

УНІФІКОВАНИЙ МЕТОД ФОРМУВАННЯ ІНФОРМАЦІЙНОГО ПРОДУКТУ В СИСТЕМАХ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

У статті описано формальну модель формування інформаційного продукту в системах електронної контент-комерції, яка спрощує технологію автоматизації процесу створення та опрацювання текстового контенту. У роботі запропоновано уніфікований метод та функціональні сервіси формування контенту.

Ключові слова: контент, інформаційний ресурс, система електронної комерції.

В статье описана формальная модель формирования информационного продукта в системах электронной контент-коммерции, которая упрощает технологию автоматизации процесса создания и обработки текстового контента. В работе предложен унифицированный метод и функциональные сервисы формирования контента.

Ключевые слова: контент, информационный ресурс, система электронной коммерции.

In this article a formal model of information products formation is described in electronic content commerce systems. It simplifies the technology to the process automation of the text content creation and processing. The paper suggests a unified method and functional services of content forming.

Key words: content, informational resource, electronic commerce system.

Вступ. Загальна постановка проблеми. Ринок із поширення контенту забезпечує технологічний процес підготовки оперативного контенту, доступного через інформаційні ресурси та залежного від сприйняття, відображення, збереження його значень. Для опрацювання і розв'язування певного кола задач модератори інформаційної системи формалізують, аналізують, форматують та структурують контент. Процес структурування – це визначення одиниць контенту, способів і порядку їх поєднання між собою та утворення більших елементів контенту з дрібніших [1; 5]. Утворений контент надходить у бази/сховища даних, де визначають його напрями/тематики, наприклад електронних публікацій із більшим коефіцієнтом попиту у відвідувачів/користувачів інформаційного ресурсу. Структурований контент зосереджений, наприклад, в ERP/CRM, а неструктурований контент – в електронній пошті, робочих документах довільного формату та інструментах забезпечення колективної роботи і зберігається, наприклад, в ECMS [1].

Аналіз останніх досліджень та публікацій. Життєвий цикл контенту (англ. Content lifecycle) – це складний процес, який проходить контент під

час управління через різні етапи/фази публікації з набором таких властивостей, як спільна робота, управління записами, цифровими активами й версіями, які підтримуються різними технологіями [1; 5; 8-21]. Найвні інструментарії е-комерції надають адміністратору/модератору системи різні можливості з управління контентом (формувати, формалізувати, структурувати, додавати, редагувати, видаляти), але не вирішують проблеми автоматичного опрацювання інформаційних ресурсів [1-7]. Тому для реалізації життєвого циклу контенту необхідний інструментарій, який реалізує процеси автоматичного формування, управління та супроводу контенту [20].

Контент характеризується часом оновлюваності/модифікації та володіє набором специфічних властивостей. Об'єм контенту вимірюють в одиницях кількості інформації. Кількість/якість контенту характеризують ступінь інтересу користувача до інформаційного ресурсу, де він розміщений [1; 4]. Економічний контент є елементом економічної діяльності суб'єкта електронного бізнесу. Ринок контенту на базі інтернет із інформаційними технологіями управління знаннями є засобами, які сприяють функціонуванню е-бізнесу з поширення

комерційного контенту та росту його прибутковості для суб'єктів електронної комерції [1]. Комерційний контент є об'єктом купівлі/продажу між учасниками е-комерції [1], наприклад інформаційні блоки, які поділяють на синдикати (курси валют, блок погоди), анонси матеріалів інших розділів/ресурсів (із посиланням), довідкова інформація (святкові дати, анонс заходу, розклад руху поїздів), розважальна інформація (анекдот дня), реклама, кнопки і посилання інформаційних партнерів, кнопки статистики.

Виділення проблем. Автори моделей життєвого циклу комерційного контенту пропонують та описують кілька етапів із набором

властивостей, які підтримуються різними технологіями та процесами (табл. 1). У деяких моделях життєвого циклу контенту передбачені концепції управління проектом/контентом/ресурсом, інформаційна архітектура, стратегії контенту, семантичний друк. Різні автори пропонують різні етапи життєвого циклу контенту [8-21]. Основні етапи (створення контенту, розроблення, перегляд, поширення та архівація) присутні майже у всіх запропонованих моделях [8-21]. Життєвий цикл процесів, дій, статусу та ролі управління контенту відрізняються в моделях залежно від організаційних стратегій, потреб, вимог і можливостей моделей [1; 20].

