

## МОДЕЛЮВАННЯ ПРОЦЕСІВ ОПРАЦЮВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

© Берко А., Висоцька В., Чирун Л., 2013

**Проаналізовано основні проблеми електронної контент-комерції та функціональних сервісів опрацювання контенту. Запропонований метод дає можливість створити засоби опрацювання інформаційних ресурсів у системах електронної контент-комерції та реалізувати підсистеми 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 method gives an opportunity to create an instrument of information resources processing in electronic commerce systems and to implement the subsystem 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**

Одним з основних завдань систем електронної контент-комерції (СЕКК) є полегшення роботи модераторів, авторів, аналітиків та адміністраторів контенту, з одного боку, та підвищення функціональності інтернет-ресурсів для їх користувачів – з іншого боку. СЕКК підбирають коло актуальних питань у вигляді множини контенту з різноманітних зовнішніх джерел для модераторів та авторів згідно з їхнім рейтингом через підсистему формування комерційного контенту. Автор створює власний комерційний контент, аналізуючи підібраний СЕКК з різних джерел актуальний контент. Модератор створює, якщо необхідно, нові правила для фільтрації контенту з різних джерел та поновлює адреси інших джерел у підсистемі формування комерційного контенту. Аналітик аналізує діяльність цільової аудиторії та функціонування СЕКК для розроблення нових правил аналізу статистики та динаміки етапів життєвого циклу комерційного контенту через підсистему супроводу комерційного контенту. Ці правила мають збільшувати коло цільової аудиторії, кількість відвідувань, кількість унікальних відвідувачів, кількість повторних відвідувань, кількість відвідувань з пошукових систем, кількість прямих відвідувань, кількість регіональних відвідувань, кількість тематичних відвідувань тощо для інформаційних ресурсів у СЕКК.

### **Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями**

Метою моделювання процесів опрацювання інформаційних ресурсів у системах електронної контент-комерції є створення загального підходу для проектування, побудови та впровадження аналогічних систем. Розв'язання такої задачі сприяє узагальненню методики опрацювання інформаційних ресурсів у СЕКК через етапи формування, управління та супроводу комерційного контенту для скорочення тривалості побудови типових систем електронного бізнесу. Впровадження систем, побудованих на основі розроблених моделей, сприяє скороченню часу на виробництво власного комерційного контенту, аналіз зовнішнього комерційного контенту, отриманого із зовнішніх джерел, аналіз динаміки життєвого циклу комерційного контенту, аналіз статистики функціонування СЕКК, аналіз статистики діяльності користувачів інформаційних ресурсів у СЕКК,

збільшення цільової аудиторії інформаційних ресурсів та розширення функціональних можливостей цих СЕКК. Результатом розроблення моделювання процесів опрацювання інформаційних ресурсів у системах електронної контент-комерції є набір функціональних вимог та стандартизованих специфікацій для створення аналогічних систем. Метою створення таких вимог є забезпечення узагальненого підходу для розроблення СЕКК типу інтернет-газета, інтернет-журнал, інтернет-видавництво, інтернет-видання, дистанційне навчання, інтернет-магазин для продажу контенту у вигляді електронних книг, фото, відео, аудіо тощо. Стандартизація вимог для побудови СЕКК забезпечує створення узагальненого підходу для розробників таких систем з метою скорочення часу для процесів створення та впровадження таких систем з уникненням етапу розроблення відповідного проекту.

### Аналіз останніх досліджень та публікацій

Процес проектування та створення СЕКК за результатами інтернет-маркетингу є ітеративним (рис. 1) 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 [1–3, 6–8, 12]. 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 великих зусиль збирають цінну інформацію, одночасно викликаючи у користувачів відчуття безпосередньої участі в процесі проектування та завойовуючи їх довіру.

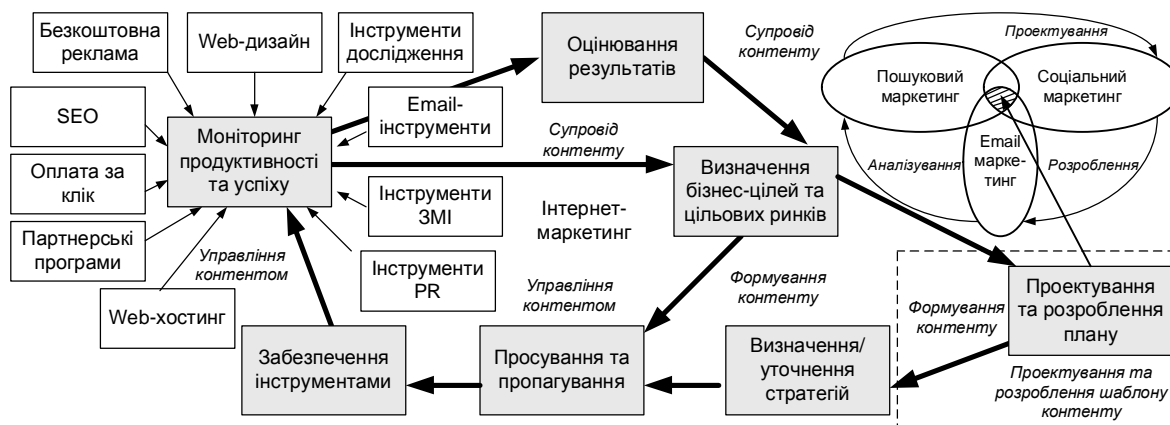


Рис. 1. Схема застосування механізмів інтернет-маркетингу в системах електронної контент-комерції

