

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 004.738.5

Висоцька В. А.¹, Чирун Л. В.²

¹ Асистент, Національний університет «Львівська політехніка», Україна, E-mail: victana@bk.ru
² Канд. техн. наук, доцент, Національний університет «Львівська політехніка», Україна

КОНЦЕПТУАЛЬНА МОДЕЛЬ СИСТЕМ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

У статті запропоновано формальну модель системи електронної контент-комерції та формальні моделі опрацювання інформаційних ресурсів. Описана розроблена загальна архітектура систем електронної контент-комерції для полегшення етапів реалізації життєвого циклу комерційного контенту. Запропоновані загальні принципи проектування архітектури систем електронної контент-комерції дають можливість реалізувати процес опрацювання інформаційних ресурсів для скорочення циклу виробництва, економії часу та розширення можливостей ведення електронної комерції.

Ключові слова: інформаційний ресурс, комерційний контент, контент-аналіз, контент-моніторинг, контентний пошук, система електронної контент-комерції.

ВСТУП

Процес проектування та моделювання систем електронної контент-комерції (СЕКК) є інтерактивним і протікає від аналізу до створення прототипу і пробних випробувань [1–3], починаючи з формування специфікацій, верстки, формування контенту та його подальше розміщення згідно структури інформаційного ресурсу [1–2]. Необхідно сконцентрувати рішення на бізнес-цілях і потребах кінцевих користувачів [3–4]. До визначення функціональних вимог і початку процесу розроблення аналізуються потреби кінцевих користувачів за допомогою листів опитування, альтернатив проектування і прототипів різного ступеня готовності, збирається цінна інформація, одночасно викликаючи у користувачів відчуття участі в процесі проектування та завойовується їх довіра [2]. Оскільки інформаційний ресурс – це нейтральне місце взаємодії самих різних користувачів, яскраво виражений візуальний зв'язок із якою-небудь компанією або оточенням викликає небажану реакцію [2, 4]. Є декілька вимог до середовища розроблення – це можливість вносити зміни до коду і тестувати внесені зміни автономно [1–4]. Створивши централізоване середовище розроблен-

ня і тестування, оптимізують роботу із кодом та іншими членами групи – час повинен витратитися на написання і тестування коду, а не на управління файлами та іншим ресурсами системи [1–4]. Вибір моделі CMS впливає на необхідність використання інших інструментальних засобів, наприклад, у випадку з Joomla! це означає використання PHP, HTML і Cascading Style Sheets (CSS) для розроблення сторінок, а також MYSQL для бази даних [2].

АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ ТА ПОСТАНОВКА ЗАДАЧІ

Інформаційний ресурс систем електронної контент-комерції – це множина даних з набором властивостей (табл. 1), які є об'єктом дій технології перетворення їх в комерційний контент [1–2]. Результат застосування однієї технології може бути інформаційним ресурсом іншої [2, 4]. Контент в галузі інформаційних технологій є формалізованими відомостями і знаннями, розміщеними у середовищі системи і, на відміну від даних, без детальної специфікації їх властивостей, способів формалізації і впорядкування [2, 5]. Перетворення різнорідних за природою, змістом та походженням даних в узгоджений централізований інформаційний ресурс є однією з важливих

проблем побудови та функціонування систем електронної контент-комерції [2]. При цьому важливими завданнями є забезпечення інформаційних потреб проблемно-орієнтованих елементів системи, підтримання доступу до даних різних категорій користувачів, дотримання правил цілісності та несуперечності даних, мінімізація та контроль надлишку даних, здатність до розвитку та зміни внутрішньої організації інформаційного ресурсу, дотримання вимог якості та ефективності даних [1–2, 4].

Важливим є забезпечення інваріантності середовища СЕКК до модифікації інформаційних ресурсів у таких змінах: способів подання, форматів, та внутрішньої організації контенту; середовища зберігання контенту, фізичних одиниць зберігання, технічних засобів; вимог користувачів, появи нових вимог та категорій користувачів; порядку розподілу контенту та способів доступу користувачів [2, 5–11].

Виникає проблема створення єдиного концептуального опису всього інформаційного ресурсу для стабільного підтримання зовнішніх/внутрішніх позначень контенту відповідно до їх завдань, вимог та змін [2, 5–6, 10]. Необхідно класифікувати інформаційні ресурси для дослідження їх природних, технологічних та споживчих якостей з метою виявлення характерних та специфічних властивостей, а також закономірностей та особливостей їх

формування та застосування [4]. За основу взято основні властивості контенту в СЕКК як синтаксис, структура та семантика та обрано основні фактори класифікації як способи подання контенту, структурування та призначення ресурсу, доступу до ресурсу [2, 5, 10]. Загальні принципи формування інформаційних ресурсів СЕКК (рис. 1) визначають порядок і способи відбору інформації із первинних джерел, її фіксації, фільтрування, перетворення до визначеного формату для формування контенту і розміщення в базі даних [2, 5].

Нехай існує деяка попередньо визначена множина N первинних джерел контенту з фіксованим або змінним складом. Кожне джерело інформації x_i , де $i = \overline{1, N}$, формує деяку множину значень, що містять відомості/знання/факти з предметної області СЕКК [2]. Результатом звернення певних технологічних засобів СЕКК до джерела x_i є генерування множини значень $X(x_i)$, яка сприймається і подається у визначеній формі. В процесі відбору і фіксації генерованих значень, згідно технологічних особливостей системи, згенеровану кожним джерелом інформації множину значень перетворюють на вхідний набір контенту, визначеного формату – X_i , де $i = \overline{1, N}$. Кожен набір контенту подається у вигляді структурованих, слабкоструктурованих даних або даних без визначеного опису структури.

