Andriy BERKO[*]*, Victoria VYSOTSKA*[**]*, Lyubomyr CHYRUN*[***]

FEATURES OF INFORMATION RESOURCES PROCESSING IN ELECTRONIC CONTENT COMMERCE

Abstract

Some solutions of main problems of content e-commerce and content management services are described in the paper. Content lifecycle model in electronic commerce systems is considered. The model gives an opportunity to create effective tolls of information resources processing in electronic commerce systems and to implement the subsystem of commercial content formation, management and support.

1. INFORMATION

The electronic content commerce systems (ECCS) creation and implementation is one of the important areas of e-business development. A characteristic of such systems is the automatic processing of information resources. This helps to increase content sales volumes for a regular user, the active involvement of potential users and the boundaries expansion of the target audience [1-2]. In particular, the principles and technologies of electronic content commerce are used actively in the systems creation for on-line/off-line sales and content analysis/exchange/saving, cloud storage/computing, online shop [2]. The problems of electronic content commerce systems design, development, implementation and maintenance are topical. This is in view to such factors as the active development of research in the e-business area, the theoretical justification lack of standardized methods and need for the software unification of information resources processing in such systems [1-2,9-19].

---

[*] Environment and Ecosystems Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE, +38 (032) 258-91-82, berkoandriy@yandex.ua
[**] Information Systems and Networks Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE, +38 (032) 258-26-38, victana@bk.ru
[***] Software Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE, +38 (032) 258-25-78, chyrunlv@mail.ru

Regularly appear the new approaches and solutions to this problem, but the important issue is the discrepancy between the known methods and means of information resources processing and principles of electronic content commerce systems building. There are lack of general approaches of ECCS creation and standardized methods of information resources processing in these systems [2]. Specialists in ECCS designing, implementation and deployment have to deal with different processes of information resources processing. They contribute an increase of content sales volumes for a regular user, the active involvement of potential users and the boundaries expansion of the target audience [1-2,4].

2. THE ANALYTICAL PROCESSES OF COMMERCIAL CONTENT PROCESSING

2.1. Set-theoretic model of electronic content commerce systems

The main processes of information resources processing in electronic content commerce systems are the content formation, management and support (Fig. 1). Their scheme of connections is the following: *content formation* → *content management* → *content support*.
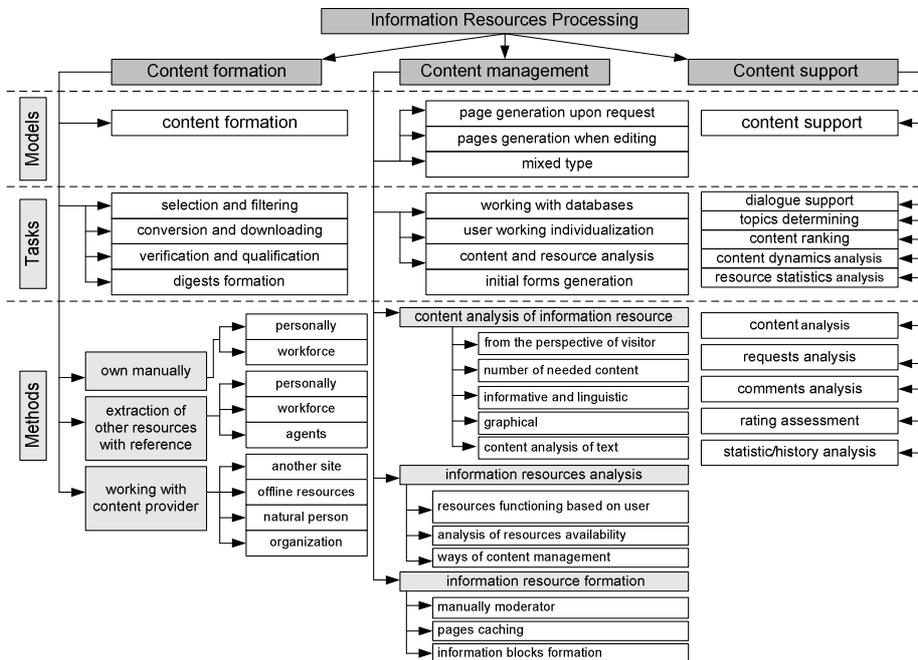

Fig. 1. Features of the information resources processing in electronic content commerce systems [source: own study]

Functioning of electronic content commerce systems is described by the following scheme of the main components links in this system:

1) for the process of an information resource formation: *content* → *commercial content formation* → *commercial content database* → *commercial content management* → *information resource of system*,

2) for the formation process of answers to a user's query: *user's query* → *commercial content management* → *information resource of system* → *commercial content support* → *system user database*,

3) for the formation process of the system working report to moderator: *a moderator request* → *commercial content support* → *users system database* → *commercial content management* → *report to moderator*,

4) for the moderation process of internal system parameters: *a moderator request* → *commercial content formation* → *commercial content support* → *rules database* → *commercial content management* → *results*.

