

УДК 004.738.5

ОПРАЦЮВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

В. А. Висоцька, Л. В. Чирун

Національний університет “Львівська політехніка”

E-mail: victana@bk.ru

Проаналізовано основні проблеми електронної контент-комерції та функціональних сервісів опрацювання комерційного контенту. Запропонований метод дає можливість створити засоби опрацювання інформаційних ресурсів у системах електронної контент-комерції та реалізувати підсистему управління комерційним контентом.

Ключові слова: *Web-ресурс, контент, контент-аналіз, контент-моніторинг, контент-ний пошук, система електронної контент-комерції.*

INFORMATION RESOURCES PROCESSING IN ELECTRONIC CONTENT COMMERCE SYSTEMS

V. A. Vysotska, L. V. Chyrun

Lviv Polytechnic National University

E-mail: victana@bk.ru

The development of unified methods and software tools for processing information resources in the electronic content commerce systems is discussed. A new detailed classification of electronic content commerce systems and a model of electronic content commerce systems are proposed. The models of information resource processing in electronic content commerce systems are also presented. Architecture and models of electronic content commerce systems are built. A new approach to application and implementation of business processes is formulated for the construction of the systems of electronic content commerce. A complex method of content creation, the operational method of content management and complex method of content support and also the software tools for content creation, management and support are developed. Design and implementation methods of electronic content commerce systems based, on online newspapers, which reflect the results of theoretical research, are developed. From the perspective of a systemic approach, the principles of applying information resources processing in electronic content commerce systems for content lifecycle implementation make it possible to develop the methods for the commercial content formation, management and support. An integrated method of commercial content formation for the time and resources reduction of content production is developed. Thus, the creation of means of information resources processing and implement subsystem of automatically generated content is possible. A method of commercial content management for the time and resources reduction of content sales, which makes it possible to implement commercial content management subsystem, was created. A method of commercial content support for the time and resource reduction of the target audience analysis in electronic content commerce systems is implemented, which makes it possible to develop a commercial content support subsystem.

Keywords: *Web resources, content, content analysis, content monitoring, content search, electronic content commerce systems*

Загальна постановка задачі. Активний розвиток Інтернету сприяє зростанню потреб в інформації як виробничого чинника та стратегічного ресурсу і реалізації нових форм інформаційного обслуговування. Документована інформація, підготовлена відповідно до потреб користувачів і призначена для їх задоволення, є ін-

формаційним продуктом або комерційним контентом (наприклад, електронний матеріал Інтернет-видавництва, маркетингові дослідження, консалтингові послуги тощо). Дії для забезпечення користувачів комерційним контентом – інформаційна послуга. Інтернет-ринок – це сукупність економічних, правових, організаційних і програмних відносин з продажу/купівлі інформаційних продуктів та послуг (ШП) між виробниками/постачальниками і користувачами. Поняття комерційний контент визначається як вміст інформаційних ресурсів (наприклад, Web-сайтів) у системі електронної контент-комерції (СЕКК); об'єкт бізнес-процесів СЕКК, наприклад, стаття, програмне забезпечення (ПЗ), книга тощо; структурована множина, логічно завершена інформація, яка є об'єктом взаємовідносин між користувачем та СЕКК; набір даних без наперед визначеної структури, які існують лише в електронному вигляді; інформація комерційного призначення та неподільна в часі; основний чинник формування області діяльності, функціонування та призначення СЕКК [1].

