COMPUTATIONAL PROBLEMS OF ELECTRICAL ENGINEERING

Vol. 3, No. 2, 2013

LIFE CYCLE MODEL OF COMMERCIAL CONTENT PROCESSING IN ELECTRONIC COMMERCE SYSTEMS

Victoria Vysotska, Lyubomyr Chyrun
Lviv Polytechnic National University, Lviv, Ukraine
victana@bk.ru

Abstract: In the given article, a functional logistic model of commercial content processing as a stage of the content life cycle in electronic commerce systems is proposed. The model of commercial content processing describes the processes of forming information resources in electronic content commerce systems, and simplifies the technology of commercial content management. In the given article, the main problems of electronic content commerce and functional services of commercial content processing are analyzed. The proposed model gives an opportunity to create an instrument of information resources processing in electronic commerce systems and to implement the subsystems of commercial content formation, management and support.

Key words: information resources, commercial content, content analysis, content monitoring, content search, electronic content commerce systems.

1. Introduction

Information technology specialists in design, implementation and deployment of electronic content commerce systems (ECCS) deal with information resources processing at various levels. They contribute to an increase in the volume of content sales to a regular user, the active involvement of potential users, and the boundaries expansion of a target audience [1, 2, 5]. The special feature of the ECCS application consists in the following [1-8]: open – access for all companies and users; global – access from anywhere in the world; unlimited in time – available at any time of the day/week/year; frankness - a low barrier to market entry; direct interaction with a user - reducing the channels of distribution and elimination of intermediaries production; information products and information services testing and implementation; automatic processing requests; automatically track information about users; reducing costs for the business operation; providing more on-line information. The need for the ECCS implementation is due to business globalization, increase in needs for content and quick access to the content for running a successful e-business; uneven operation of business processes according to regions (countries, areas, etc.); the need to promptly, regularly and periodically receive a necessary content; time-saving to obtain a desired content; personalization in rendering services in the ECCS; the integrated ECCS [2]. The advantages of the ECCS implementation are increasing the efficiency of content obtaining; reducing the cycle of production and sales; reducing costs associated with the information exchange; openness about users; automatic providing the users with the infiomation on the content; creation of alternative sales channels such as newspapers or online logs in the Internet [2, 5].

The initial information for the process of the ECCS operation is the information on the purpose and working conditions of the system which defines the main aim of ECCS simulation and makes it possible to formulate the requirements to the systems S and content processing subsystems [2] The model of ECCS is $S = \langle X, C, V, H, Function, Z, T, Y \rangle$, where X are the input data effecting the system; Q denotes the users' impact on the system; C stands for the content impact on the system; V is the external environment; H represents the internal parameters of the system; Z denotes the information resources components of the system; T is the time transaction of the content processing; Y stands for the output characteristics of the system [2]. The quantities x_i, c_r, v_l, h_k, y_j are the elements of disjoint subsets containing deterministic and stochastic components [2]. The process of the ECCS operation is described by the function $y_j(t_i + \Delta t) = Function(x_i, c_r, v_l, h_k, t_i)$ [2], where x_i is the visitor/users' request for information to ECCS. According to Google Analytics from [5], y_j is the number of visits per time period Δt; the average time of being on site (min: s); index of failures (%); achieved goal; dynamics (%); the number of all browsing; the number of page views for each visit; new visits (%); absolute unique visitors; traffic sources (%) (search engines, direct traffic or other sites). The effects of the values c_r, v_l, h_k on y_j in ECCS are unknown and unstudied [2].

The ECCS model does not reveal the mechanisms of content processing. [2] Formal models of the content management are intended for determining the processes of content stream aging (relevance), with some of them (logistics, analytical) being intended for the thematic flow analysis [2-4, 6-9]. They do not solve the problem of the content formation and maintenance. Moreover, not all the problems concerning the content management can be solved by them. These are providing an end user with the content at s/he request, history or information portfolio; thematic content identification; automatic generation of digests and information portraits; making tables of concepts relationship; rating the concepts; gathering information from various sources and its formatting; identification of key words/concepts; content duplication finding; content categorization; selective content dissemination [1-8]. The disadvantage of the content management models is the lack of connections between input data, content and output data in the ECCS [2].

The purpose of the paper is creating a functional logistic model of commercial content processing for the information resources formation in e-business systems. The work relevance is that there is a necessity of obtaining operational and objective assessment of the competition in the financial market segment of commercial content and evaluating the financial market competitiveness of the content distribution. In the paper, we shall examine the stages of information resources processing and develop an optimal life cycle for the content processing. Implementing the functional logistic model of commercial content processing enables to create means for the formation of information resources in the e-business systems.

