MEST Journal
DOI 10.12709/mest.02.02.02.29

ANALYTICAL METHODS FOR COMMERCIAL WEB CONTENT PROCESSING OF INFORMATION RESOURCE IN ELECTRONIC BUSINESS SYSTEMS

Victoria Vysotska
Lviv Polytechnic National University, Lviv, Ukraine

Lyubomyr Chyrun
Lviv Polytechnic National University, Lviv, Ukraine

JEL category: D84, L86

Abstract

The given article is devoted to the development of standardized analytical methods and software for Web content processing in e-business systems. The model of electronic content commerce systems is developed. The models of commercial Web content processing subsystems are constructed. A new approach of business processes application and implementation for the construction of electronic content commerce systems is formulated. Complex methods of formation, management and support commercial Web content are developed. Software for information resources processing in electronic content commerce systems is developed. The methods of e-business systems designing and implementation are presented as example on online newspaper and online magazine that reflect the theoretical studies results.

Keywords: *content, web content, information resource, business process, online newspaper, online magazine, e-commerce systems, e-business.*

1 INTRODUCTION

Active development of the Internet promotes the needs growth in the operative reception of data for e-business conducting. It also facilitates the new forms implementation of information services through appropriate systems and information resources (Berko, Vysotska, & Pasichnyk, 2009), (Jerk, 2001), (Uspensky, 2001). Commercial Web content is documented and according formatted unique data. They are prepared to meet the needs of information resources users (Berko, Vysotska, & Pasichnyk, 2009), (Ermakov, & Kiselev, 2005), (Pasichnyk, Scherbyna, Vysotska, & Shestakevych, 2012), (Vysotska, & Chyrun, 2011). The characteristic feature of e-business systems is an opportunity to information resources processing. This increases the goods sales

Address of the corresponding author:
Victoria Vysotska
victana@bk.ru

volume of regular user, the active involvement of potential users and expanding the boundaries of the target audience (Jerk, 2001), (Evdokimov, 2007). The e-business principles and technology are used actively in the creation of electronic content commerce systems for on-line/off-line sales and Web content analysis/exchange/saving, online shop, cloud storage/computing, etc. (Ermakov, & Kiselev, 2005), (Ivanov, & Krukovskaya, 2004), (Ivanov, 2002), (Ivanov, 1994), (Clifton, 2009). The problems of design, development, implementation and maintenance of electronic content commerce systems are actual theme in view of factors such as: active development of research in the e-business area, the theoretical justification lack of standardized methods, and the need for software unification of information resources processing (Ivanov, & Krukovskaya, 2004), (Ivanov, 2002), (Clifton, 2009), (Korneev, Hareev, Vasyutyn, & Reich, 2000), (Krupnik, 2002). There are new approaches and solutions to this problem. But there is a mismatch between the known methods and means of commercial content processing and principles of e-commerce systems construction. Lacks are general approaches of e-commerce systems creation and standardized methods for commercial Web content processing in these systems.

2 RECENT RESEARCH AND PUBLICATIONS ANALYSIS

Practical factor of commercial Web content processing in e-business systems is closely linked with the problems addressing of the content amounts growth in the Internet, rapid prevalence of accessibility to the Internet, active development of e-business, set expanding of information products and services, growth of demand for information products and services, the technology and means creation, and areas expanding of methods application for information resources processing. In this direction are actively working leading global manufacturers of commercial content resources means. In particular, such systems are developed: for on-line content selling of online newspapers (New York Times, Washington Post), online journals (Wall Street Journal, Chip), distance learning (Ashford University), Internet publication as a dictionaries/directories (Oxford English Dictionary Online), Internet publishing (Online Publishers Association), portals of entertaining (tochka.net), informative (Weather Channel, Karpaty.ua) and children (teremoc.ru, vshkolu.com) content; for off-line content selling of copywriting services (Textbroker, Sopywriting 911), Marketing Services Shop (The Copy Box, Local Internet Marketing), RSS Subscription Extension (Apple, Google, Intel, Microsoft); online shop for selling of eBooks, software, video, music, movies, picture, digital art, manuals, articles, certificates, forms, files, etc (Google, Amazon, Apple, Android, Opera, Yahoo!); for content saving on cloud storage and cloud computing (Google, Amazon, Apple, Micorsoft, Linux, iPhone, Android, Palm).

The theoretical factor of information resources processing in e-business systems is associated with the methods and means development of commercial Web content formation, management and support. In scientific studies D. Lande is mathematical models of electronic information flow studied and developed (Bolshakova, Lande, Noskov, Klyshynskyy, Peskova, & Yahunova, 2011), (Braychevskyy, & Lande, 2005), (Grigoriev, & Lande, 2005), (Lande, Furashev, Braychevskyy, & Hryhorev, 2006), (Lande, 2005-2006), (Lande, & Braychevskyy, 2005-2006), (Lande, & Litvin, 2001), (Lande, & Morozov, 2004-2005), (Lande, & Furashev, 2006), (Lande, Furashev, Braychevskyy, & Hryhorev, 2006), (Lande, Furashev , & Grigor'yev, 2006), (Furashev, Lande, Grigor'yev , & Furashev, 2005), (Furashev, Lande, & Brajchevskiy, 2005). G. Zipf is proposed an empirical regularity of words frequency distribution in natural language (Zipf 1935, 1949). And G. Salton and R. Papka is proposed new events detection in the content flow (Salton, 1979), (Papka, 1999). B. Boiko, S. McKeever, A. Rockley, G. McGovern, J. Hackos, R. Nakano, B. Doyle, Halverson are a model of the commercial content lifecycle in the works described (Boiko, 2005), (Doyle, 2005), (Hackos, 2002), (Halverson, 2009), (McGovern, & Norton, 2001), (McKeever, 2003), (Nakano, 2002), (Rockley, & Cooper, 2002), (Stone, 2003), (Vysotska, & Chyrun, 2011). In works (Pasichnyk, Scherbyna, Vysotska, & Shestakevych, 2012), (Soroka, & Tanatar, 1998), (Fedorchuk, 2005) are the content analysis methodology launched. EMC, IBM, Microsoft Alfresco, Open Text, Oracle and SAP have developed a Content Management Interoperability

Services specification for Web-services Berko, Vysotska, & Pasichnyk, 2009). They are designed to provide interaction between subsystems of commercial Web content processing in e-business systems. Interoperable Content Application Tools interacts with content from different repositories via a service interface and special superstructure as Content Management Interoperability Services Implementation.

3 PROBLEMS EXTRACTION

The number of Web content streams in e-business systems is greater than the path of goods movement in the industry. Considerable part of content flow in electronic business systems consists of easily formalized and automated procedures. The main problem is the lack of a common approach to process of e-business systems modeling, design and development for the rapid processing of commercial content.