Таблиця 1. Основні властивості інформаційних ресурсів в системах електронної контент-комерції

Назва	Властивість
Неоднорідність	Наявність складових різного походження, змісту і формату подання
Узгодженість	Відсутність суперечливих або протилежних значень контенту
Доступність формату	Доступність для всіх користувачів на основі стандартизованих методів, засобів та інтерфейсів
Відкритість	Здатність до взаємодії, обміну значеннями та спільного використання з зовнішніми ресурсами
Динамічність	Швидка актуалізація, відповідно до умов системи чи зовнішнього середовища
Масштабованість	Можливість зміни логічного/фізичного обсягу контенту (величин/понять та їх позначень)
Контрольованість	Ідентифікація зміни/використання контенту та його впливу на процеси системи

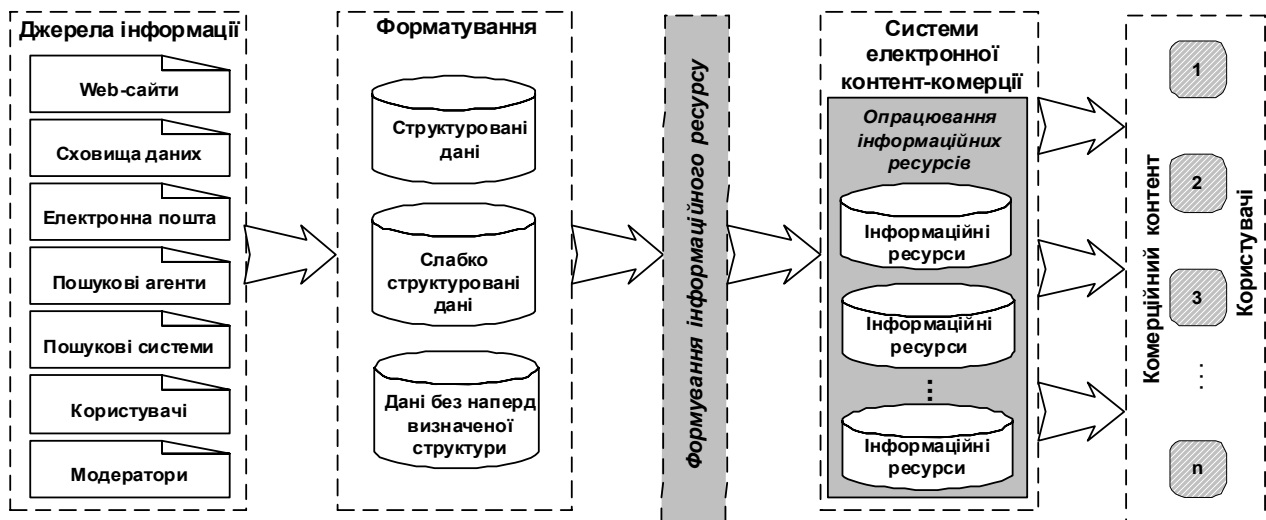


Рис. 1. Порядок формування і використання інформаційних ресурсів в системах електронної контент-комерції

Структурування контенту передбачає формування для кожного набору опису його складу, способів поєднання елементів та їх впорядкування – множини умов U_i , де $i = \overline{1, N}$. Кожен набір контенту є поєднанням множини значень у заданому форматі і множини умов $\langle X_i, U_i \rangle$, у випадку формування вхідного набору контенту без опису структури $U_i = \emptyset$. Отриманий контент перед збереженням в базі даних проходить процедуру верифікації та валідації для підтвердження його формальної і змістовної коректності/релевантності щодо вимог системи [2]. При невідповідності зазначеним критеріям частина контенту вилучається з подальшого застосування. Відфільтрований контент форматується та зберігають, після чого відповідні відомості і знання стають доступними для застосування користувачами СЕКК. Формально таку послідовність технологічних процесів подано як ланцюжок

$$x_i \rightarrow X(x_i) \rightarrow X_i \rightarrow \langle X_i, U_i \rangle \rightarrow Verification(\langle X_i, U_i \rangle) \rightarrow Qualification(\langle X_i, U_i \rangle) \rightarrow Conversion(\langle X_i, U_i \rangle) \rightarrow Downloading(\langle X_i, U_i \rangle) \rightarrow \langle X, U \rangle,$$

де $i = \overline{1, N}$, де N – кількість джерел контенту, x_i – i -те джерело, $X(x_i)$ – множина контенту x i -того джерела, X_i – набір контенту як результат відбору з i -го джерела, $\langle X_i, U_i \rangle$ – набір контенту із множиною умов, $Verification(\langle X_i, U_i \rangle)$ – оператор верифікації контенту, $Qualification(\langle X_i, U_i \rangle)$ – оператор кваліфікації контенту, $Conversion(\langle X_i, U_i \rangle)$ – оператор перетворення контенту, $Downloading(\langle X_i, U_i \rangle)$ – оператор завантаження контенту, $\langle X, U \rangle$ – загальний інформаційний ресурс в СЕКК.

ОСНОВНИЙ ВИКЛАД

Основними підсистемами опрацювання інформаційних ресурсів в СЕКК є формування, управління та суп-

ровід контенту, які мають наступні відношення:

$$\begin{aligned} & \text{контент} \rightarrow \text{формування контенту} \rightarrow \text{база даних} \\ & \rightarrow \text{управління контентом} \rightarrow \text{інформаційний ресурс,} \\ & \text{запит користувача} \rightarrow \text{управління контентом} \rightarrow \\ & \text{інформаційний ресурс} \rightarrow \text{супровід контенту} \rightarrow \text{база} \\ & \text{даних.} \end{aligned}$$

Відповідно, підсистеми мають наступний зв'язок: $\text{формування контенту} \rightarrow \text{управління контентом} \rightarrow \text{супровід контенту}$.

