Web Content Support Method in Electronic Business Systems

Victoria Vysotska[1], Vitor Basto Fernandes[2], Michael Emmerich[3]

[1]Lviv Polytechnic National University, Ukraine
[2]University Institute of Lisbon, Portugal
[3]Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands

Victoria.A.Vysotska@lpnu.ua

Abstract. In the given article content is support method as the content lifecycle stage in Web systems is proposed. The model describes the processes of information resources processing in the electronic business systems and simplifies the content automation support technology. In the given paper the main problems of e-commerce and content function management services are analyzed. The proposed method gives an opportunity to create an instrument of information resources processing and to implement the module of Web content support.

Keywords: content, context, information resources, content management system, content lifecycle, electronic content commerce system, Internet Marketing.

1 Introduction

The current development of the Internet contributes to information needs increasing such as production factor or strategic resource. It also activates the new implementation forms of information services [1-9]. Informational product (commercial content) is documented information, prepared according to user needs and designed to meet them [10-13]. Actions for commercial content user providing are information services. Internet market is set of economic, legal, organizational and programmatic relations with the sale/purchase of information products/services between producers/suppliers and users. The concept of commercial content is defined as information resource in filling in the electronic content commerce systems (ECCS); object of business processes in ECCS, such as paper, software, books, etc.; structured set of logically complete information to which the relationship between the user and ECCS; a set of data without predetermined patterns that exist only in electronic form; information for commercial purpose which indivisible in time; the main factor in the region formation of the ECCS functioning and purpose [14-21]. Information resources are a collection of documents, or arrays of instruments in information systems (libraries, archives, collections, data banks, repositories, museum storages, etc.), which are subject to ECCS. Content is a set of information resources and products that are stored in an environment information system (IS) of Virtual Library and available for application users of the system [22-26]. Content is these data without pre-defined

structure of visual presentation as opposed to structured data that usually are managed by DBMS. The development of information technology led to the fact that in the modern world economy it has become a key concept of content is an important element of the driving force of economic growth and social change. The successful development and growth of Internet e-commerce content once again proved that the information economy is dynamic and profitable. There are three fundamental content definitions [18-29]:

1. any significant information content IS (e.g., text, graphics, multimedia);
2. part of the message, which is not processed and does not change during transmission;
3. content of the document data (other than attributes) that contains text, images, videos, audio, scripts, programs, or any other similar material content of solid media.

A variety of content commercial content is this product information or content of Web-site, Internet publishing, market research, consulting services, which are subject to business processes ECCS. Content Management System (CMS) is a system that meets a specific set of requirements and is designed for content management. Such systems are used for storage and publication of a large number of documents, images, music or videos. Another case of such systems is ECCS [30-36]. Such CMS can manage text and image content, giving user's convenient storage tools and information publications. Management System Web-content is designed to generate content within the portal and have the same problem (dynamic collection, content caching, its safety, etc.) and other kinds of Web-applications. The initial information for constructing mathematical models of processes are functioning ECCS data on purpose and working conditions of CMS [18-21]. This information defines the main purpose of modeling ECCS and allows us to formulate requirements on formal models of ECCS and models of content management.

2 Formal ECCS model

Formal ECCS model $S = \langle X, C, V, H, Y \rangle$ is a tuple of variables that describe the process of the system (Table 1).

Table 1. Components of electronic content commerce systems

Set	Marked	Range	Vector form
Incoming impacts on system	$x_i \in X$	$i = \overline{1, n_X}$	$\vec{x}(t) = \left(x_1(t), \ldots, x_{n_X}(t)\right)$
Influences the flow of content to the system	$c_r \in C$	$r = \overline{1, n_C}$	$\vec{c}(t) = \left(c_1(t), \ldots, c_{n_C}(t)\right)$
External impacts environment	$v_l \in V$	$l = \overline{1, n_V}$	$\vec{v}(t) = \left(v_1(t), \ldots, v_{n_V}(t)\right)$
Internal (own) system parameters	$h_k \in H$	$k = \overline{1, n_H}$	$\vec{h}(t) = \left(h_1(t), \ldots, h_{n_H}(t)\right)$
Baseline characteristics of system	$y_j \in Y$	$j = \overline{1, n_Y}$	$\vec{y}(t) = \left(y_1(t), \ldots, y_{n_Y}(t)\right)$

The values x_i, c_r, v_l, h_k, y_j are elements of subsets that are disjoint and contain deterministic and stochastic components [18-21]. Inbox influences x_i, influences the flow of content c_r, the external environment v_l and internal system parameters h_k are independent, and source information y_j systems are dependent variables [18-21]. The process of operation S is described by a function $\vec{y}(t) = f\left(\vec{x}, \vec{c}, \vec{v}, \vec{h}, t\right)$.

3 Comparative characteristic of content management

A description models built of content flow is important and remains unexplored. When considering in [18-21] the dynamics of thematic content streams found that models are limited, opening the way for further research. In Table 2, comparative characteristics of content management are presented, which describe the process of identifying and aging research (actual) content rather than defining the stages of the content lifecycle (creation, management, implementation). Existing CMS do not support full content lifecycle (creation, management, implementation) streams and do not solve the fundamental problems of information resources processing – content development and implementation. Formal ECCS model does not disclose the content management mechanisms. Formal models of content management appointment only determine the aging process (actual) content stream, and some of them (logistics, analytical) and thematic content stream. They do not solve the problem of the content formation, implementation and not solve all problems, content management, for example, submitting a set of content by end users according to their inquiry, history or information portfolio, automatic digests creation, information portraits, automatic detection of thematic subjects, the construction of correlation tables of communication concepts, calculation of ratings terms, gathering information from different sources and its format, identifying keywords and concepts of content, automatic categorization of content, identifying duplication of content, selective dissemination of content. A basic lack of formal models of content management is the lack of links between input information, content and output information in ECCS.

Table 2. ECCS characteristic of content management models

Model	Advantages	Disadvantages
Burton-Kebler $C = \langle A, B, T \rangle$	Describes process of aging content, loss of its relevance. Equation model has the exact solution in a very simple and convenient function of the exponential. Ability to determine speed of development of different thematic areas and the entire content space.	They are quite questionable in terms of interpreting the results. Monotonically increasing functions do not describe the processes in nature and must have local extrema.
Classical space-vector $C = \langle TF, IDF \rangle$	Defining meaningful term in the whole stream of content. Ability to determine the most urgent set of	Required content ranking, use of parametric factors that depend on time.

Model	Advantages	Disadvantages
	content available.	
Linear $C=\langle Y,T,T_0,V\rangle$	Determination of the intensity of flow subject content in time (for example, as a result of aging information)	Used in the linear dynamics of subject content.
Exponential $C=\langle N,T_0,T,\lambda\rangle$	Describes the process of aging content, the loss of its relevance.	Correlation between individual content insignificant.
Logistics $C=\langle N,T,M\rangle$	It combines the relative simplicity of the formulation of the problem with the ability to vary the solution over a set of parameters that may have more or less obvious physical meaning.	A description of only a single thematic thread. The dimension parameters and their dimension is not taken into account.
Analytical $C=\langle W,D\rangle$	Describes the process of aging content, the loss of its relevance.	The presence of the dictionary of keywords.

