

ФОРМАЛЬНА МОДЕЛЬ ОПРАЦЮВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ В СИСТЕМАХ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

© Висоцька В. А., Чирун Л. В., 2015

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

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

The main problems of electronic content commerce are analyzed and functional services of commercial content management are explored. The proposed method gives an opportunity to create an instrument of information resources processing in electronic commerce systems. It also enables the implementation of the commercial content management subsystem.

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

Вступ. Загальна постановка проблеми

Документована інформація, підготовлена відповідно до потреб користувачів, є комерційним контентом. Сьогодні електронна комерція є об'єктивною реальністю та перспективним бізнес-процесом. Інтернет є бізнес-середовищем, а комерційний контент є товаром з найбільшим попитом і продажами в ньому та основним об'єктом процесів електронної контент-комерції. Комерційний контент можна відразу замовити, оформити, оплатити та отримати on-line як товар. Через Інтернет продають весь спектр комерційного контенту – наукові та публіцистичні статті, музику, книги, фільми, фото, програмне забезпечення тощо. Відомими корпораціями, які реалізують електронну контент-комерцію, є Google через Play Market, Apple – Apple Store, Amazon – Amazon.com. Більшість рішень та досліджень проведено на рівні конкретних проектів. Системи електронної контент-комерції (СЕКК) побудовані за закритим принципом як разові проекти. Сучасні СЕКК орієнтовані на реалізацію комерційного контенту, який створений за межами системи. Проектування, створення, впровадження та супровід СЕКК неможливі без використання сучасних методів та інформаційних технологій формування, управління та супроводу контенту.

Метою проекту є реалізація уніфікованих методів та апробація програмних засобів опрацювання інформаційних ресурсів в СЕКК. Розроблення загальної архітектури СЕКК сприяє узагальненню методики опрацювання інформаційних ресурсів в СЕКК через етапи формування, управління та супроводу комерційного контенту для скорочення часу при побудові типових систем електронного бізнесу. Впровадження СЕКК сприяє скороченню часу при виробництві власного комерційного контенту, аналізі зовнішнього комерційного контенту з інших джерел, аналізі динаміки життєвого циклу комерційного контенту, аналізі статистики функціонування СЕКК, аналізі статистики діяльності користувачів інформаційних ресурсів у СЕКК, збільшенні цільової аудиторії інформаційних ресурсів та розширенні функціональних можливостей цих СЕКК.

Призначенням СЕКК є формування, управління та супровід комерційного контенту на засадах опрацювання інформаційних ресурсів. СЕКК призначена для створення загальних функціональних вимог та стандартизованих специфікацій з розроблення СЕКК шляхом оптимізації етапів процесу опрацювання інформаційних ресурсів в аналогічних системах.

Розроблення технології опрацювання інформаційних ресурсів є актуальним з огляду на такі фактори, як недостатність теоретичного обґрунтування методів опрацювання потоків комерційного контенту і потреба в уніфікації програмних засобів опрацювання інформаційних ресурсів у СЕКК. Практичний чинник опрацювання інформаційних ресурсів у СЕКК пов'язаний з розв'язанням задач формування, управління та супроводу зростаючих обсягів комерційного контенту в Інтернет, активним розвитком електронного бізнесу, швидкими темпами поширення доступності до Інтернет, розширенням набору інформаційних товарів та послуг, зростанням попиту на комерційний контент. Принципи та ІТ електронної контент-комерції застосовують при створенні Інтернет-магазинів (продаж eBooks, Software, video, music, movies, picture), систем on-line (газети, журнали, дистанційне навчання, видавництва) та off-line продаж контенту (copywriting services, Marketing Services Shop, RSS Subscription Extension), cloud storage та cloud computing. В цьому напрямі працюють провідні світові виробники засобів опрацювання інформаційних ресурсів, такі як Apple, Google, Intel, Microsoft, Amazon.