Таблиця 1

Класифікація моделей життєвого циклу комерційного контенту

Автор	Модель
McKeever S.	Збирання, доставка/публікація
Bob Boiko	Збирання, управління, публікація
McGovern G.	Створення, редагування, публікація
JoAnn Hackos	Створення, архівування, складання/посадження, публікація
Ann Rockley	Створення, перегляд, управління, доставка
Russell Nakano	Подання, порівняння, оновлення, злиття, публікація
The State Victoria	Розроблення, затвердження якості, публікація, скасування, публікації, архівування
АІМ	Захоплення/поглинення, управління, накопичення, доставка, зберігання
СМР organization	Планування, розроблення, управління, розгортання, зберігання, оцінювання
Bob Doyle	Організація, створення, накопичення, workflow, управління версіями, публікація, архівування
Woods Randy	Легалізація, шаблон, створення, модифікація, управління версіями, обертання, моніторинг, управління успіхом
Halverson	Аудит, аналіз, стратегія, класифікація, структуризація, створення, перегляд, повторний перегляд, завершальний перегляд, затвердження, тестування, форматування, публікація, поновлення, архівування

Розглянуті моделі життєвого циклу контенту не вирішують проблеми його формування і супроводу, та вирішують не всі проблеми управління: подання множини контенту кінцевому користувачу згідно з його запитом, історією або інформаційним портфелем; автоматичне формування дайджестів та інформаційних портретів; виявлення тематичних сюжетів та дублювання змісту контенту; побудова таблиць взаємозв'язку та розрахунок рейтингів

контенту; збирання даних із різних джерел та їх форматування; виявлення ключових слів контенту; рубрикація та вибіркоче поширення контенту.

Формулювання мети. Існуючі системи електронної комерції не підтримують весь життєвий цикл комерційного контенту та не вирішують основних проблем опрацювання інформаційних ресурсів – формування та супроводу контенту (табл. 2).

Таблиця 2

Класифікація життєвого циклу комерційного контенту в системах електронної комерції

Формування контенту	Управління контентом	Супровід контенту
Збирання/створення контенту (збирання інформації з різних інформаційних ресурсів та її форматування)	Редагування контенту (формування бази даних, їх ротация і забезпечення доступу)	Структурування контенту (виявлення тематичних сюжетів; побудова взаємозв'язку понять)
Систематизація контенту (виявлення ключових слів, дублювання змісту, рубрикація, формування дайджестів)	Опрацювання контенту (індивідуалізація та аналіз статистики роботи користувачів)	Аналіз контенту (формування інформаційних портретів користувачів, цільової аудиторії)
Поширення контенту (визначення рейтингу модераторів та вибіркоче поширення контенту)	Подання контенту (забезпечення пошуку в базі даних; генерація вихідних форм; взаємодія з іншими базами даних)	Узагальнення контенту (розрахунок рейтингів, виявлення нових подій, їхнє відстеження й кластеризація)

Система електронної контент-комерції одержує повну інформацію про відвідувачів інформаційного ресурсу, що дозволяє будувати систему маркетингу. Система електронної контент-комерції дозволяє зібрати для аналізу максимум статистичної інформації та оперативно її використовувати (дозволяють виявити місця інформаційного ресурсу, оптимальні для розміщення рекламної

інформації, автоматизувати хід рекламної кампанії).

Аналіз отриманих наукових результатів. Процес формування комерційного контенту для інформаційного ресурсу забезпечує зв'язок між множиною вхідних даних із 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 -ий контент із джерела; X – множина даних із відповідного джерела; $Formation(u_f, x_i, t_p)$ – оператор формування комерційного 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 -ий комерційний контент; C – множина сформованого комерційного контенту; $DataBase(C)$ – база даних сформованого комерційного контенту.