Аналіз динаміки потоку контенту та побудова етапів опрацювання Web-ресурсів є важливим та актуальним. Розроблення і впровадження СЕКК є одним із стратегічних напрямів розвитку е-бізнесу. Характерною рисою таких систем є можливість автоматичного опрацювання Web-ресурсів для збільшення обсягів продажу контенту постійному користувачу, активного залучення потенційних користувачів та розширення меж цільової аудиторії. Зокрема, принципи і технології е-комерції активно застосовують під час створення систем on-line/off-line продажу та аналізу/обміну/збереження контенту, Інтернет-магазину, cloud storage/computing. Відсутність загального стандартизованого підходу до проектування СЕКК та опрацювання Web-ресурсів призводить до виникнення низки проблем під час розроблення типової структури відповідних систем. Web-ресурс є множиною структурованого контенту (наприклад, електронні видання у вигляді статей, анонсів, дайджестів, книг, репортажів, блогів, коментарів тощо) в СЕКК (наприклад, Інтернет-газета або Інтернет-журнал), призначеної для задоволення інформаційних потреб цільової аудиторії та сприйняття людиною за допомогою відповідних програмних і апаратних засобів через Інтернет. Автоматичне опрацювання Web-ресурсів контенту покращує результати аналізу текстового контенту та зменшує час пошуку необхідного контенту лише за введення відповідної адреси або ключових слів у пошукову систему [1].

Аналіз останніх досліджень та публікацій. Аналіз процесу опрацювання Web-ресурсу дозволяє зменшити витрати на формування, публікацію, зберігання, управління, супровід і донесення готового актуального та релевантного комерційного контенту постійному користувачу [1–3]. Публікація комерційного контенту на Web-ресурсах надає додаткові можливості всім авторам релевантного контенту, який актуальний для вузького кола цільової аудиторії, але не публікується через збитковість традиційними видавництвами. Автори контенту мають змогу публікуватися самостійно, формуючи тексти в редакторах і розміщуючи їх на Web-ресурсах для поширення з врахуванням авторського права. Відомим методом аналізу змісту текстової інформації (коментарів, форумів, е-листування, статей тощо) є *контент-аналіз* (рис. 1). Поняття контент-аналізу не має однозначного визначення [1], тому інформаційні системи, побудовані на основі різних підходів, є несумісними. Застосування контент-аналізу тексту в СЕКК дає ряд переваг для спрощення ведення е-бізнесу та вирішує низку проблем учасників бізнес-процесів, а саме:

- 1) фільтрація контенту користувача на Web-ресурсі;
- 2) можливість автоматичного створення *портрету постійного користувача* на основі аналізу його коментарів та історії відвідування сторінок;

- 3) можливість автоматичного створення *портрету цільової аудиторії* на основі аналізу портретів постійних користувачів;
- 4) скорочення кількості модераторів Web-ресурсу в CEKK;
- 5) скорочення часу для розміщення контенту користувача на Web-ресурсі через автоматичне його опрацювання, а не модераторами;
- 6) ліквідація мовного бар'єра через автоматичне формування словників постійного користувача та автоматичного перекладу.

Для розроблення CEKK використовують CMS (*Content management system*) [1], якщо вона відповідає визначеному набору вимог (рис. 2). CMS не вирішують основних задач опрацювання ресурсів – формування та супроводу контенту [1]. Основний недолік CMS полягає у відсутності зв'язку між вхідними даними, контентом та вихідними характеристиками функціонування системи (рис. 3) [1]. CMS дозволяють управляти контентом через зручні інструменти зберігання і публікації. Цінність контенту визначає його привабливість для споживача. Інтеграція контенту робить Web-ресурс привабливим, а інтеграція застосувань – корисним [1]. CMS – це інструмент (див. таблицю) для моделювання розгалужених структур Web-ресурсів в CEKK і управління їх контентом [1] без спеціальних технічних навичок. CMS використовують для генерування контенту для Web-ресурсів з такими проблемами, як динамічне збирання, кешування та безпека контенту [1]. CMS забезпечує доступ і внесення змін до Web-ресурсу та покликана максимально спростити управління Web-ресурсом за гнучкості налаштувань і контролю.