Теоретичний чинник опрацювання інформаційних ресурсів у СЕКК пов'язаний із розробленням ІТ опрацювання комерційного контенту. У наукових роботах Д. Ланде, В. Фурашева, С. Брайчевского, О. Григор'єва досліджено та розвинуто математичні моделі опрацювання електронних інформаційних потоків. Г. Зіпф (G. Zipf) запропонував емпіричну закономірність розподілу частоти слів природної мови в текстовому контенті для його аналізу. У роботах Б. Бойка (B. Boiko), С. Макківер (S. McKeever), Е. Роклі (A. Rockley) розроблені моделі життєвого циклу контенту. Методологію контент-аналізу для опрацювання текстових масивів даних започаткували та розвинули М. Вебер (M. Weber), Ж. Кайзер (J. Kaiser), Б. Гласер (B. Glaser), А. Стросс (A. Strauss), Г. Лассуел (H. Lasswell), О. Холсті (O. Holsti), В. Іванов, М. Сорока, А. Федорчук. У роботах В. Корнеєва, А.Ф. Гарєєва, С.В. Васюгіна, В.В. Райха запропоновані методи інтелектуального опрацювання текстової інформації. Корпорації EMC, IBM, Microsoft Alfresco, Open Text, Oracle і SAP розробили специфікації Content Management Interoperability Services на інтерфейс Web-сервісів, для забезпечення взаємодії систем управління контентом електронного бізнесу. З наукового погляду цей сегмент ІТ є мало дослідженим. Кожний окремий проект реалізують практично з початку, фактично на основі своїх ідей та рішень. У літературі надзвичайно мало висвітлені істотні теоретичні обґрунтування, дослідження, висновки, рекомендації, узагальнення для проектування СЕКК та опрацювання інформаційних ресурсів у таких системах. Виникла потреба в аналізі, узагальненні та обґрунтуванні наявних підходів реалізації електронної комерції та побудови СЕКК. Актуальною є задача створення комплексу технологічних засобів на основі теоретичного обґрунтування методів, моделей і принципів опрацювання інформаційних ресурсів у СЕКК, побудованих за принципом відкритих систем, які дозволяють керувати процесом збільшення обсягів реалізації комерційного контенту.

