

ПОБУДОВА ТА ФУНКЦІОНУВАННЯ СИСТЕМ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

© Висоцька В., Чирун Л., 2014

Проаналізовано основні проблеми електронної контент-комерції та функціональних сервісів опрацювання контенту. Запропонована модель дає можливість створити методи та засоби опрацювання інформаційних ресурсів у системах електронної контент-комерції та реалізувати процеси формування, управління та супроводу контенту.

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

General problems of electronic content commerce and functional services of content processing are analyzed in the article. Proposed model gives an opportunity to create a method and instrument of information resources processing in electronic commerce systems and to implement the processes of content formation, management and support.

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

Вступ. Загальна постановка проблеми

Розроблення і впровадження систем електронної контент-комерції є одним із важливих напрямів розвитку електронного бізнесу. Характерною рисою таких систем є можливість автоматичного опрацювання інформаційних ресурсів для збільшення обсягів продажу контенту постійному користувачу, активного залучення потенційних користувачів та розширення меж цільової аудиторії [1, 2]. Зокрема, принципи і технології електронної контент-комерції активно застосовують при створенні систем on-line/off-line продажу та аналізу/обміну/збереження контенту, Інтернет-магазину, cloud storage/computing [2]. Проблематика проектування, створення, впровадження та супроводу систем електронної контент-комерції є актуальною з огляду на такі фактори, як активний розвиток досліджень в галузі е-бізнесу та відсутність теоретичного обґрунтування стандартизованих методів і потреба в уніфікації програмних засобів опрацювання інформаційних ресурсів у таких системах [1–3]. Постійно з'являються нові підходи та способи вирішення цієї проблеми, але важливим питанням залишається невідповідність між відомими методами і засобами опрацювання інформаційних ресурсів та принципами побудови систем електронної контент-комерції. Відсутні загальні підходи до створення систем електронної контент-комерції та уніфіковані методи опрацювання інформаційних ресурсів у цих системах [2, 3]. Фахівці із проектування, реалізації та впровадження систем електронної контент-комерції мають справу із різними процесами опрацювання інформаційних ресурсів, які сприяють досягненню мети щодо збільшення обсягів продажу контенту постійному користувачу, активному залучення потенційних користувачів та розширення меж цільової аудиторії [1–5].

Формулювання мети

Метою роботи є побудова процесів опрацювання інформаційних ресурсів у системах електронної контент-комерції для визначення функціональних вимог до підсистем формування, управління та супроводу комерційного контенту. Відповідно це передбачає моделювання кожного з видів дій, які реалізують дані підсистеми з подальшим об'єднанням та

узагальненням їх в єдиній узгодженій моделі, яка описує вимоги до процесів функціонування систем електронної контент-комерції загалом.

Аналіз отриманих наукових результатів

Основними процесами опрацювання інформаційних ресурсів у системах електронної контент-комерції є формування, управління та супровід контенту (рис. 1), схема зв'язків яких є такою: *формування контенту* ® *управління контентом* ® *супровід контенту*.

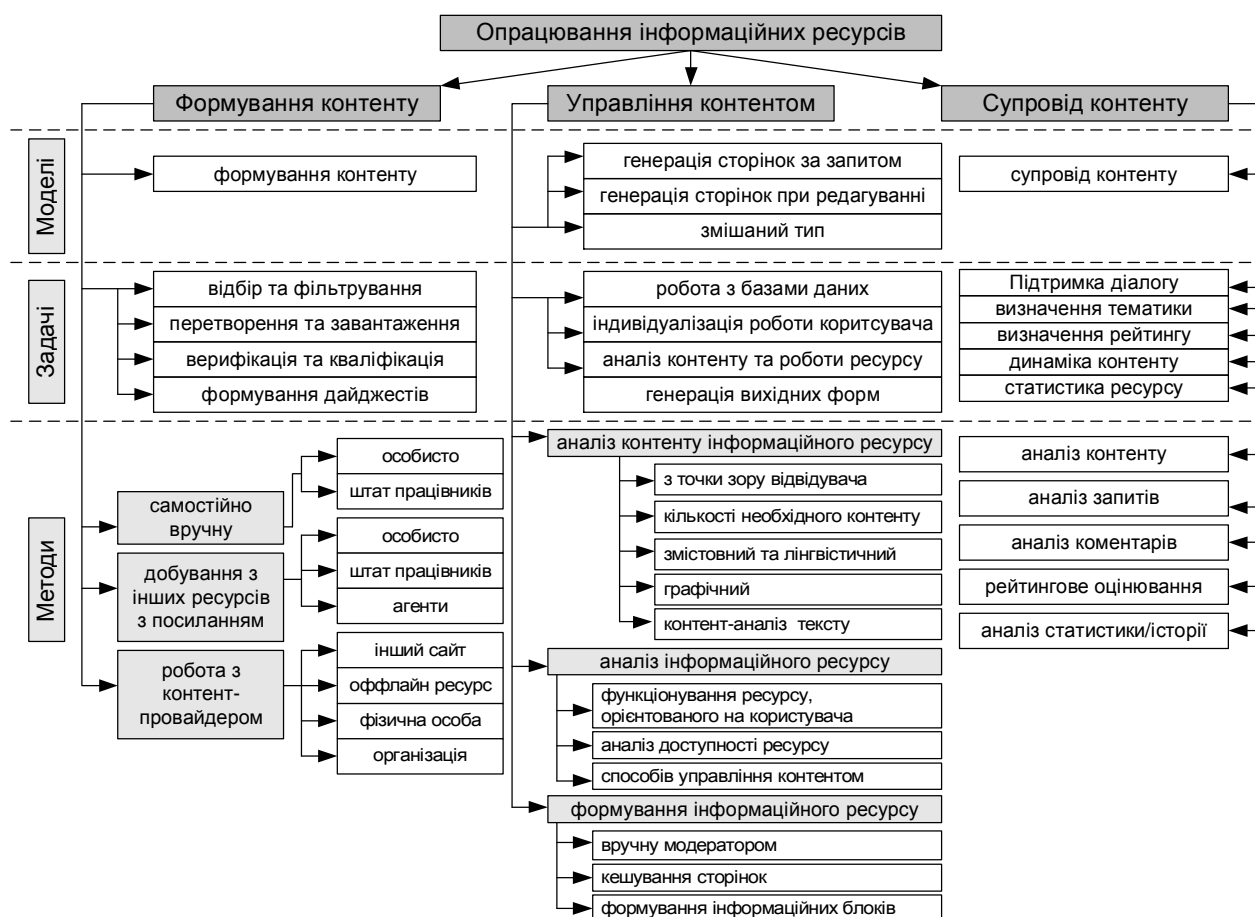


Рис. 1. Особливості процесу опрацювання інформаційних ресурсів у системах електронної контент-комерції