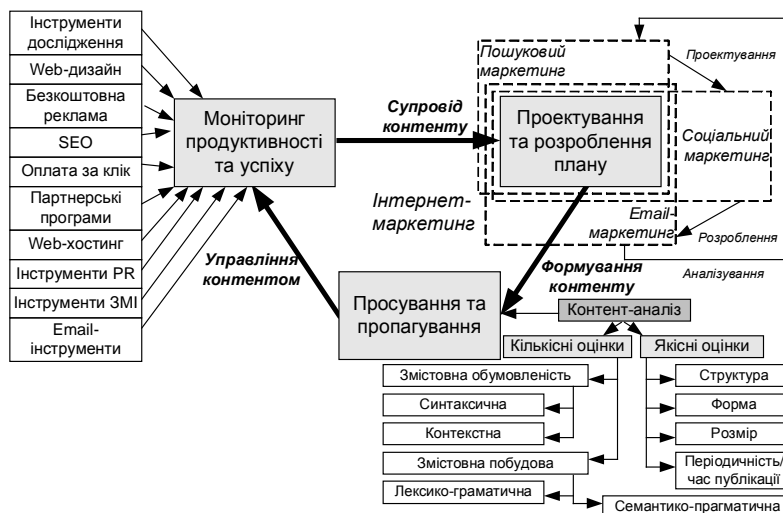


Рис. 1. Інтернет-маркетинг для систем електронної контент-комерції.

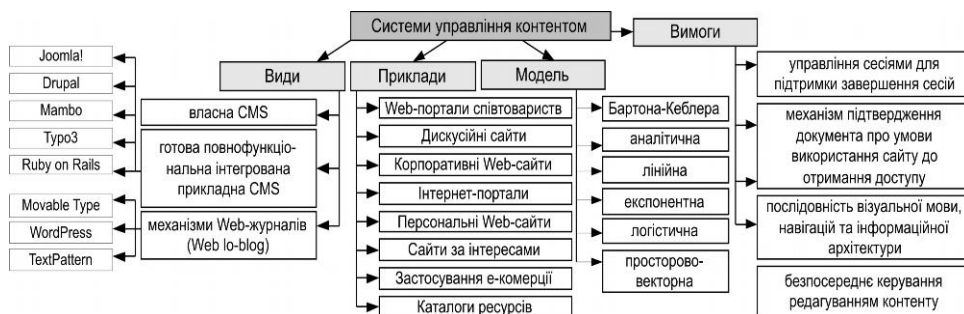


Рис. 2. Основні вимоги до систем управління контентом для побудови CEKK.

Використання CMS не вимагає встановлення ПЗ. Для редагування й адміністрування використовують браузер. Інтуїтивний інтерфейс і простота роботи зі системою полегшує управління Web-ресурсом та знижує подальші витрати на його підтримку [1]. CMS включає такі можливості: швидке поновлення та пошук контенту в Web-ресурсі; збирання даних про клієнтів та потенційних клієнтів; формування і редагування опитувань; аналіз відвідування Web-ресурсу. В CEKK присутні великі потоки та обсяги різного контенту. Значна частина цих потоків контенту складена з легко формалізованих і автоматизованих процедур та комерційного контенту. Але відсутній загальний підхід до формалізації, проектування, розроблення та впровадження CEKK. Не описані та не розкриті в таких системах зв'язки та залежності між вхідними даними, комерційним контентом, вихідними даними й процесами опрацювання Web-ресурсів.

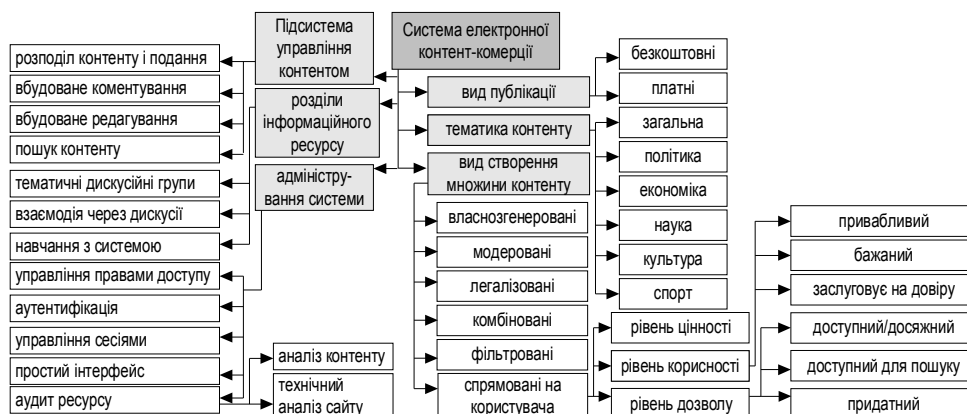


Рис. 3. Основні компоненти систем управління контентом для побудови CEKK.

