An Intelligent Method for Forming the Advertising Content of Higher Education Institutions Based on Semantic Analysis

Khrystyna Lipianina-Honcharenko [1], Taras Lendiuk [1], Anatoliy Sachenko [1], Oleksandr Osolinskyi [1], Diana Zahorodnia [1], Myroslav Komar [1]

[1] *West Ukrainian National University, Lvivska Str., 11, Ternopil, 46000, Ukraine*

Abstract

Advertising is a unique socio-cultural phenomenon: its formation is due to social, psychological, linguistic factors, features of the “aesthetic consciousness” of society and its cultural traditions. Advertising text – a special kind of text, it is a carrier and expression of information. It is important that the text is interesting to the desired audience, and it can be formed by methods of text semantic analysis and highlight the keywords on the basis of which advertising content can be generated. In this regard, it is developed an intelligent method of advertising content forming of higher education institutions on the semantic analysis basis and thus advertising manager can generate advertising content. The implementation of the method was carried out on the basis of a survey of “Computer Science” students regarding admission. Semantic analysis of documents based on LSA and LDA-method is performed. The results show that more than six keywords are present in document 0, based on the LSA method – 66%. Based on the LDA method, the vast majority of keywords are presented in document 2 – 82%. Based on the obtained keywords, the LSA and LDA methods created content for advertising of higher education institutions. The effectiveness of the generated advertising content on the basis of LSA and LDA-method was compared, a comparative experiment was conducted on Facebook on the business page “Computer Science of ZUNU”. By effectiveness comparing results of generated advertising content, the effectiveness of the ad increased by 44% and the price for the result decreased by 31%.

Keywords

Data analysis, advertising content, semantic analysis, higher education institutions, Facebook.

1. Introduction

Today, in the conditions of unstable development of the country’s economy, as well as in accordance with the new legislation in the field of higher education, the market of educational services is increasingly hard competition for potential consumers – entrants. Under these circumstances, there is an urgent need of higher education institutions to use effective communication with the market and target audiences. Such communications come in various forms, the most common of which is advertising.

In order to effectively promote advertising and educational services, higher educational institutions must use modern information and communication technologies, combining them into a well-built and strategically designed system of actions.

One such tool is the effective formation of advertising text. Advertising text has a certain structure, which consists of two main components – slogan (introductory part, title) and code (main part and conclusions), which formally divide the advertising text at the beginning and end. Advertising texts

ICTERI-2021, Vol II: Workshops, September 28 – October 2, 2021, Kherson, Ukraine
EMAIL: xrustya.com@gmail.com (A. 1); tl@wunu.edu.ua (A. 2); as@wunu.edu.ua (A. 3); oso@wunu.edu.ua (A. 4); dza@wunu.edu.ua (A. 5); mko@wunu.edu.ua (A. 6)
ORCID: 0000-0002-2441-6292 (A. 1); 0000-0001-9484-8333 (A. 2); 0000-0002-0907-3682 (A. 3); 0000-0002-0136-395X (A. 4); 0000-0002-9764-3672 (A. 5); 0000-0001-6541-0359 (A. 6)
© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

aimed at young people (entrants) can combine the functional characteristics of the product, which are complemented by a star image and together model the image of an indispensable thing for teenagers.

It is important that the text was interesting for the right audience, for this one it is needed to know the opinion of the target audience. For higher education institutions, these are entrants, but also students, who can provide information on why they chose this particular educational institution. The formation of opinion from the survey takes a long time for sociologists, but the methods of text semantic analysis can form the text main idea faster and identify keywords on the basis of which advertising content can be formed.

In this regard, we can assume that the development of an intelligent method of advertising content generating of higher education institutions based on semantic analysis is relevant and allows to increase the advertising effectiveness, and thus reduce the cost of online advertising of higher education institutions.

This paper is devoted to this topic, and it is distributed as follows. Section 2 discusses the analysis of related works. Section 3 presents an intelligent method of advertising content forming of higher education institutions based on semantic analysis. Section 4 presents the implementation of the method. Section 5 presents the conclusions of the study.

2. Related Work

Internet advertising has a large number of sites for advertising [8], namely search engines [18], social networks [19] and media advertising [20]. Given the complexity of the modern digital advertising ecosystem, there are many studies describing the impact of advertising content in social networks on customer engagement [11], using data from Facebook in: medicine [12], psychology [13], sociology [14], politics [15], food industry [16], tourism [17]. Research [9] provides an understanding of the current landscape of social networks of higher education institutions.