Set-theoretic model of electronic content commerce systems is described as:

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

where: $X = \{x_1, x_2 ..., x_{n_X}\}$ – set of input data $x_i \in X$ from different sources at $i = \overline{1, n_X}$,

$Q = \{q_1, q_2 ..., q_{n_Q}\}$ – set by user requests $q_d \in Q$ at $d = \overline{1, n_Q}$,

Formation – the operator of a commercial content formation,

$H = \{h_1, h_2 ..., h_{n_Y}\}$ – set of $h_k \in H$ internal parameters in the S system at $k = \overline{1, n_H}$,

$C = \{c_1, c_2 ..., c_{n_C}\}$ – commercial content set $c_r \in C$ at $r = \overline{1, n_C}$,

$V = \{v_1, v_2 ..., v_{n_V}\}$ – the influence parameters set $v_l \in V$ of the external environment on the S system at $l = \overline{1, n_V}$,

Management – the operator of commercial content management,

Support – operator of commercial content suport,

$Z = \{z_1, z_2 ..., z_{n_Z}\}$ – pages set $z_w \in Z$ of information resource in the S system at $w = \overline{1, n_Z}$,

$T = \{t_1, t_2 ..., t_{n_T}\}$ – transaction time $t_p \in T$ of information resource processing in the S system at $p = \overline{1, n_T}$,

$Y = \{y_1, y_2 \ldots y_{n_Y}\}$ – statistical data set $y_j \in Y$ in the S system at $j = \overline{1, n_Y}$.

The value $y_j = \{a_1, a_2, \ldots, a_g\}$ is determined by using Google Analytics, 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 – an failures indicator (%) for a time period Δt, a_4 – achieved purpose of commercial content finding, a_5 – the dynamics of the 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. [5].

2.2. Content flows in electronic commerce systems

The growth of the content volume and its spread rate contributed to define the concept of content streams. Analysis of these flows requires the new tools using of information resources processing in ECCS for commercial content formation, management and support of (Fig. 2).
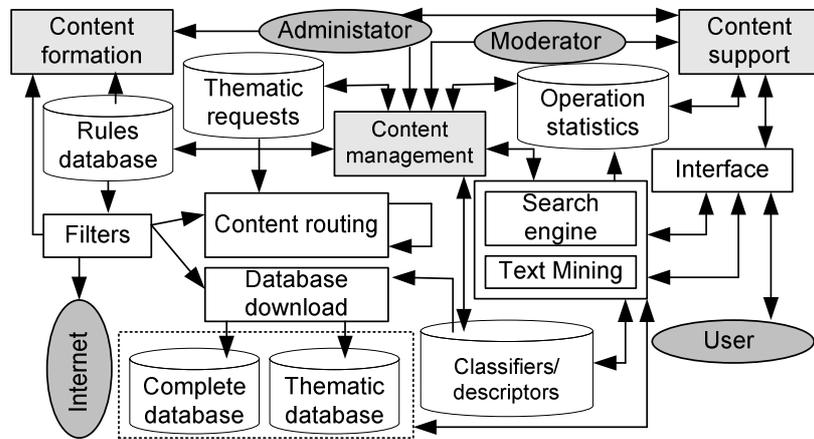
**Fig. 2. Functional diagram of the information
resources processing [source: own study]**

ECCS includes administration core, authorization/authentication subsystem, templates manager and content manager for problems solving from the user standpoint (Fig. 3).
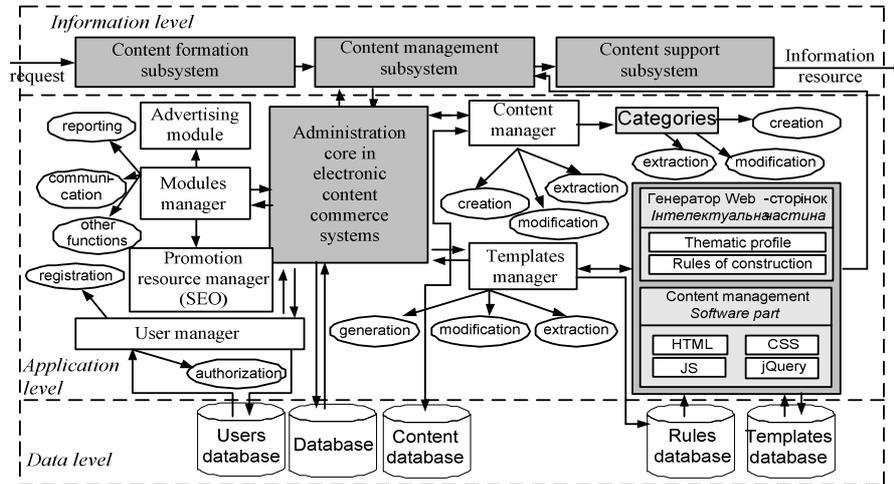
Fig. 3. The structure of electronic content commerce systems [source: own study]

