

СТВОРЕННЯ ТА ЗАСТОСУВАННЯ СИСТЕМ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

© Берко А. Ю., Висоцька В. А., Чирун Л. В., 2014

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 (СЕКК) та реалізувати підсистеми формування, управління та супроводу комерційного контенту. Процес проектування та створення СЕКК за результатами Інтернет-маркетингу є ітеративним і містить низку етапів від аналізу, проектування, розроблення плану до створення прототипу й експериментальних випробувань, починаючи з формування специфікацій, верстки, створення шаблону контенту, формування контенту, та його подальше розміщення згідно зі структурою сайта. На початкових етапах до визначення функціональних вимог і початку процесу розроблення до процесу залучають кінцевих користувачів за допомогою листів опитування, альтернатив проектування і прототипів різного ступеня готовності. Без значних зусиль збирають цінну інформацію, одночасно викликаючи у користувачів відчуття безпосередньої участі в процесі проектування та завойовуючи їх довіру. Проаналізовано способи та моделі послідовності опрацювання інформаційних ресурсів у системах електронної контент-комерції та виділено основні закономірності переходу від процесів формування комерційного контенту до його реалізації. Створено формальну модель систем електронної комерції, що дало змогу реалізувати етапи життєвого циклу комерційного контенту. Розроблено формальні моделі опрацювання інформаційних ресурсів у системах електронної контент-комерції, що дало змогу створити узагальнену типову архітектуру системи електронної контент-комерції. Запропоновано узагальнену типову архітектуру системи електронної контент-комерції, що дало змогу реалізувати процеси формування, управління та реалізації комерційного контенту.

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

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 content commerce systems (ECCS) and to implement the subsystem of commercial content formation, management and support. The process of ECCS design and creation as an Internet marketing result is iterative. It contains in its structure a number of stages (from the analysis, design and development of a plan to a prototype construction and experimental tests). The latter process begins with the specifications and layout formation, content template creation, content formation and its subsequent publishing according to the site's structure. In the initial stages (before setting functional requirements and development process initiation) regular users are involved into the process through poll letters, alternative design and prototyping of varying degrees of readiness. Thus, valuable information is collected without much effort, along with both evoking users' sense of direct involvement in the design process, as well as winning their trust. The paper analyzes sequence methods and models of information resources processing

in electronic content-commerce systems. It also allocates the basic laws of the transition from commercial content formation to its implementation. The formal model of ECCS is created, which allows implement in phases of the commercial content lifecycle. The developed formal model of information resources processing in electronic content-commerce systems allows us to create a generalized typical architecture of ECCS. The generalized typical architecture of ECCS is proposed in the paper, which helps implement the processes of commercial content formation, management and realization.

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

Вступ. Загальна постановка проблеми

Одним із основних завдань проекту створення СЕКК є розроблення інформаційної архітектури інформаційного ресурсу, яку формують за зворотною реакцією користувачів відповідно до типу поширення комерційної діяльності [1, 2]. Основні класи користувачів/персонажів інформаційного ресурсу (клієнти, керівники робочих груп і адміністратори) визначають дизайн інформаційного ресурсу і процес ухвалення рішень. Тип інформаційного ресурсу як Web-вітрина має суттєві недоліки: дає змогу організувати тільки торгівлю на замовлення; не дає змоги налагодити торгівлю через Інтернет; не зменшує операційні витрати й на утримання штату; неповоротке рішення з погляду управління і недостатньо гнучке з погляду організації маркетингових акцій; імідж компанії гірший, ніж у компанії з СЕКК; малоефективна і нерентабельна організація торгівлі [1]. Перспективним типом інформаційного ресурсу є комплексна торговельна система для організації on-line продажу комерційного контенту. На створення СЕКК необхідна більша кількість разових витрат порівняно з Web-вітриною, але вони ефективніші, оскільки використання СЕКК рентабельніше щодо обороту [2]. Існує альтернатива самостійному створенню громіздкої системи – це орендування у спеціалізованій компанії: великі разові й непродуктивні витрати рівномірно розподіляються в часі [2].

Аналіз останніх досліджень та публікацій

Процес проектування та створення СЕКК за результатами інтернет-маркетингу є інтерактивним і містить низку етапів від аналізу, проектування, розроблення плану до створення прототипу й експериментальних випробувань, починаючи з формування специфікацій, верстки, створення шаблону та формування комерційного контенту, а також його подальше розміщення згідно зі структурою інформаційного ресурсу та управління (рис. 1) [1, 2]. На початкових етапах до визначення функціональних вимог і початку розроблення до процесу залучають кінцевих користувачів за допомогою листів опитування, альтернатив проектування і прототипів різного ступеня готовності. Без значних зусиль збирають цінну інформацію, одночасно викликаючи у користувачів відчуття безпосередньої участі в процесі проектування та завойовуючи їх довіру. Головним компонентом проекту є розроблення інформаційної архітектури системи та інформаційного ресурсу. Завдяки моніторингу продуктивності та успіху інформаційного ресурсу користувачі визначають одну із вимог – якість доступу до даних та цільову аудиторію. За реакцією користувачів формують інформаційну архітектуру ресурсу відповідно до типу СЕКК [1, 2]. Основні класи користувачів/персонажів інформаційного ресурсу (користувачі як клієнти, користувачі як відвідувачі, користувачі як автори, модератори, аналітики, керівники робочих груп і адміністратори) визначають дизайн інформаційного ресурсу і процес ухвалення рішень. Визначивши важливу інформацію та її зв'язок із основними класами користувачів, створюють архітектуру і/або шаблон контенту інформаційного ресурсу, його ієрархію, способи подання і взаємодії кожного класу користувачів з цією інформацією (наприклад, дані конференції містять теми та питання порядку денного і/або сесії, планування цих питань) [1–9]. СЕКК обов'язково містить Web-вітрину з каталогом комерційного контенту (з можливістю пошуку) і необхідними інтерфейсними елементами для введення реєстраційної інформації, формування замовлення, здійснення платежів через Інтернет, оформлення доставки (e-mail, e-кабінет або on-line), одержання інформації про компанію-продавця і on-line допомоги [1].

