j k s m b št s m s d."

The possibility of recognizing (comprehending) the content after that is extremely lowered, yet we are missing only five letters from the whole alphabet.

FREQUENCY OF LETTERS IN EUROPEAN LANGUAGES

Presented results relate to the languages used by the authors in the area of former Yugoslavia. However, other languages have, in most cases, other phonemes, and thus also different sets of letters used to mark them. Each language has its own relative frequency of particular letters. For some of the languages, the data is given in the following table.

Table 5. *Relative frequencies of letters in some European languages (descending order according to the English alphabet)*

Letter	English	French	German	Spanish	Esperanto	Italian	Turkish	Swedish	Polish	Danish	Czech
e	12.702%	14.715%	16.396%	12.181%	8.995%	11.792%	9.912%	10.149%	7.352%	15.453%	7.562%
t	9.056%	7.244%	6.154%	4.632%	5.276%	5.623%	3.314%	7.691%	2.475%	6.862%	5.727%
a	8.167%	7.636%	6.516%	11.525%	12.117%	11.745%	12.920%	9.383%	10.503%	6.025%	8.421%
o	7.507%	5.796%	2.594%	8.683%	8.779%	9.832%	2.976%	4.482%	6.667%	4.636%	6.695%
i	6.966%	7.529%	6.550%	6.247%	10.012%	10.143%	9.600%*	5.817%	8.328%	6.000%	6.073%
n	6.749%	7.095%	9.776%	6.712%	7.955%	6.883%	7.987%	8.542%	6.237%	7.240%	6.468%
s	6.327%	7.948%	7.270%	7.977%	6.092%	4.981%	3.014%	6.590%	5.224%	5.805%	5.212%
h	6.094%	0.737%	4.577%	0.703%	0.384%	0.636%	1.212%	2.090%	1.015%	1.621%	1.356%
r	5.987%	6.693%	7.003%	6.871%	5.914%	6.367%	7.722%	8.431%	5.243%	8.956%	4.799%
d	4.253%	3.669%	5.076%	5.010%	3.044%	3.736%	5.206%	4.702%	3.725%	5.858%	3.475%
l	4.025%	5.456%	3.437%	4.967%	6.104%	6.510%	5.922%	5.275%	2.564%	5.229%	3.802%
c	2.782%	3.260%	2.732%	4.019%	0.776%	4.501%	1.463%	1.486%	3.895%	0.565%	0.740%
u	2.758%	6.311%	4.166%	2.927%	3.183%	3.011%	3.235%	1.919%	2.062%	1.979%	2.160%
m	2.406%	2.968%	2.534%	3.157%	2.994%	2.512%	3.752%	3.471%	2.515%	3.237%	2.446%
w	2.360%	0.049%	1.921%	0.017%	0	0.033%	0	0.142%	5.813%	0.069%	0.016%
f	2.228%	1.066%	1.656%	0.692%	1.037%	1.153%	0.461%	2.027%	0.143%	2.406%	0.084%
g	2.015%	0.866%	3.009%	1.768%	1.171%	1.644%	1.253%	2.862%	1.731%	4.077%	0.092%
y	1.974%	0.128%	0.039%	1.008%	0	0.020%	3.336%	0.708%	3.206%	0.698%	1.043%
p	1.929%	2.521%	0.670%	2.510%	2.755%	3.056%	0.886%	1.839%	2.445%	1.756%	1.906%
b	1.492%	0.901%	1.886%	2.215%	0.980%	0.927%	2.844%	1.535%	1.740%	2.000%	0.822%
v	0.978%	1.838%	0.846%	1.138%	1.904%	2.097%	0.959%	2.415%	0.012%	2.332%	5.344%
k	0.772%	0.074%	1.417%	0.011%	4.163%	0.009%	5.683%	3.140%	2.753%	3.395%	2.894%
j	0.153%	0.613%	0.268%	0.493%	3.501%	0.011%	0.034%	0.614%	1.836%	0.730%	1.433%
x	0.150%	0.427%	0.034%	0.215%	0	0.003%	0	0.159%	0.004%	0.028%	0.027%
q	0.095%	1.362%	0.018%	0.877%	0	0.505%	0	0.020%	0	0.007%	0.001%
z	0.074%	0.326%	1.134%	0.467%	0.494%	1.181%	1.500%	0.070%	4.852%	0.034%	1.503%

The source table [8] contains also data on frequency of other letters used in each of the listed 11 languages.

The observed languages have 84 letters in use, and those are:

a, á, à, â, ä, ã, å, ą, æ, b, c, ĉ, ç, č, ć, d, ď, ð, e, ě, é, è, ê, ë, ę, f, g, ğ, ĝ, h, Ĥ, i, ı, í, ì, î, ï, j, ĵ, k, l, ł, m, n, ň, ń, ñ, o, ó, ò, ô, ö, õ, ø, œ, p, q, r, ř, s, ś, ŝ, ş, ß, š, t, ť, þ, u, ŭ, ú, ù, û, ü, ů, v, w, x, y, ý, z, ź, ż, ž.

When we show the frequency of these letters for each language in a descending order, we get the following table.

Table 6. *Eleven languages and sorted order of letters used in them*

Letter	English	Letter	French	Letter	German	Letter	Spanish	Letter	Esperanto	Letter	Italian	Letter	Turkish	Letter	Swedish	Letter	Polish	Letter	Danish	Letter	Czech
e	12.70%	e	14.72%	e	16.40%	e	12.18%	a	12.12%	e	11.79%	a	11.92%	e	10.15%	a	10.50%	e	15.45%	a	8.42%
t	9.06%	s	7.95%	n	9.78%	a	11.53%	i	10.01%	a	11.75%	e	8.91%	a	9.38%	i	8.33%	r	8.96%	e	7.56%
a	8.17%	a	7.64%	s	7.27%	o	8.68%	e	9.00%	i	10.14%	i	8.60%	n	8.54%	e	7.35%	n	7.24%	o	6.70%
o	7.51%	i	7.53%	r	7.00%	s	7.98%	o	8.78%	o	9.83%	n	7.49%	r	8.43%	o	6.67%	t	6.86%	n	6.47%
i	6.97%	t	7.24%	i	6.55%	r	6.87%	n	7.96%	n	6.88%	r	6.72%	t	7.69%	n	6.24%	a	6.03%	i	6.07%
n	6.75%	n	7.10%	a	6.52%	n	6.71%	l	6.10%	l	6.51%	l	5.92%	s	6.59%	w	5.81%	i	6.00%	t	5.73%
s	6.33%	r	6.69%	t	6.15%	i	6.25%	s	6.09%	r	6.37%	ı	5.11%	i	5.82%	r	5.24%	d	5.86%	v	5.34%
h	6.09%	u	6.31%	d	5.08%	d	5.01%	r	5.91%	t	5.62%	d	4.71%	l	5.28%	s	5.22%	s	5.81%	s	5.21%
r	5.99%	o	5.80%	h	4.58%	l	4.97%	t	5.28%	s	4.98%	k	4.68%	d	4.70%	z	4.85%	l	5.23%	r	4.80%
d	4.25%	l	5.46%	u	4.17%	t	4.63%	k	4.16%	c	4.50%	m	3.75%	o	4.48%	c	3.90%	o	4.64%	l	3.80%
l	4.03%	d	3.67%	l	3.44%	c	4.02%	j	3.50%	d	3.74%	y	3.34%	m	3.47%	d	3.73%	g	4.08%	d	3.48%
c	2.78%	c	3.26%	g	3.01%	m	3.16%	u	3.18%	p	3.06%	t	3.31%	k	3.14%	y	3.21%	k	3.40%	k	2.89%
u	2.76%	m	2.97%	c	2.73%	u	2.93%	d	3.04%	u	3.01%	u	3.24%	g	2.86%	k	2.75%	m	3.24%	m	2.45%
m	2.41%	p	2.52%	o	2.59%	p	2.51%	m	2.99%	m	2.51%	s	3.01%	v	2.42%	l	2.56%	f	2.41%	u	2.16%
w	2.36%	v	1.84%	m	2.53%	b	2.22%	p	2.76%	v	2.10%	b	2.84%	h	2.09%	m	2.52%	v	2.33%	p	1.91%
f	2.23%	é	1.50%	w	1.92%	g	1.77%	v	1.90%	g	1.64%	o	2.48%	f	2.03%	t	2.48%	b	2.00%	í	1.64%
g	2.02%	q	1.36%	b	1.89%	v	1.14%	g	1.17%	z	1.18%	ü	1.85%	u	1.92%	p	2.45%	u	1.98%	z	1.50%
y	1.97%	f	1.07%	f	1.66%	y	1.01%	f	1.04%	f	1.15%	ş	1.78%	p	1.84%	ł	2.11%	p	1.76%	j	1.43%
p	1.93%	b	0.90%	k	1.42%	q	0.88%	b	0.98%	b	0.93%	z	1.50%	ä	1.80%	u	2.06%	h	1.62%	h	1.36%
b	1.49%	g	0.87%	z	1.13%	ó	0.83%	c	0.78%	h	0.64%	g	1.25%	b	1.54%	j	1.84%	å	1.19%	ě	1.22%
v	0.98%	h	0.74%	ü	1.00%	í	0.73%	ĝ	0.69%	à	0.64%	h	1.21%	c	1.49%	b	1.74%	ø	0.94%	y	1.04%
k	0.77%	j	0.61%	v	0.85%	h	0.70%	ĉ	0.66%	q	0.51%	ç	1.16%	å	1.34%	g	1.73%	æ	0.87%	ý	1.00%
j	0.15%	à	0.49%	p	0.67%	f	0.69%	ŭ	0.52%	è	0.26%	ğ	1.13%	ö	1.31%	ó	1.14%	j	0.73%	á	0.87%
x	0.15%	x	0.43%	ä	0.58%	á	0.50%	z	0.49%	ú	0.17%	c	0.96%	y	0.71%	ę	1.04%	y	0.70%	b	0.82%
q	0.10%	z	0.33%	ö	0.44%	j	0.49%	ŝ	0.39%	w	0.03%	v	0.96%	j	0.61%	h	1.02%	c	0.57%	c	0.74%
z	0.07%	è	0.27%	ß	0.31%	z	0.47%	h	0.38%	í	0.03%	p	0.89%	x	0.16%	ś	0.81%	w	0.07%	ž	0.72%
		ê	0.22%	j	0.27%	é	0.43%	ĵ	0.06%	y	0.02%	ö	0.78%	w	0.14%	ć	0.74%	z	0.03%	š	0.69%
		y	0.13%	y	0.04%	ñ	0.31%	ĥ	0.02%	j	0.01%	f	0.46%	z	0.07%	ż	0.71%	x	0.03%	é	0.63%
		ç	0.09%	x	0.03%	x	0.22%			k	0.01%	j	0.03%	q	0.02%	ą	0.70%	q	0.01%	č	0.46%
		k	0.07%	q	0.02%	ú	0.17%			x	0.00%					ń	0.36%			ř	0.38%
		û	0.06%			w	0.02%									f	0.14%			ů	0.20%
		ù	0.06%			ü	0.01%									ź	0.08%			g	0.09%
		â	0.05%			k	0.01%									v	0.01%			f	0.08%
		w	0.05%																	ú	0.05%
		î	0.05%																	x	0.03%
		ô	0.02%																	ó	0.02%
		œ	0.02%																	w	0.02%
		ë	0.01%																	ď	0.02%
		ï	0.01%																	ň	0.01%
																				ť	0.01%

For easier remembering of order of letter frequency, bibliography offers the first twelve, most frequent letters, presented in the form of two (non-existing) six-letter words. For the eleven languages of the above table, where words are shown in Table 7.

Table 7. *The most frequent letters in languages*

Language	Most frequent letters
English	etaoin shrdlc
French	esaitn ruoldc
German	ensria tdhulg
Spanish	eaosrn idltcm
Esperanto	aieonl srtkju
Italian	eaionl rtscdp
Turkish	aeinrl ıdkmyt
Swedish	eanrts ildomk
Polish	aieonw rszcdy
Danish	erntai dslogk
Czech	aeonit vsrldk

For joined text of novels from this and mentioned paper [19], which treats texts of columnists, the two words would be:

aioens rtujmk

N-GRAM WORD ANALYSIS

In computer linguistics and statistics, there is a term of n-gram. It is a continued line of n elements in the observed sample. The elements may be consecutive letters, words or selected sets of marks to be analyzed. N-grams are put in tables for the comparing of a number of repetitions of the same linguistic elements in different languages. N-grams are studied in several areas of informatics, computer linguistics and applied mathematics.

Each author has its own style in writing and some word patterns are more often found, some seldomly or not at all.

Tables 8-12 contain the most frequent bi-grams, three-grams, the most frequent words, nouns and verbs in the analyzed novels by Selimović, Nušić and Kovačić.

The most frequent word three-gram with Selimović is *kao da je* (51 times), with Nušić *Tako na primer* (32 times), and with Kovačić *kano da je* (68 times).

Table 8. *Number of repeated three-grams of words*

Three-gram of words	SELIMOVIC	Three-gram of words	NUSIC	Three-gram of words	KOVACIC
kao da je	51	Tako na primer	32	kano da je	68
mi se da	39	da je to	25	kano da se	40
da se ne	36	Razume se da	24	si ga neka	34
sam da je	33	da bi se	21	kano da si	33
ono što je	33	šta je to	20	Vrag si ga	32
kao da sam	31	da sam ja	19	da mu je	28
da je to	27	kao što je	16	i kano da	26
mi je bilo	25	na taj način	14	kano da će	22
Htio sam da	25	ja sam se	14	kano da mu	19
zato što je	25	ne bi li	13	da mi je	19
kao da se	24	docnije u životu	12	Ha ha ha	18
znao sam da	23	da sam se	12	kano da ga	18
je u meni	23	Ja ne znam	12	da se ne	17
nisam mogao da	23	Ono se smejalo	11	da je u	17
ono što sam	22	sam se u	11	i da se	17
ne mogu da	22	s tim da	11	se kano da	17
i ono što	21	ne može da	11	da mu se	16
Mislio sam da	21	sam se ja	10	da ga je	16
sve što je	21	ima ih koji	10	li je to	15
a ja sam	20	u tom pogledu	9	je sve to	14
što mi je	20	ne mogu da	9	ne bi li	14

Table 9. *The most frequent words in text*

The most frequent word	SELIMOVIC	The most frequent word	NUSIC	The most frequent word	KOVACIC
je	4,933	i	2,550	i	7,343
i	4,505	je	1,969	se	3,829
da	4,452	da	1,886	u	2,935
se	3,067	se	1,559	je	2,853
sam	2,753	u	1,334	da	2,697
u	2,239	na	900	a	1,931
ne	1,903	sam	808	na	1,854
što	1,439	to	629	ne	1,462
na	1,250	a	583	ja	866
a	1,218	ne	553	ti	844
me	975	ja	540	to	823
nije	964	bi	488	za	820
kao	938	kao	431	te	810
su	901	su	406	pa	804
to	865	pa	356	što	790
mi	837	koji	351	mi	762
ga	826	mi	350	tako	700
ali	728	tako	349	ga	649
bi	688	za	339	će	634
od	677	sa	336	sve	602
ni	677	od	334	od	596
sve	668	kad	319	mu	579
nisam	577	ali	308	li	573
za	569	te	244	on	566
kad	540	iz	240	kano	563
mu	474	nije	232	kako	562
ili	465	već	218	s	537
o	433	smo	214	iz	534
li	430	sve	200	bi	523
s	418	ti	195	sam	509

Table 10. *Number of repeated bi-grams of words*

bi-gram of words	SELIMOVIC	bi-gram of words	NUSIC	bi-gram of words	KOVACIC
da se	512	da je	216	kano da	358
da je	424	da se	181	da je	349
mi je	320	da sam	108	da se	325
što je	299	što je	108	mi se	141
sam se	277	bi se	104	da će	121
sam da	247	na primer	103	se i	114
kao da	204	sam se	101	se u	114
da ga	193	je to	94	i ne	113
ono što	176	koji je	91	mu se	112
da sam	157	u životu	84	je to	112
što se	155	mi je	81	da mu	99
što sam	152	da bi	78	što je	96
mu je	149	ja sam	77	mali kanonik	92
da me	146	sam ja	77	mi je	92
se ne	141	se u	73	je i	90
je to	137	to je	65	i da	90
i da	135	i da	62	da mi	86
je da	134	nam je	60	da ga	84
je u	129	kao što	58	se ne	83
je i	128	da su	55	ne bi	79
se da	128	kao da	55	te se	77
da ne	127	i u	55	a onda	73
mi se	117	ne bi	54	kako je	72
mogao da	116	je u	53	mu je	72
ne bi	116	se da	53	kao da	71
su se	109	me je	49	ti je	71
je bilo	105	je i	48	da si	70
to je	105	što sam	47	se na	69
da mu	103	koja je	47	sve to	67
i ne	100	što se	45	je u	66

Table 11. *The most frequent nouns in the text*

The most frequent noun	SELIMOVIC	The most frequent noun	NUSIC	The most frequent noun	KOVACIC
čovjek	201	primer	100	kumordinar	233
ljudi	190	životu	95	Ivica	186
Hasan	155	čovek	86	otac	176
riječi	114	reči	82	kanonik	172
misao	109	dana	82	Miha	170
život	83	profesor	76	oči	167
oči	83	život	65	jožica	154
vrijeme	82	reč	63	ruke	148
čovjeka	79	ljubav	59	bog	139
riječ	78	stvar	51	glavu	137
noć	77	godina	41	Laura	129

Table 12. *The most frequent verbs in the text*

The most frequent verb	SELIMOVIC	The most frequent verb	NUSIC	The most frequent verb	KOVACIC
bilo	411	bio	136	bijaše	334
bio	379	ima	117	stade	144
rekao	300	bilo	114	biti	136
mogao	262	bila	108	vidiš	123
znao	233	može	106	uhvati	94
može	195	biti	97	budi	91
mislio	190	razume	76	mogao	76
treba	157	mogao	63	ima	76
bila	149	imao	60	znam	70
mogu	144	mora	59	skoči	60
biti	142	treba	59	može	58
bude	140	znam	49	znao	56

Table 13. *The most frequent words according to number of letters*

Selimovic

	1 letter	Broj	2 letters	Broj	3 letters	Broj	4 letters	Broj	5 letters	Broj	6 letters	Broj	7 letters	Broj	8 letters	Broj	9 letters	Broj	10 letters	Broj
1	i	4,505	je	4,933	sam	2,753	nije	964	nisam	577	čovjek	201	sigurno	111	svejedno	99	zaboravio	51	neprestano	35
2	u	2,239	da	4,452	što	1,439	bilo	411	možda	318	mislio	190	govorio	109	razgovor	60	izgledalo	30	mula-jusuf	31
3	a	1,218	se	3,067	kao	938	samo	403	rekao	300	toliko	128	trebalo	104	drukčije	45	prijatelja	27	osluškivao	14
4	o	433	ne	1,903	ali	728	zbog	399	ništa	272	uvijek	126	poslije	85	pogledao	45	gledajući	24	mogućnosti	13
5	s	418	na	1,250	sve	668	kako	361	mogao	262	riječi	114	vrijeme	82	očekivao	37	odgovorio	21	pogriješio	12
6	e	8	me	975	kad	540	više	306	zašto	211	učinio	98	čovjeka	79	odjednom	34	prijatelji	21	razgovaram	11
7			su	901	ili	465	mene	293	nešto	174	njegova	97	osjećao	75	prijatelj	33	uznemiren	16	razgovarao	11
8			to	865	šta	413	koji	290	treba	157	koliko	93	izgleda	67	trenutak	32	zadovoljan	16	zaboravili	11
9			mi	837	bio	379	tako	288	hasan	155	desilo	92	siguran	51	potrebno	27	razgovora	15	zaustaviti	10
10			ga	826	bih	343	znam	278	vidio	127	gledao	88	svijetu	49	zatvoren	27	sigurnost	15	osjećajući	10
11			bi	688	bez	341	njega	260	prema	122	zaista	76	učiniti	46	vjerovao	27	pokušavao	15	uznemireno	10
12			od	677	ako	336	znao	233	dobro	119	jednom	73	ponekad	44	vjerujem	27	prijatelju	15	primijetio	10
13			ni	677	jer	323	onda	214	mnogo	115	svijet	70	razloga	44	slučajno	25	Sinanudin	15	neprijatelj	10
14			za	569	sad	270	meni	208	nikad	111	suviše	69	vremena	40	nasmijao	25	nesiguran	14	sigurnosti	9
15			mu	474	smo	233	može	195	svoju	109	uzalud	65	osjetio	40	riječima	22	zaustavio	14	zaboraviti	8
16			li	430	još	229	pred	193	misao	109	njegovo	65	svijeta	39	govoriti	21	mogućnost	13	zatvorenih	8
17			ja	395	ono	217	ljudi	190	teško	105	otišao	63	sjećanje	38	potrebna	20	opasnosti	13	ravnodušan	8
18			će	329	već	189	zato	185	svoje	99	ljudima	60	govorim	38	nemoguće	19	izmijenio	13	pokazujući	8
19			on	329	nas	150	sebe	177	jedan	99	sasvim	57	Muhamed	34	naročito	19	unaprijed	13	vrijednost	7
20			ti	309	dok	120	htio	157	kažem	95	između	57	postoji	33	dvadeset	18	razgovoru	12	oduševljenje	7

The longest word with Selimović: neprikosnovenošću – 17 letters.

Nusic

	1 letter	Broj	2 letters	Broj	3 letters	Broj	4 letters	Broj	5 letters	Broj	6 letters	Broj	7 letters	Broj	8 letters	Broj	9 letters	Broj	10 letters	Broj
1	i	2,550	je	1,969	sam	808	koji	351	jedan	129	primer	100	jednoga	59	profesor	76	profesora	29	matematika	19
2	u	1,334	da	1,886	što	501	tako	349	nisam	112	životu	95	docnije	42	neobično	49	verovatno	27	roditeljima	7
3	a	583	se	1,559	kao	431	nije	232	čovek	86	razume	76	izgleda	33	prilikom	40	gimnazije	25	upotrebiti	7
4	o	192	na	900	kad	319	koje	186	posle	85	odista	71	misliti	33	nekoliko	36	gospodine	24	geografije	6
5	s	66	to	629	ali	308	samo	179	svoje	84	koliko	67	dovoljno	28	ministar	32	otprilike	23	pozdraviti	6
6	e	33	ne	553	već	218	kako	171	život	65	toliko	55	predmet	24	naročito	31	pantalone	23	biografiju	5
7			ja	540	smo	214	koja	142	mogao	63	jednom	51	lekciju	22	uostalom	29	profesori	22	francuskog	5
8			bi	488	sve	200	onda	141	prema	63	najzad	51	mišljenje	21	dvadeset	25	zamislite	19	neprestano	5
9			su	406	jer	166	vrlo	138	jedna	61	nekako	42	međutim	18	jedanput	24	zaboravio	15	podjednako	5
10			pa	356	ili	151	bilo	114	zatim	60	godina	41	potpuno	18	odgovara	22	latinskog	15	odgovaramo	5
11			mi	350	još	151	bila	108	treba	59	stvari	41	razumem	18	razgovor	18	pantalona	12	policijski	5
12			za	339	ona	143	može	106	ljubav	59	ljubavi	41	izvesno	17	direktor	17	nepoznate	12	generacije	5
13			od	334	ako	131	biti	97	mnogo	54	četiri	37	razreda	17	istorije	16	finansija	7	količinama	5
14			te	244	taj	131	toga	95	dobro	54	sasvim	36	slučaju	17	pogdekad	15	pokušavao	7	zajednički	5
15			iz	240	šta	122	mene	93	jedno	53	razred	36	zajedno	16	ministra	15	odgovorio	7	latinskoga	5
16			ti	195	ima	117	više	92	jednu	53	života	34	trebalo	16	latinski	14	činovnika	6	praktikant	5
17			mu	183	nam	107	reči	82	možda	53	čoveka	32	odgovor	16	količina	14	stihovima	6	književnika	4
18			po	172	kod	95	dana	82	ništa	51	između	32	smejalo	15	drukčije	13	zoologije	6	biografija	4
19			on	170	vam	94	tome	75	stvar	51	recimo	31	gospođa	15	mladosti	13	profesija	6	odgovorila	4
20			me	165	ono	87	pred	72	nešto	48	kojima	29	mišljenja	15	priznaje	13	izmislili	6	strahovito	4

The longest word with Nušić: šezdesetogodišnjicu and devetnaestogodišnji – 18 letters.

Kovacic

	1 letter	Broj	2 letters	Broj	3 letters	Broj	4 letters	Broj	5 letters	Broj	6 letters	Broj	7 letters	Broj	8 letters	Broj	9 letters	Broj	10 letters	Broj
1	i	7,343	se	3,829	što	790	tako	700	svoje	305	bijaše	334	kanonik	172	gospodin	114	gospodine	57	kumordinar	233
2	u	2,935	je	2,853	sve	602	kano	563	poput	254	nikada	172	muzikaš	110	iznenada	64	gospodina	51	dobrotvora	16
3	a	1,931	da	2,697	sam	509	kako	562	nešto	195	Jožica	154	nijesam	90	kanonika	59	milostivi	41	gospodstvo	15
4	s	537	na	1,854	ali	497	nije	482	ivica	186	svojim	139	svijetu	87	nekoliko	54	zgubidane	32	rajhercera	15
5	o	326	ne	1,462	već	447	sada	382	svoga	180	svijet	103	ferkonja	87	djevojka	51	providnik	26	djevojčica	13
6	k	204	ja	866	moj	383	samo	312	svoju	177	uhvati	94	gospoda	64	muzikaša	46	dobrotvor	26	poniznosti	13
7	e	71	ti	844	ona	370	kada	297	danas	157	lijepo	88	napokon	62	naprijed	40	razumijem	21	trepetljika	12
8	nj	20	to	823	još	331	opet	291	ništa	155	rukama	88	Medonić	59	najprije	37	gospodara	21	strastveno	12
9			za	820	šta	294	onda	247	stade	144	našega	78	vrijeme	58	obrazima	37	ferkonjina	21	kažiprstom	11
10			te	810	ili	266	koji	236	glavu	137	daleko	73	poslije	54	najednom	33	Margarita	21	trepetljike	10
11			pa	804	joj	260	moja	235	Laura	129	riječi	69	starica	45	gospodar	32	gospodski	20	strastveno	12
12			mi	762	taj	216	neka	234	vidiš	123	negoli	69	svijeta	44	medonića	30	zadovoljno	19	kažiprstom	11
13			ga	649	kao	204	mene	191	ovdje	123	njegova	67	stadoše	43	ramenima	29	uhvativši	19	trepetljike	10
14			će	634	kad	202	više	187	vrata	123	Jožice	67	odijelo	37	djevojče	29	drugačije	19	neprestano	9
15			od	596	baš	180	gdje	183	dobro	120	čovjek	66	vremena	36	dječarac	26	gospodaru	19	Kičmanović	9
16			mu	579	oko	176	otac	176	prije	111	jednom	62	zajedno	36	polagano	24	Zorkovića	16	kojekakvih	9
17			li	573	dok	169	Miha	170	preko	109	veliki	60	svakoga	36	djevojke	24	Medonićev	16	providnika	9
18			on	566	oči	167	svoj	168	pravo	101	njegove	59	drugoga	35	zgubidan	21	porugljivo	14	pogledavši	9
19			iz	534	tko	156	pred	163	drugi	97	nekuda	59	osobito	33	desnicom	21	plemeniti	14	družinskoj	9
20			bi	523	oni	155	mali	160	ivice	92	natrag	59	vratima	33	rođakinja	20	gospodske	14	djevojčicu	9

The longest word with Kovačić: pedagogičkodidaktičkih– 22 letters.

