UDC 004.9

L. Chyrun[1], L. Chyrun[2], V. Vysotska[2]
Lviv Polytechnic National University
[1]Computer-Aided Design Department
[2]Information Systems and Networks Department

ONLINE NEWSPAPER CONTENT ANALYSIS BASED SEO TECHNOLOGIES

The conceptual model of the system has been developed and incoming and outgoing data has been defined. To analyze site content the following methods have been used: analysis of the visitor content, analysis of the required amount of content, graphical analysis of content.

Keywords: analysis, content analysis of contents of information resources, content-analysis, rating evaluation, content management system.

АНАЛІЗ КОНТЕНТУ ІНТЕРНЕТ-ГАЗЕТИ НА ОСНОВІ SEO-ТЕХНОЛОГІЇ

Розроблено концептуальну модель системи, визначено вхідні, вихідні дані; наведено їх опис та опис вимог до системи. Для аналізу змісту сайту використано такі методи: аналіз змісту відвідувачів, аналіз необхідної кількості контенту, графічний аналіз контенту.

Ключові слова: аналіз, контент-аналіз змісту інформаційних ресурсів, контент-аналіз, рейтингова оцінка, система управління контентом.

Introduction. The general formulation of the problem

IT development has solved a lot of problems, but has identified new tasks as well. One of these tasks is content analysis of textual information [1–30]. It is a quantitative and qualitative method for analyzing bulks of text for further meaningful interpretation of quantitative and qualitative indicators [1–30]. The Internet-technologies and related services evolution has provided users access to almost unlimited amounts of content [1–30]. Its credibility and timeliness, however, became a problem. An important characteristic of content is its adequacy, i.e. a certain level of conformity of the image created using the received information to the corresponding real-life object, process or phenomenon [1–30].Content analysis is introduced for ensuring timeliness, reliability and adequacy of content. This allows information systems (IS) to receive content in all of its activities. The result of content analysis of a text is used when determining its tone, duplication, spam detection as well as detecting new events to determine its thematic subjects flows [31–37].

Analysis of recent research and publications

Review of general concepts and aspects using the method of content analysis. Internet marketing (online-marketing) is a known technology of content management. Integration of the Internet, information management, Public Relation, service for working with customers and sales in different areas [1]. Internet-marketing uses all aspects and basic traditional marketing elements, combined with new methods of research and analysis with the help of modern technologies (Fig. 1) [1, 31–37]. Constant contact with users

is efficient by automatically tracking statistics, for analysis by using ROI (Return on Investment) and profitability of e-business (Rate of Return, ROR) and conversion information resource (conversion rate). Conversion or efficiency of visiting an information resource is the number of visitors of an information resource that completed it targeted actions (hidden / direct instructions advertisers, vendors, sponsors commercial content, i.e. purchase, registration, subscription, visit a specific page information resource, moving on Sponsored Links), the total number of visitors to the information resource. [1] Successful conversion is interpreted differently by vendors and authors (purchase performed by a consumer who became interested in the product by clicking on the appropriate banner), advertisers or content providers (actions expected from visitors of an information resource such as subscribing for a mailing listregistration of rating information resource subscription mailing list or software download).

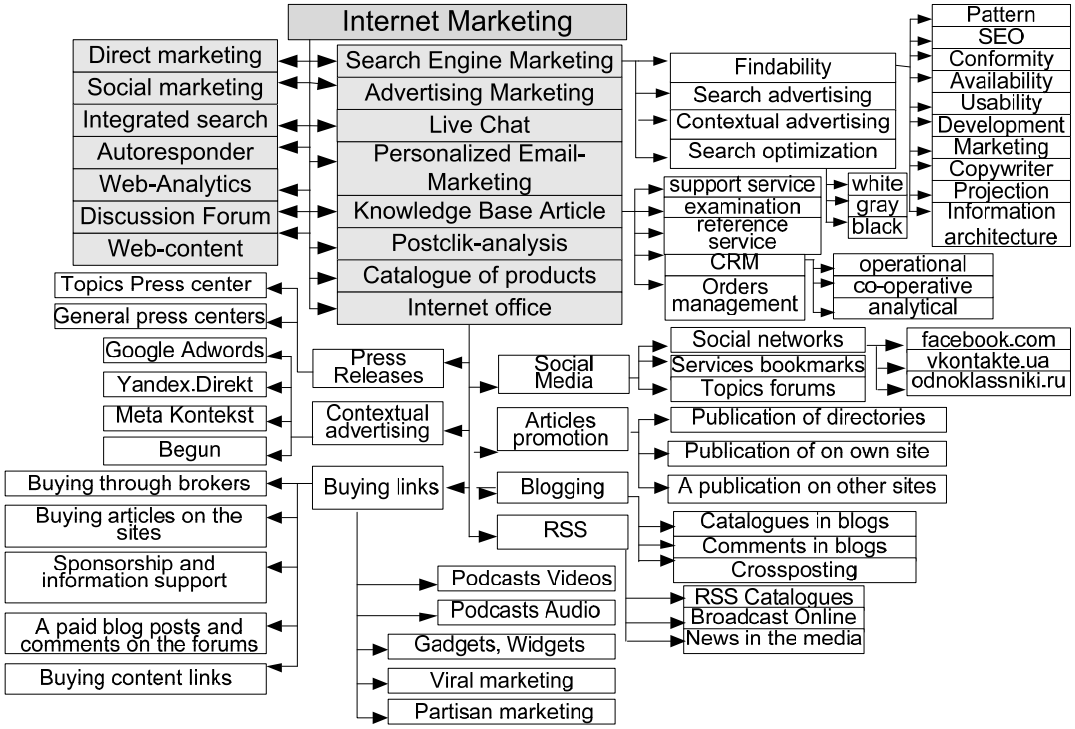

Fig. 1. Internet marketing directions [1]

The main advantages of the Internet marketing are interactivity, possibility of exact targeting, possibility of post-click analysis for maximizing performance of information resource conversion and Internet advertisement ROI / ROR [7]. The purpose of the application of Internet technology marketing is to get the maximum effect of the potential audience information resource with the ability to instantly obtain sales statistics, visits, demand, etc. (tab. 1) [1, 7, 9]

Table 1

Main advantages of the Internet and marketing (developed in [1, 7, 9])

Name	Definition
Interaction	The principle of the system is achieved in which the exchange of content between the elements of the system and the environment.
Search engine marketing	The process of increasing traffic from search engines, the lists of search results and advertisements.
Targeted advertising	Promotional mechanism to highlight the entire audience with targeted information resource that meets specified criteria to display her advertisement.
Post-click analysis	Method of Post-click marketing that maximizes efficiency and information resource conversion and ROI.