Classical mathematical apparatus and distribution tools are not able to adequately reflect the content array analysis of fixed size and navigation to it [3-4, 6-8]. ECCS architecture has the hierarchy levels. They provide an independence of stored data on the programs that use them. Also they provide an opportunity of the system development without the existing applications destruction. In Fig. 4 is presented scheme of information resources processing in ECCS, and in Fig. 5 – stages of the commercial content formation in these systems. ECCS work based on the following scheme: *Content Manager → Editing Subsystem → Database → Core → User*.
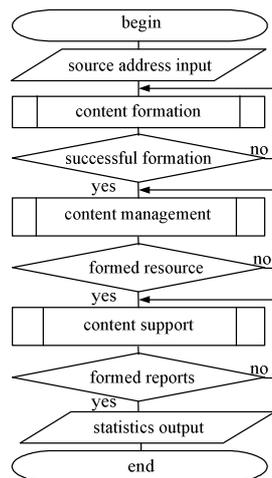

Fig. 4. Process of information resources – processing [source: own study]
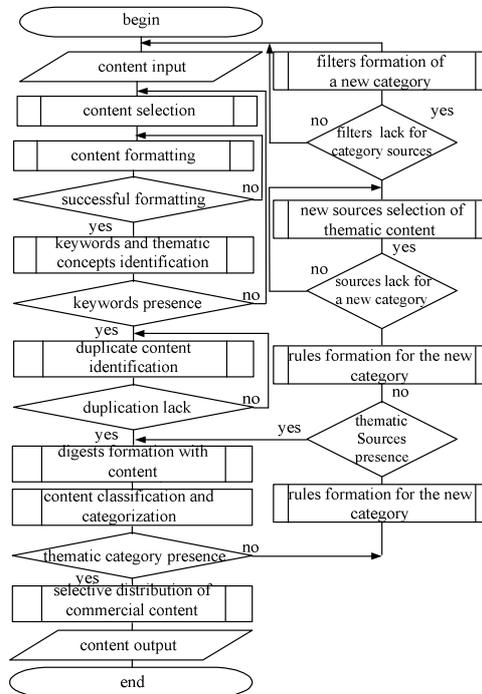
9

Fig. 5. Process of information resources – formation [source: own study]

When ECSS core creating are used an object-oriented models and abstract objects with properties/methods (Fig. 6). The interaction with the main objects of the system core implements through encapsulation. For this purpose in class implements the interface methods designed for manipulating within the object of the data/properties.
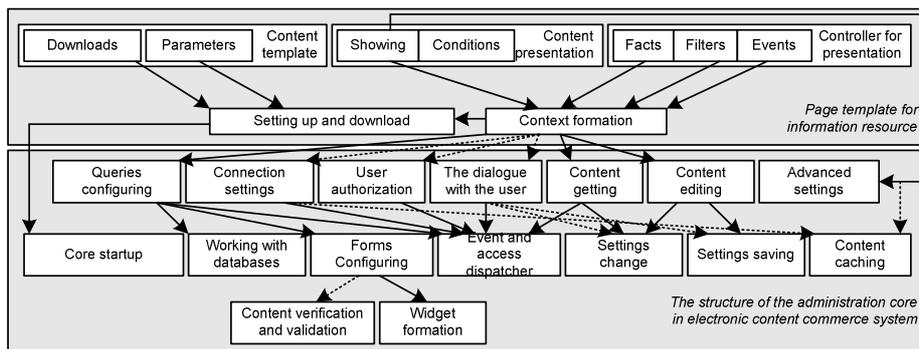

Fig. 6. The structure diagram of electronic content commerce systems [source: own study]