Paper [2] proposes a recommendations ontological system for advertising, which uses data obtained by users on social networking sites, and this approach is based on a common ontology model, which can be used to present both user profiles and advertising content. Both users and advertisements are represented by vectors created using natural language processing methods that collect ontological entities from textual content. The system of recommendations for advertising is widely tested in a simulated environment, MAP@3 and is equal to 85.6%.

Work [21] proposed advanced semantics and contextual hybrid co-filtering for event recommendations, which combines semantic content analysis and the impact of a contextual event on user neighborhood choices.

In [10] the intelligent system of advertising management in social networks is presented, on the basis of data analysis methods for automatic announcements creation. In [1], for the classification of advertising video, a video presentation is proposed, which aims to cover the hidden semantics of unattended advertising video, experiments on real advertising videos demonstrate that the proposed method can effectively differentiate advertising videos. Paper [3] proposes an approach of trust and semantic social recommendation to eliminate the advertising problems. A method [4] is proposed to help convert unreadable data into a readable structured format using machine learning classification, and clustering plays a crucial role in converting operational data into a data model and visualizing the processed information for the end user. Because organizations have specific requirements when considering them, a latent Dirichlet Distribution (LDA) and a latent semantic analysis (LSA) have been implemented that have been able to process discrete data. Document [5] uses the LDA approach to determine feedback for analyzing online shopping sentiment, as the LDA approach is designed to address LSA and PLSA issues. In order to increase the efficiency of public moods analysis in the Internet, a method of text mood analysis [6] is proposed, which combines the presentation of the text of the hidden Dirichlet distribution (LDA) and the convolutional neural network (CNN). In [7], an empirical assessment of the three most important methods of modeling topics is presented – hidden semantic analysis, hidden Dirichlet distribution and correlated topics modeling.

It should be noted that the above-mentioned works on the one hand mostly analyze the responses of users on Internet advertising. And on the other hand, a number of works – representing the study of a text semantic analysis, there are also works that use semantic analysis for online advertising (analogues).

In this regard, the purpose of this paper is to develop an intelligent method of advertising content forming of higher education institutions based on semantic analysis.

Unlike analogues [2, 21], an intelligent method of forming the advertising content of higher education institutions on semantic analysis basis has been developed, it will allow to select keywords on the basis of which it is possible to form an advertising text.

The novelty of the work is formation of the most profitable (in economic aspect) text of HEI advertising, which will reduce the advertising campaign cost.

3. Method

To reduce the time spent on choosing a popular new product, the authors have developed an intelligent method of advertising content forming of higher education institutions based on semantic analysis. The proposed method is illustrated schematically (Fig.1) and is represented by the following steps:

Step 1. Student's survey conducting (Block 1) and its conversion to .csv format (Block 2).