4 Research results analysis

The general principles of ECCS architecture design allow to realize technology of a information resources processing effective. A formal model of electronic content commerce systems is six elements as

$$S=\langle X,Formation,C,Management,Realization,Y\rangle,$$

where $X=\{x_1,x_2,\ldots,x_{n_X}\}$ is input data set, *Formation* is operator of content formation, $C=\{c_1,c_2,\ldots,c_{n_C}\}$ is a content set, *Management* is operator of content management, *Realization* is operator of content support and $Y=\{y_1,y_2,\ldots,y_{n_Y}\}$ is initial information set. Input information x_i is independent variables in vector form with like $\vec{x}(t)=(x_1(t),x_2(t),\ldots,x_{n_X}(t))$, and content c_j and source information y_k are dependent variables given via $\vec{c}(t)=(c_1(x_i,t),c_2(x_i,t),\ldots,c_{n_C}(x_i,t))$ and $\vec{y}(t)=(y_1(t+\Delta t),y_2(t+\Delta t),\ldots,y_{n_Y}(t+\Delta t))$. The formation of content to describe as $\vec{c}(x_i,t)=Formation(\overrightarrow{u_F},x_i,t)$, where $\overrightarrow{u_F}(x_i)$ is a set of conditions of the content formation, so $\overrightarrow{u_F}(x_i)=(uf_1(x_i),uf_2(x_i),\ldots,uf_{nUf}(x_i))$. Content is formed as

$$c_j=\left\{\bigcup uf_k \middle| \begin{matrix}(x_i\in X)\wedge(\exists uf_k\in U_F),\\ U_F=U_{Fx}\vee U_{F\bar{x}}, i=\overline{1,m}, k=\overline{1,n}\end{matrix}\right\}.$$

The process of content realization is described by the function

$$\vec{y}(t+\Delta t)=Realization(\overrightarrow{u_R},\vec{c},t,\Delta t),$$

where $\overrightarrow{u_R}(x_i)=(ur_1(x_i),ur_2(x_i),\ldots,ur_{n_{Ur}}(x_i))$ is a set of terms of content, so

$$y_j = \left\{ \bigcup ur_k \left| \begin{array}{l} (c_i \in C) \wedge (\exists ur_k \in U_R), \\ U_R = U_{Rc} \ \vee U_{R\bar{c}} \ , i = \overline{1,m}, k = \overline{1,n} \end{array} \right. \right\}.$$

Formal model of the formation of information resources formation will be provided as

$$Formation =< X, C, Gathering, Formatting,$$
$$KeyWords, Categorization, Backup, Dissemination >$$

where X is the set of incoming information from Web-sites or from moderators, C is the set of content, $Gathering$ is a set of functions of collecting information from sources, $Formatting$ is the set of functions of formatting information, transforming the set of content, $KeyWords$ is a set of functions identify keywords, concepts, $Categorization$ is the set features an automatic categorization, $Backup$ is the set of functions identify duplication of content, $Dissemination$ is a set of functions of selective dissemination of content.

The set of the commercial content is shaped by dependence $C_F = Formatting\left(Gathering\left(X, U_G\right), U_F\right)$, where U_G is a finite set of conditions for collecting information from various sources, U_F is the set of finite terms of formatting information. Detection duplication of content $C = Backup\left(C_F, U_B\right)$, where U_B is finite set of conditions identifying duplication of content. Detection of duplicate content within the meaning of ECCS based on linguistic methods is to identify general terms, chains which constitute verbal signature content. Automatic categorization $C_{Ct} = Categorization\left(KeyWords\left(C, U_K\right), U_{Ct}\right)$, where $KeyWords\left(C, U_K\right)$ is detection of keywords, concepts; U_K is a set of conditions out the keywords, concepts; U_{Ct} is set an automatic categorization; C_{Ct} is a set of classified content.

Processing a set of content C to identify important keywords built on the principle of detecting keywords in content (terms), which is based on Zapf's law and reduced to the choice of words with an average frequency of occurrence (most common words are ignored by "stop-dictionary", and rare words with text messages not include). The function of selective commercial textual content distribution $C_D = Dissemination\left(C_{Ct}, U_{Ds}\right)$, where C_D is set of selectively distributed content; U_{Ds} is a set of conditions of selective dissemination of content. Selective distribution list content $C_{Ds} = \sup(C_{Ct})$ depends on the level of demand for this content. Associative rules forming the list of contents consists of a list of content C_{Ct} (original list) and the list of content selected from the original list C_{Ds}, so $C_{Ct} \to C_{Ds}$. Formation of association rules is formation content formed as a result of initial and original lists. The probability of formation of association rules is the probability with which the content of the original list C_{Ct} will appear in the content of the original list

C_{Ds} in the database, so

$$cf(C_{Ct} \to C_{Ds}) = \frac{\sup(C_{Ct} \cup C_{Ds})}{\sup(C_{Ct})}.$$

Models of resource management. Main tasks of the content management module are creation, rotation of the database and providing access to it, forming the operational and retrospective database; personalization of users, storing personal user queries and sources of statistics, ensuring the search in database, generation of output forms, information interaction with other DB.

Classification models of content management.

1. Formal model *generating pages on demand* is

$$Management_Q = \langle X, C, Q, R, Edit, Y \rangle,$$

where $X = \{x_1, x_2, \ldots, x_{n_X}\}$ is the set of incoming information; $C = \{c_1, c_2, \ldots, c_{n_C}\}$ is the set of content; $Y = \{y_1, y_2, \ldots, y_{n_Y}\}$ is the prevailing set of pages; $Q = \{q_1, q_2, \ldots, q_{n_Q}\}$ is a set of requests; R is feature formation and presentation pages; $Edit$ is editing and modifying content. The process of editing and updating content is described by $\vec{c}(x_i, t) = Edit(\vec{c}, x_i, t)$. The formation of pages is described as $\vec{y}(t + \Delta t) = R(\vec{q}, \vec{c}, \vec{w}, t, \Delta t)$, that is,

$$y_i(t + \Delta t) = \left\{ \bigcup (c_j, \vec{w}) \middle| \begin{matrix} (\forall c_j \in C_q) \wedge (\exists q_i(t) \in C_q), \\ C = C_q \vee C_{\bar{q}}, j = \overline{1, m}, i = \overline{1, n} \end{matrix} \right\},$$

where $\vec{w}(t)$ is the content weight of the block.

2. *Generation of pages while editing.* When making changes to the contents of a site create a set of static pages then ignores interactivity between the visitor and content. Formal model of the generation of pages while editing

$$Management_E = \langle C, Edit, Y \rangle,$$

where $C = \{c_1, c_2, \ldots, c_{n_C}\}$ is the set of content; $Y = \{y_1, y_2, \ldots, y_{n_Y}\}$ is a set of static pages; $Edit$ is editing and modifying content. The formation of static pages to describe $Edit$ appearance $\vec{y}(t) = Edit(\vec{c}, \vec{w}, t)$.

3. *Mixed type* combines the advantages of the first two types. It can be realized by caching is submission module generates a page once; it still several times faster downloads from the cache. Cache can be updated either automatically after a certain period or when amending certain sections of the site and a team administrator. Another approach is the preserving of certain information blocks at the stage of collecting and editing site pages with these blocks when requested by the relevant page. Formal model of mixed type is then given as

$$Management_M = \langle X, C, Q, R, Edit, Caching, Y \rangle,$$

where $X = \{x_1, x_2, \ldots, x_{n_X}\}$ is the set of incoming information; $C = \{c_1, c_2, \ldots, c_{n_C}\}$ is the set of content; $Y = \{y_1, y_2, \ldots, y_{n_Y}\}$ is the prevailing set of pages $Q = \{q_1, q_2, \ldots, q_{n_Q}\}$ is a set of requests; R is feature formation and presentation pages; *Edit* is editing and updating content; *Caching* is function cache formation. The formation of cache pages

$$Cache = Caching(\vec{y}, \vec{w}, t, \Delta t),\ Cache = \left\{ \bigcup y_i \,\middle|\, y_i \in Y, t + \Delta t, i = \overline{1,n} \right\}.$$

Formal model of information resources will be provided as

$$Realization =< X, C, BuInfPortr, BuDigest,$$
$$IdThemTop, ConCorrTablConc, CalRankConc, Y >,$$

where X is the set of incoming information; C is the set of content; *BuInfPortr* is a set of functions forming information portraits; *BuDigest* is a set of functions forming digest; *IdThemTop* is set of functions identifying thematic stories; *ConCorrTablConc* is a set of functions build tables relationship concepts; *CalRankConc* is a set of functions calculating ratings of concepts; Y is a set of initial information.