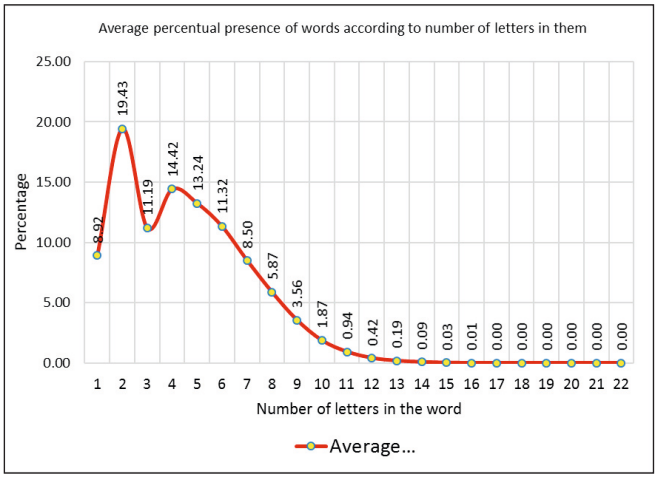
Observing individual words, it is noticeable that with all three authors there are no nouns, adjectives and numbers among the most frequent words. The dominant ones are conjunctions, auxiliary verbs, prepositions (non-lexical, function words).

With Selimović, the most frequent noun is čovjek **(man),** and it is found only at the 70th place of used words. With Nušić it is a noun **primer (example),** at the 64th place, and with Kovačić **kumordinar** at 63rd place. It is similar with the use of verbs.

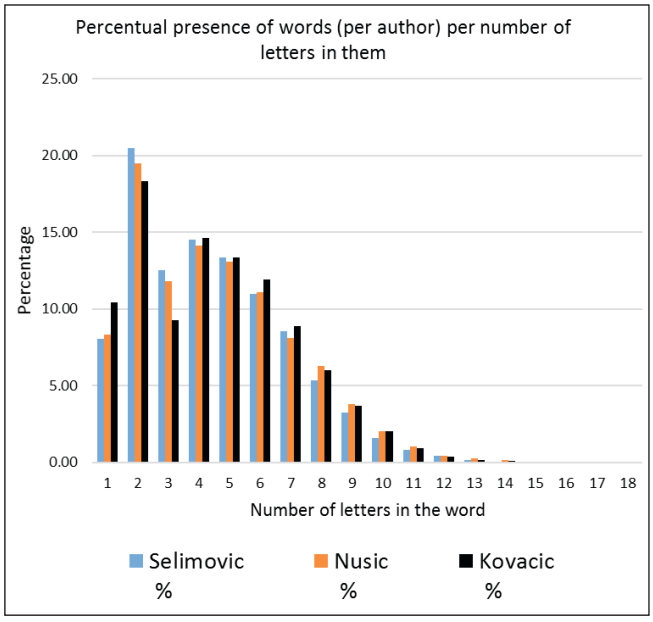
Considering that analyzed novels are of different size in terms of number of paragraphs, sentences, words and characters, the frequency of words per number of characters in them is best compared in percentages. The results are given in Table 14 and Pictures 5 and 6.

Table 14. *Absolute and percentual presence of words according to number of letters in them*

Number of letters in the word	Selimovic total	Selimovic %	Nusic total	Nusic %	Kovacic total	Kovacic %	Prosjek %
1	8,712	8.04	4,732	8.30	13,198	10.41	8.92
2	22,184	20.48	11,116	19.49	23,216	18.31	19.43
3	13,557	12.52	6,731	11.80	11,716	9.24	11.19
4	15,687	14.48	8,060	14.13	18,566	14.64	14.42
5	14,426	13.32	7,456	13.08	16,906	13.33	13.24
6	11,900	10.99	6,307	11.06	15,097	11.91	11.32
7	9,227	8.52	4,629	8.12	11,253	8.87	8.50
8	5,759	5.32	3,584	6.29	7,625	6.01	5.87
9	3,503	3.23	2,155	3.78	4,663	3.68	3.56
10	1,715	1.58	1,150	2.02	2,557	2.02	1.87
11	912	0.84	590	1.03	1,186	0.94	0.94
12	483	0.45	254	0.45	458	0.36	0.42
13	167	0.15	138	0.24	225	0.18	0.19
14	45	0.04	85	0.15	95	0.07	0.09
15	13	0.01	22	0.04	33	0.03	0.03
16	9	0.01	8	0.01	9	0.01	0.01
17	1	0.00	4	0.01	0	0.00	0.00
18	0	0.00	2	0.00	4	0.00	0.00
19	0	0.00	0	0.00	0	0.00	0.00
20	0	0.00	0	0.00	0	0.00	0.00
21	0	0.00	1	0.00	1	0.00	0.00
22	0	0.00	0	0.00	1	0.00	0.00



Picture 5. *Average percentual presence of words according to number of letters in them*



Picture 6. *Percentual presence of words (per author) per number of letters in them*

Each author has its own style of writing, characterized with different features. One of them is the length of used words. On the presented charts, we can see the similarity of distribution of percent of single-letter words, words with two letters, three letters, etc., with all three authors.

In the analyzed novels, we could perform a separate analysis of lexical density to measure how informative the text is. Lexical density is defined as number of lexical words (or content words) divided by total number of words. Lexical words give meaning to a text. Those are nouns, adjectives, verbs and adverbs. Other types of (function) words, such as

auxiliary verbs, prepositions, conjuncts, are more of grammatical nature and give little or none information about what is going on in a text.

Conclusion

Encyclopedia Britannica on language, inter alia, gives the following: "The most important individual characteristic of human language (i.e. each individual language) in relation to all other known ways of communication among animals, is its infinite productivity and creativity." [12]

Noam Chomsky, an American linguist and political author, believes that to know a language means to be capable to produce an infinite number of sentences that have never been uttered before and understand sentences that have never been heard before. Chomsky calls this feature a "creationist aspect" of language. [15]

Published expert linguistic works, printed and electronic, have analyzed language and its characteristics in many places. Naturally, analyses have most often been done for dominant world languages, Chinese, Spanish, English, Hindi, Arab, Bengali, Portuguese, Russian, Japanese [3], but works like this are rare, detailed and visually illustrated analyses for languages in the area of former Yugoslavia.

In this paper, we statistically analyzed the language of three famous writers, Mesa Selimović, Branislav Nušić and Ante Kovačić, used in their best-known novels. These authors originate from different geographical areas (Bosnia and Herzegovina, Croatia, Serbia). Novels differ in content (psychological and philosophical novel, a comedy with autobiographical content, a novel from the time of realism). The time span between publishing of the novels is almost eighty years. Nevertheless, computer analysis showed great similarities in certain elements of text.

Contemporary informatics and its branch, computer linguistics, enable analyses that were unthinkable until recently, and it will probably influence the creation of new branches of linguistics and bring about new insights of language. Considering large endangerment of languages in present world and the fact that many languages are dying out, every contribution, even the smallest one, to the study of language, particularly language of small nations, is precious.

References

[1] Avramović Zoran Ž., Marinković Dražen, Lastrić Igor, Use of Computer Search Algorithms in the Research of Statistical, Semantic and Contextual Rules of Language in Digital Information Space, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 7(2017) 1:5-8, Volume 7, Number 1, Banja Luka, june 2017 (1-56), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

[2] Čubelić, T. *Književni leksikon*. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu, 1972.

[3] Ethnologue: *Jezici svijeta*, sedamnaesto izdanje. Dallas, Teksas: SIL International. Online verzija: http://www.ethnologue.com. (24.10.2019)

[4] http://gimnazija-sb.com/portal/wp-content/uploads/2015/02/Kovačića_uregistraturi.pdf (23.9.2019)

[5] http://www.politika.rs/scc/authors/texts/901 (25.3.2019)

[6] https://avaz.ba/tag/4975/muhamed-filipovic (28.3.2019)

[7] https://biblioteka.elektronskaknjiga.com/dervis_i_smrt.php (23.9.2019)

[8] https://en.wikipediam.org/wiki/Letter_frequency (26.10.2019)

[9] https://klubcitalaca.files.wordpress.com/2010/12/branislav-Nušić-autobiografija.pdf (23.9.2019)

[10] https://norvig.com/mayzner.html (27.4.2019)

[11] https://www.ardahan.edu.tr/CUAConference2014/ (27.10.2019)

[12] https://www.britannica.com/topic/language (24.10.2019)

[13] https://www.jutarnji.hr/autori/miljenko-jergovic (28.3.2019)

[14] https://www.merriam-webster.com/dictionary/language (23.9.2019)

[15] https://www.sk.com.br/sk-chom.html (24.10.2019)

[16] https://www.vijesti.me/autor/miodrag-lekic (28.3.2019)

[17] https://www.worldometers.info/ (24.10.2019)

[18] Jahić, Dževad. *Trilogija o bosanskom jeziku*. Knjiga 3, Školski rječnik bosanskog jezika, Sarajevo: Ljiljan biblioteka Linguos, 1999.

[19] Smailović, Nedim. Statistical Analysis of Texts of the Balkans Electronic Media Columnists, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 9(2019) 1:5-16, (UDC: 659.3/.4:316.776]:004.738.5), (DOI: 10.7251/JIT1901005S), Volume 9, Number 1, Banja Luka, June 2019 (1-48), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 00

Submitted: November 21, 2019
Accepted: May 17, 2020.

ABOUT THE AUTHORS

Nedim Smailovic was born in Tuzla. He has been living in Banja Luka since 1973. He graduated from the Faculty of Electrical Engineering, department of Telecommunications. Since 1982 he has worked in RO PTT traffic of Bosnia and Herzegovina, and a series of organizational transformation it is now called Mtel doo Banja Luka. His first work experience was in designing and maintaining the PTT capacities. He obtained his Master's degree from Pan European University 'Aperion' Banja Luka, in 2005. There he also defended his doctoral thesis titled: Computer information graphics in presenting Bosnia and Herzegovina on the road to accessing the European Union. He was elected Associate Professor in 2013 and he has been teaching since in three universities in Bosnia and Herzegovina subjects relating to computer technology. He is an author and co-author of several books from the field of information technology and mathematics. He is married, father of two daughters.

Zoran Ž. Avramović is the Rector of Pan-European University APEIRON Banja Luka. He was born in Loznica (Serbia). He graduated from the University of Belgrade - Faculty of Electrical Engineering, with an average grade of 9.72 (out of 10). He received his master's degree at this faculty (average grade 10 out of 10), and then received a PhD in technical sciences (1988). He studied mathematics at the Mathematical faculty in Belgrade. He was the champion of Serbia in mathematics ("first prize") and Yugoslavia in electrical engineering ("gold medal"). He is member of the Russian Transport Academy, of the Russian Academy of Natural Sciences, of the Engineering Academy of Serbia, of the Academy of Electrotechnical Sciences of the Russian Federation.

FOR CITATION

Nedim Smailović, Zoran Ž. Avramović, Some Possibilities of Computer Linguistics on an Example of Analysis of Novels, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 10(2020) 1:5-16, (UDC: 004.82:81`322, 821.163.4`322:004.9), (DOI: 10.7251/JIT2001005S), Volume 10, Number 1, Banja Luka, June 2020 (1-68), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Improving the Spatial Data Quality in the Geographical Information System of the Telecom Operator

Dragan Čubrilović[1], Miloš Ljubojević[2]

[1]*Mtel d.o.o, Banja Luka, Bosnia and Herzegovina, dragan.cubrilovic@mtel.ba,*

[2]*Faculty of Electrical Engineering, University of Banja Luka, B&H; ljubojevic.milos@gmail.com*

Contribution to the State of the Art

https://doi.org/10.7251/JIT2001017C UDC: 007:528.9]:004, 007:912]:004

Abstract: Spatial data about telecommunication infrastructure facilities represent the inevitable resources of each telecom operator. The precision of collected spatial data, used in the geographic information system (GIS) of telecom operators, is very important, especially when it is about urban environments. In this paper, we have presented the possibility of correcting the positions of telecommunication facilities obtained using the Global Positioning System (GPS). Factors affecting the accuracy and quality of spatial data have been analyzed and solutions for quality improvement proposed. We have shown that using permanent stations can achieve the required level of spatial position correction. A fast and efficient position correction allows updating data in GIS of telecom operators, providing correct, accurate, and timely information about telecom infrastructure.

Keywords: GPS, GIS, spatial data quality, differential correction, permanent station.

Introduction

Modern and competitive environment causes that telecommunications companies are finding a way to adapt to new trends and technologies. Along with the growth and development of technology, the number of users of telecom operators is growing. With the increase in the number of users and networks of telecommunication, infrastructure becomes more complex. This the main reason why telecom operators are finding the most appropriate model for efficient monitoring and management of telecommunications infrastructure. Therefore, it is necessary to provide accurate spatial and attribute data on elements of the communication infrastructure that are easy to access, analyze, and optimally present.

For telecom operators, geographic information systems or GIS represents a good solution to business challenges related to the surveying and representation of spatial data on the existing telecommunications infrastructure, as well as a solution to capture the geographical location data about their users, i.e. service subscribers. Spatial data represent information about objects that contain the geographical location of these objects and contains attribute (descriptive) information about those objects. The great advantage of GIS is that it allows the integration of data that has been collected using different acquisition techniques.

To improve the quality and precision of spatial data in GIS, a GPS is used for their collection. GPS system enables efficient collection of accurate and up-to-date position data of objects of interest. Considering the problems related to the accuracy of obtained data, previous experiences have shown that Differential Global Positioning Systems (DGPS) method is the optimal positioning solution. For acquiring the precise GPS position, this method is using also the additional receivers whose coordinates are already known and where permanent GPS stations are located. This approach enables to subsequently correct and adequately integrate the ob-

tained data into various applications, in this case, the telecom operator's GIS system [7].

The accuracy and quality of the spatial data used by telecom operators are crucial for the use-value of GIS. Their analysis is an important task of this paper. Therefore, the main research interest of this paper is improving the quality of spatial data and appropriate visualization of spatial data and their attributes.

This paper is organized as follows: After the introduction section, in the second section, the main important characteristics of GIS systems used by telecom operators are presented. In the third section, the methodology for improvement of the quality and precision of spatial data using permanent stations is presented. The experimental results and discussion of obtained results are presented in the fourth section. Finally, the conclusions and future research directions are outlined.

The Characteristics of the Telecom Operators Gis Systems

Geographic Information System is using for collecting, storing, searching, analyzing, and displaying geographic data. Each object within the GIS system refers to a specific geographic location and can be adequately mapped, meaning that all data in the GIS is georeferenced.

Benefits of the GIS use for telecom operators

Every object in the GIS is, in addition to the location, defined by a set of attributes that are describing the characteristics of that object. Those attributes are defining both, the graphical and non-graphic data that are important for GIS system use. [2, 3].

In order to manage and to display the data in the urban planning process with the GIS system, all necessary data should have a spatial component. It means that they should be spatially oriented (they are associated with the corresponding coordinates, address, or name of the area) and can be adequately shown on the map [4].

Telecommunication companies belong to the group of companies that own and manage objects of complex geometric shapes and positions in space. Whether they are copper or fiber optic networks, cable ducts, shafts, connectors, or users themselves, the data that are describing them are incomplete without additional information about their location in the space. The information system that is used for the management of telecommunication resources must be capable to effectively combine the capturing of geometric and spatial features with classical object descriptive data and data on relationships and connections to other objects. Besides the logging and listing of objects, the geographic information system must be able to display data to different user groups, in a way that integrates tabular and spatial views through interactive maps [18].

This research is closely related to a geographic information system that is using in the Mtel company and it is continually upgrading. The main task of the GIS that is implemented for Mtel is to capture all physical elements of the infrastructure in a spatial and descriptive (attribute) form. The process of logging existing infrastructure through GIS depends most on the quality and reliability of existing documentation. The key task of the GIS project implementation was the introduction of GIS software, with the name TeleCAD-GIS software, as the main tool for electronically documenting the current state of telecommunications infrastructure in Mtel. This implies that all documentation of the currently realized installations must be entered into the GIS software as soon as it is received, as well as any changes resulted from regular, preventative, or interventional maintenance [9].

The benefits of accurate information related to spatial data on telecommunications infrastructure and geographical location information of the users themselves for telecom operators are multiple. This data are important because they provide real-time information about the network structure and thus enable the telecom operator to monitor the network, test network elements, plan network capacity, maintain, as well as provide timely service and eliminate any interference or problems. Spatial data on telecommunications infrastructure, in addition to providing detailed information about existing users (status and history of users, existing network infrastructure, signal quality in a specific demographic area, etc.), can also be used to determine telecommunication market demand in the future [2, 6].

Collecting GPS data for the GIS system

There are several ways to collect GIS data: digitizing or scanning existing maps, manually entering text into databases, using aerial photos and photogrammetry, transferring files from other sources, etc. Still, the GPS technique proved itself as one of the fastest and most accurate techniques for collecting field data.

The geographical accuracy of the existing data is often a problem during the implementation of GIS. Poor quality of existing data, inaccurate instruments or operator error reduces position accuracy below the required level that is necessary for use in the telecom operator business. All the data entered into the GIS must be accurate. This ensures both the overall accuracy of the GIS and an effective system to provide support in service providing or system maintenance and optimal decision making. Therefore, the GPS system is a great and easy for use tool for collecting accurate and precise geographic data [11, 12]. In addition to these many benefits of GPS for collecting GIS data and creating maps, it should be noted that GPS data is also often used to control and verify existing maps and GIS products.

Improving the quality and precision of spatial data using permanent stations

Permanent stations are a great solution, both for their ease of use and for the very good results they provide for correction of the measured GPS position. Unlike Telekom Srbija, where the entire network of the permanent station is used, the Mtel company uses few individual permanent GPS stations in the Republic of Srpska. They are used for generating a differential signal to perform the correction of GPS measurements. In this way, satisfactory GPS measurement accuracy is obtained and the ability to create spatial data of high accuracy about objects is achieved.

The permanent stations that are used in Mtel contain a GPS receiver, an indoor unit, a separate robust GPS antenna for facility mounting, and a permanent station software. The permanent station used at Mtel is the Trimble 5700 L1 permanent station. GPS permanent station the Trimble 5700 L1 is a GPS receiver that is very easy to use and specially developed for geodetic surveying. This system offers modern Trimble GPS technology flexibly and cost-effectively [19]. One permanent station can simultaneously provide service data for the correction of an unlimited number of field devices.

This research paper presents how permanent stations can contribute to improving the accuracy of spatial data captured by GPS devices. The Mtel company has a total of three permanent stations, so the choice of locations for permanent stations depended on the radius they can cover and the geographical area of the Republic of Srpska, which needs to be covered by the GPS correction signal. Therefore, these three permanent stations were set up at appropriate locations: one in Banja Luka, the other in Lopare, and the third in Gacko. Due to this arrangement of those permanent stations, the signal for differential correction is available in a large part of the territory of the Republic of Srpska, so coverage is over 85% of the territory [1]. For the purpose of the experimental investigation in this paper, a permanent station in Lopare was used as a reference station.

Differential correction of GPS signals

GPS positioning accuracy depends on various factors, such as atmospheric effects that affect the speed of radio-waves propagation, multiple signal paths, inaccuracy of satellite clocks, inaccuracy of satellite position data, numerical errors in calculation, satellite speed, Earth's gravity, etc. GPS positioning accuracy is also affected by radiation that can be natural (radiation from the Sun and geomagnetic storms) or artificial (powerful TV antennas nearby, GPS signal generators, etc). When all the above factors are taken into account, the total positioning error can be over 10 m [2,7].

Due to the above factors, there is a real need to improve the accuracy of GPS positioning by using additional correction techniques. Differential correction is one of the techniques that significantly increases the accuracy of GPS data collected, and its use requires the simultaneous use of a receiver at a known location, i.e. the use of permanent base stations and collection of GPS positions at unknown locations, with other moving receivers [2, 10].

For DGPS, receivers are used on fixed, i.e. permanent stations whose coordinates are already known.

The current position of these receivers is constantly calculated based on the current GPS signals, and this calculated position is compared to the actual/real, i.e. known position. Due to the influence of various error factors described above, the actual and calculated positions are not the same. By comparing the values of these two positions, an error is obtained in calculating the position at the permanent station location and it is assumed that the same type and the same or very similar error value will occur at locations near the permanent station. This means that data collected at a known position can be used to determine the errors contained in satellite data at locations close to that known position. GPS receivers near these fixed stations subsequently receive this information and use it to correct the error of their position. There are two basic techniques for differential correction: real-time differential correction and post-processing differential correction [8, 15].

In the experimental research, differential correction with post-processing was used, in which the data are taken from the permanent base station, and the data collected by the mobile GPS receiver (rover) are subsequently processed. This gives a differential corrected position for the captured spatial data. It is assumed that, the closer the receivers are to these permanent stations, the more corrected their measurements will be, and the accuracy of the correction decreases as mobile receivers move away from the location of the permanent station. One of the tasks of this paper is to analyze the influence of the distance of a permanent station from the rover on the precision of positioning and error correction.

Methodology for Improving the Positioning Accuracy and Differential Correction With Post-Processing

As presented earlier, several factors affect positioning accuracy and differential positional correction. Considering the importance of precisely positioning objects of interest to telecom operators, and especially in the case of urban areas, the subject of research is the identification and analysis of those factors that affect differential correction with subsequent processing. This technique is often used in everyday work as part of increasing the accuracy of the data collected for telecom operator GIS systems.

Methodology for testing different impacts on the quality and precision of spatial data

In the experimental part of the paper several factors that affect positioning precision and spatial data quality, when using the post-processing differential correction methodology, were identified and analyzed. On this basis, a methodology for testing different impacts on the quality and precision of spatial data has been proposed. Following the adopted methodology, the following influences were analyzed in details:

- The effect of the distance of the rover from the permanent station on the percentage of error correction for the collection of telecommunication infrastructure data
- The influence of the selected surveying method (continuous surveying - *Kinematic method* or surveying with breaking points - *Stop&Go method*) on positioning accuracy is analyzed.

Examining the above impacts represent major research tasks in the experimental part of the paper. Within the experimental part of the paper, the main focus was on those levels of positioning accuracy that are relevant to the needs of the telecom operator. Those levels of accuracy have a strong influence on the area of implementation of the GIS system regarding the characteristics of telecommunication infrastructure.

The methodology for testing the previously described factors was proposed considering the different types of measurements. Measurements were defined depending on the identified factors that affect positioning precision.

Based on the previous analysis, two different types of measurements were performed:

1. Measurements at three locations, in cities located at different distances from the permanent station used in the experimental research - the permanent station in the city of Lopare. Measurements were made in the cities of Banja Luka, Ugljevik, and Bijeljina. In this way, the influence of the distance of the rover from the permanent station (for distances of about 15km - Ugljevik, about 31 km – Bijeljina and over 100 km - Banja Luka) was examined.
2. Measurements that have to take into account the impact of the surveying method (continu-

ous surveying along the route or surveying at breakpoints) for line objects, for each of the three locations where measurements were taken.

Therefore, when considered the measurement techniques described above, measurements were performed for telecommunication objects of interest in three different cities, using two different methods. For each of these measurements, using the post-processing differential correction technique, the originally obtained GPS position was corrected.

The principle of collecting data about objects of interest

The methodology of GPS measurement and data collection covers several activities. It mainly consists of creating new files, selecting codebooks, collecting data in the field using appropriate surveying methods, and storing the collected and properly corrected data in the created files. One of the important contributions of this paper is the analysis of the precision and efficiency of different surveying methods and their impact on the quality of the obtained spatial data. In this paper, two commonly used methods and their dependence on various factors are discussed.

In the experimental part of the research were used a GPS device, GPS permanent station, and software applications necessary for collection and correction of the collected GPS data. The hardware and software configuration of the GPS device is a combination of a Trimble GPS receiver and a PDA (Personal Digital Assistant) with a Microsoft Windows Mobile operating system. TerraSync software is used to measure positions and collect attribute data of objects of interest.

Trimble 5700 L1 permanent station with single-frequency GPS antennas Trimble A3, located at city Lopare, was used to increase the accuracy of spatial data on telecommunication infrastructure. The Trimble GeoXH GPS mobile device was used as the rover receiver, which enables the accuracy of code measurements in the range of 0.3 m to 1 m and accuracy of phase measurements in the range of 0.1 m to 0.2 m. These are field devices used to operate at Mtel and they are GPS devices for spatial data acquisition for the telecom operator's GIS system. It is important to mention that the preparation of the device for fieldwork, in addition to the initial settings, requires the transfer of appropriate coordinate systems and georeferenced substrates to the device.