Характеристика основних CMS для розроблення CEKK

Назва системи управління контентом	Вимоги до ПЗ			Простота установки	Крива навчання	Управління сесіями	Управління користувачами	Розширюваність	Масштабованість	Використання тем	xHTML/CSS
	Web-сервер	База даних	Мова								
Ruby	Apache	MySQL	Ruby	+/-	+/-	+/-	+/-	+	+	+/-	+
Drupal	Apache IIS	MySQL	PHP	+/-	+/-	+	+	+	+	+	+
Mambo	Apache IIS	Apache IIS	PHP	+	+/-	+/-	+	+/-	+	+/-	+
Туро3	Apache IIS	Apache IIS	PHP	-	-	+	+	+	+	+/-	
Movable Type	Apache IIS	Apache IIS	Perl	+	+/-	-	+/-	+/-	+	+/-	-
Word Press	Apache	Apache	PHP	+	-	-	+/-	+/-	-	+/-	+
Text Pattern	Apache	Apache	PHP	+	-	-	+/-	-	-	+/-	+
Joomla!	Apache	MySQL	PHP	+	+	+	+	+	+	+	+

Виділення проблем. Більшість визначень контент-аналізу конструктивні, тобто процедурні. Через різні початкові підходи вони породжують різноманітні алгоритми, які часом суперечать один одному [1–3]. Найбільші сумніви викликає

ігнорування ролі контексту. Практичне значення методу дозволяє уникнути багатьох суперечностей. Об'єднання засобів і методів та їх природний відбір шляхом багатократного оцінювання отриманих результатів дають можливість виділити або підтвердити знання і фактичну силу/корисність інструментарію. Контент-аналіз – кількісно-якісний аналіз масивів тексту для подальшої змістовної інтерпретації отриманих кількісно-якісних закономірностей. Метод полягає у формуванні з різноманіття тексту абстрактної моделі змісту контенту. Його застосовують під час аналізу джерел, інваріантних за структурою/змістом у вигляді несистематизованого, неструктурованого тексту [1–3]. Результат контент-аналізу тексту використовують під час визначення тональності тексту, дублювання контенту, наявності спаму та виявлення нових подій для встановлення тематичних сюжетів потоків контенту. Визначення тональності контенту на основі аналізу текстів є складнішим за виявлення спаму. Для виявлення спаму розглядають дві гіпотези (спам, не спам), а для визначення тональності перевіряють емоційне забарвлення (позитивне, негативне, нейтральне) та їх комбінації. У методі Байеса для визначення спаму використовують базу оцінок – двох корпусів контенту, один із яких складений зі спаму, а інший – ні [2, 3]. Для кожного контенту підраховують частоту використання кожного слова та вагову оцінку (від 0 до 1), тобто умовну імовірність, що контент з цим словом є спамом [2, 3]. Значення ваг, близькі до 1/2, не враховують під час інтегрованого розрахунку, тому слова з такими вагами ігнорують і видаляють. Простір гіпотез містить такі тональності: негативну T^{-1} , нейтральну T^0 і позитивну T^{+1} . При гіпотезах T^{+1} із множини з позитивною тональністю обирають терми, характерні для цього контенту. Із них обирають терми t із ймовірністю, розрахованою за формулою Байеса та більшою за 1/2. Рішення про тональність контенту приймають з урахуванням різниці значень T^{+1} і T^{-1} [2, 3]. Для знаходження подібного контенту і дублікату справедливе правило рефлексивності, але не виконується умова транзитивності. Для визначення дублювання використовують поняття симетричності та транзитивності. Відношення із властивостями рефлексивності, симетричності та транзитивності є відношенням еквівалентності [2, 3], тобто відношенням дублювання. Для виявлення нових подій з потоку контенту, що послідовно надходить на вхід СЕКК від засобів сканування або контентного роутера та відбирається за тематичним запитом, визначають нові події, описані в контенті [1–3]. Для них формують сюжетні ланцюжки подібного контенту. Контент, що відображає нові події, є основою кластерів взаємозалежного контенту [1–3].

