SCITECH
RESEARCH ORGANISATION|

Journal of Information Sciences and Computing Technologies
www.scitecresearch.com

The Means Structure of Information Resources Processing in Electronic Content Commerce Systems

Victoria Vysotska[1], Lyubomyr Chyrun[2]

[1] Information Systems and Networks Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12.

[2] Software Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12.

Abstract

Some of principal problems of electronic content commerce and functional services of content processing are analyzed in the article. Proposed method gives an opportunity to form resources processing tools for electronic commerce systems so as implement subsystems for content formation, management and support.

Keywords: Information resources; commercial content; content analysis; content monitoring; content search; electronic content commerce systems.

1. Introduction. General statement of the problem

The basis for the design of the complex processing of information resources in electronic content commerce systems (ECCS) a three-level architecture client/server [1] was selected. The process of the content processing occurs on a "client-server application – database". The request is processed by an application server that communicates with the database and payment system, and when you are connected to the business process of the organization communicates with the appropriate systems. From a technical point of view ECCS is a combination of Web-showcase as a front system and a trading system as the back-office. The main functions of ECCS are information service of the purchaser, order processing, payments, collection and analysis of statistical information.

2. Connection issues with important scientific and practical tasks

Active development of the Internet has increased the need of operational data production/strategic nature and implementation of new forms of information services [1]. Documented information prepared in accordance with the needs of the users of an information product or commercial content and the main object of processes of electronic content commerce. The treatment benefits of information resources by means of ECCS stipulated by the growth in the volume of content on the Internet, the rapid development of electronic business, rapid growth of Internet accessibility, expanding of the set of information goods and services, increased demand for commercial content [1-9]. Principles and techniques of electronic content commerce are used when creating the online stores (selling eBooks, Software, video, music, movies, picture), on-line systems (newspapers, magazines, e-learning, publishing houses) and off-line distribution of content (copywriting services, Marketing Services Shop, RSS Subscription Extension), cloud storage and cloud computing. In this area are working the world's leading manufacturers of means of information resources processing, such as Apple, Google, Intel, Microsoft, Amazon, Android, Opera. Factors that hindered the implementation of information resources processing in ECCS associated primarily with the lack of scientifically based methods and tools for creating, administering and maintenance of content [1]. A number of scientific studies is dedicated to this area. In particular, in his works D. Lande researched and developed mathematical models of electronic information [2, 3, 6, 7]. G. Zipf proposed the empirical law of distribution of word frequencies in natural language. In the works of B. Boiko, S. McKeever, A. Rockley the models of content life cycle were describes. The methodology of content analysis was founded and developed by J. Kaiser, Glaser, H. Lasswell, O. Holsti. The EMC Corporation, IBM, Microsoft, Alfresco, Open Text, Oracle and SAP have developed specifications of Content Management Interoperability Services for Web services interface that enables interoperability between content management systems of e-business.

3. Analysis of recent research and publications

The ECCS software forms the interface with the buyer, and the system functionality based on the company needs. A potential customer has the opportunity to get answers to any question at any time (conditions of after-sales service, advice on the payment specifics, etc) that accompanies the process of buying/selling. Registration/authorization occurs before/after content selecting. In the first case you create a registration entry for ECCS customers for whom you are implementing a special scheme of service and payment. The possibility of registration after selecting the content allows the buyer to remain anonymous and saves time. The system protects the personal information of the buyer, using data transmission over secure channels. ECCS receives full information about Web site visitors that allows you to build marketing system in accordance with her. CMS allows you to collect a lot of statistics for analysis and use it operational (allow to identify areas of our site that are optimal for advertising information, automate the course of the advertising campaign). The publication of additional information is implemented using a separate application server (the area of publications) and electronic databases. ECCS must support the content lifecycle. ECCS involves using IT to communicate trading companies with retail customers, providing a full cycle of content sales. In ECCS the participants are much wider, except for users (sellers, buyers) it includes a number of financial institutions (the issuing Bank, the Bank of the seller/buyer, the acquiring Bank), computer centers, etc. Users are primarily individuals and institutions, social institutions, other types of consumers (legal entity). Sellers in ECCS are different organizational forms of the trade content. The communication network consists of providers, servers, processing centers. The delivery system is the Internet. All components interact in a system of relationships. This is the guarantee of stability and reliability of ECCS. The important elements of ECCS are organizational forms of electronic content commerce, having a single focus onto retail sales providing, but differ in the composition, structure, purpose in ECCS. ECCS provides the introduction, the selection of content categories, checkout, payments settlements, tracing the execution of the order. For the SECC operation has hardware and software components: Web-storefront (front office) on the Web server; electronic catalogues; payment system; CMS. Web-showcase has an active content, is based on the conventional static HTML files or dynamic with the display of the database content. Web-showcase contains information about the name, the profile, the status of the ECCS owner, the range of content and services, means of payment, discounts, guarantees, and terms of content delivery.