Finite set of content $C = \langle C_P, C_D, C_T, C_C, C_R \rangle$, where C_P is a set of content information of portraits; C_D is a set of content digest, C_T is a set of thematic content of plots; C_C is a set of tables of content relationship concepts; C_R is a set of content ratings of concepts.

The set of content information of portraits $C_P = BuInfPortr(X, U_P, C_T)$, where U_P is a finite set of conditions for the formation of information portraits. The set of content digests $C_D = BuDigest(X, C_T, U_D)$, where U_D is a finite set of conditions the formation of digests.

The set of thematic content subjects $C_T = IdThemTop(X, U_T)$, where U_T is a finite set of conditions identifying thematic subjects. Terms of identifying thematic scenes $U_T = sim(c_i, PD) > \alpha$, where c_i is current content monitoring; PD is dictionary; $sim(c_i, PD)$ is a measure of proximity of content i to content dictionary; α is empirically conditioned setting.

The set of tables of commercial content relationship concepts $C_C = ConCorrTablConc(C_T, U_C)$, where U_C is finite set of conditions for building a table relationship concepts. The set of content ratings of concepts $C_R = CalRankConc(C_C, Th, U_R, Sp)$, where U_R is finite set of parameters calculating rankings concepts Th is key criteria of content; Sp is a function definition of spam. The set of initial information $Y = Realization(C_D, C_R)$. In the case of as-

sessment content key hypotheses space will include: $Th = H_{-1}$ is a negative tone; $Th = H_0$ is the tone neutral and $Th = H_1$ is positive tone. With a set of content with a positive tone selected terms that are typical for this content. Of these selected terms with values of probability $p(t \mid H_1) \geq 1/2$.

5 Content analysis of textual information

The system provides the introduction, the content categories choice, content ordering, mutual settlements implementation, the order tracing. There are different classes of ECCS and models variety of such systems. Nevertheless the article highlights the main patterns of transition from the processes of information resources formation to them support. Accordingly formal models of information resources formation, management and maintenance were proposed. These models allow realizing the ECCS architecture optimally. The complex method of the content forming provides a information gathering from different Web-site and its formatting; content keywords and concepts identify; automatic categorization of content; content duplication identify; selective dissemination of content. Operative method of the content management provides database formation and access to it; operational and retrospective database formation; database rotation; user experience personalized; personal needs and sources maintaining; counting; statistical analysis of the users operation in the system; content search in the database; initial forms generation; information interaction with the database of other subsystems. Complex method of content support provides information portraits formation; digests formation; thematic content identification; concepts relationship tables construction; concepts ratings calculation; new developments identification, their tracking and clustering. The fully functional ECCS is characterized complex system of interrelated operations, methods, techniques. The annotations database creating is a search images database creating of original content and their clustering (to the content group formation with close on some criteria of search images). Annotation base (search pattern of clusters used in the search process) is associated with the cluster base. Each entry corresponds to its specified cluster and description containing. This description is made of automatic abstracting methods.

The content accompany step *Realization* described in the operator form

$$y(t_p + \Delta t) = Realization(v_l, h_k, c_r, z_w, t_p, \Delta t),$$

where v_l is conditions set of content support and external influences of environment on the system, i.e. $v_l = (v_1(q_i, h_k, c_r, z_w, t_p), \ldots, v_{n_V}(q_i, h_k, c_r, z_w, t_p))$; h_k is the set of content management conditions, i.e. $H = \{h_1(c_{i+1}, q_d), \ldots, h_{n_H}(c_{i+n_H}, q_d)\}$; $Q = \{q_1, q_2 \ldots, q_{n_Q}\}$ is set of user queries; value $Z = \{z_1, z_2 \ldots, z_{n_Z}\}$ is set of information resource pages $z_w \in Z$ of in the system S at $w = \overline{1, n_Z}$; the value $T = \{t_1, t_2 \ldots, t_{n_T}\}$ is time $t_p \in T$ transaction of information resource processing in the system S when $p = \overline{1, n_T}$; the value $Y = \{y_1, y_2 \ldots y_{n_y}\}$ is set of statistical data

$y_j \in Y$ in system S at $j = \overline{1, n_Y}$.

Output statistics implemented are as

$$y_j = \left\{ \bigcup_l v_l \middle| \begin{array}{l} (\exists q_d \in Q) \wedge (\exists z_w \in Z) \wedge \\ \wedge (\forall v_l \in V) \wedge \left(\forall \left(c_r \wedge q_d\right) \in h_k\right), \\ V = V_{q_d} \vee V_{\overline{q_d}}, d = \overline{1, n_Q}, l = \overline{1, n_V}, \\ w = \overline{1, n_Z}, r = \overline{1, n_C}, k = \overline{1, n_H} \end{array} \right\}.$$

Content analysis of textual information allows determine the signs prevalence of researched content in ECCS. It is important not as an absolute, but the relative importance of attributes, i.e. characteristics of the place (shares) among other features. For example, this is the percentage of forum users discussion of economic Issues concerning political or the percentage of positive comments on articles regarding the negative and in respect of all comments on this category of articles in online newspapers. The degree of correlation between the features in the texts provides empirical data to understand of the functional links between elements reflected in the texts reality eg the audience mood determination of online newspaper about the economic or political situation in the country and/or world. If there are texts that have chronological sequence received number fixed in time portraits of the investigated reality (change in demand for a content category according to the season, such as winter read more fiction and detective stories – summer) or the target audience portraits (change in demand for a content category according to the article, for example, demand for political articles before the election). This allows to put forward hypotheses of the prognostic character about the system functioning. In digest formation using content analysis with regard to frequency weights of words from the concepts dictionary generated. The digest formation consists are algorithms of the concepts dictionary forming, of the content duplicate definition and of the digest create (alg. 1).

Algorithm 1. A digest create.

Stage 1. Content Select based on its weight.

Step 1. Digest size C input.

Step 2. The algorithm 1 implementation.

Step 3. The weight consistent determine of each content as the weights sum of its individual words that $W = \sum w_i$.

Step 4. The input content stream sort from the weights values.

Step 5. Meaningful content duplicates definition for statistical criterion of text uniqueness $U \geq 0{,}9$.

Step 6. Content filter of unsuitable for digests building (when $W \leq l$, where l – Content removal threshold value by the self-education rules of content structuring and moderating) and statistically substantial duplicates.

Step 7. The choice of $V = q$ content with greater weight where $q = const$ and the moderator given.

Stage 2. Digest text construction of selected content.

Step 1. Dictionary construction of selected content.

Step 2. Content analysis application to the text (Table 1).

Step 3. Sentences filtration that do not meet the semantic rules of content structur-

ing and moderating.

Step 4. Hypertext presentation formation of digest, its contents and a link to the original source.

Stage 3. Generated text edit of digest.

Step 1. The check amount of generated content c_i. If $c_i < C$, then step 2, otherwise stage 4.

Step 2. Content delete from the input stream that is used to the digest formation.

Step 3. Steps 1-2 implementation.

Step 4. Resulting append to the pre-formed digest and move to step 1.

Stage 4. Digest text formation as a separate content and its maintaining in the database with reference on the source.