Типи джерел контенту для підсистеми формування комерційного контенту: список адрес інформаційних ресурсів із довірою та необхідними даними; список адрес інформаційних ресурсів із підпискою на їх контент; множина контенту від

$$Formation = \left\langle X, Gathering, Formatting, KeyWords, Backup, \right. \\ \left. Categorization, BuDigest, Dissemination, T, C \right\rangle,$$

де $X = \{x_1, x_2, \dots, x_{n_x}\}$ – множина вхідних даних $x_i \in X$ із різних інформаційних ресурсів або від модераторів при $i = \overline{1, n_x}$; *Gathering* – оператор збирання/створення контенту з різних джерел; *Formatting* – оператор форматування контенту; *KeyWords* – оператор виявлення ключових слів і понять контенту; *Categorization* – оператор автоматичної рубрикації контенту; *Backup* – оператор виявлення дублювання змісту контенту; *BuDigest* – оператор формування дайджестів контенту; *Dissemination* – оператор вибіркового поширення контенту; $T = \{t_1, t_2, \dots, t_{n_T}\}$ – час $t_p \in T$ транзакції формування контенту при $p = \overline{1, n_T}$; $C = \{c_1, c_2, \dots, c_{n_C}\}$ – множина комерційного контенту $c_r \in C$ при $r = \overline{1, n_C}$.

Процес формування комерційного контенту описано оператором вигляду $c_r = Formation(u_f, x_i, t_p)$, де u_f – множина умов формування контенту, тобто $u_f = \{u_1(x_i), K, u_{n_U}(x_i)\}$. Комерційний контент подано так:

модераторів та авторів контенту; список запитів із ключовими словами для пошукових систем. Підсистема формування контенту забезпечує збирання інформації з різних інформаційних ресурсів та її форматування, виявлення ключових слів та дублювання, формування дайджесту, рубрикацію та вибіркоче поширення контенту. Основними етапами процесу формування комерційного контенту в системах електронної контент-комерції є форматування, рубрикація та поширення контенту, які мають таку схему зв'язків: *джерело контенту* → *збирання/створення контенту* → *база даних* → *форматування контенту* → *база даних* → *виявлення ключових слів та понять* → *база даних* → *рубрикація контенту* → *база даних* → *виявлення дублювання контенту* → *база даних* → *формування дайджесту контенту* → *база даних* → *вибіркоче поширення контенту* → *модератор*.

Модель процесу формування контенту в системах електронної контент-комерції подано як:

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

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

Аналіз лексико-граматичної та семантико-прагматичної побудови тексту використовується в модулі автоматичної рубрикації контенту, головне завдання якого полягає в знаходженні в контентному потоці за допомогою контент-аналізу текстової інформації, яка найбільше відповідає тематикам сайту та потребам користувачів. Виконання перерахованих у табл. 3 етапів призводить до формування тематично підібраних масивів текстової інформації, у яких акумулюється інформація про висвітлення всіх аспектів досліджуваної проблеми, враховуючи різноманітність думок і поглядів.

Таблиця 3

Основні етапи роботи модуля рубрикації контенту

Назва	Призначення етапу
Підготовка	Визначення тематики, мети та об'єкта аналізу, його хронологічні та географічні рамки, принципи відбору
Класифікація	Формування класифікатора відбору ключових цитат та інструкції для кодувальника
Кодування	Кодування фрагментів текстової інформації
Архівація	Збереження фрагментів текстової інформації в базі даних
Аналіз	Автоматичне опрацювання фрагментів текстової інформації

Основною формою подання і збереження інформації є природна мова, тому ефективність процесу рубрикації контенту залежить від розв'язання проблеми автоматичного опрацювання текстового контенту, кінцева мета якого – розпізнавання їх змісту (табл. 4).