Аналіз отриманих наукових результатів. Підсистемами опрацювання Web-ресурсів в СЕКК є формування, управління, супровід контенту та управління Web-ресурсом, схема зв'язків яких є такою: *формування контенту* → *опрацювання Web-ресурсу* → *управління контентом* → *опрацювання Web-ресурсу* → *супровід контенту* → *опрацювання Web-ресурсу*. Необхідно розв'язати науково-практичну задачу розроблення методу опрацювання комерційного контенту шляхом автоматизації управління Web-ресурсом в СЕКК на основі нечіткої логіки.

І. Метод формування комерційного контенту – комплекс заходів забезпечення контролю опрацювання даних з різних джерел інформації для створення комерційного контенту з набором додаткових значень (актуальність, достовірність, унікальність, повнота, точність тощо). Створення комерційного контенту описано оператором $C_0 = \alpha_0(X, U_C, T)$, де U_C – множина умов створення комерційного контенту. Задачу збирання інформації з джерел описано оператором вигляду $C_0 = \alpha_1(X, U_G, T)$, де U_G – множина умов збирання даних з різних джерел. Задачу виявлення дублювання змісту комерційного контенту описано оператором α_2 вигляду $C_1 = \alpha_2(\alpha_0(X, U_C, T), U_B)$ та $C_1 = \alpha_2(\alpha_1(X, U_G, T), U_B)$, або

$C_1 = \alpha_2(C_0, U_B)$, де U_B – множина умов виявлення дублювання змісту комерційного контенту. Виявляють дубльований за змістом комерційний контент в СЕКК за допомогою лінгвостатистичних методів знаходження загальних термів, ланцюжки яких утворюють словесні сигнатури комерційного контенту (текст є унікальним за коефіцієнта унікальності $\geq 80\%$). Задачу сканування комерційного контенту та приведення до загального формату в XML описано оператором α_3 вигляду

$$C_2 = \alpha_3(\alpha_2(C_0, U_B), U_{FR}), \quad (1)$$

де U_{FR} – множина умов форматування комерційного контенту.

Опрацювання множини контенту C_2 для виявлення значущих ключових слів побудоване на принципі знаходження ключових слів за змістом (термами), базується на законі Зіпфа і зводиться до вибору слів зі середньою частотою появи (найуживаніші слова ігноруються за допомогою стоп-словника, а рідкісні слова з текстів повідомлень не враховують). Знаходження ключових слів та понять з використанням словників визначається оператором $\alpha_4(c_2, U_K)$ вигляду $C_3 = \alpha_4(\alpha_3(\alpha_2(C_0, U_B), U_{FR}), U_K)$ при $U_K = \{U_{K1}, U_{K2}, U_{K3}, U_{K4}\}$, де U_K – колекція умов виявлення ключових слів та понять у тексті; U_{K1} – множина всіх термів (термом є основа іменника, іменник, словосполучення іменників або прикметника з іменником); U_{K2} – множина частот вживання терму в тексті комерційного контенту; U_{K3} – множина коефіцієнтів вживання термів з врахуванням кількості знаків без пробілів (за 2000...3000 знаків частота ключових слів у межах 4...6%, до 2000 знаків – 6...8%, понад 3000 знаків – 2...4%), U_{K4} – множина термів, які відповідають умовам належності до ключових слів.