The methods of data collection depend on the type of objects of interest, how the positions are collected, and the number of points of interest whose exact positions affect the quality of the spatial data of the object being collected. Collecting line object positions (the same goes for polygon objects) can be done in several ways:

- By surveying individual break points of an object;
- Continuous surveying of the object (the operator moves along the route of the object, while the GPS device registers positions at regular time intervals);
- Map Digitization - this method is used as an alternative to registering GPS positions. Digitization from a map implies that a georeferenced map is positioned in place for a particular location, which is loaded as a background file. The cursor on the screen marks the positions of the breaking points of the line (polygon) object, following the actual situation on the ground, and the route is automatically plotted on the map. This method is not used in the practical part of the paper, so it will not be analyzed in more detail.

The method based on surveying with breaking points - *Stop&Go method*

This procedure is using for capturing the points along with the entire object, which is subsequently connecting by straight segments, and this merging is performing automatically by the software. The procedure for capturing data consists of several related tasks.

First, the operator is positioning at the location of the starting point of the line object, which practically represents the first point to be captured. A GPS device is initialized to detect a sufficient number of satellites, and then the operator is defining a line object to be captured. The object consists of multiple points of interest. The principle of operation of TerraSync software that is installed in the GPS device enables that, after defining and opening the line ob-

ject and selecting the point of interest an adequate form will be opened. Using this form operator is entering necessary attribute data for the opened, selected object. In the device setup itself, it is useful to include the *Repeat* option, which makes it easy to enter attribute data by entering values from the previous object in the attribute fields when opening each subsequent object of the same type. In practice, this is a very useful feature that saves a lot of time during the GPS survey, since these values are in most cases very similar to objects of the same type [14, 20].

The following figure shows a graphical representation of the surveying procedure of line objects by using the point of interest method, with a detailed description of the tasks to be accomplished when measuring.

The software automatically starts capturing GPS positions for the first point of interest, i.e breaking point, and as final result calculates the arithmetic mean of all captured positions. After capturing the sufficient number of measurements (the more, the more accurate), the operator manually finishes calculation the position of the first point of interest.

In the experimental part of the research, during the surveying of line objects by the method of capturing the point of interests, ie. breaking points. A characteristic of this method is that, even after the first recorded point of interest, the line object remains open, and the continuation of the measurement implies that the operator is moving to the location of the next point of interest and repeats the procedure. The procedure is repeated successively to the last point of interest that belongs to the line object being surveying. The attributes that are entered in on the first point will be valid for the entire line object until it is closed, no matter how many points of interest were captured. After closing the last breaking point, the complete line object that was the subject of the shooting is being closed [20].

Therefore, this method is characterized by the fact that the line object closes and the measurement ends only after completing all points of interest along with that line object, regardless of how many points of interest the object contains.

The method based on continuous surveying - Kinematic method

This procedure is used in cases where it is necessary to survey a line object by moving the operator along its route, while the GPS device, with a predefined time interval, captures positions during the movement of the operator. The process of surveying a linear (polygonal) object using the aforementioned method is started by selecting the line object that needs to be surveyed. After defining the line object and selecting the method of surveying by continuous method, the necessary attributes are entered in and the capturing of the precise position is starting.

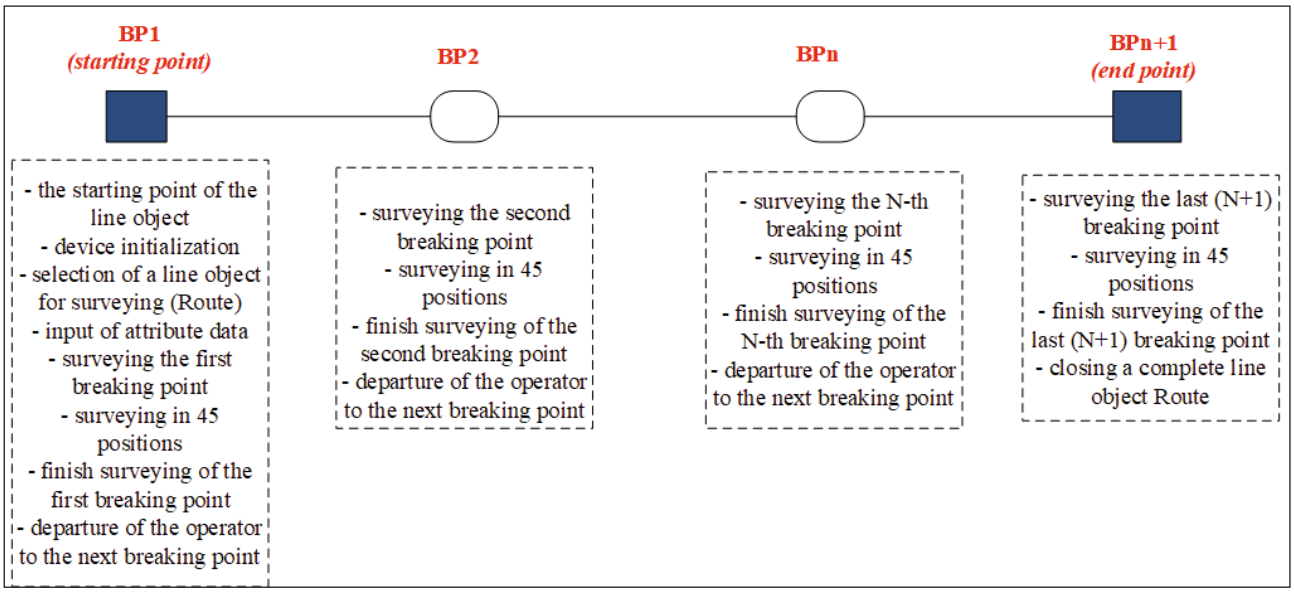


Figure 1: *Method based on surveying with breaking points - Stop&Go method*

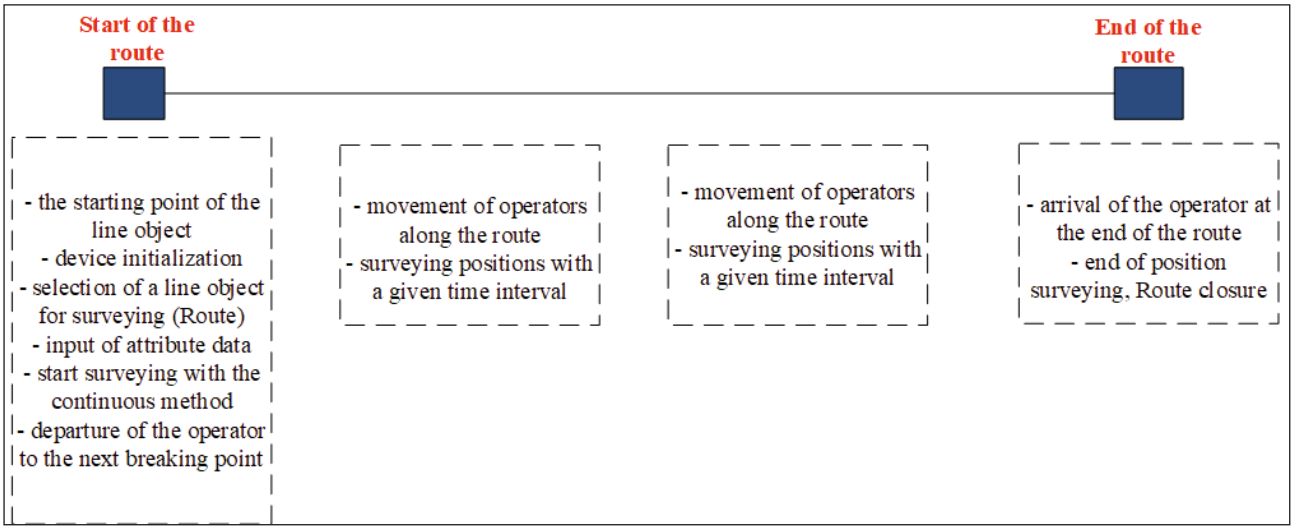


Figure 2: *Method based on continuous surveying - Kinematic method*

The GPS device starts capturing the positions, captures positions with a predefined time interval (this is usually one captured position per second, though this recording interval can be adjusted as needed). This implies that the operator in the field is moving along the object, and only upon reaching the endpoint of the object that is surveying it is being closed [20]. An illustration of this surveying method and the operator's tasks when using the continuous surveying method is given in Figure 2.

Practical aspects of differential GPS correction technique with post-processing

After completed practical measurements on the field with the rover receiver (Trimble GeoXH), the measured data files represent the basis for the post-processing differential correction process. The methodology for differential correction of GPS positions requires two important datasets. First are data that were collected by the rover receiver and second a set of base files from a permanent station, whose position is closest to the location of the field data recording with the moving receivers. These files obtained from a permanent station must be generated at the same time as the field data files were collected.

There are several ways of processing data in the post-processing differential correction process, depending on what kind of base file set is available. Although *Carrier and Code Processing* is the most complete processing method that produces the most accurate results (because it also contains the phase carrier data), this paper uses the *Code Processing* data processing method, since the base file from Trimble 5700 L1 permanent station does not have data for phase carrier [13].

The differential correction of field data was performed using the Pathfinder Office software package. As first, step the field data files that will be differentially corrected must be selected, and then the processing type and the set of correction files from the nearest permanent station can be used. Thereafter, the software initiates differential correction and after completing the differential correction process, a report was generated. The results of that correc-

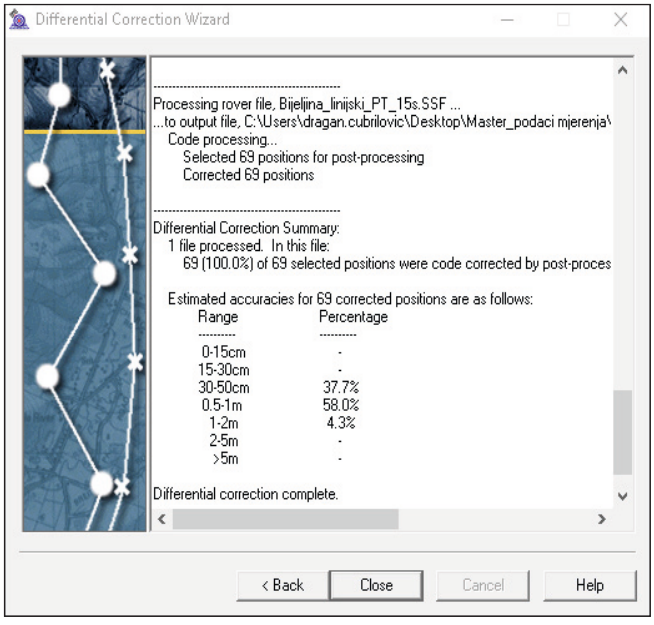


Figure 3: *The report on the differential correction*

Table 1: *The results of differential correction considering the selected surveying method*

Location of measurement	Method of measurement	Correction for range 0.3-0.5m; [%]	Correction for range 0.5-1m; [%]	Correction for range 1-2m; [%]
Ugljevik	Continuous measurement	32.7	67.3	0
Ugljevik	The point of interest measurement	60	40	0
Bijeljina	Continuous measurement	15.9	59.1	25
Bijeljina	The point of interest measurement	37.7	58	4.3
Banja Luka	Continuous measurement	0	62.5	37.5
Banja Luka	The point of interest measurement	0	72.6	27.4

tion are presented in Figure 3 [13]. An example is presenting a report on the differential correction of data captured at the Bijeljina location, using the breakpoint method with a holding interval of 15 seconds.

From the differential correction report itself it can be seen that in the specific example, 100% of the positions have been corrected, in which case there is no need to analyze the correction note. A correction note is a corresponding file that contains a complete history of data processing by differential correction. If all positions are not corrected, this note gives the user an insight into a more detailed description of the correction itself and helps to find the source of the error as well as to resolve it. Data corrected like this can be exported to TeleCAD-GIS that is used in the company Mtel.

Results and Discussion

Based on the described methodology for testing different impacts on the quality and precision of spatial data of interest for the telecom operator, the obtained results are analyzed and systematically and graphically presented for each of the individual measurements, at each site. Comparative analysis of the obtained results of differential correction revealed the influence of previously identified factors.

The influence of the selected surveying method on the accuracy and quality of spatial data

To examine the impact of different surveying methods, the results of the continuous surveying method were compared with the surveying method which takes into account the point of interest. In doing so, an occupation time of 45s for every point of interest was used when capturing with Stop&Go method. Although these are two different surveying methods, it is important to note that both methods were used for capturing the same type of objects, along the same route. Also, only line objects were captured.

The Stop&Go method involves capturing the individual point of interest along with the entire object, which the software subsequently interconnects with straight segments, while in continuous capturing the line object is surveying by the operator moving along the line object, and the GPS device with predefined time interval is capturing the positions during the movement of the operator.

Surveying of line objects of interest for the telecom operator, using the methods described above, was realized at three different locations. A total of 6 measurements were taken, three for continuous surveying method and three for surveying objects with the point of interest method (Stop&Go method). After collecting GPS positions for line objects, differential correction of GPS positions was performed with post-processing. A comparative overview of the obtained differential correction results for all three locations for the previously described surveying methods is presented in Table 1.

To make a clearer and more transparent analysis of the comparison of the results of differential correction of continuous surveying of GPS positions of line objects and surveying GPS positions of line objects with Stop&Go method, the following figures provide a graphical presentation of the obtained results, Figure 4.

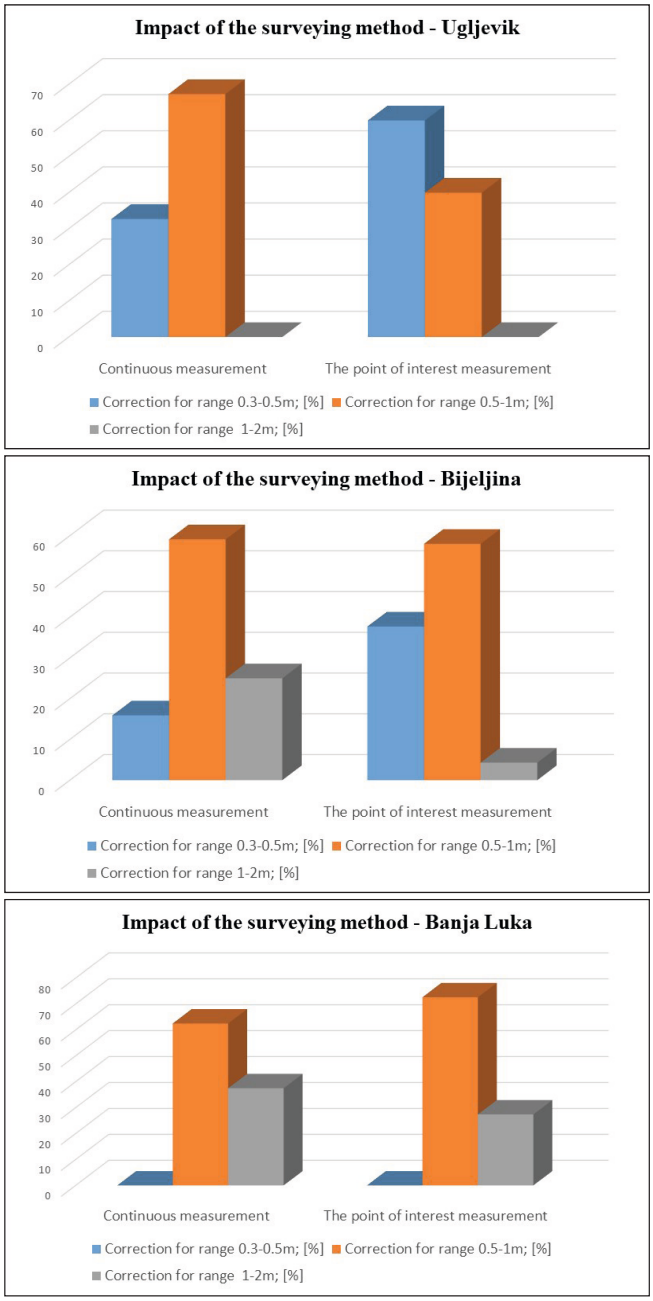


Figure 4: *The graphical presentation of results of differential correction of data depending on the surveying method for different locations (Ugljevik, Bijeljina, Banja Luka)*

The continuous surveying method is not exactly the most reliable and is rarely used to capturing telecommunications infrastructure at Mtel. There are several reasons for this: the speed at which an operator moves with a device from point A to point B is not constant, the significant influence of reflected signals mixed with direct signals from the satellite - multipath signals, the captured route itself may be such that it is simply impossible for the operator to cross it on foot - natural obstacles, bays, rivers, etc.

The Stop&Go method is significantly more accurate and more convenient to use than the continuous surveying method (kinematic method) for detailed geographic surveying. One of the main reasons is that the final position of the data collected by the point of interest method is obtained from the mean values of the positions of that data, which were collected over some time, at each collection point. Generally, the point of interest method provides more accurate solutions compared to the continuous method since the GPS capturing time in the point of interest method is a minimum of several seconds (in the experimental part of the work, the capturing interval was 45 seconds) at each captured point (which automatically results in a smaller error rate), while GPS capturing duration is only 1 second per point captured for the continuous surveying method [17].

The obtained results show, and the graphs clearly are illustrating, that the Stop&Go method is more accurate than the continuous method. This was especially expressed for capturing in Bijeljina and Ugljevik sites, at locations that are closer to the permanent station. At a more distanced remote location in the Banja Luka site, the Stop&Go method is also more accurate. The presented comparisons of the obtained results of differential correction, depending on the surveying method, are completely in line with the results presented in previous studies presented in the literature [5, 17].

This research has shown that, in addition to being more accurate, the Stop&Go method is a significantly more stable method for measurement. As the rover's distance from the permanent station increases, much smaller deviations are obtained with this method than with the measurements obtained by the continuous method. The results show that by using the continuous method, with increasing distance from the base station, the dominant correction remains in the range of 0.5-1 m, but neighboring bands of 1-2 m appear with a slightly higher percentage, with less accurate corrections. From this, it can be concluded that the continuous method has a higher dispersion of the measurement results than the Stop&Go method.

Table 2: *The rover distance influence- surveying of line objects with the continuous method*

Location of measurement	Type of objects	Method of measurement	Correction for range 0.3-0.5m; [%]	Correction for range 0.5-1m; [%]	Correction for range 1-2m; [%]
Ugljevik	line objects	Continuous measurement	32.7	67.3	0
Bijeljina	line objects	Continuous measurement	15.9	59.1	25
Banja Luka	line objects	Continuous measurement	0	62.5	37.5

Table 3: *The rover distance influence - surveying of line objects with the Stop&Go method using the capturing interval of 45 seconds at each point of interest*

Location of measurement	Type of objects	Method of measurement	Correction for range 0.3-0.5m; [%]	Correction for range 0.5-1m; [%]	Correction for range 1-2m; [%]
Ugljevik	line objects	The point of interest measurement	59.2	40.8	0
Bijeljina	line objects	The point of interest measurement	47.3	52.7	0
Banja Luka	line objects	The point of interest measurement	0	93.1	6.9

Table 4: *The rover distance influence - surveying of line and dot objects with the Stop&Go method using the capturing interval of 45 seconds at each point of interest*

Location of measurement	Type of objects	Method of measurement	Correction for range 0.3-0.5m; [%]	Correction for range 0.5-1m; [%]	Correction for range 1-2m; [%]
Ugljevik	line + dot objects	The point of interest measurement	18.4	81.3	0.3
Bijeljina	line + dot objects	The point of interest measurement	0	96.3	3.7
Banja Luka	line + dot objects	The point of interest measurement	0.2	99.1	0.7

The influence of the distance of the rover from the permanent station on the accuracy and quality of spatial data

Another very important factor that affects the accuracy of positioning and correction is the distance of the rover, at the time of data collection, from the permanent station, the so-called "baseline length". In the experimental part of the paper, an analysis of this effect was carried out, to indicate the need for increasing the number of permanent stations, to improve the precision of positioning of telecommunication objects, and therefore the quality of spatial data in the GIS system of telecom operators.

Based on the proposed research methodology, three groups of measurements were performed for different rover distances from the permanent station, and measurements were performed for different methods. In the first group of measurements, the distance of the rover from the permanent station is about 15 km (location Ugljevik). The second group of measurements was realized at a distance of the rover from the permanent station in the range from 30 to 35 km (location Bijeljina), and the third in the range over 100 km (location Banja Luka). After the measurements were realized and the differential correction was performed with subsequent processing, the obtained correction results were systematized and presented in tables and with graphs.

A comparative overview of the results of differential correction at different locations, with different surveying methods, as a function of the distance of the rover from the permanent station, is shown in Tables 2, 3, and 4.

To make a clearer and more transparent analysis of the comparison of the results of differential correction of surveying GPS positions of line and line-dot objects by different methods as a function of the rover distance from the permanent station, the following figures give a graphical presentation of the obtained results, Figure 5.

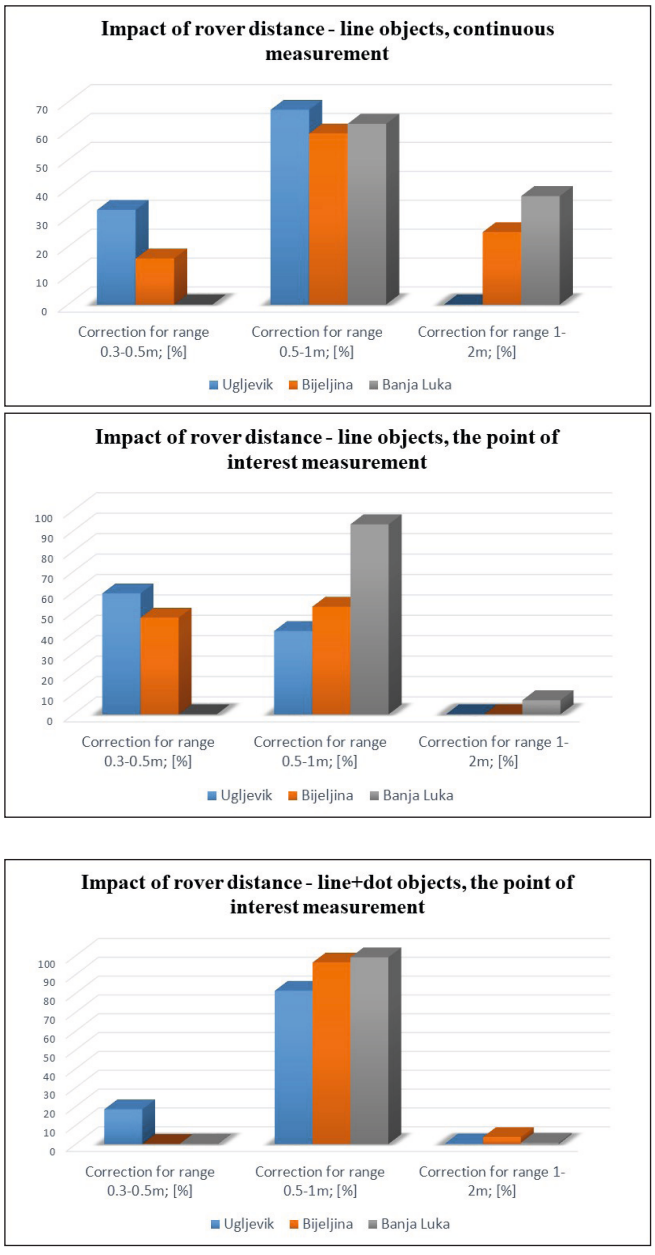


Figure 5: *The graphical presentation of results of differential correction of data depending on rover distance and illustrated for different surveying methods*

The analysis of the obtained results shows and the graphs illustrate, that the efficiency of differential correction is negatively affected by an increase in the distance of the rover from the permanent station (i.e. an increase of the *baseline*), which is in line with the results of previous studies presented in the literature [16, 21]. This degradation of the correction results with increasing distance of the rover from the permanent station is understandable, given the principle on which the differential correction is based. Also, the fact that the same type but also approximately the same range of error values will occur at locations near the permanent station as at the location of the permanent station itself, which is later used to correct data recorded by the rover receiver, will have a significant impact.

The further away the rover receiver is from the location of the permanent station, i.e. the longer the baseline, the more factors affecting the measurement error and the conditions for field surveying, especially the atmospheric one, are increasingly changing compared to those related to measurements at the permanent station location. It is expected that the error occurring at remote location measurements is different from the error occurring at the permanent station location and that the error at the permanent station location cannot be effectively used to correct the positions of too remote rover receivers.

From the graphs, it can be seen that the correction of those measurements performed at smaller distances from the permanent station, enables the correcting of the accuracy of the positions for lower ranges, such as the 0,3-0,5 m range. In the case of longer rover distances from the permanent station, this is not achieved, i.e. at the Banja Luka location, which is more than 100 km away from the permanent station, no correction was made for this range of measurement accuracy.

Also, it was mentioned earlier in the paper that the coverage radius, for which the permanent station provides the best results, is about 70 km, which, by analyzing the obtained results of differential correction of data in Banja Luka, proved to be correct. These measurements were made at a rover distance from a permanent station of more than 100 km, and although these measurements have been corrected to some extent and a certain percentage, they still do not fully meet the required level of data accuracy for the telecom operator. Based on the above, it can be concluded that the need to increase the number of permanent stations, which would be used to correct and improve the positioning accuracy of telecommunication infrastructure, is fully justified.

Conclusion

The precision of spatial data positions is one of the most important information in a telecom opera-

tor's GIS system, especially when it comes to underground telecommunications infrastructure in urban areas. The possibility of correcting the positions of telecommunication facilities obtained using the GPS represents an important research direction.

In this research paper, the various impacts on the quality and precision of spatial data that are of interest to telecom operators were identified and a methodology for their analysis and possibilities for quality improvement was proposed. The results obtained by differential correction are systematically and graphically presented for each of the individual measurements at each location. A comparative analysis of the results of the differential correction revealed the influence of the most important factors.

To examine the impact of different recording methods, the results of the continuous surveying method were compared with the surveying method which takes into account the breaking points, and from the analysis of the obtained results, it can be concluded that the Stop & Go method is more accurate and accurate than the continuous one.

Also, the influence of the distance of the rover at the moment of data collection from the permanent station (so-called "baseline length") on the precision of positioning and correction was analyzed in detail. The analysis of the obtained results shows that the efficiency of the differential correction is negatively affected by the increase in the distance of the rover from the permanent station. It has been shown that by reducing the distance, an increasing correction percentage can be achieved for the accuracy bands that meet the needs of the telecom operator. It can be concluded that, by approaching the rover to the permanent station, besides increasing the percentage of corrected positions of interest to the telecom operator (positions with an accuracy of 0.5-1 m), position correction was achieved with more accurate levels of accuracy.