Таблиця 4

Основні етапи опрацювання текстів для розпізнавання змісту

Опрацювання	Призначення етапу
Формальне	Перетворення фрагментів тексту без звернення до аналізу його змісту. Цей етап – основа всіх існуючих інформаційних технологій у чинних системах опрацювання контенту. Морфологічна інформація забезпечує доступ до змісту, опосередкованого через співвідношення одиниць змісту з одиницями виразу
Змістове (семантичне)	Розпізнається зміст окремих елементів і логіко-семантичних відношень між ними для подання семантики повідомлення. Використання додаткової семантичної інформації, яка експліцитно не виражена у тексті. Цей етап – поле для теоретичних й експериментальних досліджень автоматичного семантичного аналізу
Синтаксичне	Автоматично за наявності лексико-граматичної та граматичної інформації до кожного слова синтаксично прив'язуються словоформи в реченні
Морфемне	Морфемне сегментування тексту, де виділення префіксів можливе без знання частин мови, а суфіксів – ні: потрібні різні їх набори та процедури відсікання суфіксів для іменників, дієслів, прикметників, прислівників

Обов'язковою складовою частиною лінгвістичного забезпечення процесу опрацювання контенту є морфологічний аналіз, тобто визначення для одиниць тексту інформації місця в морфологічній системі відповідної мови та ідентифікація словоформ однієї лексеми (рис. 1). Унаслідок морфологічного аналізу кожній словоформі контенту приписують коди частин мови і значення граматичних категорій (рід, число,

відмінок, вид, час, особа тощо). Характер цієї інформації, її обсяг й методи, за якими встановлюється морфологічна інформація, залежать від мети дослідження, у межах якого здійснюють аналіз, від орієнтації на характер аналізованих текстів. Аналіз присутній на всіх етапах аналізу тексту, так як морфемний, синтаксичний, семантичний аналіз не обходяться без визначення частин мови.



Рис. 1. Діаграма варіантів використання для процесу рубрикації інформаційного продукту

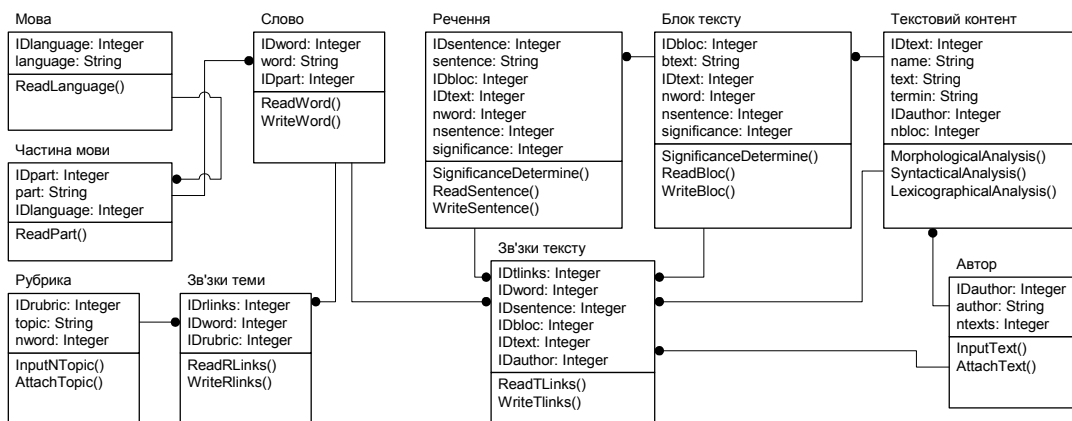


Рис. 2. Діаграма класів для процесу рубрикації інформаційного продукту

Реалізація автоматичного кодування слів тексту, тобто приписування їм кодів граматичних класів, пов'язане з граматичною класифікацією (рис. 2). Об'єктом аналізу є структура слова, форми словозміни, способи виразу граматичних значень.