Задачі класифікації та розподілу контенту реалізують через інформаційно-пошукову систему вибіркового поширення контенту. Комерційний контент аналізують на відповідність запитам з використанням результатів рубрикації. Оператор рубрикації комерційного контенту, згідно з виявленими ключовими словами, описаний як $C_4 = \alpha_5(\alpha_4(C_2, U_K), U_{CT})$ при $U_{CT} = \{U_{CT1}, U_{CT2}, U_{CT3}, U_{CT4}\}$, де U_{CT} – колекція умов рубрикації; U_{CT1} – множина тематичних ключових слів із словника; U_{CT2} – множина частот вживання ключових слів у комерційному контенті; U_{CT3} – множина залежностей вживання ключових слів різних тематик (коефіцієнти визначає модератор згідно з належністю ключового слова до певної тематики в межах $[0, 1]$); U_{CT4} – множина частот вживання тематичних ключових слів у контенті. Множину дайджестів C_5 формують залежністю $C_5 = \alpha_6(C_4, U_D)$, де U_D – множина умов формування дайджестів контенту, тобто $C_5 = \alpha_6(\alpha_5(\alpha_4(C_2, U_K), U_{CT}), U_D)$. Релевантний контент розсилають користувачам та завантажують у бази даних. Вибіркове поширення контенту описано так: $C_6 = \alpha_7(C_5, U_{DS})$, де U_{DS} – множина умов вибіркового поширення контенту. Вибіркове поширення списку комерційного контенту $C_7 = \max(C_6)$ залежить від рівня попиту на цей контент. Асоціативне правило формування списку контенту складається з сервісного списку контенту C_6 і списку контенту, обраного із первісного – похідний список C_7 , тобто $C_6 \rightarrow C_7$. Формування асоціативного правила – це формування списку контенту, утвореного внаслідок об'єднання первісного і похідного списків. Асоціативне правило появи контенту із первісного списку C_6 разом із контентом із списку C_7 в базі даних є таким оператором:

$$P = \varphi(C_6, C_7) = \frac{\max(C_7 \cup C_6)}{\max(C_6)}, \quad (2)$$

де $d_s = \max(C_6)$ – максимум множини контенту C_6 при $\forall c_{6i} \in C_6 \Rightarrow c_{6i} \leq d_s$;
 $d = \max(C_7 \cup C_6)$ – максимум $C = C_6 \cup C_7$ при $\forall c_i \in C = (C_6 \cup C_7) \Rightarrow c_i \leq d$, де

$$d_s = \max(C_6) \Leftrightarrow d_s \in D_{C_6} \wedge \forall c_{faset} \in D_{C_6} : d_s \leq c_{faset}, \quad (3)$$

$$d = \max(C_6 \cup C_7) \Leftrightarrow d \in D_{C_6 \cup C_7} \wedge \forall c_{faset} \in D_{C_6 \cup C_7} : d \leq c_{faset}, \quad (4)$$

де $D_{C_6} = \{c_{faset} \in C_{faset} \mid \forall c_{6i} \in C_6 : c_{6i} \leq c_{faset}\}$ – множина максимумів для C_6 при значеннях потужностей множин $\rho_{faset} = |C_{faset}|$, $\rho_6 = |C_6|$, $\rho_{faset} \geq \rho_6$; $D_{C_6 \cup C_7} = \{c_{faset} \in C_{faset} \mid \forall c_i \in (C_6 \cup C_7) : c_i \leq c_{faset}\}$ – множина максимумів для $C_6 \cup C_7$ при $C_{faset} \geq (C_6 \cup C_7)$. Показники прибутковості I_p та зростання попиту I_g на комерційний контент використовують для визначення його актуальності та розраховують так:

$$I_p = \kappa(C_7, P) = P \frac{\max(C_6 \rightarrow C_7)}{\max(C_7)}, \quad (5)$$

$$I_g = \chi(C_7, P) = \frac{\max(C_6 \rightarrow C_7) - \max(C_7)}{\max(C_6 \rightarrow C_7)[1 - P]}. \quad (6)$$

Похідний список обраних правил визначає модератор. Часто вживаним списком контенту є список, формування якого перевищує мінімальний рівень. За встановлення цієї умови формується велика кількість списків часто вживаного контенту відповідно до запитів. Результат обмежується під час вибірки операторами

$$I_r = \psi(C_6, C_7, P) = \min(\forall C'_6 \subset C_6, P - P'), \quad (7)$$

$$I_r = \min(\forall C'_6 \subset C_6, \varphi(C_6, C_7) - \varphi(C'_6, C_7)). \quad (8)$$

II. Метод управління комерційним контентом – комплекс заходів забезпечення підтримки визначальних параметрів комерційного контенту (актуальність, повнота, релевантність, автентичність, достовірність) відповідно до визначених вимог до набору критеріїв управління комерційним контентом. Класифікація процесів управління комерційним контентом така.