Функціонування системи електронної контент-комерції описується такими схемами зв'язків основних компонентів цієї системи:

1) для процесу формування інформаційного ресурсу системи: *контент* ® *формування комерційного контенту* ® *база даних комерційного контенту* ® *управління комерційним контентом* ® *інформаційний ресурс системи*;

2) для процесу формування відповіді на запит користувача: *запит користувача* ® *управління комерційним контентом* ® *інформаційний ресурс системи* ® *супровід комерційного контенту* ® *база даних користувачів системи*;

3) для процесу формування звіту роботи системи для модератора: *запит модератора* ® *супровід комерційного контенту* ® *база даних користувачів системи* ® *управління комерційним контентом* ® *звіт для модератора*;

4) для процесу модерації внутрішніх параметрів системи: *запит модератора* ® *формування комерційного контенту* ® *база правил* ® *супровід комерційного контенту* ® *база правил* ® *управління комерційним контентом* ® *результат*.

Модель системи електронної контент-комерції S описується як

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

де величина $X = \{x_1, x_2 \mathbf{K}, x_{n_X}\}$ – множина вхідних даних $x_i \in X$ з різних джерел інформації при $i = \overline{1, n_X}$; величина $Q = \{q_1, q_2 \mathbf{K}, q_{n_Q}\}$ – множина запитів $q_d \in Q$ користувачів при $d = \overline{1, n_Q}$; величина $Formation$ – оператор формування комерційного контенту; величина $H = \{h_1, h_2 \mathbf{K}, h_{n_H}\}$ – множина внутрішніх параметрів $h_k \in H$ системи S при $k = \overline{1, n_H}$; величина $C = \{c_1, c_2 \mathbf{K}, c_{n_C}\}$ – множина комерційного контенту $c_r \in C$ при $r = \overline{1, n_C}$; величина $V = \{v_1, v_2 \mathbf{K}, v_{n_V}\}$ – множина параметрів впливу $v_l \in V$ зовнішнього середовища на систему S при $l = \overline{1, n_V}$; величина $Management$ – оператор управління комерційними контентом; величина $Support$ – оператор супроводу комерційного контенту; величина $Z = \{z_1, z_2 \mathbf{K}, z_{n_Z}\}$ – множина сторінок $z_w \in Z$ інформаційного ресурсу в системі S при $w = \overline{1, n_Z}$; величина $T = \{t_1, t_2 \mathbf{K}, t_{n_T}\}$ – час $t_p \in T$ транзакції опрацювання інформаційного ресурсу в системі S при $p = \overline{1, n_T}$; величина $Y = \{y_1, y_2 \mathbf{K}, y_{n_Y}\}$ – множина статистичних даних $y_j \in Y$ роботи системи S при $j = \overline{1, n_Y}$.

Величина $y_j = \{a_1, a_2, \dots, a_8\}$, де a_1 – кількість відвідувань за період часу Δt , a_2 – середній час перебування на інформаційному ресурсі (хв:с) за період часу Δt , a_3 – показник відмовлень (%) за період часу Δt , a_4 – досягнута мета пошуку комерційного контенту, a_5 – динаміка життєвого циклу комерційного контенту (%), a_6 – загальна кількість переглядів сторінок за період часу Δt , a_7 – кількість переглядів сторінок за одне відвідування, a_8 – нові відвідування (%), a_9 – абсолютно унікальні відвідувачі, a_{10} – джерело трафіка (прямі переходи, переходи з пошукових систем, переходи з інших сайтів тощо) у % тощо [5].

Ріст обсягу контенту і швидкості його поширення породив поняття контентних потоків, аналіз яких вимагає використання нового інструментарію опрацювання інформаційних ресурсів у системах електронної контент-комерції для формування, управління та супроводу контенту (рис. 2).

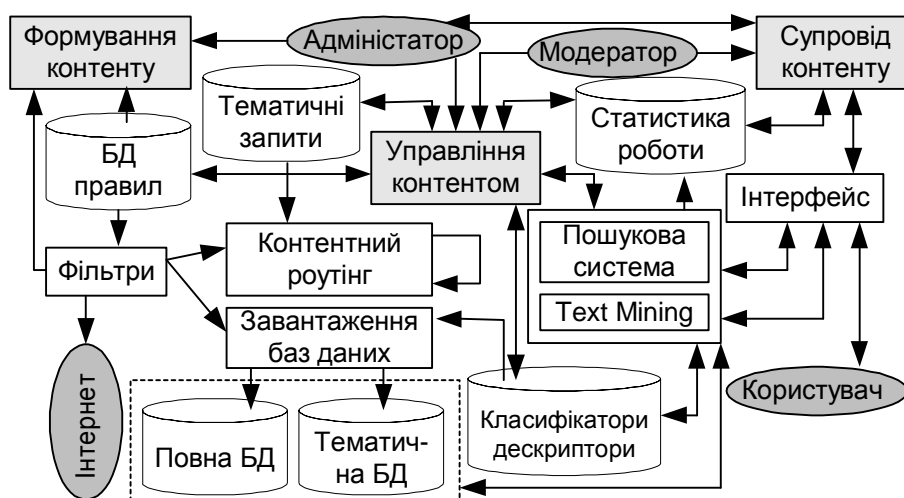


Рис. 2. Функціональна схема процесу опрацювання інформаційних ресурсів в системах електронної контент-комерції