Search Engine Marketing (tab. 2) has the following characteristics [1, 7, 9]: work with specific queries (keywords); connections with the search (search engines, search information resource); increase the content findability level on an information resource; the context analysis (subject content, information resources, etc.). Usability (use ease) or the general convenience factor object of use; concept development of user interfaces Software security is focused on maximum psychological / visual experience for the user; coefficient of performance design menus and navigation system information resources; usability use, friendly interface and usability of software support. Search engine marketing may not reach advertising goals due to the following factors [1, 7, 9]: the advertised product is not always noted directly; the goods / services sales goal is not always pursued; increasing brand awareness is difficult; it is impossible to bring a new product / service to the market. The using search engine marketing results are [1, 7, 9] 1) attract users information resource, where each case is different audience, so attracted wide (increasing attendance rate information resource) or interested audience; 2) content resource distribution in search engines.

Table 2

Search engine marketing technologies (developed according to [1, 7, 9])

Technology	Appointment of search engine marketing technology
Search advertising	Dissemination of information in search engines by placing ads with keywords.
Search engine optimization (Search engine optimization)	Actions to change the state information resources (promotion, promotion) and elements of the environment to obtain high positions in search results.
Contextual advertising	Placing ads on relevant information resources.

The criteria for a well-chosen search engine marketing strategy is the number of visitors of the information resource and correlation between quality of the obtained and required audience. A simple criterion for checking popular information resource is the dynamics of the number of external links to information resources and increase mentions the name of the product / service or trade mark of the Internet. The limiting case of search and contextual advertising by placing ads in search results on thematic information resource. The development of the Internet contributed to the emergence of new technologies of search engine marketing for social networking (social marketing optimization) and video (video search marketing). Separation of search engine marketing as a separate independent strategy linked to the [1, 7, 9]: continued growth in Internet-markets; growing market of contextual / search advertising; use search engine optimization [1, 7, 9]; need optimal navigation (surfing) the content space, containing text/visual/ animated/audio content and/or experience of users of the system; support complex life cycle process content that it is in the management through various stages of publication [1, 7, 9]. For an effective analysis of the content appropriate techniques, technologies, tools, methods are used, followed by definition of objective conclusions [1]. Content Management System (CMS) is a software complex for building web sites or other IS [4]. Content is new information taken and evaluated by a person, new knowledge received by a user as a result of the perception and processing of certain information / facts or as a result of observation. The main content properties are reliability, efficiency and adequacy. Adequacy is an important characteristic of content. This is generated according image of the real world using the information received. There is a semantic, syntactic and pragmatic as well as the adequacy [3]. Pragmatic adequacy reflects compliance of the content to the content management purpose. Semantic adequate measures relevancy of the object image and of the object itself that is described by corresponding content. This form is used for the formation of notions and ideas relevant content and synthesis of relevant concepts. Syntax adequacy reflects the formal structural characteristics under content and does not affect its semantic content. This form contributes to the perception of external structural characteristics, i.e. syntactic sides respectively content. Each form has its own content adequacy retaliation of information and about volume of data. Pragmatic measure determines the value of content the user to achieve the goal.

Thesaurus is a collection of specific information, etc. at any subordinate that user as well. Syntax measure content operates with a certain type of information does not represent the semantic relation to the object of the object or phenomenon or process [3].

Rules for content quality design and evaluation. Content ought to be of interest for the user, to draw their attention and hold them for the largest amount of time. For this content must be a high degree of uniqueness for all common search engines. Content should be optimized for the most common search engines – add the perfect amount of relevant keywords and terms, expressions and phrases. Content ought en be so placed second on the page to users in was it is convenient to operate [2]. Popularity of content analysis based on the ability to measure human behavior. Unlike the appropriate method of questioning. Two main types of analysis are quantitative and qualitative. [3] The first one is aimed at identifying the frequency of certain themes, words, phrases and symbols contained in the given user content, the second one uses a method for finding certain certain expressions [3].

The history of the development of content analysis. Many areas of application of content analysis exist nowadays. This is linked to the and considerable popularity, ease and efficiency of the method. The method of content analysis was founded in 1640 in Sweden. [3] Later the term was mentioned in the second article of the famous J. Speed – it was the first result of this method, analysis and trans dyvsya based editions of the New York newspapers on every Sunday from 1881 to 1883. J. Speed measured the amount of related materials, broke the relevant category of each subject, and then using the devices measured in inches and compared to the corresponding results. It turned out that with each year newspapers allocated more space to gossip, scandals and scenes, but less space to literary, scientific, political and research articles [6]. This method is used in the analysis of textual information, media information, data visualization and more. Also, utilizing its research focus to the newspaper. Not all the information serves as an object for content analysis. Today there are three main areas where this method is used [7]. In developing the method and its practical application are the following stages. After the formation of the research topic and its objectives, define a category of analysis is the most common, key concepts relevant research goals. Increasingly this method is spreading in the sphere of information and Internet technologies. Analysis and content management, its modeling, processing and integration is one of the most informative methods for the quantitative study of the dynamics of individual thematic areas and carrying out technical analysis site (Fig. 2). The change in the content management indicates the speed individual thematic areas of and the entire space [7]. Analysis of content is growing rapidly, due primarily to the development of information and Internet technologies, where this method is widely used.

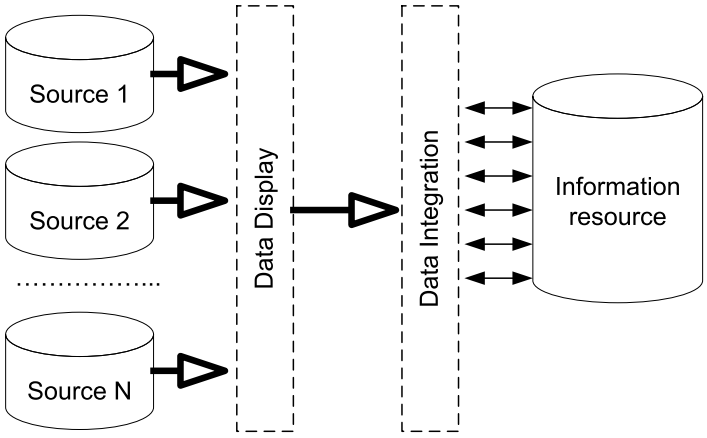

Fig. 2. Construction of information resource by integrating content

Peculiarities of designing IS for information resource processing. Process of designing and creating an IS processing information resources (ISOIR) on the basis of Internet –marketing is an iterative and proceeds from analysis, design, development plan to build will a prototype and experimental testing, since the formation of specifications, layout, template creation , content development and placement under information resource structure (Fig. 1–3). Concentrate the solution to business goals and needs of end users. In the initial stages to determine / develop functional requirements connected users through questionnaires, design alternatives and prototypes of varying readiness of degrees. That is collecting valuable information, causing the user a feeling of direct involvement in the design process and gaining their trust. the content analysis use for text in ISOIR offers several advantages to simplify maintenance of e–business and decides to lower the problems facing the participants of business processes, namely:

1) content filtering on user information resource;

2) the ability to automatically create a portrait of a permanent user by analyzing his comments;

3) the ability to automatically create a portrait of the target audience on the basis of analysis of portraits of everyday users;

4) the number of reducing moderators information resource in ISOIR;

5) reducing the time for posting content to user information resource because of its automatic processing, not the moderator;

6) elimination of language barrier due to automatic creation of user dictionaries and automatic translation.

Most definitions of content analysis are constructive or procedural. Due to different initial approaches they induce different algorithms that sometimes contradict each other [1–10]. Ignoring context raises the largest doubts. Practical value of the method avoids many contradictions. Consolidation of means and methods of natural selection and by repeated evaluation of the results make it possible to release or confirm factual knowledge and power / useful tools. Content analysis – a quantitative and qualitative analysis sets the text for further meaningful interpretation of the quantitative and qualitative patterns. The method is the formation of a variety of text content abstract model of content. It is used in the analysis of sources invariant structure / unstructured content in the form of randomly organized text [1–10]. The result of content analysis of content used in text determining the tone of the text, duplicate content, presence of spam detection and new events to determine the content of stories text content.

The main components of content analysis. Definition of key content by analyzing the texts are difficult for spam detection. In identifying spam considering two hypotheses (spam – not spam), and in determining the emotional tone check (positive, negative, neutral), and combinations thereof. In the method of Bayesian detection using base spam assessments – two buildings of content, one of which is made up of spam, and the other one is not [11–12]. For each content count frequency of use of each word and Measures OVU score (from 0 to 1), that the conditional probability that the content of the word spam [11–12]. The weights of value close to ½, do not take into account when calculating the integrated because words such weights ignored and removed. The hypotheses space contains tonalities $Tonality = H_{-1}$ (negative), $Tonality = H_0$ (neutral), $Tonality = H_1$ (positive). In H_1 hypotheses of the set positive tone chosen terms characteristic of this content. Of them choose terms with probability calculated by the formula and Bayes greater than ½. the decision to adopt the tone of content considering the difference of weight estimates hypotheses H_1 and H_{-1} [11–12]. For the purpose of identification of such content c_i and its duplicate c_j reflectivity rule is valid, but the transitivity condition $c_i \prec c_j, c_j \prec c_k \not\Rightarrow c_i \prec c_k$ is not met. The content is similar to the text from the sample which includes it, but the sample is not similar to the text. Alternatively the content may be similar to the other two which it was compiled from, but the originals are significantly different. For duplication ratio performed symmetry and transitivity, ie $c_i \equiv c_j \Rightarrow c_j \equiv c_i$ and $c_i \equiv c_j, c_j \equiv c_k \Rightarrow c_i \equiv c_k$. The ratio of reflexivity, symmetry and transitivity

properties is the ratio of equivalence[11–12], ie the ratio of duplication. Each content c_i matches the is associated with element vector according to the algorithm of signature term matching provided above [11–12]: $a_{ij} = \begin{cases} 1, & c_i \equiv c_j, \\ 0, & \textit{інакше}, \end{cases}$ at $U_B = \begin{vmatrix} a_{11} \dots a_{1n} \\ \dots\dots\dots\dots \\ a_{n1} \dots a_{nn} \end{vmatrix}$. Analysis of the In similarity criteria used terms symmetry $\forall i, j : a_{ij} = a_{ji}$ and transitivity $\forall i, j, k : a_{ij} = 1, a_{jk} = 1 \Rightarrow a_{ik} = 1$ changing volume terms for comparison of the relevant factors, ie $\left(\left(\sum_i^N \sum_j^N \left| a_{ij} - a_{ji} \right| \right) \Big/ \left(\sum_i^N \sum_j^N a_{ij} \right) \right) \to 0$ and $\left(\left(\sum_i^N \sum_j^N \sum_k^N a_{ij} a_{jk} a_{ik} \right) \Big/ \left(\sum_i^N \sum_j^N a_{ij} \right) \right) \to \max$, where N stands for content quantity. Asymmetry quotient is approximately associated with the duplicates definition and the transitivity level is associated fully.

Identifying new events In from the stream ads of content that consistently the input ISOIR by means of scanning or content router and selected thematically results reveal new developments in the description The content of the [11–12]. Ago They form a chain of this story content. that reflects new Content developments are the basis of clusters interconnected content [11–12]. the cluster is forming the basis of the plot in the chain $u_{t_1} = sim(c_i, Dictionary) > \alpha$, $u_{t_2} = \sum_{j=2}^{n} sim(c_i, Dictionary) > \beta$, where n stands for the amount of content flow; c_1 – current content; c_n – the latest content; c_i – i–th content; $sim(c_i, c_j)$ – the degree of closeness of content i to j; $sim(c_i, Dictionary)$ – The degree of closeness of content i to the dictionary, α and β – Empirically conditioned parameters [11–12]. If the content is a set of terms $c_i = \{w_{ik}\} = \{w : w \in c_i\}$, then $c_i + c_j = \{w : w \in c_i \mid w \in c_j\}$ is the combination of terms from the content c_i and c_j into vector $E_i = \{e_{ik}\}$ with N dimension, which is defined as $e_{ik} = \begin{cases} 1, & w_{ik} \in c_i, \\ 0, & w_{ik} \notin c_i, \end{cases}$ i.e. $E = \begin{vmatrix} e_{11} \dots e_{1n} \\ \dots\dots\dots\dots \\ e_{n1} \dots e_{nn} \end{vmatrix}$. The degree of closeness is defined as $sim(c_i, c_j) = \left(\sum_{k=1}^{N} e_{ik} e_{jk} \right) \Big/ N$. In [11–12] degree of proximity is determined using the conditional probabilities instrument (occurrence of some of term w in content c_i, provided it is in content c_j).