Model of e-business system is given as

$$S = \langle X, Q, C, V, H, Function, T, Z, Y \rangle, \qquad (1)$$

where:

- $X = \{x_1, x_2 \ldots, x_{n_X}\}$ – content set from various sources;
- $Q = \{q_1, q_2 \ldots, q_{n_Q}\}$ – set of information users requests;
- $C = \{c_1, c_2 \ldots, c_{n_C}\}$ – commercial Web content set;
- $V = \{v_1, v_2 \ldots, v_{n_V}\}$ – set of content support conditions;
- $H = \{h_1, h_2 \ldots, h_{n_H}\}$ – the conditions set of content formation and management,
- $Z = \{z_1, z_2 \ldots, z_{n_Z}\}$ – the components set of an information resource,
- $T = \{t_1, t_2 \ldots, t_{n_T}\}$ – transaction time of content processing,
- $Y = \{y_1, y_2 \ldots y_{n_Y}\}$ – statistics data set of system operation,
- $Function$ – the operator of the output data formation in the system (Berko, Vysotska, & Pasichnyk, 2009), (Sovetov, & Yakovlev, 1998).

The process of the e-business system operation is described by the S operator as:

$$y_j(t_p + \Delta t) = Function(x_i, q_d, c_r, v_l, h_k, t_p, z_w) \quad (2)$$

where x_i – visitors/users requests to the e-business system. Value y_j is described as:

$$y_j = \{a_1, a_2, \ldots, a_g\}$$

where

- a_1 – visits number over a time period Δt,
- a_2 – the average time spent on an information resource (min : c) for a time period Δt,
- a_3 – a refusals indicator (%) for a period time Δt,
- a_4 – achieved purpose of commercial content finding,
- a_5 – the dynamics of commercial content lifecycle (%),
- a_6 – the total number of page views for a time period Δt,
- a_7 – the number of page views per visit,
- a_8 – new visits,
- a_9 – absolute unique visitors,
- a_{10} – traffic sources (direct referrals, referrals from search engines, referrals from other sites, etc.) in % etc. (Clifton, 2009).

The value influences x_i, q_d, c_r, v_l, h_k on z_w, and y_j value as a result of the e-business system operation are unknown and unexplored. The model does not reveal relationships between the input data, commercial Web content, the original data and the information resources processing in e-business system S. This substantiates the purpose, relevance, appropriateness and research directions of the commercial content processing in e-business systems. It is necessary to solve the problems of the Web content formation, management and support as components of the information resources processing in e-business systems in the form of theory-based concept.

4 GOALS FORMULATION

This work aim is to expand the functionality and the efficiency improving of e-business systems. This is achieved by system attendance increasing through the use of standardized methods and software for commercial Web content processing.

The purpose is identified the need to the following tasks solving:

- models creation of e-business systems to the shortcomings determine of existing methods and tools for commercial Web content processing;
- standardized methods development of commercial Web content processing for the architecture designing of e-business systems;
- a generalized architecture development of e-business system for stages implementation of the commercial Web content lifecycle;
- software implementation for commercial Web content processing in e-business systems (for the time and costs reducing on commercial content formation, management and support) by the quality increasing of information resources processing.

5 OBTAINED SCIENTIFIC RESULTS ANALYSIS

The main stages of the commercial Web content processing in e-business systems is the content formation, management and support with the following links: *Web content* $\rightarrow$ *Web content formation* $\rightarrow$ *database* $\rightarrow$ *Web content management* $\rightarrow$ *information resource* or *user request* $\rightarrow$ *Web content management* $\rightarrow$ *information resource* $\rightarrow$ *Web content support* $\rightarrow$ *database*, i.e.

$$Function = Formation \cup Management \cup Support \quad (3)$$

where $Function$ – the operator of the output data formation in the system, $Formation$ – the operator of the Web content formation, $Management$ – operator of Web content management, $Support$ – operator of Web content support.

Under the model of e-business system is presented as

$$S = \left\langle \begin{matrix} X, Q, Formation, H, C, V, \\ Management, Support, Z, T, Y \end{matrix} \right\rangle \quad (4)$$

where X – Web content set from various sources, Q – set of user queries, $Formation$ – the operator of the content formation, H – the conditions set of Web content formation and management, C – set of commercial Web content, V – the conditions set the content support and external influences on the system, $Management$ – Web content management operator, $Support$ – Web content support operator, Z – the components set of an information resource, T – transaction time of information resources processing, Y – statistical data set of the system operation.

Phase of commercial Web content formation is described operator $Formation$ as follows

$$c_r = Formation(u_f, x_i, t_p) \quad (5)$$

where u_f – conditions set of commercial Web content formation, ie. $u_f = \{u_1(x_i), \ldots, u_{n_U}(x_i)\}$.

Commercial Web content is presented as follows:

$$c_r = \left\{ \bigcup_f u_f \left| \begin{matrix} (x_i \in X) \wedge (\exists u_f \in U), \\ U = U_{x_i} \vee U_{\overline{x_i}}, i = \overline{1,m}, f = \overline{1,n} \end{matrix} \right. \right\} \quad (6)$$

Stage of commercial Web content managing is described operator $Management$ as follows

$$z_w = Management(q_d, c_r, h_k, t_p) \quad (7)$$

where Q – set of information users requests, H – set of commercial Web content management conditions, ie. $H = \{h_1(c_{i+1}, q_d), \ldots, h_{n_H}(c_{i+n_H}, q_d)\}$.

Commercial Web content management is presented as

$$z_w = \left\{ \bigcup_{k=1}^{n_H} h_k(c_{i+1}, q_d) \left| \begin{matrix} (c_{i+k} \in C) \wedge (q_d \in Q) \wedge \\ \wedge (h_k \in H_q), \\ H = H_{q_d} \vee H_{\overline{q_d}}, i = \overline{1,n_C}, \\ d = \overline{1,n_Q}, k = \overline{1,n_H} \end{matrix} \right. \right\} \quad (8)$$

Stage of commercial Web content support is described operator $Support$ as follows

$$y(t_p + \Delta t) = Support(v_l, h_k, c_r, z_w, t_p, \Delta t) \quad (9)$$

where v_l – conditions set of commercial Web content support and the environment impact on the system, i.e.

$$v_l = (v_1(q_i, h_k, c_r, z_w, t_p), \ldots, v_{n_V}(q_i, h_k, c_r, z_w, t_p))$$

Output data are implemented

$$y_j = \left\{ \bigcup_l v_l \left| \begin{matrix} (\exists q_d \in Q) \wedge (\exists z_w \in Z) \wedge \\ \wedge (\forall v_l \in V) \wedge (\forall (c_r \wedge q_d) \in h_k), \\ V = V_{q_d} \vee V_{\overline{q_d}}, d = \overline{1,n_Q}, l = \overline{1,n_V}, \\ w = \overline{1,n_Z}, r = \overline{1,n_C}, k = \overline{1,n_H} \end{matrix} \right. \right\} \quad (10)$$

A complete functional electronic business system is characterized a complex system of interrelated operations, methods and techniques of commercial Web content processing. This is shown in Fig. 1.