Класичний математичний апарат та поширені інструментальні засоби не здатні адекватно відображати аналіз контентного масиву фіксованого розміру та навігацію в ньому [3, 4, 6–8]. Архітектура СЕКК має рівні ієрархії, які забезпечують незалежність збережених даних від програм, що їх використовують, та можливість розвитку системи без руйнування існуючих застосувань. На рис. 3, а подано схему процесу опрацювання інформаційних ресурсів в СЕКК, а на рис. 3, б – етапи формування комерційного контенту в системах електронної контент-комерції. СЕКК працюють на основі такої схеми: *контент-менеджер* → *підсистема редагування* → *база даних* → *ядро* → *користувач*.

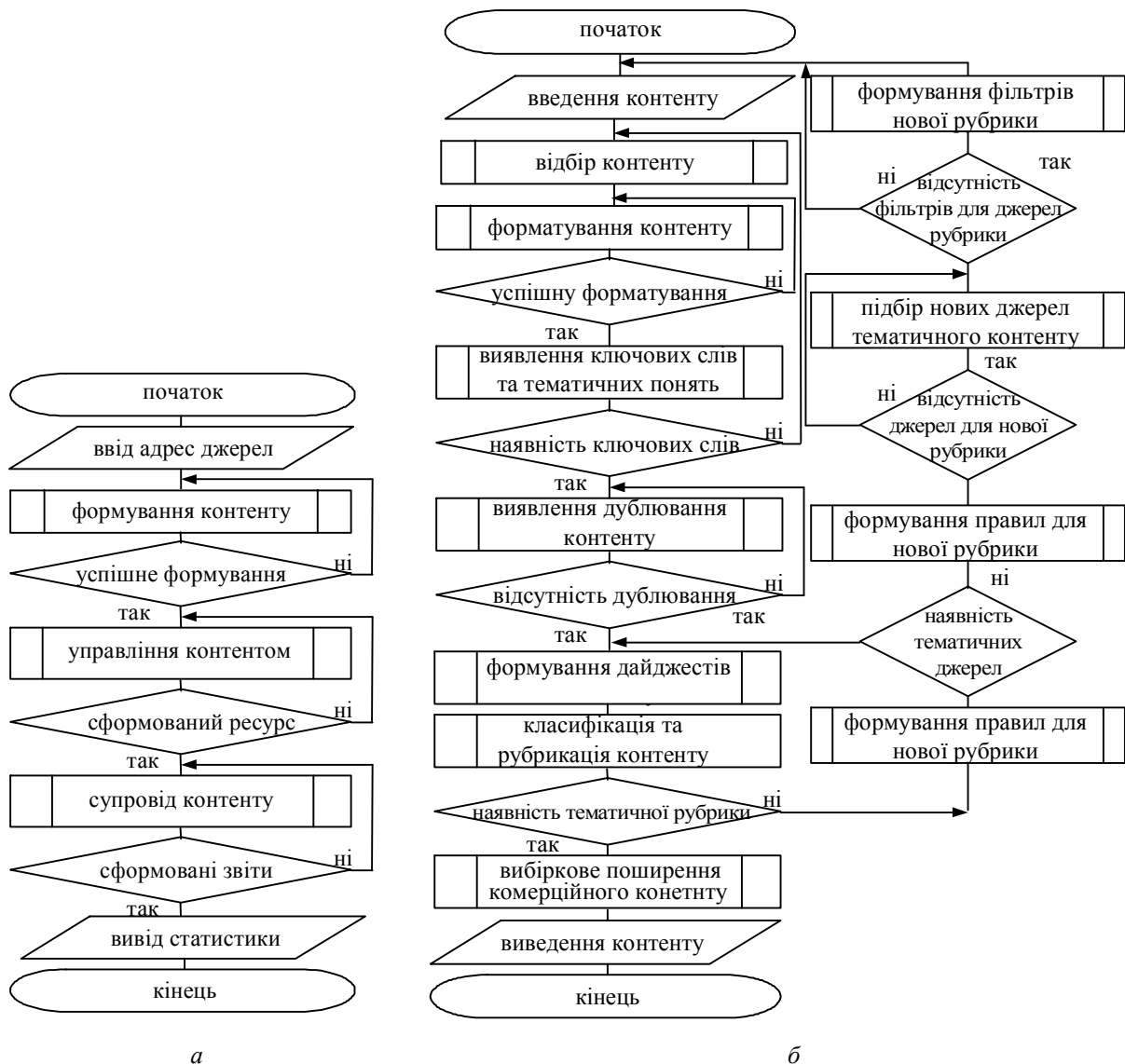


Рис. 3. Процес опрацювання (а) та формування інформаційних ресурсів (б)

Система електронної контент-комерції містить ядро адміністрування, підсистему авторизації/аутентифікації, менеджер шаблонів і менеджер контенту для розв'язування задач з позиції користувача (рис. 4).

При створенні ядра системи електронної контент-комерції використовують об'єктно-орієнтовану модель та абстрактні об'єкти зі властивостями/методами (рис. 5). Взаємодія з основними об'єктами ядра системи реалізують через інкапсуляцію. Для цього в класах реалізують інтерфейсні методи, призначені для маніпуляцій всередині об'єкта з даними/властивостями.