This impact was analyzed to point out the need to increase the number of permanent stations, to improve the precision of positioning of telecommunication objects and, consequently, the quality of spatial data in the telecommunication operator's GIS system. Therefore, it is desirable to constantly increase the number of permanent stations. In this way, a significant reduction in the distance of the rover to the nearest permanent station can be achieved, thereby providing the best position correction. The position obtained by the GPS and corrected by the use of the post-processing differential correction technique effectively can be integrated with other information in the telecom operator's GIS, achieving constant improvement of the spatial data quality.

Future research directions are related to the analysis of the impact of different GPS positioning approaches and methodologies on the effectiveness of differential positional correction. Improvement of the quality of spatial data from the constant development of a dedicated GIS data dictionary also represents future directions of interest.

REFERENCES

[1] Cubrilovic D i Maksimovic S (2010) "Prijedlog tehničkog rješenja za postavljanje permanentnih stanica za korekciju GPS mjerenja u Telekomu Srpske", tehnička dokumentacija, Telekom Srpske, Banja Luka

[2] Cubrilovic D, Ljubojevic M, Pantovic V (2019), Poboljšanje preciznosti prostornih podataka u geografsko informacionom sistemu telekom operatora, ICT conference & exibition, INFOTECH, pp 1-6, Aranđelovac, Srbija.

[3] Djuric Z, Radulovic V, Radulovic C (2012), Online GIS i kolaborativno mapiranje kao novi trendovi u razvoju GIS-a i njihova primjena u savremenom informacionom sistemu, INFOTEH-Jahorina vol. 11, pp 621-626, Jahorina, BiH.

[4] Djuzo F i Taletovic J (2007), GIS u prostornom planiranju, I kongres o katastru u Bosni i Hercegovini, pp 1-12, Neum, BiH.

[5] Farah A (2017), Accuracy assessment study for kinematic PPP using low-cost GPS receiver, Al Azhar 14th international conference on: engineering, architecture, technology, Cairo, Egypt.

[6] Fry C, (1999), GIS in telecommunications. In P. Longley, M. Goodchild, D. Maguire, and D. Rhind, eds., Geographical information systems volume 2: Management, John Wiley and Sons, New York, USA.

[7] https://www.gps.gov/ (accessed on 20.04.2020)

[8] https://www.trimble.com/ (accessed on 25.04.2020)

[9] INOVA (2010), TeleCAD-GIS Enterprise - Implementacija za Telekom Srpske, tehnička dokumentacija, INOVA informatički inženjering d.o.o., Srbija.

[10] Jha K, Singh S, Upadhyay N and Khare N (2016), Comparative study of online GPS post processing services and effects on DGPS data processing, International research journal of management science & technology, vol 7, issue 1, pp.29-35.

[10] Lange AF and Gilbert C (1999), Using GPS for GIS data capture. Geographical information systems (ed. by P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind), 2nd edn, pp. 467–476. John Wiley and Sons, New York.

[12] Livona d.o.o (1998), Sistemi za kartiranje-mapping systems, tehnička dokumentacija, Livona d.o.o. Trimble navigation limited, Beograd. Srbija.

[13] Livona d.o.o (1998), Pathfinder Office - uputstvo za upotrebu, tehnička dokumentacija, Livona d.o.o. Beograd, Trimble Navigation Limited, Srbija.

[14] Livona d.o.o (2006), Uputstvo za snimanje telefonskih instalacija sa real-time korekcijom sa AGROS GPS mreže, tehnička dokumentacija, Livona d.o.o., Beograd, Srbija.
[15] Mahdi A S (2018), Accurate three dimensional coordinates measurements using differential GPS real time kinematic mode", Iraqi journal of science, vol.59, no.2C, p.p. 1146-1151.
[16] Matsushita T, Tanaka T (2010), Improving Measurement Accuracy of Long Baseline Differential GPS, SICE journal of control, measurement, and system integration, Vol. 3, No. 3, pp. 157–163.
[17] Sestić M (2016), INOVA GIS platform and TeleCAD-GIS 2016, INOVA informatički inženjering, d.o.o., Banja Luka, BiH.
[18] Shouny A E, Yakoub N and Hosny M (2017), Evaluating the performance of using PPK-GPS technique in producing topographic contour map, Marine geodesy, vol 42, no 4, pp 224-238.
[19] Trimble (2003), 5700/5800 GPS receiver user guide, Trimble Navigation Limited, USA
[20] Trimble (2007), GPS receiver Trimble GEOEXPLORER 2005-user guide, Trimble Navigation Limited, USA.
[21] Weston N D, Soler T and Mader G M (2009), Rover station positional accuracies from OPUS as a function of reference station spacing and rover station occupation time, Proc. FIG Working Week, Eilat, Israel.

Submitted: March 11, 2020
Accepted: May 17, 2020.

About the authors

Dragan Čubrilović received BSc degree, majoring in Electronics and Telecommunications, at the Faculty of Electrical Engineering, University of Banja Luka, Bosnia and Herzegovina. He is currently attending postgraduate studies at the Faculty of Electrical Engineering in Banja Luka, for obtaining a MSc degree. He has been employed at Mtel since 2008.

Miloš Ljubojević, PhD received a BSc and MSc degree at the University of Banja Luka, Faculty of Electrical Engineering. He received a PhD degree in the field of information technology at the University of Belgrade, Faculty of Organizational Sciences. He is employed at the Ministry for Scientific and Technological Development, Higher Education and Information Society, Government of the Republic of Srpska, and as an assistant professor at the University of Banja Luka.

For citation

Dragan Čubrilović, Miloš Ljubojević, Improving the Spatial Data Quality in the Geographical Information System of the Telecom Operator, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 10(2020) 1:17-29, (UDC: 007:528.9]:004, 007:912]:004), (DOI: 10.7251/JIT2001017C), Volume 10, Number 1, Banja Luka, June 2020 (1-68), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

ANALYSIS OF STAGES OF DEVELOPMENT, CURRENT STATE AND PROSPECTS OF THE EXPERT SYSTEMS

Nazila A. Rahimova, Vugar H. Abdullayev

Azerbaijan State Oil and Industry University, rahimova.n.a@gmail.com, abdulvugar@mail.ru

A General Survay

https://doi.org/10.7251/JIT2001030R UDC: 004.891.2

Abstract: The objects of the study are stages of development and modern state. In general terms, expert systems are knowledge-based systems. This paper focuses on the components and principles of expert systems. Expert systems are also described. The components of expert systems include knowledge base, logical impact mechanism, user interface and decision-making. In addition, this article describes the capabilities of expert systems. One challenge is to identify the future prospects of expert systems. The research examined the expert system and its significance. It also focuses on generations of expert systems. The first generation of expert systems includes systems created before 1990. This article discusses SAINT, DENDRAL and HEARSAY-1. The features of this expert systems are also discussed here. First-generation expert systems are research prototypes. As a result, the foundations of artificial intelligence were developed. Mostly first-generation expert systems were used as a passive assistant expert. The second generation of expert systems refers to systems created since 1990. Features of second-generation expert systems include dynamism, interactivity, and processing of disparate knowledge. Unlike first-generation expert systems, these systems are able to test the completeness of the knowledge base, to process fuzzy knowledge. Their main difference is the ability to integrate second-generation expert systems with existing systems. At the moment, statistical and dynamic expert systems are distinguished. This article describes the current status of both types. Here are also discussed the tools of statistical and dynamic expert systems. At the end, possible prospects of expert systems are received.

Keywords: expert systems, knowledge-based systems, perspective expert systems.

INTRODUCTION

Expert systems are complex programs that reflect the expertise of a particular subject matter and provide advice to low-skilled users. Such systems help decision makers with less experience in specific subject areas. Therefore, such systems are sometimes referred to as systems that support decision-making. Expert system is a type of application program that solves problems, gives recommendations and even makes decisions based on certain information and analytical rules set up by experts in the field (such as finance and medicine). Expert-people apply their factual knowledge, using their ability to judge when solving problems. In the expert system, these two fundamental principles are realized as two interrelated components: the knowledge base and the logical consequence mechanism. The knowledge base provides specific facts and rules related to the subject matter, while the logical result (output) mechanism executes judgments that allow expert systems to draw conclusions. Expert systems also include additional tools such as user interface and decision-making. User interfaces, like other applications, allow you to form queries, provide information, and interact with the system in other ways. One of the most interesting components of expert systems is that reasoning tools allow you to explain the results that you bring to the system. These tools allow developers to test the performance of those systems [1].

Expert systems, in other words, are called «knowledge-based systems». These systems, which include a variety of software tools, provide the description and use of actual knowledge to solve problems. For example, diagnostics of diseases, diagnostics of technologic systems, advance information on mineral deposits, production planning, and so on [2].

Therefore, it is practical to study the possibilities of expert systems. Thus, *the objects of the research* are selected stages of development, the current state of expert systems, whereas *the aim of the study* is to determine the future prospects of expert systems.

Methods of research

Expert system development stages

The first practical result of scientific works in the field of artificial intelligence since the 60s of the last century was the creation of experimental systems of experimental type in the 1970s [3]. As a result of the evolution of knowledge-based systems, including expert systems, a theoretical framework has been established, methodological and practical frameworks have been developed for the efficient collection of information, descriptions and processing problems, and programs have been developed to keep up with the possibilities of modern computer technology. Expert systems have been divided into two phases so far, and they are called first and second generation systems [4].

Expert systems of the first generation

Expert systems (ES), which were created until the late 90s of the last century, belong to the first generation. These systems are basically designed as research prototypes to explore and substantiate the theoretical foundations of artificial intelligence. The research carried out during this period was fundamental and focused on the acquisition of knowledge, the use of tools, and the establishment of a mechanism for extracting knowledge by various means [5].

The ES of this generation was intended primarily for autonomous application. As a result, methods and tools have been developed that form the basis of artificial intelligence as a scientific direction. However, all this does not mean that any universal technology, which takes into account the variety of features and their specificity, solved through computer technology. On the contrary, the experience has shown that universal strategies do not succeed because they require many restrictions to solve complex problems. Therefore, many researchers have come to the conclusion that the use of narrow frameworks with the help of expert systems can be more effective and promising [6].

Due to the origin of the structure, subject areas, inheritance of ideas, methods and tools, I generation expert systems and related tools can be divided into several families:

1. The SAINT program, designed for mathematical transformations at the Massachusetts Institute of Technology (USA), later became the basis for the creation of SIN, MATLAB and MACSYMA systems. The MACSYMA system, which implements differentiation and integration in the form of symbols, is often dominated by mathematicians and widely used by mathematicians and physicists worldwide [7].

2. The DENDRAL and META-DENTAL systems developed under the Stanford (USA) project are used to derive the probable structure of an unknown chemical compound based on nuclear magnetic resonance, mass-spectrographic and other chemical experimental data. META-DENDRAL automates the process of acquiring knowledge for the DENDRAL system. With its help, the rules for the formation of fragments of chemical structures are generated [8].

3. The HEARSAY-1 and HEARSAY-2 systems were created to recognize the conjugate human speech. The words used in the speech are taken from a known dictionary. Although the quality of speech recognition through these systems is not high (comparing this quality to a 10-year-old child). These systems have laid the groundwork for artificial intelligence in speech recognition, and the ideas and methods used in these systems are subsequently used for instrumental tools, including HERSAY-3 and AGE form the basis of instrumental systems [9].

The common features, disadvantages of the first generation expert systems, their main areas of application and their heritage can be summarized as follows:

- The first-generation expert systems were intellectual activities at the level of the user's passive assistant. They only had knowledge acquired from experts and put into the system by «knowledge engineers». The ES responds to user requests by analyzing this knowledge and imitating the logical extraction process.
- In most systems of the first generation, it was not possible to evaluate the knowledge included in the knowledge base, to detect contradictions in it, to automatically determine the correctness of the knowledge and to generate new knowledge.
- The first-generation expert systems were widely used in medical diagnostics [2]. These sys-

tems include functional medical knowledge and the expertise of professional physicians. These systems played the role of a questionnaire developed by experienced experts and were widely used in the training of medical personnel. Further development of medical expert systems was due to the use of fuzzy extracts.

- The main role of the first-generation ES in the development of artificial intelligence, including expert systems, is that theoretical methods of artificial intelligence and ES have been developed and tested in these systems.

Expert systems of the second generation

Expert systems established since the 1990s refer to the II generation. Realization of features inherent in this generation of expert systems, such as dynamism, interactivity, processing of scattered knowledge, new achievements in computer and communication technologies during this period, lead to the creation of high-performance and broadband personal computers and meframes, local and global computer networks, including Internet. To gain the dynamism of the expert system, new components were added to its architecture: subsystem modeling, communication with the external environment, and the logic of processing events.

The 2nd generation expert systems have the means to acquire new knowledge from the data entered during their creation and operation. It is possible to detect contradictions between new knowledge and data entered into the system with the knowledge stored in this ES, ie checking the completeness of the knowledge base.

In more advanced 2nd generation systems, the processing of fuzzy knowledge is being implemented. Such systems are called fuzzy expert systems. The use of fuzzy logic greatly improves the processing of fuzzy judgments and the quality of logical conclusions based on them. In practical applications, the number of fuzzy systems exceeds that of traditional systems. The main difference between real-time ES and first-generation systems is their integration.

At present, there is a tendency to create expert systems without the knowledge engineer. The mass use of personal computers and the rise of expert users' computer literacy have made it possible for ES to integrate knowledge and expertise in the system. It is expected that through expert dialogue with the participation of the knowledge engineer, as well as experimental protocols, dialogues, articles, instructions, pointers and guidelines, schemes, etc. knowledge acquisition will be the main focus of knowledge formation for the creation of modern perspective expert systems.

Research results and discussion

Current state of statistical expert systems

The current state of statistical expert systems can be summarized as follows:

- Most expert systems are integrated and open according to application types; scattered ESs are usually used by client-server architecture.
- According to the scale (type of computer) more than 82 % of expert systems are implemented on workstations and personal computers, 12 % of the characters in computers are 6 %.
- Static ES 1,2,3,3 is created for the type of problem environment; For the 1st and 2nd environment, simple ESs performed on PCs and workstations are used, and for those types of environments, ESs on workstations and meframes are used [1].

Many of the tools for static expert systems have been implemented simultaneously on different types of computers (personal, workstations, characters, and mefream), which can be attributed to different types of tools. However, the tool is usually the most commonly used type. For example, the GUBU Instrument (Cover) works on PC PCs and Micro VAX computers.

World-wide static ES instrumental tools include:

- Large instrumental tools: Aion DS 5.1 (Trinsic), KBMS (Trinzic), ART (İnference).
- Medium instrumental tools: Nexpert Object (Neuron Data), ProKappa (İntellicorp), Art-İM, ART Enterprice (İnference), Level 5 Object (İBİ), CLİPS.
- Small instrumental tools: Expert (Paperback Software), İST Class (İst Class Expert System), Personal Consultant Easy, Procedure Consultant, Crystal.
- Simvol instrumental tools: KEE (İntellicorp), ART (İnference), Gold Works (GoldenHill), Mercury (AГT).

Most of these tools can use packages (Lotus 1-2-3) integrated into databases (dBase, DB-2, Oracle, IMS, etc.). Some of them (eg East Class, KDS2&2, Super Expert, etc.) use inductive methods of knowledge acquisition. Most of the instruments are in C and Pascal. There are tools such as Instant Expert +, Intelledgent Dev eloper, Level 5 that are linked to hypertext processing tools (Hypertext, HyperCard).

Current state of dynamic expert systems

Dynamic expert systems, including real-time ESs, are developing more intensively. Currently, the sales of these systems worldwide make up 70 % of knowledge-based problem/problem-oriented systems [10]. The importance of these systems is not only in the continuous management of commercial production processes (oil transportation and refining, chemical, metallurgy, formakology, etc.), but also in aerospace research, nuclear and thermal power plants, financial operations, communications and so on. apps that are strategically important.

Recently, dynamic tools have been created to simulate the intellectual imitation used in reengineering business processes. Interest in this type of ES and instrumental tools comes from the fact that, unlike static ES and instrumental tools (IV), used to automate the current state of business, dynamic ES and IV are used to solve more important and complex issues in business process reorganization.

The first instrumental tools to create a dynamic ES was introduced in 1985 by Lisp Macjine Inc. Made in the firm. This instrumental tools Symbolics symbol called Picon was implemented on a computer. The success of this instrumental tools prompted its developers to create a private firm called Gensun in 1986. The firm developed Picon's ideas and in 1988 created the 1st version of the G2 IV (G2 1.0). Currently, versions 5.0, 6.0 and 7.0 of this system are widely used.

Perspective expert systems

The main feature of perspective systems is their dispersion, the processing and application of scattered knowledge. The basis for the establishment of perspective expert systems are methods for detecting regularities, recognition of copies, structural and logical analysis of data and knowledge, results in mathematical linguistics, as well as accumulated experience in creation of expert systems. The aforementioned factors play a part in the expert systems established today [2].

Perspective ESs should provide not only knowledge and data processing, but also meaning (semantics). These systems should be able to analyze the sentences of the natural language and build a network structure that corresponds to their semantic content. The ES should be able to understand the meaning of the information contained in the natural language and to make sentences relating to the subject matter under consideration. To this end, the problem of automatic recognition of texts and situations becomes more urgent. An important feature of this problem is that the recognition result should reflect the situation described by the user, the expert, and the decision maker. To solve this problem, a rich theoretical and practical framework in the field of artificial intelligence has been created.

The perspective expert system should build the model of the applied problem area, ie its theory. Also ES has to construct the model of the user (learner and tutor) and his/her own to optimize the process of forming the model of the investigated action (situational) in the thinking of the learner.

The main function of future expert systems is to make valid conclusions based on the discovery of the relationship between the data processing and the characteristic of the inputs that represent the known characteristics.

In addition to experts as a source of knowledge, you can use experimental protocols, articles, appointments, scientific and methodological guidelines and guidelines, schemes, etc. should be used. Thus, it is possible to automatically acquire new knowledge based on existing knowledge.

At present and in the near future, updating the concept of creation of expert systems and use of artificial intelligence is connected with the transition from local artificial intelligence systems to information processing and multidimensional intellectual systems.

Finally, we can point out that the logical output mechanism of a promising ES should be able to simulate human judgments based on similarity, to find the proximity of the studied and stored data sets with the computer's memory. This method can greatly accelerate the process of logical extraction in big databases.

CONCLUSIONS

1. This article has theoretical significance. The research analysed the characteristics of modern statistical and dynamic expert systems. Attention was also paid to expert systems of the first generation and their difference from expert systems of the second generation.
2. The research examined the stages of development of expert systems, which consist of two phases (first generation and second generation). The result of the research is an overview of the current state of expert systems in the world. The following promising directions of expert systems were also considered: automatic discovery and extraction of knowledge, methods of checking inconsistencies and completeness, processing of non-factor information, automatic forecasting of missing data in the database.

REFERENCES

[1] Beemer, B. A., Gregg, D. G. (2008). Advisory Systems to Support Decision Making. *Handbook on Decision Support Systems 1,* 511–527. doi:10.1007/978-3-540-48713-5_24

[2] Angeli, C. (2010). *Diagnostic Expert Systems: From Expert's Knowledge to Real-Time Systems.* Available at: https://www.researchgate.net/publication/238740574_Diagnostic_Expert_Systems_From_Expert's_Knowledge_to_Real-Time_Systems

[3] Petrov, N., Vasileva, S. (2014). History and Advances of the Artificial Intelligence. *Science and culture.* Available at: https://www.researchgate.net/publication/261132956_History_and_Advances_of_the_Artificial_Intelligence

[4] Tan, C. F., Wahidin, L. S., Khalil, S. N., Tamaldin, N., Hu, J., Rauterberg, G. W. M. (2016). The application of expert system: a review of research and applications. *ARPN Journal of Engineering and Applied Sciences, 11 (4),* 2448-2453.

[5] Rubin, S., Murthy, S., Ceruti, M., Malinova, M. (2000). Third-Generation Expert Systems. *2nd International ISCA Confference on Information Reuse and Integration.* Available at: https://www.researchgate.net/publication/275641302_Third-Generation_Expert_Systems

[6] Everitt, T., Hutter, M. (2018). Universal Artificial Intelligence. *Studies in Systems, Decision and Control,* 15–46. doi:10.1007/978-3-319-64816-3_2

[7] Slagle, J. (2005). *A heuristic program that solves symbolic integration problems in freshman calculus: symbolic automatic integrator (SAINT).* Available at: https://www.researchgate.net/publication/37603605_A_heuristic_program_that_solves_symbolic_integration_problems_in_freshman_calculus_symbolic_automatic_integrator_SAINT

[8] Tan, H. (2017). A brief history and technical review of the expert system research. *IOP Conference Series: Materials Science and Engineering, 242,* 012111. doi:10.1088/1757-899x/242/1/012111

[9] Ho, H. (1999). A theory of hearsay. *Oxford Journal of Legal Studies, 19 (3),* 403–420. doi:10.1093/ojls/19.3.403

[10] Lu, J.-W., Chang, N.-B., Liao, L. (2013). Environmental Informatics for Solid and Hazardous Waste Management: Advances, Challenges, and Perspectives. *Critical Reviews in Environmental Science and Technology, 43 (15),* 1557–1656. doi:10.1080/10643389.2012.671097

Submitted: March 8, 2020
Accepted: May 13, 2020.

ABOUT THE AUTHORS

Nazila Rahimova, Assos.Prof., PhD (candidate of technical science). Study: Azerbaijan Social Oil and Industry University (1975-1981) – Automation and telemechanics, electric engineer. Work experiences: 1976-1985 - Information Center of the Ministry of Trade of Azerbaijan – Operator, engineer, senior engineer, department head. From 1985 to 1991: Azerbaijan Scientific Space Research Institute – Software engineer. From 1991 until today: Azerbaijan Social Oil and Industry University (ASOIU) – Operator, engineer, senior engineer, teacher-assistant, senior-teacher, department head.

Vugar Abdullayev, PhD, was born on September 1, 1977, Sumgayit city (Azerbaijan). Married, with two children. From 1994 to 1998: Bachelor degree in Manufactory Automation, Control Engineering, Teleautomatics and Electronics on ASOA (Azerbaijan State Oil Academy). From 1998 to 2000: Master degree in Manufactory Automation and Informatics on ASOA. From 2000 to 2003: Aspirant, PhD student on Azerbaijan Cybernetics Institute. On February 7, 2007, he received PhD diploma – Candidate of Technical Sciences. On February 29, 2016, he received the title Assoc.Professor in the Speciality of Computer Sciences.

FOR CITATION

Nazila Rahimova, Vugar Abdullayev, Analysis of Stages of Development, Current State and Prospects of The Expert Systems, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 10(2020) 1:30-34, (UDC: 004.891.2), (DOI: 10.7251/JIT2001030R), Volume 10, Number 1, Banja Luka, June 2020 (1-68), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Ambient Intelligence and e-Learning

Julijana Vasiljević[1], Željko Stanković[2]

[1]PhD candidate at Pan-European University APEIRON, Banja Luka, Bosnia and Herzegovina

[2]Pan-European University APEIRON, Banja Luka, Bosnia and Herzegovina

julija2921968@gmail.com, zeljko.z.stankovic@apeiron-edu.eu

A General Survay

https://doi.org/ 10.7251/JIT2001035V UDC: 37.018.43:004, 371.3:004.738.5

Abstract: The use of ambient intelligence knowledge inevitably leads to a new education concept particularly in creating an environment towards the implementation of teaching as well as the process of education. The process of teaching and education, besides conventional and physical elements of the environment, will be enriched with elements regarding modern information technology.

Ambient intelligence will be presented in this paper as a result of the artificial algorithm neural networks, through the following contexts: e-learning environment, identification, and security.

The key role in raising students' achievements as well as competency levels belongs to modern information technology which works towards creating ambient intelligence. It is also executed through the concept of e-learning onto one of the convenient learning management platforms.

Survey results indicate that with the use of ambient intelligence, better results are achieved, especially in mathematics taught at the elementary school level. Furthermore, learned lessons are memorized by students for a long period, which is proved by higher levels of students' knowledge and skill acquisition in terms of general progress.

Keywords: ambient intelligence, e-learning, neural networks.

Introduction

The subject of the research represents analyses of how ambient factors influence the success of students who use e-learning platforms in the teaching process.

Ambient intelligence should provide students with optimal conditions for monitoring classes. It should also allow teachers to keep track of how long and successfully involved students are in a particular activity. Recommendations towards enhancing activities can be electronically transmitted to convenient platforms or even to students' wristwatches [1].

The significance of ambient intelligence has been recognized as well as time applied by leading world-class companies such as Siemens, which has invested heavily in smart development buildings and autonomies in the production process in factory halls [2]. Nokia company has also invested significant resources in communication development which do not necessarily relate to smart homes only but have a much wider application for reasons of developed mobile applications [3].

Ambient intelligence [4, 5] is being developed within multidisciplinary fields thus allowing the benefit of research to be used for a variety of purposes. This paper deals with ambient intelligence in the function of enhancing e-learning.

The basic idea behind the concept of ambient intelligence is the adaptation of the environment with the help of information and communication technology for e-learning towards students' needs to achieve better results as well as greater achievement.

E-learning system should be built so that it adjusts teaching content to students' needs based on information in real-time.

Contemporary teaching, especially electronically shaped educational learning processes, and teaching, is getting closer and closer to the concept of stu-

dents' creation within the pedagogical framework activities. This increasingly signifies that the concept of constructionism is expanding. Regardless of the importance of constructing, rather than imposing knowledge, elements are still needed likewise instructive and constructivist approach in all teaching situations-from the traditional classroom-to-classroom teaching system, to the virtual classroom as well as e-learning [6,7,8,9].

Learning with modern strategies that use the concept of e-learning in comparison with traditional teaching is characterized by completely different relationships between students and teachers likewise modified forms of the process of teaching organization [10].

A student qualified for self-education represents one of the key goals in the process of education. Those methods of work that contribute to a more active attitude of students towards teaching content, should be given preference. Training students to use different sources of knowledge as well as becoming independent in the learning process is also very significant [11].