Головним компонентом проекту є розроблення інформаційної архітектури. Завдяки моніторингу продуктивності та успіху користувачів визначають одну із вимог – доступ до даних про конференції або цільову аудиторію. За реакцією користувачів формують інформаційну архітектуру сайта (рис. 2) відповідно до типу системи торгівлі контентом [1–2, 12–14]. Основні класи користувачів/персонажів сайта (клієнти, керівники робочих груп і адміністратори) визначають дизайн сайта і процес ухвалення рішень. Визначивши важливу інформацію та її зв'язок із основними класами користувачів, створюють архітектуру контенту сайта, його ієрархію, способи подання і взаємодії кожного класу користувачів з цією інформацією (наприклад, дані конференції містять теми та питання порядку денного або сесії, планування цих питань). Класифікація електронних каталогів: загальний (забезпечує широкий список різних сайтів), комерційний (концентрує увагу на характеристиках ринку, постачальників, конкурентах) та спеціалізований (охоплює одну вузьку тематику) [1, 2].

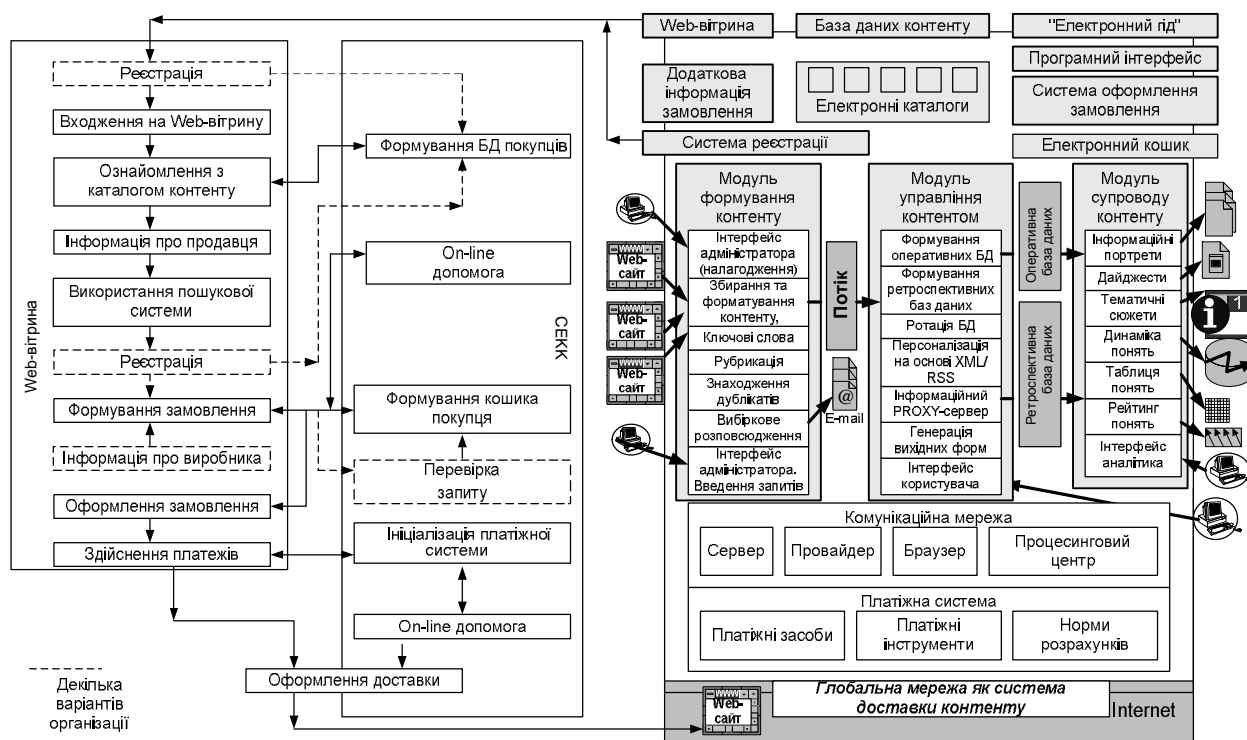


Рис. 2. Схема взаємодії компонентів системи електронної контент-комерції

На рис. 3 подано типову схему опрацювання інформаційних ресурсів у СЕКК. Після вибору контенту покупець заповнює форму із зазначенням методу оплати/доставки. Для захисту персональної інформації взаємодія здійснюється через захищений канал SSL. Після закінчення формування замовлення і реєстрації зібрана інформація про покупця надходить в СЕКК з ініціюванням запиту до платіжної системи. Після повідомлення про здійснення on-line платежу в СЕКК формується замовлення доставки.