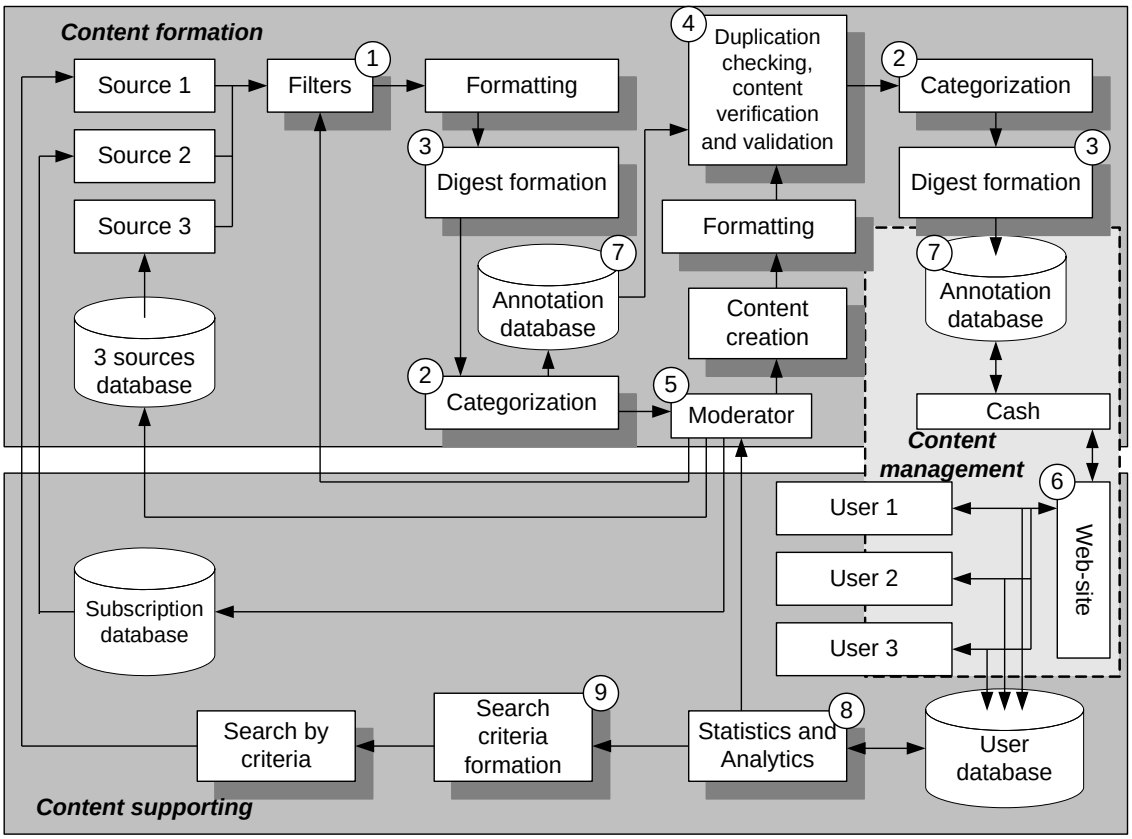


Fig. 1. Methods of commercial Web content processing in e-business systems

6 METHOD OF COMMERCIAL WEB CONTENT FORMING

The commercial Web content formation for an information resource provides communication between inputs data set from different sources and generated content set that is stored in the appropriate database of e-business systems, i.e.:

$$Source(x_i) \to x_i \to X \to Formation(u_f, x_i, t_p) \to c_r \to C \to DataBase(C),$$

where:

- $Source(x_i)$ – the x_i content source,
- x_i – the *i*-th content from data source,
- X – content set from corresponding sources of data,
- $Formation(u_f, x_i, t_p)$ – the operator of the a commercial content formation at x_i in a fixed time t_p at the conditions u_f, c_r – formed under conditions u_f of *r*-th commercial content,
- C – a set of commercial content,
- $DataBase(C)$ – database of commercial Web content.

Model of commercial Web content formation is presented as

$$Formation = \left\langle \begin{matrix} X, Gathering, Formatting, \\ KeyWords, Backup, \\ Caterization, BuDigest, \\ Dissemination, T, C \end{matrix} \right\rangle \quad (11)$$

where:

- X – set of input data $x_i \in X$ from different information resources or moderators;
- $Gathering$ – operator of content collecting/creating from various sources;

- $Formatting$ – content formatting operator; $KeyWords$ – the operator of content keywords and concepts identification;
- $Caterization$ – the operator of automatic categorization of content;
- $Backup$ – the operator of content duplication identification;
- $BuDigest$ – the operator of the content digest formation;
- $Dissemination$ – operator of selective distribution of Web content;
- T – the transaction time of $t_p \in T$ the content formation;
- C – a Web content set $c_r \in C$.

The optimal solution can help navigate in the dynamic input data from different sources. This optimal solution provide the process of information gathering from the sources and its fragments distribution according to the users' needs $C = Gathering(X, U_G, T)$, where:

- X – content set from different sources of data;
- U_G – a conditions set of data gathering from a different sources;
- $Gathering$ – the operator of content gathering/creation;
- T – the content gathering/creation time.

Content duplication identification is described by the operator

$$C = Backup(Gathering(X, U_G, T), U_B) \quad (12)$$

where:

- X – content set from different sources of data,
- U_B – conditions set content duplication identification,
- $Backup$ – the operator of content duplication identification,
- C – a content set.

Content duplication identification in the e-business systems implements linguistic statistical methods of the general terms searching, which form chain of a verbal signatures in content.

Content syndicate is in data gathering programs training of structural features at individual sources (with information resources, by moderators, users, visitors, journalists, and/or editors), direct scanning of commercial content and his addiction to a common XML format:

$$C = Formatting(Backup(Gathering(X, U_G, T), U_B), U_{FR}) \quad (13)$$

where:

- $Formatting$ – the operator of content formatting,
- U_{FR} – a conditions set of commercial content formatting.

The content set processing C for meaningful keywords identification are based on the principles of the keywords finding in content (terms). This are based on Zipf's law and reduced to the words choice which an average frequency of appearance (the most used words are ignored by stop-dictionary using and rare words from messages text not included). The keywords and concepts identification are defined by the operator:

$$C = KeyWords(Formatting(Backup(Gathering(X, U_G, T), U_B), U_{FR}), U_K) \quad (14)$$

where:

U_K – conditions set of key words and concepts identification. Commercial Web content classification and distribution implement through a information retrieval system of selective content distribution (the containing router). Content is analyzed content categorization operator

$$C_{Ct} = Categorization(KeyWords(C, U_K), U_{Ct}) \quad (15)$$

for compliance with thematic requests, where:

- $Categorization$ – operator of content categorization according to the detected keywords,
- U_{Ct} – a conditions set of an automatic categorization,
- C_{Ct} – set of classified relevant content.

The digest set C_D form depends of $C_D = BuDigest(C_{Ct}, U_D)$,

where:

– $BuDigest$ – the digest formation operator

– U_D – conditions set of digest formation, i.e.

$$C_D = BuDigest(Categorization(KeyWords(C, U_K), U_{Ct}), U_D) \quad (16)$$

Relevant content send users and are loaded into the database. Selective distribution of commercial content is described

$$C_{Ds} = Dissemination(C_D, U_{Ds}) \quad (17)$$

7 METHOD OF COMMERCIAL WEB CONTENT MANAGEMENT

The process of commercial Web content management is presented this general scheme links:

$$User(q_d) \to q_d \to Q \to H(c_{i+1}, q_d) \to$$
$$Management(q_d, c_r, h_k, t_p) \to z_w \to User(z_w)$$

Methods for commercial content management is divided into different classes according to the model of pages generating of information resources in e-business systems.