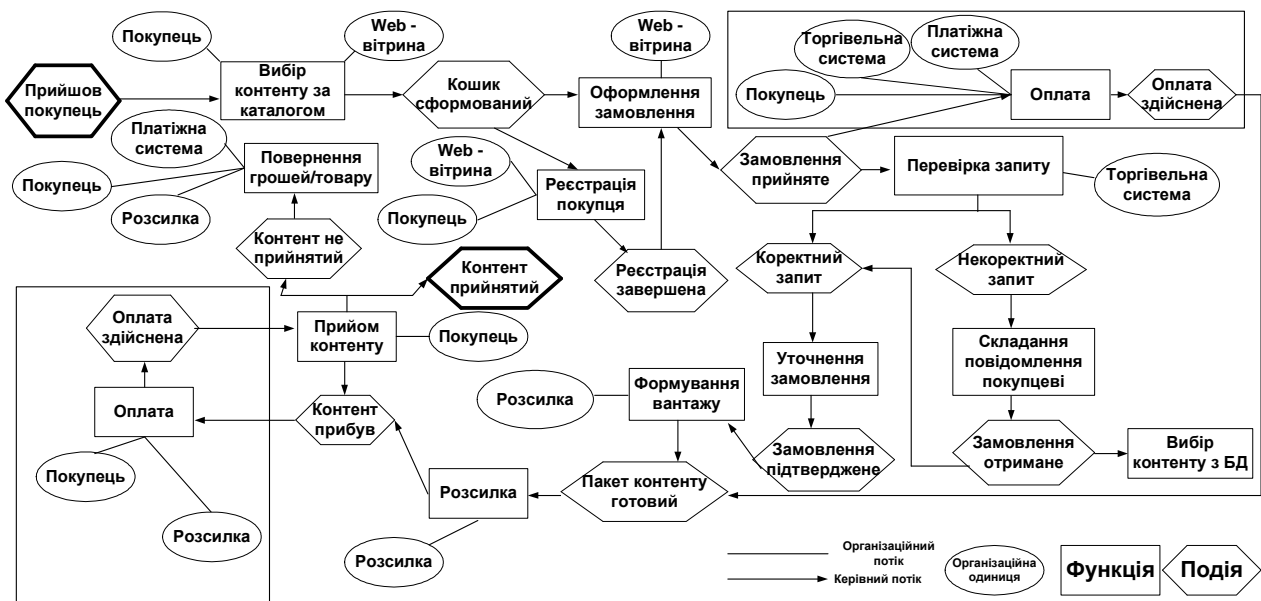


Рис. 1. Структура підсистеми управління комерційним контентом

Процес опрацювання інформаційних ресурсів складається з таких підсистем: формування (збирання/створення, систематизація, поширення), управління (опрацювання, аналіз, подання), супровід (структурування, модерація, узагальнення) комерційного контенту [1–9]. Визначивши пріоритетні тематики контенту та їх зв'язок із основними класами користувачів, створюють архітектуру інформаційного ресурсу, ієрархію комерційного контенту, його способи подання і взаємодії з ним кожного класу користувачів (наприклад, теми та питання порядку денного або сесії, планування цих питань) [1–9]. Інформаційний ресурс СЕKK – це множина даних з набором властивостей, які є об'єктом дій інформаційної технології (ІТ) з перетворення їх на комерційний контент. Результат застосування однієї ІТ є інформаційним ресурсом іншої [2]. Контент (від англ. *Content* – зміст, вміст, суть) у галузі ІТ є формалізованими відомостями і знаннями, розміщеними у середовищі системи і, на відміну від даних, без детальної специфікації їх властивостей, способів формалізації і впорядкування [2]. Перетворення різнорідних за природою, змістом та походженням даних на узгоджений централізований інформаційний ресурс є однією з важливих проблем побудови та функціонування СЕKK. 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 [2].

Важливим є забезпечення інваріантності середовища СЕKK 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].

Виділення проблем

Виникає проблема створення єдиного концептуального опису всього інформаційного ресурсу для стабільного підтримання зовнішніх/внутрішніх позначень контенту відповідно до їх завдань, вимог та змін. 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. За основу взято такі основні властивості комерційного текстового контенту в СЕKK, як синтаксис, структура та семантика, та вибрано основні фактори класифікації – способи його подання, структурування та призначення інформаційного ресурсу, організація доступу до нього. Загальні принципи формування інформаційних ресурсів СЕKK визначають порядок і способи відбору інформації із первинних джерел, її фіксації, фільтрування, перетворення до визначеного формату для формування комерційного контенту і розміщення в БД.

Формулювання мети

Метою є розроблення формальних описів СЕКК і таких процесів опрацювання інформаційних ресурсів, як формування, управління та супровід комерційного контенту. Необхідно розробити загальну структуру СЕКК для полегшення реалізації етапів життєвого циклу контенту. Також необхідно розробити вимоги до проектування структури СЕКК для реалізації опрацювання інформаційних ресурсів зі скороченням циклу виробництва, економією часу та розширенням можливостей е-комерції.

Аналіз отриманих наукових результатів

Основними підсистемами опрацювання інформаційних ресурсів у СЕКК є формування, управління та супровід контенту, які мають такі відношення: *контент* → *формування контенту* → *база даних* → *управління контентом* → *інформаційний ресурс*, або *запит користувача* → *управління контентом* → *інформаційний ресурс* → *супровід контенту* → *база даних*. Відповідно, підсистеми мають такий зв'язок: *формування контенту* → *управління контентом* → *супровід контенту*. Процес опрацювання інформаційних ресурсів виконується за такою схемою: *формування контенту* → *управління контентом* → *супровід контенту*, тому СЕКК подано як

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

де величини кортежу описують процес функціонування системи (табл. 1) [10].

Таблиця 1

Величини моделі системи електронної контент-комерції

№	Назва	Позначення	Діапазон	Множина
1	Множина вхідних даних з різних джерел інформації	$x_i \in X$	$i = \overline{1, n_X}$	$X = \{x_1, x_2, \dots, x_{n_X}\}$
2	Множина запитів користувачів	$q_d \in Q$	$d = \overline{1, n_Q}$	$Q = \{q_1, q_2, \dots, q_{n_Q}\}$
3	Оператор формування комерційного контенту	<i>Formation</i>	–	–
4	Множина внутрішніх параметрів системи	$h_k \in H$	$k = \overline{1, n_H}$	$H = \{h_1, h_2, \dots, h_{n_H}\}$
5	Множина комерційного контенту	$c_r \in C$	$r = \overline{1, n_C}$	$C = \{c_1, c_2, \dots, c_{n_C}\}$
6	Множина параметрів впливу зовнішнього середовища	$v_l \in V$	$l = \overline{1, n_V}$	$V = \{v_1, v_2, \dots, v_{n_V}\}$
7	Оператор управління комерційним контентом	<i>Management</i>	–	–
8	Оператор супроводу комерційного контенту	<i>Support</i>	–	–
9	Множина сторінок інформаційного ресурсу системи	$z_w \in Z$	$z = \overline{1, n_Z}$	$Z = \{z_1, z_2, \dots, z_{n_Z}\}$
10	Час транзакції опрацювання інформаційних ресурсів	$t_p \in T$	$p = \overline{1, n_T}$	$T = \{t_1, t_2, \dots, t_{n_T}\}$
11	Множина статистичних даних роботи системи	$y_j \in Y$	$j = \overline{1, n_Y}$	$Y = \{y_1, y_2, \dots, y_{n_Y}\}$

Оператор формування комерційного контенту *Formation* – відображення контенту в новий стан, що відрізняється від попереднього появою нової частини контенту, яка доповнює попередній стан

$$Formation : (x_i, c_r, u_f, t_p, \Delta t) \rightarrow c_{r+1}, \quad (2)$$

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

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

Оператор управління комерційним контентом *Management* – відображення контенту в новий стан, який відрізняється від попереднього стану значеннями визначальних параметрів (актуальність, повнота, релевантність, автентичність, достовірність), що задовольняють наперед визначені вимоги

$$Management : (q_d, z_w, c_r, h_k, u_M, t_p, \Delta t) \rightarrow (c'_r, z_{w+1}), \quad (4)$$

де $q_d \in Q$, $h_k \in H$, $H = \{h_1(c_{i+1}, q_d), \dots, h_{n_H}(c_{i+n_H}, q_d)\}$ – множина умов управління комерційним контентом. Управління комерційним контентом подано як

$$z_w = \left\{ \bigcup_{k=1}^{n_H} h_k(c_{i+1}, q_d) \mid \begin{array}{l} (c_{i+k} \in C) \wedge (q_d \in Q) \wedge (h_k \in H_q), \\ H = H_{q_d} \vee H_{q_d}^-, i = \overline{1, n_C}, d = \overline{1, n_Q}, k = \overline{1, n_H} \end{array} \right\}. \quad (5)$$

Оператор супроводу комерційного контенту *Support* – відображення комерційного контенту в колекцію значень, що утворюються як результат аналізу, моніторингу, оцінювання взаємодії з користувачем, пошуковими системами та іншими інформаційними ресурсами, що є основою для прийняття рішень щодо формування та управління комерційним контентом.