Previous research has not completely provided comprehensive answers to questions regarding the impact of the environment on e-learning, identification, and students' security within e-space. To properly perceive the impact of the environment on e-learning likewise the predictions of students' success at work, an analysis of the factors relevant to the success of the teaching process using neural network algorithms have been performed.

METHOD

While exploring the impact of ambient intelligence on the e-learning process, this paper has defined the largest factor, most relevant to students' achievements in the context of an e-learning environment, with the use of analytical hierarchical methods.

Analyses of the e-learning environment

Defined factors are further processed as input variables within an artificial neural network. Results obtained have been used to develop ambient intelligence in the e-learning enhancement function. In order to define the factors of greatest importance towards students' achievements, the authors of this paper have used extended AHP fusion method of triangular numbers that are performed in four steps [12] as follows:

Let X = {X_1X_1, X_2X_2, X_3X_3, ..., X_nX_n} be a set of objects, and G = {G_1G_1, G_2G_2, G_3G_3 ....., G_mG_m} be set of goals.

Each object is analyzed for each objective, respectively. Accordingly, *m* values for each object can be expressed as follows:

$M_{gi}^1M_{gi}^1$, $M_{gi}^2M_{gi}^2$, ..., $M_{mgi}^mM_{mgi}^m$, i = 1, 2, ..., n.

Where all the $M_{gi}^jM_{gi}^j$, (j = 1, 2, ..., m) represent triangular fuzzy numbers.

The AHP fusion steps are:

STEP 1:

The fusion value of the synthetic respective domains with the respect to the *i-th* object is defined as:

$$S_i = \sum_{j=1}^m M_{gi}^j * \left[\sum_{i=1}^n \sum_{j=1}^m M_{gi}^j\right]^{-1} (S_i = \sum_{j=1}^m M_{gi}^j * \left[\sum_{i=1}^n \sum_{j=1}^m M_{gi}^j\right]^{-1} (1)$$

To obtain $\sum_{j=1}^m M_{gi}^j$,$\sum_{j=1}^m M_{gi}^j$, we perform the fusion operation by adding *m* values for a certain matrix according to the following statement ,

$$\sum_{j=1}^m M_{gi}^j = \left(\sum_{j=1}^m l_j, \sum_{j=1}^m m_j, \sum_{j=1}^m u_j\right) \sum_{j=1}^m M_{gi}^j = \left(\sum_{j=1}^m l_j, \sum_{j=1}^m m_j, \sum_{j=1}^m u_j\right) (2)$$

To obtain values,

$$\left[\sum_{i=1}^n \sum_{j=1}^m M_{gi}^j\right]^{-1} \left[\sum_{i=1}^n \sum_{j=1}^m M_{gi}^j\right]^{-1} (3)$$

perform the fusion addition operation,

$$M_{gi}^j (j = 1,2, \dots, m)\, M_{gi}^j (j = 1,2, \dots, m) (4)$$

thus obtaining values,

$\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^{j} = \left(\sum_{j=1}^{n} l_i, \sum_{j=1}^{n} m_i, \sum_{j=1}^{n} u_i\right) \sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^{j} = \left(\sum_{j=1}^{n} l_i, \sum_{j=1}^{n} m_i, \sum_{j=1}^{n} u_i\right)$ (5)

The inverse vector for a given statement is calculated as follows,

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^{j}\right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} u_i}, \frac{1}{\sum_{i=1}^{n} m_i}, \frac{1}{\sum_{i=1}^{n} l_i}\right)\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^{j}\right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} u_i}, \frac{1}{\sum_{i=1}^{n} m_i}, \frac{1}{\sum_{i=1}^{n} l_i}\right) \quad (6)$$

STEP 2:

If M_1M_1 = (l_1l_1, m_1m_1, u_1u_1) and M_2M_2 = (l_2l_2, m_2m_2, u_2u_2) stand for two fuzzy triangular numbers, the degree possibilities M_2M_2 = (l_2l_2, m_2m_2, u_2u_2) ≥ M_1 M_1 = (l_1l_1, m_1m_1, u_1u_1) is defined as:

$$V(M_2 \geq M_1) = \sup_{y \geq x}[min(\mu_{M_1}(x),)\mu_{M_2}(y),] \; V(M_2 \geq M_1) = \sup_{y \geq x}[min(\mu_{M_1}(x),)\mu_{M_2}(y),] \quad (7)$$

An equivalent statement for (7) is given by the following,

$$V(M_2 \geq M_1) = hgt(M_1 \cap M_2) = \mu_{M_2} = 1 \; if m_2 \geq m_1, 0 \; if \; l_1 \geq u_2, otherwise \; \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}$$

$V(M_2 \geq M_1) = hgt(M_1 \cap M_2) = \mu_{M_2} = 1 \; if m_2 \geq m_1, 0 \; if \; l_1 \geq u_2, otherwise \; \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}$

$$V(S_b \geq S_a) = 1 \; if m_b \geq m_a, 0 \; if \; l_a \geq u_b, otherwise \; \frac{l_a - u_b}{(m_b - u_b) - (m_a - l_a)}$$

$V(S_b \geq S_a) = 1 \; if m_b \geq m_a, 0 \; if \; l_a \geq u_b, otherwise \; \frac{l_a - u_b}{(m_b - u_b) - (m_a - l_a)}$ (8)

STEP 3:

Possibility degree of a convex fusion number to be greater than k , a convex fusion number can be defined as follows:

$$VV(M \geq M_1, M_2, \ldots M_k) = V\left[(M \geq M_1) and \; (M \geq M_2) and \ldots and \; (M \geq M_k)\right] = min(M \geq M_i), i = 1,2,3, \ldots, k \quad (9)$$

$VV(M \geq M_1, M_2, \ldots M_k) = V\left[(M \geq M_1) and \; (M \geq M_2) and \ldots and \; (M \geq M_k)\right] = min(M \geq M_i), i = 1,2,3, \ldots, k$ (9)

Supposedly,

$d(A_i) = minV(S_i \geq S_k) d(A_i) = minV(S_i \geq S_k)$, *for the k*= 1,2, …, $n: k \neq i$,= 1,2, …, $n: k \neq i$,

the weight vector is given as follows,

$$W' = \left(d'(A_1), d'(A_2), \ldots, d'(A_n)\right)^T \quad (10) \; W' = \left(d'(A_1), d'(A_2), \ldots, d'(A_n)\right)^T \quad (10)$$

Where $A_i (i = 1,2, \ldots, n) A_i (i = 1,2, \ldots, n)$ represents the number of *n* elements.

STEP 4:

In normalization, the normalized weight vectors would be:

$$W = \left(d(A_1), d(A_2), \ldots, d(A_n)\right)^T (11) W = \left(d(A_1), d(A_2), \ldots, d(A_n)\right)^T (11)$$

where *W* does not represent a fuzzy number.

This paper considers attributes as criteria for analyzing the influence of factors on the development of ambient intelligence towards enhancing e-learning as shown in Table 1.

Table 1.

CRITERIA	
C1	Audio-video communication possibility
C2	Possibility for continuous teachers' guidance through teaching content
C3	Environment likewise e-learning system provide conditions for dynamic realization of teaching contents
C4	Multimedia presentation of teaching contents
C5	Teaching content design

Table 2 displays, as an alternative, the teaching environment.

Table 2.

ALTERNATIVES	
A1	Traditional teaching
A2	Hybrid teaching (Traditional and E-learning)
A3	Guided e-learning supported by simulations
A4	Pure e-teaching
A5	Traditional teaching with the use of multimedia content

Upon the implementation of the fusion AHP method as Chang's analytical method [14], we come to results as shown in Table 3:

Table 3.

CRITERION	WEIGHTED VALUES	ALTERNATIVES				
		A1	A2	A3	A4	A5
C1	**0.14368**	0.237334	0.202872	0.225072	0	0
C2	**0**	0.109059	0.151158	0.130807	0.021366	0.11039
C3	**0.26754**	0.34287	0.303967	0.319555	0.38745	0.36427
C4	**0.22053**	0.154462	0.201578	0.148429	0.38745	0.36427
C5	**0.36825**	0.156274	0.140425	0.176136	0.203734	0.16105
WEIGHTS OBTAINED		**0.217443**	**0.206638**	**0.215428**	**0.264128**	**0.23710**

Table 3 allows us to come to the conclusion that criterion C3 "Environment and e-learning system provides conditions for the dynamic realization of teaching content" is weighted by the highest value while criterion C2 "Ongoing teachers' guidance through teaching content" is weighted by the lowest value. In order to determine the value of alternatives, a convergence consensus model has been applied in this paperwork.

The consensus convergence model [15] has been developed for the sake of decision making on the effectiveness of ambient factors regarding students' achievements within the e-environment. This model is based on determining the differences in the "weights" of decision-makers on the basis of values assigned by each decision-maker to relevant elements (criteria, sub-criteria, and/or alternatives) [16]. Table 4 displays the calculated values of alternatives.

Table 4.

ALTERNATIVES	CONSENSUS WEIGHT VECTORS	RANK
A1	0,060	5
A2	0,069	2
A3	0,077	1
A4	0,061	4
A5	0,062	3

Based on the calculation, we can conclude that the alternative A3 "Guided e-simulation-supported teaching" is ranked first regarding the importance of the effective development of ambient intelligence in order to enhance e-learning.

Identification and security context analyses

Identification context implies observational procedures relating to the notification of students' reactions to the created teaching content likewise the teaching process.

Security context implies defining procedures that minimize the possibility of teaching content as well as users' personal information misuse.

Stated contexts in this paper have been analyzed by the same method as the context of the environment for e-learning with the maximum assurance of the prescribed quality of e-learning standards [17].

An alternative to the emerging value identification context is "Guided e-simulation-supported teaching". An alternative with the highest possible value for context security is "Traditional teaching using multimedia".

Criteria considered for all three contexts are: C1-Possibility of audio-video communication, C2-Teachers' ongoing guidance through the teaching content process, C3-Both e-learning environment, and e-learning system provides conditions for the dynamic realization of the teaching content, C4-Multimedia syllabus presentation, C5-Teaching content design.

Comparative analyses have found that "Guided e-simulation-supported teaching" represents the optimal environment for e-learning. Considered criteria are the basics for the development of ambient intelligence.

Required and sufficient conditions for the C1 criterion "Audio-video communication possibility" can be the existence of compatible audio-video equipment for teachers as well as for students.

Required and sufficient conditions for criterion C2 " Possibility for continuous teachers' guidance throughout teaching content" can be: installed interactive platform which is able to track students' work, and of course, when needed, informs the teacher regarding the difficulty in mastering the teaching content.

Required and sufficient conditions for C3 criterion " Both e-learning environment and e-learning system provide conditions for the dynamic realization of the teaching content " can be the existence of e-learning platform as well as appropriate microclimate conditions towards the realization of dynamic content and sensors installed for monitoring basic parameters of general conditions like students' pulse, temperature, blood pressure, etc.

Required and sufficient conditions for the C4 criterion "Multimedia syllabus presentation" can be: the existence of e-platform which allows you to create, set up, or use the multimedia presentation of teaching content.

Required and sufficient conditions for C5 criterion "Teaching content design" can be the existence of an e-platform which enables the design of teaching content according to the topics as well as the timing of a teaching process.

Ambient intelligence shall be developed throughout these contexts with the use of a neural network whose final result will be expressed through the fusion of neural network results, observed for each criterion individually. Basic connections within the system for the development of ambient intelligence are presented in Figure 1.

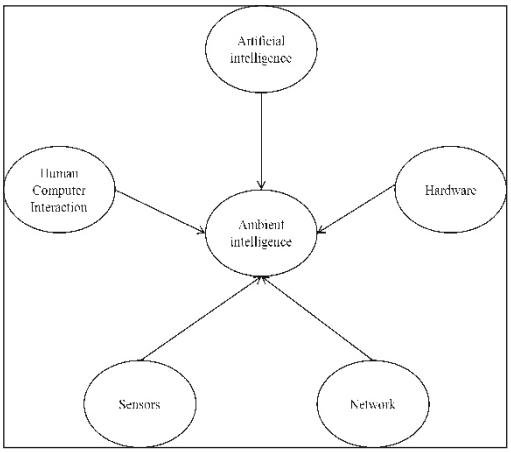


Figure 1: *Connections in the ambient intelligence development system*

This paper will explain the neural network model used for criterion C1 "Audio-video communication possibility".

Video equipment should allow the monitoring of students' facial expressions. Based on these expressions, the system will propose corrective actions within the ambient environment. For this purpose, Convolution neural networks models will be used, also known as non-cognitive models [18]. For the sake of analyses, these models will use data deployed in advanced databases [19].

Convolution neural networks

This term "Convolution neural networks" is in correlation with convolution, respectively operator, used in image and signal processing [20]. We use convolution filters in the field of artificial intelligence and neural networks to sharpen or blur images likewise to detect edges in contrasting terms.

Convolution neural networks are most frequently used when data represent images that are pixel matrix and they are presented by their width, height, and pixel values. For color images, each of the three "RGB" channels is usually represented by value pixels within the range 0-255.

In convolution neural networks, the convolution filter represents a generalized linear model for the image region to which it is applied [21]. Apart from the filter name, the context of convolution neural networks, the name convolution kernel is also used. The filter is represented by a two-dimensional matrix of small dimensions, compared to the image to which it is applied and is consisted of real values.

An important concept of convolution neural networks is displayed in maximal compression as a form of nonlinear pixel reduction [22]. Depending on a compression type, a pixel with a particular value in a particular region is selected. As for reduction operation, the pixel with the highest value is chosen for the maximum. The compression layer is used to reduce progressively the size of the image and therefore the number of features, thus leading to a decrease in the complexity of the calculation. An example of maximum compression is shown in Figure 2.

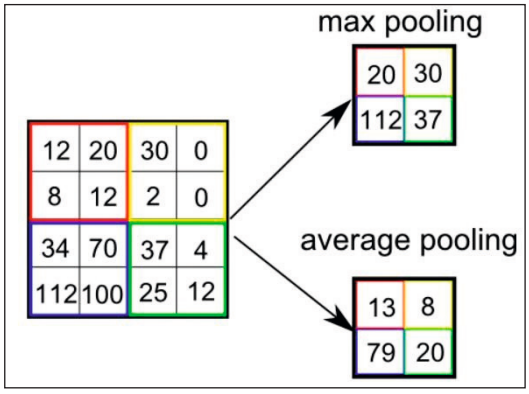


Figure 2: *Example of maximum compression*

Description of the general architecture relating to convolution neural networks.

The difference of convolution neural networks about other types of architecture neural networks reflects in the existence of a layer where convolution occurs. Convolution layer is essential for the functioning of a convolution neural network. It also performs various demanding calculations [23,24,25].

The essence of using convolution neural networks is to enable the response in real-time regarding activities that take place in distance learning systems [26,27,28,29].

Figure 3 shows the architecture of convolution neural networks.

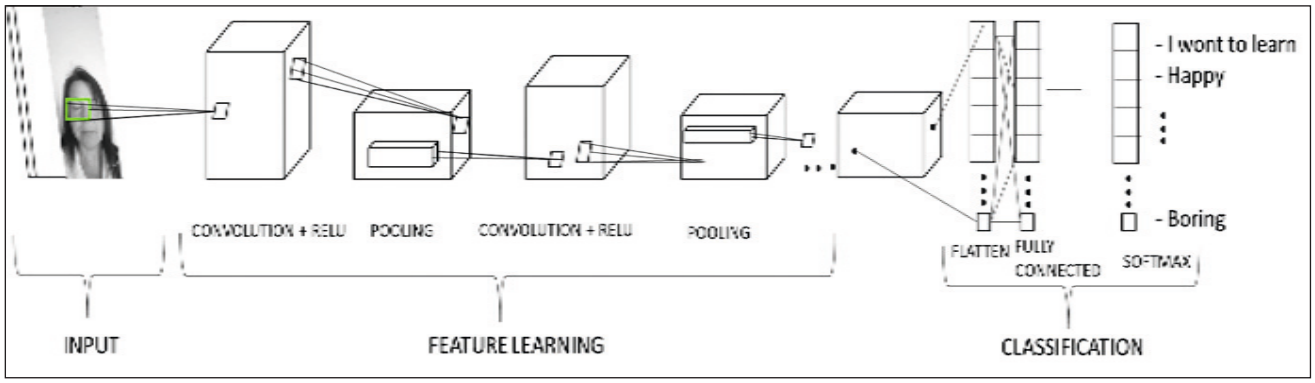


Figure 3: *The architecture of convolution neural networks*

The initial author used to construct and recognize facial expressions with the use of his photos as input data towards processing within convolution neural networks. The initial author of the paper has also tried to present different states of facial expressions as shown in Figure 4.

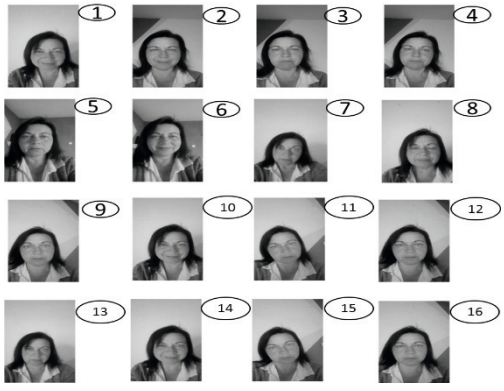


Figure 4: *Different facial expressions*

Position 8 in figure 4 shows the state in which the person expresses disinterest in teaching activities, manifested by eyes closed as well as lips folded. The lack of interest in teaching activities can be a consequence either of fatigue or poor ambient conditions.

To compare the results of artificial neural networks, comparative analyses of the obtained values for the criterion C1 has been conducted by the author. The analyses are the result of image processing via convolution neural networks likewise through artificial "backpropagation" neural networks of which activation function is the sigmoid function. For the research, Neuroph studio has been used in which neural network has been created as shown in Figure 5.

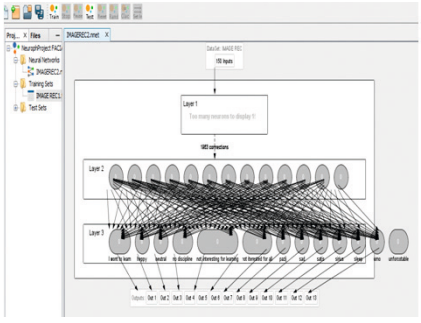


Figure 5: *A neural network created in a Neuroph studio*

After 10000 interactions, an error value of 0.01 has been reached while in Figure 6 a graph of the total network error after 136 interactions is shown.

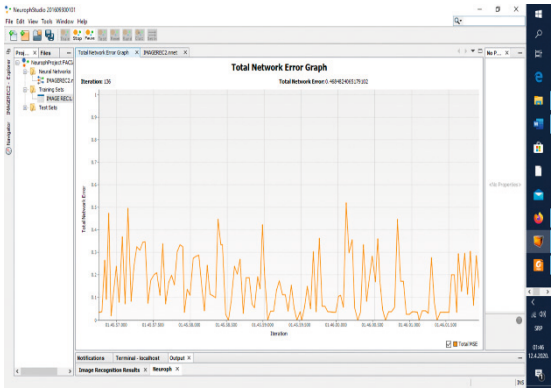


Figure 6: *Total network error graph after 136 interactions*

RESULT

During the testing of the convolution neural network, facial expression detection has been performed, relating to disinterest in teaching in 97,7 percent of cases based on 16 different facial expression images.

By testing artificial backpropagation of a neural network whose activation function is sigmoid recognition of facial expression, detection of facial expression relating to disinterest in teaching in 87,7 percent of cases has been conducted, based on 16 different photos of facial expressions.

Further development of the e-learning platform should provide the teaching staff with information on which corrective measures should be taken to create ambient conditions to achieve better results in mastering the curriculum.

The initial author, Maths professor, has applied corrective methods within the organization of e-teaching for the sake of mathematics course, which implied a higher level of achievement in ambient students' intelligence in comparison with students taught traditionally.

DISCUSSION

Ambient intelligence evolves through contexts: e-learning environment, identification, and security.

This paper displays the influence of the e-learning environment through the criterion C1 "Audio-video communication possibility" where sufficient as well as required conditions are listed as the existence of installed compatible audio-video equipment for students likewise for teachers.

Survey results indicate that with the use of ambient intelligence better results are achieved, especially in elementary schools mathematics.

CONCLUSION

Contemporary trends, as well as lifestyles, are increasingly suggesting that e-learning will be applied on a much larger scale in comparison to the traditional way of schooling. Suppose that we see learning as a type of business or production, the final product represents knowledge, we have to be objective and conclude that in case the final result is good, the production system will be simpler and cheaper. Therefore, there is no reason not to accept it.

Research results indicate that the convolution neural network is more suitable for facial expression recognition in comparison with artificial backpropagation neural networks.

Modern information technology in the creation of ambient intelligence has a key role in raising students' achievements as well as competency levels. It is also being executed through the concept of e-learning onto one of the convenient learning management platforms.

Research results also indicate that with the use of ambient intelligence, better results are achieved, especially in elementary school mathematics education.

REFERENCES

[1] Augusto, J. C., & McCullagh, P. 2007. Ambient intelligence: Concepts and applications. *Computer Science and Information Systems*, *4*(1), (pp.1-27).

[2] Siemens, 2006. http://networks.siemens.de/smarthome/en/index.htm .

[3] Nokia, 2006. research.nokia.com/research/projects/sensorplanet/.

[4] ST Advisory Group., 2010. The european union report, scenarios for ambient intelligence in, 2001. ftp.cordis.lu/pub/ist/docs/istagscenarios.

[5] C. Augusto and D. Cook., 2007. Ambient Intelligence: applications in society and opportunities for AI. Lecture Notes, tutorial given during 20th International Joint Conference on Artificial Intelligence – IJCAI'07, Hyderabad, India.

[6] Бјекић Д., Крнета, Р., Милошевић, Д., 2008. Компетенције за е-наставу у систему професионалних компетенција наставника основне школе, Иновације у настави, (pp.7-20).

[7] Webster D., 2006. Knowledge Presenter: Learning about e-Learning, Kookaburra Studios, Pty Ltd. http://www.knowledgepresenter.com/assets/freeguide.htm

[8] Nakajima K., 2006. Is „e-Teaching" Web Zero or Potentially Web 2.1?, http://www.cccties.org/access/toukou/nakajima_20061117_2.pdf

[9] Williams, M., 2005. E-Learning. History and prospects, http://apps.hct.ac.ae/ewb2005/2005/ewb_newsletter_04_10.pdf

[10] Lin, H. S., Chiu, H.L. 2000. Using Computers to Support a Beginning Teacher's Professional Development, Journal of Science Education and Technology, 4, Springer Stable, (pp. 367-373).

[11] Sutherland, R., 2004. Designs for learning: ICT and knowledge in the classroom, Computers & Education, 43, (pp. 5–16).

[12] Saaty, T., 2008. Decision making with the analytic hierarchy process. International Journal of Services Science, (pp.83–98).

[13] Chang, D.Y., 1996. Applications of the extent analysis method on fuzzy AHP. European Journal of Operational Research, (pp 649–655).

[14] Chang, D.Y. 1996. Applications of the extent analysis method on fuzzy AHP. European Journal of Operational Research, (pp. 649–655).

[15] Srdjevic, B. & Medeiros, Y. 2008. Fuzzy AHP assessment of water management plans. Water Resources Management (pp. 877-894).
[16] Triantaphyllou, E. 2000. Multi-criteria decision making methods. In Multi-criteria decision making methods: A comparative study (pp. 5-21). Springer, Boston, MA.
[17] Tomić S., Drljača D., DDLM - Quality Standard for Electronic Education Programs in Higher Education of Bosnia and Herzegovina, JITA – Journal of Information Technology and Applications Banja Luka, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 9(2019) 2:67-79, (UDC: 004.735.8:621.39(497.6), (DOI: 10.7251/JIT1902067T), Volume 9, Number 2, Banja Luka, december 2019 (49-128), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004
[18] Padgett, C., & Cottrell, G. W. 1998. A simple neural network models categorical perception of facial expressions. Proceedings of the 20th Annual Conference of the Cognitive Science Society (pp.806–807). Mahwah,NJ: Erlbaum.
[19] Nedeljko Šikanjić, Zoran Ž. Avramović, Esad F. Jakupović, Implementation of the Neural Network Algorithm in Advanced Databases, JITA – Journal of Information Technology and Applications, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 8(2018) 2:54-63, (UDC: 004.738.5:551.588:551.506), (DOI: 10.7251/JIT1802054S), Volume 8, Number 2, Banja Luka, december 2018 (45-96), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004
[20] K. He, X. Zhang, S. Ren, J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition", Computer Vision and Pattern Recognition, 2014, https://arxiv.org/abs/1406.4729/.
[21] Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. 2020. Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104, 106189.
[22] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, "Recent Advances in Convolutional Neural Networks", Computer Vision and Pattern Recognition, 2015, https://arxiv.org/abs/1512.07108/.
[23] C. Gulcehre, "Convolutional Neural Networks (LeNet)", http://deeplearning.net/tutorial/lenet.html/, Feb, 2017.
[24] J. Johnson, A. Karpathy, "Convolutional Neural Networks", Stanford Computer Science, http://cs231n.github.io/convolutional-networks/, Feb, 2017.
[25] Wang, Y., Li, Y., Song, Y., & Rong, X. 2020. The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition. Applied Sciences, 10(5), 1897.
[26] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", MIT Press, 2016, https://www.deeplearningbook.org/.
[27] Saito, T., & Watanobe, Y. 2020. Learning Path Recommendation System for Programming Education Based on Neural Networks. International Journal of Distance Education Technologies (IJDET), 18(1), 36-64.
[28] Syed, M. R. 2020. Methods and applications for advancing distance education technologies: International issues and solutions. Simulation, 348.
[29] Agrawal, A., & Mittal, N. 2020. Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. The Visual Computer, 36(2), 405-412.
[30] An, F., & Liu, Z. 2020. Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM. The Visual Computer, 36(3), 483-498.

Submitted: February 2, 2020
Accepted: June 1, 2020.

About the authors

Julijana Vasiljević was born on February 29, 1968, in Bitola, Republic of North Macedonia. She graduated from the Faculty of Natural Sciences and Mathematics the Ss. Cyril and Methodius University in Skopje in 1991. At the Faculty of Technical Sciences in Čačak she completed her Master studies in 2012. She is currently a Ph.D. candidate at Pan-European University "Apeiron", Banja Luka, Bosnia and Herzegovina, Europe.