2. Research results analysis

The main subsystems of information resources processing in ECCS are the content formation, management and support, the scheme of their connections being as follows [2]: content formation → content management → content support. The model of the electronic content commerce systems is presented as

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

where the value $X = \{x_1, x_2 \ldots, x_{n_X}\}$ is the set of the input data $x_i \in X$ from different sources of information at $i = \overline{1, n_X}$; the value $Q = \{q_1, q_2 \ldots, q_{n_Q}\}$ is the set of user queries $q_d \in Q$ at $d = \overline{1, n_Q}$; the value $Formation$ denotes the operator of content formation; the value $H = \{h_1, h_2 \ldots, h_{n_Y}\}$ represents the set of internal parameters $h_k \in H$ of the system S when $k = \overline{1, n_H}$; the value $C = \{c_1, c_2 \ldots, c_{n_C}\}$ stands for the set of commercial content $c_r \in C$ at $r = \overline{1, n_C}$; the value $V = \{v_1, v_2 \ldots, v_{n_V}\}$ is the set of the influence parameters of the environment $v_l \in V$ on the system S at $l = \overline{1, n_V}$; the value $Management$ denotes the operator of content management; the value $Support$ is the operator of commercial content support; value $Z = \{z_1, z_2 \ldots, z_{n_Z}\}$ is the set of information resource pages $z_w \in Z$ in the system S at $w = \overline{1, n_Z}$; the value $T = \{t_1, t_2 \ldots, t_{n_T}\}$ represents the time $t_p \in T$ transaction of information resource processing in the system S when $p = \overline{1, n_T}$; the value $Y = \{y_1, y_2 \ldots y_{n_Y}\}$ is the set of statistical data $y_j \in Y$ in the system S at $j = \overline{1, n_Y}$.

The ECCS operation is described by the following schemes of its main components relationships [2]:

1) to form information resources: *data → content formation → content database → content management → information resource of the system*;

2) to answer a user's query: *user's query → content management → information resource → content support → users database*;

3) to prepare a report on the system operation to the moderator: *moderator's query → content support → users database → content management → report*;

4) to moderate the internal parameters of the system: *query → content formation → base of rules → content support → base of rules → content management → result*.

The content formation is described by the operator of the given form $c_r = Formation(u_f, x_i, t_p)$, where u_f is the set of conditions for the content formation, ie $u_f = \{u_1(x_i), \ldots, u_{n_U}(x_i)\}$. The content is presented as follows:

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

The formation of commercial content for the information resource provides a link between the input data from different sources and the formed commercial content stored in an appropriate database in the electronic content commerce systems, i.e.

$$Sourse(x_i) \rightarrow x_i \rightarrow X \rightarrow Formation(u_f, x_i, t_p) \rightarrow c_r \rightarrow C \rightarrow DataBase(C),$$

where $Sourse(x_i)$ is the sourse of x_i content, x_i is the i-th content from the source, X represents the set of the data from a relevant source, $Formation(u_f, x_i, t_p)$ stands for the operator of the content formation at a fixed time t_p under the appropriate conditions u_f, c_r is the r-th commercial content formed under the conditions u_f, C is the set of the generated content, $DataBase(C)$ denotes the database of the generated commercial content.

The model of forming the content in electronic content commerce systems can be showed as

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

where $X = \{x_1, x_2 \ldots, x_{n_X}\}$ is the set of input data $x_i \in X$ from different information resources or moderators at $i = \overline{1, n_X}$; $Gathering$ denotes the content collecting/creating operator from various sources; $Formatting$ is the content formatting operator; $KeyWords$ represents the operator of identifying key words and concepts; $Categorization$ is the content categorization operator; $Backup$ stands for the operator of the content duplicate detection; $BuDigest$ is the content digest formation operator; $Dissemination$ is the selective content distribution operator; $T = \{t_1, t_2 \ldots, t_{n_T}\}$ is the content forming transaction time $t_p \in T$ when $p = \overline{1, n_T}$; $C = \{c_1, c_2 \ldots, c_{n_C}\}$ denotes the set of the commercial content $c_r \in C$ when $r = \overline{1, n_C}$.

The content formation is described by the operator in the form $c_r = Formation(u_f, x_i, t_p)$, where u_f is the set of conditions to generate the content, i.e. $u_f = \{u_1(x_i), \ldots, u_{n_U}(x_i)\}$.