$$Support : (c_r, q_d, v_l, h_k, z_w, u_s, t_p, \Delta t) \rightarrow y_i, \quad (6)$$

де $v_l = (v_1(q_i, h_k, c_r, z_w, t_p), \dots, v_{n_v}(q_i, h_k, c_r, z_w, t_p))$ – множина умов супроводу комерційного контенту та впливів середовища на систему. Вихідні дані реалізовано

$$y_j = \left\{ \bigcup_l v_l \left| \begin{array}{l} (\exists q_d \in Q) \wedge (\exists z_w \in Z) \wedge (\forall v_l \in V) \wedge (\forall (c_r \wedge q_d) \in h_k), \\ V = V_{q_d} \vee V_{z_w}, d = \overline{1, n_Q}, l = \overline{1, n_V}, w = \overline{1, n_Z}, r = \overline{1, n_C}, k = \overline{1, n_H} \end{array} \right. \right\}. \quad (7)$$

I. Підсистему формування контенту реалізують у вигляді контент-моніторингу зі збирання контенту з різних джерел за визначеними методами. Для підсистеми формування комерційного контенту забезпечують створення бази даних контенту відповідно до інформаційних потреб постійних/потенційних користувачів: *Модератор* → *створення контенту* → *база даних* → *систематизація контенту* → *база даних* → *поширення контенту* → *редактор, і/або Інформаційний ресурс (джерело)* → *збирання контенту* → *база даних* → *систематизація контенту* → *база даних* → *поширення контенту* → *модератор*. У результаті збирання і первинного опрацювання контент приводиться до єдиного формату, класифікується відповідно до визначеного рубрикатора та йому приписують дескриптори з ключовими словами. Підсистема забезпечує постійне поповнення бази даних контентом, ефективний одночасний доступ до бази даних користувачами, зручні засоби пошуку необхідного контенту. Підсистема дає змогу економити на ресурсах адміністрування та аналізу функціонування системи; підтримувати анонімність; не перевантажувати інтернет-трафік; самостійного сканувати Інтернет. Процес формування комерційного контенту для інформаційного ресурсу забезпечує відображення множини вхідних даних з різних джерел інформації в множину сформованого комерційного контенту та збереженого у відповідній базі даних у СЕКК

$$Source(x_i) \rightarrow x_i \rightarrow X \rightarrow Formation(u_f, x_i, t_p) \rightarrow c_r \rightarrow C \rightarrow DataBase(C), \quad (8)$$

де $Source(x_i)$ – джерело x_i контенту, x_i – i -й контент з джерела; X – множина даних з відповідного джерела; $Formation(u_f, x_i, t_p)$ – оператор формування комерційного контенту з x_i в фіксований час t_p за умов u_f ; c_r – сформований згідно з умовою u_f r -й комерційного контент; C – множина комерційного контенту; $DataBase(C)$ – база даних комерційного контенту.

Формування комерційного контенту подано суперпозицією функцій

$$Formation = Dissemination \circ BuDigest \circ Caterization \circ KeyWords \circ Formatting \circ Backup \circ Creation, \quad (9)$$

або, відповідно,

$$Formation = Dissemination \circ BuDigest \circ Caterization \circ KeyWords \circ Formatting \circ Backup \circ Gathering, \quad (10)$$

де $Creation$ – оператор створення комерційного контенту; $Gathering$ – оператор збирання контенту з різних джерел; $Formatting$ – оператор форматування комерційного контенту; $KeyWords$ – оператор виявлення ключових слів і понять комерційного контенту; $Caterization$ – оператор автоматичної рубрикації комерційного контенту; $Backup$ – оператор виявлення дублювання змісту комерційного контенту; $BuDigest$ – оператор формування дайджестів комерційного контенту; $Dissemination$ – оператор вибіркового поширення комерційного контенту.

Процес формування комерційного контенту подано як

$$Formation = \left\langle \begin{array}{l} X, Creation, Gathering, Formatting, KeyWords, Backup, \\ Caterization, BuDigest, Dissemination, T, C \end{array} \right\rangle. \quad (11)$$

Множина операторів $Gathering, Creation, Backup, Formatting, KeyWords, Categorization, BuDigest, Dissemination$ є адекватною у процесі формування комерційного контенту.

II. Підсистема управління контентом складається із таких відношень: користувач → опрацювання контенту → база даних → аналіз контенту → база даних → подання контенту → користувач. Аналіз контенту та інформаційного ресурсу, їхнє моделювання та подання є одним з найінформативніших методів кількісного дослідження динаміки окремих тематичних потоків комерційного контенту. За зміною величин управління контентом визначають швидкість розвитку окремих тематичних напрямів або всього контентного простору. Стійкі статистичні зв'язки між окремим контентом свідчать про кореляції окремих тематик, ефективність посилань на публікації попередників, більш ранні цитування, републікації тощо. Механізми, що ґрунтуються на узагальнених методах кластерного аналізу, виявляють повідомлення в потоках контенту, що формують нові тематичні напрями. Кластерний аналіз, теорія фракталів і автотельних процесів, якщо їх коректно застосовують, кількісно оцінюють ступінь зв'язку в тематичних контентних потоках.

Процес управління комерційним контентом подано такою загальною схемою зв'язків:

$$User(q_d) \rightarrow q_d \rightarrow Q \rightarrow H(c_{i+1}, q_d) \rightarrow Management(q_d, c_r, h_k, t_p) \rightarrow z_w \rightarrow User(z_w), \quad (12)$$

де $User(q_d)$ – формування запиту користувачем системи; q_d – запит від користувача системи; Q – множина запитів користувачів; $H(c_{i+1}, q_d)$ – множина параметрів управління комерційним контентом c_{i+1} відповідно до запиту q_d ; $Management(q_d, c_r, h_k, t_p)$ – оператор управління комерційним контентом; z_w – сформована сторінка інформаційного ресурсу як відповідь на запит q_d ; $User(z_w)$ – перегляд користувачем відповіді на запит q_d .

Управління комерційним контентом подано суперпозицією функцій

$$Management = Presentation \circ Weight \circ Edit, \quad (13)$$

де $Presentation$ – оператор формування та подання сторінки інформаційного ресурсу; $Edit$ – оператор редагування та модифікації контенту; $Weight$ – оператор визначення ваги блока контенту.

Процес управління комерційним контентом подано як

$$Management = \langle H, C, Q, Presentation, Edit, Weight, T, Z \rangle. \quad (14)$$

1. Оператор редагування та модифікації комерційного контенту

$$Edit : (c_r, h_k, u_l, t, \Delta t) \rightarrow c_{r+1}. \quad (15)$$

2. Оператор визначення ваги блока комерційного контенту

$$Weight (c_{r+1}, y_j, u_l, t, \Delta t) \rightarrow c'_{r+1}, \quad (16)$$

3. Оператор формування та подання сторінки інформаційного ресурсу

$$Presentation (c'_{r+1}, h_k, z_w, q_d, t, \Delta t) \rightarrow z_{w+1}. \quad (17)$$

де $h_k \in H$, $h_k = \{h_{1k}, h_{2k}, \dots, h_{mk}\}$ – множина параметрів процесу управління комерційним контентом (h_{1k} – актуальність, h_{2k} – релевантність, h_{3k} – повнота, h_{4k} – автентичність, h_{5k} – достовірність комерційного контенту); $u_l \in U$, $u_l = \{u_{1l}, u_{2l}, \dots, u_{nl}\}$ – множина критеріїв процесу управління комерційним контентом (u_{1l} – коефіцієнт розташування блока в контенті, u_{2l} – коефіцієнт ключових слів у блоці, u_{3l} – коефіцієнт статистичної важливості ключових слів, u_{4l} – коефіцієнт наявності ключових слів із запиту користувача, u_{5l} – коефіцієнт обсягу ключових слів із запиту).