Аналіз останніх досліджень та публікацій

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

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

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

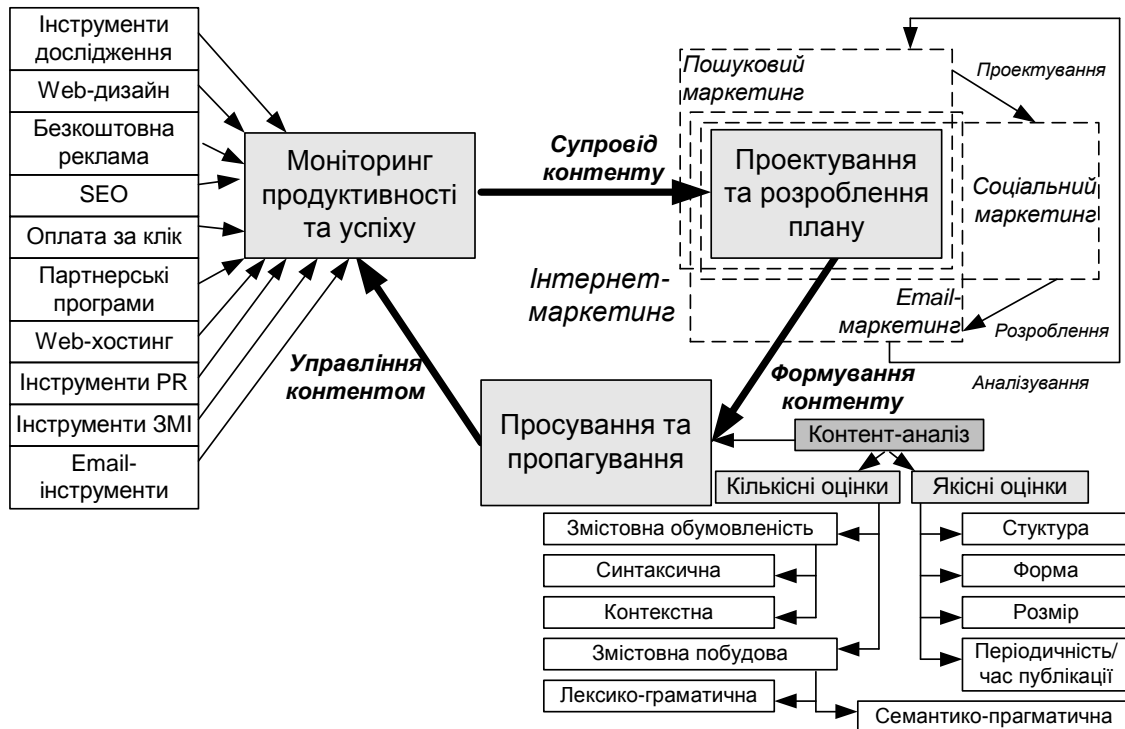


Рис. 1. Інтернет-маркетинг для систем електронної контент-комерції

Для розроблення СЕКК використовують CMS [1], якщо вона відповідає визначеному набору вимог (рис. 2).

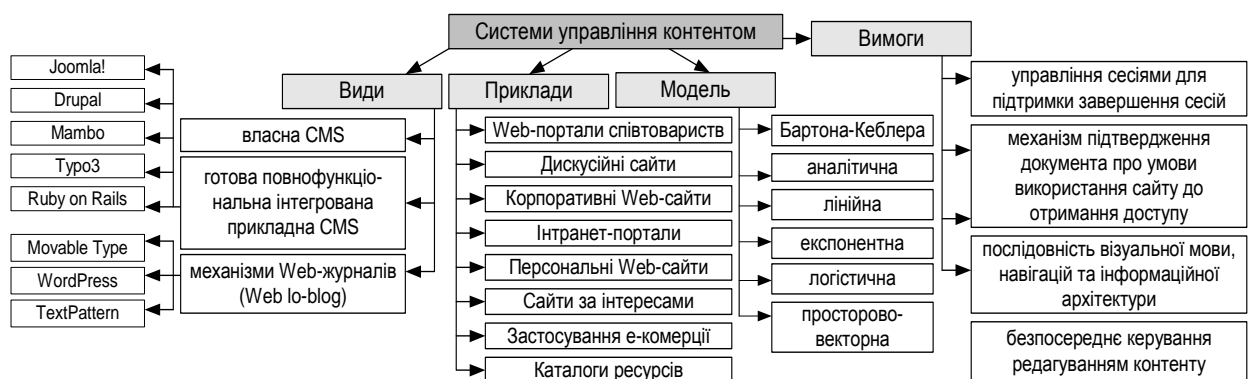


Рис. 2. Основні вимоги до систем управління контентом для побудови СЕКК

CMS не розв'язують основних задач опрацювання ресурсів – формування та супроводу контенту [1]. Основний недолік CMS полягає у відсутності зв'язку між вхідними даними, контентом та вихідними характеристиками функціонування системи (рис. 3) [1]. CMS дозволяють управляти контентом через зручні інструменти зберігання і публікації. Цінність контенту визначає його привабливість для споживача. Інтеграція контенту робить Web-ресурс привабливим, а

інтеграція застосувань – корисним [1]. CMS – це інструмент (табл. 1) для моделювання розгалужених структур Web-ресурсів в CEKK і управління їх контентом [1] без наявності спеціальних технічних навичок. CMS використовують для генерування контенту для Web-ресурсів з такими проблемами, як динамічне збирання, кешування та безпека контенту [1]. CMS забезпечує контроль доступу і внесення змін до Web-ресурсу та покликана максимально спростити управління Web-ресурсом при гнучкості налаштувань і контролі.