Table 3. Content analysis stages of textual information

Stage	Stage characteristics of content analysis
Total sources or content determination	Using a set of defined criteria which corresponds to each content: given type of source; one type of content; given the parties which involved in the communication process; message size matched (minimum / length); messages appearance frequency; messages distributing method; messages distribution space; messages appearance time, etc.
Content analysis selection	The selected set of content is formed on the criteria a limited sampling from a larger array of information. Its forming using the procedure from a set of precisely defined actions for processing without any changes of all objects study.
Linguistic units identifying	Compliance with strict requirements concerning the linguistic units choice for content analysis: large enough to interpret meaning; small enough not to interpret the many meanings; easily identified; units number is large enough for sampling. When taking of the themes analysis unit take into account that its size does not go beyond a paragraph; new theme arises with the new characteristics appearance of linguistic units.
Computation units finding and classifier formation	Computing units may coincide with semantic units, or have specific characteristics.In the first case, the analysis procedure is to the frequency calculate of the selected content unit use. Otherwise, the researcher proposes computation unit (physical length of the texts; text area, filled with informative units; the number of rows, paragraphs, characters, columns in text; the file size / type; pictures number with a certain content and story) based on the analyzed material and research purposes.
Procedure for calculating	Standard techniques for classification of selected groups of mathematical statistics and probability theory formulas.
Results interpretation	This includes all extracted text fragments. When forming conclusions do not take into account of the some results, without exception all. Here are identified and measured the text characteristics. They allow drawing conclusions about that wanted to emphasize or hide its author. Or they predict changes in demand for content based statistical set of calculated coefficients for the time period of specified category.

Content with new themes is the new groups basis of interdependent content in thematic stories identifying with the following procedures:

- control within the system is level destination of user access to different content;
- content integration is content moving to a new decision;
- content support of various types is content storage and sorting in a central repository;
- detailed documentation and context-intelligent help support;
- rating system of site articles evaluation;

- template changes is general formatting changes to the content of the part site reflects the entire site;
- workflow support is automated business processes create for specific content;
- content marking is new categories and markers adding to content before / after saving;
- version control is new versions creating, view and return to the previous versions of content;
- content analysis of text streams in the system;
- visual administration tool is easy authors management of content, without resorting to programming, typically implemented using HTML-forms;
- concepts relationship tables construction.

The concepts ratings calculation is based on procedure for results calculating of content analysis, taking into account the ratio coefficient c of positive and negative (for the selected item) estimates, opinions, arguments, as described in the user comments on the content ECCS. In Table 6 presents a available stages list of the commercial content lifecycle on the developed system. In Table 2 also presented the operation results of the developed systems from Google Analytics. They imply that the presence of all stages of the content lifecycle on the site significantly increases the amount of visits and unique users.

Table 4. The functioning results of electronic commerce systems

Characteristic	Information resource				
	Fotoghalereja-Vysocjkykh	Vgolos	Tatiana	PressTime	AutoChip
Content formation	+/–	+	–	+/–	–
Content management	+	+	–	+	+
Content support	+/–	+	–	+	+/–
for the period from 10.2011 till 11.2011					
Visiting	142	199873	43	124653	372
Average Time of Online Spent	2:04	2:48	3:40	2:18	2:49
Refusals Index	61,27	65,99	60,47	59,89	58,06
Achieved goal	6	0	0	0	54
Dynamics (%)	-8,97	39,74	-58,6	23,18	17,72
Pageviews	349	425576	98	245632	1013
Pages number per visit	2,46	2,13	2,28	2,09	2,72
New Visits (%)	70,42	36,84	55,81	23,54	75,27
Absolute Unique Visitors	112	98845	27	45321	290
Traffic Sources - search systems	76,76	41,91	60,47	42,75	54,03
Traffic Sources - Direct Traffic	11,27	10,50	39,53	10,50	26,34
Traffic Sources - Other Sites (%)	11,97	47,34	0	46,75	19,62
for the period from 10.2010 till 11.2011					
Visiting	2033	1813928	186	913929	2423
Average Time of Online Spent	7:03	3:08	6:52	2:58	2:56
Refusals Index	46,34	62,45	48,68	62,23	48,68

Characteristic	Information resource				
	Fotoghalereja-Vysocjkykh	Vgolos	Tatiana	PressTime	AutoChip
Achieved goal	253	0	0	0	449
Dynamics (%)	8,97	39,74	8,65	23,18	17,72
Pageviews	12694	4249331	802	2149567	8423
Pages number per visit	6,24	2,34	4,31	2,12	3,48
New Visits (%)	55,53	35,34	22,04	25,65	65,37
Absolute Unique Visitors	1152	671308	41	334536	1592
Traffic Sources - search systems	49,09	42,77	64,52	40,75	49,94
Traffic Sources - Direct Traffic	20,17	15,76	22,04	16,53	29,67
Traffic Sources - Other Sites (%)	30,74	41,21	13,44	42,72	20,39

In Fig. 1 presents the functioning results of the developed systems from Google Analytics in graphs and charts. They imply that the presence of all stages of the content lifecycle on the site significantly increases the amount of visits and unique users.

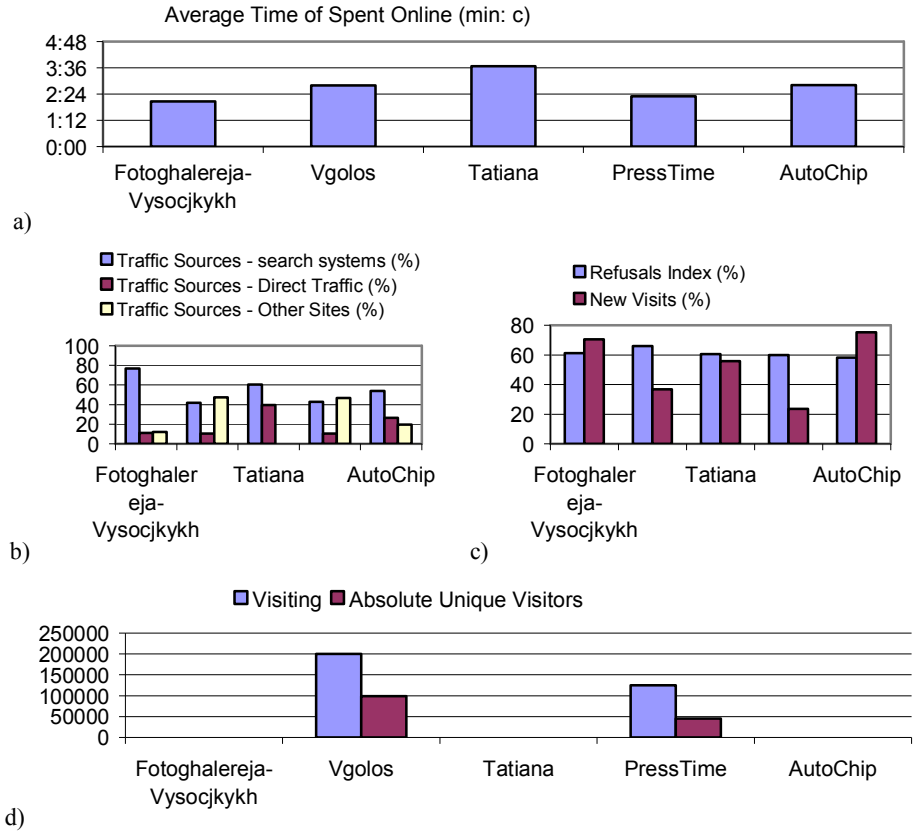

Fig. 1. Distribution of a) the residence time, b) traffic, c) new visits and refusals, d) visit to the resource

The lifecycle model of commercial content implement in the form of content-monitoring complexes to content collection from data various sources and provide a content database according to the users information needs. As a result, content harvesting and primary processing its lead to a single format, classified according to the Categories and he is credited tags with keywords. This facilitates the commercial content management process implementation. In text analyzing explore its layered structure: the source text as a characters linear sequence; morphological structure linear sequence, statements linear sequence, related unity net. The text preliminary study provides for the text division into individual tokens that carry out the finite automata method. Entry information is text in natural language text as a characters sequence, and output information – analyzed text partition, sentences and tokens table. There is the following relationship: the more unique content in the electronic content commerce systems, the more the visitor's information resource in its system. Commercial content support subsystem reduces the time to fill out unique content an information resource and increases the volume in a short time at this unique content in information resources and the queries number from search engines. These data take into account when creating or updating information resource and improve the electronic content commerce systems architecture. The service of the statistics of visits to the information resource of the ECCS allows us to estimate the increase in the volume of sales of content from the direct proportional dependence of the increase in the number of visits to the information resource, the number of regular users, and the promise of marketing activities (Figures 2).