The process of information resources processing in ECCS is the content formation, management and support (Fig. 7).
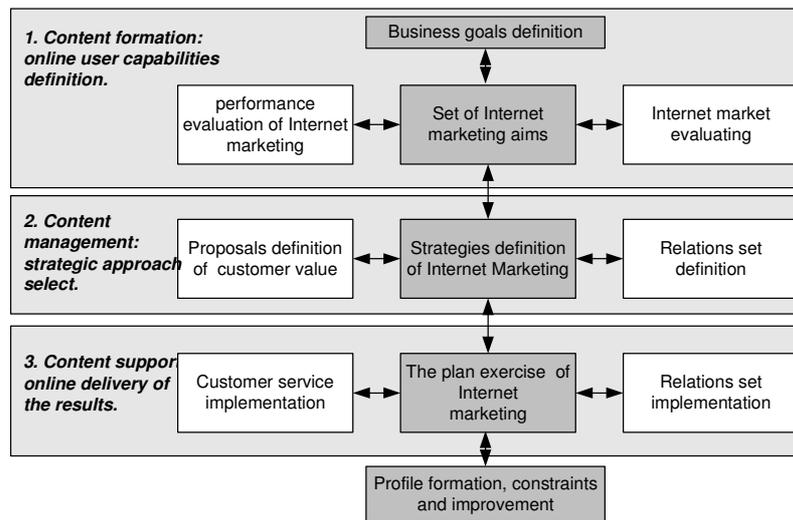

Fig. 7. Basic processes of information resources processing [source: own study]

The obtained data taken into account when an information resource creating or updating and the ECCS architecture improvement. Earlier moderator is independently searched and worked out needed content: content collection from different sources of data, content analysis and filtering, commercial content formation as the final product according to individualized data from system user. The process implementation of information resources processing in ECCS allows facilitating of moderator work by automating the content gathering from sources, content analysis and filtering.

2.3. Commercial content formation

The content formation consists of several stages: *moderator → content creating → database → content systematization → database → content distributing → editor* or *information resource (source) → content gathering → database → content systematization → database → content distribution → moderator*. This is realized in the form of content monitoring and database creation according to the information needs of consumers (Fig. 8).

Content leads to a common format after its gathering and primary processing. This is classified according to the specified classifier, including categories. And Content is attributed descriptors, including keywords.

When using online marketing for content systematization stages are provided constant updating of available data in databases, effective concurrent access of many users to the database, convenient means of the desired content finding.
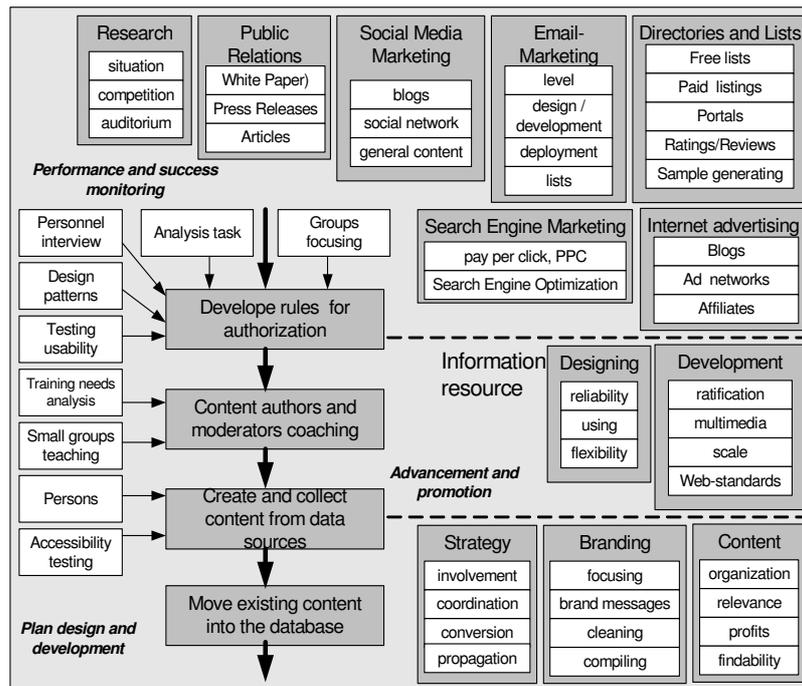
Fig. 8. Methods of content gathering/creation and marketing [source: own study]

The division into stages of the commercial content formation in the ECCS increases efficiency in the system administration, is provides economy of resources, Internet traffic and users anonymous, auto scanning of data sources.

2.4. Commercial content management

Content management of information resource and visitors (Fig. 9), their modelling is a quantitative research process of the individual thematic areas dynamics and technical analysis of information resource.

Fig. 9. Scheme of visiting analysis process of information resource [source: own study]