Формальна модель системи електронної контент комерції подано як

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

де величини моделі описують процес функціонування системи [3] і утворюють підмножини, подані в табл. 2.

Процес формування контенту описано функцією вигляду

$$c_r = Formation(u_f, x_i, t_p), \quad (2)$$

де u_f – множина умов формування контенту, тобто $u_f = \{u_1(x_i), \dots, u_{n_U}(x_i)\}$.

Контент подано наступним чином:

$$c_r = \left\{ \bigcup_f u_f \mid (x_i \in X) \wedge (\exists u_f \in U), U = U_{x_i} \vee U_{x_i}^-, i = \overline{1, m}, f = \overline{1, n} \right\}. \quad (3)$$

Етап управління контентом описано функцією вигляду

$$z_w = Management(q_d, c_r, h_k, t_p), \quad (4)$$

де $Q = \{q_1, q_2, \dots, q_{n_Q}\}$ – множина запитів користувачів,

h_k – множина умов управління контентом, тобто

Таблиця 2. Величини моделі системи електронної контент-комерції

Назва	Позначення	Діапазон	Множина
Множина вхідних даних з різних джерел	$x_i \in X$	$i = \overline{1, n_X}$	$X = \{x_1, x_2, \dots, x_{n_X}\}$
Множина запитів користувачів	$q_d \in Q$	$d = \overline{1, n_Q}$	$Q = \{q_1, q_2, \dots, q_{n_Q}\}$
Оператор формування контенту	<i>Formation</i>	–	–
Множина внутрішніх параметрів системи	$h_k \in H$	$k = \overline{1, n_H}$	$H = \{h_1, h_2, \dots, h_{n_H}\}$
Множина комерційного контенту	$c_r \in C$	$r = \overline{1, n_C}$	$C = \{c_1, c_2, \dots, c_{n_C}\}$
Множина параметрів впливу середовища	$v_l \in V$	$l = \overline{1, n_V}$	$V = \{v_1, v_2, \dots, v_{n_V}\}$
Оператор управління контентом	<i>Management</i>	–	–
Оператор супроводу контенту	<i>Support</i>	–	–
Множина сторінок інформаційного ресурсу	$z_w \in Z$	$z = \overline{1, n_Z}$	$Z = \{z_1, z_2, \dots, z_{n_Z}\}$
Час транзакції опрацювання ресурсів	$t_p \in T$	$p = \overline{1, n_T}$	$T = \{t_1, t_2, \dots, t_{n_T}\}$
Множина статистичних даних роботи системи	$y_j \in Y$	$j = \overline{1, n_Y}$	$Y = \{y_1, y_2, \dots, y_{n_Y}\}$

$h_k = \{h_1(x_i, q_d), \dots, h_{n_H}(x_i, q_d)\}$. Управління контентом подано як

$$c_j = \left\{ \bigcup_{m_k} (q_i \in Q) \wedge (\exists u_{m_k} \in U_M), U_M = U_{M_q} \vee U_{T_q}, i = \overline{1, m}, k = \overline{1, n} \right\}. \quad (5)$$

1 МОДЕЛЬ ПРОЦЕСУ ФОРМУВАННЯ КОМЕРЦІЙНОГО КОНТЕНТУ

Підсистема формування контенту забезпечує збирання інформації з різних інформаційних ресурсів та її форматування, виявлення ключових слів, понять та дублювання, формування дайджесту, рубрикацію та вибіркове поширення контенту.

Основними етапами процесу формування комерційного контенту в СЕКК є форматування, рубрикація та поширення контенту, які мають наступні відношення:

джерело контенту → *збирання/створення контенту* → *база даних* → *форматування контенту* → *база даних* → *виявлення ключових слів та понять* → *база даних* → *рубрикація контенту* → *база даних* → *виявлення дублювання контенту* → *база даних* → *формування дайджесту контенту* → *база даних* → *вибіркове поширення контенту* → *модератор*.

Формальна модель процесу формування контенту в СЕКК подано як

$$Formation = \left\langle X, Gathering, Formatting, KeyWords, Backup, Categorization, BuDigest, Dissemination, T, C \right\rangle, \quad (6)$$

де величини моделі утворюють підмножини та оператори, подані в табл. 3.

Процес формування комерційного контенту описано функцією вигляду

$$c_r = Formation(u_f, x_i, t_p), \quad (7)$$

де u_f – множина умов формування контенту, тобто

$$u_f = \{u_1(x_i), \dots, u_{n_U}(x_i)\}.$$

Комерційний контент подано наступним чином:

$$c_r = \left\{ \bigcup_f u_f \mid (x_i \in X) \wedge (\exists u_f \in U), U = U_{x_i} \vee U_{x_i}^-, i = \overline{1, m}, f = \overline{1, n} \right\}, \quad (8)$$

проходячи наступні етапи перетворення з множини даних у множину релевантного форматowanego, класифікованого та валідованого контенту: $x_i \in X \rightarrow Gathering(u_f, x_i, t_p) \rightarrow Backup(c_r, u_b, t_p) \rightarrow Formatting(c_r, t_p) \rightarrow KeyWords(c_r, t_p) \rightarrow Categorization(c_r, t_p) \rightarrow BuDigest(c_r, t_p) \rightarrow Dissemination(c_r, t_p) \rightarrow c_r \in C$.