Main directions of content analysis. When creating effective ISOIR much attention is payed to content management as well as analysis of the content used in CMS for automation and reducing the amount of costs and time spent.There are several stages in managing content, namely: content analysis, content processing and presentation content. To work effectively ISOIR first conduct an analysis of content, then work out the corresponding results and make conclusions, then work out very content. and the final step is the presentation of content. Content analysis carried out by the following methods: content text analysis articles, analysis articles comments, rating of this evaluation, statistical analysis of visits [2, 3]. *analysis of comments* uses tion for analysis, updating and monitoring of moods Members ISOIR that in his comments, write: reviews of the system advantages and disadvantages or to adjust operational and liquidity information. *Analysis of statistics* uses tion under observation and also processing and results to determine the operative ness and liquidity of content. for example, if a paper was visited 100 users, and another – 1, then it safe to say is, that before the first article content is Forum more quickly than in the second. *rating of this evaluation* uses tion to rank the same articles and conducted away via survey, evaluation users, etc. (button Fig. 3).

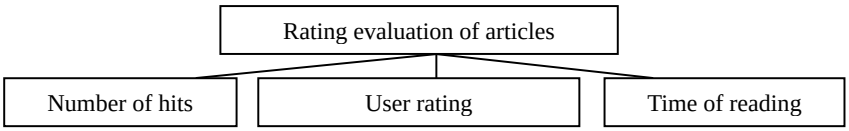

Fig. 3. Components of the rating of this evaluation items

In analyzing the content in most cases using two methods, namely: content analysis and technical analysis website ISOIR. In analyzing the content ISOIR using the following components: analysis of content from a user's perspective, an analysis of the number of required content, content and content analysis of linguistic, graphical analysis of content, content analysis of the text for keywords. [6] In analyzing the content in terms of user monitor its interests in the course of ISOIR Forum more. In analyzing the amount of content and determine the content of information. Information should be comprehensive and liquid. When performing semantic and linguistic analysis the main attention is payed to segment structure, super-segment structure, inter-segment interaction as well as functional interaction between super-segments. The linguistic analysis is a process of investigation of corresponding structural and system characteristics [5]. Graphical analysis of content is caused by the fact that most image information is perceived by users much faster than the text. This can be traced in for example in providing charts, graphs and histograms. When providing content, for example, in the form of tables information perception tends to be slower. When performing content analysis of textual information corresponding methods are used, and in this case based on two types of analysis: *quantitative and qualitative* (tab. 3) [5, 6]. However, only quantitative measures are crucial for making conclusions. Indicators of these values are either similar or different. In these cases, estimates are performed using simple calculations and applications [5]. Qualitative content analysis is focused on deep research of text or other material, including in terms of context. The results in this analysis are formulated based correlation bonds content elements and their relative importance in the structure of the text. Depending on the challenges set in the investigated, this analysis can be extended and some elements of other types of analysis, such as quantitative content analysis [5]. There are stages of analysis:

Table 3

Stages of quantitative and qualitative content analysis [1-10]

Q	What are the stages	Characteristics stage
quantitative	Bold analysis unit	Convert linguistic units on the form for processing.
	Counting frequency units	Of interdependence Detection relationships between linguistic units.
	Categorization	Definition of finite sets of redundant categories and to obtain quantitative data of their appearance.
	Data Mining	Detection of stream content via multiple quantitative assessments of new knowledge with further qualifications as their categories.
	Interpretation of results	Getting content, filled semantically results using mathematical methods and semantic formalizatoriv.
qualitative	Breaking into blocks of text	Formation of integrated content units for encoding and processing.
	Reconstruction stream content	Reconstruction of values, ideas, beliefs and evidence of each source text.
	Forming opinions	Withdrawal generalizations by comparing the individual system values.

Preparation Program document review. This stage typically formulated the so-called theory of empirical research. So, in preparation for analysis, hypotheses existing in the context of this subject matter are systematized and the ones not amenable to verification using the collected data are eliminated.

Selection of sources for analysis. It is necessary to determine the range of sources that contain materials and information.

Determination of empirical models of analysis of the sample (ie selection of communication, choice of materials for certain periods of time in determining types of messages such as sample).

Developing methodology for this particular analysis.

Pilot studies, check of methodology reliability.

Collection of primary empirical data;

Quantitative processing of data collected.

Interpretation of the obtained results, the conclusions of the study.

Problem allocation

Analysis of the content is the basis of journalism and mass communication that reveals the use of technology in areas such as education, linguistics, history, anthropology, philology and literary analysis. In general, the content analysis techniques use in these areas anyway floor is linked with the use within the sociological studies [6, 7]. The content management method uses the following components: generation pages when prompted, generating pages while editing, and mixed type. The technical analysis method in system uses the following components: technical analysis system availability, technical analysis of the system, technical analysis methods of information content management system. With the availability of technical analysis studies how systems perform system available to users, or it can run on different platforms and so on. If the analysis and the system perform research that if the system works correctly, or errors in the operation of the system and so on. Technical analysis of ways to manage information content of the system used for see how much liquidity and operational information, that system if it is popular th among users of the system [8]. In order to increase the popularity of the system usage of only filtered content is required.

Formulation of goals

The goal is to analyze the methods of content analysis and development of ISOIR, which have carried out a content analysis of textual information online newspaper, a ranking of articles, duplicate content, perform monitoring of users, after the assessment of the results of the analysis. Basic research carried out by methods Content Assist Analysis of textual information. of As a result, developed and improved methods of content analysis to process text data sets, further developed methods for ranking articles. The findings and results of the methods considered in updating information resource and improve the architecture ISOIR. the practical of value of the results lies in the development and testing of new methods and tools for system content analysis of textual information, saving human and financial resources and in building convenient for end-user software tools and systems.

Analysis of the research results

Main stages of applying content analysis of textual information

1. Sources aggregate determination, investigated, or messages according to preset criteria, which corresponds to each message: type of source (forum, e-mail, Internet newspaper, chat, Internet magazine); message type (article, email, banner, comment) parties involved in the communication process (sender, recipient); Mapping message the size (minimum volume or length); frequency of messages, method of distribution of messages; place dissemination of messages; time of occurrence reports and forum.

2. Formation of a limited sample messages.

3. Identification of linguistic units of analysis. There are strict requirements for the selection of possible linguistic unit of analysis: large enough to interpret meaning; small enough not to interpret many meanings; easily identified; LH sufficient number of units Ica for sampling. If analysis unit of the topics such rules take into account: the size of the theme does not extend beyond the paragraph; there is a new theme by changing the purpose theme, destination, category and persons for which created theme.