1. Model of commercial Web content management such as pages generating by the user requests is presented as

$$Management_Q = \left\langle \begin{matrix} H, C, Q, Presentation, \\ Edit, Weight, T, Z \end{matrix} \right\rangle \quad (18)$$

Where:

– C – Web content set,
– Z – set of generated pages on information resource,
– Q – a set of user queries,
– $Presentation$ – the operator page formulation and presentation on information resource,
– $Edit$ – operator of content editing and modification,
– $Weight$ – the weight of the content block, T – transaction time of commercial Web content management,
– H – a conditions set of commercial content management.

Step of content editing and modification is presented by the operator $c_j(t_{r+1} + \Delta t) = Edit(c_j(t_r), t_{r+1}, \Delta t)$ Stage of pages set forming is described by operator:

$$Z(t_r + \Delta t) = Presentation(q_i, C, Weight(C), t_r, \Delta t)$$

where:

– $Weight(C)$ – the total weight of content block on information resource, i.e.

$$z_i = \left\{ \begin{matrix} \bigcup c_j(q_i, t_r) \left| (\forall c_j \in C_q) \wedge (\exists q_i \in C_q) \wedge \right. \\ \wedge Weight(C_q), \\ C = C_q \vee \bar{C_q}, j = \overline{1,m}, i = \overline{1,n} \end{matrix} \right\} \quad (19)$$

2. Model of pages generation while the information resource editing in the system is presented as

$$Management_E = \langle C, H, Edit, Weight, T, Z \rangle \quad (20)$$

where:

– C – Web content set,
– H – conditions set of content management,
– Z – static pages set of information resource,
– $Weight$ – the weight of the Web content block,
– $Edit$ – operator of Web content editing/updating.

Phase of pages formation is described operator

$$Z(t) = Edit(C, Weight, H, t) \quad (21)$$

3. Model of mixed management of commercial content is presented as

$$Management_M = \left\langle \begin{matrix} C, Q, H, Presentation, \\ Edit, Weight, Caching, T, Z \end{matrix} \right\rangle \quad (22)$$

where:

– C –content set,
– Z – generated pages set of information resource,
– Q – the set of user queries,
– H – the set of content management conditions,
– $Presentation$ – the operator of page formulation and presentation on information resource,
– $Edit$ – operator of content editing and modification,
– $Caching$ – the operator of the cache or information content blocks formation where the cache is presented as

$$Cache = Caching(C, Weight, t, \Delta t) \quad (23)$$

or

$$Cache = Caching(Z, Weight, t, \Delta t) \quad (24)$$

i.e.

$$Cache = \left\{ \bigcup c_i \left| \begin{matrix} (\forall c_i \in C) \wedge Weight(c_i), \\ t = t + \Delta t, i = \overline{1,n} \end{matrix} \right. \right\} \quad (25)$$

$$Cache = \left\{ \bigcup z_i \left| \begin{matrix} (\exists z_i \in Z) \wedge (\forall c_j \in C) \wedge \\ \wedge Weight(c_j) \wedge (\forall c_j \in z_i), \\ t = t + \Delta t, i = \overline{1,n} \end{matrix} \right. \right\} \quad (26)$$

8 METHOD OF COMMERCIAL WEB CONTENT SUPPORTING

Process of commercial Web content support is presented as follows scheme links

$$User(q_d, z_w) \rightarrow q_d \rightarrow z_w \rightarrow V(q_d, z_w, t_p) \rightarrow Support(v_l, h_k, c_r, z_w, t_p, \Delta t) \rightarrow y_j \rightarrow Moderator(y_j).$$

The method of commercial Web content support looks like

$$Support = \left\langle \begin{matrix} Q, C, H, V, T, BuInfPort, \\ IdThemTop, ConCorrTablConc, \\ CalRankConc, Z, Y \end{matrix} \right\rangle \quad (27)$$

where:

- Q – set of user queries,
- C – the Web content set,
- $BuInfPort$ – the operator of the information portraits formation,
- $IdThemTop$ – the operator of thematic storyline identification,
- $ConCorrTablConc$ – operator content relationship tables building,
- $CalRankConc$ – the operator of content ratings calculation,
- H – set of internal parameters in the system,
- V – the parameters set of the external environment,
- Z – page set of information resource,
- T – set of transaction time for information resources processing,
- Y – set of statistical data of the system operation.

The analysis of results of the S e-business systems functioning and C commercial Web content support has formed set $Y = \{Y_P, Y_T, Y_C, Y_R\}$ under the conditions $V = \{V_P, V_T, V_C, V_R\}$,

where:

- $Y_P = Y_{Pc} \vee Y_{Pq}$ – a subset of the information portraits of Y_{Pc} content and Y_{Pq} users,
- Y_T – a subset of thematic storyline of content,
- Y_C – subset of content relationship tables,
- Y_R – a subset of the ratings content,
- $V_P = V_{Pc} \vee V_{Pq}$ – the conditions set of information portraits formation,
- V_T – a conditions set for thematic storyline identification,
- V_C – the conditions set of the content relationship construct tables,
- V_R –the parameters set of the content ratings calculation.

The information portraits set Y_{Pc} of content, presented as $Y_{Pc} = BuInfPort(V_{Pc}, C, H, Q, T)$, and Y_{Pq} set of the users' portraits are given as:

$$Y_{Pq} = BuInfPort(V_{Pq}, Q, H, Z, T) \quad (28)$$

where:

- $V_P = V_{Pc} \vee V_{Pq}$ – set of the portraits formation conditions,
- $BuInfPort$ – the operator of the portraits formation as $Y_P = Y_{Pc} \vee Y_{Pq}$.

The thematic storyline set for the Y_T content is presented as

$$Y_T = IdThemTop(C, H, Q, V_T, T) \quad (29)$$

where:

- V_T – conditions set for content storyline identification,
- $IdThemTop$ – the operator of Y_T thematic storyline definition.

The Y_C set of relationship content tables is presented as $Y_C = ConCorrTablConc(C, V_c, T)$, where:

- V_C –the conditions set of the relationship tables building,
- $ConCorrTablConc$ – tables constructing operator.

The Y_{Rc} set of the content ratings is presented as:

$$Y_{Rc} = CalRankConc(C, Q, H, Y_C, V_{Rc}, Spam, Tonality, T) \quad (30)$$

and a Y_{Pq} set of a moderators ratings is presented as

$$Y_{Rm} = CalRankConc(C, Q, H, Y_C, V_{Rm}, T) \quad (31)$$

where:

- $V_R = V_{Rc} \vee V_{Rm}$ – the parameters set for the content ratings calculation,
- $Tonality(Q^+, Q^0, Q^-, T, H)$ – the tonality criteria for commercial content,
- $Spam(Q, T)$ – operator of comments filtering definition,
- $CalRankConc$ – operator of content and moderators ratings identification $Y_R = Y_{Rc} \vee Y_{Rm}$.