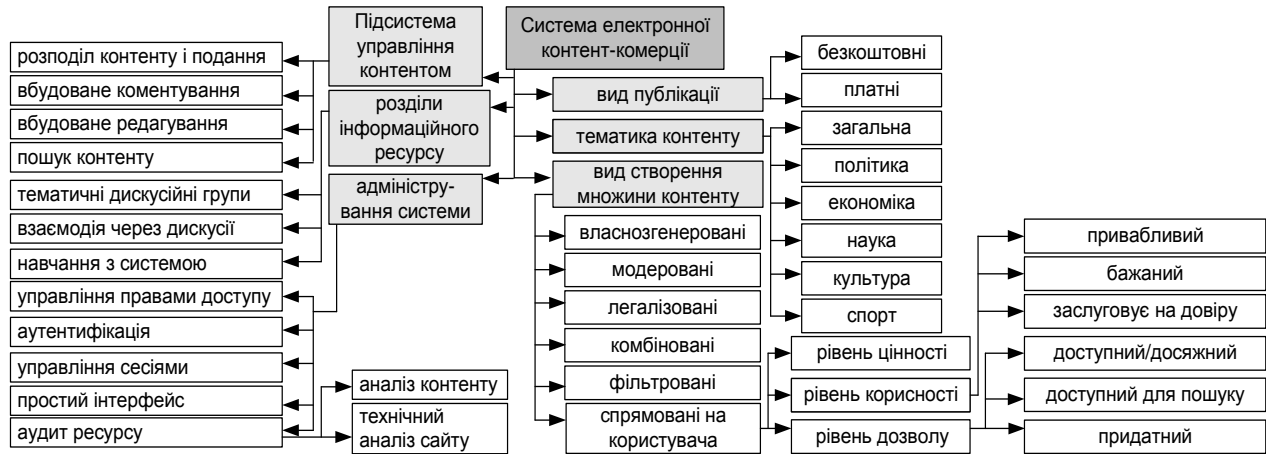


Рис. 3. Основні компоненти систем управління контентом для побудови CEKK

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

Таблиця 1

Характеристика основних CMS для розроблення CEKK, розроблено за [1]

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

Виділення проблем

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

Аналіз отриманих наукових результатів

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

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

комерційного контенту в СЕКК виконують за допомогою лінгвостатистичних методів знаходження загальних термів, ланцюжки яких утворюють словесні сигнатури комерційного контенту (текст є унікальним при коефіцієнті унікальності $\geq 80\%$). Задачу сканування комерційного контенту та приведення до загального формату в XML описано оператором α_3 вигляду

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

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

Опрацювання множини контенту C_2 для виявлення значущих ключових слів побудоване на принципі знаходження ключових слів за змістом (термами), ґрунтується на законі Зіпфа і зводиться до вибору слів із середньою частотою появи (найбільш вживані слова ігноруються за допомогою стоп-словника, а рідкісні слова з текстів повідомлень не враховують). Виявлення ключових слів та понять з використанням словників визначається оператором $\alpha_4(C_2, U_K)$ вигляду $C_3 = \alpha_4(\alpha_3(\alpha_2(C_0, U_B), U_{FR}), U_K)$ при $U_K = \{U_{K1}, U_{K2}, U_{K3}, U_{K4}\}$, де U_K – колекція умов виявлення ключових слів та понять у тексті, U_{K1} – множина всіх термів (термом є основа іменника, іменник, словосполученням іменників або прикметника з іменником), U_{K2} – множина частот вживання терму в тексті комерційного контенту, U_{K3} – множина коефіцієнтів вживання термів з врахуванням кількості знаків без пробілів (при 2 000 – 3 000 знаків частота ключових слів в межах 4–6 %, до 2 000 знаків – 6–8 %, понад 3 000 знаків – 2–4 %), U_{K4} – множина термів, які відповідають умовам належності до ключових слів.

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

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

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

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

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

де $D_{C_6} = \{c_{faset} \in C_{faset} \mid \forall c_{6i} \in C_6 : c_6 \leq c_{faset}\}$ – множина максимумів для C_6 при значеннях потужностей множин $\rho_{faset} = |C_{faset}|$, $\rho_6 = |C_6|$, $\rho_{faset} \geq \rho_6$; $D_{C_6 \cup C_7} = \{c_{faset} \in C_{faset} \mid \forall c_i \in (C_6 \cup C_7) : c_i \leq c_{faset}\}$ – множина максимумів для $C_6 \cup C_7$ при $C_{faset} \geq (C_6 \cup C_7)$. Показники прибутковості I_p та зростання попиту I_g на комерційний контент використовують для визначення його актуальності та розраховують відповідно як