1. Управління контентом з генерацією сторінок за запитом користувача СЕКК подано так:

$$\beta_Q = \langle C, Q, H, U, T, Z, \beta_1, \beta_2, \beta_3, \beta_4 \rangle. \quad (9)$$

Етап редагування та модифікації комерційного контенту подано оператором вигляду $c_j(t_{r+1}) = \beta_1(c_j, t_r, h_k, u_l)$ при $c_j(t_{r+1}) \in C$. Етап формування множини сторінок описано оператором $Z(t_r) = \beta_4(q_i, C, \beta_3(\beta_2(C), t_r))$, де

$$z_i = \left\{ \bigcup_{j=1}^m c_j(q_i, t_r) \mid \begin{array}{l} \forall c_j \in C_q, c_j \notin C_q^-, C_q = \beta_3(\beta_2(C_q)), \exists q_i \in Q_c, \exists h_k \in H_c, h_k \notin H_c^-, \\ C = C_q \cup C_q^-, Q_c \subset Q, H = H_c \cup H_c^-, k = \overline{1, n_H}, i = \overline{1, n}, r = \overline{1, w} \end{array} \right\}.$$

Вагу блоку визначають як суму коефіцієнтів ваг комерційного контенту:

$$\omega = \|C\| = \beta_2(C, \omega_1, \omega_2, \omega_3, \omega_4, \omega_5), \quad (10)$$

де $\omega_1(c_j)$ – коефіцієнт розташування блоку у комерційному контенті; $\omega_2(c_j)$ – коефіцієнт ключових слів у блоці; $\omega_3(c_j)$ – коефіцієнт статистичної важливості термів; $\omega_4(c_j)$ – коефіцієнт додаткових термів; $\omega_5(c_j)$ – коефіцієнт наявності та обсягу термів із запиту користувача.

2. Управління контентом з генерацією сторінок за редагування Web-ресурсу модератором

$$\beta_E = \langle C, H, T, Z, \beta_1, \beta_2, \beta_3 \rangle. \quad (11)$$

Етап формування Web-сторінок описано оператором $Z(t_r) = \beta_3(C, H, t_r, \beta_1, \beta_2)$.

3. Змішаний тип управління комерційним контентом подано так:

$$\beta_M = \langle C, Q, H, T, Z, W, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 \rangle, \quad (12)$$

де W – множина кешованого комерційного контенту; β_5 – оператор формування множини кешованого контенту або інформаційних блоків з нього при $W = \beta_5(C, \beta_3(\beta_2(\beta_1(C, t_r, H, U)), t_{r+1}))$ або $W = \beta_5(Z, \beta_3(\beta_2(\beta_1(C, t_r, H, U)), t_{r+1}))$, де

$$w_l = \left\{ \bigcup_{i=1}^n c_i \mid \forall c_i \in C_Q, C_Q \subset C, C_Q = \beta_3(\beta_2(C)) \right\}, \quad (13)$$

$$w_l = \left\{ \bigcup_{j=1}^m z_j \mid \begin{array}{l} z_j \in Z_c, \forall c_j \in C_z, \exists c_j \in Z_c, \forall c_j \in z_j, C_z = \beta_3(\beta_2(C)), \\ C_z \subset C, Z_c \subseteq Z, i = \overline{1, n} \end{array} \right\}.$$