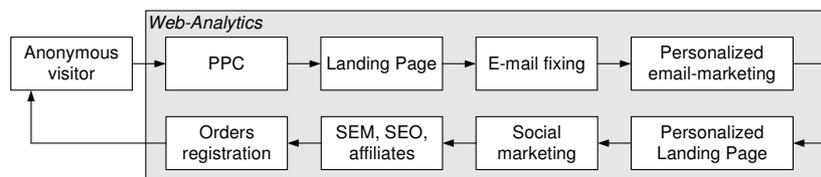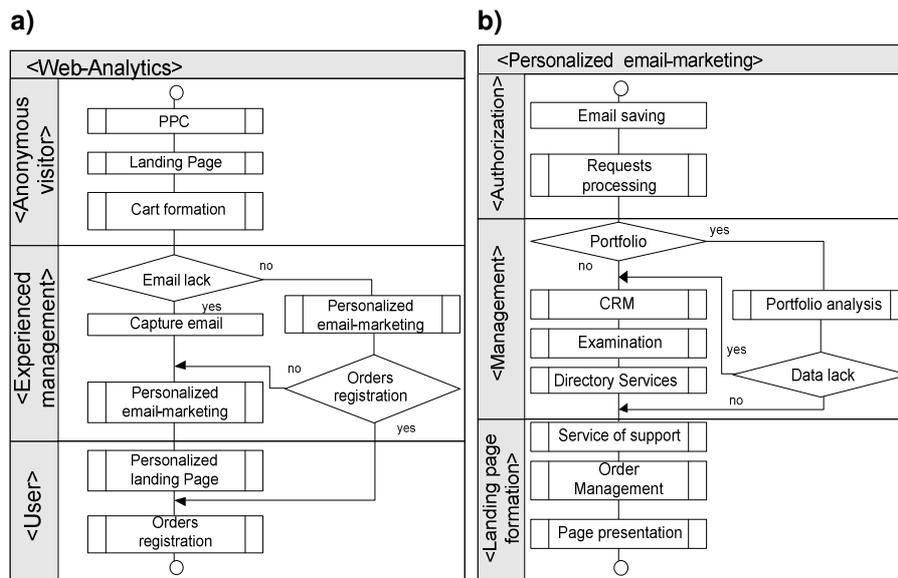Content management process consists of several stages (Fig. 10): *the user →
content processing → database → content analysis → database → content
presentation → user.*

a)

<Web-Analytics>	
<Anonymous visitor>	○
	PPC
	Landing Page
	Cart formation
<Experienced management>	Email lack — no → Personalized email-marketing
	↓yes
	Capture email
	Personalized email-marketing ← no — Orders registration
	↓yes
<User>	Personalized landing Page
	Orders registration
	○

b)

<Personalized email-marketing>	
<Authorization>	○
	Email saving
	Requests processing
<Management>	Portfolio — yes →
	↓no
	CRM — Portfolio analysis
	Examination yes
	Directory Services — Data lack
	↓no
<Landing page formation>	Service of support
	Order Management
	Page presentation
	○

**Fig. 10. Content analysis scheme of a) visitors
and b) an information resource [source: own study]**

Results analysis of the content management affects the speed thematic areas
and content space development. Stable static relationships between content
indicate a correlation between themes, the effectiveness of links to sources
publishing, earlier citation, and republications others. Mechanisms, based on
the generalized method of cluster analysis, identify content in the flow that
forms the around a new thematic areas. Cluster analysis, fractals and self-similar
processes theory in their correct application assess quantitatively the commu-
nication degree in thematic content streams. Operational analysis of visitor's
content contributes to the implementation the content management process using
pages generation through information blocks. It is divided into types: content,
on last accessed, and combined.

2.5. Commercial content support

The content support process is operational stages of commercial content
generalization, moderation and structuring (Fig. 11), i.e. *user → content
structuring → database → content moderation → database → content
generalization → moderator.* With the content support subsystem ECCS has

13

the following features: commercial content rankings formation; information portraits formation of permanent user in ECCS; characteristics analysis (comments, reviews, suggestions, etc.) on the commercial content of by the user in ECCS; information collection, storage and processing about the needs of the end/ potential user in ECCS and content consumer; a information portrait formation of content flows.
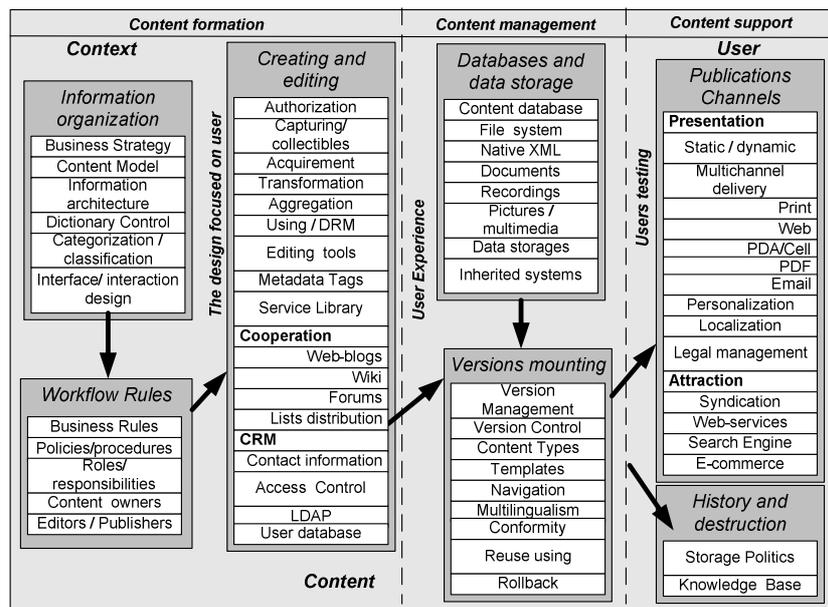

**Fig. 11. Subsystems of content formation, management and support [source: own study]**

The electronic content commerce systems greatly facilitate the work of moderator in the commercial content formation, management and support (Fig. 12). It has the following main stages of content processing of information resources.