1. Генерація сторінок за запитом. Підсистема подання генерує сторінку із змістом у разі запиту на нього на основі контенту з бази даних. У режимі діалогового доступу до бази даних забезпечується перегляд, пошук і відображення контенту, надається можливість звертання до оригіналів контенту в Інтернеті. Контент у базі даних змінюється за допомогою модуля редагування. Сторінки наново створює сервер для кожного запиту, що збільшує навантаження на ресурси системи. Навантаження знижується у разі використання засобів хешування у сучасних Web-серверах. Завдання повнотекстового пошуку у великих масивах контенту є неефективним. Проблема точності вирішує пошук в анованому контенті. Замість пошуку в повному контенті доцільно вести пошук за

анотаціями – пошуковими образами контенту. Квазіреферат для великого масиву контенту є утворенням, що віддалено нагадує вихідний контент та найчастіше не сприймається людиною. Але пошуковий образ такого контенту зі зваженими ключовими словами і фразами приводить до цілком адекватних результатів за повнотекстового пошуку. Квазіреферат будується з фрагментів контенту із найбільшими ваговими значеннями. Управління комерційним контентом із генерацією сторінок за запитом користувача системи електронної контент-комерції подано як

$$Management_Q = \langle H, C, Q, Presentation, Edit, Weight, T, Z \rangle. \quad (18)$$

2. Генерація сторінок під час редагування – це набір статичних сторінок, які безпосередньо редагує модератор прямо в базі даних системи електронної контент-комерції. В такій підсистемі відсутня інтерактивність між відвідувачем та контентом інформаційного ресурсу в СЕКК. Управління контентом з генерацією сторінок під час редагування інформаційного ресурсу модератором подано як

$$Management_E = \langle C, H, Edit, Weight, T, Z \rangle. \quad (19)$$

1. **Змішаний тип** об'єднує особливості попередніх процесів управління (рис. 2) та подається як

$$Management_M = \langle C, Q, H, Presentation, Edit, Weight, Caching, T, Z \rangle. \quad (20)$$

Переваги хешування сторінок з контентом: підсистема подання генерує сторінку один раз (рис. 2); сторінка в кеші існує час Δt – поки актуальний контент; готова сторінка швидше завантажується з кешу; кеш оновлюється періодично вручну/автоматично (після закінчення деякого терміну часу Δt ; або у разі модифікації розділів інформаційного ресурсу чи контенту). Види реалізації процесу генерації сторінок змішаного типу: хешуванням (рис. 3); через формування інформаційних блоків (рис. 4):

- а) редагування інформаційного ресурсу (збереження інформаційних блоків);
- б) формування сторінок (збирання сторінки із блоків згідно із запитом користувача).

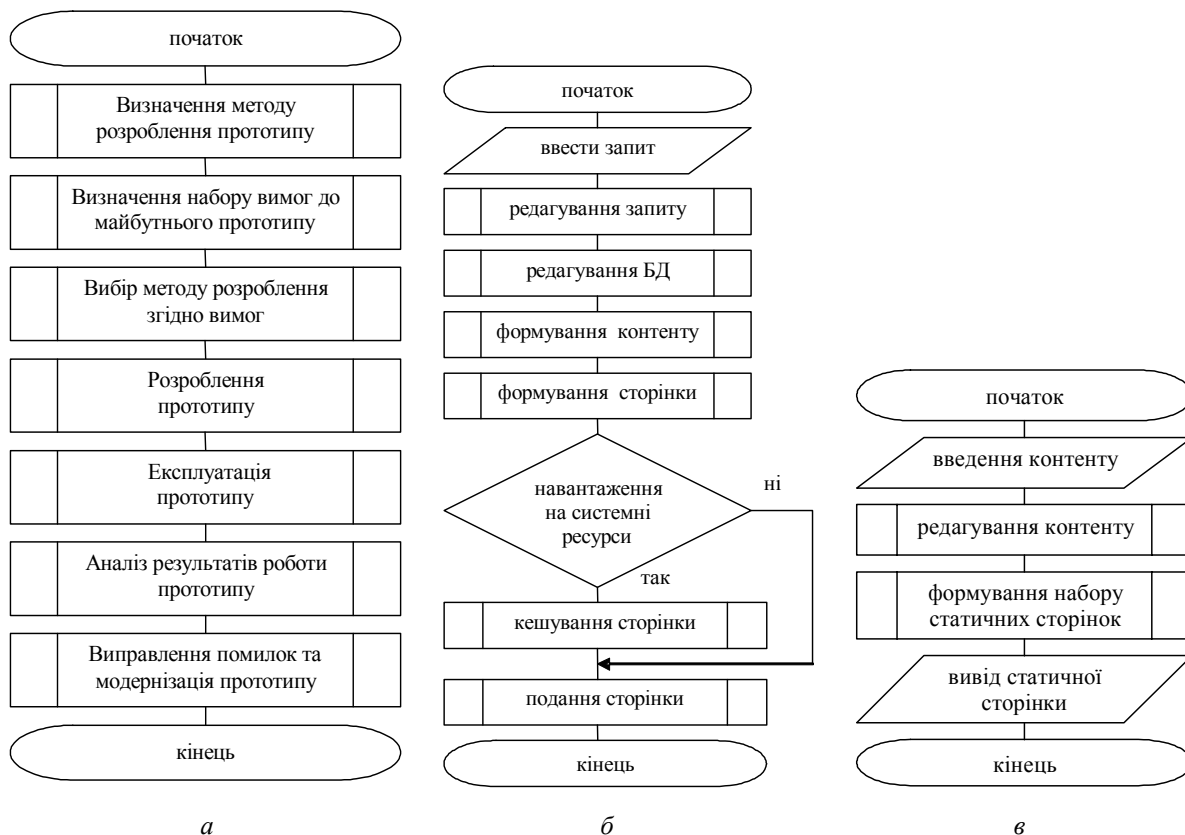


Рис. 2. Процес: а – створення прототипу, генерації сторінок; б – інформаційного ресурсу; в – під час редагування