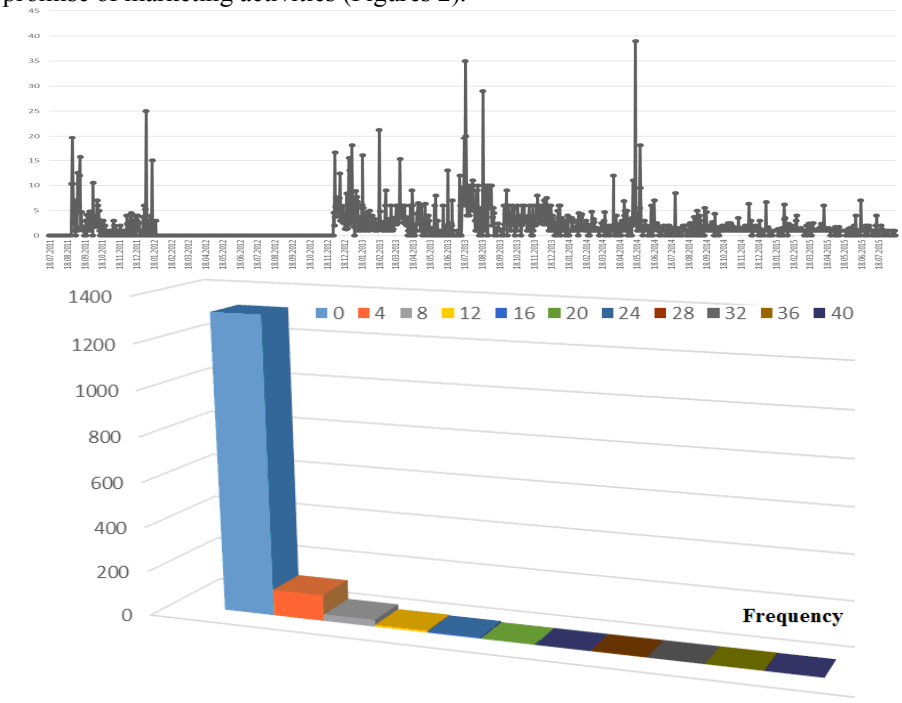

Fig. 2. For information resource victana.lviv.ua visiting pages per session

6 Discussion

The presence of different types ECCS increases in sales of goods and / or services a regular user of 9%, the active involvement of unique visitors, potential users and expanding the boundaries of the target and regional audience by 11%, page views by 12%, time of visit information resources on 7%. So cumulative and median filtering for information resource victana.lviv.ua visit pages per session is shown in Fig. 3. For an agglomeration hierarchical cluster analysis, it is advisable to choose the strategy of the "closest neighbor" for unification (since the proximity matrix consists of only 5 objects). The distance between the two groups is defined as the distance between the two closest elements of these groups (Figure 4). This strategy is monotonous and greatly compresses the space of signs.

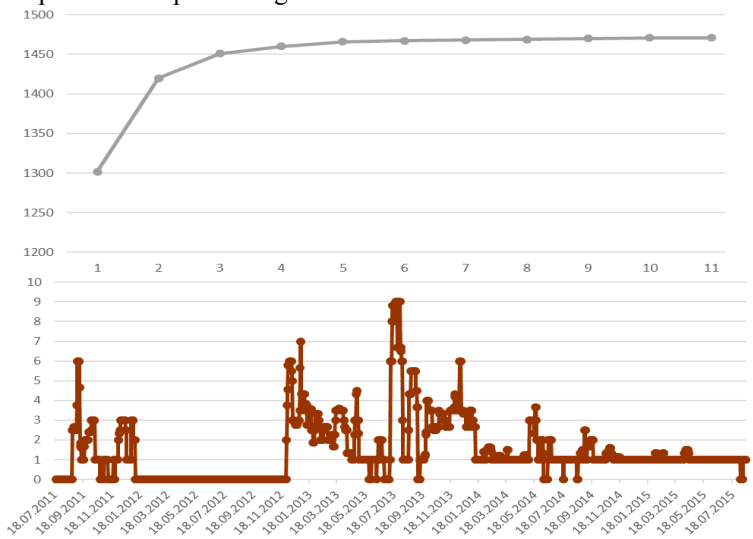

Fig. 3. Cumulate and Median Filtering to visit information resource victana.lviv.ua

	2011	2012	2013	2014	2015
2011	0	1,5	2,1	2,3	2
2012	1,5	0	2,4	1,6	1,8
2013	2,1	2,4	0	2	2,8
2014	2,3	1,6	2	0	2,4
2015	2	1,8	2,8	2,4	0

	[2011,2012]	2013	2014	2015
[2011,2012]	0	2,1	1,6	1,8
2013	2,1	0	2	2,8
2014	1,6	2	0	2,4
2015	1,8	2,8	2,4	0

	[2011,2012,2014]	2013	2015
[2011,2012,2014]	0	2	1,8
2013	2	0	2,8
2015	1,8	2,8	0

	[2011,2012,2014,2015	2013
[2011,2012,2014,2015]	0	2
2013	2	0

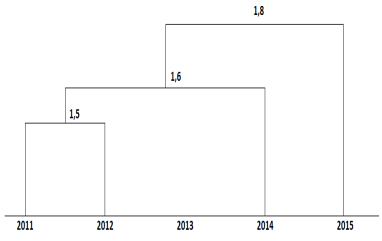

Fig. 4. Stages of the “closest neighbor” strategy and the dendrogram for visiting resource victana.lviv.ua

We will now analyze the information resource tatjana.in.ua from the ECCS type B2C (Figure 5). In fig. 6-7 gives the result of finding the trend in a time series of smoothing methods for the information resource tatjana.in.ua without B2C.