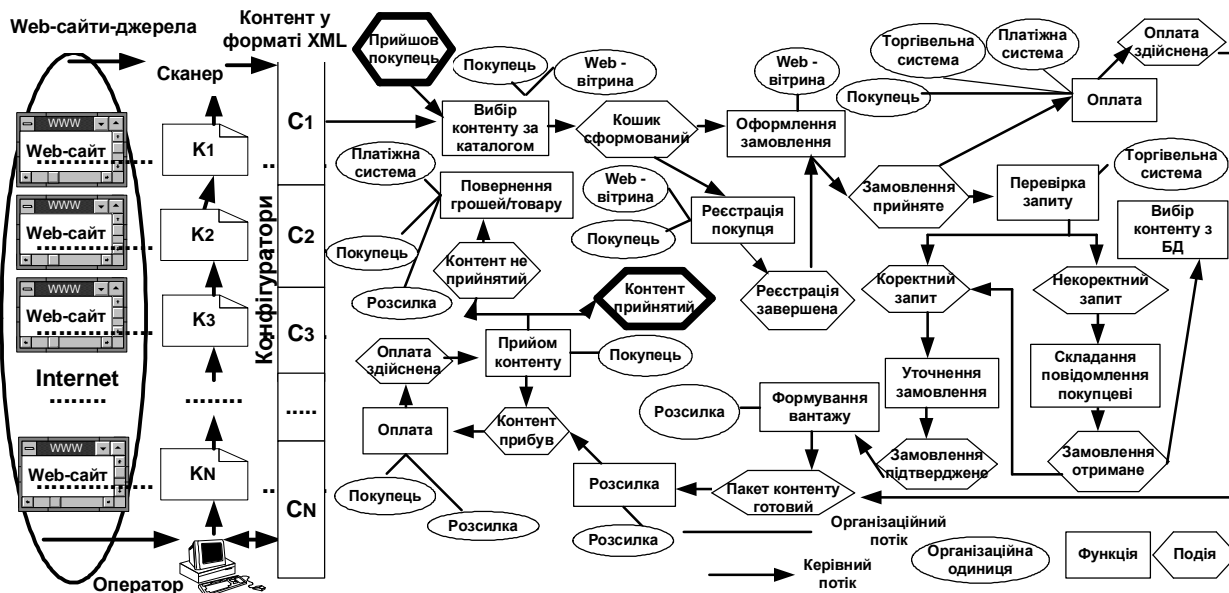


Рис. 3. Загальна схема процесу управління контентом

Структура комплексу управління СЕКК реалізується у вигляді триланкової архітектури клієнт/сервер. Процес опрацювання контенту відбувається за схемою "клієнт-сервер застосувань -

БД". Запит опрацьовується сервером застосувань, який зв'язується з БД і платіжною системою, а за наявності підключення до бізнес-процесу організації обмінюється даними з відповідними системами. З технічного погляду СЕКК – це сукупність Web-вітрини й торговельної системи – фронт-системи і бек-офісу. Основні функції СЕКК – це інформаційне обслуговування покупця, опрацювання замовлень, здійснення платежів, збирання і аналіз статистичної інформації [1–8, 10–12]. Програмний комплекс СЕКК формує інтерфейс із покупцем та функціональні можливості системи, урахуваючи потреби компанії. Потенційний клієнт має можливість у будь-який момент одержати відповіді на будь-яке запитання (умови післяпродажного сервісу, консультації з особливостей схем оплати тощо), що супроводжує процес купівлі/продажу. Реєстрація/авторизація відбувається до/після вибору контенту. У першому випадку створюється реєстраційний вхід для постійних клієнтів СЕКК, для яких реалізується спеціальна схема обслуговування й оплати. Можливість реєстрації після вибору контенту дає змогу покупцеві зберегти анонімність і заощаджує час. Система забезпечує особисту інформацію покупця, користуючись для передавання даних захищеними каналами (протоколи SSL або SET). СЕКК одержує повну інформацію про відвідувачів Web-сайта, що дає змогу будувати відповідно до неї систему маркетингу. CMS дозволяє зібрати для аналізу максимум статистичної інформації та оперативно її використовувати (виявити місця сайта, оптимальні для розміщення рекламної інформації, автоматизувати хід рекламної кампанії). Публікація додаткової інформації реалізується за допомогою окремого сервера застосувань (область публікацій) і БД. СЕКК має підтримувати життєвий цикл контенту (табл. 1) [1–3, 6–8, 10–12].