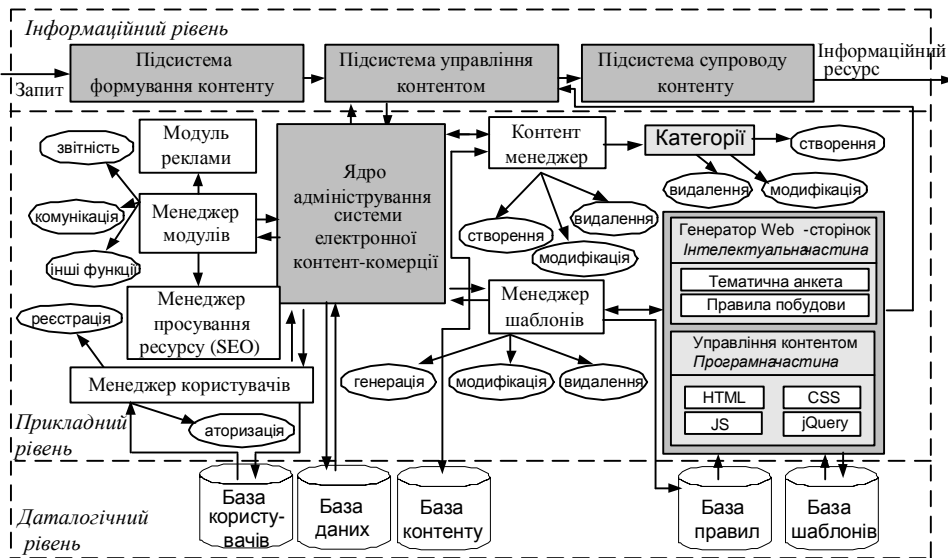


Рис. 4. Структура системи електронної контент-комерції

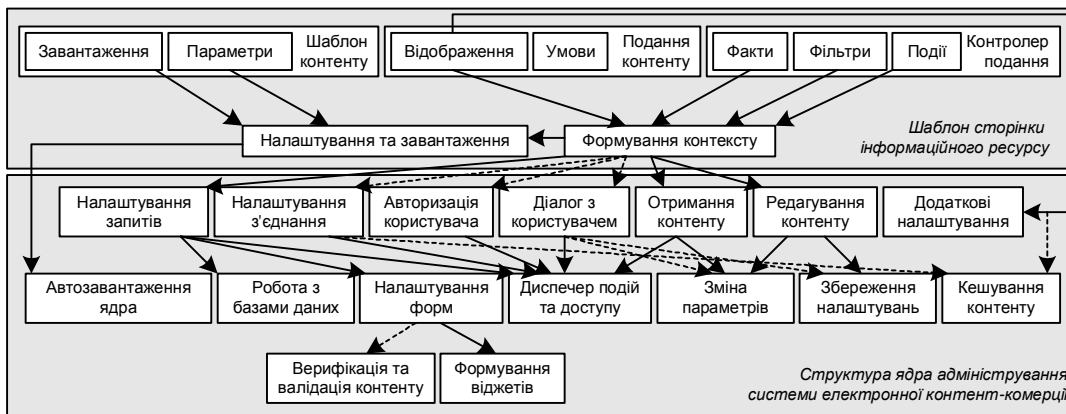


Рис. 5. Діаграма структури системи електронної контент-комерції

Процесами опрацювання інформаційних ресурсів у системах електронної контент-комерції є формування, управління та супровід контенту (рис. 6).

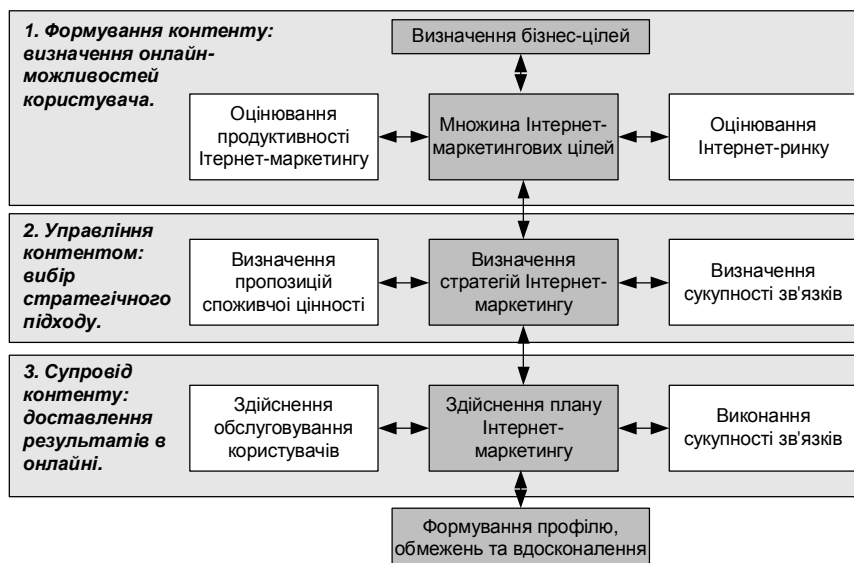


Рис. 6. Основні процеси опрацювання інформаційних ресурсів у системах електронної контент-комерції

Отримані дані враховують при створенні або оновленні інформаційного ресурсу та удосконаленні архітектури СЕКК. Раніше модератор самостійно шукав та опрацював необхідний контент: збір контенту з різних джерел даних, аналіз та фільтрація контенту, формування комерційного контенту як кінцевого продукту згідно із індивідуалізованими даними користувача системи. Реалізацією процесу опрацювання інформаційних ресурсів в СЕКК полегшують роботу модератора, автоматизуючи збирання контенту з джерел, аналіз та фільтрацію контенту.

Процес формування контенту складається із декількох етапів: *модератор* ® *створення контенту* ® *база даних* ® *систематизація контенту* ® *база даних* ® *поширення контенту* ® *редактор, або Інформаційний ресурс (джерело)* ® *збирання контенту* ® *база даних* ® *систематизація контенту* ® *база даних* ® *поширення контенту* ® *модератор* і реалізується у вигляді контент-моніторингу контенту та створення бази даних відповідно до інформаційних потреб споживачів (рис. 7). Після збирання і первинного опрацювання контент приводять до єдиного формату, класифікують відповідно до визначеного рубрикатора та йому приписують дескриптори, включаючи ключові слова. При застосуванні інтернет-маркетингу етапи систематизації контенту забезпечують постійне поповнення бази даних оперативними даними, ефективний одночасний доступ багатьох користувачів до бази даних, зручні засоби пошуку необхідного контенту.