1. Content gathering from different data sources.
2. Created and/or collected content filtration.
3. Content analysis (keywords identification, classification, digests formation).
4. Obtained data analysis for the moderator attention concentration on necessary issues. Data preparation for operational work with them.
5. Needs analysis of permanent and potential users of system.
6. End product formation according to the analyzed data and the needs of potential/regular users.
7. The list updating of data sources by users' queries analyzing.
8. Filter rules and dictionaries updating by content analysis.
9. New thematic categories identification for content flow.

14
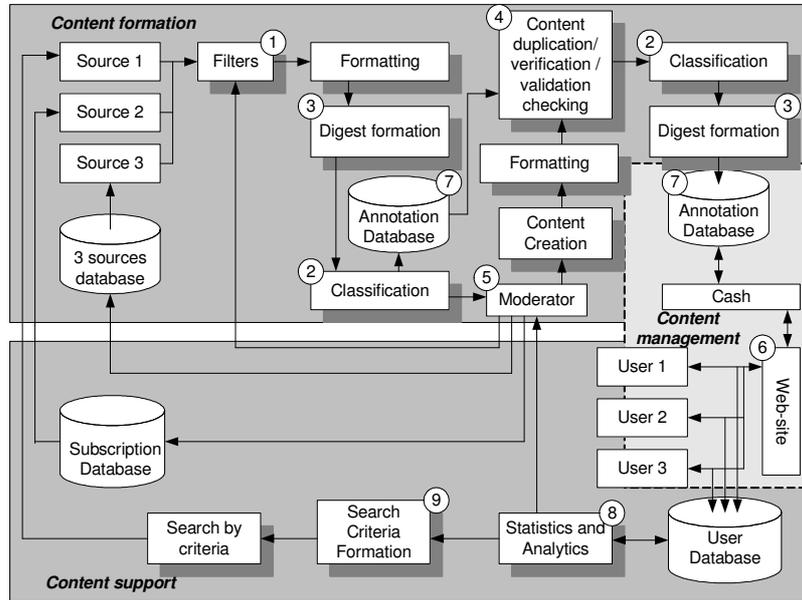
Fig. 12. Scheme of subsystems interaction of information resources processing [source: own study]

In Fig. 13 provides for a developed methods classification of information resources processing in ECCS with a detailed list of the implemented stages of commercial content formation, management and support of in these systems.
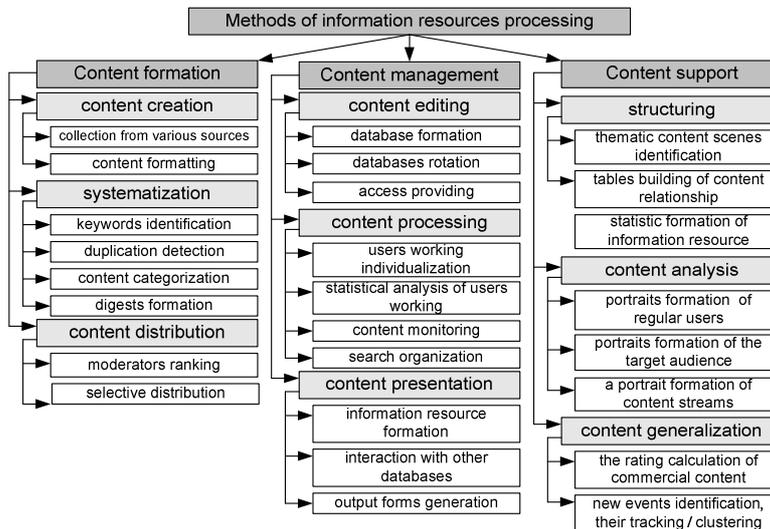
Fig. 13. Methods of information resources processing in electronic content commerce systems [source: own study]

Developed methods of information resources processing in ECCS make it possible to generate requirements for routines of information resources processing.
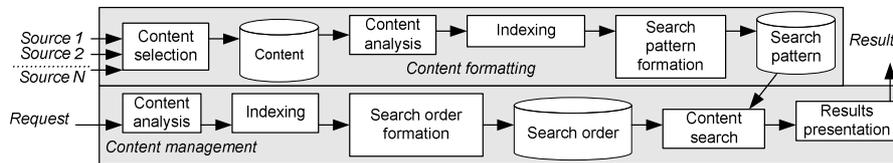
2.5. Content analysis of information resources

The modernity feature is the constant growth of content production rate. This process is objective and positive, but there was a problem: progress in the content production leads to a decrease in the general level of the potential user awareness. Increase in volumes content leads to the impossibility of his immediate processing. Moreover content distribution speed arise specific problems number (table 1).