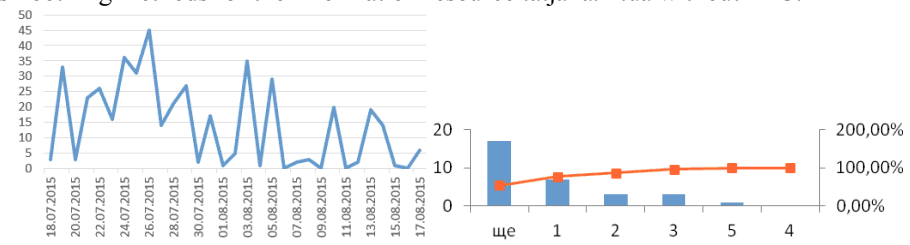

Fig. 5. Schedule and cumulative of visits

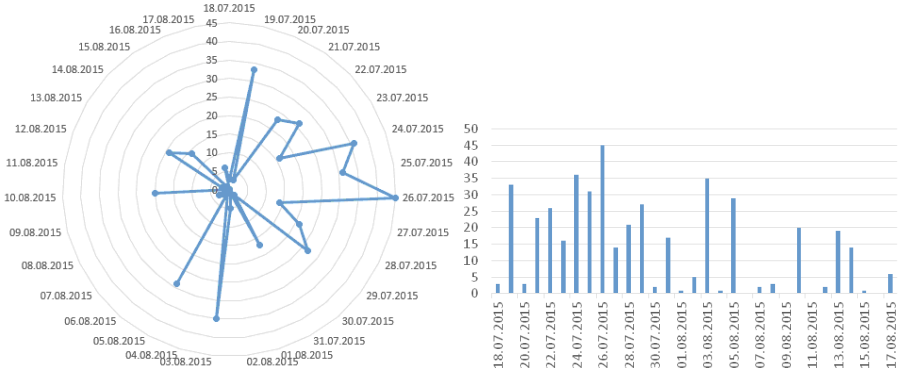

Fig. 6. Visit schedule and histogram

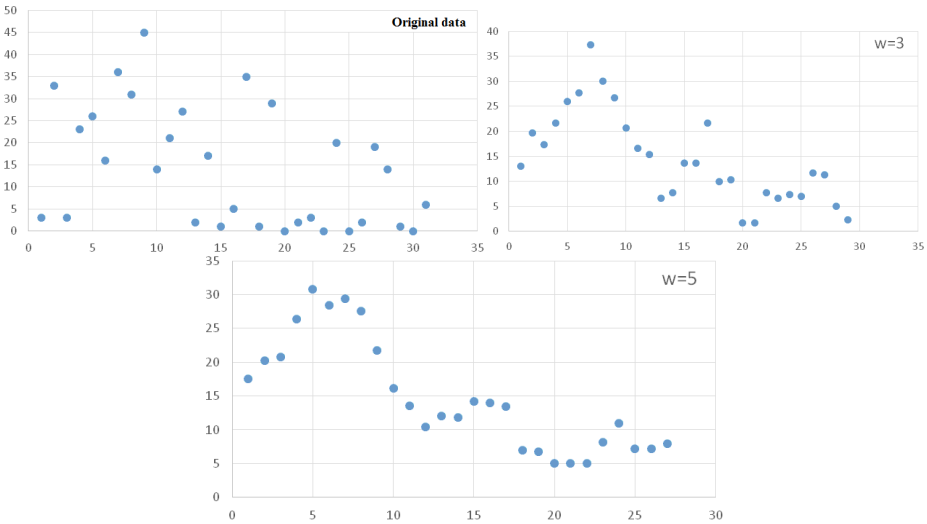

Fig. 7. Smoothing schedule for the visit

In fig. 8-9 results of median filtration are presented. To construct graphs and histograms, the largest and smallest smoothing interval is selected, so that the dependence of the degree of data smoothing depending on the interval can be seen. In fig. 10-12 expositional smoothing is given.

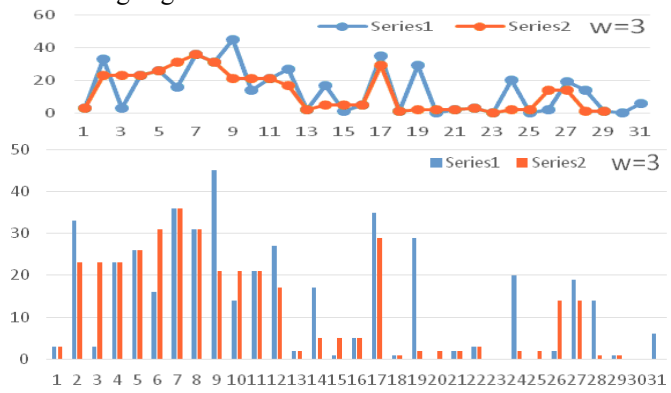

Fig. 8. Graph and histogram at interval 3

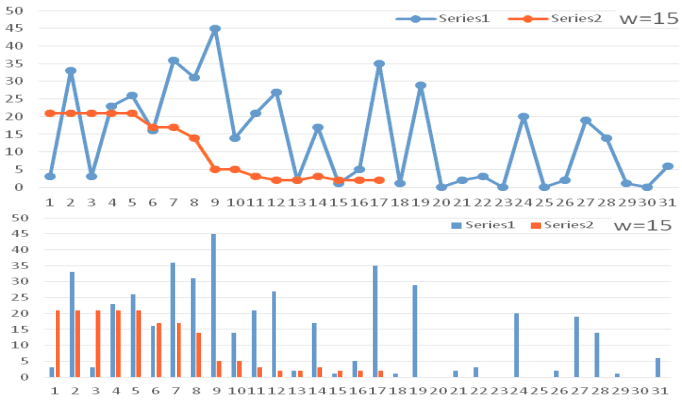

Fig. 9. Graph and histogram at interval 15

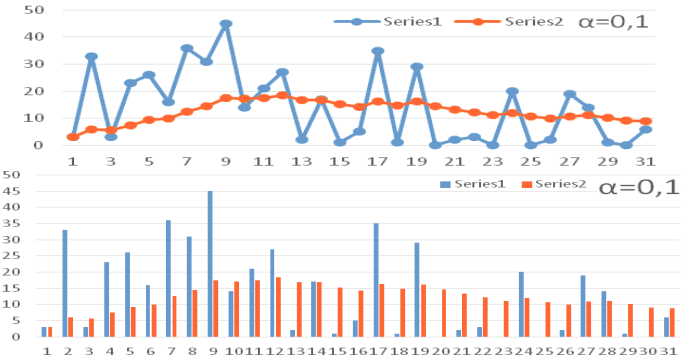

Fig. 10. Graph and histogram for α = 0,1

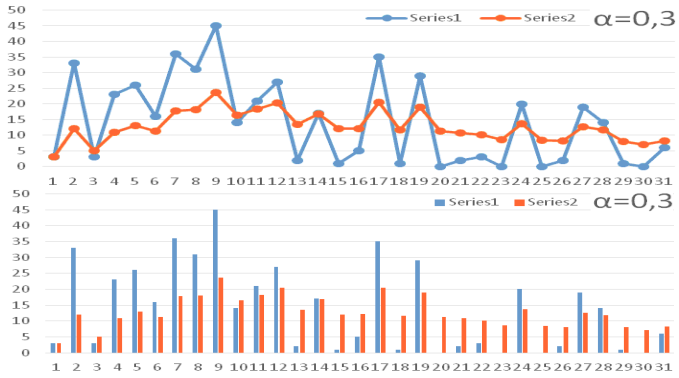

Fig. 11. - Graph and histogram at α = 0,3

In fig. 12-13 provides smoothing formulas from Kendel. In fig. 14 - dendrogram, and in fig. 15 is a correlation field.

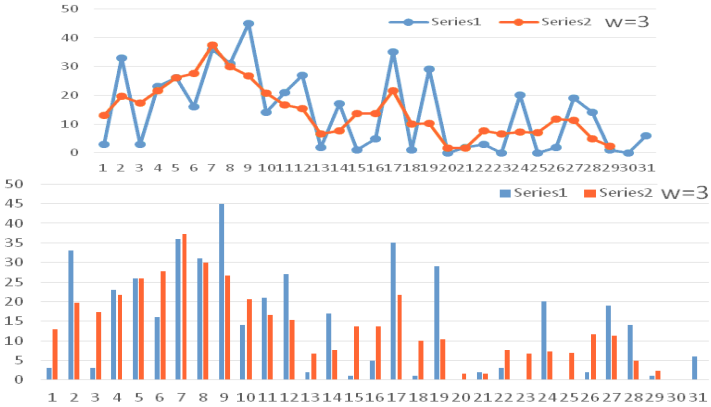

Fig. 12. Graph and histogram at w = 3

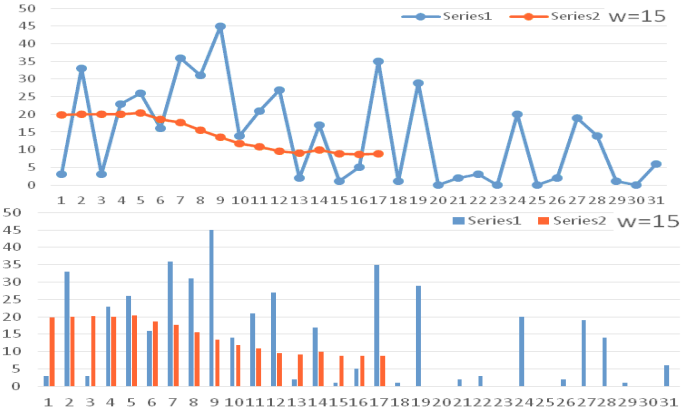

Fig. 13. Graph and histogram at w = 15

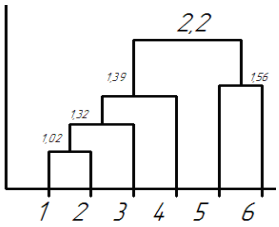

Fig. 14. Visit the dentogram tatjana.in.ua

We will now analyze the functioning of the information resource vgolos.com.ua as an example of the ECCS (Figure 16).