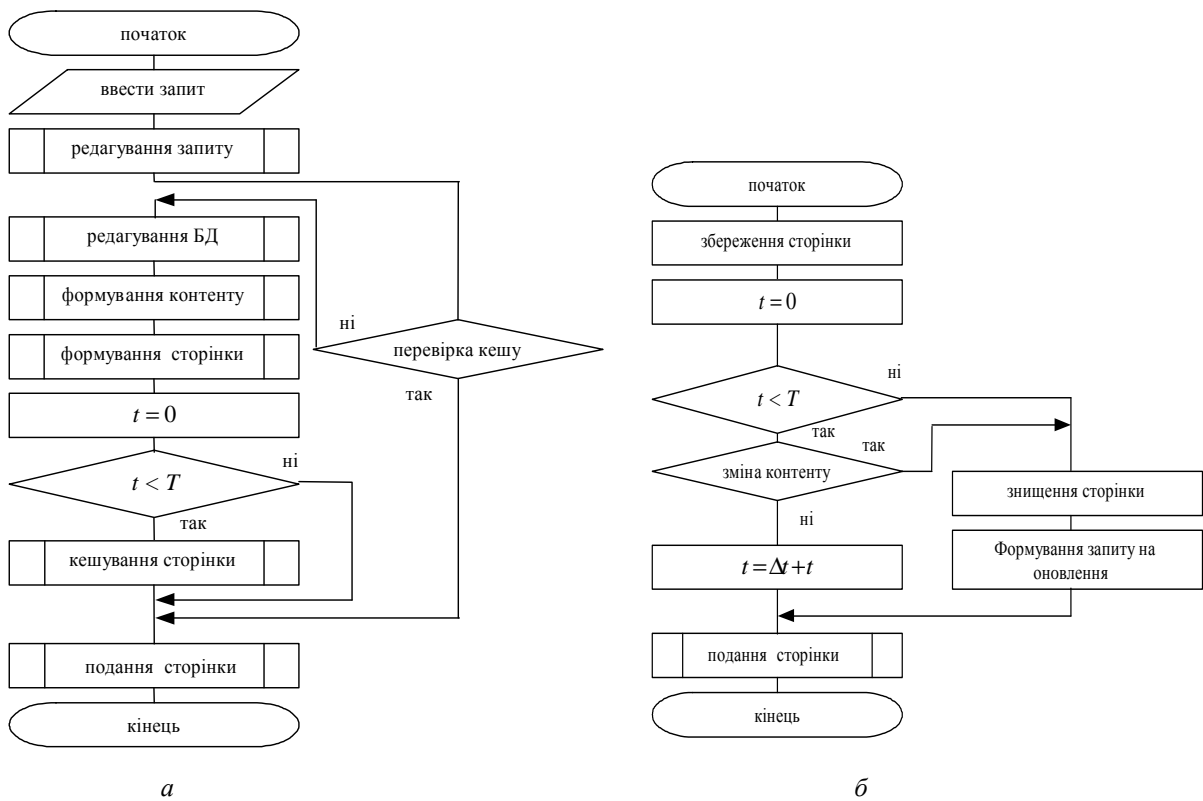


Рис. 3. Процес: а – генерації; б – хешування сторінок змішаного типу

На рис. 3, а подано етапи процесу генерації сторінок змішаного типу, а на рис. 3, б – кроки етапу хешування сторінки для процесу генерації сторінок змішаного типу. На рис. 4, а наведено кроки етапу генерації сторінок змішаного типу через формування інформаційних блоків під час редагування інформаційного ресурсу (збереження інформаційних блоків), а на рис. 4, б – кроки генерації сторінок змішаного типу через формування інформаційних блоків під час формування сторінок (збирання сторінки із інформаційних блоків у разі запиту контенту користувачем). Підсистему управління контентом змішаного типу складніше реалізувати, але вона оперативніше працює з користувачем через гнучкий інтерфейс, надаючи необхідний комерційний релевантний контент без надлишку інформації.

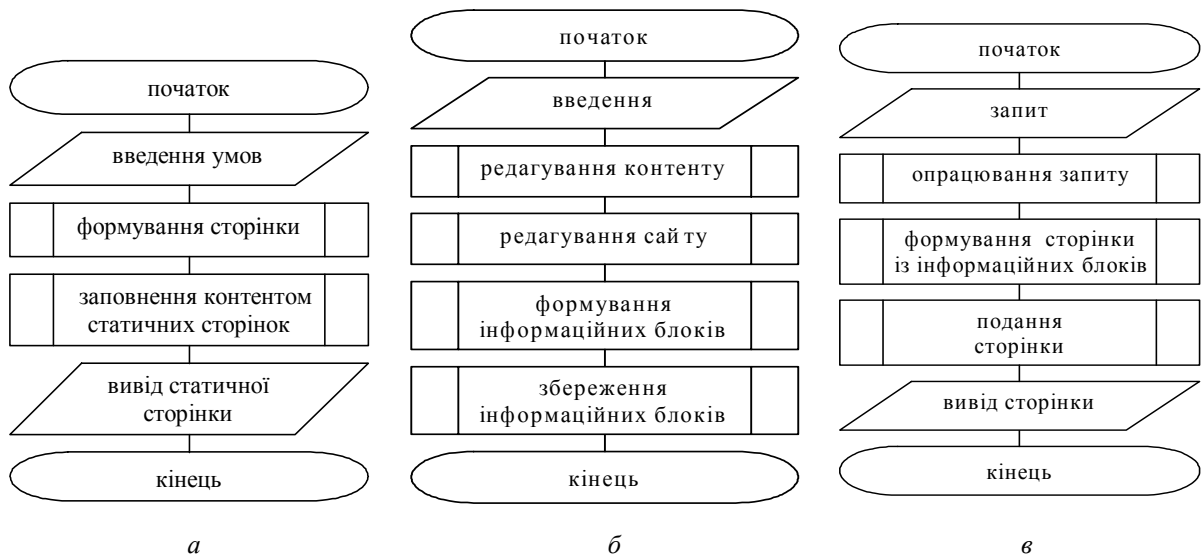


Рис. 4. Процес: а – заповнення сторінок під час редагування; б – першого; в – другого етапу генерації сторінок

III. Підсистема супроводу контенту – це комплексні послідовні процеси: користувач → структурування контенту → база даних → модерація контенту → база даних → узагальнення контенту → модератор. Процес технологічного супроводу комерційного контенту подано такою загальною схемою зв'язків: $User(q_d, z_w) \rightarrow q_d \rightarrow z_w \rightarrow V(q_d, z_w, t_p) \rightarrow Support(v_l, h_k, c_r, z_w, t_p, \Delta t) \rightarrow y_j \rightarrow Moderator(y_j)$. Супровід комерційного контенту подано суперпозицією функцій

$$Support = StatistAnal \circ CalRankConc \circ ConCorrTablConc \circ IdThemTop \circ BuInfPort, \quad (21)$$

де *BuInfPort* – оператор формування інформаційних портретів, *IdThemTop* – оператор виявлення тематичних сюжетів комерційного контенту, *ConCorrTablConc* – оператор побудови таблиць взаємозв'язку комерційного контенту, *CalRankConc* – оператор розрахунку рейтингів комерційного контенту, *StatistAnal* – оператор аналізу статистики функціонування системи.

Процес технологічного супроводу комерційного контенту подано як

$$Support = \left\langle \begin{array}{l} Q, C, H, V, T, BuInfPort, IdThemTop, \\ ConCorrTablConc, CalRankConc, StatistAnal, Z, Y \end{array} \right\rangle, \quad (22)$$

Актуальність розроблення загальної структури модуля супроводу комерційного контенту полягає у необхідності отримувати оперативні й об'єктивні оцінювання рівня конкуренції на сегменті фінансового ринку контенту; оцінювати рівень конкурентів та міру їх конкурентоспроможності на фінансовому ринку з поширення контенту. Причиною актуальності є швидкий темп збільшення попиту в поширенні контенту; постійна взаємодія суспільства і науковців з новітніми ІТ; можливість збільшення прибутку існуючих/потенційних клієнтів, збільшення рейтингу; покращення е-бізнесу через розуміння користувача як клієнта. Проблемаю супроводу комерційного контенту є відсутність загальних підходів до автоматичного аналізу функціонування СЕКК. З метою залучення більшої кількості користувачів на інформаційний ресурс СЕКК додають опції аналізу контенту (рейтинги, відгуки і коментарі користувачів). Інформаційний ресурс із відгуками про контент приваблює більшу кількість відвідувачів, але ця функціональність не охоплює модерації та призводить до надмірної інформації на Web-вітрині. Кінцевий споживач змушений сегментувати конструктивну інформацію, що вимагає зусиль та відштовхує потенційного клієнта. Контент, який потрапляє на інформаційний ресурс, не фільтрується, тобто взаємодія користувачів із порталом є односторонньою (табл. 2).