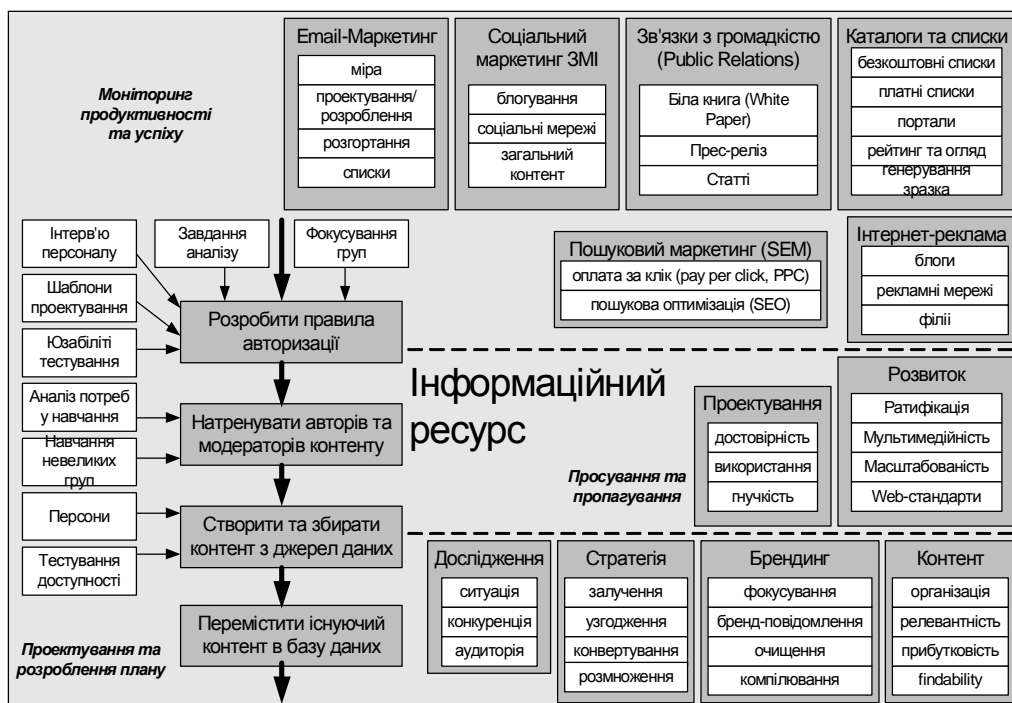


Рис. 7. Методи збирання/створення контенту та маркетингу

Поділ на етапи формування комерційного контенту в СЕКК підвищує ефективність при адмініструванні системи; забезпечує економію ресурсів, інтернет-трафіку та анонімність користувачів; автоматичне сканування джерел даних.

Управління контентом інформаційного ресурсу та відвідувачів (рис. 8–9), їх моделювання є процесом кількісного дослідження динаміки окремих тематичних напрямів та технічний аналіз інформаційного ресурсу.

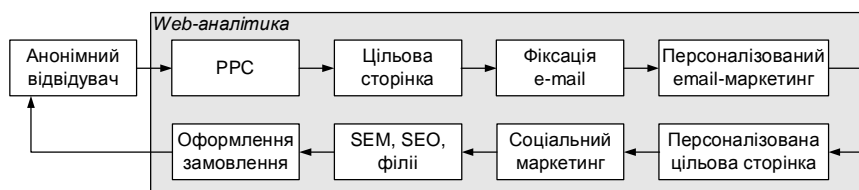


Рис. 8. Схема процесу аналізу відвідування інформаційного ресурсу

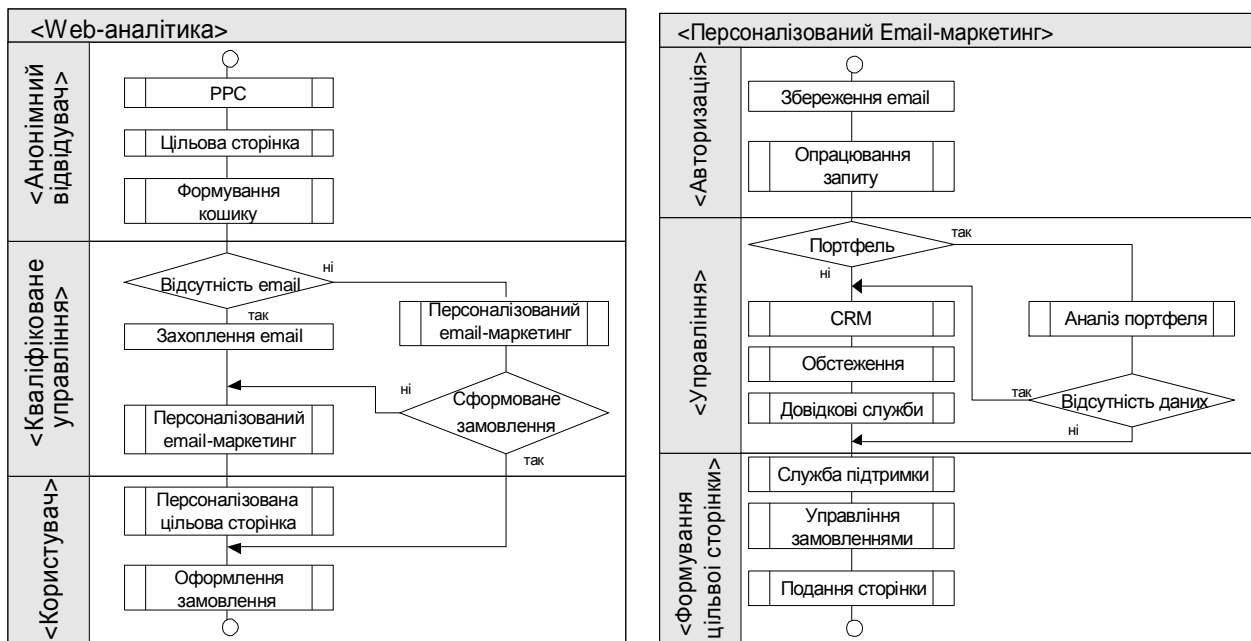


Рис. 9. Схема аналізу контенту відвідувачів (а) та інформаційного ресурсу (б)

Процес управління контентом складається із декількох етапів (рис. 9): користувач @ опрацювання контенту @ база даних @ аналіз контенту @ база даних @ подання контенту @ користувач.