4. Bold calculation units, which may coincide with meaningful units or wear specific. In the first case, the analysis procedure is reduced to counting the frequency of the selected content units in another - from researcher analyzed material and research purposes raises calculating unit, which may be the physical length of texts; text area filled with meaningful units; the number of lines (paragraphs, characters, text columns); the size and type of file; number of pictures of certain content, plot and so on. In some cases, researchers use other elements calculation. Fundamental importance at this stage of content analysis has a strict definition of its operators.

5. Immediately calculation procedure. In general terms, similar to the standard methods of classification groups selected from the formulas of mathematical statistics and probability theory. There are also special calculation procedures concerning content analysis.

6. The interpretation of the results according to specific goals and objectives of the study. At this stage, are evaluated and the following characteristics of texts that allow to draw conclusions that would pidkr eslyty or hide the author [1, 2, 6–8].

Most important in the internal system is optimizing its text content, as content- analysis of submitted central event in a number of actions. In terms of global network Internet of value of each property is determined that it is the content. Analysis of the site from its content can only give the current assessment of the site as well as content over time, as well as any other information on the Web, ceases to be relevant, that gets old. The analysis procedure is quite simple, but the difficulty is how it should be changed to achieve the desired result. In other words, you can always point to determine the current status, the rating of this of each site, although in most cases not possible to predict optimization will effect what. So just making shots and comparing the result with what was assumed in the previous stage of optimization, you can see how correct was selected optimization direction after the current analysis was conducted site content [1–3]. Feature content- Another analysis to is the inability to direct “treatment”. That is, any content can be removed from the site without problems, but often do not necessary : much wiser it would be to change or edit, add Forum more. On this floor solved one of the most common “defects” of many sites, where content is edited, updated, very often pages are transferred to archives or lost leading to “broken links”. This means that previously stored reference pages do not correspond the information with [6].

Successful design of ISOIR content analysis of textual information requires correct identification of what provides effectiveness of the system. To solve this task a decision tree is built. The main aim is to develop ISOIR that includes the following sub-goals: choice of methods of solution. Liabilities, systems analysis and choice of development system and the best technology about the object – is oriented e design. The final goal is the choice of development ISOIR with sub pit, as the choice of the medium of writing programs, choice of with creating a database and creating a system interface. The aim sub-goals is to create a friendly end-user system interface that was easy to use. The users want to receive information in a concise, clear and correct manner. Do not want to read a couple of letters of the text and in the end did not get any information. For example, if the text will be composed of a large number of definitions and terms or formulas, it will be difficult chytabel it and the user will not be interesting to read. On other the hand if the text is a large amount of useless information, the user will spend Forum more time reading it. This is why information in newspaper articles should be adequate, prompt and understandable to the user.

For successful analysis using only one method is not possible, one must use a set of techniques, so we will use another credit rating of this evaluation and statistical analysis. Analysis of statistics of visiting the resource is used to monitor and process the results are used to determine the operative ness and liquidity information. Rating assessment used for the same ranking systems and items made using surveys, evaluation users and other methods. for rating of this evaluation popularly used three criteria: number of downloads, reading articles and user evaluation. These three criteria for rating of this assessment form articles. For k at the best understanding of the subject area construct hierarchy. Fig. 4

present hierarchy in the case of three alternatives analysis content. The hierarchy has four levels: objectives, criteria for higher and lower levels and alternatives. The numbers in Fig. 4 indicate priority elements of the hierarchy in terms of objectives, which are calculated in the MAI based on paired comparisons of elements on each level connected with them elements of a higher level. in our case known top-level priority criteria, providing the most weight criterion content analysis of text and rating of this assessment, since it about objectively reflects the quality of work. Criterion statistical analysis provide less weight because it does not fully reflect the quality of work. In the lower level criteria, providing the most weight criterion number of applications, this criterion displays as user interest in a particular material. the next weight criterion User rating of this that reflects the assessment of users for a specific material. the smallest scales provide criteria for reading and registering number of visits. So evaluate all alternatives for each criterion separately. the calculations follows that the first alternative is the best for the selected criteria evaluation. A very important factor for the system is the availability of inputs. In this case, the input is paper. Article – a journalistic or scientific work. Adding, editing and deleting items has administrator Internet -resursu. The source of information for articles are encyclopedias, periodicals, books and other articles, etc. [1]. Another important factor is the availability of users. Users – this individuals, who use information resource searching articles, read them and carry the vote. No user quality control aspect can not be content.

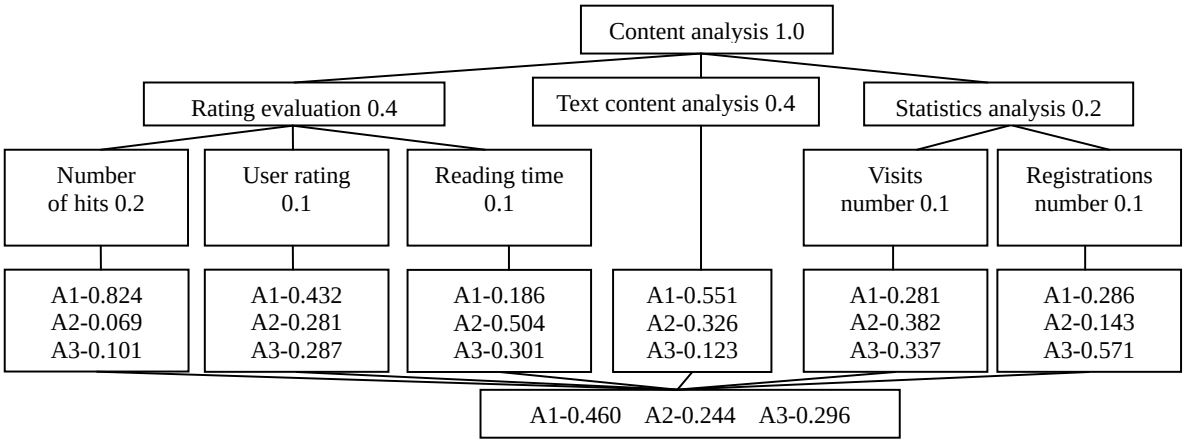

Fig. 4. The hierarchy in the case of three alternatives for content analysis

Fig. 5 shows the chart options for the use Tanna content analysis of Internet – newspapers. The essence of the chart is as follows: the designed system is supplied as a set of entities or actors that interact with the system through the so-called use cases [1]. Fig. 7 is displayed chart sequence that shows the interaction about objects, ordered by time of their execution. T aki charts show involved about objects and consistency of messages sent. In Figure 8 displayed chart cooperation, through which describes the overall context of interaction as a kind of time slices collection of objects that interact to a task or perform business purpose ISOIR. Unlike diagrams the we in sequence it serves only the relationship between objects that play a role in the interaction. On the other hand, this chart does not specify a separate time dimension.