Таблиця 1

#### Класифікація життєвого циклу контенту

№	Етапи опрацювання інформаційних ресурсів		
	Формування контенту	Управління контентом	Супровід контенту
1	Збирання/створення контенту (збирання інформації з різних Web-сайтів та її форматування).	Опрацювання контенту (формування БД, їх ротація і забезпечення доступу).	Структурування контенту (виявлення тематичних сюжетів; побудова таблиць взаємозв'язку понять).
2	Систематизація контенту (виявлення ключових слів та понять, дублювання змісту, автоматична рубрикація).	Аналіз контенту (персоналізація та аналіз статистики роботи користувачів).	Модерація контенту (формування інформаційних портретів та дайджестів).
3	Поширення контенту (визначення рейтингу модераторів та вибіркоче поширення контенту).	Подання контенту (забезпечення пошуку в БД; генерація вихідних форм; взаємодія з іншими БД).	Узагальнення контенту (розрахунок рейтингів понять, виявлення нових подій, їх відстеження й кластеризація).

СЕКК передбачає використання ІТ для взаємодії торговельних компаній з роздрібними покупцями, забезпечення повного циклу продажу контенту. У СЕКК склад учасників набагато ширший, крім користувачів (продавців, покупців) до нього входить низка фінансових установ (банк-емітент, банк продавця/покупця, банк-еквайр), комп'ютерних центрів тощо. Користувачами виступають переважно фізичні особи, а також установи, соціальні заклади, інші види споживачів (юридичні особи). Продавцями в СЕКК є різні організаційні форми контент-торгівлі. Комунікаційну мережу формують провайдери, сервери, процесингові центри тощо. Систему доставки утворюють відповідні канали обміну та ресурси Інтернету [1, 2]. Усі складові елементи взаємодіють у системі взаємозв'язків. У цьому гарантія стійкості й надійності СЕКК. Важливими елементами СЕКК є організаційні форми електронної контент-торгівлі, що мають єдину спрямованість – забезпечення процесу роздрібної купівлі-продажу, але відрізняються складом, структурою, призначенням у СЕКК [1, 2]. Web-вітрина належить до простих організаційних форм продажу контенту, має самостійний характер або є складовою СЕКК. Web-вітрина є інструментом для залучення покупців та не забезпечує повного циклу купівлі-продажу товару. Це завдання

виконує СЕКК, що забезпечує повний цикл купівлі/продажу контенту через Інтернет в інтерактивному режимі з використанням електронного каталогу. Електронні каталоги – це довідники з можливістю пошуку контенту за такими характеристиками, як назва, призначення, дата, категорія тощо; множина систематизованого контенту з доступною системою навігації) [1, 2]. СЕКК забезпечує ознайомлення, вибір категорії контенту, оформлення замовлення, здійснення взаєморозрахунків, відстеження виконання замовлення. Для функціонування СЕКК має програмно-апаратні компоненти: Web-вітрину (фронт-офіс) на Web-сервері; електронні каталоги; платіжну систему; СЕКК. Web-вітрина має активний зміст, є статичною на базі звичайних HTML-файлів або динамічною з відображенням контенту з БД. Web-вітрина містить інформацію про назву, профіль, статус власника СЕКК, асортимент контенту та послуг, засоби платежів, знижки, гарантії, умови доставки контенту [1, 2].

### Виділення проблем

Процес опрацювання інформаційних ресурсів виконується за такою схемою *формування контенту* → *управління контентом* → *супровід контенту*, тому модель систем електронної контент-комерції подамо як  $S = \langle X, Formation, C, Management, Realization, Y \rangle$ , де  $X = \{x_1, x_2, \dots, x_{n_X}\}$  – множина вхідних даних, *Formation* – оператор формування контенту,  $C = \{c_1, c_2, \dots, c_{n_C}\}$  – множина контенту, *Management* – оператор управління контентом, *Realization* – оператор супроводу контенту та  $Y = \{y_1, y_2, \dots, y_{n_Y}\}$  – множина вихідних даних. Вхідні дані  $x_i$  є незалежними змінними, а множина контенту  $c_j$  та вихідні дані  $y_k$  є залежними [2, 10, 11]. Етап формування контенту описується оператором  $c_j(x_i, t) = Formation(U_F, x_i, t, \Delta t)$  за умов  $U_F = \{u_{f_1}, u_{f_2}, \dots, u_{f_m}\}$  з результатом  $c_j = \left\{ \mathbf{U}u_{f_k} \mid (x_i \in X) \wedge (\exists u_{f_k} \in U_F), U_F = U_{F_x} \vee U_{F_{\bar{x}}}, i = \overline{1, m}, k = \overline{1, n} \right\}$ . Етап управління контентом описується оператором  $c_j(q_1, t) = Management(U_M, q_1, t, \Delta t)$  при запитах користувачів  $Q = \{q_1, q_2, \dots, q_{n_Q}\}$  з умовами  $U_M = \{u_{m_1}, u_{m_2}, \dots, u_{m_p}\}$ . Управління комерційним контентом відбувається як  $c_j = \left\{ \mathbf{U}u_{m_k} \mid (q_i \in Q) \wedge (\exists u_{m_k} \in U_M), U_M = U_{M_q} \vee U_{M_{\bar{q}}}, i = \overline{1, m}, k = \overline{1, n} \right\}$ . Етап супроводу контенту описується як  $y_k(t + \Delta t) = Realization(U_R, c_j, q_1, t, \Delta t)$  за умов  $U_R = \{u_{r_1}, u_{r_2}, \dots, u_{r_z}\}$ , тобто  $y_j = \left\{ \mathbf{U}u_{r_k} \mid (q_i \in Q) \wedge (c \in C) \wedge (\exists u_{r_k} \in U_R), U_R = U_{R_c} \vee U_{R_{\bar{c}}}, i = \overline{1, m}, k = \overline{1, n} \right\}$ . Отже, побудова ефективної та адекватної моделі процесів формування, управління та супроводу контенту дає змогу виробити єдині уніфіковані підходи до побудови систем електронної контент-комерції. Виконання такого завдання, своєю чергою, значно спростить і прискорить процеси створення і запровадження таких систем.