Таблиця 2

Підсистема супроводу комерційного контенту в системі електронної контент-комерції

Назва	Характеристика модуля Web-застосування
Рейтинги/ відгуки	Підсистема оцінювання рейтингу продукції СЕКК за наперед визначеною шкалою. Користувач може ознайомитись з рейтингом контенту і має право залишити свій відгук, що дозволяє побудувати постійну взаємодію між клієнтами та системою.
Питання/ відповіді	Підсистема організації спілкування клієнтів між собою, що в результаті надає повнішу інформацію про контент і, як наслідок, збільшує обсяги продажів.
Історія	Підсистема, принцип роботи якої базується на застосуванні «питання та відповіді», тільки працює в односторонньому режимі.

Із такою підсистемою СЕКК має різні можливості: підвищення рейтингу комерційного контенту; аналіз характеристик (коментарі, відгуки, побажання тощо) комерційного контенту з боку користувача; збирання, накопичення та опрацювання інформації про потреби кінцевого/потенційного користувача комерційного контенту (табл. 3).

Таблиця 3

Основні компоненти підсистеми супроводу комерційного контенту

Назва	Характеристика компонентів модуля супроводу контенту
Customer	Потенційний/існуючий клієнт із власним інформаційним ресурсом, де вмонтовується аплікація CMS з метою підвищення рейтингу комерційного контенту та кращого розуміння потреб кінцевого користувача через аналіз контенту типу питання/відповіді, рейтинги, відгуки, статті тощо.
OpenMQ	Черга типу FIFO.
CMS	Підсистема аплікації фільтрування контенту за наперед визначеними правилами індивідуально для кожного користувача як клієнта.
Workbench	Індивідуальний кабінет користувача як клієнта з можливістю перегляду контенту та аналізу статистики.
Solar	Підсистема індексування для швидкого пошуку комерційного контенту.

Аналізують статистику функціонування СЕКК згідно з аналізом реакцій на цю систему постійного/потенційного користувача (відвідування, запити, пошук за ключовими словами тощо). Застосування операцій елімінування, секвентування та паралелення у відповідній послідовності в певні проміжки часу сприяє ефективному проведенню аналізу реакції цільової та потенційної аудиторії на функціонування СЕКК та прогнозуванню відповідних змін щодо попиту на комерційний контент. Бази даних підсистеми супроводу комерційного контенту подано в табл. 4.

Таблиця 4

Бази даних підсистеми супроводу комерційного контенту в системі електронної контент-комерції

Назва	Характеристика баз даних модуля супроводу контенту
Master	Головна база даних, яка є основною ланкою системи і для розподілу навантаження реплікує дані на дочірні бази.
Display Submit	Дочірня база даних від Master, з якою працює інформаційний ресурс клієнта на зчитування того контенту, який має дозвіл на публікацію. Запобігає дублюванню інформації на першому етапі.
Reporting	На основі цієї бази даних формуються статистичні звіти для клієнтів.
Backup	Додаткове сховище даних.
Portal	База даних призначена для роботи з аплікацією Workbench.

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, поява нових вимог та категорій користувачів; порядку розподілу контенту та способів доступу користувачів.

Основні компоненти підсистеми супроводу комерційного контенту в СЕКК та їх взаємозв'язок подано на рис. 5. Для підвищення попиту на комерційний контент не весь контент викладають на Web-сторінках інформаційного ресурсу, а лише відфільтрований та структурований згідно з правилами структуризації комерційного контенту в системі електронної контент-комерції (рис. 5, табл. 5).

Таблиця 5

Правила структуризації комерційного контенту в системі електронної контент-комерції

Аналіз цензури	База даних нецензурної лексики
Підрахунок символів і відтинання тексту (залишається лише рейтинг) у разі недостатності, щоб називатися відгуком.	Конкретне правило на кількість символів.
Блокування контенту з назвами конкурентних товарів.	Конкурентні товари визначає модератор.
Блокування визначених відправників контенту.	База даних IP адрес.
Формування власного правила.	Визначає клієнт за бажанням.

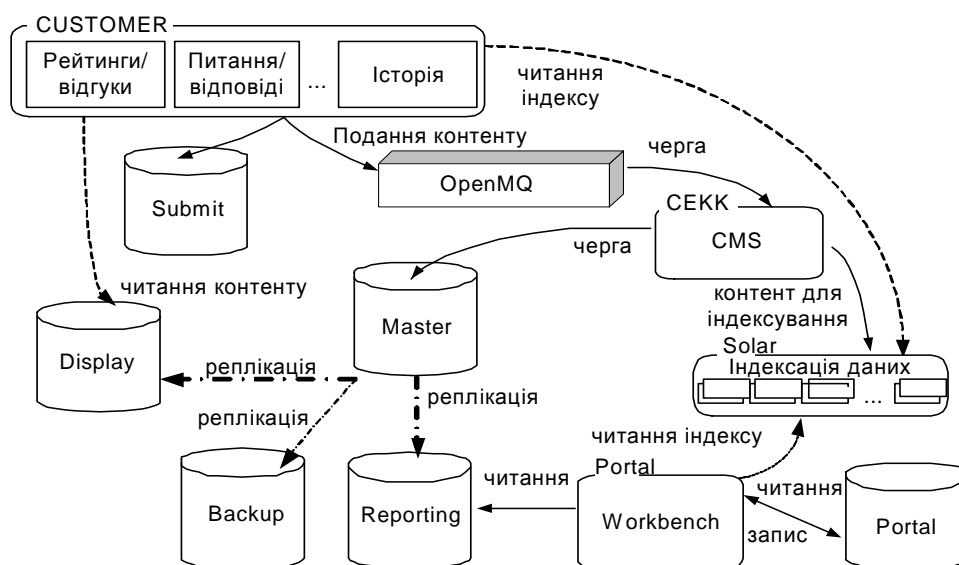


Рис. 5. Схема потоків даних у підсистемі супроводу контенту в системі електронної контент-комерції

У результаті аналізу функціонування системи електронної контент-комерції S та технологічного супроводу комерційного контенту C формується множина $Y = \{Y_P, Y_T, Y_C, Y_R\}$ відповідно до умов $V = \{V_P, V_T, V_C, V_R\}$, де $Y_P = Y_{Pc} \vee Y_{Pq}$ – підмножина інформаційних портретів комерційного контенту Y_{Pc} та постійних користувачів Y_{Pq} , Y_T – підмножина тематичних сюжетів комерційного контенту, Y_C – підмножина таблиць взаємозв'язку комерційного контенту, Y_R – підмножина рейтингів контенту, $V_P = V_{Pc} \vee V_{Pq}$ – 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. Множину інформаційних портретів комерційного контенту Y_{Pc} подано як $Y_{Pc} = BuInfPort(V_{Pc}, C, H, Q, T)$, а множину портретів постійних користувачів Y_{Pq} – як $Y_{Pq} = BuInfPort(V_{Pq}, Q, H, Z, T)$, де $V_P = V_{Pc} \vee V_{Pq}$ – множина умов формування портретів, $BuInfPort$ – оператор формування портретів $Y_P = Y_{Pc} \vee Y_{Pq}$.