Аналіз результатів управління контентом впливає на швидкість розвитку тематичних напрямів та контентного простору. Стійкі статичні зв'язки між контентом свідчать про кореляцію тематик, ефективність посилань на публікації джерел, більш ранні цитування, републікації тощо. Механізми, основні на узагальнених методах кластерного аналізу, виявляють контент у потоках, що формує навколо себе нові тематичні напрями. Кластерний аналіз, теорія фракталів і автотельних процесів при їхньому коректному застосуванні кількісно оцінюють ступінь зв'язку в тематичних контентних потоках. Оперативний аналіз контенту відвідувачів сприяє реалізації процесу управління контентом за допомогою генерації сторінок через інформаційні блоки, який поділяється на типи: тематична; по останніх зверненнях; комбінована.

Процес супроводу контенту – це оперативні етапи узагальнення, модерації та структурування комерційного контенту (рис. 10), тобто *Користувач @ структурування контенту @ база даних @ модерація контенту @ база даних @ узагальнення контенту @ модератор*. Із підсистемою супроводу контенту СЕКК має такі можливості: формування рейтингу комерційного контенту; формування інформаційного портрету постійного користувача СЕКК; аналіз характеристик (коментарі, відгуки, побажання тощо) на комерційний контент з боку користувача СЕКК; збирання, накопичення та опрацювання інформації про потреби кінцевого/потенційного користувача СЕКК та споживача контенту; формування інформаційного портрету контентних потоків.

Система електронної контент-комерції значно полегшує роботу модератора з формування, управління та супроводу комерційного контенту (рис. 11). Вона передбачає такі основні етапи опрацювання контенту інформаційних ресурсів:

1. Збирання контенту з різних джерел даних системою.
2. Фільтрація створеного і/або зібраного контенту системою.
3. Аналіз (визначення ключових слів, рубрикація, формування дайджестів) контенту системою.
4. Аналіз отриманих даних для концентрування уваги модератора на необхідній тематиці.

Підготовка даних для оперативної роботи з ними.

5. Аналіз потреб потенційних та постійних користувачів системи.

6. Формування кінцевого продукту згідно із проаналізованими даними та потребами потенційних/постійних користувачів.
7. Поповнення списку джерел даних згідно із результатами аналізу запитів користувачів.
8. Поповнення правил та словників фільтрів згідно із результатами аналізу контенту.
9. Виявлення нових тематичних рубрик потоку контенту.

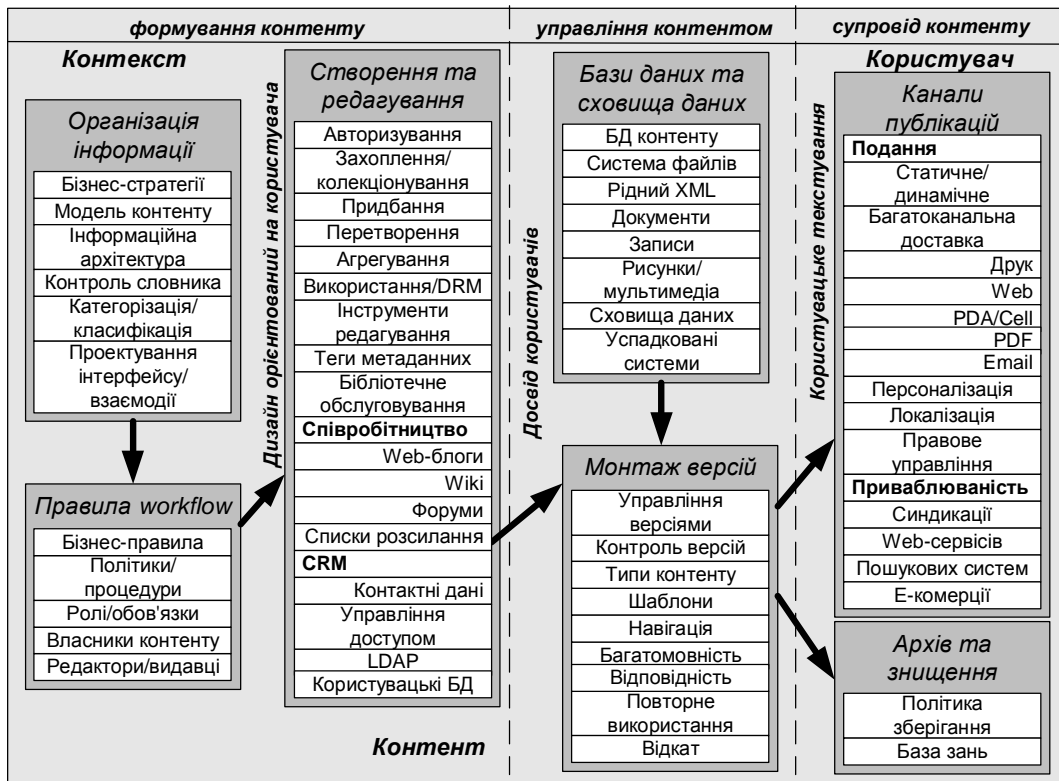


Рис. 10. Підсистеми формування, управління та супроводу контенту в системах електронної контент-комерції