Оптимальне рішення, здатне допомогти орієнтуватися в динамічній вхідній інформації з різних джерел, надають процеси синдикації даних (збирання інформації з джерел і наступне поширення її фрагментів відповідно до потреб користувачів):

$$C = Gathering(X, U_G, T), \quad (9)$$

де X – множина контенту з різних джерел даних, U_G – множина умов збирання даних з різноманітних джерел, $Gathering$ – оператор збирання/створення контенту, T – час збирання/створення контенту.

Функція виявлення дублювання змісту контенту описано функцією вигляду

$$C = Backup(Gathering(X, U_G, T), U_B), \quad (10)$$

де X – множина контенту з різних джерел даних, U_B – множина умов виявлення дублювання змісту контенту, $Backup$ – оператор виявлення дублювання змісту контенту, C – множина контенту. Виявлення дубльованих за змістом контенту у СЕКК виконується на основі лінгвостатистичних методів виявлення загальних термів, ланцюжки яких утворюють словесні сигнатури контенту.

Таблиця 3. Величини моделі формування комерційного контенту

Назва	Позначення	Діапазон	Множина
Множина вхідних даних з різних інформаційних ресурсів або від модераторів	$x_i \in X$	$i = \overline{1, n_X}$	$X = \{x_1, x_2, \dots, x_{n_X}\}$
Оператор збирання/створення контенту з джерел	<i>Gathering</i>	–	–
Оператор форматування контенту	<i>Formatting</i>	–	–
Оператор виявлення ключових слів контенту	<i>KeyWords</i>	–	–
Оператор автоматичної рубрикації контенту	<i>Categorization</i>	–	–
Оператор виявлення дублювання змісту контенту	<i>Backup</i>	–	–
Оператор формування дайджестів контенту	<i>BuDigest</i>	–	–
Оператор вибіркового поширення контенту	<i>Dissemination</i>	–	–
Час транзакції формування контенту	$t_p \in T$	$p = \overline{1, n_T}$	$T = \{t_1, t_2, \dots, t_{n_T}\}$
Множина комерційного контенту	$c_r \in C$	$r = \overline{1, n_C}$	$C = \{c_1, c_2, \dots, c_{n_C}\}$

Технологія синдикації контенту містить процес навчання програм збирання даних структурним особливостям окремих джерел (з інформаційних ресурсів, від модераторів, користувачів, відвідувачів, журналістів, редакторів), безпосереднє сканування контенту та приведення до загального формату в XML:

$$C = \text{Formatting}(\text{Backup}(\text{Gathering}(X, U_G, T), U_B), U_{FR}), (11)$$

де Formatting – оператор форматування контенту; U_G – множина умов збирання даних з різноманітних джерел; Gathering – оператор збирання/створення контенту; U_{FR} – множина умов форматування інформації; T – час збирання контенту.

Опрацювання множини контенту C для виявлення значущих ключових слів побудоване на принципі знаходження ключових слів за змістом (термами), базується на законі Зіпфа і зводиться до вибору слів із середньою частотою появи (найбільш вживанні слова ігноруються за допомогою стоп-словника, а рідкісні слова з текстів повідомлень не враховуються). Виявлення ключових слів та понять визначається оператором $\text{KeyWords}(C)$ та описано оператором вигляду:

$$C = \text{KeyWords}(\text{Formatting}(\text{Backup}(\text{Gathering}(X, U_G, T), U_B), U_{FR}), U_K), (12)$$

де Formatting – оператор форматування контенту, U_G – множина умов збирання даних з різноманітних джерел, Gathering – оператор збирання/створення контенту, U_{FR} – множина умов форматування інформації, T – час збирання контенту, U_K – множина умов виявлення ключових слів та понять.

Засоби класифікації та розподілу контенту являють собою інформаційно-пошукову систему вибіркового поширення контенту (контентний роутер). Контент аналізується на відповідність тематичним запитам оператором рубрикації

$$C_{Cl} = \text{Categorization}(\text{KeyWords}(C, U_K), U_{Cl}), (13)$$

де $\text{KeyWords}(C, U_K)$ – оператор виявлення ключових слів і понять, Categorization – оператор рубрикації контенту згідно виявлених ключових слів та понять, U_K – множина умов виявлення ключових слів і понять, U_{Cl} – множина умов автоматичної рубрикації, C_{Cl} – множина рубрикованого релевантного контенту.

Множина дайджестів C_D формується за допомогою наступної залежності:

$$C_D = \text{BuDigest}(C_{Cl}, U_D), (14)$$

де BuDigest – оператор формування дайджестів, U_D – множина умов формування дайджестів, C_{Cl} – множина рубрикованого релевантного контенту, тобто

$$C_D = \text{BuDigest}(\text{Categorization}(\text{KeyWords}(C, U_K), U_{Cl}), U_D). (15)$$

Релевантний контент розсилається користувачам, а також завантажується в тематичні бази даних. Вибіркове поширення контенту описується так

$$C_{Ds} = \text{Dissemination}(C_D, U_{Ds}), (16)$$

де C_{Ds} – множина вибірково поширеного контенту, U_{Ds} – множина умов вибіркового поширення контенту, Dissemination – оператор вибіркового поширення контенту.