The commercial content is represented as follows:

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

It passes the following steps of the convertion of the dataset into a set of a relevant, formatted, classified and validated content:

$$x_i \in X \rightarrow Gathering(u_f, x_i, t_p) \rightarrow Backup(c_r, u_b, t_p) \rightarrow Formatting(c_r, t_p) \rightarrow KeyWords(c_r, t_p) \rightarrow Categorization(c_r, t_p) \rightarrow BuDigest(c_r, t_p) \rightarrow Dissemination(c_r, t_p) \rightarrow c_r \in C.$$

The decisions that can help to navigate in the dynamic input information from different sources are provided by the data syndication $C = Gathering(X, U_G, T)$, i.e. by gathering information from the sources and further distribution of its fragments according to the users' needs, where X is the set of the content different information sources, U_G is the set of conditions needed for data collecting from various sources, $Gathering$ represents the content collecting/creating operator, T is the time of the content collecting/creating.

The detection of content duplicating is denoted by the operator such as $C = Backup(Gathering(X, U_G, T), U_B)$, where X is the set of content from different data sources, U_B is the set of conditions needed to detect the content duplication, $Backup$ is the operator of detecting the content duplication, C stands for the content set. The identification of content duplication is based on the linguistic statistical methods for detecting general terms, whose chains form the verbal content signatures.

Content syndication technology contains the process of filtering information when collecting the data with structural characteristics from individual sources (from information resources, from moderators, users, visitors, journalists and editors), direct scanning of the content and bringing to the total:

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

where $Formatting$ is the content formatting operator, U_G denotes the set of conditions needed to collect information from various sources, $Gathering$ is the content collecting/creating operator, U_{FR} represents the set of conditions needed for information formatting, T is the content collection time.

Processing of the set of content C which is built on the principle of finding keywords in the content (terms) is based on the Zipf's law and reduced to the choice of the words with an average frequency of their use (the most used words are ignored by a stop-dictionary, and rare-used words from the messages texts are not taken

into account). The detection of keywords and concepts is defined by the operator $KeyWords(C)$ and described by the operator below:

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

where $KeyWords$ is the operator for the identification of the content keywords and concepts that is implemented as a set of processes; $Formatting$ is the content formatting operator; U_G stands for the set of conditions for data collecting from various sources; $Gathering$ represents the content collecting/creating operator; U_{FR} is the set of conditions for data formatting; T is the content collection time; U_K denotes the set of conditions for identifying key words and concepts.

The database of terms/morphemes, as well as the rules of text analysis are used for searching the terms. Based on the generated grammar rules we can perform the correction of the term by using its context. Classification and content distribution means are the information retrieval system for the selective content distribution (Content Router).

The content is analyzed for compliance to thematic requests

$$C_{Ct} = Categorization\left(KeyWords(C,U_K),U_{Ct}\right),$$

where $KeyWords(C,U_K)$ is the keywords identification operator, $Categorization$ is the operator of content categorization according to the keywords identified, U_K stands for the set of conditions for the keywords identification, U_{Ct} is the set of categorization conditions, C_{Ct} is the set of rubricated relevant content. The digest set C_D is formed by such a dependence as $C_D = BuDigest\left(C_{Ct},U_D\right)$, where $BuDigest$ is the digests forming operator, U_D represents the set of conditions for the digests formation, C_{Ct} is the set of rubricated relevant content, i.e.

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

The content is sent to users and uploaded into thematic databases. The selective distribution of the content is described as $C_{Ds} = Dissemination\left(C_D,U_{Ds}\right)$, where C_{Ds} is the set of the selectively distributed content, U_{Ds} is the set of conditions for the selective content distribution, $Dissemination$ is the selective content distribution operator.

The step of managing the content is described by the operator of the following form:

$$z_w = Management(q_d,c_r,h_k,t_p),$$

where $Q = \{q_1,q_2\ldots,q_{n_Q}\}$ is the set of user's queries; h_k is the set of conditions for the content management, ie $H = \left\{h_1(c_{i+1},q_d),\ldots,h_{n_H}(c_{i+n_H},q_d)\right\}$.