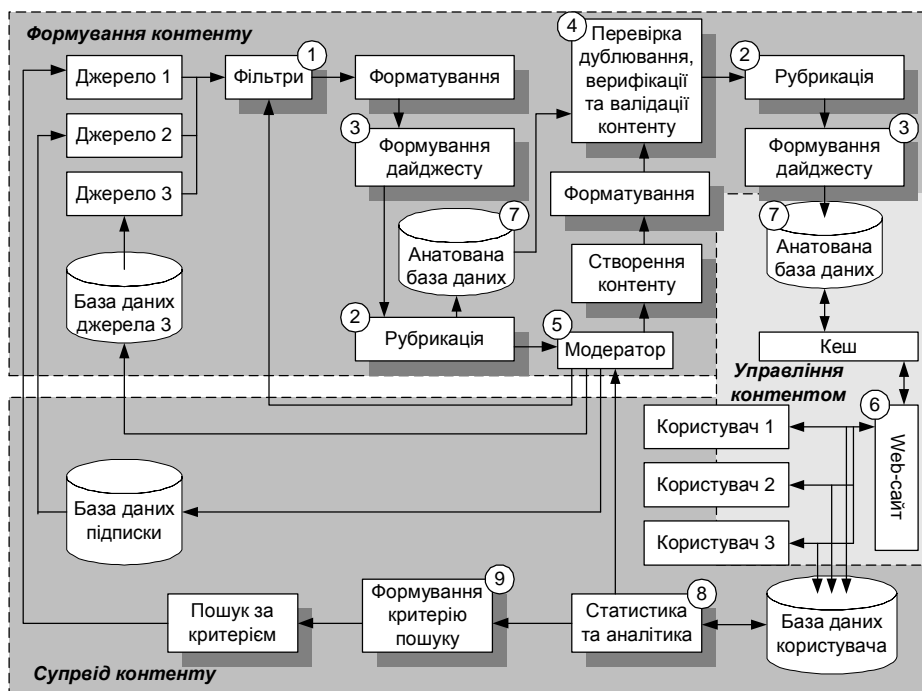


Рис. 11. Схема взаємодії підсистем опрацювання інформаційних ресурсів

На рис. 12 подано класифікацію розроблених методів опрацювання інформаційних ресурсів у СЕКК з детальним переліком реалізованих етапів формування, управління та супроводу комерційного контенту в цих системах.



Рис. 12. Методи опрацювання інформаційних ресурсів у системах електронної контент-комерції

Розроблені методи опрацювання інформаційних ресурсів в системах електронної контент-комерції дають можливість сформулювати вимоги до підпрограм опрацювання інформаційних ресурсів.

Контент-аналіз інформаційних ресурсів

Особливістю сучасності є постійний ріст темпів виробництва контенту. Цей процес є об'єктивним і позитивним, але виникла проблема: прогрес у галузі виробництва контенту призводить до пониження загального рівня інформованості потенційного користувача. Крім збільшення обсягів контенту до масштабів, яке призводить до неможливості його безпосереднього опрацювання та швидкості його поширення, виникає низка специфічних проблем (таблиця).

Основні негативні чинники формування комерційного контенту

Назва	Основна причина	Рішення
Інформаційний шум	Структурованість масивів контенту.	Фільтри, контент-моніторинг, аналіз сайту, контент-аналіз.
Паразитичний контент	Поява як додатків.	Фільтри, контент-моніторинг, контент-аналіз.
Нерелевантність контенту	Невідповідність потребам користувачів.	Створення анотованої бази даних, пошукових образів первинного контенту та їх кластеризація, контент-аналіз.
Дублювання контенту	Дублювання в джерелах.	Контент-аналіз, сканери і фільтри на базі статистики та критеріїв.
Навігація в потоці контенту	Швидкий ріст обсягу і поширення контенту.	Аналіз сайту, фільтри, контент-моніторинг, контент-аналіз.
Надмірність пошуку	Дублювання і нерелевантність.	Анотований пошук, контент-аналіз та реферування.

Негативні чинники у формуванні комерційного контенту ускладнюють процес пошуку необхідних даних при скануванні різних джерел інформації. Збільшення фізичного обсягу та змінна актуальності/динаміки контентних потоків (постійне систематичне та нерегулярне оновлення) призводить до дублювання, інформаційного шуму та надмірності результатів пошуку контенту. Охоплення та узагальнення великих динамічних потоків контенту, які неперервно генерують в Інтернет-джерелах, вимагає якісно нових методів/підходів пошуку – таких, як контент-моніторинг (рис. 13). Вхідною інформацією для контент-моніторингу є текст природною мовою як послідовність символів, вихідна інформація – це таблиці розділів, речень і лексем аналізованого тексту. Контент-моніторинг є програмним засобом автоматизації знаходження найважливіших складових у потоках контенту. Це змістовний аналіз потоків контенту з метою постійного отримання необхідних якісних/кількісних зрізів протягом наперед не визначеного проміжку часу.

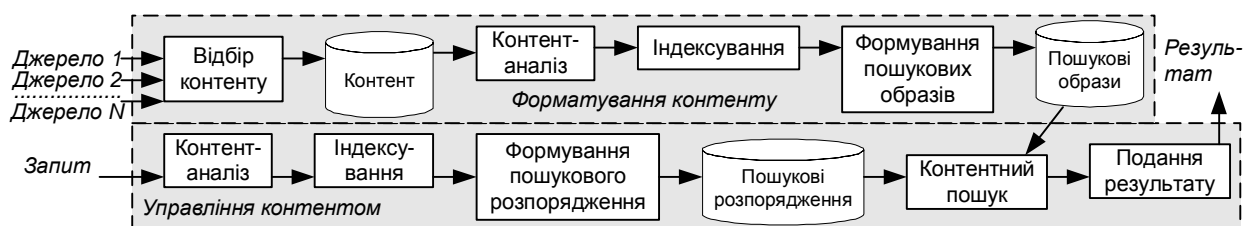


Рис. 13. Структурна схема процесу контент-моніторингу під час формування комерційного контенту

Складовою контент-моніторингу є контентний пошук та контент-аналіз тексту. Контент-аналіз призначений для пошуку контенту в масиві даних за змістовими лінгвістичними одиницями (алг. 1). Одиниця рахунку є кількісною мірою одиниці аналізу, що дає змогу реєструвати частоту (регулярність) появи ознаки категорії аналізу в тексті (кількість певних слів або їх поєднань, рядків, друкованих знаків, сторінок, абзаців, авторських аркушів, площа тексту тощо).