Вибіркове поширення списку контенту $C_{Dm} = \text{sup}(C_{Ds})$ залежить від рівня попиту на цей контент. Асоціативне правило формування списку контенту складається зі первісного списку контенту C_{Ds} і списку контенту, обраного із первісного – похідний список C_{Dm} , тобто $C_{Ds} \rightarrow C_{Dm}$. Формування асоціативного правила – це формування контенту, утвореного внаслідок об'єднання первісного і похідного списків. Асоціативне правило появи контенту із первісного списку C_{Ds} разом із контентом із списку C_{Dm} в базі даних є наступним

$$P_{Dm} = \text{Demand}(C_{Ds}, C_{Dm}) = \frac{\text{sup}(C_{Dm} \cup C_{Ds})}{\text{sup}(C_{Ds})}. (17)$$

Показники прибутковості та зростання розраховують відповідно як

$$I_p = \text{Profitability}(C_{Dm}, P_{Dm}) = P_{Dm} \frac{\text{sup}(\emptyset)}{\text{sup}(C_{Dm})}, (18)$$

$$I_g = \text{Growth}(C_{Dm}, P_{Dm}) = \frac{\text{sup}(\emptyset) - \text{sup}(C_{Dm})}{\text{sup}(\emptyset)[1 - P_{Dm}]}. (19)$$

Похідний список обраних правил визначає модератор. Часто вживаним списком контенту є список, формування якого перевищує мінімальний рівень. Навіть при встановленні цієї умови списків часто вживаного контенту в запитах буде багато. Результат обмежується при вибірці

$$I_r = \text{Restriction}(C_{Ds}, C_{Dm}, P_{Dm}) = \min(\forall C'_{Ds} \subset C_{Ds}, P_{Dm} - P'_{Dm}), (20)$$

або

$$I_r = \min(\forall C'_{Ds} \subset C_{Ds}, \text{Demand}(C_{Ds}, C_{Dm}) - \text{Demand}(C'_{Ds}, C_{Dm})). (21)$$

Модель формування комерційного контенту реалізується у вигляді комплексів контент-моніторингу зі збирання контенту з різних джерел даних, і забезпечують створення бази даних контенту відповідно до інформаційних потреб користувачів. В результаті збирання і первинного опрацювання контент приводиться до єдиного формату, класифікується відповідно до визначеного рубризатора та йому приписують дескриптори з ключови-

ми словами. Це полегшує реалізацію процесу управління комерційним контентом.

2 МОДЕЛІ ПРОЦЕСУ УПРАВЛІННЯ КОМЕРЦІЙНИМ КОНТЕНТОМ

Завдання підсистеми управління контентом: формування, ротація баз даних і забезпечення доступу до неї; формування оперативних і ретроспективних баз даних; персоналізація роботи користувачів, збереження персональних запитів користувачів і джерел, ведення статистики роботи; забезпечення пошуку в базах даних; генерація вихідних форм; інформаційна взаємодія з іншими базами даних; формування інформаційного ресурсу.

Процес управління контентом описано функцією вигляду

$$z_w = Management(q_d, c_r, h_k, t_p), \quad (22)$$

де $Z = \{z_1, z_2, \dots, z_{n_z}\}$ – множина інформаційних ресурсів, $Management$ – оператор управління комерційним контентом, $C = \{c_1, c_2, \dots, c_{n_c}\}$ – множина комерційного контенту, $Q = \{q_1, q_2, \dots, q_{n_q}\}$ – множина запитів користувачів, $T = \{t_1, t_2, \dots, t_{n_t}\}$ – час транзакцій управління контентом, множина комерційного контенту, h_k – множина умов управління контентом, тобто $h_k = \{h_1(x_i, q_d), \dots, h_{n_H}(x_i, q_d)\}$.

На рис. 2 подана класифікація моделей управління комерційним контентом.

Управління контентом подано як

$$c_j = \left\{ \bigcup_{m_k} \left((q_i \in Q) \wedge (\exists u_{m_k} \in U_M) \right), U_M = U_{M_q} \vee U_{Fq}, i = \overline{1, m}, k = \overline{1, n} \right\}. \quad (23)$$

1. **Генерація сторінок за запитом** реалізовується відношенням

Модератор → *редагування контенту* → *база даних* → *подання контенту* → *інформаційний ресурс*.

Формальна модель етапу генерації сторінок за запитом подано як

$$Management_Q = \langle X, H, C, Q, Presentation, Edit, T, Z \rangle, \quad (24)$$

де $X = \{x_1, x_2, \dots, x_{n_x}\}$ – множина вхідних даних; $C = \{c_1, c_2, \dots, c_{n_c}\}$ – множина контенту;



Рис. 2. Особливості підсистеми управління комерційним контентом

$Z = \{z_1, z_2, \dots, z_{n_z}\}$ – множина сформованих сторінок; $Q = \{q_1, q_2, \dots, q_{n_q}\}$ – множина запитів; $Presentation$ – оператор формування та подання сторінки; $Edit$ – оператор редагування та модифікації контенту; $T = \{t_1, t_2, \dots, t_{n_t}\}$ – час транзакцій управління контентом; $h_k = \{h_1(x_i, q_d), \dots, h_{n_H}(x_i, q_d)\}$ – множина умов управління контентом.