Автоматичний аналіз контенту розпадається на три складові частини: автоматичне виділення основи у словоформі тексту; пошук основи у словнику основ; порівняння структури словоформи з даними про її основу, які містяться у словнику основ. Кожна словоформа тексту аналізується за допомогою заздалегідь укладених словників основ, коренів, префіксів, суфіксів, флексій. У процесі аналізу відбувається ідентифікація значень слів та синтагматичних відношень між словами

вхідного тексту. Інструментами аналізу є словник основ, словник закінчень, словник омонімічних основ, таблиці семантико-синтаксичної сполучуваності компонентів прийменникових конструкцій, зняття лексичної омонімії, семантичний аналіз іменних безприйменникових конструкцій, таблиці семантичної сполучуваності іменників і прикметників, алгоритми автоматичного аналізу, які визначають певну послідовність перевірок і звертань до словника і таблиць. Процедура автоматичного індексування розбита на блоки, які працюють послідовно: морфологічний аналіз, синтаксичний аналіз, семантико-синтаксичний аналіз прийменникових конструкцій та варіювання смислового запису запиту (рис. 3).



Рис. 3. Діаграма послідовності для процесу рубрикації інформаційного продукту

Основу системи складають словники і засоби їх обслуговування: система поділу слів тексту на флексію й основу; словник еквівалентностей (тезаурус), призначений для заміни еквівалентних слів одним або кількома номерами понять, які слугують ідентифікаторами змісту замість основ слів; тезаурус у вигляді ієрархії понять, який забезпечує пошук для певного поняття загального або асоційованого з ним поняття; словники статистичних і синтаксичних словосполучень;

система обслуговування словників. Система автоматичного індексування складається з таких етапів: виділення інформативних слів і словосполучень з тексту; розшифрування абрєвіатури; заміна слів з основами-дескрипторами на код дескриптора; зняття омонімії. Процедура індексування складається із двох етапів: індексування за дескрипторним словником; індексування за допомогою автоматичного інформаційно-пошукового тезауруса (рис. 4,а).

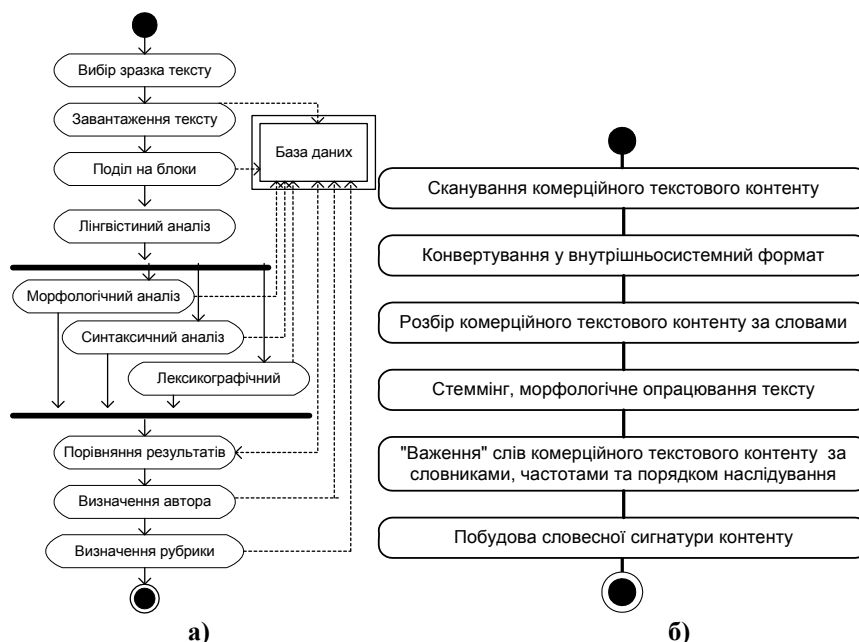


Рис. 4. Діаграма а) діяльності підсистеми рубрикації та б) виявлення дублів інформаційного продукту

Дескрипторний словник має структуру таблиці, що складається з трьох колонок: основи слів; набори дескрипторів, приписані кожній основі; граматичні ознаки дескрипторів. Рубрикація контенту реалізується з частковим морфологічним аналізом.

1. Побудова ієрархічної структури властивостей кожної лексичної одиниці мови, що містить граматичну та семантичну інформацію, і проектування лексику з ієрархічною організацією типів властивостей, де кожен тип-нащадок успадковує та перевизначає властивості предка.

2. Уніфікація – базовий механізм побудови синтаксичної структури.