The Y set of output statistical data is presented as

$$Y = \{Y_P, Y_T, Y_C, Y_R\} = \\ = Support(V, C, Q, H, Z, T, \Delta T) \quad (32)$$

$$Y = \{Y_P, Y_T, Y_C, Y_R\} = \\ = Support(V_P, V_T, V_C, V_R, C, Q, H, Z, T, \Delta T) \quad (33)$$

where

- $Y_P = Y_{Pc} \vee Y_{Pq}$ – the information portraits subset of content and users,
- Y_T – a subset of the thematic storyline in content,
- Y_C – a subset of content relationship tables,
- $Y_R = Y_{Rc} \vee Y_{Rm}$ – a subset of the content and moderators ratings,
- $Support$ – operator of content support.

The obtained data taken into account when an information resource creation or updating and the architecture improvement of e-business systems. Actuality of subsystem development for commercial content processing consists in the necessity obtain operational/objective assessment of the competition level in the financial market segment of commercial content, assess the competition level and their competitiveness degree in the financial market with the content distribution. In Table. 1 lists the developed systems with the lifecycle support for commercial content.

Table 1. Subsystems of commercial content processing in the developed e-business systems

№	Information Resource Name	System Appointment	Address	Subsystem availability		
				Formation	Manage-ment	Support
1	Fotoghalereja-vysocjkykh	Online shop	fotoghalereja-vysocjkykh.com	+/–	+	+/–
2	Vgolos	Online Newspaper	vgolos.com.ua	+	+	+
3	Tatjana	Online Newspaper	tatjana.in.ua	–	+/–	+/–
4	Presstime	Online Newspaper	presstime.com.ua	+/–	+	+
5	AutoChip	Online shop	autochip.vn.ua	–	+	+/–
6	Kursyvalyut	Online Journal	kursyvalyut.com	+	+	–
7	Good morning, accountant!	Online Newspaper	dobryjranok.com	+	+	+/–
8	Victana	Online Journal	victana.lviv.ua	+/–	+/–	+/–
9	Information for Businesses	Online Journal	goodmorningua.com	–	–	–

There are developed software for commercial Web content formation, management and support. There are posted software of developed systems implementation with subsystems of information resources processing in e-business organization over the Online Newspaper and Online Journal.

Table 2. The results of systems operation in the time period from 10.2010 to 03.2013

Characteristics	***The developed e-business systems***								
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	***9***
Visiting	4144	5997052	646	3654456	7838	9124	8724	517	25
Visits time	4:50	2:14	6:22	2:04	1:58	1:30	2:27	8:38	8:12
Failures/refusals	55.62	71.90	47.5	83.08	54.45	82.76	68.15	35.8	48.0
Goal Conversion	4.35	0.00	0.00	0.00	10.97	0.00	0.00	0.00	0.00
Unique visitors	2674	2501402	256	1501202	5704	7632	4996	250	7
Repeat visits	1470	3495641	393	2153254	2134	1500	3728	267	18
Page-views	19082	11588861	2924	11588861	20828	16952	18892	3760	81
Pages/visit	4.60	1.93	4.51	1.67	2.66	1.86	2.17	7.27	3.24
% New Visits	64.48	41.68	39.3	39.88	72.75	83.56	57.23	47.9	28.0
New Visitor	64.53	41.71	39.5	34.76	72.77	83.58	57.27	48.4	28.0
Returning Visit	35.47	58.29	60.5	65.24	27.23	16.42	42.73	51.6	72.0
Search traffic	56.47	36.10	32.5	26.12	69.67	79.68	24.34	23.0	20.0
Direct Traffic	19.50	11.20	38.5	31.22	17.15	14.17	48.73	32.9	80.0
Transitions (%)	24.03	52.48	28.97	42.46	13.18	6.15	26.94	44.10	0.00
Language – RU	45.15	55.57	33.90	43.54	66.87	54.39	69.01	49.7	52.0
Language – EN	36.20	23.70	57.3	33.34	10.23	25.51	10.49	39.3	48.0
Ukraine	81.25	91.22	83.9	87.56	67.17	66.01	89.47	87.8	60.0
Russia	9.10	1.84	6.78	4.67	24.04	5.57	0.62	0.39	0.00
Poland	0.19	0.29	0.01	0.12	0.11	0.61	0.11	0.19	0.00
United States	1.21	0.66	0.01	0.32	0.01	0.57	0.07	0.01	0.00
Germany	0.34	0.32	0.15	0.11	0.20	0.26	0.08	0.01	0.00
City – Kyiv	17.25	23.62	29.1	25.45	20.09	22.86	36.16	28.4	48.0
City – L'viv	38.92	43.41	48.4	34.54	10.17	14.43	9.70	45.1	12.0
Dnipropetrovsk	2.82	2.72	0.00	0.45	3.22	2.51	6.63	1.35	0.01
City – Kharkiv	1.50	2.82	0.62	0.65	3.56	1.65	6.08	1.16	0.01
City – Ternopil	2.24	0.01	0.01	0.01	0.01	2.17	0.01	0.01	0.01

Table 2 shows the comparative results characteristics of the developed e-business systems operation over the time period 10-11.2012. Statistical data is obtained from Google Analytics, which provides advanced features for data analysis. Google Analytics is a free service for visits maintenance to Web-sites. It allows us to estimate the information resources traffic in e-business systems and the effectiveness of marketing activities, such as for Online Newspaper and Online Journal. Fig. 2-3 present the work results of the developed systems in the form of graphs. So the all stages presence of the commercial content lifecycle significantly increases the visits and unique users amount of information resources in e-business systems.