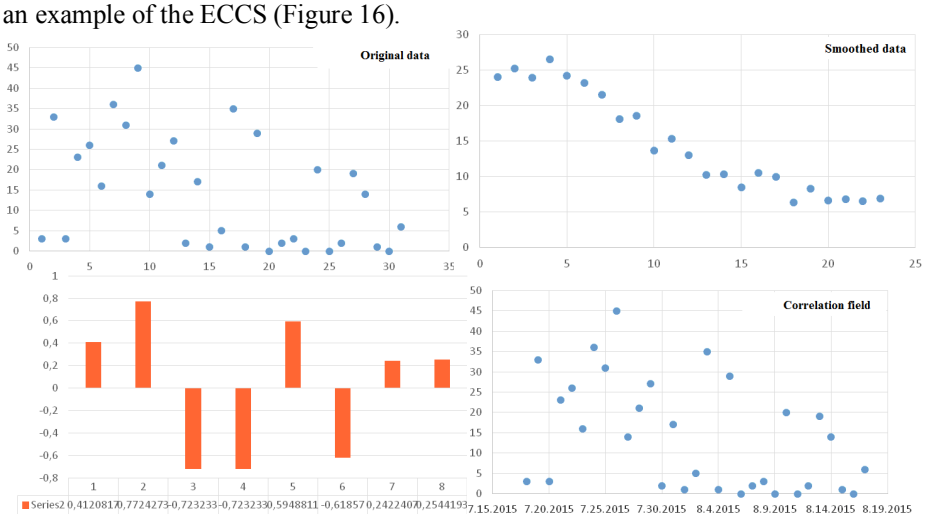

Fig. 15. Correlation field at tatjana.in.ua

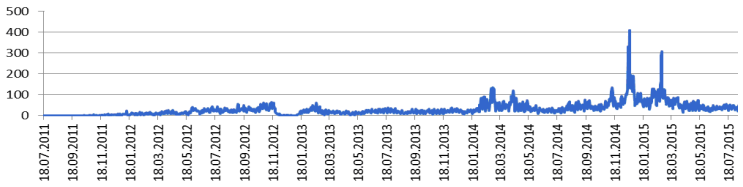

Fig. 16. Information resource vgolos.com.ua

To construct curls (Fig. 17) a table is analyzed for the computation of the frequency of sorted data. On the basis of the correlation coefficients for each table with the values of the smoothed functions, the corresponding correlation matrices were constructed (Figures 18-20). From the provided data, a property object table is generated. Data was split by years (2011-2015). For a graphical representation, a dendrogram has been constructed that shows the formed clusters (Fig. 21).

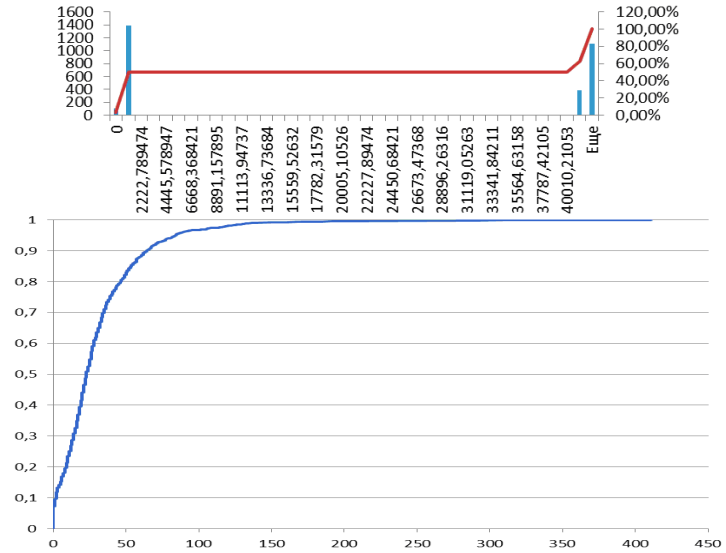

Fig. 17. Analysis of the redistribution of vgolos.com.ua

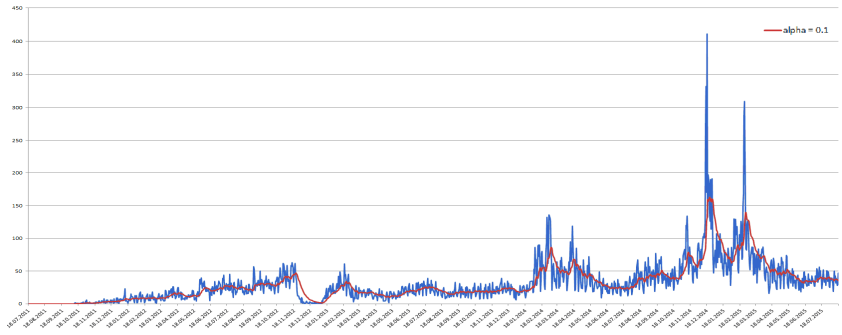

Fig. 18. Exponential smoothing with the parameter α = 0.1 for vgolos.com.ua

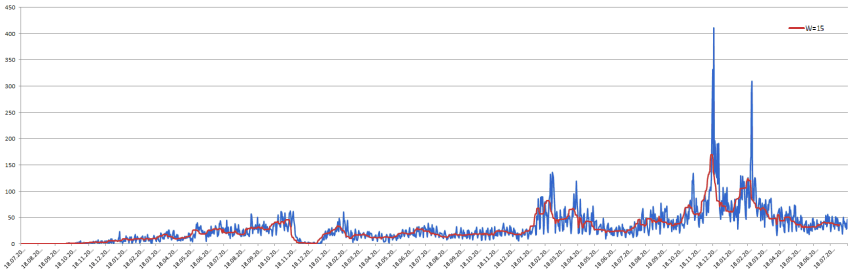

Fig. 19. Median smoothing at w = 15 for vgolos.com.ua

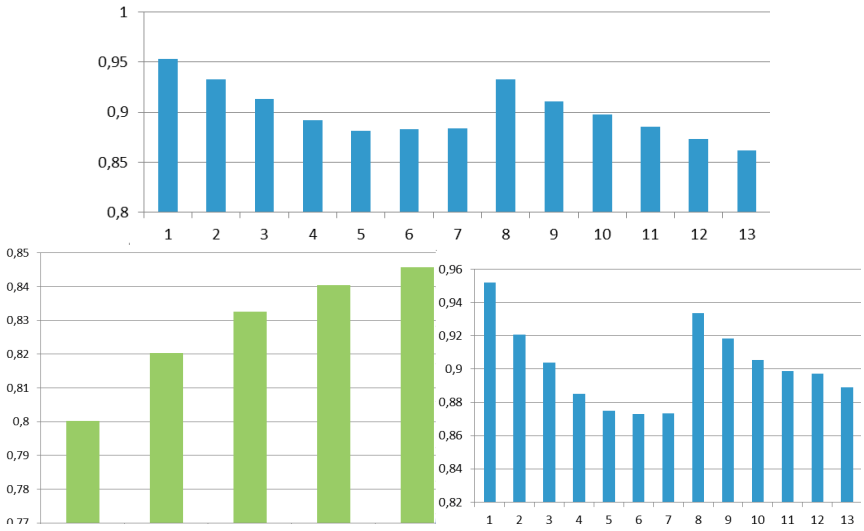

Fig. 20. Corelograms for smoothing visits to vgolos.com.ua

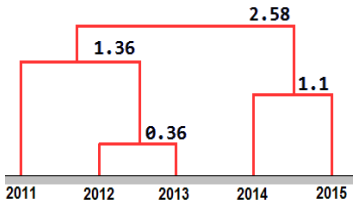

Fig. 21. Dendogram for vgolos.com.ua

7 Conclusion

This paper solves the problem of current scientific research and development of methods and means of information resources processing in ECCS. In the upshot used the classification, mathematical and software architecture and overall ECCS. In the paper the terminology and electronic content commerce systems are analyzed and classified to determine their characteristic patterns, trends, process design and simulation. The authors developed a formal model ECCS for shortcomings determining of existing methods and means of resources processing. They developed also standardized methods for information resources processing in ECCS. The architecture of modules is implementing stages of commercial content lifecycle in ECCS is proposed. Software for information resources processing increased attendance of electronic content commerce systems is suggested. From the perspective of systemic approach is applying the principles of information resources processing in ECCS. This made it possible to develop methods for the commercial content formation, management and support. The proposed integrated method of content support makes it possible to develop a module of commercial content realization. In the present paper, the