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

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

Похідний список обраних правил визначає модератор. Часто вживаним списком контенту є список, формування якого перевищує мінімальний рівень. При встановленні цієї умови формується велика кількість списків часто вживаного контенту відповідно до запитів. Результат обмежується при вибірці операторами

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

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

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

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

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

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

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

Вагу блока визначають як суму коефіцієнтів ваг комерційного контенту:

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

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

2. Управління контентом з генерацією сторінок при редагуванні Web-ресурсу модератором

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

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

3. Змішаний тип управління комерційним контентом подано як

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

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

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

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

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

Колекцію вихідних статистичних даних Y подано як

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

Колекція Y описує процес функціонування СЕКК з основними процесами опрацювання Web-ресурсів як формування, управління та супровід контенту. Аналіз статистики роботи СЕКК проводять згідно з аналізом реакцій на цю систему постійного/потенційного користувача (відвідування, запити, пошук по ключових словах тощо). Це сприяє ефективному проведенню аналізу реакції цільової та потенційної аудиторії. Для детального аналізу функціонування СЕКК типу Інтернет-газета та Інтернет-журнал було розроблено та впроваджено вісім різних систем відповідно з підтриманням для кожної з них різною кількістю етапів життєвого циклу контенту. Тобто для різних реалізованих систем були розроблені не всі компоненти або взагалі не були розроблені підсистеми опрацювання Web-ресурсів як формування, управління та супровід контенту. Для збереження статистики функціонування системи електронної контент-комерції, результатів аналізу цієї статистики та конфігурацій інформаційного ресурсу використовуються окремі файли конфігурації (рис. 4).