Tab. 1. The main negative factors in the commercial content formation

Name	Rationale	Decision
Information noise	Content array structuring.	Filters, content monitoring, site analysis, content analysis.
Parasitic content	Appearances as applications.	Filters, content monitoring, content analysis.
Content irrelevance	User needs inconsistency.	Creation of annotated database, search pattern of initial content and their clustering, content analysis.
Content duplication	Duplication in sources.	Content analysis, scanners and filters based on statistics and criteria.
Navigating through the content stream	Rapid growth of the content volume and distribution.	Site analysis, filters, content monitoring, content analysis.
Search redundancy	Duplication and irrelevance.	Annotated search, content analysis and abstracting.

Negative factors in the commercial content formation make it difficult to find the necessary data when different sources scanning. The increase in volumes and relevance/dynamics variable of content streams (permanent regular and non regular updates) leads to duplication, information noise and results redundancy of content search. Large dynamic content streams coverage and generalization that continuously generate Internet sources requires new methods/approaches of search as a content monitoring (Fig. 14). Input information for content-monitoring is the text in natural language as a sequence of characters, output information – a table of partition, sentences and lexemes of analyzed text. Content monitoring is software tool of finding automation of the most important components in the content flow. This is an informative analysis of content flow in order to constantly obtain of the required qualitative/quantitative cuts over the advance not defined period of time.

Fig. 14. Block diagram of the content monitoring process of the commercial content formation in the electronic content commerce systems [source: own study]

As part of the content monitoring is the content search and content analysis of the text. Content analysis is designed to search content in a data array by semantic linguistic units (Alg. 1). The account unit is a quantitative measure of the analysis unit. It allows symbol appearance frequency (regularity) registering of analysis category in text (number of specific words, or its combinations, lines, characters, pages, paragraphs, copyright pages, text area, etc.).

Algorithm 1. Commercial textual content analysis.
Stage 1. Criteria set determination for commercial textual content.
Step 1. Criteria set formation as the source type (forum, email, online newspaper, chat, and online magazine), type content (article, email, banner, comments), communication participants (sender, recipient, moderator).
Step 2. Content size (minimum volume or length), occurrence frequency, and distribution method/space and appearance time determination.
Step 3. Content stream filtration by generated set of criteria and the identified relevancy content storing.
Stage 2. Content analytic selection. The content sample formation on the criteria's of the limited sample from a larger array.
Stage 3. Meaningful units identification of commercial text content analysis (word combination, phrases, sentences, theme, idea, author, character, and social situation, the text part that is clustered content by analysis categories). Requirements for the linguistic unit selection of the analysis: large enough for values interpretation, small enough so as not to interpret many values, is easily identified, the units number is large enough for the sample conducting.
Stage 4. Account units bold of textual content analysis.
Step 1. If the account units are the same as analysis units, then find the occurrence frequency of the selected content unit, otherwise go to step 2.
Step 2. Moderator on the basis analyzed content offers account units, for example, the text length, text area, filled with meaningful units, the lines number (paragraph, symbols, marks, columns of text), the file size/type, the pictures number with a particular content/storyline and more.
Stage 5. Meaningful analytic units comparison with the unit account.
Step 1. Units classification by groups with scales evaluation of meaningful categories in total text. Classifier is a common table, which consolidated all categories of analysis and units analysis. It is fixed categories expression units.

Step 2. Statistical calculations are for content understandability and attractiveness identification.

Stage 6. The tool development of content analysis.

Step 1. Coded protocol of content creates for compact representation of data and rapid comparison of the different content analysis results.

Step 2. Content protocol is filled properties (author, publication time, volume, etc.).

Step 3. Commercial content protocol is filled analysis its results (number of usage for specific units of analysis and conclusions regarding the analysis categories). Each content protocol is filled based on data collection of all his registration cards.

Stage 7. It is content analysis table development. Table type is defined in a form of analytic coordinated and subordinated categories system: each category (issue) provides a features number (responses), which quantifies the textual content.

Stage 8. The coding matrix development for content analysis.

Step 1. If the sample size is ≥ 100 units, then is analyzed a set of matrix sheets, else perform step 2.