main regularities of ECCS classes and models opportunities are defined. The authors identified the transition from the information resources processing for their implementation. Formal models of the information resources formation, management and implementation are proposed. They allow realizing optimally the architecture of ECCS. The most significant factor taken into account in the formalization and algorithm models is the use of research tools for hardware and software IT.

References

1. Mobasher, B.: Data mining for web personalization. In: The adaptive web, Springer, 90-135 (2007)
2. Dinucă, C.E., Ciobanu, D.: Web Content Mining. In: University of Petroşani, Economics, 85 (2012)
3. Kravets, P.: The control agent with fuzzy logic, Perspective Technologies and Methods in MEMS Design, MEMSTECH'2010, 40-41 (2010)
4. Basyuk, T.: The main reasons of attendance falling of internet resource, Computer Science and Information Technologies, 91-93 (2015)
5. Pospelov, D.: Situatsionnoe upravlenie: teoriya i praktika (1986)
6. Golota, Y.: Logika antonimov i nechetkaya logika: skhodstva i razlichiya, Soft Computing and Measurement, 208-210 (1998)
7. Grinyaev, S.: Nechetkaya logika, www.computerra.ru/offline/2001/4.
8. Zade, L." Ponyatie lingvisticheskoy peremennoy i ego primenenie, (1976)
9. Osgood, C.: The nature and measurement of meaning, Psychological Bulletin, vol. 49, 197-237 (1952)
10. Peleshko, D., Rak, T., Izonin, I.: Image Superresolution via Divergence Matrix and Automatic Detection of Crossover. In: International Journal of Intelligent Systems and Application, 8(12), 1-8 (2016)
11. Rashkevych, Y., Peleshko, D., Vynokurova, O., Izonin, I., Lotoshynska, N.: Single-frame image super-resolution based on singular square matrix operator. In: IEEE 1th Ukraine Conference on Electrical and Computer Engineering (UKRCON), 944-948 (2017)
12. Rusyn, B., Lutsyk, O., Lysak, O., Lukeniuk, A., Pohreliuk, L.: Lossless Image Compression in the Remote Sensing Applications In: Proc. of the IEEE First Int. Conf. on Data Stream Mining & Processing (DSMP), 195-198 (2016)
13. Kowalik, Dagmara: Polish vocational competence standards for the needs of audult education and the Europen labour marke. In: International Conference on Advanced Information Engineering and Education Science, ICAEES, 95-98 (2013)
14. Mykich, K., Burov, Y.: Uncertainty in situational awareness systems, Modern Problems of Radio Engineering, Telecommunications and Computer Science, TCSET, 729-732 (2016)
15. Mykich, K., Burov, Y.: Algebraic Framework for Knowledge Processing in Systems with Situational Awareness, Advances in Intelligent Systems and Computing, 217-228 (2016)
16. Davydov, M., Lozynska, O.: Information System for Translation into Ukrainian Sign Language on Mobile Devices. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 48-51 (2017).
17. Davydov, M., Lozynska, O.: Linguistic Models of Assistive Computer Technologies for Cognition and Communication. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 171-175 (2017)
18. Su, J., Vysotska, V., Sachenko, A., Lytvyn, V., Burov, Y.: Information resources processing using linguistic analysis of textual content. In: Intelligent Data Acquisition and Advanced Computing Systems Technology and Applications, Romania, 573-578, (2017)

19. Lytvyn, V., Vysotska, V., Pukach, P., Vovk, M., Ugryn, D.: Method of functioning of intelligent agents, designed to solve action planning problems based on ontological approach. In: Eastern-European Journal of Enterprise Technologies, 3/2(87), 11-17 (2017)
20. Vysotska, V., Chyrun, L., Chyrun, L.: Information Technology of Processing Information Resources in Electronic Content Commerce Systems, Computer Science and Information Technologies, 212–222 (2016)
21. Vysotska, V., Hasko, R., Kuchkovskiy, V.: Process analysis in electronic content commerce system. In: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT 2015, 120-123 (2015)
22. Gonçalves, M.A., Fox, E.A., Watson, L.T., Kipp, N.A.: Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. In: ACM Transactions on Information Systems (TOIS), 22(2), 270-312 (2004)
23. Gonçalves, M.A.: Streams, Structures, Spaces, Scenarios, and Societies (5S): A Formal Model for Digital Library Framework and Its Applications. PhD thesis, Virginia Polytechnic Institute and State University, (2004).
24. Pérez, A., Enrech, M.: Defining library services for a virtual community. In: Libraries Without Walls Conference, Lesvos, Grecia, Centre for research in Library and Information Management (1999)
25. Pérez, A., Enrech, M.: Virtual Library Services for a Virtual University: User-Oriented Virtual Sites in an Open Library. In: EADTU, Paris, 1999.
26. Stoffle, C.J.: The Emergence of Education and Knowledge Management as Major Functions of the Digital Library. In: Follet Lecture Series, http://www.ukoln.ac.uk/services/papers/follett/stoffle/paper.html.
27. Tkachenko, R., Tkachenko, P., Izonin, I., Tsymbal, Y.: Learning-based image scaling using neural-like structure of geometric transformation paradigm. In: Studies in Computational Intelligence, 730, Springer Verlag, 537–565 (2018)
28. Maksymiv, O., Rak, T., Peleshko, D.: Video-based Flame Detection using LBP-based Descriptor: Influences of Classifiers Variety on Detection Efficiency. In: International Journal of Intelligent Systems and Applications, 9(2), 42-48 (2017)
29. Chernukha, O., Bilushchak, Y.: Mathematical modeling of random concentration field and its second moments in a semispace with erlangian disrtibution of layered inclusions. In: Task Quarterly, 20(3), 295-334 (2016)
30. Zhezhnych, P., Markiv, O.: Linguistic Comparison Quality Evaluation of Web-Site Content with Tourism Documentation Objects. In: Advances in Intelligent Systems and Computing, 689, 656-667 (2018)
31. Mobasher, B.: Data mining for web personalization. In: The adaptive web, Springer, 90-135 (2007)
32. Dinucă, C.E., Ciobanu, D.: Web Content Mining. In: University of Petroşani, Economics, 85 (2012)
33. Xu, G., Zhang, Y., Li, L.: Web content mining. Web Mining and Social Networking. In: Springer, 71–87 (2011)
34. Jivani Anjali Ganesh: A Comparative Study of Stemming Algorithms. In: Int. J. Comp. Tech. Appl., 6, 1930-1938 (2011)
35. Chen, J., Dosyn, D., Lytvyn, V., Sachenko, A.: Smart Data Integration by Goal Driven Ontology Learning. In: Advances in Big Data. Advances in Intelligent Systems and Computing. – Springer International Publishing AG 2017. P. 283-292 (2017).
36. Basyuk, T.: The Popularization Problem of Websites and Analysis of Competitors". In: Advances in Intelligent Systems and Computing II. CSIT 2017. Advances in Intelligent Systems and Computing, vol 689. Springer, Cham pp. 54-65 (2017)