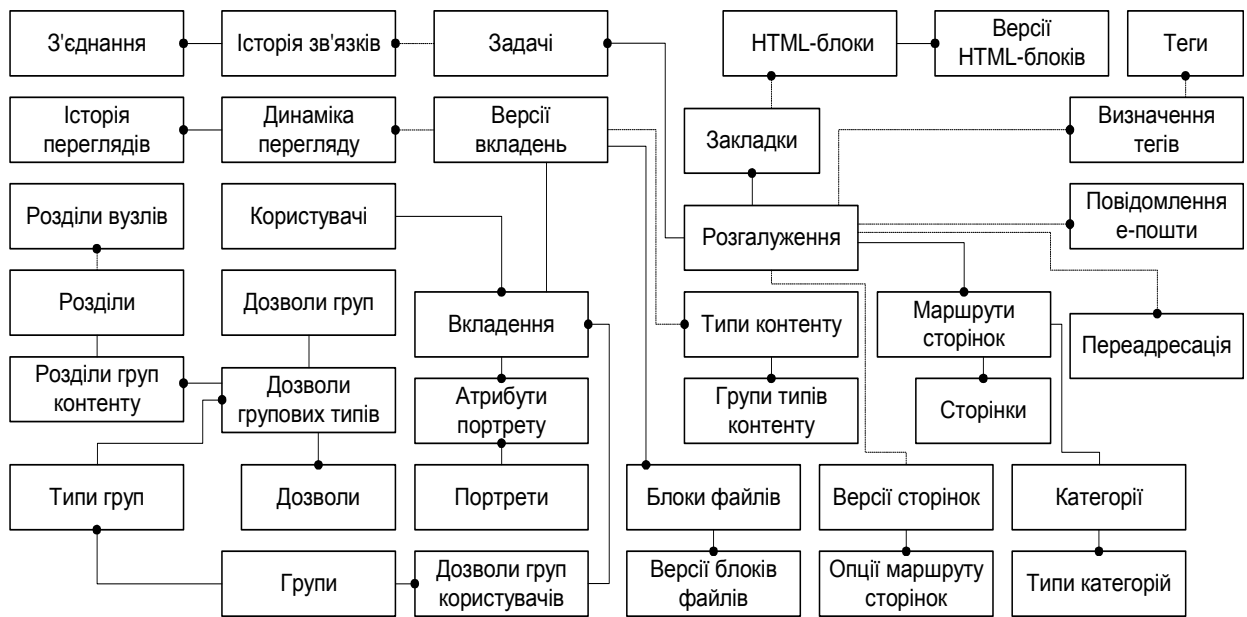


Рис. 4. Структура інформаційного ресурсу систем електронної контент-комерції

Для супроводу комерційного контенту аналізують статистику та динаміку функціонування інформаційного ресурсу за певний період, наприклад, дані про відвідування, пошукові трафіки, активність запитів, кількість переглядів сторінок, кількість і час читання контенту та тематик контенту, популярність авторів та рубрик контенту, активність відвідувань за регіонами та мовами, категорії з вибору налаштувань, завантажені файли, налаштування стосовно типів контенту, структури секції груп контенту, повідомлення електронної пошти, коментарі та відгуки, кількість і типи груп контенту, групи контенту та користувачів, Web-шаблони, наповнення Web-шаблонів, відвідування Web-сторінки, структуру інформаційного ресурсу, дані користувачів та їхні налаштування тощо.

В інформаційний ресурс вмонтовується аплікація СЕКК з метою формування рейтингу комерційного контенту та кращого розуміння потреб кінцевого користувача через аналіз контенту типу питання/відповіді, рейтинги, відгуки, статті тощо. Через компоненту Черга типу FIFO запити користувачів потрапляють в аплікацію фільтрування контенту за наперед визначеними правилами індивідуально для кожного клієнта (табл. 2). Індивідуальний кабінет клієнта надає можливість перегляду контенту та аналізу статистики. Компонента Індекссування даних призначена для швидкого пошуку комерційного контенту.

Таблиця 2

Правила супроводу контенту в системі електронної контент-комерції

Аналіз цензури	База даних забороненої 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.	Визначається клієнтом за бажанням.

Аналізом, вибіркою, побудовою та модерацією статистичних даних керує підсистема супроводу контенту за правилами (табл. 4.4) та алгоритмами модерації/узагальнення комерційного контенту.

**Правила модерції комерційного контенту
в системі електронної контент-комерції із самонавчанням**

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

Діаграма класів на рис. 5 показує блоки об'єктно-орієнтованої підсистеми супроводу комерційного контенту. На діаграмі класів зображено статичне подання частини моделі супроводу контенту, що описує атрибути і поведінку підсистеми.