Алгоритм 1. Контент-аналіз текстового комерційного контенту.

Етап. 1. Визначення набору критеріїв для текстового комерційного контенту.

Крок 1. Формування набору критеріїв як тип джерела (форум, електронна пошта, Інтернет-газета, чат, Інтернет-журнал); тип контенту (стаття, електронний лист, баннер, коментарій); учасники комунікації (відправник, одержувач, реципієнт).

Крок 2. Визначення розміру (мінімальний обсяг або довжина), частоти появи, способу/місця розповсюдження та час появи контенту.

Крок 3. Фільтрування згідно сформованим набором критеріїв контентного потоку та зберігання ідентифікованого релевантного контенту.

Етап. 2. Контент-аналітичний відбір. Формування вибіркової сукупності контенту за критеріями обмеженої вибірки з більшого масиву.

Етап. 3. Виявлення змістовних одиниць аналізу текстового комерційного контенту (словосполучення, речення, тема, ідея, автор, персонаж, соціальна ситуація, частина тексту, кластеризованна за змістом категорії аналізу). Вимоги до вибору лінгвістичної одиниці аналізу: достатньо велика для інтерпретації значення; достатньо мала, щоб не інтерпретувати багато значень; легко ідентифікується; кількість одиниць достатньо велика для проведення вибірки.

Етап. 4. Виділення одиниць рахунку аналізу текстового контенту.

Крок 1. Якщо одиниці рахунку збігаються з одиницями аналізу, то знаходять частоти появи виділеної змістовної одиниці, інакше перейти до кроку 2.

Крок 2. Модератор на основі аналізованого контенту висуває одиниці рахунку, наприклад, протяжність текстів; площа тексту, заповнена змістовними одиницями; кількість рядків (абзаців, знаків, колонок тексту); розмір/вид файлу; кількість рисунків з певним змістом/сюжетом тощо.

Етап. 5. Порівняння змістовних одиниць аналізу з одиницями рахунку.

Крок 1. Класифікація за угрупованнями із оціненням ваги змістовних категорій в загальному обсязі тексту. Класифікатором є загальна таблиця, в яку зведені всі категорії аналізу і одиниці аналізу. Фіксують одиниці виразу категорій.

Крок 2. Статистичні розрахунки зрозумілості та атрактивності контенту.

Етап. 6. Розроблення інструменту контент-аналізу.

Крок 1. Створення закодованого протоколу контенту для компактності подання даних та швидкого порівняння результатів аналізу різного контенту.

Крок 2. Заповнення протоколу контенту властивостями (автор, час видання, обсяг тощо).

Крок 3. Заповнення протоколу комерційного контенту підсумками його аналізу (кількість вживання в ньому певних одиниць аналізу і висновки щодо категорій аналізу). Протокол кожного контенту заповнюється на основі підрахунку даних всіх його реєстраційних карток.

Етап. 7. Розроблення таблиці контент-аналізу. Тип таблиці визначають у вигляді системи скоординованих і субординованих категорій аналізу: кожна категорія (питання) передбачає ряд ознак (відповідей), за якими квантифікується зміст тексту.

Етап. 8. Розроблення кодувальної матриці контент-аналізу.

Крок 1. Якщо обсяг вибірки ≥ 100 одиниць, то аналізується набір матричних листів, інакше виконати крок 2.

Крок 2. Якщо вибірка < 100 одиниць, то проводиться двовимірний аналіз. В цьому випадку для кожного контенту формується кодувальна матриця.

Етап. 9. Проведення аналізу тексту згідно із створеними кодувальними матрицями.

Етап. 10. Інтерпретація результатів. Виявляють і оцінюють характеристики контенту на основі статистичного набору підрахованих коефіцієнтів за певний період часу на визначену категорію. Охоплює всі здобуті фрагменти тексту, висновки спираються не на частину результатів, а враховуються всі без винятку.

Застосування контент-аналізу при моніторингу інтернет-джерел даних дає змогу автоматизувати процес знаходження найважливіших складових в потоках контенту при відборі даних з цих джерел. Це усуває дублювання контенту, інформаційний шум, паразитичний контент, надмірність результатів пошуку тощо. Цей метод застосовують в подальших етапах формування комерційного контенту для отримання більш точного релевантного результату – створення унікального комерційного контенту, який користується попитом серед користувачів СЕKK.

Висновки і перспективи подальших наукових розвідок

З позиції системного підходу застосовано принципи опрацювання інформаційних ресурсів у системах електронної контент-комерції для реалізації життєвого циклу контенту, що дало змогу розробити методи формування, управління та супроводу комерційного контенту. Розроблено комплексний метод формування комерційного контенту для скорочення часу та зменшення ресурсів виробництва контенту, що дає можливість створити засоби опрацювання інформаційних ресурсів та реалізувати підсистему автоматичного формування контенту. Створено оперативний метод управління комерційним контентом для скорочення часу та зменшення ресурсів продажу контенту, що дає можливість реалізувати підсистему управління комерційним контентом. Реалізовано комплексний метод супроводу контенту для скорочення часу та зменшення ресурсів аналізу цільової аудиторії системи електронної контент-комерції, що дає можливість розробити підсистему супроводу 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. Клифтон Б. Google Analytics: профессиональный анализ посещаемости веб-сайтов / Б. Клифтон. – М: Вильямс, 2009. – 400 с.
5. Корнеев В. Базы данных. Интеллектуальная обработка информации / В. Корнеев, А. Гареев, С. Васютин, В. Райх. – М: Нолидж, 2000. – 352 с.