4. Problems allocation

The issues of design, creation, implementation and maintenance of electronic content commerce is relevant, taking into account such factors as the lack of theoretical justification of standardized methods and the need for the standardization of software tools to process information resources. There is a mismatch between methods and means of the information processing resources and the principles of systems of electronic content commerce construction. One of the urgent tasks in the framework of this problem is the development of a common model structure of electronic content commerce.

5. Statement of purpose

The aim of this work is the definition of functional requirements to subsystems of processing of information resources in ECCS, such as the formation, management and support of commercial content. The subsystem of commercial content development provides facilitates the work of the ECCS author and moderators. The subsystem of content management facilitates the work of the administrators and moderators of ECCS, and supports a variety of functionality for users of these systems. The subsystem of content support makes work easier for ECCS analysts.

6. Main results in research

The processing of information resources is performed according to scheme: *content formation* $\rightarrow$ *content management* $\rightarrow$ *content maintenance*, so the model of electronic content commerce is given as $S = \langle X, Formation, C, Mamagement, Realization, Y \rangle$, where $X = \{x_1, x_2, ..., x_{n_X}\}$ – set of input data, *Formation* – content formation operator, $C = \{c_1, c_2, ...c_{n_C}\}$ – set of content, *Mamagement* – content management operator, *Realization* – content maintenance operator and $Y = \{y_1, y_2, ..., y_{n_Y}\}$ – set of input data. Input data x_i are independent variables, and the set of content c_j and output data y_k are dependent.

I. The stage of the content formation is implemented according to the scheme *build/create content* $\rightarrow$ *content organization* $\rightarrow$ *content distribution* in the form of content monitoring system of collect content from the source on a pre-defined methods (fig. 1), and enable the creation of the database (fig. 1) according to the information needs of consumers. Stage of content formation is described by the operator $c_j(x_i, t) = Formation(U_F, x_i, t, \Delta t)$ under conditions $U_F = \{u_{f_1}, u_{f_2}, ...u_{f_m}\}$ with result $c_j = \left\{ \bigcup u_{f_k} \middle| (x_i \in X) \wedge (\exists u_{f_k} \in U_F), U_F = U_{Fx} \vee U_{F\bar{x}}, i = \overline{1,m}, k = \overline{1,n} \right\}$.

```
                          Methods of content formation

Independently manually    Extraction other      Working with          Structuring content
                          resources with link   content provider

    personally                personally          other website        analysis of statistics/history

    workforce                 workforce           offline resource     analysis of comments

                              agents              natural person       rating evaluation

                          technical analysis      organization         analysis of the site's content
                          of website
                                                              Content analysis from the standpoint of visitor
                    Technical analysis of the functioning system
                    of website with the direction of the user      analysis of the number of the necessary content

                    Technical analysis of accessibility website    informative and linguistic analysis of content

                    Technical analysis of expedients of the web    graphical analysis of content
                    site content management
                                                                   content analysis of text (by keywords)
```
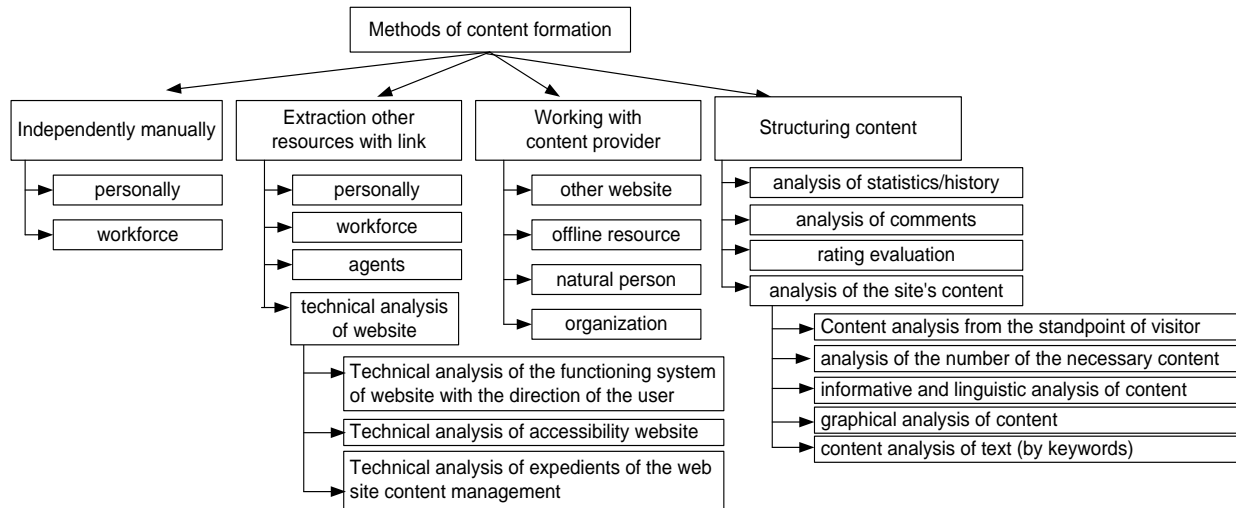
Fig. 1. Methods commercial content forming