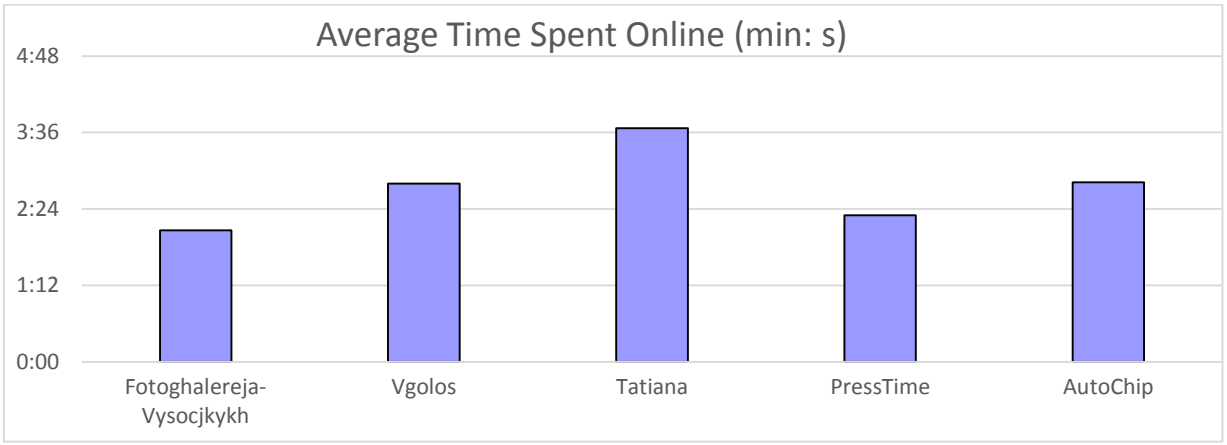


Fig. 2a. Distribution of time of arrival at the site

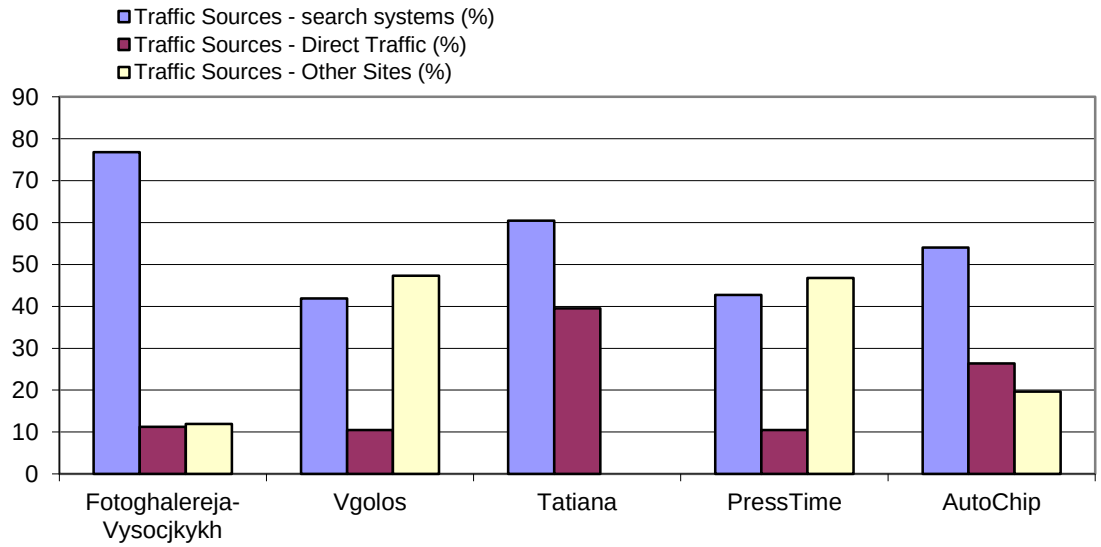


Fig. 2b. Distribution of traffic

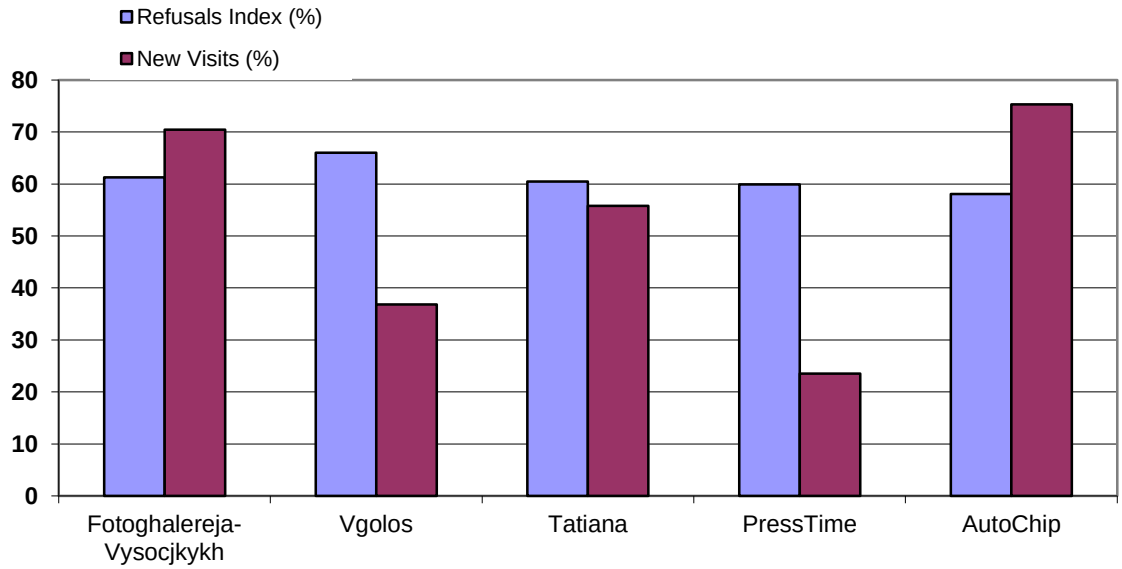


Fig. 3a. Distribution of new visits and failures

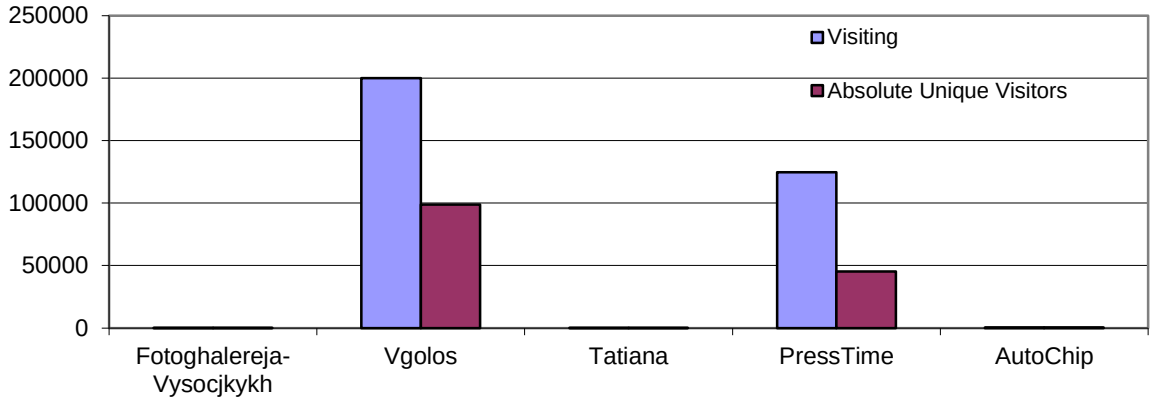
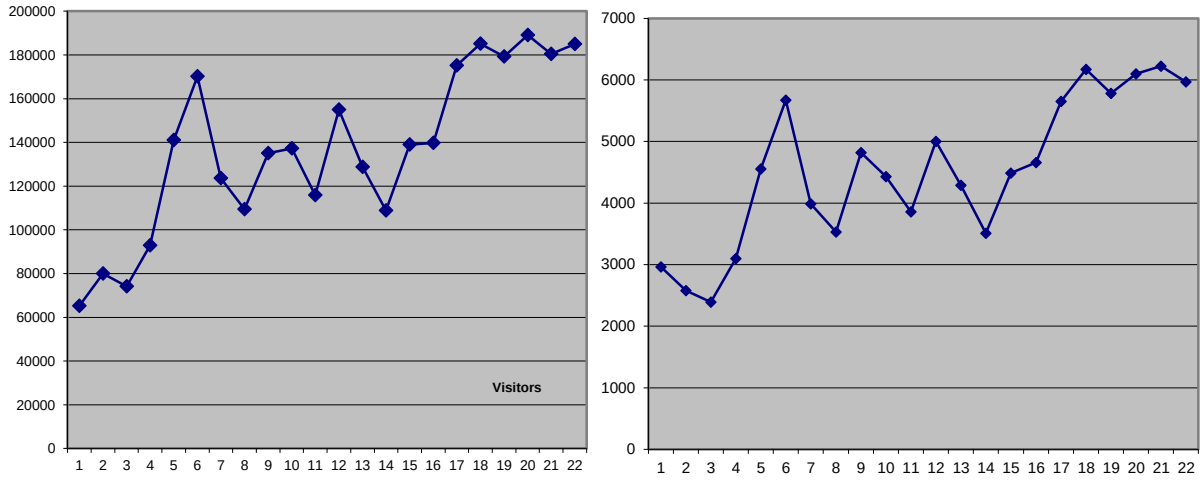


Fig. 3b. Distribution of information resources visiting

Service for accesses counting of information resource allows us to estimate the information resource traffic and marketing activities efficiency, such as for the Online Newspaper Vgolos (Fig. 4).



a) *b)*

Fig. 4. a) Overall and b) the average number of visitors for years 2010-2012

The subsystems presence of commercial Web content creation, management and support in e-business systems increases the number of unique visitors, viewed pages, visiting time and information resources, the target and regional audiences increasing.