### Виокремлення мети

Мета роботи – побудова моделі 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 ці 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

Інформаційний ресурс у СЕКК – це множина даних з набором властивостей, які є об'єктом дій технології перетворення їх на контент. Результат застосування однієї технології може бути інформаційним ресурсом іншої. Контент у галузі інформаційних технологій є формалізованими

відомостями і знаннями, розміщеними у середовищі системи і, на відміну від даних, без детальної специфікації їх властивостей, способів формалізації і впорядкування. Перетворення різних за природою, змістом та походженням даних на узгоджений централізований інформаційний ресурс є однією з важливих проблем побудови та функціонування СЕКК. Порядок формування і використання інформаційних ресурсів у СЕКК визначає способи відбору даних із первинних джерел, їх фіксацію, фільтрування, перетворення до визначеного формату для формування контенту і розміщення в базі даних. Цінність контенту визначає його привабливість для споживача. Інтеграція контенту робить інформаційний ресурс привабливим, а інтеграція застосувань – корисним. Використання СЕКК не потребує встановлення програмного забезпечення. Для редагування й адміністрування використовують браузер. Інтуїтивний інтерфейс і простота роботи із системою полегшує управління інформаційним ресурсом і знижує подальші витрати на його підтримку. СЕКК надає такі можливості: швидке поновлення та пошук контенту в інформаційному ресурсі; збирання даних про клієнтів та потенційних клієнтів; формування і редагування опитувань; аналіз відвідування інформаційного ресурсу.

Розв'язання задач вивчення, аналізу та опису основних процесів опрацювання інформаційних ресурсів у системах електронної контент-комерції забезпечує побудова низки формальних моделей. Метою кожної з моделей є специфікація маніпуляцій, дій та перетворень, які виконуються над комерційним контентом. Основними процесами, моделювання яких описано в цій роботі, є процеси формування, управління та реалізації комерційного контенту.

**Моделювання процесу формування контенту.** Цей процес забезпечує отримання інформації з різноманітних Web-сайтів та її форматування; виявлення ключових слів і понять контенту; рубрикацію контенту; виявлення дублювання змісту контенту; вибіркоче поширення контенту.

Процес формування комерційного контенту для інформаційного ресурсу забезпечує зв'язок між множиною вхідних даних з різних джерел інформації та множиною сформованого комерційного контенту, збереженого у відповідній базі даних у системах електронної контент-комерції, тобто  $Source(x_i) \rightarrow x_i \rightarrow X \rightarrow Formation(u_f, x_i, t_p) \rightarrow c_r \rightarrow C \rightarrow DataBase(C)$ , 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 контенту з  $x_i$  контенту джерела в фіксований час  $t_p$  за умов формування контенту  $u_f$ ;  $c_r$  – сформований згідно з умовою  $u_f$   $r$ -й комерційний контент;  $C$  – множина сформованого комерційного контенту;  $DataBase(C)$  – база даних сформованого комерційного контенту. Формування контенту реалізується у вигляді комплексів контент-моніторингу зі збирання контенту з різних джерел даних, і забезпечують створення бази даних контенту відповідно до інформаційних потреб користувачів. У результаті збирання і первинного опрацювання контент приводять до єдиного формату, класифікують відповідно до визначеного рубрикатора та приписують йому дескриптори з ключовими словами. Загальну модель формування контенту подамо як кортеж виду

$$Formation = \langle X, C, Gathering, Formatting, Key, Categorization, Backup, Dissemination \rangle,$$

де  $X = \{x_1, x_2, \dots, x_{n_x}\}$  – множина вхідної інформації з Web-сайтів або від модераторів;  $C = \{c_1, c_2, \dots, c_{n_c}\}$  – множина контенту;  $Gathering$  – оператор збирання інформації з джерел;  $Formatting$  – оператор форматування інформації, перетворення на множину контенту;  $Key$  – оператор виявлення ключових слів;  $Categorization$  – оператор автоматичної рубрикації;  $Backup$  – оператор виявлення дублювання контенту;  $Dissemination$  – оператор вибіркового поширення контенту.