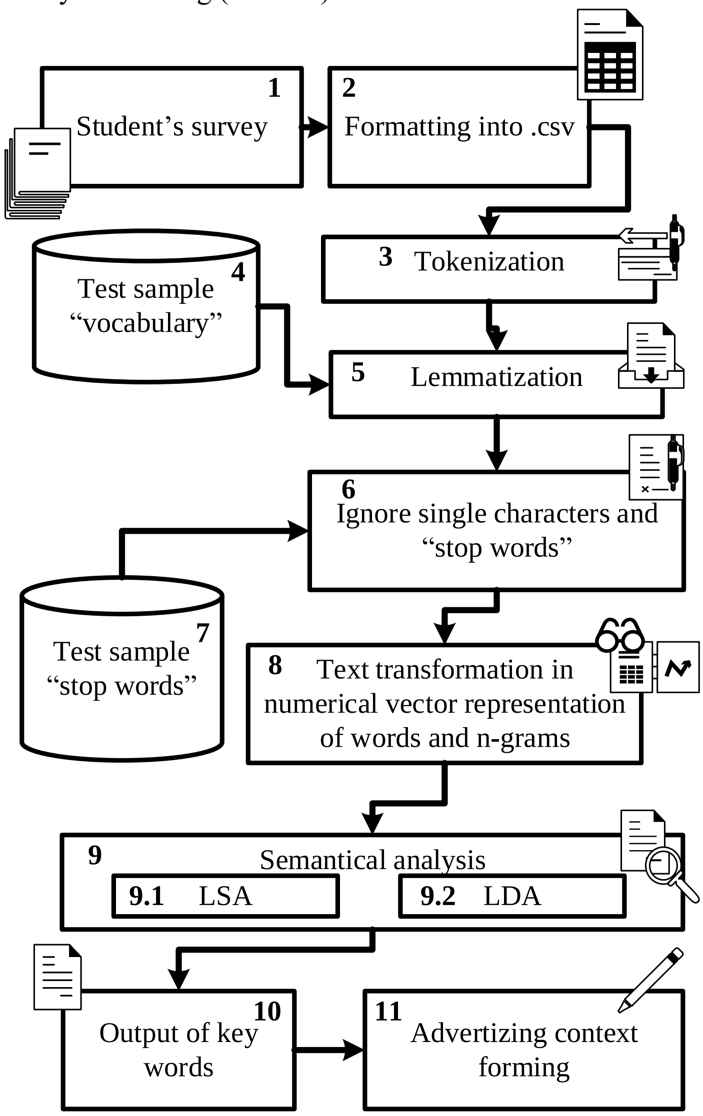


Figure 1: The structure of the intelligent method of forming the advertising content of higher education institutions on the basis of semantic analysis

Step 2. Tokenization of the document (Block 3). Tokenization is a process by which a large amount of text is divided into smaller parts, so called tokens. These tokens are very useful for finding patterns and are considered as a basic step for justification and lemmatization. Tokenization also helps to replace sensitive data elements with insensitive data elements.

Step 3. Lemmatization of the document (Block 5). Lemmatization is the process of transforming a word into its basic form (Block 4). This approach is highly dependent on the exact definition of the lexical category (part of speech). Although there is a partial overlap between the normalization rules for some lexical categories, the indication of an incorrect category or the inability to determine the correct category reduces to comment the advantage of this approach over the algorithm of endings truncation.

Step 4. The process of ignoring single characters (Block 6) (".", ";", ":", "!", """, "?", ",", "'"", "()", "[]"), which interfere with the program on the base on a ready-made set of "stop words" Stop words are very common words such as "if", "but", "we", "he", "she" and "they" (Block 7). Usually, it is possible to delete these words without changing the text semantics (but not always), improving the model performance.

Step 5. Document vectorization (Block 8). Convert a collection of raw documents into a matrix of TF-IDF functions. TF (term frequency) – IDF (inverse document frequency) – a statistical indicator used to assess the words importance in a document context that is part of a collection of documents or corpus. The weight (significance) of a word is proportional to the number of uses of that word in the document, and inversely proportional to the frequency of use of the word in other documents of the collection.

$$TF\text{-}IDF = TF \times IDF, \tag{1}$$

where, $TF = \frac{n_i}{\sum_k n_k}$, where n_i – is the number of the word occurrences in the document, and in the denominator – the total number of words in the document.

$$IDF = \log \frac{|D|}{|(d_i \supset t_i)|},$$

where, $|D|$ – number of documents in the collection; $|(d_i \supset t_i)|$ – the number of documents in which the word t_i occurs (when $n_i \neq 0$).

Step 6. Semantic analysis (Block 9).

Step 6.1. Reducing the data dimensionality (Block 9.1). Dimension reduction is the process of reducing the number of random variables by obtaining a set of main variables. For semantic analysis (LSA), the dimension reduction method is the method of single value decomposition (SVD), which calculates only the *k* largest unit values, where *k* is a user-defined parameter.

Topics modeling in the text based on the LSA method. When SVD is applied to term document matrices, this transformation is known as latent semantic analysis (LSA) because it converts such matrices into a low-dimensional "semantic" space. In particular, it is known that LSA fights the effects of synonymy and polysemy (both of which roughly mean that the word has several meanings), which causes excessive matrices sparseness of urgent documents and shows low similarity in such indicators as the cosine's similarity.

According to the singular decomposition theorem, any material rectangular matrix can be decomposed into a product of three matrices:

$$\mathrm{A{=}USV^T}, \tag{2}$$

where, the matrices *U* and *V* are orthogonal, and *S* is a diagonal matrix, the values on the diagonal of which are called singular values of the matrix *A*. The letter T in the expression *VT* means the matrix transposition. This decomposition has a remarkable feature: if in the matrix *S* leave only *k* largest singular values, and in the matrices *U* and *V* – only the columns corresponding to these values, then the product obtained matrices *S*, *U* and *V* will be the best approximation of the initial matrix *A* to the matrix $\hat{A}$ rank *k*:

$$\widehat{\mathrm{A}} \approx \mathrm{A}=\mathrm{USV}^{\mathrm{T}}. \tag{3}$$

The main idea of latent semantic analysis is that if the term-on-documents matrix was used as the matrix *A*, then the matrix $\hat{A}$, containing only *k* of the first linearly independent components *A*, reflects the basic structure of various dependencies present in the original matrix. The structure of dependencies is determined by the weight functions of the terms.