Zeljko Stankovic was born in Belgrade. He finished primary and secondary school in his hometown. He received his higher education in Cleveland, Ohio, USA, where he graduated in 1981. In 2006, he defended his master's thesis at the University of Novi Sad. He defended his doctoral dissertation at Singidunum University in 2010. He has been programming since 1984. making programs for his first computer Commodore 64. Robotics and bioengineering are a long-standing field of work and interest.

For citation

Julijana Vasiljević, Zeljko Stankovic, Ambient Intelligence and e-Learning, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 10(2020) 1:35-43, (UDC: 37.018.43:004, 371.3:004.738.5), (DOI: 10.7251/JIT2001035V), Volume 10, Number 1, Banja Luka, June 2020 (1-68), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

INTELLIGENT DISTANCE LEARNING SYSTEMS

Dragan Vasiljević[1], Branko Latinović[2]

[1]PhD candidate at Pan-European University APEIRON, Banja Luka, Bosnia and Herzegovina

[2]Pan-European University APEIRON, Banja Luka, Bosnia and Herzegovina

vasiljevicdj68@gmail.com, branko.b.latinovic@apeiron-edu.eu

A General Survay

https://doi.org/10.7251/JIT2001044V UDC: 004.8.032.26

Abstract: Models used for creating intelligent systems based on artificial non-chromic networks indicate to the teachers which educational as well as teaching activities should be corrected. Activities that require to be corrected are performed at established distance learning systems and thus can be: lectures, assignments, tests, grading, competitions, directed leisure activities, and case studies. Results regarding data processing in artificial neural networks specifically indicate a specific activity that needs to be maintained, promoted, or changed in order to improve students' abilities and achievements. The developed models are also very useful to students who can understand their achievements much better as well as to develop their skills for future competencies. These models indicate that students' abilities are far more developed in those who use some of the mentioned distance learning systems in comparison with the students who learn due to the traditional classes system.

Keywords: neural networks, distance learning system, achievements, competencies.

INTRODUCTION

Predicting future events is one of the basic areas within artificial application neural networks [1].

Hereby, these future events are necessary abilities as well as students' achievements. Unlike classic methods, based on models, artificial neural networks belong to the class of self-adaptive methods based on data with only a few model assumptions for problems being studied [2]. The research problem is in correlation with activities in the teaching process that can be promoted. Artificial neural networks throughout data testing and training should provide a projection of students' abilities as well as competence [3]. The subject of research in this paper relates to the development and application of algorithms towards the prediction of neural networks use in distance learning systems.

Most research work concerning to the use of artificial neural networks in distance learning systems is one-dimensional just because it focuses on methods and techniques applied to either the electronic learning system only or especially to students and teachers [4-6] while fewer papers and research deal with the problem of learning based on past events [7-8].

So far, this area has not been sufficiently explored, especially the field of e-learning as indicated by the great number of scientific work published in 2019 which refers to the robotics, economy as well as natural phenomena [9-13]. To create intelligent distance learning systems, this paper will display the method of analyzing data using artificial neural networks.

METHOD

Artificial neural networks are an integral part of artificial intelligence primarily used for numerical prediction [14,15], classification [16,17] and pattern recognition [18,19, 20].

The data should be divided into three samples: the one for coaching, cross-validation, and testing [20]. Upon model defined, the input is prepared, the algorithm selected, learning rule, and required functions established, the network should be taught or trained on the basis of prepared data in order to identify the connection among data likewise to be capable of predicting values based on input values. The learning phase is a process of adjusting network weights that take place in multiple iterations

or passing through the network. In the cross-validation phase, the network tends to observe the length training, the number of hidden neurons as well as parameters. Network testing is the third phase of neural network operation and it is crucial for network evaluation. The difference relating to the learning and the testing phase is that in this phase the network is no longer in the learning process. Network evaluation is done by calculating the error in a way to compare the network output with the actual outputs [22].

Result

Lectures, assignments, tests, grading, competitions, directed leisure activities are input variables important for in-depth data analyses as well as model creation neural networks [23]. The output variable is a satisfactory level of achievement. Defined variables data refer to 102 elementary school students (eighth graders) analyzed during the school year 2018-2019. The data has been updated into the Neural designer studio database for the purpose of this paper due to further processing and creating an artificial neural network. The statistics regarding defined input variables are important information for designing models since they can alert to the presence of false data. Table 1 displays the minimum, maximum, middle, and standard deviation of all variables in the data set.

Table 1.

	Minimum	Maximum	Mean	Deviation
Lectures	1.00	5.00	4.04	1.09
assignments	1.00	5.00	3.93	1.23
tests	1.00	5.00	4.06	1.19
grading	1.00	5.00	2.96	0.61
competitions	1.00	5.00	2.60	0.94
directed leisure activities	1.00	5.00	2.25	0.88
level of achievement - satisfactory	0.00	1.00	0.51	0.50

Table 2 displays the value of the correlations of all input variables. The minimum correlation is -0.440712 between the lecture and competition variables. The maximum correlation is 0.9555376 between tasks and tests.

Table 2.

	Lectures	assignments	tests	grading	competitions	directed leisure activities
Lectures	1	0.83	0.75	0.24	-0.44	-0.29
assignments		1	0.96	0.56	-0.19	-0.048
tests			1	0.6	-0.0052	0.19
grading				1	0.65	0.33
competitions					1	0.67
directed leisure activities						1

An artificial neural network is defined as a single layer with forwarding propagation while the activation function is a Hyperbolic Tangent function which can be mathematically expressed as follows:

$$tghX = \left(\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}\right) tghX = \left(\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}\right) \quad (1)$$

Where X represents an independent variable size. Graphic representation of the Hyperbolic Tangent of the function is given in Figure 1.

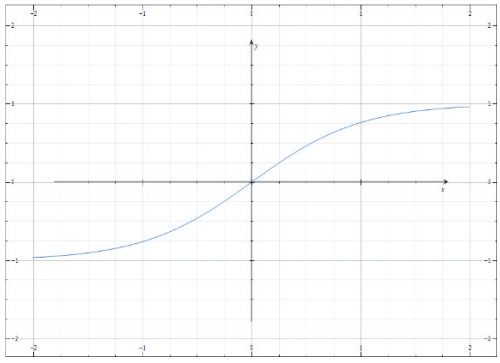


Figure 1. *Graphic representation of Hyperbolic Tangent function*

The total number of variables for a neural network is seven, of which one variable is a target variable referring to a satisfactory level of achievement. The training strategy over the data set is defined by the optimization algorithm using the Quasi-newton method. The optimization process is approximately the same as the ordinary Newton method with Hessian matrix step modification.

The following is a brief numerical example of one type of Quasi-newton method that uses the original Hessian inverse matrix for each iteration.

Target function:

$$\min f(x) = 2x_1^2 + 3x_2^2 \min f(x) = 2x_1^2 + 3x_2^2 \quad (2)$$

Starting point selection:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (3)$$

Hessian matrix inverse calculation :

$$\nabla\nabla f(x) = \begin{bmatrix} 2x_1 \\ 3x_2 \end{bmatrix}\begin{bmatrix} 2x_1 \\ 3x_2 \end{bmatrix} \quad (4)$$

$$\nabla^2\,\nabla^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \quad (5)$$

$$H^{-1} = \begin{bmatrix} 0{,}5 & 0 \\ 0 & 0{,}3 \end{bmatrix} H^{-1} = \begin{bmatrix} 0{,}5 & 0 \\ 0 & 0{,}3 \end{bmatrix} \quad (6)$$

Finding a new value for variable x:

$$x^{k+1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0{,}5 & 0 \\ 0 & 0{,}3 \end{bmatrix} x^{k+1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0{,}5 & 0 \\ 0 & 0{,}3 \end{bmatrix} \quad (7)$$

Specifying a new value for variable x :

$$x^{k+1} x^{k+1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (8)$$

Determining the value in case the function converges:

$$\nabla\nabla f(x) = 0 \quad (9)$$

The training of the data set has been performed within a thousand interactions during one hour of testing time. Upon the performed training, Figure 2 displays the training as well as selection error in each iteration. Blue line represents a training error and the orange one represents a selection error. The home value of the training error is 2.33562 and the final value after 135 epochs is 9.72136e-6. The initial value of the selection error is 3.444445 and the final value after 135 epochs is 0,382761.

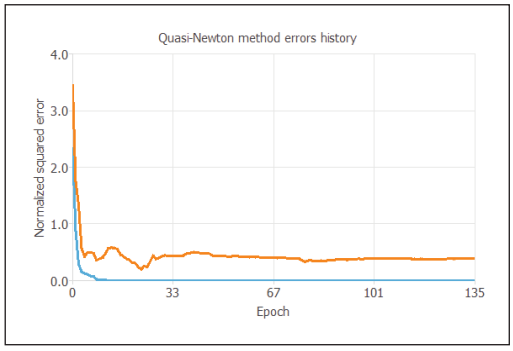


Figure 2. *Graphic representation of error history using the Quasi-newton method*

Figure 3 provides a graphic representation of the resulting deep artificial architecture neural networks. It contains a scaling layer, a neural network as well as a non-scaling layer. The yellow circles represent the scaled neurons, the blue ones represent perceptron neurons and the red ones represent neurons. The number of inputs is 6 and the number of outputs is 1. The number of hidden neurons is 1.

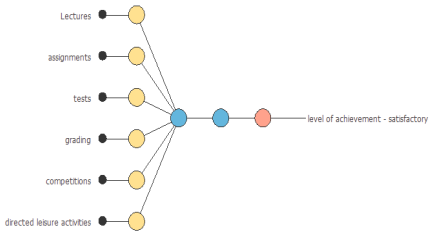


Figure 3 *- Graphic representation of artificial neural network architecture*

The testing has been performed on the data relating to the lecture variable with dedicated value 1. Table 3 displays the value for the target variable with value 0.0911 whose correlation with original values has been proved.

Table 3.

	Value
Lectures	1
assignments	3.93137
tests	4.05882
grading	2.96078
competitions	2.59804
directed leisure activities	2.2451
level of achievement - satisfactory	0.0911282828

Figure 4 displays a graph illustrating the dependence of the target variable with input variables.

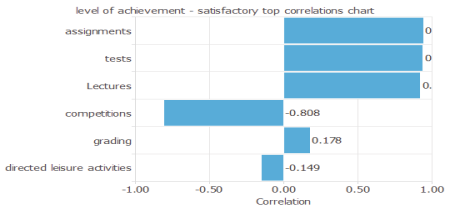


Figure 4- *graph view illustrating the dependency of the target variable with the input variables*

In order to test the model, we have assigned random variables to the input variables by assigning the value 1 to three variables. The value of the target vari-

able is 0.459, which is, closer to 0 than 1, which proves that the model of artificial neural networks provides expected results. Table 4 displays the allocated input values as well as the target variable value.

Table 4.

	Value
Lectures	1
assignments	1
tests	3
grading	2
competitions	1
directed leisure activities	4
level of achievement - satisfactory	0.459540169

Within creating intelligent distance learning systems, based on the results obtained by the data processed within an artificial neural network of the observed input variables: lectures, assignments, tests, grading, competitions as well as directed free activities represent necessary facilities in achieving a satisfactory level of accomplishment.

Discussion

Defined input variables and the data processing results, using artificial neural networks, are the basis of creating a distance learning system that needs to provide users with unhindered access to all available resources on the portal distance learning systems [24].

Content organization on the distance learning platform should provide conditions for that clear, transparent, and logical organization of teaching content through lectures, which in form and content, should be tailored to the target audience. Upon the lectures execution on the platform, tasks related to the completed After-lectures are assigned. Further follow-up of classes is enabled only after tasks have successfully been solved.

Tests represent a separate unit in the distance learning system, periodically organized. Due to their form, they can be classified into self-evaluative and formal ones. Assessment is a constant activity whereby the lecturer monitors all the work as well as students' engagement.

Introducing contest-related content likewise directed leisure activities are innovations within distance learning systems that provide trainees with competency development likewise greater learning motivation.

Due to the analyses including 102 elementary school pupils (eighth graders) who have used the distance learning system as an adjunct to traditional teaching, within the school year of 2018-2019, as well as 85 elementary school pupils within the school year 2017-2018, taught in a traditional way, the greater success of the pupils using the distance learning system is accomplished and evident.

Conclusion

The research presented in this paper indicates that by application of the neural network in creating intelligent distance learning systems, the achievement and competences of primary school pupils can be significantly improved. This paper displays a realized model of artificial neural networks in the function of the development towards organizing teaching content likewise activities on a distance learning platform.

A comparative analysis regarding two-generation eighth-graders at their first year of high school education indicates the implementation efficiency of the above-mentioned contents of distance learning.

The developed distance learning model has been applied at the elementary school level. However, it can be successfully applied both at a high school or at an academic level.

Further development directions for distance learning relate to the application, besides the methods of neural networks likewise other methods of in-depth data analyses.

The aforementioned research results should form the basis for further development of the distance learning systems.

References

[1] Sharda, R., 1994. Neural networks for the MS/OR analyst: An application bibliography. Interfaces 24 (2), 116–130.
[2] White, H., 1989. Learning in artificial neural networks: A statistical perspective. Neural Computation 1, 425–464
[3] Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.
[4] McClure, P., & Kriegeskorte, N., 2016. Representational distance learning for deep neural networks. Frontiers in computational neuroscience, 10, 131.
[5] Saito, T., & Watanobe, Y., 2020. Learning Path Recommendation System for Programming Education Based on Neural Networks. International Journal of Distance Education Technologies (IJDET), 18(1), 36-64.
[6] Syed, M. R., 2020. Methods and applications for advancing

distance education technologies: International issues and solutions. Simulation, 348.
[7] Heidari, A. A., Faris, H., Mirjalili, S., Aljarah, I., & Mafarja, M., 2020. Ant lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks. In Nature-Inspired Optimizers (pp. 23-46). Springer, Cham.
[8] S. Milinković, M. Maksimović, 2013. Case study: Using decision tree classifier for analyzing students' activities, volume 3 number 2 (pp. 87-95). Journal of Information Technology and Applications, Aprerion, Banja Luka, Republic of Srpska, BiH.
[9] Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J., 2019. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sensing, 11(2), 196.
[10] Gil, C. R., Calvo, H., & Sossa, H., 2019. Learning an efficient gait cycle of a biped robot based on reinforcement learning and artificial neural networks. Applied Sciences, 9(3), 502.
[11] Nguyen, H., & Bui, X. N., 2019. Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. Natural Resources Research, 28(3), 893-907.
[12] Liu, S., Oosterlee, C. W., & Bohte, S. M., 2019. Pricing options and computing implied volatilities using neural networks. Risks, 7(1), 16.
[13] Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., & White, E., 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. Remote Sensing, 11(11), 1309.
[14] Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufman: Burlington, MA, USA.
[15] Mair, C., Kadoda, G., Le, M., Phalp, K., Scho, C., Shepperd, M., Webster, S., 2000. An investigation of machine learning based prediction systems. J. Syst. Softw. 53, (pp. 23–29).
[16] Kotsiantis, S.B., 2007. Supervised Machine Learning: A Review of Classification Techniques. Informatica 31, (pp. 249–268).
[17] Carpenter, G.A., Grossberg, S., Reynolds, J.H., 1991. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. Neural Netw. 4, (pp. 565–588).
[18] Bishop, C.M., 2006. Pattern Recognition and Machine Learning; Springer: Berlin, Germany, Volume 4.
[19] Raudys, S.J.; Jain, A.K., 1991. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Trans. Pattern Anal. Mach. Intell. 13, (pp. 252–264).
[20] Марковић, Љ., 2009. Примена backpropagation алгоритма за обучавање неуронских мрежа, Симпозијум о операционим истраживањима, SYM-OP-IS 2009 Београд, (pp.1-4), ISBN 978-86-80593-43-2.
[21] Dr. David D. Pokrajac, 2013. Predicting students' results in higher education using neural networks, Conference: Applied Information and Communication Technologies, At Jelgava, Latvia.
[22] Зекић-Сушац, М., Фрајман-Јакшић, А. и Дрвенкар, Н., 2009. Неуронске мреже и стабла одлучивања за предвиђање успјешности студирања. Економски вјесник, XXII, (pp.314-327).
[23] I. Isaković., 2013. CRM performances accented with the implementation of data warehousing and data mining technologies, volume 3 number 2 (pp. 107-112). Journal of Information Technology and Applications, Aprerion, Banja Luka, Republic of Srpska, BiH.
[24] Tomić S., Drljača D., DDLM - Quality Standard for Electronic Education Programs in Higher Education of Bosnia and Herzegovina, JITA – Journal of Information Technology and Applications Banja Luka, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 9(2019) 2:67-79, (UDC: 004.735.8:621.39(497.6), (DOI: 10.7251/JIT1902067T), Volume 9, Number 2, Banja Luka, December 2019 (49-128), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004.

Submitted: January 30, 2020
Accepted: June 2, 2020.

About the authors

Dragan Vasiljević was born on September 16, 1968, in Skopje, Republic of North Macedonia. He graduated from the Military Academy in Rajlovac in 1991. At the faculty of Technical Sciences Čačak he enrolls a master's degree that ends in 2012. He is currently a Ph.D. candidate at Pan-European University "Apeiron", Banja Luka, Bosnia and Herzegovina, Europe.

Branko Latinović was born on April 28, 1956, in Prijedor, Bosnia and Herzegovina. He graduated from the Faculty of Economics in Banja Luka in 1980. At the same faculty he enrolls a master's degree that ends in 1994, and in 1997 successfully defended his doctoral dissertation. He works as a dean of the Faculty of Information Technologies of Pan-European University "Apeiron" in Banja Luka.

For citation

Dragan Vasiljević, Branko Latinović, Intelligent Distance Learning Systems, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 10(2020) 1:44-48, (UDC: 004.8.032.26), (DOI: 10.7251/JIT2001044V), Volume 10, Number 1, Banja Luka, June 2020 (1-68), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Implications of Implementation of Artificial Intelligence in the Banking Business in Relation to the Human Factor

Krunoslav Ris, Željko Stanković, Zoran Ž. Avramović

Pan-European University APEIRON, Banja Luka, Republic of Srpska, Bosnia and Herzegovina

krunoslav.ris@gmail.com, zeljko.z.stankovic@apeiron-edu.eu, zoran.z.avramovic@apeiron-edu.eu

Critical Review

https://doi.org/10.7251/JIT2001049R UDC: 004.8:336.7]:007.52, 004.5

Abstract: The banks are known as monetary management institutions because they deal with money. There is a number of customers that a bank daily interacts with. In this technology era, when everything is moving to automation from the beginning step to final product manufacture, medical checkups, medical reports, and evaluation, the banking system is still working on the legacy system. Instead, with the participation and implementation of new Virtual Assistant-powered with AI and Machine Learning technology in the banking sector, the institutions are again using the legacy system or may be bound to use the legacy system. This research will help to elaborate and emphasize the impact of the implementation of automation, using artificial intelligence in the banking business process. This research will be based on the quantitative as well model base prof of system performance using different analytical tools like SPSS. This automation process will help the institutions to enhance profitability, performance, and reduce human dependency. In a nutshell, Virtual Assistants powered with Artificial Intelligence will improve the business process performance in every sector of business, especially the banking sector.

Keywords: AI, Machine Learning, Automation, Banking Systems, Virtual Assistants, Chatbots.

Introduction

The whole banking sector is based on rules and regulations. Artificial intelligence helps to process the low level of information but the high level of volume in the data process, which results in lower the cost but also provides accurate and quick results instead of a human. By using the Algorithms, the AI is implemented in the banking sector with great success in customer management services, Credit Information Service, FAQ service, Financial Assistance Series, etc. All these types of services implemented using the Chatbot, AI Agents, and Self-learned system. Nowadays, the Banking sector is trying to enhance the performance by applying the AI in all fields like cybersecurity, risk management, fraud management, sales, Internal Audit, Financial Assistance, Asset Management, Loan Management, and Customer Management [5].

This research paper will elaborate on the framework of an AI system that will be implemented in all fields of banking sectors that has been discussed earlier. These models will be pre-testing using historical data and survey forms. These models can be applied in any or all fields of the banking sector to automate the business process using a robot – processing, Chatbot, and AI agent.

To enhance the customer services, the customer doesn't need to visit the branch, they just need to visit the website, and the pre-define Chatbot will get the required data according to the model. When the primary data is gathered, the Chatbot will ask about the problem statement and check if it meets the edibility criteria to communicate or stop communication. If the problem statement is matched with the Chatbot algorithm, then the Chatbot will proceed and guide the customer according to the written instructions. But in another case, the Chatbot will guide the customer to visit the nearby branch physically to get the problem solved.

A Chatbots will pull the data from the data warehouse much faster than a human, which leads us to performance. Performances are another critical aspect of the data warehouse as it is based on proper structure and great query engines that are fully optimized for reading and are supporting various incremental changes in the data. Another great characteristic is usability as users may not be familiar with how to get information from source data, but with the analytical approach of the data warehouse, users can by transforming, filtering, or slicing the data to find the information they need. In this way, users are getting a single source of data, which will be processed by AI instead of matching various sources of data, trying to find the information they are looking for [8].

To prevent the fraud, an AI system will monitor all the traffic and transaction within the organization, and if the blacklist transaction allowed by the employee multiple times, the system will record the logs, and by using the woods, the AI data analysis will catchup the fraud. This system will work intelligently without influence the working of the transaction; this system will also learn by experience in the system.

Finding information using computational methods gives immeasurably better results compared to data retrieval that relies only on classical methods [2].

To improve the operation efficiency, the front-end system will be integrated with the AI system in which the user will only input the single field information, and all the other fields will populate by using the intelligent learning. All the data will be filled in all the fields automatically. It will provide insight and analysis very fast so that an immediate decision can be taken. The process of tracing documents in real-time is much time consuming, but AI makes this process more comfortable and more efficient [7].

The commercial banks are moving to cutting edge technologies and process automation. Unicredit Bank has launched its Emerging Opportunities Engine (EOE) that is used to predict the investment recommendation and primarily focuses on the capital market. It also guides the customer to purchase the shares or sales in the stock market, and it is beneficial for the trading brokers. The initial results of this software are good, and now this software is going to implement in the financial institutes [3][4].

The artificial intelligence adaption can help to bank and finance industry to make the consistent and faster engagement of customers to their business by following the quick address of their problems. All the AI base software are used Machine Learning Algorithms that help the system to understand the procedure of problem and solve it according to the algorithm set of instructions. The natural language Chatbot and search engines can overcome client issues and allow the customer to get the desired information [1].

Rapid developments of techniques and technologies have driven the needs for the application of new knowledge in workplaces. In modern businesses, increasing standards, automation, and technologies have led to vast amounts of data becoming available. This explosive growth has generated an even more urgent need for techniques and tools that can assist us to convert this data into useful information and knowledge which can meet customer requirements [6]. All those data should be processed fast and in real-time. The only tool that can do this at present is AI powered with Machine learning algorithms.

Problem Statement

Management of traditional manual banking practices with the implementation of state-of-the-art AI systems Hypothesis.

The problem with the existing and old banking system is to decide based on massive data, is very expensive in term of cost as well as about twenty to thirty percent decision goes wrong due to incomplete and inappropriate information on the organizational plan. The AI state of the art system will deal with these issues intelligently and monitor all the information related to stakeholders to process the reports. This AI system will use real-time data to coordinate and guide the customer to take immediate decisions and govern according to the rules and regulations. This system will also maintain the profitability of the organization by increasing the credit by conducting multiple customers at the same time on the right way to immediately invest the money in the banking sector.

Hypothesis

Information systems of financial institutions with a strong artificial intelligence are an autonomous set of individual subsystems of all stakeholders of a financial organization, which has to cognitively/logically make the relevant decision regarding a particu-

lar procedure based on the investigated parameters at the macro and microeconomic level. The system should achieve interoperability of all components in order to function smoothly and safely. Interoperability at the organizational, logical, research and technological level creates essential technical and human preconditions for achieving full system interoperability as ideal. The assumption is that the system may be so ideal that it can completely replace the human factor through defined safety limits by manually checking individual procedures when resolving financial procedures.

The use of controls of organizational, financial and technological interoperability through the regular security procedures of the information system ensures a smooth, fast, accessible and secure operation of the information system in the domain of financial institutions.

Such hypotheses will be tested by research to be conducted in a leading bank in the CEE market, based on the example of the Informatics Division.

Objectives of the Research

The purpose of this research is to highlight the need and opportunities to improve the efficiency, effectiveness of service delivery and increase the profits of financial institutions and replace human factors with automatic Virtual Assistants and Chatbots. Full automation electronic services are the future of the functioning of all systems within financial organizations. On the other hand, the research goals are oriented towards the improvement of the model of information systems audit in public administration and its specific business.

One goal is to show how financial institutions with widespread machine learning models (which replicate human intuition and intelligence) provide digital platforms (information systems) with a way to reduce costs, reduce the workforce, improve banking experience and maximize profits.

The second objective of the research is to further improve the existing information system modules in order for the results to be realistic and, in scientific postulates, a finding that will improve the system and improve the degree of customer service provision.

The third objective of the research is to make a scientific and professional contribution to the development of artificial intelligence within financial institutions. This will be achieved by the aforementioned enhancement of the model that is currently generally and widely applicable.

The fourth objective of the research will be to contribute to improving the quality of service delivery of financial institutions at all levels, since the model will ensure system interoperability.

The offered extension and improvement do not refer to the steps in the internal processes of financial institutions, that is, does not create a completely new model of internal processes that has some special or special steps, but rather this improvement relates to its essence, subject and scope. In other words, the model offered will follow existing practices and contribute to expanding the scope of issues that test system interoperability and the effects of interoperability implementation in line with other entity functionalities, and in particular, the security of the system itself.

The offered model will be specific for financial institutions, but it will be possible to apply it to other types of organizations (agencies, credit unions, credit rating companies). The issue of semantic interoperability is a challenge that is central to new and emerging technologies (especially when it comes to the "Internet of Things" paradigm) but also to ensuring "intelligibility" between different systems. It is a universal paradigm, and therefore can be applied to different systems, thus creating an environment in which the system will function and communicate in an understandable "language".