As a result of collecting and primary processing the content is reduced to the unified format, classified according to certain categories and a number of descriptors are attributed to it, including keywords. Such complexes provide a constant replenishment of DB by operational messages, efficient concurrent access to the database by many users, a convenient means of finding the desired content. Level 1 provides access to the cache that is filled with intelligent scanner. This cache is accessed by end users using information retrieval system; users can reach the content directly at the Internet. At level 2, content is downloaded from the cache of level 1 and data base is filled with data scanned directly from the Internet. This approach has advantages: savings in administration resources; anonymity; Internet traffic saving; the possibility of self-scanning the Internet.

II. The stage of content management is described by the operator $c_j(q_l,t) = Management\,(U_M, q_l, t, \Delta t)$ with queries $Q = \{q_1, q_2, \ldots q_{n_Q}\}$ under conditions $U_M = \{u_{m_1}, u_{m_2}, \ldots u_{m_p}\}$. The commercial content management occurs as

$$c_j = \left\{ \bigcup u_{m_k} \middle| (q_i \in Q) \wedge (\exists u_{m_k} \in U_M), U_M = U_{Mq} \vee U_{F\bar{q}}, i = \overline{1,m}, k = \overline{1,n} \right\}$$

. The content management of the site is implemented as *content processing* → *content analysis* → *content presentation*; its modelling is one of the most informative methods for quantitative studies of the dynamics of different thematic areas and technical analysis of the site. There are three models of content management: page generation at user information request based on the information from the databases (fig. 2, a); static pages generation when editing with the lack of interactivity between the visitor and the site content (fig. 2, b); mixed pages generation type (fig. 2 c). Changes in the parameters of content management will determine the speed of development of individual topic areas or the entire content space. Stable statistical relationships between individual content indicated a correlation of some topics, about the effectiveness of links to the publications of his predecessors, the earlier citation, republication, etc. Mechanisms