Ефективне вирішення, здатне допомогти орієнтуватися в динамічній вхідній інформації з різних джерел, забезпечує процеси синдикації даних (збирання інформації з джерел і поширення надалі її фрагментів відповідно до потреб користувачів). Технологія синдикації контенту містить

процес навчання програм збирання даних структурним особливостям окремих джерел (з Web-сайтів, від модераторів, користувачів, відвідувачів, журналістів, редакторів), безпосереднє сканування контенту та приведення до загального формату в XML, рубрикацію. Засоби класифікації та розподілу контенту є інформаційно-пошуковою системою вибіркового поширення контенту (контентний роутер). Отриманий контент аналізують на відповідність тематичним запитам. Релевантний контент розсилається користувачам та завантажується в тематичні БД. Множина контенту формується оператором  $C = Formatting(Gathering(X, U_G), U_{FR})$ , де  $U_G$  – множина умов збирання даних з різних джерел,  $U_{FR}$  – множина умов форматування даних. Оператор виявлення дублювання змісту контенту  $C = Backup(C, U_B)$ , де  $U_B$  – множина умов виявлення дублювання змісту контенту. Виявлення дубльованих за змістом контенту в СЕKK виконується на основі лінгвостатистичних методів, що полягають у виявленні загальних термів, ланцюжки яких утворюють словесні сигнатури контенту. Аналіз властивостей контенту на рефлексивність, симетричність ( $\forall i, j : a_{ij} = a_{ji}$ ) та транзитивність ( $\forall i, j, k : a_{ij} = 1, a_{jk} = 1 \Rightarrow a_{ik} = 1$ ) використовує відношення еквівалентності, тобто відношення збігу змісту або дублювання. Властивість дублювання контенту є жорсткішим критерієм подоби, наприклад, збіг 3, 4 або 5 термів свідчить про деяку близькість змісту. Кожному контенту  $c_j$  за наведеним вище алгоритмом збігу термів у сигнатурах ставиться у відповідність вектор з елементами  $a_{ij}$ , який за значення  $c_j \equiv c_k$  дорівнює 1, в іншому випадку 0. Досліджують критерії подоби, змінюючи обсяг у сигнатурах термів порівняння для досягнення коефіцієнтів асиметричності (попереднє визначення дублікатів) і транзитивності (визначення дублікатів з повнотою). Ці коефіцієнти використовують для перевірки наближення відношення до властивостей еквівалентності.

Оператор рубрикації  $C_{Ct} = Categorization(Key(C, U_K), U_{Ct})$ , де  $Key(C, U_K)$  – функція виявлення ключових слів і понять,  $U_K$  – множина умов виявлення ключових слів і понять,  $U_{Ct}$  – множина умов автоматичної рубрикації,  $C_{Ct}$  – множина рубрикованого контенту. Опрацювання множини контенту  $C$  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 (найвживаніші слова ігноруються за допомогою стоп-словника, а рідкісні слова з текстів повідомлень не враховуються). Оператор вибіркового поширення контенту  $C_D = Dissemination(C_{Ct}, U_{Ds})$ , де  $C_{Ds}$  – множина вибірково поширеного контенту,  $U_{Ds}$  – множина умов вибіркового поширення контенту. Вибіркове поширення списку контенту  $C_{Ds} = sup(C_{Ct})$  залежить від рівня попиту на цей контент. Асоціативне правило формування списку контенту складається з первісного списку контенту  $C_{Ct}$  і списку контенту, вибраних із первісного – похідний список  $C_{Ds}$ , тобто  $C_{Ct} \rightarrow C_{Ds}$ . Формування асоціативного правила полягає у формуванні контенту, утвореного внаслідок об'єднання первісного і похідного списків. Ймовірність утворення асоціативного правила – це ймовірність, з якою контент із первісного списку  $C_{Ct}$  з'являється разом із контентом із  $C_{Ds}$  в БД, тобто  $Conf(C_{Ct} \rightarrow C_{Ds}) = \frac{sup(C_{Ct} \cup C_{Ds})}{sup(C_{Ct})}$ . Похідний список вибраних

правил визначає користувач. Частовживаним списком контенту є список, формування якого перевищує мінімальний рівень. Навіть у разі виконання цієї умови частовживаного контенту в запитах багато. Результат обмежується при  $Imp(C_{Ct} \rightarrow C_{Ds}) = \min(\forall C_{Ct} \subset C_{Ct}, Conf(C_{Ct} \rightarrow C_{Ds}) - Conf(C_{Ct} \rightarrow C_{Ds}))$ .

**Модель управління інформаційними ресурсами.** Головними завданнями управління контентом є: формування, ротація баз даних комерційного контенту і забезпечення доступу до них; 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 інформаційними ресурсами. Моделювання процесу управління контентом реалізується за такою схемою.