Принцип виявлення ключових слів за змістом (термами) базується на законі Зіпфа і зводиться до вибору слів із середньою частотою зустрічання (найбільш зустрічальні слова ігноруються за допомогою «стоп-словника», а рідкісні слова з текстової інформації не враховуються).

Усунення контенту, що дублюється (рис. 4, б), у контентних потоках потрібне далеко не завжди. Існує ряд задач, у яких використовується факт дублювання контенту з різних джерел, наприклад при визначенні важливості контенту (якщо повідомлення багаторазово дублюється на сайтах) або при визначенні ефективності PR-кампаній (підрахунок републікацій прес-релізів тощо). При побудові моделі досліджується рівень дублювання контенту з одних Web-сайтів, які мають посилання

на інші сайти-джерела. Виявлення дубльованих за змістом текстової інформації в системах електронної контент-комерції виконується на основі лінгвостатистичних методів, що полягають у виявленні в різному контенті загальних термів, ланцюжки яких утворюють словесні сигнатури контенту.

Модуль вибіркового поширення контенту реалізує розподіл навантаження між модераторами систем електронної контент-комерції, що розвиваються, розширюються, охоплюють усе більшу аудиторію читачів та примножують кількість оперативної інформації. Зазвичай розподіл контенту виконується модераторами. Автоматичне вибіркоче поширення контенту скорочує час зменшує ресурси для подальшого функціонування систем електронної контент-комерції.

Контент містить тему та короткий опис події (дайджест). Модуль вибіркового поширення контенту автоматично розподіляє дайджести між модераторами згідно з автоматичним рейтингуванням якості їх роботи.

На рис. 5 подано діаграму класів системи автоматичного розподілу дайджестів між модераторами, що містить основні її класи, їх дані та методи, а також відношення між ними.

На рис. 6 подано діаграму кооперації, яка ілюструє взаємодію модератора з модулями оцінювання матеріалу, рейтингування журналістів та розподілу дайджестів.

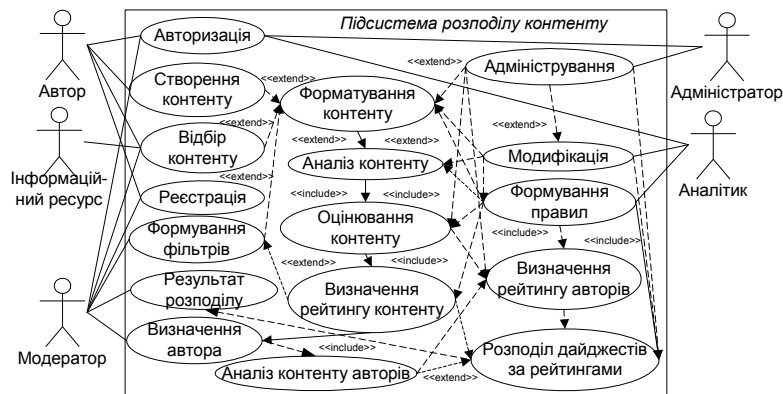


Рис. 5. Діаграма варіантів використання для процесу розподілу інформаційного продукту



Рис. 6. Діаграма кооперації для процесу розподілу інформаційного продукту

Модуль рейтингування оцінює якість роботи за сукупністю декількох критеріїв, що забезпечує об'єктивність та стимулювання якісної роботи. Частина роботи модератора в таких рутинних

функціях, як розподіл однотипної інформації, її сортування, оцінка та аналіз, зменшується. Це дозволяє звільнити людські ресурси від механічної роботи та покращити результати завдяки

об'єктивності оцінювання якості виконаних завдань. Спочатку модуль отримує готові дайджести з віддалених мережевих ресурсів за допомогою RSS. Далі дайджести розподіляються між модераторами за їх рейтингом: першими можливість обрати дайджести для опрацювання отримують працівники з найвищим рейтингом. Рейтинг працівників визначається динамічно і вказує на продуктивність та результативність роботи кожного з працівників окремо. Впливають на нього такі критерії, як процент унікальності

контенту, кількість переглядів контенту, оцінка користувача та час перебування на сторінці з контентом, що дозволяє об'єктивно оцінити якість роботи модераторів. На діаграмі послідовності (рис. 7) подано взаємодію об'єктів, упорядкованих за часом їх прояву. Діаграма надає чітке уявлення про процеси, що протікають у системі.