Step 6.2. Topics modeling of in the text based on the LDA method (Block 9.2). Latent Dirichlet Allocation (LDA) is a method of matrices decomposing. In the vector space, any corpus (documents collection) can be represented as a matrix of document terms. The following matrix (4) shows the body of *N* documents D_1, D_2, D_3… D_n and the size of the dictionary of *M* words W_1, W_2 … W_n. The value of cell *i*, *j* gives calculation of the word frequency W_j in the document D_i.

$$\begin{pmatrix} D_1W_1 & D_1W_2 & \dots & D_1W_n \\ D_2W_1 & D_2W_2 & \dots & D_2W_n \\ \dots & \dots & \dots & \dots \\ D_nW_1 & D_nW_2 & \dots & D_nW_n \end{pmatrix} \tag{4}$$

The LDA converts (4) the document matrix term into two matrices of lower size – M_1 and M_2.

M_1 is a matrix of document topics (5), and M_2 is a matrix of topics (6) – terms with dimensions (*N*, *K*) and (*K*, *M*), respectively, where *N* is the documents number, *K* is the topics number, and *M* is vocabulary size.

$$M_1 = \begin{pmatrix} D_1K_1 & D_1K_2 & \dots & D_1K_n \\ D_2K_1 & D_2K_2 & \dots & D_2K_n \\ \dots & \dots & \dots & \dots \\ D_nK_1 & D_nK_2 & \dots & D_nK_n \end{pmatrix} \tag{5}$$

$$M_2 = \begin{pmatrix} K_1W_1 & K_1W_2 & \dots & K_1W_n \\ K_2W_1 & K_2W_2 & \dots & K_2W_n \\ \dots & \dots & \dots & \dots \\ K_nW_1 & K_nW_2 & \dots & K_nW_n \end{pmatrix} \tag{6}$$

These matrices (5, 6) already provide topics distribution by topics and words, however, this distribution needs improvement, which is the main goal of the LDA. The LDA uses sampling methods to improve these matrices. After a series of iterations, a stable state is reached, when the distribution of the document topic and the topic terms is quite good. This is the point of LDA convergence.

Step 7. Output of keywords (Block 10) on documents based on LSA and LDA method.

Step 8. Advertising context forming (Block 11) of higher education institutions.

For a better understanding of the developed method, it is necessary to conduct an experimental study.

4. Experimental Results and Discussion

Python language was chosen for semantic analysis to conduct an intelligent method of advertising content forming of higher education institutions. The following libraries were used: pandas, numpy, matplotlib, nltk, stop_words.

As input, we used students survey of majoring in Computer Science, regarding admission. 152 students took part in the survey and answered 10 questions. All student feedback is formed in .csv format, answers to each questionnaire on a separate line of text, then defined as a document.

The document was lemmatized using the WordNetLemmatizer function. Using the nltk.tokenize.word_tokenize () method, markers were extracted from the character string using the tokenize.word_tokenize () method.

Next, the data was cleaned for the presence of characters that do not affect the text content: “.”, “;”, “:”, “!”, “'”, “?”, “,”, “"”, “()”, “[]”. The next step is to remove the Ukrainian “stop words” (Fig. 2).

0	з соцмереж самостійно знайшов інформацію зац...
1	до нас в школу представники спеціальності приї...
2	самостійно знайшов інформацію мені зателефону...
3	з соцмереж пройшов тут на державне навчання
4	порадили знайомі родичі бачив багато цікавої...

Figure 2: Lemmatized and tokenized responses of students without “stop words” and symbols

Using the TfidfVectorizer function, which converts a collection of raw documents into a matrix of TF-IDF functions (Fig. 3).

```
(0, 21)      0.249923
(0, 5)       0.236487
(0, 40)      0.249923
(0, 52)      0.249923
(0, 48)      0.249923
(0, 0)       0.249923
(0, 32)      0.264944
(0, 42)      0.249923
```

Figure 3: Part of the TF-IDF matrix

In the next step (Fig. 4) we reduce the dimension using truncated SVD (aka LSA). This transformer reduces the linear dimension using truncated special value decomposition (SVD). Unlike PCA, this estimator does not center data before calculating a special value decomposition. This means that it can work efficiently with sparse matrices. In particular, the truncated SVD works on the TF-IDF matrices returned by the vectorizers. In this context, it is known as laten semantic analysis (LSA).