Research

In the modern era of technology, automation, and utilization of technology in the field of banking has increased. Because the world is moving to the Global Village concept in which anyone from anywhere can get its desired product without lifting the boxes of cash in hand. The banking sector now a day have been completely transferred over the technology; due to this, it would make it possible to purchase any product from anywhere in the world. Several algorithms are working behind the technology to process the information and customer request on all types of banking services like online banking, android apps banking, digital banking, ATM transfer, Wire transfer, etc. All these types of services have been synchronized with the banks using machine

learning algorithms that automatically track data, verify the integrity of data, and synchronize it with the central database of the concerned bank. For all demanded that we mentioned, we need to utilize workforce which needs to be available 24/7. Availability of manpower 24h in a day increases banks operational costs and decreases profit. The purpose of the research is to find out will customers accept to use robots, Virtual Assistants and Chatbots instead of real human beings.

ALGORITHMS

Several algorithms can be used for data mining in Machine learning, but some of specific that those are only used within the banking sector. There are two main types of machine learning data mining algorithms that we use in building new AI Framework.

1. Supervised Learning Algorithms
2. Unsupervised Learning Algorithms

DATASETS

Dataset consists of 517 observations are obtained from the Survey. This dataset consists of 37 different questions from the personal info to business info, and Chatbot uses. The survey questions which are most related to our research belong to the Chatbot. Such as:

- Have you ever chatted with a Chatbot?
- Has your bank had offered Chatbot functionality/support?
- Do you prefer using Chatbots in a Banking Business?
- If you can do most of your Banking Business through Chatbot, would you still go to the Bank branch office?
- Would you like that you can talk to your Bank virtual assistant like you are chatting with a Chatbot?

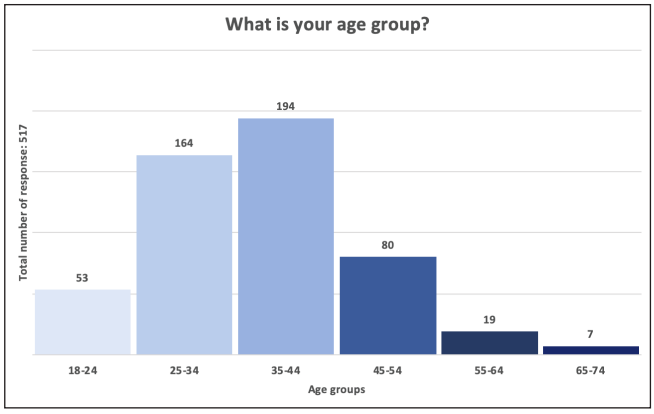
As they are within the scope of research topic, these questions are most important in the statistical analysis. This hypothesis testing will be done using the IBM SPSS V.23 statistical analysis tool. SPSS is an abbreviation of Statistical Package for the Social Science that has many versatile and automated analysis algorithms for both the qualitative and quantitative analysis. There are many features of SPSS like it can handle a large amount of data and manipulate the human-understandable logical results according to the applied algorithm. It used elementary statistics function as well as advanced features to measure the output. The most appropriate models like ANOVA, Correlation, Means, Chie Square Test, and multiple regression test, can be implemented in SPSS.

METHODS AND TECHNIQUES

In order to analyze the validity of the dataset as well as the hypothesis, there is the dataset is survey responses and an ordinal dataset, so two models are designed on this dataset. For the validity of the dataset, the frequencies algorithm is applied to validate the ratio of each attribute. Later on, the Multinomial Logistic Regression is applied to learn the validity of the hypothesis statement; either the hypothesis is accepted or rejected.

Frequencies Analysis

What is your age group?

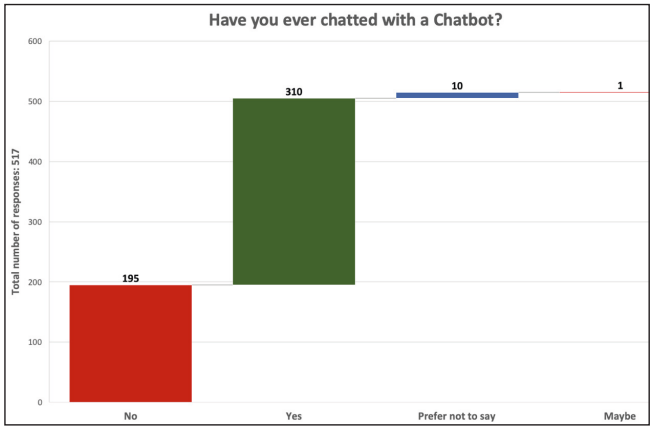


		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18-24 years old	53	10.1	10.1	10.1
	25-34 years old	164	31.7	31.7	41.8
	35-44 years old	194	37.5	37.5	79.3
	45-54 years old	80	15.5	15.5	95.0
	55-64 years old	19	3.7	3.7	98.6
	65-74 years old	7	1.4	1.4	100.0
	Total	517	100.0	100.0	

Figure 1: *Graph and frequency: What is your age group?*

In the above table, 6 different age groups are given in which the first group response consists of 53, while the second group consists of 164, and third group consists of 194 participants. The second column shows the frequency of how many participants are there in each age group. The percent column displays the percentage of overall data, while the valid percentage column shows the data validity percentage of total dataset. The total valid percentage dataset set is 100%.

Have you ever chatted with a Chatbot?

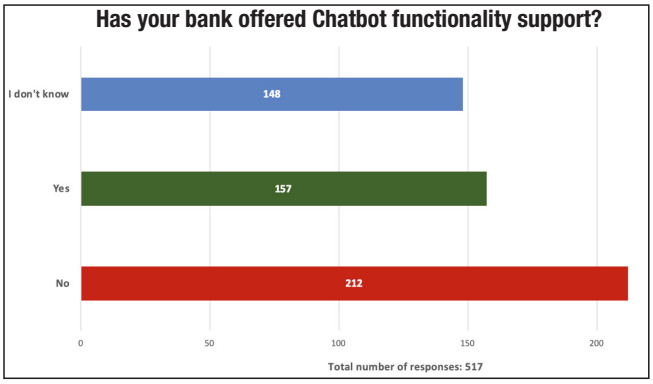


		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		1	.2	.2	.2
	Maybe	1	.2	.2	.4
	No	195	37.7	37.7	38.1
	Prefer not to say	10	1.9	1.9	40.0
	Yes	310	60.0	60.0	100.0
	Total	517	100.0	100.0	

Figure 2: *Graph and frequency: Have you ever chatted with a Chatbot?*

In the given table, the status of chat over Chatbot is discussed. According to it, most persons use Chatbot for chat purposes for meeting their frequently asked question. The second column shows the frequency of how many participants are there in the related group. The percentage column displays the percentage of overall data, while the valid percentage column shows the data validity percentage of the total dataset. The total valid percentage dataset set is 100%.

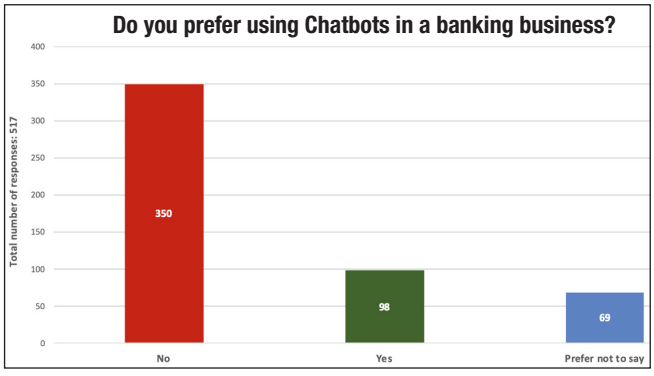
Has your bank offered Chatbot functionality support?



		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	I don't know	148	28.6	28.6	28.6
	No	212	41.0	41.0	69.6
	yes	1	.2	.2	69.8
	Yes	156	30.2	30.2	100.0
	Total	517	100.0	100.0	

Figure 3: *Graph and frequency: Has your bank offered Chatbot functionality/support?*

In the given table, the total values are 517, out of which the majority belong to no answer. That means the results are not favorable in this table. The second column shows the frequency of how many participants are there in the related group. The percentage column displays the percentage of overall data, while the valid percentage column shows the data validity percentage of the total dataset. The total valid percentage dataset set is 100%.

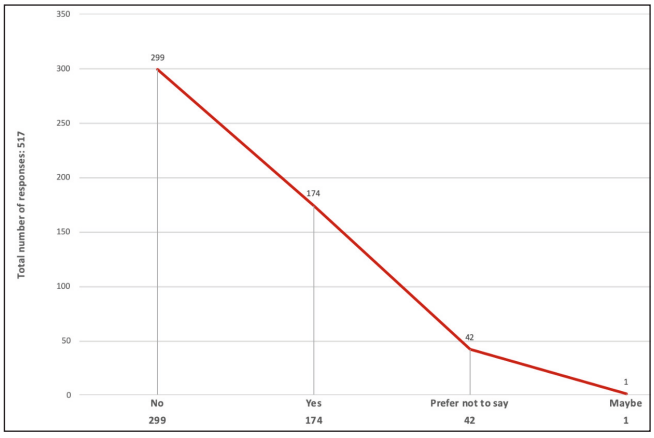


		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	350	67.7	67.7	67.7
	Prefer not to say	69	13.3	13.3	81.0
	Yes	98	19.0	19.0	100.0
	Total	517	100.0	100.0	

Figure 4: *Graph and frequency: Do you prefer using Chatbots in Banking Business*

In the given table, the total values are 517, out of which the majority belong to no answer. That means the results are not favorable in this table. The second column shows the frequency of how many participants are there in the related group. The percentage column displays the percentage of overall data, while the valid percentage column shows the data validity percentage of the total dataset. The total valid percentage dataset set is 100%.

Doing banking business with Chatbot, or going to the bank branch office?

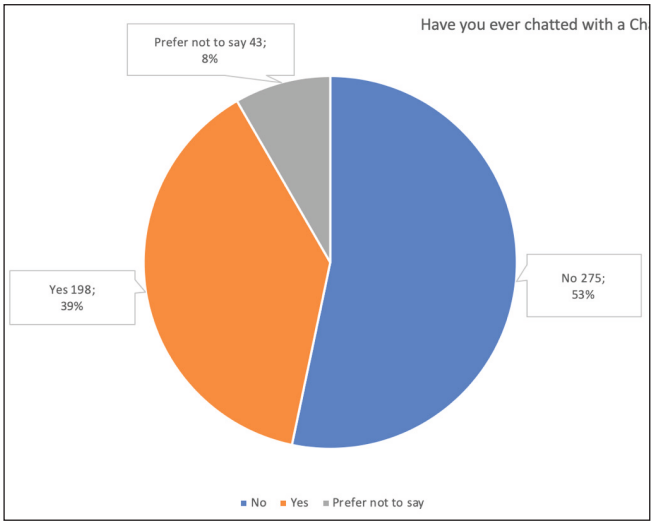


		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		1	.2	.2	.2
	Maybe	1	.2	.2	.4
	No	299	57.8	57.8	58.2
	Prefer not to say	42	8.1	8.1	66.3
	Yes	174	33.7	33.7	100.0
	Total	517	100.0	100.0	

Figure 5: *Graph and frequency: Doing banking business with Chatbot, or going to the bank branch office?*

In the given table, most of the responses belong to the Chatbot in which the user prefers to use Chatbot instead to visit the branch, as about 299 are disagree with visiting branch, but they prefer to use Chatbot for the problem solution. The second column shows the frequency of how many participants are there in the related group. The percent column displays the percentage of overall data, while the valid percentage column shows the data validity percentage of the total dataset. The total valid percentage dataset set is 100%.

If you could talk anytime with VA, would you prefer to go to branch office?



		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		1	.2	.2	.2
	No	275	53.2	53.2	53.4
	Prefer not to say	43	8.3	8.3	61.7
	Yes	198	38.3	38.3	100.0
	Total	517	100.0	100.0	

Figure 6: *Graph and frequency: If you could talk anytime with VA, would you prefer to go to branch office?*

In the given table, most of the want to use Chatbot to talk as a virtual assistant and get prompt responses instead of waiting for a long time. The second column shows the frequency of how many participants are there in the related group. The percent column displays the percentage of overall data, while the valid percentage column shows the data validity percentage of the total dataset. The total valid percentage dataset set is 100%.

Multinomial Logistic Regression

Multinomial Logistic Regression is used to predict a nominal dependent variable given one or more independent variables. As the dataset consists of the many string values, so the best model that fits for the analysis of dataset validity is multinomial Logistic Regression.

For the purpose of model validity, two variables are selected for this test.

- Sex
- Do you prefer using Chatbot in a banking business?

The first variable is used as an independent variable, and the second is used for the dependent variable. The results of the test are given below.

Case Processing Summary

		N	Marginal Percentage
Do_you_prefer_using_ChatBots_in_a_Banking_Business	No	351	67.8%
	Prefer not to say	69	13.3%
	Yes	98	18.9%
Sex	Female	197	38.0%
	Male	318	61.4%
	Prefer not to say	3	0.6%
Valid		518	100.0%
Missing		0	
Total		518	
Subpopulation		3[a]	

a. The dependent variable has only one value observed in 1 (33.3%) subpopulations.

In the above result table, column N provides the number of observations used in the model. Three types of observations are used, male, female, and prefer not to say. The next column, Marginal Percentage, provides the list of valid observation funds in each outcome variable. The validity is calculated by dividing the total observation with the number of observations per category. In this table, the valid means of how much data or observations are valid from the total observations. The dataset validity rate is 100%. The next result indicates the number of subpopulations contained in the data. A subpopulation of the data consists of one combination of the predictor variables specified for the model.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	41.760			
Final	21.299	20.461	4	.000

In this table, the first column is a model that contains two parameters - the Intercept Only and Final. This indicates the parameters of the model for which the model fit is calculated. "Intercept Only" describes a model that does not control for any predictor variables and simply fits an intercept to predict the outcome variable. "Final" describes a model that includes the specified predictor variables and has arrived at through an iterative process that maximizes the log-likelihood of the outcomes seen in the outcome variable. By including the predictor variables and maximizing the log-likelihood of the outcomes seen in the data, the "Final" model should improve upon the "Intercept Only" model. This can be seen in the differences in the -2(Log-Likelihood) values associated with the models.

The second column is about the model fitting criteria by which the model results are produced. This is the product of -2 and the log-likelihoods of the null model and fitted "final" model. The likelihood of the model is used to test whether all predictors' regression coefficients in the model are simultaneously zero and in tests of nested models.

Chi-Square – This is the Likelihood Ratio (L.R.) Chi-Square test in which at least one of the predictors' regression coefficient is not equal to zero in the model. The L.R. Chi-Square statistic can be calculated by -2 * L(null model) – (-2*L(fitted model)) = 41.760 – 21.299 = 20.299, where L(null model) is from the log-likelihood with just the response variable in the model (Intercept Only) and L(fitted model) is the log-likelihood from the final iteration (assuming the model converged) with all the parameters.

df – This indicates the degrees of freedom of the chi-square distribution used to test the L.R. Chi-Square statistic and is defined by the number of predictors in the model.

Sig. – This is the probability of getting a Multinomial Regression test statistic being as extreme as, or more so, than the observed statistic under the null hypothesis. This model indicates the p-value. The p-value is used to determine the model authentication and hypothesis; either the hypothesis is accepted or not, and the results are significant or not. The standard value of p should be p-value <= 0.05 for the significance of the model. In this case, the sig value is lower than 0.05 which indicates that the null hypothesis is rejected, and results are significant.

Parameter Estimates

Do_you_prefer_using_ChatBots_in_a_Banking_Business[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
No	Intercept	15.916	.146	11916.015	1	.000			
	[Sex=Female]	-14.826	.236	3959.175	1	.000	3.639E-7	2.293E-7	5.776E-7
	[Sex=Male]	-14.547	.000	.	1	.	4.811E-7	4.811E-7	4.811E-7
	[Sex=Prefer not to say]	0[b]	.	.	0	.	.	.	.
Prefer not to say	Intercept	-.351	1967.015	.000	1	1.000			
	[Sex=Female]	.425	1967.015	.000	1	1.000	1.530	.000	.[c]
	[Sex=Male]	-.431	1967.015	.000	1	1.000	.650	.000	.[c]
	[Sex=Prefer not to say]	0[b]	.	.	0	.	.	.	.

a. The reference category is: Yes.

b. This parameter is set to zero because it is redundant.

c. Floating point overflow occurred while computing this statistic. Its value is therefore set to system missing.

In this table, the second column indicates the relationship between the dependent and independent variables. These are the estimated multinomial logistic regression coefficients for the models. An important feature of the multinomial logit model is that it estimates k-1 models, where k is the number of levels of the outcome variable. At the bottom of this table, it is indicated that the YES is used as a reference category that is significant with the results. The standard error of the individual regression coefficients for the two respective models estimated. In the next column, Wald chi-square test that tests the null hypothesis that the estimated equals. On the other hand, the df is the list of columns that have the degree of freedom of each variable in the dataset. The degree of freedom used in this dataset is 1.

The most important column in this table is sig that is also known as the p-value table. It is the coefficient of the probability that is given in the model. For the best fit and result from the model, sig value should be sig <= 0.05, which is given in the table. According to this, it can be concluded that the model is significant and the null hypothesis is rejected.

CONCLUSION

Interpretation of the obtained results proved the hypothesis, using statistical models such as ANOVA, T-Test, and the latest possibilities of SPSS tools such as multinominal regression tests and regular regression tests. In addition to this statistical evidence, this paper shows how the unification of human work functions with the artistic intelligence of the present, and there is no word on the concept of the future.

The future is today. On the examples of world banking conglomerates, it has emerged how artificial intelligence tools currently on the market are used. It is evident that substantial financial resources are being invested in reducing the company's operating expenses and increasing productivity.

The only possibility for such a thing in the coming times is explored in artificial intelligence that will cover vacation, pre-war sites, maternity leave, and all the other things a human being needs. Virtual assistants, Chatbots, holograms, physical robots will upgrade over the years to flood the market due to all the cheap technologies. A man as an individual will become conceptually obsolete and will not be able to fulfill the need to practice a large amount of information that will overwhelm us daily. Each of us will need a personal assistant who will take care of things for us, from arranging a meeting, reminders, providing information, all the way to a mentor or educator for a particular area of interest.

An army of people who, about ten years ago, processed data in the FinTech industry and pulled that data from databases, will be engaged with other creative jobs. Repetitive manual work can be disassembled into a segment that artificial intelligence can absorb and serve as a ready-made solution in any number of users. Today it is no longer essential to have banks in a small community. Today, it is only relevant that this small community has an Internet connection, and spare time that they can utilize as they want.

A survey of 517 respondents resulted in the majority of users who worked while sitting at home and setting up their virtual existence to do the work

for them. The days of visiting commercial banks and wasting time waiting for someone behind the counter to shout "Next" are already gone.

References

[1] AI Innovation Playbook, URL: https://www.pymnts.com/wp-content/uploads/2019/06/AI-Innovation-Playbook_June-2019.PDF, date of access: 11.12.2019.

[2] Avramović Zoran Ž, Dražen Marinković, Igor Lastrić (2017), Use of computer search algorithms in the research of statistical, semantic and contextual rules of language in digital information space. JITA – Journal of Information Technology and Applications Banja Luka, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA Volume 7, Number 1 (2017)

[3] HINA, *UniCredit predstavlja strateški plan, 2019.,* URL: https://www.glasistre.hr/gospodarstvo/unicredit-predstavio-strateski-plan-ukida-8000-radnih-mjesta-do-2023-608293 , date of access: 11.12.2019.

[4] Pavlou LIe, D.; *What Drives Mobile Commerce; 2007., URL:* https://doi.org/10.2139/ssrn.2380712 date of access: 28.11.2019.

[5] Parsons, G.; Productivity and Computers in Canadian Banking; *Productivity Issues in Services at the Micro Level;* 150., 1993.

[6] Rostami, N. A.; *Integration of Business Intelligence and Knowledge Management – A literature review*, Isfahan University, Iran Journal of Intelligence Studies in Business Vol 4, No 2 30-40; 2014.; URL: https://pdfs.semanticscholar.org/1c9b/85e5519bd0bdee7ebbbddfea77695ddef2fb.pdf, date of access: 28.11.2019.

[7] Schank, R. (2016) „Once upon a time in AI " URL: https://www.sciencedirect.com/science/article/pii/S0004370206000919?via%3Dihub date of access: 28.11.2019.

[8] Šikanjić N., Avramović Ž. Z., New Approach of Storing and Retrieving Large Data Volumes, JITA – Journal of Information Technology and Applications Banja Luka, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 9(2019) 2:89-98, (UDC: 004.6:658.4]:519.8), (DOI: 10.7251/JIT1902089S), Volume 9, Number 2, Banja Luka, december 2019 (49-128), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Submitted: January 2, 2020
Accepted: June 1, 2020.

About the authors

Krunoslav Ris is a PhD candidate at the Faculty of Information Technologies at the Pan-European University APEIRON Banja Luka and has MA in information technology and Computer Science. His main research interests are in Digital Transformation, Artificial Intelligence, Industry 4.0. He is currently engaged as a Head of IT development at Unicredit Group - Zagrebačka Banka.

Zeljko Stankovic was born in Belgrade. He finished primary and secondary school in his hometown. He received his higher education in Cleveland, Ohio, USA, where he graduated in 1981. In 2006, he defended his master's thesis at the University of Novi Sad. He defended his doctoral dissertation at Singidunum University in 2010. He has been programming since 1984. making programs for his first computer Commodore 64. Robotics and bioengineering are a long-standing field of work and interest.

Zoran Ž. Avramović has published 319 scientific and professional papers:

- 9 papers in WoS (SCI) journals,
- 14 papers in journals cited in Scopus bibliographic database,
- 25 papers in international scientific journals,
- 31 papers in domestic scientific and peer-reviewed journals,
- 124 papers on international conferences held abroad and
- 116 papers on domestic scientific and professional conferences (38 of which are by invitation).

As a leader, author, co-author or contributor, he participated in the design of 150 studies and projects. He has applied for five patents. He has published three scientific monographs, one university textbook and one manual.

For citation

Krunoslav Ris, Željko Stanković, Zoran Ž. Avramović, Implications of Implementation of Artificial Intelligence in the Banking Business in Relation to the Human Factor, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 10(2020) 1:49-57, (UDC: 004.8:336.7]:007.52, 004.5), (DOI: 10.7251/JIT2001049R), Volume 10, Number 1, Banja Luka, June 2020 (1-68), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

THE ROLE OF POSTPROCESSOR IN THE TRANSLATION OF VECTOR GRAPHICS UNDERSTANDABLE TO THE CNC MACHINE CONTROLLER

Boris Pauković, Dražen Marinković
Pan-European University APEIRON Banjaluka
paukovic@hotmail.com, drazen.m.marinkovic@apeiron-edu.eu

Critical Review
https://doi.org/ 10.7251/JIT2001058P UDC: 621.9-52:004.388, 004.382:621.01

Abstract: The purpose of this paper is to show the post-processor role and importance in the creation of programming code which CNC machine controller can understand and proceed. Today, the use of CNC technology introduces the Industry 4.0 principles in the production, thus increasing the productivity and precision of the produced parts. It is very important to optimize all the production steps, from choosing the right CAD software for vectors drawing, defining tools and generating tool paths, creating and optimizing post-processor, to translating the tool paths in the programming code which controller of the CNC machine can understand, as well as educating the operators to be able to calibrate the machine and understand and run the CNC programming code properly. When all the abovementioned steps are correctly defined, the production can be optimized and best results are guaranteed.

Keywords: CNC technology, CAD/CAM software, postprocessor, CNC programming.

WHAT IS CNC PROGRAMMING

By defining the programming of CNC machines as creating commands that define pre-conceived actions to the control unit to obtain the final product, which in many cases is very expensive, we can conclude that it is very important that all steps are well analyzed and optimized. It is necessary to know the control unit that is located on the CNC machine and based on this information to enable the creation of program code.

There are two basic ways to create a program code:

1. Manual programming - the programmer writes the CNC program independently using commands known to the control unit. This way of programming is quite demanding, very slow and prone to errors because there is no simulation, which is a very important step, especially if the program will be executed on a CNC machine for the first time.
2. Automatic programming - this way you create drawings (2D or 3D models) in one of the CAD programs. The drawing is entered into the CAM program in which important parameters are defined such as: thickness of machining material, tool, starting point, position of the draft/model in the material, idle speed, speed of tool passing through the material, tool rotatin frequency, and with the help of a *processor* as a group of programs that determines the toolpath on the basis of input predefined parameters. In order for the control unit itself to successfully execute the pre-generated path, it is necessary to use the *postprocessor* as a compiler that converts the defined path into a language understandable to the control unit.

THE ROLE OF POSTPROCESSOR IN CREATING A PROGRAM CODE

A CNC postprocessor is a set of commands that converts a entered tool path into a language understandable to the control unit so that the tool can

move safely and consistently along a predefined path.

Most CAM software is designed to be independent of specific control units but to allow customization to each control unit. CAM software generates files that define tool paths and other necessary information to obtain the final product from the material, but this information is completely unusable on a specific CNC machine without the participation of postprocessors.

The main role of the postprocessor is to read the path from the files generated by the CAM software and convert it to G-code that will be understandable to the control unit.

It is important to emphasize that the translation into the program code for optimal use is not enough without knowing the mechanical components of the machine itself, because improper adjustment leads to mechanical failures that are usually very expensive and not covered by the manufacturer's warranty. Creating and optimizing postprocessors is not an easy task - it is necessary to have a very good command of programming skills and to be familiar with the principles of mechanical engineering, kinematics and mechatronics.

The following graphics represent a schematic representation of the process of generating program code from vector graphics and the role of postprocessor [2]:

What a postprocessor looks like - an example of translation of a program code

G-code is the most widely used language for numerical control today. The machine control unit uses the G-code to start the machine and execute the commands. The G-code was originally developed in 1958 at MIT (Massachusetts Institute of Technology) and is considered the industry standard for programming numerically controlled control units, known as RD-274. A finite number of basic control commands are available but depending on the software manufacturer can be expanded and specified according to customer needs.

A few basic commands which will be used in the representative postprocessor presented in this paper are:

G0 - linear movement that is simultaneous and coordinated for all axes that are present, in practice it is used to move and define the coordinates of the axles at idle.