based on generalized cluster analysis techniques, encounter messages in the streams of content are shaping new themed areas. Cluster analysis and fractal theory processes at their correct application provide opportunities for quantitative assessment of the level of connection in thematic content streams. In the dialog mode the access to database allows to browse, search and view content, and having access to the original content in the Internet (fig. 2, d). Information in the database is changed by the editing module (fig. 3, a). Pages are re-created by the server with each request, which increases the load on system resources (fig. 3, b). The load is reduced due to the use of caching in modern servers (fig. 2, c). The benefits of pages caching with content: the presentation module generates a page once (fig. 3, d); page exists in cache for the period of time Δt – until content is relevant; complete page loads from the cache faster; the cache is updated periodically manually/automatically: after a certain period of time Δt; when modifying sections of the site or content. Types of realization of process of mixed type pages generation: by caching (fig. 3, c); through the formation of information blocks (Fig. 3, d): site editing (preservation of information blocks); pages creation (page assembling of the information blocks at the request of appropriate content by the user).
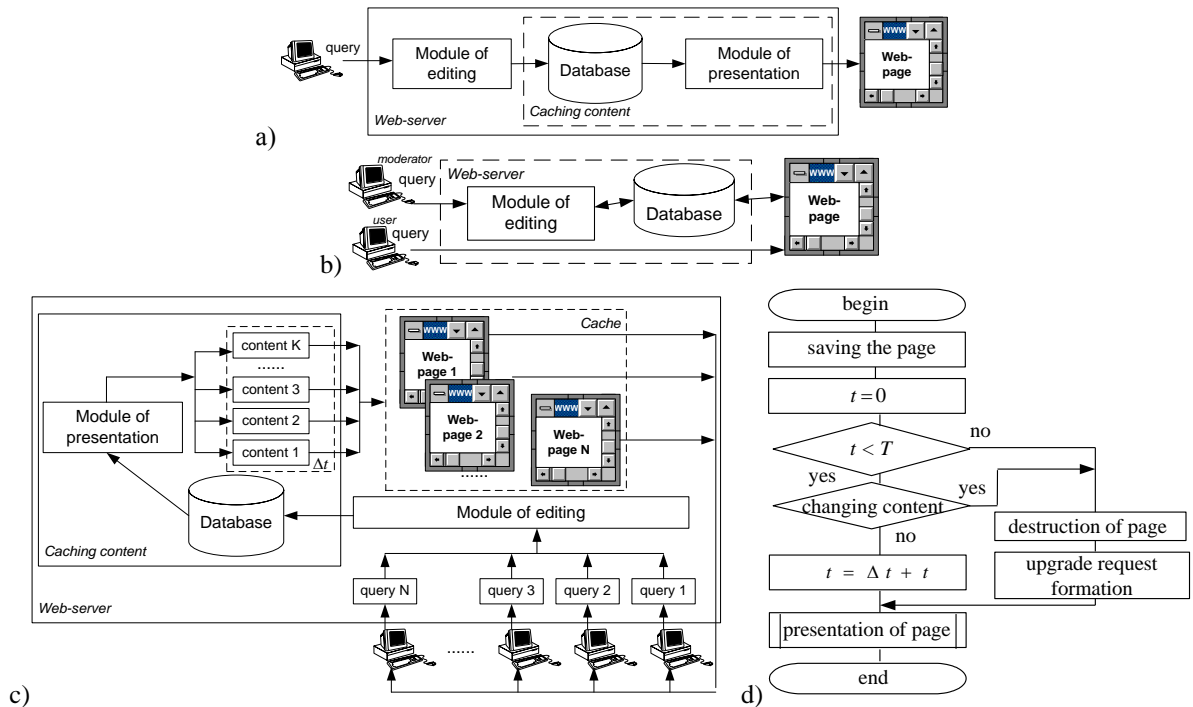
Fig. 2. Scheme pages generation (a) upon request, b) when editing, c) mixed type and d) the algorithm of commercial content presentation
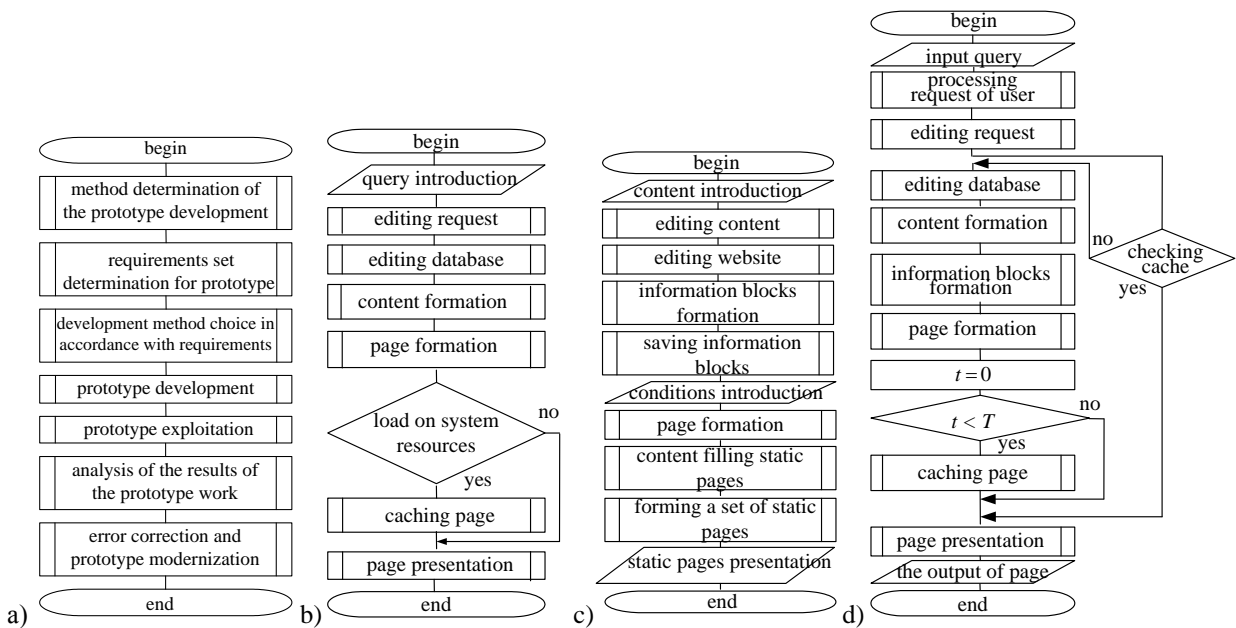
Fig. 3. Process a) content editing to create the prototype, b) page view c) pages generation and d) caching of mixed type pages

The task of full-text search in large arrays of content is ineffective. The accuracy issue solves search in annotated content. Instead of searching the full content, it is advisable to search by annotation – as images of content. Quasi-abstract for great content is a formation that resembles the original content and often is not perceived by the person. Search image of such content with weighted keywords and phrases leads to adequate results in full-text search. Quasi-abstract is built from blocks of content with large weight values.