Множину тематичних сюжетів комерційного контенту Y_T подано як $Y_T = IdThemTop(C, H, Q, V_T, T)$, де V_T – множина умов виявлення сюжетів комерційного контенту, $IdThemTop$ – оператор визначення тематичних сюжетів комерційного контенту Y_T . Множину таблиць взаємозв'язку комерційного контенту Y_C подано як $Y_C = ConCorrTablConc(C, V_c, T)$, де V_C – множина умов побудови таблиць взаємозв'язку комерційного контенту, $ConCorrTablConc$ – оператор побудови таблиць взаємозв'язку комерційного контенту. Множину рейтингів комерційного контенту Y_{Rc} подано як $Y_{Rc} = CalRankConc(C, Q, H, Y_C, V_{Rc}, Spam, Tonality, T)$, а множину рейтингів модераторів Y_{Rm} – як $Y_{Rm} = CalRankConc(C, Q, H, Y_C, V_{Rm}, T)$, де $V_R = V_{Rc} \vee V_{Rm}$ – множина параметрів для розрахунку рейтингів комерційного контенту, $Tonality(Q^+, Q^0, Q^-, T, H)$ – критерій тональності комерційного контенту, $Spam(Q, T)$ – оператор визначення фільтрування коментарів, $CalRankConc$ – оператор визначення рейтингів комерційного контенту та модераторів $Y_R = Y_{Rc} \vee Y_{Rm}$.

Множину вихідних статистичних даних Y подано як

$$Y = \{Y_P, Y_T, Y_C, Y_R\} = Support(V, C, Q, H, Z, T, \Delta T), \quad (23)$$

$$Y = \{Y_P, Y_T, Y_C, Y_R\} = Support(V_P, V_T, V_C, V_R, C, Q, H, Z, T, \Delta T), \quad (24)$$

де $Y_P = Y_{Pc} \vee Y_{Pq}$ – підмножина інформаційних портретів комерційного контенту та постійних користувачів, Y_T – підмножина тематичних сюжетів комерційного контенту, Y_C – підмножина таблиць взаємозв'язку комерційного контенту, $Y_R = Y_{Rc} \vee Y_{Rm}$ – підмножина рейтингів комерційного контенту та модераторів, $Support$ – оператор супроводу комерційного контенту. Отримані дані враховують під час створення або оновлення/модифікації інформаційного ресурсу в СЕКК (рис. 6).

Аналізом, вибіркою, побудовою та модерациєю статистичних даних (рис. 6) керує CMS на етапі супроводу контенту за правилами (табл. 6) та алгоритмами модерациї/узагальнення комерційного контенту (рис. 7). Підсистема супроводу контенту виконує процес підвищення рейтингу СЕКК із інтерактивним інтерфейсом. Вона забезпечує підтримку збирання маркетингової інформації з поширення контенту; підтримку інтерактивного діалогу з кінцевим споживачем контенту; формування каталогу комерційного контенту та послуг СЕКК; інформаційну підтримку кінцевого споживача контенту; формування віртуального кошика клієнта комерційного контенту; реєстрування кінцевого споживача комерційного контенту; опрацювання замовлення на комерційний контент.

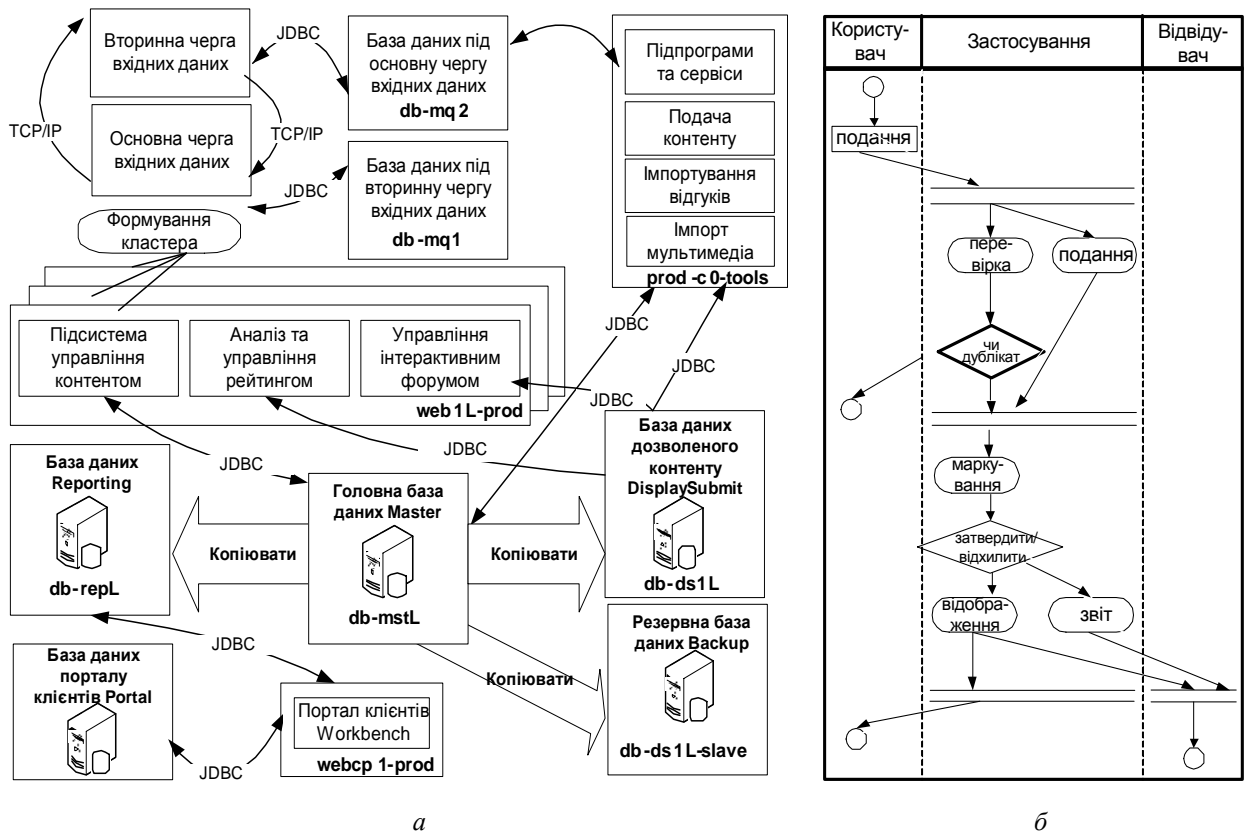


Рис. 6. Детальна схема (а) та діаграма активності супроводу комерційного контенту в СЕКК (б)

Таблиця 6

Правила модерації комерційного контенту із самонавчанням

Фільтр	Правило	Режим роботи
Нецензурної лексики	Перевірка за словником та заборона у разі TRUE	Напівавтоматичний, у разі FALSE перевіряє модератор, словник періодично поповнюють модератори.
За кількістю символів	Підрахунок кількості символів, порівняння із лімітом та заборона при FALSE	Автоматичний, ліміт визначає адміністратор або власник інформаційного ресурсу.
За наявністю URL-посилань	Визначення наявності URL-посилань, перевірка за списком та заборона у разі TRUE	Напівавтоматичний, у разі FALSE остаточно перевіряє модератор, список періодично поповнюють модератори.
За "blacklist" списком	Перевірка за списком та заборона у разі TRUE	Напівавтоматичний, у разі FALSE остаточно перевіряє модератор, список періодично поповнюється автоматично у разі перевищення ліміту.
За рейтингом	Підрахунок рейтингу, порівняння із лімітом та заборона у разі FALSE	Автоматичний, період і ліміт визначає адміністратор або власник інформаційного ресурсу.
За посиланням на конкурента	Перевірка за списком та заборона у разі TRUE	Напівавтоматичний, у разі FALSE перевіряє модератор, список періодично поповнюють модератори.
За IP адресами	Перевірка за списком та заборона у разі TRUE	Автоматичний, список періодично поповнюється автоматично у разі перевищення ліміту, ліміт визначає адміністратор або власник інформаційного ресурсу.
За ID користувачів	Перевірка за списком та заборона у разі TRUE	Автоматичний, список періодично поповнюється автоматично у разі перевищення ліміту, ліміт визначає адміністратор або власник інформаційного ресурсу.