Етап редагування та модифікації контенту подано функцією вигляду

$$c_j(x_i, t_r) = Edit(c_j, x_i, t_r). \quad (25)$$

Етап формування множини сторінок описано функцією вигляду

$$Z(t_r + \Delta t) = Presentation(q_i, C, Weight(C), t_r, \Delta t), \quad (26)$$

де $Weight(C)$ – загальна вага контентного блоку, тобто

$$z_i = \left\{ \bigcup c_j \left((\forall c_j \in C_q) \wedge (\exists q_i \in C_q) \right), C = C_q \vee C_{\bar{q}}, j = \overline{1, m}, i = \overline{1, n} \right\}. \quad (27)$$

Вагу блоку визначають як суму коефіцієнтів, які характеризують контент:

$$\Delta c = \|C\| = Weight(C, Location, KeyWords, Static, Addterm, Preferences), \quad (28)$$

де $Location(c_j)$ – коефіцієнт розташування блоку у контенті, $KeyWords(c_j)$ – коефіцієнт ключових слів в блоці, $Static(c_j)$ – коефіцієнт статистичної важливості термів, $Addterm(c_j)$ – коефіцієнт наявності додаткових термів, $Preferences(c_j)$ – коефіцієнт наявності та обсягу термів із запиту користувача [7–8].

2. **Генерація сторінок при редагуванні** полягає у створенні набору статичних сторінок при внесенні змін до контенту інформаційного ресурсу, тобто

Модератор → *редагування контенту (інформаційного ресурсу)* → *база даних* → *інформаційний ресурс*.

В моделі не враховується інтерактивність між відвідувачем і ресурсом.

Формальна модель етапу генерації сторінок при редагуванні має вигляд

$$Management_E = \langle C, H, Edit, T, Z \rangle, \quad (29)$$

де $C = \{c_1, c_2, \dots, c_{n_c}\}$ – множина контенту;

$h_k = \{h_1, h_2, \dots, h_{n_H}\}$ – множина умов управління контентом, $Z = \{z_1, z_2, \dots, z_{n_z}\}$ – множина статичних сторінок; $Edit$ – оператор редагування/модифікації контенту інформаційного ресурсу.

Етап формування сторінок описано оператором вигляду

$$Z(t) = Edit(C, Weight, H, t). \quad (30)$$

3. **Змішаний тип** поєднує в собі переваги перших двох типів:

Модератор → *редагування контенту* → *база даних* → *збирання блоків* → *кеш* → *подання контенту* → *інформаційний ресурс*.

Формальна модель етапу управління контентом подано як

$$Management_M = \langle X, C, Q, H, Presentation, Edit, Caching, T, Z \rangle, \quad (31)$$

де $X = \{x_1, x_2, \dots, x_{n_X}\}$ – множина вхідних даних;

$C = \{c_1, c_2, \dots, c_{n_C}\}$ – множина контенту;

$Z = \{z_1, z_2, \dots, z_{n_Z}\}$ – множина сформованих сторінок;

$Q = \{q_1, q_2, \dots, q_{n_Q}\}$ – множина запитів;

$h_k = \{h_1(x_i, q_d), \dots, h_{n_H}(x_i, q_d)\}$ – множина умов управління контентом, *Presentation* – оператор формування та подання сторінки; *Edit* – оператор редагування та модифікації контенту; *Caching* – оператор формування кешу або інформаційних блоків, де кеш

$$Cache = Caching(Z, Weight, t, \Delta t), \quad (32)$$

тобто

$$Cache = \left\{ \bigcup_i z_i \mid z_i \in Z, t = t + \Delta t, i = \overline{1, n} \right\}. \quad (33)$$

Підсистема управління комерційним контентом реалізована шляхом кешування (підсистема подання генерує сторінку один раз, надалі вона в декілька разів швидше завантажується з кешу, який оновлюється автоматично по закінченню деякого терміну часу або при внесенні змін до певних розділів інформаційного ресурсу, або вручну по команді адміністратора) або формуванням інформаційних блоків (збереження блоків на етапі редагування інформаційного ресурсу і збирання сторінки з цих блоків при запиті відповідної сторінки користувачем).

3 МОДЕЛЬ ПРОЦЕСУ СУПРОВОДУ КОМЕРЦІЙНОГО КОНТЕНТУ

Підсистема супроводу контенту забезпечує формування інформаційних портретів; виявлення тематичних сюжетів контенту; побудову таблиць взаємозв'язку контенту; розрахунок рейтингів контенту, виявлення нових подій в контентних потоках, їхнє відстеження й кластеризацію.

Підсистема супроводу контенту *Support* описана як

$$y(t_p + \Delta t) = Support(v_l, h_k, c_r, z_w, t_p, \Delta t), \quad (34)$$

де v_l – множина умов супроводу контенту та зовнішніх впливів середовища на систему, тобто $v_l = (v_1(q_i, t_p), \dots, v_{n_V}(q_i, t_p))$. Вихідна інформація реалізована як

$$y_j = \left\{ \bigcup_l v_l \mid (q_d \in X) \wedge (\exists v_l \in V), V = V_{q_d} \vee V_{q_d}^-, d = \overline{1, m}, l = \overline{1, n} \right\}. \quad (35)$$

Формальна модель супроводу комерційного контенту має вигляд

$$Support = \left\langle X, Q, C, H, V, T, BuInfPort, IdThemTop, ConCorrTablConc, CalRankConc, Z, Y \right\rangle, \quad (36)$$

де величини моделі подані в табл. 4.

Множина комерційного контенту C подано як

$$C = \langle C_P, C_T, C_C, C_R \rangle, \quad (37)$$

де C_P – підмножина інформаційних портретів, C_T – підмножина тематичних сюжетів комерційного контенту, C_C – підмножина таблиць взаємозв'язку комерційного контенту, C_R – підмножина рейтингів комерційного контенту.