III. The content implementation phase form the complex processes of content analysis. The content accompany stage is described as $y_k(t + \Delta t) = Realization(U_R, c_j, q_l, t, \Delta t)$ under conditions $U_R = \{u_{r_1}, u_{r_2}, \ldots u_{r_z}\}$, meaning

$$y_j = \left\{ \bigcup u_{r_k} \left| (q_i \in Q) \wedge (\vec{c} \in C) \wedge (\exists u_{r_k} \in U_R), U_R = U_{Rc} \vee U_{R\bar{c}}, i = \overline{1, m}, k = \overline{1, n} \right. \right\}$$

. The problem with content marketing is the lack of common approaches for the automatic analysis of the commercial activities of the respective companies

(marketing research). The relevance of developing the overall architecture of the module of content marketing is the need to obtain operational and objective assessments of the competition level in the financial segment of content market; to assess the level of competitors and the degree of competitiveness in the financial market of content distribution. The obtained data are taken into account when creating or updating/modifying Web-portal. The reason for urgency is a fast paced demand growth of content distribution; continuous interaction between society and scientists with the latest IT; the possibility of profits increasing of existing/potential customers, rating increase; the improvement of the business through understanding client according to this scheme: *content structuring → content moderation → content generalization*. With the aim of attracting more customers on the Web page ECCS add functions of content analysis (ratings, user feedbacks and comments) about the content (article, book, etc.). A site with user reviews of the content attracts more visitors, but this functionality does not include the moderation and leads to redundant information on the Web showcase. The end user is forced to segment meaningful information that requires effort and pushes the potential client away. Content that gets to the site, is not filtered, that is user interaction with the portal is one-way. The module or content implementing improves the ECCS ranking with interactive interface and performs the following tasks: collect marketing information on the content distribution; support dialogue with the end user of the content; the formation of the catalogue of commercial content and services of ECCS; the information support of the content consumer; the formation of a virtual shopping cart of the commercial content customer; registration of the content end user; processing orders for commercial content. With this module (fig. 4, a) ECCS has the following capabilities: raising the commercial content rating; characteristics analysis (comments, feedback, suggestions etc.) on commercial content from the user (fig. 4, b); collection, storage and processing of information about the needs of the end/potential content user (Fig. 4., c).
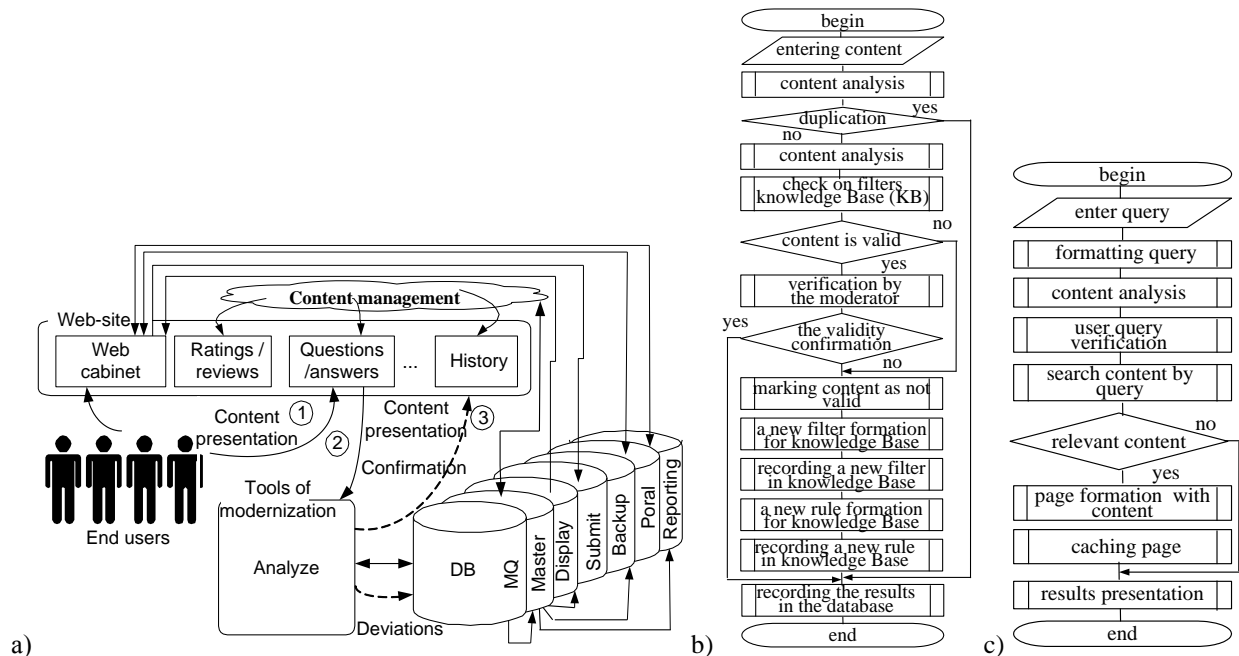