III. Метод супроводу комерційного контенту – комплекс заходів забезпечення функціонування СЕКК згідно з визначеними вимогами та їх подальших змін. У результаті аналізу функціонування СЕКК та технологічного супроводу комерційного контенту C формують колекцію $Y = \{Y_P, Y_T, Y_C, Y_R\}$ відповідно до умов $V = \{V_P, V_T, V_C, V_R\}$, де $Y_P = Y_{Pc} \cup Y_{Pq}$ – множина інформаційних портретів комерційного контенту Y_{Pc} та постійних користувачів Y_{Pq} , Y_T – множина тематичних сюжетів комерційного контенту; Y_C – множина таблиць взаємозв'язку комерційного контенту, $Y_R = Y_{Rc} \cup Y_{Rm}$ – множина рейтингів комерційного контенту Y_{Rc} та модераторів Y_{Rm} , $V_P = V_{Pc} \cup V_{Pq}$ – множина умов формування інформаційних портретів; V_T – множина умов виявлення тематичних сюжетів; V_C – множина умов побудови таблиць взаємозв'язку комерційного контенту; V_R – множина параметрів розрахунку рейтингів комерційного контенту. Формування портретів комерційного контенту подано так: $Y_{Pc} = \gamma_1(V_{Pc}, C, H, Q, T)$, а портретів постійних користувачів – $Y_{Pq} = \gamma_2(V_{Pq}, Q, H, Z, T)$. Формування тематичних сюжетів комерційного контенту подано так: $Y_T = Y_{Tc} \cup Y_{Tq}$, де $Y_{Tc} = \gamma_3(C, H, X, V_T, T)$ – виявлення тематичних сюжетів у множині нового комерційного контенту та $Y_{Tq} = \gamma_4(C, H, Q, V_T, T)$ – виявлення тематичних сюжетів комерційного контенту за множиною запитів користувачів. Формування таблиць взаємозв'язку комерційного контенту – $Y_C = \gamma_5(C, V_C, T)$. Формування рейтингів комерційного контенту – $Y_{Rc} = \gamma_6(C, Q, H, Y_C, V_{Rc}, T, \theta, \xi)$, а множину рейтингів модераторів Y_{Rm} подано так: $Y_{Rm} = \gamma_7(C, Q, H, Y_C, V_{Rm}, T)$, де $V_R = V_{Rc} \cup V_{Rm}$ – множина параметрів для розрахунку рейтингів комерційного контенту, $\theta(Q^+, Q^0, Q^-, T, H)$ – критерій тональності комерційного контенту, $\xi(Q, T)$ – оператор визначення фільтрування коментарів. Колекцію вихідних статистичних даних Y подано так:

$$Y = \{Y_P, Y_T, Y_C, Y_R\} = \gamma(V_P, V_T, V_C, V_R, C, Q, H, Z, T). \quad (14)$$

Колекція Y описує функціонування СЕКК з основними процесами опрацювання Web-ресурсів такими, як формування, управління та супровід контенту.

Аналізують статистику роботи СЕКК відповідно до аналізу реакцій на цю систему постійного/потенційного користувача (відвідування, запити, пошук за ключовими словами тощо). Це сприяє ефективному аналізу реакції цільової та потенційної аудиторій. Для детального аналізу функціонування СЕКК типу Інтернет-газета та Інтернет-журнал розроблено та впроваджено вісім різних систем відповідно з підтриманням для кожної з них різною кількістю етапів життєвого циклу контенту. Тобто для різних реалізованих систем розроблені не всі компоненти або взагалі не розроблені підсистеми опрацювання Web-ресурсів такі, як формування, управління та супровід контенту.

1. Берко А., Висоцька В., Пасічник В. Системи електронної контент-комерції. – Л.: НУЛП, 2009. – 612 с.
2. Основы моделирования и оценки электронных информационных потоков / Д. Ландэ, В. Фурашев, С. Брайчевский, О. Григорьев. – К.: Інжиніринг, 2006. – 348 с.
3. Ландэ Д. Основы интеграции информационных потоков: монографія. – К.: Інжиніринг, 2006. – 240 с.

Одержано 24.02.2015