Таблиця 4. Величини моделі супроводу комерційного контенту

Назва	Позначення	Діапазон	Множина
Вхідні дані від модераторів/користувачів	$x_i \in X$	$i = \overline{1, n_X}$	$X = \{x_1, x_2, \dots, x_{n_X}\}$
Множина запитів користувачів	$q_d \in Q$	$d = \overline{1, n_Q}$	$Q = \{q_1, q_2, \dots, q_{n_Q}\}$
Множина комерційного контенту	$c_r \in C$	$r = \overline{1, n_C}$	$C = \{c_1, c_2, \dots, c_{n_C}\}$
Оператор формування портретів	<i>BuInfPort</i>	–	–
Оператор виявлення тематичних сюжетів	<i>IdThemTop</i>	–	–
Оператор побудови взаємозв'язку контенту	<i>ConCorrTablConc</i>	–	–
Оператор розрахунку рейтингів контенту	<i>CalRankConc</i>	–	–
Множина внутрішніх параметрів системи	$h_k \in H$	$k = \overline{1, n_H}$	$H = \{h_1, h_2, \dots, h_{n_H}\}$
Множина параметрів впливу середовища	$v_l \in V$	$l = \overline{1, n_V}$	$V = \{v_1, v_2, \dots, v_{n_V}\}$
Множина сторінок інформаційного ресурсу	$z_w \in Z$	$z = \overline{1, n_Z}$	$Z = \{z_1, z_2, \dots, z_{n_Z}\}$
Час транзакції опрацювання ресурсів	$t_p \in T$	$p = \overline{1, n_T}$	$T = \{t_1, t_2, \dots, t_{n_T}\}$
Множина статистичних даних роботи системи	$y_j \in Y$	$j = \overline{1, n_Y}$	$Y = \{y_1, y_2, \dots, y_{n_Y}\}$

Множина інформаційних портретів C_P подано як

$$C_P = \text{BuInfPortr}(X, U_P, C_P), \quad (38)$$

де U_P – множина умов формування інформаційних портретів, BuInfPortr – оператор формування інформаційних портретів.

Множина тематичних сюжетів комерційного контенту C_T подано як

$$C_T = \text{IdThemTop}(X, C_T, U_T), \quad (39)$$

де U_T – множина умов виявлення тематичних сюжетів комерційного контенту, IdThemTop – оператор визначення тематичних сюжетів комерційного контенту C_T .

Множина таблиць взаємозв'язку комерційного контенту C_C подано як

$$C_C = \text{ConCorrTablConc}(C, U_C), \quad (40)$$

де U_C – множина умов побудови таблиць взаємозв'язку комерційного контенту, ConCorrTablConc – оператор побудови таблиць взаємозв'язку комерційного контенту.

Множина рейтингів комерційного контенту C_R подано як

$$C_R = \text{CalRankConc}(C_C, \text{Tonality}, U_R, \text{Spam}), \quad (41)$$

де U_R – множина параметрів розрахунку рейтингів комерційного контенту, Tonality – критерій тональності контенту, Spam – функція визначення спаму, CalRankConc – оператор визначення рейтингів комерційного контенту C_R .

Множина вихідних статистичних даних Y подано як

$$Y = \text{Support}(C) = \text{Support}(C_P, C_T, C_C, C_R), \quad (42)$$

де C_P – підмножина інформаційних портретів, C_T – підмножина тематичних сюжетів комерційного контенту, C_C – підмножина таблиць взаємозв'язку комерційного контенту, C_R – підмножина рейтингів комерційного контенту, Support – оператор супроводу комерційного контенту.

Модель СЕКК є універсальною та простою для опису процесу функціонування типових систем. Моделі процесів опрацювання інформаційних ресурсів дають змогу уніфікувати основні функції системи як формування, управління та реалізація контенту. На основі розроблених моделей, зокрема, розв'язують задачі побудови типової архітектури систем електронної контент-комерції, формують специфікацію функціональних та нефункціо-

нальних вимог щодо СЕКК та їх складових таких як засоби взаємодії з кінцевим користувачем, реалізують процедури аналізу контенту та збору і опрацювання статистичних даних відвідування інформаційних ресурсів. Моделі процесів формування, управління та реалізація контенту дають змогу спростити процеси проектування СЕКК шляхом уніфікації та стандартизації методів опрацювання інформаційних ресурсів.

ВИСНОВКИ

У статті вибрано підходи та моделі послідовності опрацювання інформаційних ресурсів в системах електронної контент-комерції для проектування та створення таких систем, що дало змогу виділити основні закономірності переходу від процесів формування комерційного контенту до його супроводу. Побудовано формальну модель систем електронної комерції для реалізації етапів життєвого циклу комерційного контенту, що дало змогу визначити основні компоненти системи електронної контент-комерції та їх зв'язки між собою.

Розроблено формальні моделі опрацювання інформаційних ресурсів в системах електронної контент-комерції для реалізації формування, управління та супроводу комерційного контенту, що дало змогу створити узагальнену архітектуру системи електронної контент-комерції. Розроблено узагальнену архітектуру системи електронної контент-комерції для спрощення етапів розроблення таких систем, що дало змогу реалізувати процеси формування, управління та супроводу комерційного контенту.