Fig. 4. a) support scheme, b) process analysis and b) the process of commercial content finding

Ratings/reviews is a module of evaluation of the rating of ECCS products on a predetermined scale (fig. 5). The user can see the rating of the content and has a right to leave feedback, allowing to build a constant communication between clients and the system. Module Questions/answers is a module of communication between clients themselves, resulting in more complete information on the content and, as a consequence, increases the sales. The story is a module of the content archive, the principle of operation is based on the use of "Questions and answers", only operates in one-way mode. Customer – potential/existing customer with its own website, with build in CMS in order to raise the rating of commercial content and a better understanding of the end user needs through the content analysis of questions/answers type, ratings, feedbacks, articles, and more. OpenMQ – it is a queue of FIFO type. CMS – module of application content filtering on pre-defined rules individually for each client. Workbench – individual account of the client with the ability to view content and for statistical analysis. Solar – indexing module for fast content search.
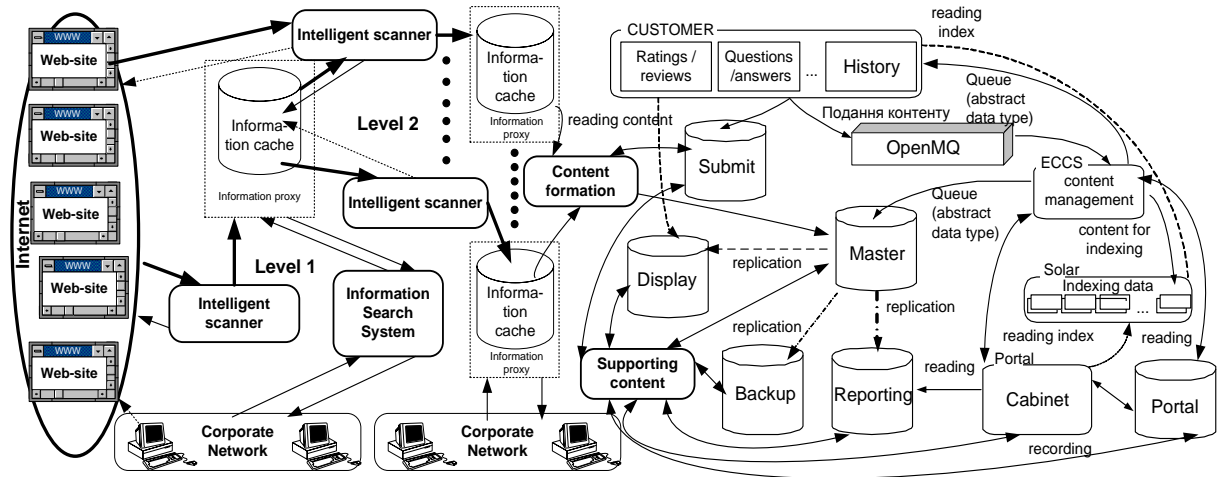
Fig. 5. The scheme of data flow in the electronic content commerce