Using the TruncatedSVD function, the following parameters are defined: the desired dimension of the source data (documents) – n_components = 5; SVD method algorithm – algorithm = 'randomized'; number of iterations for randomized SVD – n_iter = 10; the number of reproducible results after several calls to the functions -random_state = 100.

```
[[ 6.96660197e-01 -5.73572092e-02 -4.06682974e-01 ... -2.00406853e-18
  -7.54197595e-18  1.32222428e-17]
 [ 7.83147373e-02  2.58859227e-01 -1.10413796e-01 ... -6.93889390e-18
  -5.20417043e-18 -1.38777878e-17]
 [ 3.18620187e-01  1.45234702e-01 -6.98410495e-02 ... -7.26415456e-18
   2.25514052e-17  1.73472348e-18]
```

Figure 4: Part of the matrix of the LSA method

Next, the text was analyzed based on the LDA method (Fig. 5). Using the LatentDirichletAllocation function, the following parameters are defined: the desired dimension of the original data (documents) – n_components = 5; method used to update _component, if the amount of data is large, online update will be much faster than batch update – learning_method = 'online'; the maximum number of iterations is max_iter = 1.

```
[[0.04455599 0.8221565  0.04445858 0.04441181 0.04441713]
 [0.7910456  0.05214722 0.05216473 0.05240864 0.05223381]
 [0.0572413  0.05729839 0.77093604 0.05725154 0.05727273]
 [0.07148326 0.07139567 0.71293007 0.07197358 0.07221742]
 [0.0459054  0.81552225 0.04597863 0.04597552 0.04661821]
 [0.04668559 0.04725528 0.04661619 0.04676363 0.81267932]
```

Figure 5: Part of the matrix of the LDA method

LSA

```
Document 0 :
Topic  0  :  65.75057550474627 %
Topic  1  :  40.182712167943826 %
Topic  2  :  23.376891919092163 %
Topic  3  :  -21.590881614430092 %
Topic  4  :  -28.60402955510097 %
```

LDA

```
Document 0:
Topic  0 :  4.594951994120869 %
Topic  1 :  4.59272498367418 %
Topic  2 :  81.48129348616614 %
Topic  3 :  4.7410176809086515 %
Topic  4 :  4.590011855130164 %
```

Figure 6: The probability of the keyword's presence in individual documents by LSA and LDA-method

Figure 6 shows that most keywords are present in document 0, based on the LSA method – 66%. Based on the LDA method, the vast majority of keywords are presented in document 2 – 82%.

Now it is needed to get a list of important words for each of the 5 documents (Fig. 7) based on the LSA and LDA method. Let's choose 5 words for each topic for simplicity.

Based on the obtained keywords by LSA and LDA methods (see Fig. 6 and 7), it is possible to create the following content for advertising higher education:

"Get a cool public training. More information in messenger"

To compare the effectiveness of the generated advertising content based on the LSA and LDA-method, a comparative experiment was conducted on Facebook on the business page "ZUNU Computer Science" (Fig. 8). The first version of advertising (Fig. 8a), developed on the basis of the rules identified in previous research, namely:

- the greatest interaction with the video Facebook advertising "WUNU Computer Science" had male in the age category 18-25, 35-55 [22];
- The largest interaction with the business Facebook page "WUNU Computer Science" had male and female clients in the age category 40-55 [23].

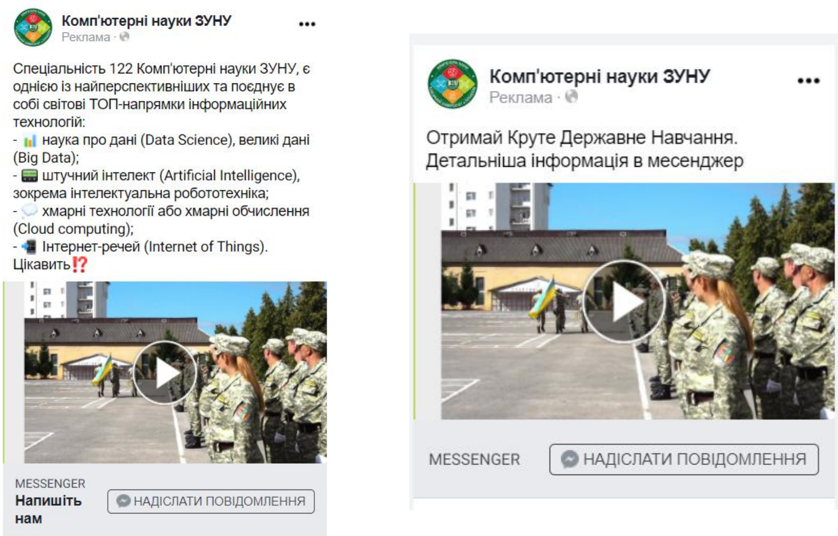
LSA

```
Topic 0:
навчання державне пройшов інформацію круто

Topic 1:
круто соцмережах цікавої інформації бачив

Topic 2:
спеціальності зацікавило представлення представниками представники

Topic 3:
вчитись зателефонували переконали цікаво інформацію

Topic 4:
знайомі порадили родичі влаштувала вартість
```

```
Topic 0:
спеціальності представлення представниками зацікавило інформацію круто

Topic 1:
представлення інформацію спеціальності зацікавило карантином представниками

Topic 2:
інформації круто соцмережах бачив цікавої вирішив

Topic 3:
державне пройшов навчання спеціальності порадили родичі

Topic 4:
знайомих навчається тернополі сфер спеціальність подобається
```