The subsystem of Web content formation is implemented as a content-monitoring complex for content gathering from different sources of data which provides a content database creation according to the information needs of users. As a gathering and primary processing result of content is reduced to a single format, classified according to the specified categories. And he is credited descriptors with keywords. This facilitates the process implementation of content management.

Tasks of Web content management subsystem are: database formation, rotation and providing access to it; the operational and retrospective databases formation; the user experience personalization; personal user queries and sources storing; operation statistics analysis; search providing in database; initial forms generation on information resources; information interaction with other databases; the an information resource formation. Commercial Web content management subsystem is implemented through caching (representation module generates a page once; then it is several times faster loaded from the cache, which is updated automatically after a certain period of time or when making changes to specific sections of an information resource, or manually by administrator

command) or information blocks formation (blocks conservation in the information resources editing stage and page collection from these blocks at the user request of the relevant page).

Web content support subsystem provides information portraits formation, thematic storyline identification in content flows, the content relationship tables building, content rankings calculation, new events identification in their content flows, their tracking and clustering.

9 CONCLUSIONS

The paper is solved the actual scientific problem of methods and means research and development for commercial Web content processing in e-business systems by using the developed mathematical software for the appropriate systems creation. It is substantiated the necessity of methods and means development for the content processing in e-business systems by the system architecture improving to the processes automate of the content formation, management and support. In the work is analyzed e-business terminology for determination of their characteristic patterns, trends, process of e-business systems design and simulation and the shortcomings of existing methods and tools for content processing. It is developed e-business system model, which is helped to develop the commercial content lifecycle, the generalized system architecture and standardized methods for content processing in these systems. There is improved overall system architecture for e-business systems, which is different from the existing architectures of availability of commercial content processing modules. This is enabled the stage implementation of the content lifecycle.

In this paper is developed a complex methods of commercial Web content formation, management and support to achieve the work effect at the developer level (the time and costs reducing for development, quality improvement through the use of proven solutions). It is developed the general guidelines for the architecture designing of e-business systems, which are different from the existing systems more detailed steps and the presence of commercial Web content processing modules They allow effectively implement a information resources processing at the developer level of e-business systems (the time and resources reducing to the development, the quality improvement of e-business systems operation). In this work is developed the modules architecture of e-business systems for the stages implementation of the commercial Web content lifecycle. It is developed and implemented a software application of commercial Web content formation, management and support to achieve the effect of work at the level of the owner (profitability increasing, user interest growing) and user (clarity, the interface simplification standardization, unification, choice expanding) of e-business systems.

WORKS CITED

Berko, A., Vysotska, V., & Pasichnyk, V. (2009). Systemy elektronnoyi kontent-komertsiyi. Lviv, Ukraine: NULP Publ.

Boiko, B. (2005). Content Management Bible. Indianapolis: Wiley Publishing.

Bolshakova, E., Lande, D., Noskov, A., Klyshynskyy, E., Peskova, O., & Yahunova, E. (2011). Avtomatycheskaya obrabotka tekstov na estestvennom yazyke y kompyuternaya lynhvystyka. Moscow, Russia: MYEM Publ.

Braychevskyy, S., & Lande, D. (2005). Sovremennye informatsionnye potoki. Nauchnotehnicheskaya informatsiya(11), 21-33.

Clifton, B. (2009). Google Analytics: Professional'nyj analiz posešaemosti veb-sajtov = Advanced Web Metrics with Google Analytics. M., Russia: Vil'jame Publ.

Doyle, B. (2005). Seven Stages of the CM Lifecycle. EcontentMag.com. Retrieved from EContent: http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=13554&AuthorID=155

Ermakov, A., & Kiselev, S. (2005) Lingvisticheskaya model dlya kompyuternogo analiza tonalnosti publikatsiy SMI. Kompyuternaya lingvistika i intellectualnye tehnologii (pp. 282-285). Moscow, Russia: Nauka Publ.

Evdokimov, N. (2007) Raskrutka Web-saytov. Effektivnaya Internet-kommertsiya. Moscow, Russia: Williams Publ.

Fedorchuk, A. G. (2005). Kontent-monitoring informacionnyh potokov. Kiev: Ukraine: B-ki nac. akad. nauk: probl. funkcionirovanija, tendencii razvitija. Retrieved from http://www.nbuv.gov.ua/articles/2005/05fagmip.html

Furashev, B., Lande, D., Grigor'yev , O., & Furashev, O. (2005) Elektronne informatsiyne suspilstvo Ukraini. Akademiya pravovih nauk Ukraine., Kiev, Ukraine: Inzhiniring Publ.

Furashev, B., Lande, D., & Brajchevskiy, S. (2005) Sistemnaya informatizatsiya izbiratelnyh i referendumnyh protsessov (pp. 11-15). OIKIT № 29, Kharkiv, Ukraine: NAKU HAI Publ.

Grigoriev, A., & Lande, D. (2005) Adaptivnyy interfeys utochneniya zaprosov k sisteme kontent-monitoringa InfoStream (pp.109-111). K:Dialog'2005 Publ.

Grigoriev, A., & Lande, D. (2005) Sistema monitoringa InfoStream – informatsionnoe prostranstvo iz odnih ruk. Postroenie informatsionnogo obschestva: resursy i tehnologii (pp. 17-20). Kyiv, Ukraine: UkrISTEI Publ.

Hackos, J. T. (2002). Content Management for Dynamic Web Delivery. Hoboken, NJ, USA: Wiley.

Halverson, K. (2009). Content Strategy for the Web. Reading, Mass: New Riders Press.

Hartman, E. (2006). Content management lifecycle poster. Retrieved from CM Professionals: http://www.google.rs/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=0CDsQFjAC&url=http%3A%2F%2Fwww.jrtcllc.com%2Ffiles%2Freference-files%2Fcm-poster%3Fformat%3Draw&ei=Iru2UtLrH8LeygPPw4D4DQ&usg=AF Q jCNHN0mXF1wved45chqXh7AhomQ0hpA&sig2=IuCGFtOmMepRZK0Ah3zUYw&

Ivanov, S., & Krukovskaya, N. (2004) Statisticheskiy analiz documentaryyh informatsionnyh potokov. Nauchno-tehnicheskaya informatsiya. Inform. protsessy i sistemy (pp. 11-14) ser . 2, № 2.