In the processes of ECCS functioning the following databases are used: Master – main DB, which is the basic element of the system, which replicates data to child bases, Display – child database from the Master DB, that the website of client works with to read the content that has permission to publish, Submit – prevents duplication of information in the first phase, Reporting – basic DB on which are generated statistical reports for clients, Backup – additional data warehouse, Portal – DB is designed to work with application Office. The proposed model of content implementation allows: to build models of communication with the user; to control the dynamic user material; user to influent the client's business; to implement a new type of perception of the world alternate realities. To increase the demand for commercial content only the filtered content is spread on Web pages (fig. 6).
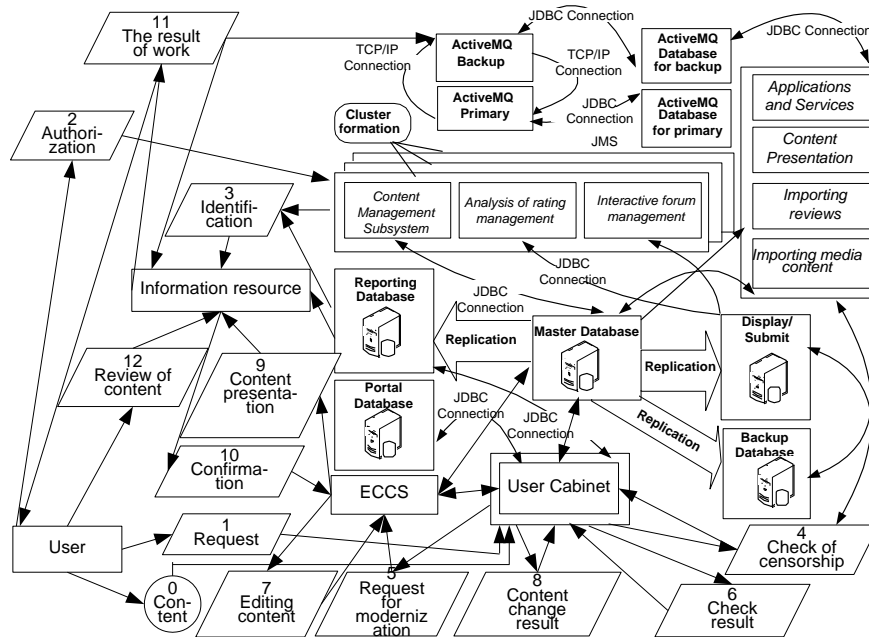

Fig. 6. The scheme of information resources processing in electronic content commerce systems

Analysis of censorship is the following: count the characters and text dropping in the case of failure to be called a review (specific rule on the number of characters); lock of content with the names of competitive products (competitive goods determines moderator); blocking certain senders (database IP addresses); the forming of its own rules (as determined by the client). Analysis, sampling, construction and statistical data moderation is managed by CMS at the implementation phase of content managing according to the rules and algorithms of moderation/synthesis of content (tab. 1).

Table 1. The rules of support of self-learning content

Rule name	Characteristics	Mode
The profanity filter	Check dictionary and ban if TRUE	Semi-automatic, if FALSE checked by moderator, the dictionary is periodically updated by moderators
Filter by number of characters	Counting the number of characters, comparison with a limit and ban if FALSE	Automatic, limit defined by adminstrator or site owner
The filter for the presence of the URL links	Detect URL links, check the ban list and ban if TRUE	Semi-automatic, if FALSE finally checked by moderator, the list is periodically updated by moderators
Blacklist filter	Check the list and ban if TRUE	Semi-automatic, if FALSE finally checks the moderator, the list is periodically updated in case of limit exceeding
Filter by rating	The rating calculation, comparison with the limit and the ban in if FALSE	Automatic, period and limit determined by the administrator or the website owner
Filter on the links of competitors	Check the list and ban if TRUE	Semi-automatic, if FALSE checked by moderator, the list is periodically updated by moderators
The IP addresses filter	Check the list and ban if TRUE	Automatic, the list is periodically updated automatically if the limit is exceeded, the limit is determined by the administrator or the website owner.
Filter by user ID	Check the list and ban if TRUE	Automatic, the list is periodically updated automatically if the limit is exceeded, the limit is determined by the administrator or the website owner.

The design ECCS process is iterative and includes the stages from analysis to prototyping and proof testing, starting with the formation of specs, layout, content formation and its subsequent placement according to the structure of the site (fig. 7,a). The main factors that affect project decisions are the business purpose of the designed system and the needs of end users. Before the definition of functional requirements and the beginning of the development process, the needs of end-users are analyzed by questionnaires, design alternatives and prototypes of varying degrees of readiness, the valuable information is gathered, at the same time causing users a sense of participation in the design process and their trust is gained.
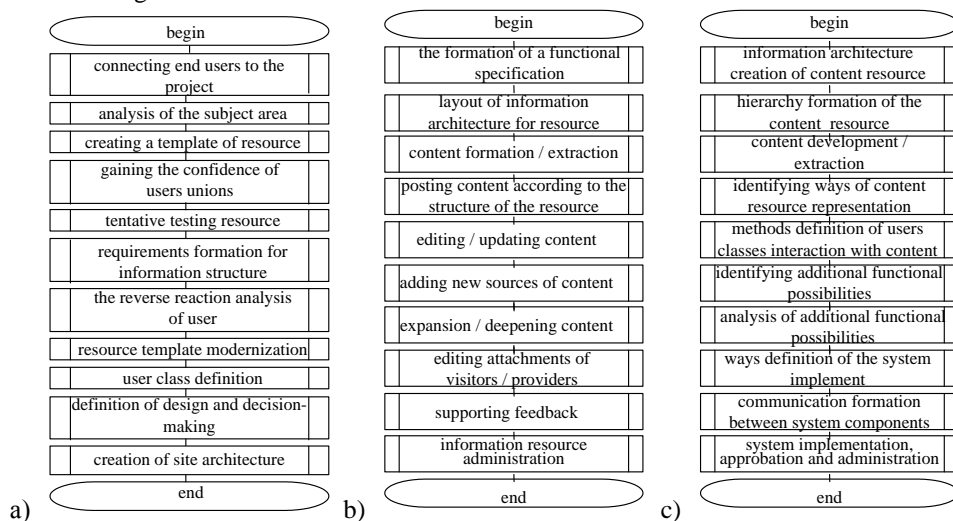
Fig. 7. Process of a) ECCS design, b) organization of site content and b) analysis of the site architecture