L D A

Figure 7: Keywords for the advertising content forming in individual documents by LSA and LDA-method



a) Previous version b) formed using LSA and LDA-method

Figure 8: “Computer Science” advertising of Western Ukrainian National University on Facebook

Table 1 presents effectiveness comparison of the generated advertising content based on the LSA and LDA method. In the period from Jan 4, 2021 – Jan 31, 2021 was conducted an advertising campaign with a lot of text content, see Fig. 8a. In the period from Feb 1, 2021 – Feb 28, 2021 an advertising campaign was conducted with text content developed on the basis of LSA and LDA-method, see Fig. 8b.

Table 1
Comparison of the effectiveness of the generated advertising content

Indicator	Jan 4, 2021 – Jan 31, 2021	Feb 1, 2021 – Feb 28, 2021	Change
Coverage	6792	7350	558
Frequency	1,18	1,19	0,02
Results	9,00	13,00	4,00
Price for the result	1,42	0,98	-0,44
Amount of expenses (USD)	12,77	12,75	-0,02
CPM (Cost per 1000 views)	1,60	1,46	-0,14
Link clicks	46,00	43,00	-3,00
CPC (cost-per-click)	0,28	0,30	0,02
CTR (clickability)	0,58	0,49	-0,09

Coverage increased by 558 people (see Table 1). The average number of times each user saw your ad (Frequency) increased by 0.02 per person. The number of times the ad reached its goal, goal setting, and settings (Results) increased by 4 instant messaging. This is really not much, but taking to account small budget of the advertising campaign and the target audience specifics, the result is quite good. Accordingly, the price for the result decreased by 0.44 cents, which in the amount taking into account the increase in coverage, by 0.02 cents less. Accordingly, the average cost per 1000 impressions (CPM) decreased by 0.14 cents. The number of clicks on links in advertisements that resulted in landing pages selected by the administrator, within or outside Facebook, decreased by 3. This resulted an increase in the cost-per-click (CPC) of the link, which is not critical, respectively, the proportion of people who viewed ad and click the link (CTR) decreased by 0.09. In percentage, the comparison of performance indicators of the generated advertising content is presented in Fig. 9.

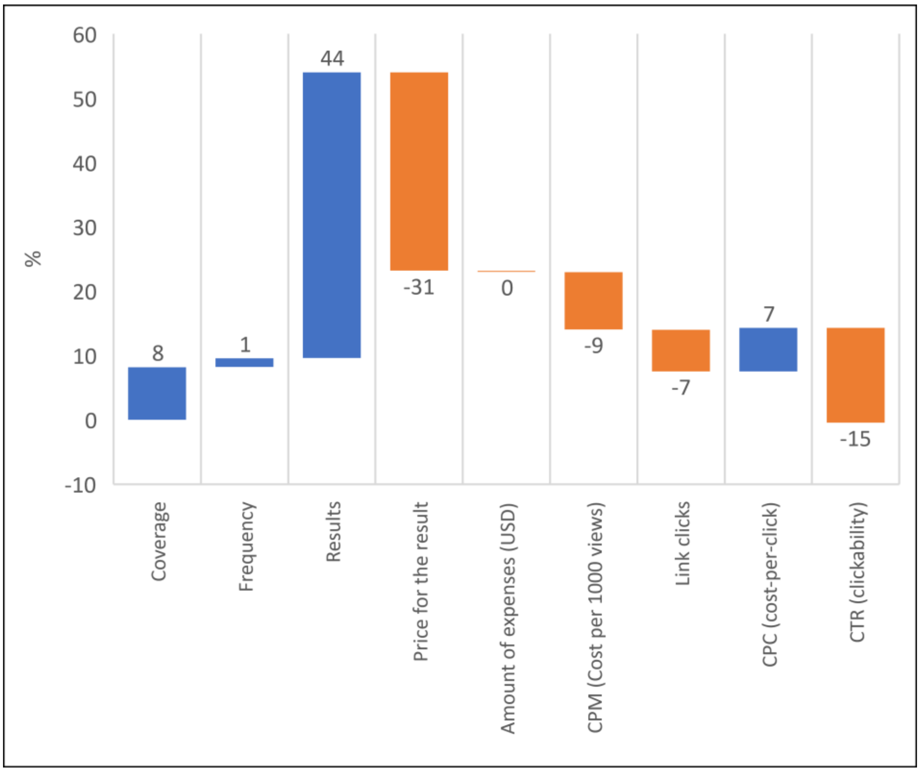


Figure 9: Change (%) of the advertising campaign effectiveness for the periods Jan 4, 2021 –Jan 31, 2021 and Feb 1, 2021 – Feb 28, 2021