Ivanov, S. (2002) Stohasticheskie fraktaly v informatike. Nauchno-tehnicheskaya informatsiya, (pp. 7-18), ser. 2, № 8.

Ivanov, V. (1994) Kontent-analiz: Metodologiya i metodika doslidzhennya ZMK. Moskalenko A. Sciences. Ed. Kyiv, Ukraine: Engineering Publ.

Jerk N. (2001) Razrabotka prilozheny dlya elektronnoy kommertsii. S-P, Russia: Piter Publ.

Korneev, V., Hareev, A., Vasyutyn, S., & Reich, B. (2000). Bazy dannyh. Intellektualnaya obrabotka informatsii. Moscow, Russia: Nolydzh Publ.

Krupnik, A. (2002) Biznes v Internet. Moscow, Russia: Mikroart Publ.

Lande, D., & Brajchevskiy, S. (2005) Opredelenie tematicheskoy napravlennosti zaprosov putem analiza nabora reytingovyh istochnikov (pp. 169-174). OIKIT № 29, Kharkiv, Ukraine: NAKU HAI Publ.

Lande, D., & Brajchevskiy, S. (2006) Prognozno-analiticheskie issledovaniya na osnove sistemy kontent-monitoringa InfoStream. Informatsiya, analiz, prognoz – strategicheskie rychagi effektivnogo gosudarstvennogo upravleniya (pp. 147-152). № 5. Kyiv, Ukraine: UkrISTEI Publ.

Lande, D. (2005) Nekotorye metody analiza novostnyh informatsionnyh potokov (pp. 277-287). Donetsk, Ukraine: IKVT, Donetsk National Technical University Publ., № 93.

Lande, D. (2005) Poisk znany v Internet. Moscow, Russia: Williams Publ.

Lande, D. (2005) Semanticheskiy veb. Telecom, № 6, pp. 60-65.

Lande, D. (2005) Skaner sistemy kontent-monitoringa InfoStream (pp. 53-58). OIKIT № 28, Kharkiv, Ukraine: NAKU HAI Publ.

Lande, D. (2006). Osnovy intehratsii informatsyonnyh potokov. Kyiv, Ukraine: Engineering Publ.

Lande, D., & Litvin, A. (2001) Fenomeny sovremennyh informatsionnyh potokov (pp. 14-21). Seti i biznes, № 1.

Lande, D., & Morozov, A. (2004) Chitayte novosti, batenka. CHIP, № 7, pp. 82-85.

Lande, D., & Morozov, A. (2005) Novostnoy Internet. Telecom. № 1-2, pp. 58-62

Lande, D., & Furashev, V. (2006) Voprosy postroeniya i ispolzovaniya mnogokriterialnoy modeli vybora istochnikov informatsii (pp. 76-85). OIKIT № 30, Kharkiv, Ukraine: NAKU HAI Publ.

Lande, D., Furashev, V., Braychevskyy, S., & Hryhorev, O. (2006). Osnovy modelirovaniya i otsenki elektronnyh informatsionnyh potokov. Kyiv, Ukraine: Engineering Publ.

Lande, D., Furashev , V., & Grigor'yev , O. (2006) Programno-aparatny kompleks informatsiynoyi pidtrymky priynyattya rishen. Kyiv, Ukraine: Inzhiniring Publ.

McGovern, G., & Norton, R. (2001) Content Critical. Upper Saddle River, NJ: FT Press Publ.

McKeever, S. (2003). Understanding Web content management systems: evolution, lifecycle and market. Industrial Management & Data Systems, 103(9), 686 - 692.

Nakano, R. (2002). Web content management: a collaborative approach. Boston: Addison Wesley Professional.

Papka, R. (1999) On-line News Event Detection, Clustering, and Tracking. Ph. D. Thesis, University of Massachusetts at Amherst, September

Pasichnyk, V., Scherbyna, Y., Vysotska, V., & Shestakevych, T. (2012). Matematychna linhvistyka. Lviv, Ukraine: Novyy Svit - 2000 Publ.

Rockley, A., & Cooper, C. (2002). Managing Enterprise Content: A Unified Content Strategy (Second ed.). Berkeley: Reading Mass New Riders Press. Retrieved from http://www.managingenterprisecontent.com/

Salton, D. (1979) Dinamicheskie bibliotechno-informatsionnye sistemy. Moscow, Russia: Mir Publ.

Soroka, M., & Tanatar, N. (1998) Vikoristannya metodu kontent-analizu pri stvorenni avtomatizovanih informatsiynih sistem (pp. 318-322). BNKI № 1.

Sovetov, B., & Yakovlev, S. (1998). Modelyrovanye system. Moscow, Russia: VS Publ.

Stone, W.R. (2003). Plagiarism, Duplicate Publication and Duplicate Submission: They Are All Wrong! IEEE Antennas and Propagation, 45(4). Retrieved from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1241310&url= http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D1241310

Uspensky, I. (2001) Entsiklopediya Internet-biznesa. S-P, Russia: Piter Publ.

Vysotska, V., & Chyrun, L. (2011). Commercial Web Content Lifecycle Model. CSIT'2011 (pp. 160-163). Lviv, Ukraine: LPNU

Zipf, G. (1949) Human Behavior and the Principle of Least Effort. Addison-Wesley.

Zipf, G. (1935) The Psychobiology of Language. Houghton-Mifflin.

Received for publication: 14.04.2014
Revision received: 17.06.2014
Accepted for publication: 03.07.2014

How to cite this article?

*Style – **APA** Sixth Edition*

Vysotska, V., & Chyrun, L. (2014, 07 15). Analitical methods for commercial web content processing of information resources in electronic business systems. (Z. Čekerevac, Ed.) *MEST Journal, 2*(2), 285-299. doi:10.12709/mest.02.02.02.29

*Style – **Chicago** Fifteenth Edition:*

Vysotska, Victoria, and Lyubomyr Chyrun. 2014. "Analitical methods for commercial web content processing of information resources in electronic business systems." Edited by Zoran Čekerevac. *MEST Journal* (MESTE) 2 (2): 285-299. doi:10.12709/mest.02.02.02.29.

*Style – **GOST** Name Sort:*

Vysotska Victoria and Chyrun Lyubomyr Analitical methods for commercial web content processing of information resources in electronic business systems [Journal] // MEST Journal / ed. Čekerevac Zoran. - Belgrade : MESTE, 07 15, 2014. - 2 : Vol. 2. - pp. 285-299.

*Style – **Harvard** Anglia:*

Vysotska, V. & Chyrun, L., 2014. Analitical methods for commercial web content processing of information resources in electronic business systems. *MEST Journal,* 15 07, 2(2), pp. 285-299.

Style – **ISO 690** Numerical Reference:

Analitical methods for commercial web content processing of information resources in electronic business systems. **Vysotska, Victoria and Chyrun, Lyubomyr.** [ed.] Zoran Čekerevac. 2, Belgrade : MESTE, 07 15, 2014, MEST Journal, Vol. 2, pp. 285-299.