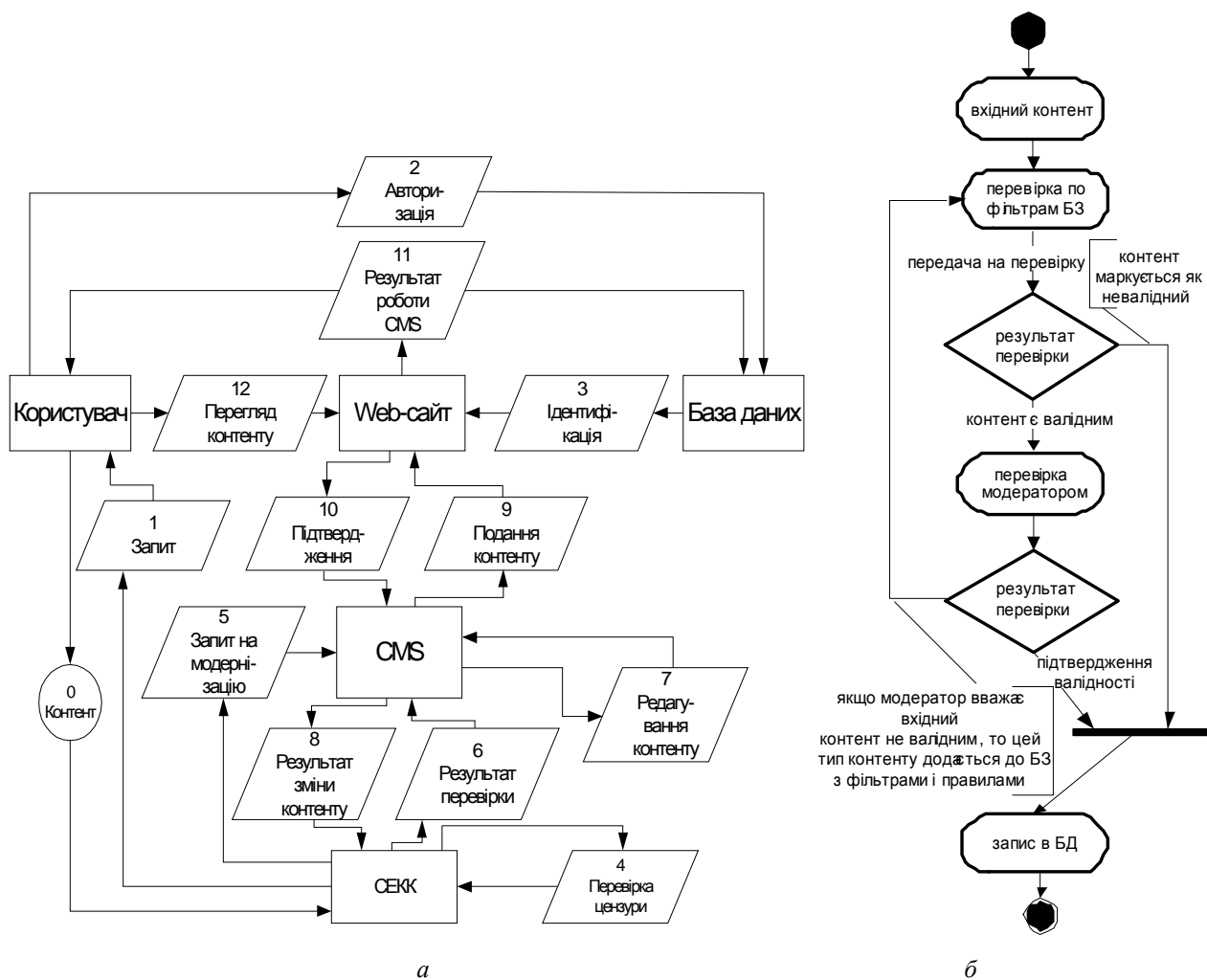


Рис. 7. Схема модерації контенту (а) та узагальнення комерційного контенту в СЕКК (б)

Запропонована концептуальна модель супроводу комерційного контенту дає змогу: будувати спрямовані моделі спілкування з користувачем; керувати динамічним матеріалом користувача; користувачам скеровувати бізнес клієнта; реалізовувати новий тип сприйняття світу альтернативної реальності. Підсистему формування комерційного контенту реалізовано у вигляді комплексів контент-моніторингу зі збирання контенту з різних джерел даних, що забезпечують створення бази даних контенту відповідно до інформаційних потреб користувачів. У результаті збирання і первинного опрацювання контент приводять до єдиного формату та класифікують відповідно до визначеного рубрикатора. Контенту приписують дескриптори з ключовими словами. Це полегшує реалізацію процесу управління комерційним контентом. Завдання підсистеми управління комерційним контентом: формування та ротація баз даних; 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. Підсистема управління комерційним контентом реалізована хешуванням (підсистема подання генерує сторінку один раз. Надалі вона завантажується з кешу, який оновлюється автоматично після закінчення деякого терміну часу або внесення змін до певних розділів інформаційного ресурсу, або вручну за командою адміністратора, або формуванням інформаційних блоків (збереженням блоків на етапі редагування інформаційного ресурсу і збиранням сторінки з цих блоків у разі запиту відповідної сторінки користувачем). Підсистема супроводу контенту забезпечує: формування інформаційних портретів; виявлення тематичних сюжетів контенту; побудову таблиць взаємозв'язку контенту; розрахунок рейтингів контенту, виявлення, відстеження й кластеризацію нових подій у потоках контенту.

Висновки і перспективи подальших наукових розвідок

У статті вибрано підходи та концептуальні моделі послідовності опрацювання інформаційних ресурсів у СЕКК для проектування та створення таких систем, що дало змогу виділити основні закономірності переходу від процесів формування комерційного контенту до його супроводу. Побудовано формальний опис СЕКК для реалізації етапів життєвого циклу комерційного контенту, що дало можливість визначити основні компоненти системи електронної контент-комерції та їх зв'язки між собою. Розроблено формальні описи процесів опрацювання інформаційних ресурсів у СЕКК для реалізації формування, управління та супроводу комерційного контенту, що дало змогу створити узагальнену структуру СЕКК для спрощення етапів розроблення таких систем та реалізувати процеси формування, управління та супроводу комерційного контенту.

1. Береза А. Електронна комерція / А. Береза, 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. Большакова Е. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Е. Большакова, Д. Ландэ, А. Носков, Э. Клышинский, О. Пескова, Е. Ягунова. – М: МИЭМ, 2011. – 272 с.
4. Брайчевский С. Современные информационные потоки / С. Брайчевский, Д. Ландэ // Научно-техническая информация. – 2005. – № 11. – С. 21–33.
5. Клифтон Б. Google Analytics: профессиональный анализ посещаемости веб-сайтов / Б. Клифтон. – М: Вильямс, 2009. – 400 с.
6. Корнеев В. Базы данных. Интеллектуальная обработка информации / В. Корнеев, А. Гареев, С. Васютин, В. Райх. – М: Нолидж, 2000. – 352 с.
7. Ландэ Д. Основы моделирования и оценки электронных информационных потоков / Д. Ландэ, В. Фурашев, С. Брайчевский, О. Григорьев. – К: Інжиніринг, 2006. – 348 с.
8. Ландэ Д. Основы интеграции информационных потоков: монография / Д. Ландэ. – К: Інжиніринг, 2006. – 240 с.
9. Пасічник В. Математична лінгвістика / В. Висоцька, В. Пасічник, Ю. Щербина, Т. Шестакевич. – Л.: Новий Світ–2000, 2012. – 359 с.
10. Советов Б. Моделирование систем / Б. Советов, С. Яковлев. – М: ВШ, 1998.