From the results of comparing the effectiveness of the generated advertising content (see table 1 and Fig. 9), we can conclude that the effectiveness of the ad increased by 44% and the price for the result decreased by 31%.

Thus, the intelligent method of the advertising content forming of higher education institutions on the basis of semantic analysis, allows to increase the advertising effectiveness on social networks by at least 44%. The authors believe that extension of student surveys number will increase the semantic analysis quality, respectively, get the best keywords for the content forming, and thus increase the advertising effectiveness and reduce its costs.

In contrast to analogues [2, 21], the developed intelligent method of advertising content forming of higher education institutions on the basis of semantic analysis will allow to select keywords on the basis of which it is possible to form advertising text and ad effectiveness increased by 44% and at the same time the price for the result decreased by 31%.

5. Conclusions

An intelligent method of the advertising content forming of higher education institutions based on semantic analysis has been developed, on the basis of which the advertising manager can form advertising content, and, thus, the effectiveness of the announcement has increased. Also, the proposed method reduces the time spent on assessing student questionnaires and determining the main idea regarding admission.

The implementation of the method is based on students' survey of majoring in "Computer Science" regarding admission. 152 students took part in the survey. Cleaning of unnecessary symbols and "stop words" was conducted. After tokenization, vectorization and lemmatization of documents were conducted. Documents semantic analysis based on LSA and LDA method was performed. Most keywords present in document 0, based on the LSA method was 66%. Based on the LDA method, the vast majority of keywords presented in document 2 was 82%. Based on the obtained keywords, LSA and LDA methods created content for advertising of higher education institutions: "Get a cool public education. More information is in the messenger. The effectiveness of the generated advertising content on the basis of LSA and LDA-method was compared, a comparative experiment was conducted on Facebook on the "WUNU Computer Science" business page. In the period from Jan 4, 2021 to Jan 31, 2021, an advertising campaign with a lot of textual content was conducted. In the period from Feb 1, 2021 to Feb 28, 2021, an advertising campaign was conducted with text content developed on the basis of LSA and LDA-method. From the results of comparing the effectiveness of the generated advertising content, the effectiveness of the ad increased by 44% and the price for the result decreased by 31%.

In contrast to analogues [2, 21], an intelligent method of forming the advertising content of higher education institutions based on semantic analysis has been developed, what allowed to select keywords on the basis of which it is possible to form an advertising text.

6. References

[1] S. Hou, S. Zhou, W. Liu, et al., Classifying advertising video by topicalizing high-level semantic concepts, Multimed Tools Appl, Springer 77 (2018) 25475–25511. https://doi.org/10.1007/s11042-018-5801-3.

[2] F. García-Sánchez, R. Colomo-Palacios, R. Valencia-García, A social-semantic recommender system for advertisements, Information Processing & Management, Elsevier 57 (2020) 102153. doi: 10.1016/j.ipm.2019.102153.

[3] J. Shokeen, C. Rana, A trust and semantic based approach for social recommendation, J Ambient Intell Human Comput, Springer (2021). Doi: 10.1007/s12652-020-02806-1.

[4] Y. Kalepalli, S. Tasneem, T. Phani, D. Pasupuleti, S. Manne, Effective comparison of LDA with LSA for topic modelling, in: Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 13-15 May 2020, pp. 1245–1250. doi: 10.1109/ICICCS48265.2020.9120888.

[5] Parveen, N., Santhi, M. V. B. T., Burra, L. R., Pellakuri, V., & Pellakuri, H.: Women's e-commerce clothing sentiment analysis by probabilistic model LDA using R-SPARK. Materials Today: Proceedings. Elsevier (2021). https://doi.org/10.1016/j.matpr.2020.10.064.

[6] L.-X. Luo, Network text sentiment analysis method combining LDA text representation and GRU-CNN, Pers Ubiquit Comput, Springer 23 (2019) 405–412. https://doi.org/10.1007/s00779-018-1183-9.