Step 2. If the sample size is <100 units, then is carried out a two-dimensional analysis. In this case, for the each content is formed coding matrix.

Stage 9. The text analysis is conducted by the created coding matrix.

Stage 10. Results are interpretation. The content characteristics identify and evaluate based on a statistical set of calculated coefficients for a certain period of time on specified category. This includes all extracted textual fragments; conclusions do not rely on some results and accounted for all without exception.

The content analysis using in data Internet sources monitoring automates the process of the most important components finding in the content flow based on the data selection from these sources. This eliminates content duplication, information noise, parasitic content search results redundancy and more. This method is used in subsequent stages of the commercial content formation for more accurate relevant results – a unique commercial content creation that is in demand among ECCS users.

3. CONCLUSIONS

From the perspective of systemic approach, is conducted the principles applying of information resources processing in ECCS for content lifecycle implementation, that enabled to develop methods for the commercial content formation, management and support. It is developed an integrated method of commercial content formation for the time and resources reduction of content production. This makes it possible to create a means of information resources processing and implement subsystem of automatically generated content.

In this paper is created operational method of commercial content management for the time and resources reduction of content sales, which makes it possible to implement commercial content management subsystem.

It is implemented a comprehensive method of commercial content support for the time and resource reduction of the target audience analysis in ECCS, which makes it possible to develop a commercial content support subsystem.

REFERENCES

[1] BEREZA A., KOZAK I., LEVCHENKO F.: *Elektronna komertsiya* (in Ukraine). KNEU, Kyiv, 2002.

[2] BERKO A., VYSOTSKA V., PASICHNYK V.: *Systemy elektronnoyi kontent-komertsiyi* (in Ukraine).NULP, Lviv, 2009.

[3] BOLSHAKOVA E., LANDE D., NOSKOV A., KLYSHINSKY E., PESKOVA O., YAGUNOVA E.: *Avtomaticheskaya obrabotka tekstov na estestvennom yazyke I kompyuternaya lingvistika.* (in Russian). MIEM, Moskva, 2011.

[4] BRAYCHEVSKY S., LANDE D.: *Sovremennye informatsionnye potoki* (in Ukraine). Nauchno-tehnicheskaya informatsiya, no. 11, 2005, pp. 21-33.

[5] CLIFTON B.: *Google Analytics: professionalny analiz poseschaemosti web-saytov* (in Russian). Vilyams, Moskva, 2009.

[6] KORNEEV V., GAREEV A., VASYUTIN S., FREIHAM V.: *Bazy dannyh. Intellektu-alnaya obrabotka informatsii* (in Russian). Nolidzh, Moskva, 2000.

[7] LANDE D., FURASHEV V., BRAYCHEVSKY S., GRIGOREV O.: *Osnovy modelirovaniya i otsenki elektronnyh informatsionnyh potokov* (in Ukraine). Inzhiniring, Kyiv, 2006.

[8] PASICHNYK V., SCHERBYNA J., VYSOTSKA V., SHESTAKEVYCH T.: *Matema-tychna linhvistyka* (in Ukraine). "Novyy Svit – 2000", Lviv, 2012.

[9] BOIKO B.: *Content Management Bible* (in USA). Hoboken, 2004.

[10] CM LIFECYCLE POSTER: *Content Management Professionals* (in USA). Retrieved 20, 2010. (http://www.cmprosold.org/resources/poster)

[11] DOYLE B.: *Seven Stages of the CM Lifecycle* (in USA). EcontentMag.com, 2010. (www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=13554&AuthorID=155)

[12] HACKOS J.: *Content Management for Dynamic Web Delivery* (in USA). NJ: Wiley, Hoboken, 2002.

[13 HALVERSON K.: *Content Strategy for the Web* (in USA). Reading, Mass: New Riders Press, 2009.

[14] MCGOVERN G., NORTON R.: *Content Critical* (in USA). Upper Saddle River, NJ: FT Press, 2001.

[15] MCKEEVER S.: *Understanding Web content management systems: evolution, lifecycle and market* (in USA). Industrial Management & Data Systems (MCB UP), 103 (9), 2003, pp. 686–692.

[16] NAKANO R.: *Web content management: a collaborative approach* (in USA). Boston, Addison Wesley Professional, 2002.

[17] PAPKA R.: *On-line News Event Detection, Clustering, and Tracking* (in USA). Ph. D. Thesis, University of Massachusetts at Amherst, September, 1999.

[18] WOODS R.: *Defining a Model for Content Governance*. 2010. (www.contentmanager.net/magazine/article_785_defining_a_model_for_content_governance.html)

[19] ROCKLEY A.: *Managing Enterprise Content: A Unified Content Strategy* (in USA). Reading, Mass: New Riders Press, 2002.