Administrator only about object in the system, cue can carry up giving or removing items you stake check for new s tate if he is not provided data, performs development of new articles. In constructing diagrams cooperation was held definition about objects of the system conducted peo their interaction. of As shown in Fig. 8 diagram, there are fifteen parallel streams in this system. Every stream forwards appropriate messages and data on about the object to about object and button Fig. 6 reflected chart for rating of this evaluation. Chart activity can be considered a special case of state diagram. She can realize the procedural features and simultaneous control caused by the completion of internal activities and

actions. the system goes into action these states depending on the choice of the client system. As shown in the fig. 9, content analysis process is performed after the introduction of the article. Once all actions performed interpretation of the results of the content analysis.

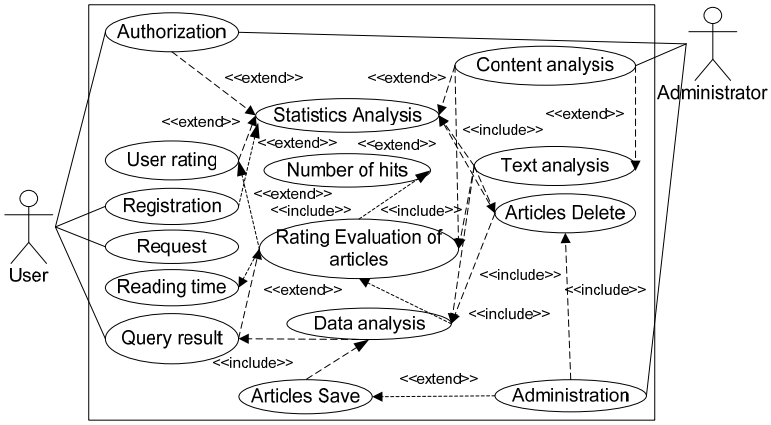

Fig. 5. Use Case Diagram

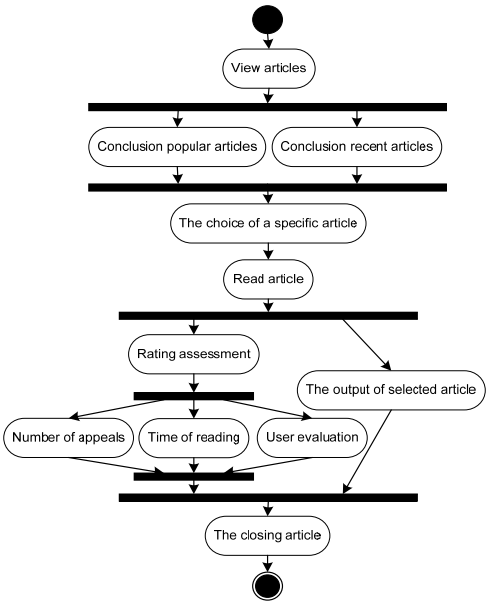

Fig. 6. Activity Diagram

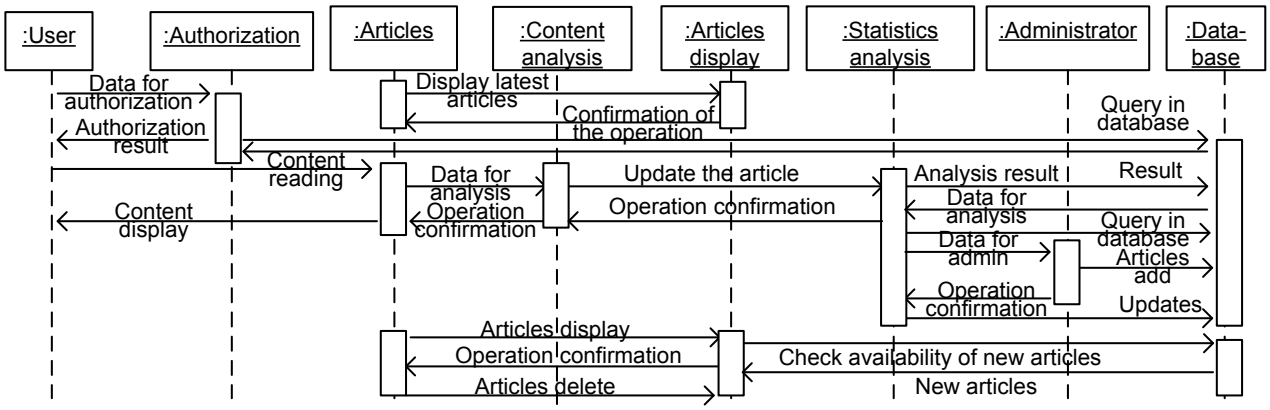

Fig. 7. Sequence Diagrams

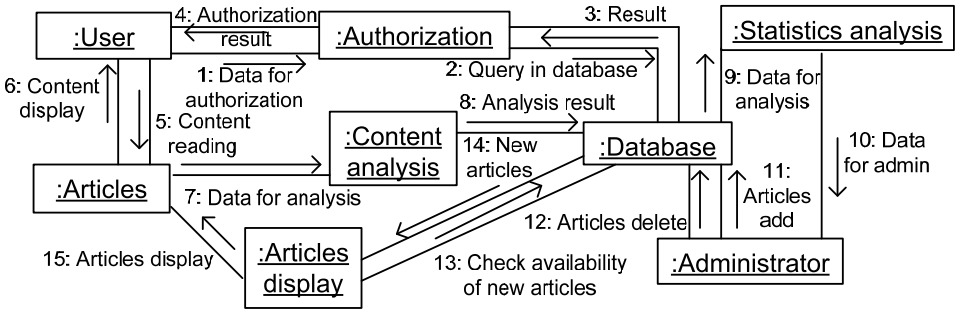

Fig. 8. Cooperation Diagram

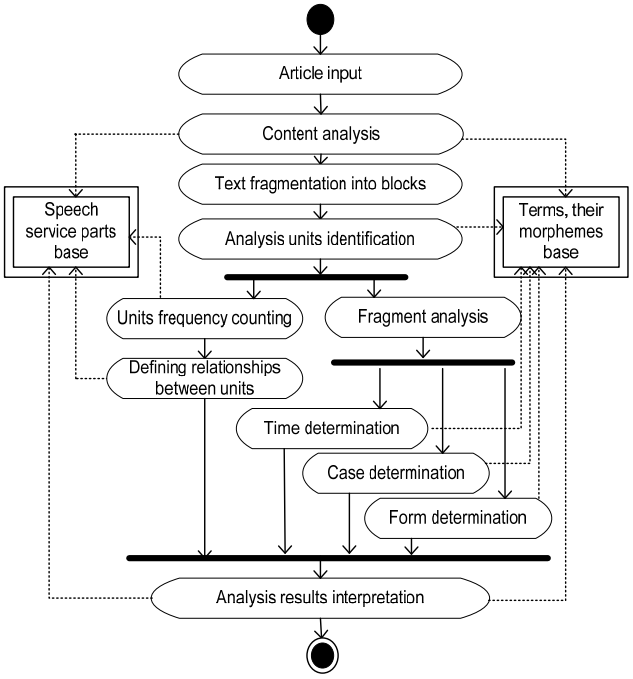

Fig. 9. Diagram of process content- analysis

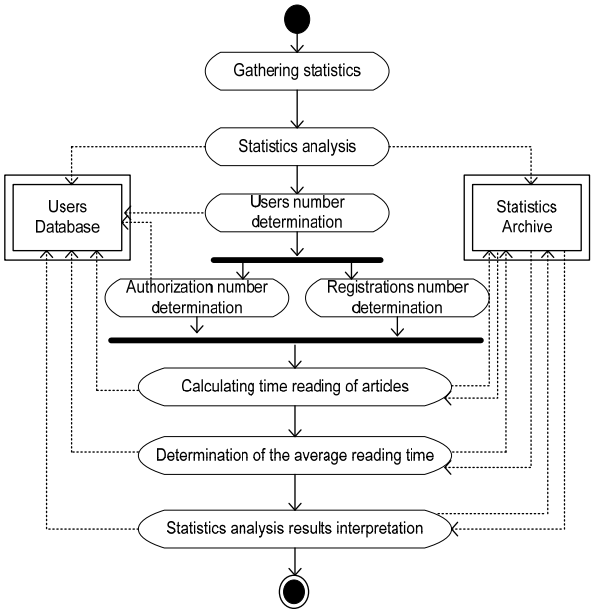

Fig. 10. Diagram of process analysis statistics

Figure 10 shows a diagram for the process of analyzing statistics. Statistical analysis is performed after the system will collect data for statistics. The main effect of the actions is determining the number of system users. Also determine the authorizations and registrations number. The amount system determines of time reading articles and average reading. Once all actions performed interpretation of the test statistics.

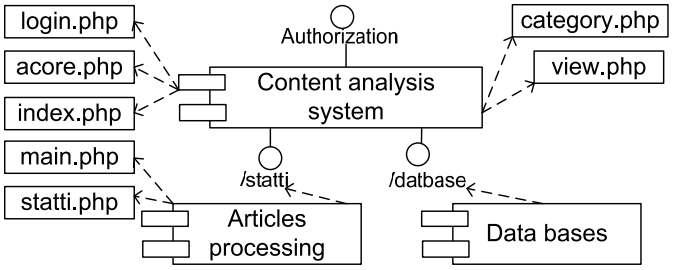

Fig. 11. Chart component

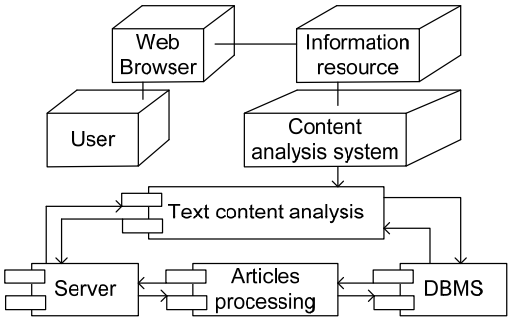

Fig. 12. Chart deployment

In Fig. 11 displayed chart components. Chart Component to determine the architecture of the developed system by installing dependencies between software components in a role which may be the original The, binary executable below code. With a user interface in its authorization system. However, power andis th component of the information system analysis of content online newspaper is a component of content analysis, which analyzes content away via respective classes. Another component is the component database that reflects the database and used with unity, as well as sharing information. Another component is the component processing items that used to carry me appropriate action on items. button Fig. 12 reflected deployment diagram, which is designed to visualize the elements and program components that are at the stage of implementation. The chart shows a part of system of content analysis of textual information online newspaper.