Syntax example:

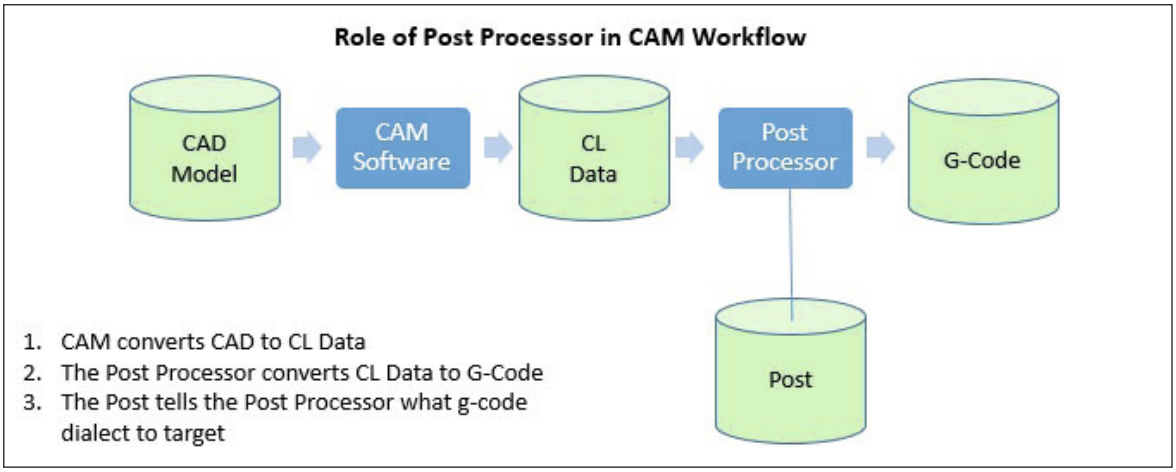
G0 [X] [Y] [Z] [F], where X, Y, Z are the coordinates of the three-axis CNC machine, F is the offset

Example of a program code (taken from a generated representative program code):

G0 X103.5887 Y94.4212 Z5.0000

G1 - linear motion that is simultaneous and coordinated for all axes present. In practice, it is used to move and define coordinate positions when the tool is in the material.

Figure 1: *The role of postprocessor in the G-code creation process*



Syntax example:

G1 [X] [Y] [Z] [F], where X, Y, Z are the coordinates of the three-axis CNC machine, F is the offset

Example of a program code (taken from a generated representative program code)

G1 Z-1.3333 F720.0

G1 X125.3425 F2520.0

G1 X116.7836 Y117.2038

G2 - Circular interpolation clockwise

Syntax example:

G2 [X] [Y] R [Radius] [F], where X, Y, Z are the coordinates of the three-axis CNC machine, R is the radius, F is the offset

Example of a program code (taken from a generated representative program code)

G2 X114.1735 Y124.8695R184.856

G2 X110.9580 Y114.1032R92.383

G3 - Circular interpolation counterclockwise

Syntax example:

G2 [X] [Y] R [Radius] [F], where X, Y, Z are the coordinates of the three-axis CNC machine, R is the radius, F is the offset

Example of a program code (taken from a generated representative program code)

G3 X187.3316 Y102.6650R38.223

G3 X192.2292 Y103.7325R20.928

G3 X194.6210 Y104.9136R11.432

There are also M-codes, with the help of which the control unit sends signals for starting the spindle, the direction of rotation of the spindle, measuring the tool, replacing the tool, etc.

M code	Description	Syntax example
M3	Start spindle rotation	M03 [S], where the variable S is the spindle rotation speed
M5	Stop spindle rotation	M5
M6	Tool replacement	M06 T [T], where the variable T is the number of tools in the tool holder
M30	Arrival at the starting point	M30

In the postprocessor, we define all the necessary parameters that are necessary for the G-code to be completely understandable to the control unit. Only then can the control unit execute all desired commands. The postprocessor consists of several parts:

1. Basic information - where the name of the postprocessor is most often defined (indicates the name of the control unit for which the postprocessor is in charge and will be displayed in the CAD/CAM software when selecting a postprocessor for G-code generation); extension of this unit of measurement:

Example:

```
DESCRIPTION = "OSAI OPEN CONTROL(*.cnc)"
FILE_EXTENSION = "cnc"
UNITS = mm
```

2. Defining the numbering of program blocks:

Example:

```
; Block numbering;
LINE_NUM_START = 10
LINE_NUM_INCREMENT = 10
LINE_NUM_MAXIMUM = 9999999
START = "N[N] G300 F4"
```

3. Defining the format of variables:

Example:

```
; Line numbering
FORMAT = [N|@|N|1.0]
; Spindle Speed
FORMAT = [S|@|S|1.0]
; Feed Rate
FORMAT = [F|#|F|1.1]
; Tool moves in x,y and z
FORMAT = [X|#|X|1.4]
FORMAT = [Y|#|Y|1.4]
FORMAT = [Z|#|Z|1.4]
```

4. Program header where it is defined which tool will be used first in machining, material dimensions, starting point, starting of the spindle and so on.

Example:

```
START = ";First tool:[T] [TOOLDESC]"
START = ";Material dimensions: X=[XSIZE],
   Y=[YSIZE], Z=[ZMATERIAL]"
START = "[N] G90"
START = "[N] G40"
START = "[N] G80"
START = "[N] (UAO,1)"
START = "[N] M06 T[T]"
START = "[N] M03 [S]"
```

5. Definition of movement:

Example:

```
;Linear section:
RAPID_RATE_MOVE = "[N] G0 [X] [Y] [Z]"
FIRST_FEED_RATE_MOVE = "[N] G1 [X] [Y] [Z] [F]"
FEED_RATE_MOVE = "[N] G1 [X] [Y] [Z]
; Arc Section:
CW_ARC_MOVE = "G2 [X] [Y]R[Radius] [F]"
CCW_ARC_MOVE = "G3 [X] [Y]R[Radius] [F]"
```

6. End of the program (spindle shutdown, return to starting point or predefined point with other possible options).

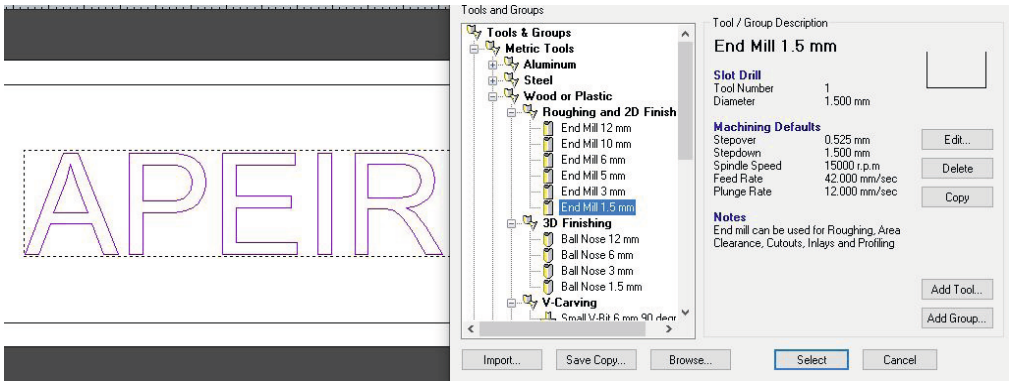
Example:

```
; End of file
END = "[N] G80"
END = "[N] M05"
END = "[N] M30"
```

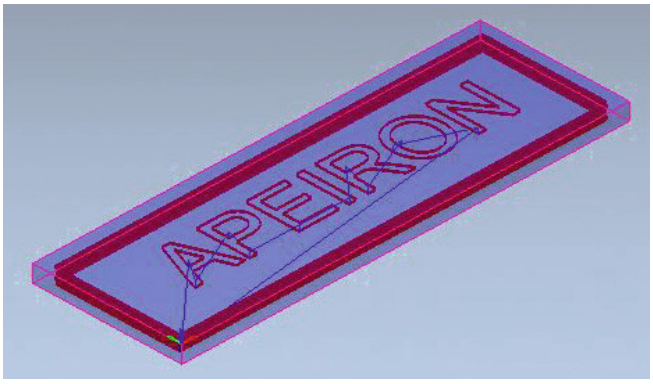
An example of a draft that has been converted into a G-code understandable to the control unit of the Italian manufacturer OSAI with the help of a postprocessor created on the basis of the previously described basic procedures.

1. Step: Use CAD software to create or import an existing vector graphic:

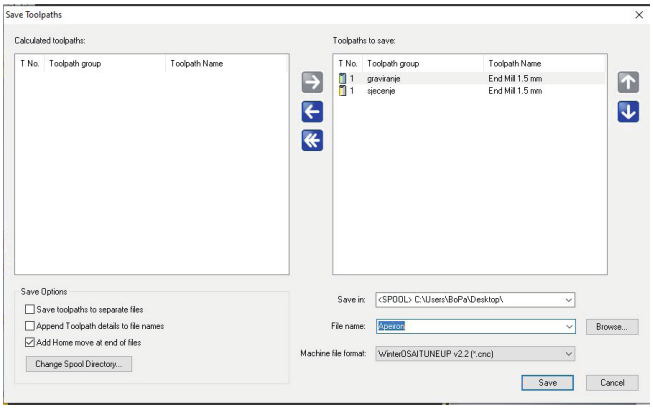
2. Select the tools:



3. Create a toolpath:



4. Generate a G-code with the help of a postprocessor:



Header and the first few lines of the code:

```
N10 G300 F4
;First tool 1 1.500 mm dia. slot drill
;Material dimensions : X=600.000, Y=200.000,
    Z=25.000
N40 G90
N50 G40
N60 G80
N70 (UAO,1)
N80 M06 T1
N90 M03 S15000
N100 M49
N110 h1
N120 G27
N130 G17
N140 G0 X103.5887 Y94.4212 Z5.0000
N150 G1 Z-1.3333 F720.0
N160 G1 X125.3425 F2520.0
N170 G1 X116.7836 Y117.2038
G2 X114.1735 Y124.8695R184.856
G2 X110.9580 Y114.1032R92.383
N200 G1 X103.5887 Y94.4212
N210 G1 Z-2.6667 F720.0
N220 G1 X125.3425 F2520.0
N230 G1 X116.7836 Y117.2038
G2 X114.1735 Y124.8695R184.856
G2 X110.9580 Y114.1032R92.383
N260 G1 X103.5887 Y94.4212
N270 G1 Z-4.0000 F720.0
N280 G1 X125.3425 F2520.0
N290 G1 X116.7836 Y117.2038
G2 X114.1735 Y124.8695R184.856
```

OSAI OPENCONTROL CONTROLLER HARDWARE STRUCTURE

OpenControl is a family of controllers designed by the Italian manufacturer OSAI. The most powerful configurations can support up to 64 digital axes. It is installed and controls numerous types of machines; CNC milling machine, sharpener, marble processing machine, laser cutters and other. Thanks to the "OPEN" philosophy, it is quite flexible and is very popular among manufacturers, as well as final users due to its simple and intuitive user interface. The user interface is known as "WinNBI" (Windows Network Based Interface) based on the Windows Embedded CE operating system, known for its stability and durability.

The Windows Embedded CE system in this environment retains the basic functionality of the Windows operating system with the ability to run the software required to manage the machine itself.

OpenControl systems are modular based and can be combined as needed. In the most widespread and popular setting, the system consists of an operator panel where parameters are manipulated and commands to start the machine. The following figure [2] shows the logical setup of the system consisting of a panel connected to an industrial computer, an OpenControl controller, a central part of the system that communicates using TCP/IP protocol with the industrial computer and further with the IO module using the standard fieldbus protocol.) and

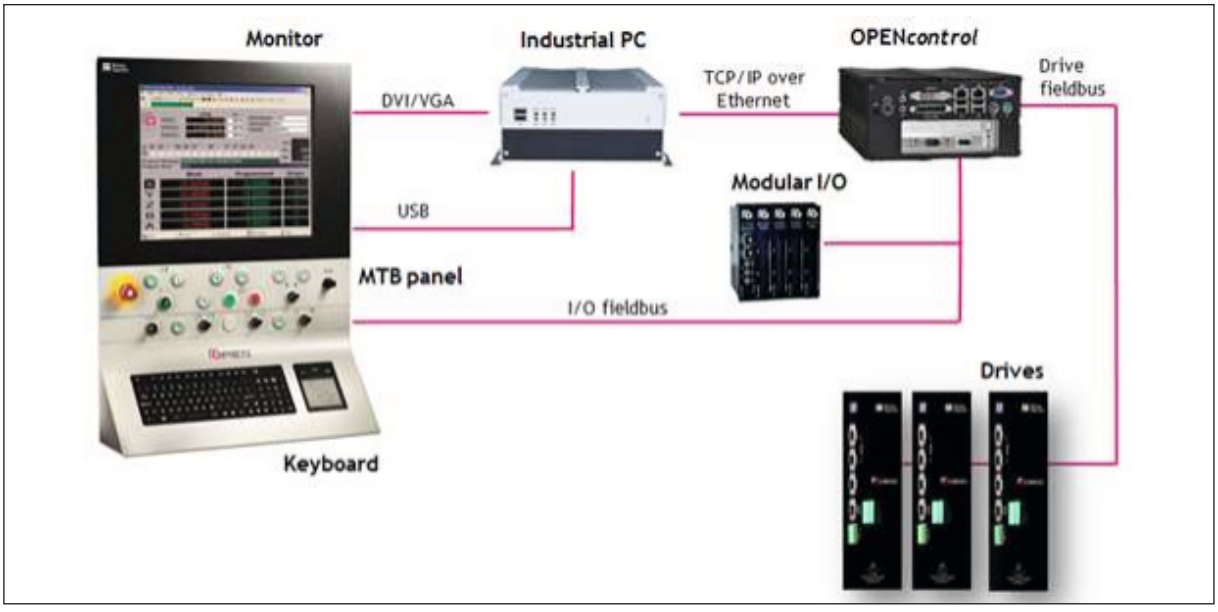


Figure 2: *A display of OSAI OpenControl system*

recently with the help of EtherCAT protocol (Ethernet for Control Automation Technology) - Ethernet based fieldbus protocol, created by the German manufacturer Beckoff Automation, standardized as IEC61158, offering great stability and reliability required by real-time computing and automation, with drivers that allow the machine to start and execute commands.

A practical example of motion optimization of a CNC machine for embossing using paramacro subroutines

Optimization of CNC machine movement is crucial when creating specific projects. The speed of production, along with the quality of the finished product with minimal wear of the components of the machine itself is crucial during installation and initial setup. This is often not taken into account in today's production.

The following example will show the basic functions of paramacro subroutines with a practical example where the processing speed is increased by up to 40% thus satisfying the demand of the final customer.

Based on the instructions for OSAI controller developers, we come to the conclusion that paramacro subroutines can be used in user-defined cycles and are called with a three-digit code.

Basic syntax of paramacro subroutine [1]:

Gn parameter-name-1 [value-1].... [parameter-name-n] [value-n]..... [string]
whereby:

n	a number from 300 to 998
parameter-name-1....parameter-name-n	letters (defined based on Table 2
value-1....value-n	can be a number or E parameter
string	a string of characters (maximum 99)

There are two basic groups of paramacro subroutines:

1. Non-modal paramacros (G300-G699)
2. Modal paramacros (G700 - G998)

The main difference is that modal paramacros can only be active in blocks that do not have M functions. No other function or code can be active or executed while the modal paramacro is active.

In this example, we will use the non-modal paramacro subroutine G300 that will be called in the postprocessor and will perform optimal acceleration of the axles in order to obtain the speed in operation:

```
;PARAMACRO G300
"1" ;SLOW
ERF=0.05
MDA=80
;VELOCITY LEAP
(CPA,W,X,92,100)
(CPA,W,Y,92,100)
(CPA,W,Z,92,100)
;WORK JRK
(CPA,W,X,16,10000)
(CPA,W,Y,16,10000)
(CPA,W,Z,16,20000)
;WORK DEJRK
(CPA,W,X,122,10000)
(CPA,W,Y,122,10000)
(CPA,W,Z,122,20000)
;SPLINE
TBS(X)=0.01
TBSG0(X)=0.2
TBS(Y)=0.01
TBSG0(Y)=0.2
TBS(Z)=0.01
TBSG0(Z)=0.2
(GTO,END)
;
```

In the above example, we see the difference in the parameters VELOCITY, JERK, DEJERK, which are directly responsible for the speed of the machine itself. The solution presented using subroutines represents an elegant change of essential parameters with small corrections within the program code of the postprocessor, while manual adjustment and

reprogramming would result in several hours of adjustment with a high possibility of errors that often cause major mechanical failures and production downtime. Also, the extension and additional increase of speed is enabled by adding new F parameters and expanding the subroutine in a simple way.

After the successful definition of paramacro subroutines, it is necessary to perform the integration with the postprocessor used for the generated program G-code recognizable by the CNC machine controller itself.

Most often, the paramacro subroutine is called and defined in the postprocessor header itself, as in the following example:

```
"2" ;FAST
ERF=0.15
MDA=80
;VELOCITY LEAP
(CPA,W,X,92,200)
(CPA,W,Y,92,200)
(CPA,W,Z,92,200)
;WORK JRK
(CPA,W,X,16,20000)
(CPA,W,Y,16,15000)
(CPA,W,Z,16,30000)
;WORK DEJRK
(CPA,W,X,122,20000)
(CPA,W,Y,122,15000)
(CPA,W,Z,122,30000)
;SPLINE
TBS(X)=0.05
TBSG0(X)=0.2
TBS(Y)=0.05
TBSG0(Y)=0.2
TBS(Z)=0.05
TBSG0(Z)=0.2
```

```
DESCRIPTION = "OSAITUNEUP v2.2 (*.cnc)"
;
FILE_EXTENSION = "cnc"
;
UNITS = mm
;
; Carriage return - line feed at end of each line
;
END_OF_LINE = "[13][10]"
;
; Block numbering
;
LINE_NUM_START = 10
LINE_NUM_INCREMENT = 10
LINE_NUM_MAXIMUM = 9999999
START = "N[N] G300 F2"; here we can see that the called paramacro is G300 and ;defined speed F2, FAST
;
; Set up default formatting for variables
; .....
```

CONCLUSION:

The development of CNC machines and affordability in recent years has enabled this technology to come into widespread use. Modernization of production of CNC machines enables greater efficiency and profitability. Machines are becoming more common in many industries. Postprocessor as one of the more important items in this system is credited with the proper transmission of information and the creation of commands that the machine understands. By using a postprocessor, errors in writing programs are minimized. Proper setup and optimization of the postprocessor allows you to get the most out of the machine, speed up production, increase productivity and thus profit.

REFERENCES:

Web Page

[3] https://www.cnccookbook.com/cnc-post-processor-cam/, [Accessed: 12. April 2020]

PDF Offprint

[4] Prima Electro SpA; Open Control Programming Manual, Document Number 45004607K, Release Ref. V3.1, Edition 04 February2013

[5] Prima Electro SpA; Open Control End User Manual, Document Number 4602H, Edition 00, June 2013.

Submitted: April 4, 2020
Accepted: June 8, 2020.

About the authors

Boris Paukovic was born in Sanski Most in 1988. He is currently a master candidate at the Pan-European University APEIRON at the College of Information Technologies.
From 2010 to 2012 he worked as an Information Technology Support Specialist at the company Computing SYSTEMS d.o.o in Banjaluka.
Since 2013 he has lived and worked in Leipzig, Germany and gained experience in the usage of information technologies in CNC Machinery and the administration of modern software solutions for architects.

Dražen M. Marinković was born in Banja Luka, Republika Srpska - Bosnia and Herzegovina. He finished elementary school in Prnjavor. He finished secondary school of Electrical Engineering in Prnjavor. He graduated at the Faculty of Information Technologies at Pan-European University "APEIRON" in Banja Luka. Master studies completed in 2015. Doctoral studies of the third degree enrolled in the academic 2015/2016. He lives and works in Prnjavor.

For citation

Boris Pauković, Dražen M. Marinković, The Role of Postprocessor in the Translation of Vector Graphics Understandable to the CNC Machine Controller, *JITA – Journal of Information Technology and Applications Banja Luka*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 10(2020) 1:58-65, (UDC: 621.9-52:004.388, 004.382:621.01), (DOI: 10.7251/JIT2001058P), Volume 10, Number 1, Banja Luka, June 2020 (1-68), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Instructions for Authors

The *Journal of Information Technology and Application (JITA)* publishes quality, original papers that contribute to the methodology of IT research as well as good examples of practical applications.

Authors are advised that adherence to the Instructions to Authors will help speed up the refereeing and production stages for most papers.

- Language and presentation
- Length of submissions
- Submission
- Contact details/biographies
- Title of the paper
- Abstract and keywords
- Figures and tables
- Sections
- Footnotes
- Special characters
- Spelling
- References
- Proofs
- PDF offprint
- Copyright and permissions
- Final material
- Correspondence
- Publication ethics

Language and presentation

Manuscripts should be written in English. All authors should obtain assistance in the editing of their papers for correct spelling and use of English grammar. Manuscripts should have double spacing, with ample margins and pages should be numbered consecutively. The Editors reserve the right to make changes that may clarify or condense papers where this is considered desirable.

Length of submissions

Papers should not normally exceed 12 Journal pages (about 8000 words). However, in certain circumstances (e.g., review papers) longer papers will be published.

Submission

Manuscripts must be submitted through the JITA online submission system.

Please read the instructions carefully before submitting your manuscript and ensure the main article files do not contain any author identifiable information.

Although PDF is acceptable for initial submission original source (i.e. MS Word) files will be required for typesetting etc.

Contact details/biographies

A separate file containing the names and addresses of the authors, and the name and full contact details (full postal address, telephone and e-mail) of the author to whom correspondence is to be directed should be uploaded at the time of submission (you should select Contact details/Biographies as the file type). This file is not shown to reviewers. This file should also contain short biographies for each author (75 words maximum each) which will appear at the end of their paper.

The authors' names and addresses must not appear in the body of the manuscript, to preserve anonymity. Manuscripts containing author details of any kind will be returned for correction.

Title of the paper

The title of the paper should not be longer than 16 words.

Abstract and keywords

The first page of the manuscript should contain a summary of not more than 200 words. This should be self-contained and understandable by the general reader outside the context of the full paper. You should also add 3 to 6 keywords.

Figures and tables

Figures which contain only textual rather than diagrammatic information should be designated Tables. Figures and tables should be numbered consecutively as they appear in the text. All figures and tables should have a caption.

Sections

Sections and subsections should be clearly differentiated but should not be numbered.

Footnotes

Papers must be written without the use of footnotes.

Special characters

Mathematical expressions and Greek or other symbols should be written clearly with ample spacing. Any unusual characters should be indicated on a separate sheet.

Spelling

Spelling must be consistent with the Concise Oxford Dictionary.

References

References in the text are indicated by the number in square brackets. If a referenced paper has three or more authors the reference should always appear as the first author followed by et al. References are listed alphabetically. All document types, both printed and electronic, are in the same list. References to the same author are listed chronologically, with the oldest on top. Journal titles should not be abbreviated.

- Journal

[1] Avramović Z.Ž. (1995). Method for evaluating the strength of retarding steps on a marshalling yard hump. *European Journal of Operational Research*, 85(1), 504–514.

- Book

[2] Walsham G. (1993). *Interpreting Information Systems in Organizations*. Wiley, Chichester.

- Contributed volume

[3] Huberman A.M. and Miles M.B. (1994). Data Management and analysis methods. In *Handbook of Qualitative Research* (Denzin NK and Linkoln YS, Eds), pp 428-444, Sage, Thousand Oaks, California.

- Conference Paper

[4] Баранов Л.А. (2017). Принципы построения и алгоритмы интеллектуальных автоматических систем управления движения поездов, функционируемых в рамках городских транспортных систем. In G. Radić & Z.Ž. Avramović (Eds.), *Proceedings of a IX International Scientific Conference "Information Technology for e–Education"*, (pp.9-18). Pan-European University APEIRON, Banjaluka, 29–30.9.2017. Republic of Serpska, B&H

- Unpublished reports/theses

[5] Nandhakumar J.J. (1993). The practice of executive information systems development: and in-depth case study. *PhD Thesis*, Department of Engineering, University of Cambridge.

Proofs

Proofs of papers will be sent to authors for checking. Alterations to diagrams should be avoided where possible. It will not be possible to accept major textual changes at this stage. Proofs must be returned to the publishers within 48 hours of receipt by fax, first-class post, airmail or courier. Failure to return the proof will result in the paper being delayed.

PDF offprint

Corresponding authors will receive a PDF of their article. This PDF offprint is provided for personal use. It is the responsibility of the corresponding author to pass the PDF offprint onto co-authors (if relevant) and ensure that they are aware of the conditions pertaining to its use.

The PDF must not be placed on a publicly-available website for general viewing, or otherwise distributed without seeking our permission, as this would contravene our copyright policy and potentially damage the journal's circulation. Please visit http://www.jita-au.com to see our latest copyright policy.

Copyright and permissions

The copyright of all material published in the Journal is held by Pan-European University APEIRON. The author must complete and return the copyright form enclosed with the proofs.

Authors may submit papers which have been published elsewhere in a foreign language, provided permission has been obtained from the original publisher before submission.

Authors wishing to use material previously published in JITA should consult the publisher.

Final material

All final material must be submitted electronically in its original application format (MS Word is preferred). The file must correspond exactly to the final version of the manuscript.

Correspondence

Business correspondence and enquiries relating to advertising, subscriptions, back numbers or reprints should be addressed to the relevant person at:

Pan-European University APEIRON
Journal JITA
Pere Krece 13, P.O.Box 51
78102 Banja Luka
Bisnia and Hercegovina / RS
E-mail: jita@apeiron-edu.eu

Publication ethics

We take an active interest in issues and developments relating to publication ethics, such as plagiarism, falsification of data, fabrication of results and other areas of ethical misconduct. Please note that submitted manuscripts may be subject to checks using the corresponding service, in order to detect instances of overlapping and similar text.

JITA

Journal of Information Technology and Applications
Pan-European University APEIRON Banja Luka
www.jita-au.com
jita@apeiron-edu.eu

PUBLISHER: **Pan-European University APEIRON**, Banja Luka
College of Information Technology Banja Luka, Republic of Srpska, B&H
www.apeiron-uni.eu
Darko Uremović, Person Responsible for the Publisher
Aleksandra Vidović, PhD, Editor of University Publications

ISSN 2232-9625