СПИСОК ЛІТЕРАТУРИ

1. *Берега, А.* Електронна комерція / А. Берега, І. Козак, Ф. Левченко. – К. : КНЕУ, 2002. – 326 с.
2. *Берко, А.* Системи електронної контент-комерції / А. Берко, В. Висоцька, В. Пасічник. – Л. : НУЛП, 2009. – 612 с.
3. *Советов, Б.* Моделирование систем / Б. Советов, С. Яковлев. – М. : ВШ, 1998. – 319 с.
4. *Клифтон, Б.* Google Analytics: профессиональный анализ посещаемости веб-сайтов / Б. Клифтон. – М. : Вильямс, 2009. – 400 с.
5. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / [Е. Большакова, Д. Ландэ, А. Носков и др.]. – М. : МИЭМ, 2011. – 272 с.
6. *Брайчевский, С.* Современные информационные потоки / С. Брайчевский, Д. Ландэ // Научно-техническая информация. – 2005. – № 11. – С. 21–33.
7. *Корнеев, В.* Базы данных. Интеллектуальная обработка информации / В. Корнеев, А. Гареев, С. Васютин, В. Райх. – М. : Нолидж, 2000. – 352 с.
8. Основы моделирования и оценки электронных информационных потоков / [Ландэ Д., Фурашев В., Брайчевский С., Григорьев О.]. – К. : Інжиніринг, 2006. – 348 с.
9. *Ландэ, Д.* Основы интеграции информационных потоков: монография / Д. Ландэ. – К. : Інжиніринг, 2006. – 240 с.
10. Математична лінгвістика / [Висоцька В., Пасічник В., Щербина Ю., Шестакевич Т.] – Л. : «Новий Світ-2000», 2012. – 359 с.

11. Федорчук, А. Г. Контент-мониторинг информационных потоков [Электронный ресурс] / А. Г. Федорчук // Б-ки Нац. акад. наук: пробл. функционирования, тенденции развития. – К.,

2005. – Вып. 3. – Режим доступа: <http://www.nbuv.gov.ua/articles/2005/05fagmip.html>. – Загол. с экрана.

Стаття надійшла до редакції 13.12.2013.
Після доробки 07.03.2014.

Высоцкая В. А.¹, Чирун Л. В.²

¹Ассистент, Национальный университет «Львовская политехника», Украина

²Канд. техн. наук, доцент, Национальный университет «Львовская политехника», Украина

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ СИСТЕМ ЭЛЕКТРОННОЙ КОНТЕНТ-КОММЕРЦИИ

В статье предложена формальная модель системы электронной контент-коммерции и формальные модели обработки информационных ресурсов. Описана разработанная общая архитектура систем электронной контент-коммерции для облегчения этапов реализации жизненного цикла коммерческого контента. Предложенные общие принципы проектирования архитектуры систем электронной контент-коммерции дают возможность реализовывать процесс обработки информационных ресурсов для сокращения цикла производства, экономии времени и расширения возможностей ведения электронной коммерции.

Ключевые слова: информационный ресурс, коммерческий контент, контент-анализ, контент-мониторинг, контентный поиск, система электронной контент-коммерции.

Vysotska V.¹, Chyrun L.²

¹Assistant, National University «Lviv Polytechnic», Ukraine

²Ph.D, Associate Professor, National University «Lviv Polytechnic», Ukraine

CONCEPTUAL MODEL OF ELECTRONIC CONTENT COMMERCE SYSTEMS

The given article is devoted to the development of standardized methods and software of Web content processing in e-business systems. The model of electronic content commerce systems is developed. The article is devoted to the development of unified methods and software tools for processing information resources in the electronic content commerce systems. A new detailed classification of electronic commerce systems and electronic content commerce systems is proposed. A formal model and generalized typical architecture of electronic content commerce systems are declared. 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 systems of electronic content commerce. A complex method of content forming, the operational method of content management and complex method of content support are developed. Software tools of content formation, management and support are developed. Methods of designing and implementation of electronic content commerce systems on the example of online newspapers, which reflects the results of theoretical research, are developed.

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

REFERENCES

- Bereza A., Kozak I., Levchenko F. Elektronna komertsiya. Kiev, KNEU, 2002, 326 p.
- Berko A., Vysotska V., Pasichnyk V. Systemy elektronnoyi kontent-komertsiyi. Leningrad, NULP, 2009, 612 p.
- Sovetov B., Yakovlev S. Modelirovanie sistem. Moscow, VS, 1998.
- Clifton B. Google Analytics: professionalny analiz poseschaemosti web-saytov. Moscow, Vilyams, 2009, 400 p.
- Bolshakova E., Lande D., Noskov A., Klyshinsky E., Peskova O., Yagunova E. Avtomaticheskaya obrabotka tekstov na estestvennom yazyke i kompyuternaya lingvistika. Moscow, MIEM 2011, 272 p.
- Braychevsky S., Lande D. Sovremennye informatsionnye potoki, *Nauchno-tehnicheskaya informatsiya*, 2005, No. 11, pp. 21–33.
- Korneev V., Gareev A., Vasyutin S., Freihim V. Bazy dannyh. Intellektualnaya obrabotka informatsii. Moscow, Nolidzh, 2000, 352 p.
- Lande D., Furashev V., Braychevsky S., Grigorev O. Osnovy modelirovaniya i otsenki elektronnyh informatsionnyh potokov. Kiev, Inzhiniring, 2006, 348 p.
- Lande D. Osnovy integratsii informatsionnyh potokov: monografiya. Kiev, Inzhiniring, 2006, 240 p.
- Pasichnyk V., Pasichnyk V., Scherbyna J., Shestakevych T. Matematychna linhvistyka / V. Vysotska, V. Pasichnyk, J. Scherbyna, T. Shestakevych. Leningrad, Novyy Svit-2000, 2012, 359 p.
- Fedorchuk A. G. Kontent-monitoring informatsionnyh potokov, B-ky nats. akad. nauk: probl. funktsionirovaniya, tendentsii razvitiya, Kyiv, 2005, No. 3, Mode of access: <http://www.nbuv.gov.ua/articles/2005/05fagmip.html>