The articles published by the system administrator will serve as the information used by the system. After the presentation of papers by using a custom aspect of construction is carried ü analysis.

Conclusions and prospects for further scientific studies

In the article textual information content analysis ISOIR for an online newspaper is designed. The main purpose of this system is to improve the quality and timeliness of content of an online newspaper. Relevance of creating a system for content analysis of textual information in an online newspaper is caused by increasing user requirements of such systems and is caused by the following factors: rapid rate of demand growth for reliable and adequate information, the need for forming a set of operational information as well as automatic filtering of unwanted information and spam. Methods and design tools have been analyzed. With the help of methods and technologies charts have been made in order to display logical and physical model of how the system functions. Objects and usage options provided by the system have been defined as well as the basic functions of the system, the purpose of system development and system application. The conceptual model of the system has been elaborated. Input and output data have been defined and described. System requirement description has been provided. The designed ISOIR textual information content analysis system for an online newspaper uses statistics analysis, rating evaluation and content analysis in its activities, making it a powerful tool for analysis. The main advantages of these methods are: accuracy and reliability of results, simplicity of mathematical calculations as well as a wide field of application. In fulfilling the practical implementation a software implementation of the system has been written, database has been created, description of the task implementation has been made and analysis of the obtained results has been performed. System interface is provided in corresponding forms and is designed for filling the database and testing system functioning. Additionally verification and validation has been performed and corrections of any identified deviations and defects in the functioning of the system have been made. Thus, this ISOIR of content analysis of textual information in an online newspaper is a powerful tool in management and successful operation of any newspaper.