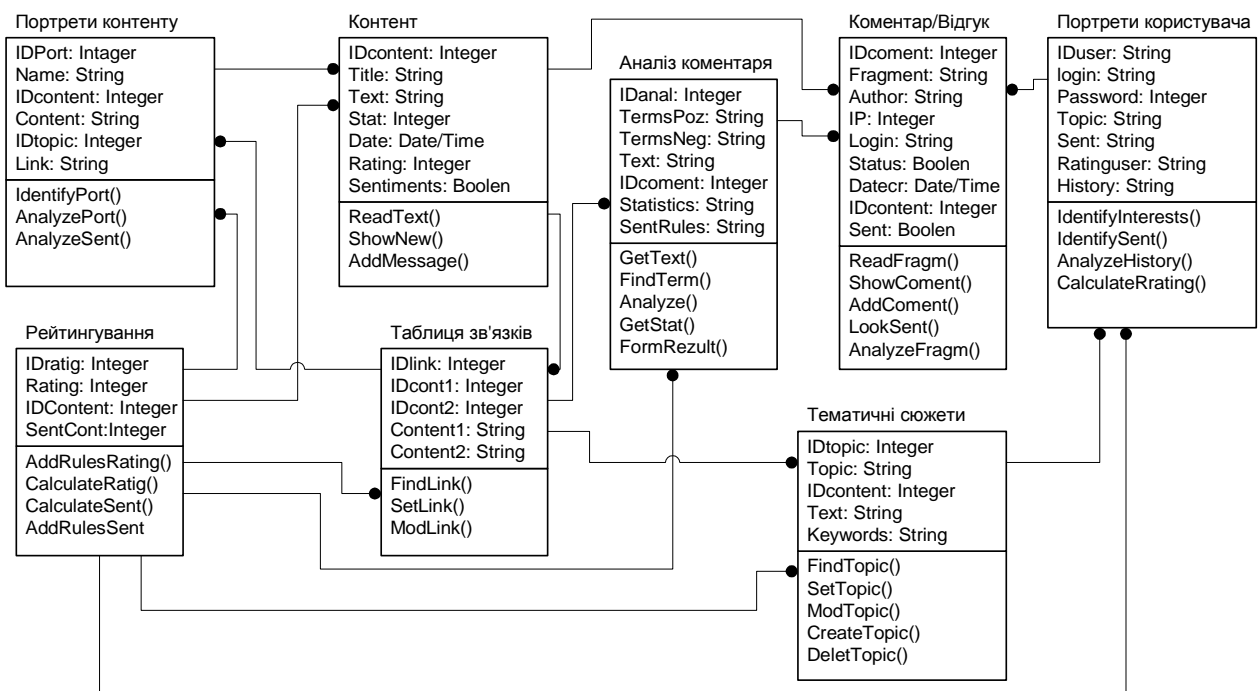


Рис. 5. Діаграма класів для процесу супроводу комерційного контенту
в системах електронної контент-комерції

Аналіз динаміки контенту сприяє формуванню висновків щодо можливості впливу на ріст цільової аудиторії, тенденцій збільшення з часом річного обсягу відвідувань інформаційного ресурсу. Досліджувати можна динаміку зміни світового, національного, регіонального, галузевого, проблемно тематичного потоку контенту або динаміку росту однорідної за будь-якою ознакою частини потоку. Базою для отримання даних про динаміку зміни та росту цільової аудиторії є вторинні джерела, наприклад, GoogleAnalytics. Для покращення точності результатів аналізу динаміки контенту реалізують такі етапи: аналіз інформаційних потреб користувачів та контенту, аналіз тематичних сюжетів контенту, формування таблиць взаємозв'язку контенту, аналіз тональності коментарів, рейтингу/історії контенту та статистики інформаційного ресурсу.

Висновки і перспективи подальших наукових розвідок

Стаття присвячена питанням розроблення методів та програмних засобів опрацювання інформаційних ресурсів у системах електронної контент-комерції. Вдосконалена класифікація систем електронної контент-комерції. Розроблено формальний опис системи електронної контент-комерції та процесу опрацювання комерційного контенту. Вдосконалена структура систем електронної контент-комерції. Сформульовано новий підхід застосування та впровадження бізнес-процесів для побудови систем електронної контент-комерції. Розроблено методи формування управління та технічного супроводу комерційного контенту. Розроблено програмні засоби формування, управління та технічного супроводу комерційного контенту. Розроблено методи проектування і реалізації системи електронної контент-комерції на прикладах Інтернет-газети та Інтернет-журналу, які відображають результати теоретичних досліджень. З позиції системного підходу проведено застосування принципів опрацювання інформаційних ресурсів у системах електронної контент-комерції для реалізації життєвого циклу комерційного контенту, що дало змогу розробити методи формування, управління та супроводу комерційного контенту. Проаналізовано основні проблеми електронної комерції та функціональних сервісів управління комерційним контентом. Розроблено метод формування комерційного контенту для скорочення часу та зменшення ресурсів виробництва контенту, що дає можливість створити засоби для опрацювання інформаційних ресурсів та реалізувати підсистему автоматичного формування комерційного контенту. Створено метод управління комерційним контентом для скорочення часу та зменшення ресурсів продажу контенту, що дає можливість реалізувати підсистему управління комерційним контентом. Реалізовано метод супроводу комерційного контенту для скорочення часу та зменшення ресурсів аналізу цільової аудиторії системи електронної контент-комерції, що дає можливість розробити підсистему супроводу комерційного контенту.

1. Берко А. Системи електронної контент-комерції / А. Берко, В. Висоцька, В. Пасічник. – Львів: НУЛП, 2009. – 612 с. 2. Клифтон Б. Google Analytics: профессиональный анализ посещаемости веб-сайтов / Б. Клифтон. – М.: Вильямс, 2009. – 400 с. 3. Основы моделирования и оценки электронных информационных потоков / Д. Ландэ, В. Фурашев, С. Брайчевский, О. Григорьев. – К.: Инжиниринг, 2006. – 348 с. 4. Основы интеграции информационных потоков: монография / Д. Ландэ. – К.: Инжиниринг, 2006. – 240 с.