1. Генерація сторінок за запитом відбувається на основі послідовності дій – “редагування контенту → запис до бази даних → подання контенту”. Модель етапу генерації сторінок за запитом користувача подамо як  $Management_Q = \langle X, C, Q, R, Edit, Y \rangle$ , де  $X$  – множина вхідної інформації;  $C$  – множина контенту;  $Y$  – множина сформованих сторінок;  $Q$  – множина запитів;  $R$  – функція формування та подання сторінки;  $Edit$  – функція редагування контенту, тобто  $c(x_i, t) = Edit(c, x_i, t)$ . Формування сторінок описується оператором  $y(t + \Delta t) = R(q, c, Weight, t, \Delta t)$ , де  $Weight$  – загальна вага контентного блока, тобто  $y_i = \left\{ \bigcup c_j \mid (\forall c_j \in C_q) \wedge (\exists q_i \in C_q), C = C_q \vee C_{\bar{q}}, j = \overline{1, m}, i = \overline{1, n} \right\}$ .

2. Генерація сторінок під час редагування полягає у створенні набору статичних сторінок при внесенні змін до змісту сайта, тобто за схемою “модуль редагування → база даних”. При цьому не враховується інтерактивна взаємодія між відвідувачем і вмістом сайта. Модель етапу генерації сторінок під час редагування подамо як  $Management_E = \langle C, Edit, Y \rangle$ , де  $C$  – множина контенту;  $Y$  – множина статичних сторінок;  $Edit$  – функція редагування/модифікації контенту, тобто  $y(t) = Edit(c, Weight, t)$ .

3. Змішаний тип поєднує переваги перших двох типів та реалізується за схемою “редагування → база даних інформаційних блоків → збирання → подання” одним з двох способів.

1. Шляхом кешування модуль подання генерує сторінку один раз, надалі вона значно швидше довантажується з кешу, який оновлюється автоматично після закінчення деякого терміну часу, під час оновлення певних розділів сайта або вручну за командою адміністратора.

2. Формуванням інформаційних блоків – збереження блоків на етапі редагування сайта і збирання сторінки з цих блоків за запитом користувача.

Модель змішаного типу подамо як  $Management_M = \langle X, C, Q, R, Edit, Caching, Y \rangle$ , де  $X$  – множина вхідної інформації;  $C$  – множина контенту;  $Y$  – множина сформованих сторінок;  $Q$  – множина запитів;  $R$  – функція формування та подання сторінки;  $Edit$  – функція редагування та модифікації контенту,  $Caching$  – функція формування кешу або інформаційних блоків, де кеш  $Cache = Caching(y, Weight, t, \Delta t)$ , тобто  $Cache = \left\{ \bigcup y_i \mid y_i \in Y, t + \Delta t, i = \overline{1, n} \right\}$ .

**Модель процесу реалізації контенту.** В межах цього процесу забезпечується формування інформаційних портретів; формування дайджестів (короткий зміст публікацій); виявлення тематичних сюжетів; побудова таблиць взаємозв'язку понять; розрахунок рейтингів понять; виявлення нових подій, їх відстеження та кластеризація. Формальну модель реалізації контенту задає кортеж вигляду

$$Realization = \langle X, C, BuInfPortr, BuDigest, IdThemTop, ConCorrTablConc, CalRankConc, Y \rangle,$$

де  $X$  – множина вхідної інформації,  $C$  – множина контенту,  $BuInfPortr$  – функція формування інформаційних портретів,  $BuDigest$  – функція формування дайджестів,  $IdThemTop$  – функція виявлення тематичних сюжетів,  $ConCorrTablConc$  – функція побудови таблиць взаємозв'язку понять,  $CalRankConc$  – функція розрахунку рейтингів понять,  $Y$  – множина вихідної інформації. Множина контенту – це  $C = \langle C_P, C_D, C_T, C_C, C_R \rangle$ , де  $C_P$  – підмножина контенту інформаційних портретів,  $C_D$  – підмножина контенту дайджестів,  $C_T$  – підмножина контенту тематичних сюжетів,  $C_C$  – підмножина контенту таблиць взаємозв'язку понять,  $C_R$  – підмножина контенту рейтингів понять. Множина контенту інформаційних портретів



$C_P = BuInfPortr(X, U_P, C_P)$ , де  $U_P$  – множина умов формування інформаційних портретів. Множина контенту дайджестів  $C_D = BuDigest(X, U_D)$ , де  $U_D$  – множина умов формування дайджестів. Множину тематичних сюжетів визначають як  $C_T = IdThemTop(X, C_T, U_T)$ , де  $U_T$  – множина умов виявлення тематичних сюжетів.