The management of commercial content is presented as

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

The step of the content support $Support$ is described by the operator of the following form

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

where v_l is the set of conditions for the content support and external influences of the environment on the system, ie

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

The output statistics is implemented as follows

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

The following formula reflects the commercial content support in the ECCS, that is describes the analysis of the users' reaction to the information resources processing in such systems. The functional logistic model describes the process of the ECCS operation with the basic processes of information resources processing as the content formation, management and maintenance. A statistics analysis of the ECCS operation is conducted according to the analysis of the regular/potential user' reactions to the system (visit, requests, search for keywords, etc). Applying the operations of elimination, sequencing and parallelization in the proper order facilitates an effective analysis of the target/potential audiences' response to the ECCS functioning. It also helps to predict relevant changes concerning the demand for commercial content. The proposed general design principles of the ECCS

architecture allow the process of information resources processing to be implemented in order to expand the functional capabilities of ECCS.

3. Conclusion

In the given paper, a functional logistic model of commercial content processing in e-business systems is developed. The model is based on the multilayered structure of the processes. This model involves the division of the overall process into the following stages: collecting/creating a content from different sources; formatting a content; identifying key words and concepts; categorizing a content; detecting the content duplication; forming digests; selective distribution of the content between moderators and users of ECCS. The model is based on the principles of content analysis. This enables to automatize various steps of creating an information product of this type without losing the content and lowering the quality. The method effectiveness is confirmed by the results of its application to the development of a number of commercial content projects. The developed means for the automation commercial content processing allow the processes of content formation, management and maintenance to be speeded up. These means also contribute to the increase in rating of the generated commercial information resources.

References

[1] A. Bereza , I. Kozak, and F. Levchenko, *Ecommerce*. Kyiv, Ukraine: KNEU, 2002. (Ukrainian)

[2] A. Berko , V.Vysotska and V.Pasichnyk*, Electronic content commerce systems*. Lviv, Ukraine: NULP, 2009. (Ukrainian)

[3] E. Bolshakova , E. Klyshynskyy , D. Lande , A. Noskov , O. Peskova , and E. Yahunova , *Automatic processing naturally texts on language and computer linguistics*. Moscow, Russia: MYEM, 2011. (Russian)

[4] S. Braychevskyy and D. Lande , "Modern information streams", *Naukovo-tekhnichna Informatsiya*, no. 11, pp. 21-33, Kyiv, Ukraine, 2005. (Ukrainian)

[5] B. Clifton , *Google analytics: professional attendance analysis web sites*. Moscow, Russia: Williams, 2009. (Russian)

[6] B. Korneyev , A. Gareyev , S. Vasiutin , and W. Reich , *Databases. Intelligent processing of information*. - Moscow, Russia: Znaniye, 2000. (Russian)

[7] D. Lande , V. Furashev , S. Braychevskyy, and A. Grigoriev, *Modeling and evaluation electronic information flows fundamentals*. Kyiv, Ukraine: Tekhnika, 2006. (Ukrainian)

[8] D. Lande *Fundamentals of information flows integration*. Kyiv, Ukraine: Tekhnika, 2006. (Ukrainian)

[9] B. Pasichnyk , Y. Sherbyna , V.Visotska and T. Shestakevich, *Mathematical linguistics*. Lviv, Ukraine: "Novyi Svit - 2000", 2012. (Ukrainian)

МОДЕЛЬ ЖИТТЄВОГО ЦИКЛУ ОПРАЦЮВАННЯ ЕЛЕКТРОННОГО КОМЕРЦІЙНОГО КОНТЕНТУ В СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

Вікторія Висоцька, Любомир Чирун

Запропоновано функціонально-логістичну модель опрацювання комерційного контенту як етап життєвого циклу контенту в системах електронної комерції. Модель опрацювання комерційного контенту описує процеси формування інформаційних ресурсів у системах електронної контент-комерції та спрощує технологію управління комерційним контентом. Проаналізовано основні проблеми електронної контент-комерції та функціональних сервісів опрацювання комерційного контенту. Запропонована модель дає можливість створити засоби опрацювання інформаційних ресурсів у системах електронної контент-комерції та реалізувати підсистеми формування, управління та супроводу комерційного контенту.

Victoria Vysotska – an assistant of Department of Information systems and networks at Lviv Polytechnic National University, Ukraine.

Research intersts: information systems and networks; e-commerce; information resources; commercial content; content analysis; content monitoring; content search; electronic content commerce systems; software systems; models, algorithms, analysis, methods and strategies of systems design.

Lyubomyr Chyrun – Ph.D. associate professor of the Department of Software at Lviv Polytechnic National University, Ukraine.

Research intersts: information systems and networks; e-commerce; information resources; commercial content; content analysis; content monitoring; content search; electronic content commerce systems; software systems; models, algorithms, analysis, methods and strategies of systems design.