[7] P. Kherwa, P. Bansal, A comparative empirical evaluation of topic modeling techniques, in: Gupta D., Khanna A., Bhattacharyya S., Hassanien A., Anand S., Jaiswal A. (eds). International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, Springer, Singapore, 2021, vol. 1166, pp. 289-297. https://doi.org/10.1007/978-981-15-5148-2_26.

[8] B. Aslam, & H. Karjaluoto, Digital advertising around paid spaces, E-advertising industry's revenue engine: A review and research agenda, Telematics and Informatics, Elsevier 34 (2017) 1650-1662. https://doi.org/10.1016/j.tele.2017.07.011.

[9] A. Peruta, & A. B. Shields, Social media in higher education: understanding how colleges and universities use Facebook, Journal of Marketing for Higher Education 27 (2017) 131-143. Doi: 10.1080/08841241.2016.1212451.
[10] J. Aguilar, and G. Garcia, An adaptive intelligent management system of advertising for social networks: A case study of Facebook, IEEE Transactions on Computational Social Systems 5 (2018) 20-32. Doi: 10.1109/TCSS.2017.2759188.
[11] D. Lee, K. Hosanagar, and H. S. Nair, Advertising content and consumer engagement on social media: Evidence from Facebook, Management Science 64 (2018) 5105-5131. https://doi.org/10.1287/mnsc.2017.2902.
[12] A. M. Jamison, D. A. Broniatowski, M. Dredze, Z. Wood-Doughty, D. A. Khan, S. C. Quinn, Vaccine-related advertising in the Facebook Ad Archive, Vaccine, Elsevier 38 (2019) 512-520. doi: 10.1016/j.vaccine.2019.10.066.
[13] S. Youn, S. Kim, Understanding ad avoidance on Facebook: Antecedents and outcomes of psychological reactance, Computers in Human Behavior, Elsevier 98 (2019) 232-244. https://doi.org/10.1016/j.chb.2019.04.025.
[14] C. L. White, B. Boatwright, Social media ethics in the data economy: Issues of social responsibility for using Facebook for public relations, Public Relations Review, Elsevier 46 (2020) 101980. https://doi.org/10.1016/j.pubrev.2020.101980.
[15] A. Gitomer, P.V. Oleinikov, L.M. Baum, et al., Geographic impressions in Facebook political ads, Appl Netw Sci, Springer 6 (2021) 18. https://doi.org/10.1007/s41109-020-00350-7.
[16] A. Galati, M. Crescimanno, S. Tinervia, & F. Fagnani, Social media as a strategic marketing tool in the Sicilian wine industry: Evidence from Facebook, Wine Economics and Policy, Elsevier 6 (2017) 40-47. https://doi.org/10.1016/j.wep.2017.03.003.
[17] A. Sharma, S. Sharma, & M. Chaudhary, Are small travel agencies ready for digital marketing? Views of travel agency managers, Tourism Management, Elsevier 79 (2020) 104078. https://doi.org/10.1016/j.tourman.2020.104078.
[18] G. Szymanski, & P. Lininski, Model of the effectiveness of Google Adwords advertising activities, in: Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), September 2018, vol. 2, pp. 98-101. Doi: 10.1109/STC-CSIT.2018.8526633.
[19] J. Phua, J. S. E. Lin, & D. J. Lim, Understanding consumer engagement with celebrity-endorsed E-Cigarette advertising on Instagram, Computers in Human Behavior, Elsevier 84 (2018) 93-102. https://doi.org/10.1016/j.chb.2018.02.031.
[20] H. Gupta, S. Singh, & P. Sinha, Multimedia tool as a predictor for social media advertising – a YouTube way, Multimedia Tools and Applications, Springer 76 (2017) 18557-18568. https://doi.org/10.1007/s11042-016-4249-6.
[21] M. Xu, S. Liu, Semantic-enhanced and context-aware hybrid collaborative filtering for event recommendation in event-based social networks, IEEE Access 7 (2019) 17493–17502. doi: 10.1109/ACCESS.2019.2895824.
[22] H. Lipyanina, S. Sachenko, T. Lendyuk, A. Sachenko, Targeting model of HEI video marketing based on classification tree, in: Proceedings of the 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020, Kharkiv, Ukraine, 6-10 October 2020, CEUR Workshop Proceedings, vol. 2732, pp. 487-498. http://ceur-ws.org/Vol-2732/20200487.pdf.
[23] H. Lipyanina, A. Sachenko, T. Lendyuk, S. Nadvynychny, S. Grodskyi, Decision tree based targeting model of customer interaction with business page, in: Proceedings of the third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), April 27 – May 1, 2020. CEUR Workshop Proceedings, vol. 2608, pp. 1001–1012. http://ceur-ws.org/Vol-2608/paper75.pdf.