To define the information architecture (fig. 7, b) thanks to the users the information about the working groups, past/future conference and all members of the community is determined. On the reverse reaction of users becomes also clear that we need clear and simple architecture. There are three classes of users (or characters) that visiting the site: customers, heads of working groups and administrators that determine site design and decision making process. They wisenut information what is important and how it relates to the main classes of users. They create content site architecture, its hierarchy, methods of presentation and ways of interaction of each user class with this information. For example, information about the conference contains the issues on the agenda or session, the planning of these questions, themes and issues of the conference. During the analysis the additional functionality of SECC (Fig. 7) is

formed. To maintain an active community are added: the discussions and comments on the content, the support for contextual feedback and interaction, the use of unique but neutral brand or visual identity. Web site is a neutral place for interactions between various users, a distinct visual connection with any company or environment raises unwanted reaction. The basic requirements for the ECCS development are the ability to make changes to code and test the changes you made. After testing the changes, this code is available for groups of developers. Interactive development cycle forces you to use a remote version of control system CVS, which allows synchronizing with members of the development team and manage sharing database of source code. Creating a centralized development and test environment optimizes working with code and other members of the group – the time should be spent on writing and testing code, and not the management of files and other system resources. The selection of CMS model affects the need for other tools, for example, in the case of Joomla! it means using PHP, HTML, and Cascading Style Sheets (CSS) to design pages, and MYSQL for the database.

7. Conclusions and further research prospects

In article the general principles of structure construction of electronic content commerce systems, that are implemented formal models of information resources processing. The general architecture of ECCS, described in this work, is developed to facilitate the implementation stages of the life cycle of commercial content. General design principles of ECCS architecture are proposed to implement the information resources processing to reduce the production cycle, time saving and e-commerce empowering. In the work on the basis of the analysis of the ECCS basic tasks the tools, information technologies and software to build such systems are analyzed and summarized. The functional scheme of ECCS with subsystems of information processing resources is developed. The overall architecture of ECCS is describes in detail, its tasks and principles of realization. The most important functional elements of the system and patterns of action of the main mechanisms according to GOST 24.204.80, GOST 24.201-79, 19.201-78, GOST 34.602-89, IEEE Std 1233, 1998 Edition, IEEE Std 830-1998 are described. The software for generation, management and content support are developed. The software implementation of SECC is described, including subsystems of information resources processing for the organization of e-commerce in online newspapers and online magazines.

References

[1]   Berko, A., Vysotska, V., Pasichnyk, V. (2009). Systemy elektronnoyi kontent-komertsiyi. Lviv: NULP.

[2]   Bolshakova, E., Lande, D., Noskov, A., Klyshinsky, E., Peskova, O., Yagunova, E. (2011). Avtomaticheskaya obrabotka tekstov na estestvennom yazyke i kompyuternaya lingvistika. Moskva: MIEM.

[3]   Braychevsky, S., Lande, D. (2005). Sovremennye informatsionnye potoki. *Nauchno-tehnicheskaya informatsiya*, № 11, 21-33.

[4]   Clifton, B. (2009). Google Analytics: professionalny analiz poseschaemosti web-saytov. Moskva: Vilyams.

[5]   Korneev, V., Gareev, A., Vasyutin, S., Rayh, V. (2000). Baza dannyh. Intellektualnaya obrabotka informatsii. Moskva: Nolidzh.

[6]   Lande, D., Furashev, V., Braychevsky, S., Grigorev, O. (2006). Osnovy modelirovaniya i otsenki elektronnyh informatsionnyh potokov. Kyiv: Inzhiniring.

[7]   Lande, D. (2006). Osnovy integratsii informatsionnyh potokov: monografiya. Kyiv: Inzhiniring.

[8]   Sovetov, B., Yakovlev, S. (1998). Modelirovanie sistem. - Moskva: VS.

[9]   Fedorchuk, A. (2005). Kontent-monitoring informatsionnyh potokov. Kyiv.