Множина таблиць взаємозв'язку понять  $C_C = ConCorrTablConc(C, U_C)$ , де  $U_C$  – множина умов побудови таблиць взаємозв'язку понять. Множина рейтингів понять  $C_R = CalRankConc(C_C, Tonicity, U_RSpam)$ , де  $U_R$  – множина параметрів розрахунку рейтингів понять,  $Tonicity$  – критерій тональності контенту,  $Spam$  – оператор визначення спаму. Множина вихідної інформації  $Y = Realization(C_D, C_R)$ . Під час оцінювання тональності контенту простір гіпотез містить тональності  $Tonicity = H_{-1}$  (негативна),  $Tonicity = H_0$  (нейтральна) і  $Tonicity = H_1$  (позитивна). Наприклад, за гіпотез  $H_1$  (тональність позитивна) та  $\overline{H_1}$  (тональність не позитивна) із множини з позитивною тональністю вибирають терми, характерні для цього контенту. Із них вибирають терми  $t$  із ймовірністю  $> 0,5$ , наприклад 0,6, яка розраховується за формулою Байеса. Такі терми називають тонально-забарвленими, або просто тональними, які мають вартісну семантику та  $\forall t$  вагу  $Weight = a$ . Позитивна і негативна тональності є антагонізмами, тому остаточне рішення про тональність контенту приймають з урахуванням різниці значень вагових оцінок гіпотез  $H_1$  і  $H_{-1}$ .

Модель СЕКК є універсальною та простою для опису процесу функціонування типових систем. Моделі процесів опрацювання інформаційних ресурсів дають змогу уніфікувати такі основні функції системи, як формування, управління та реалізація контенту. На основі розроблених моделей, зокрема, розв'язують задачі побудови типової архітектури систем електронної контент-комерції, формують специфікацію функціональних та нефункціональних вимог щодо СЕКК та їх складових, таких як засоби взаємодії з кінцевим користувачем, реалізують процедури аналізу контенту та збору й опрацювання статистичних даних відвідування інформаційних ресурсів. Моделі процесів формування, управління та реалізація контенту дають змогу спростити процеси проектування СЕКК за допомогою уніфікації та стандартизації методів опрацювання інформаційних ресурсів.

### **Висновки і перспективи подальших наукових розвідок**

У статті 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.

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. Овсяк В. Алгоритми: методи побудови, оптимізації, дослідження вірогідності / В. Овсяк. – Л: Світ, 2001. – 160 с. 10. Пасічник В. Математична лінгвістика / В. Висоцька, В. Пасічник, Ю. Щербина, Т. Шестакевич. – Л: Новий Світ – 2000, 2012. – 359 с. 11. Советов Б. Моделирование систем / Б. Советов, С. Яковлев. – М: ВШ, 1998. 12. Федорчук А. Контент-мониторинг информационных потоков / А. Федорчук. – Киев, 2005. – № 3. 13. Molga A. Problem solving on absolute value – relevance of visualisation by means of TI-Nspire graphic calculator / Molga A., Wójtowicz M. // Education-Technology-Computer Science, Main problems of informatics and information education, Scientific annual No/2/2011/Part2, Uniwersytet Rzeszowski. – Wydawnictwo Oświatowe FOSZE, Rzeszów 2011. – S. 242–252. 14. Molga A. Use of computer technology in the design process / Molga A., Wójtowicz M. // Education-Technology-Computer Science, Main problems of informatics and information education, Scientific annual No/4/2013/Part2, Uniwersytet Rzeszowski. – Wydawnictwo Oświatowe FOSZE, Rzeszów 2013. – S. 418–423.

УДК 004

В. Гнатушенко, Данладі Алі

Національна металургійна академія України,  
кафедра інформаційних технологій та систем

## ДОСЛІДЖЕННЯ САМОПОДІБНИХ ПРОЦЕСІВ ПЕРЕДАЧІ ТРАФІКУ НА ОСНОВІ ON/OFF МОДЕЛІ

© Гнатушенко В., Данладі Алі, 2013

**Розглянуто взаємозв'язок різних характеристик випадкового процесу, пов'язаних з ефектом самоподібності. Проведено імітаційне моделювання процесу передавання трафіку. Отримано результати оцінки показника Херста різними методами.**

**Ключові слова:** випадковий процес, on/off модель, трафік, самоподібність, мережа.

**The article examines the relationship of the various characteristics of the random process associated with the effect of self-similarity. A simulation of the transfer traffic. The results Hurst parameter estimation by different methods.**

**Key words:** random process, on/off model, traffic, self-similarity, network.

### Постановка проблеми

Під час дослідження та вивчення різних властивостей і характеристик систем передачі інформації виникає завдання побудови моделей вхідного трафіку, характеристики яких були б найближчими до характеристик реальних потоків даних. Дослідження різних типів мережевого трафіку за останні роки доводять, що мережевий трафік є самоподібним (self-similar) або фрактальним (fractal) за природою [1–5]. З цього випливає, що використовувані методи моделювання та розрахунку мережевих систем, основані на використанні пуассонівських потоків, не дають повної та точної картини того, що відбувається в мережі.

### Аналіз останніх досліджень

У роботі [6] на основі аналізу великої кількості робіт з дослідження трафіку в IP мережах наведено класифікацію трафіку, і кожному виду трафіку увідповіднений закон розподілу. У роботі