1. Berko A. Systems e-commerce content: monograph / A. Berko, V. Vysotska, V. Pasichnyk. – Lviv: Publishing House of the National University "Lviv Polytechnic", 2009. – 612 p. 2. Content of technique "content analysis" [Electronic resource] / T. Khoroshylova l // Applied inhvistyka. – Access: http://studentstpl.ucoz.ru/publ/teorija_vozdejstvija/metodika_content_analiza/soderzhanie_metodiki_content_analiz/12-1-0- 116. 3. Mathematical linguistics. [Book 1. The quantitative linguistics] teach. guide / [V. Pasichnyk, Y. Shcherbina, V. Vysotska, T. Shestakevych] // Series "Komp'yutynr". – Lviv: "New World - 2000", 2012. – 359 p. 4. Grigoriev C. Conducting content analysis [electronic resource]: Tutorial / S. Grigoriev. – Access: http : //www.psyfactor.org/lib/k-a2.htm. 5. Content analysis as a method of research [electronic resource] / Psi Factor is OTManayev // Access: http://psyfactor.org/lib/ analysis3.htm-content. 6. Birch A. E-commerce / A. Birch, J. Kozak, F. Levchenko. – K: Kyiv National Economic University, 2002. – 326 p. 7. Clifton B. Google Analytics: professyonalnыy poseschaemosty of web sites analysis / B. Clifton. – Moscow: Williams, 2009. – 400 p. 8. Derevya solutions – Sharing principles of work [electronic resource] / Akobyr Shahydy. – Access: http://www.basegroup.ru. 9. Analysis of the statistics, multivariate statistics methods. – Access: http://christsocio.info/content/view/492/102/ – the name of the title User. screen. 10. Ivanov VF Content analysis: Methodology and methods of research QMS: teach. / VFIvanov; [nauk guide. Ed. Moskalenko A.]. – K., 1994. – 112 p. 11. Fundamentals of modeling and otsenki elektronnыh ynformatsyonnыh flows / [D. Landэ V. Furashev, S. Braychevskyy, A. Grigoriev]. – K.: Engineering, 2006 – 348 p. 12. Landэ D. Principles of integration flows ynformatsyonnыh: monograph / D. Lande. – K. : Engineering, 2006. – 240 with. 13. Pospelov D. Sytuatsyonnoe Management: Theory and Practice / D. Pospelov. – M.: Science, 1986. – 288 p. 14. CM Lifecycle Poste r / Content Management Professionals. – 2010. Retrieved 20 July – Access: http://www.cmprosold.org/ resources/poster/. – Title from the screen. 15. EMC. Content Management Interoperability Services. Appendices. 0.5 the the Version / EMC, IBM, Microsoft. – Hopkinton: EMC, 2008. – 17 p. 16. EMC. Content Management Interoperability Services. I. the Version 0.5 Part the / EMC, IBM, Microsoft. – Hopkinton: EMC, 2008. – 76 p. 17. EMC. Content Management Interoperability Services. Part II – REST protocol binding. 0.5 the

the Version / EMC, IBM, Microsoft. – Hopkinton : EMC, 2008. – 79 p. 18. EMC. Content Management Interoperability Services. Part II - SOAP protocol binding. 0.5 the the Version / EMC, IBM, Microsoft. – Hopkinton: EMC, 2008. – 37 p. 19. Hackos J. Management for Dynamic Web Content Delivery / J. Hackos. – Hoboken: Wiley, 2002. – 432 p. 20. Halvorson K. Strategy for the Web Content / K. Halvorson. – Reading: New Riders Press, 2009. – 192 p. 21. McGovern G. Critical Content / G. McGovern, R. Norton. – Upper Saddle River: FT Press, 2001. – 256 p. 22. McKeever S. Web content management systems Understanding: evolution, lifecycle and market / S. McKeever // Industrial Management & Data Systems (MCB UP), 2003. – number 103 (9). – P. 686–692. 23. Nakano R. Content of web management of: a collaborative approach / R. Nakano. – Boston : Addison Wesley Professional, 2002. – 222 p. 24. Osgood C. Nature and measurement The meaning of / C. Osgood // Psychological Bulletin, 49 (1952). – P. 197–237. 25. Papka R. Line News Event Detection -On, Clustering, and Tracking: thesis for the degree of doctor philosophy / R. Papka. – Amherst: Massachusetts University, 1999. – 154 p. 26. Woods R. A Model for Defining Content / R. Woods. – 2010. – Access: http://www.contentmanager. net/magazine/article_785_defining_a_model_for_ content_governance.html. – Title from the screen. 27. Rockley A. Enterprise Managing Content: A Unified Content Strategy / A. Rockley. – Reading: New Press Riders, 2002. – 592 p. 28. Stone WR Plagiarism, Duplicate and Duplicate Publication Submission: ago ago They Are All Wrong! / WR Stone // IEEE Antennas and Propagation, 2003. – Vol. 45. – number 4. – p. 47–49. 29. Sullivan D. Web Gets Deeper Invisible / D. Sullivan // Search Engine Report. – 2002. 30. The Content Management Possibilities Poster [Electronic resource] / Metatorial Services, Inc. – 2010. Retrieved 20 July – Access mode: http://metatorial.com/pagea.asp?id=poster. 31. Methods based on ontologies for information resources processing : Monograph / [Vasyl Lytvyn, Victoria Vysotska, Lyubomyr Chyrun, Dmytro Dosyn] // LAP Lambert Academic Publishing. Saarbrücken, Germany. – ISBN-13: 978-3-659-89905-8, ISBN-10: 3659899054, EAN: 9783659899058. – 2016. – 324 p. 32. Vysotska V. Methods and means of processing information resources in the electronic commerce content: Abstract dissertation for the degree of Ph.D. 05.13.06 – Information Technology / Victoria A.; National University "Lviv Polytechnic". – Lviv, 2014. – 27 p. 33. Berko A. Features of information processing resources in electronic commerce content / Andriy Berko, Victoria Vysotska, Lyubomyr Chyrun // Applied Computer Science. ACS journal. – Volume 10, Number 2. – Poland, 2014. – ISSN 2353-6977 (Online), ISSN 1895-3735 (Print). – P. 5–19 [Online]. 34. Vysotska Victoria. Web Content Processing Method for Electronic Business Systems / Victoria Vysotska, Lyubomyr Chyrun // International Journal of Computers & Technology. – Vol 12, No. 2. – December 2013. – P. 3211–3220. – ISSN 2277-3061. – [Online] http://cirworld.org/journals/index.php/ijct/article/view/3299. 35. Vysotska, V. Modeling stages of the life cycle of commercial of web-content / V. Vysotska, L.Chyrun, L. Chyrun // Information networks and systems. Proceedings of the National University "Lviv Polytechnic". – Lions 2011. – N 715. – P. 69–87. 36. Vysotska, V. Features design and implementation of e-commerce. / V. Vysotsky // Information systems networks and. Proceedings of the National University "Lviv Polytechnic". – Lions. 2008. – number 631. – P. 55–84. 37. Vysotska Victoria. And Analysis evaluation of risks in electronic commerce / Victoria Vysotska, Ihor Rishnyak, Lyubomyr Chyrun CAD Systems in // Microelectronics, CADSM'07, 9th International Conference. – The Experience of Designing and Applications of CAD Systems in Micro e lectronics. – Lviv, 2007. – 24 February – P. 332–333.