На рис. 8 подано діаграму класів модуля розподілу дайджестів, яка відображає основні класи системи.

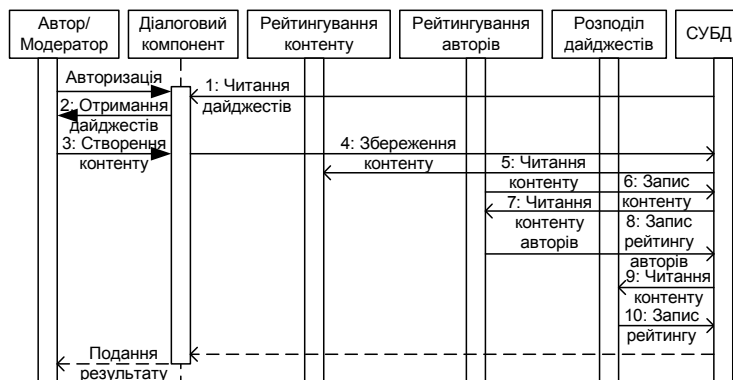


Рис. 7. Діаграма послідовності для процесу розподілу інформаційного продукту

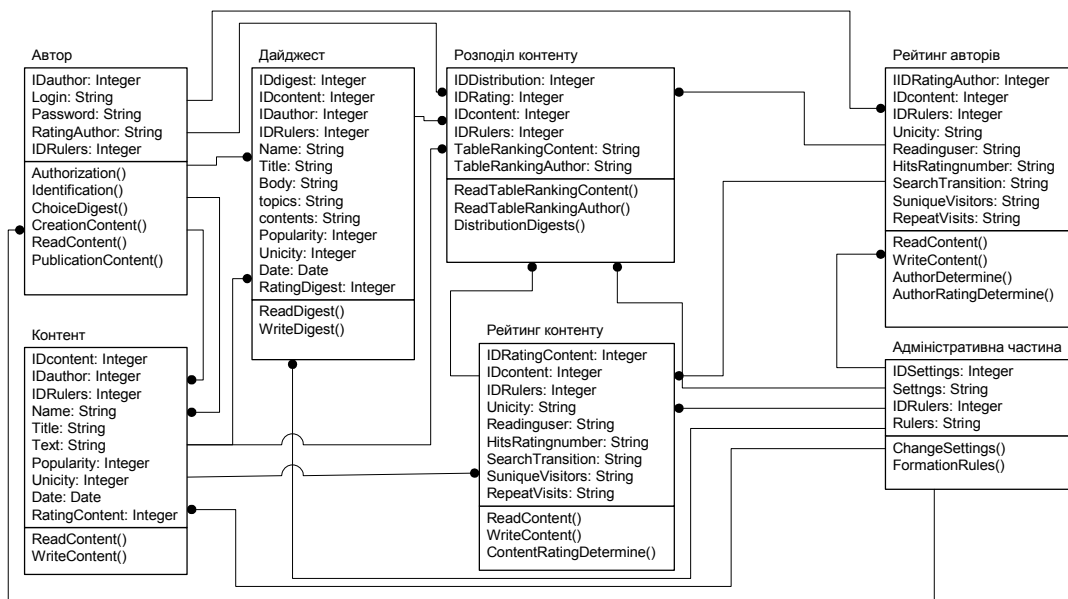


Рис. 8. Діаграма класів для процесу розподілу інформаційного продукту

Контент містить тему та дайджест. Підсистема розподілу контенту вибірково розсилає дайджести між авторами згідно з рейтингуванням якості їх роботи. Підсистема рейтингування оцінює якість роботи за сукупністю критеріїв, що забезпечує об'єктивність та стимулювання якісної роботи. Частка роботи модератора, а саме: розподіл однотипних даних, її сортування, оцінювання та аналіз, зменшується. Це зменшує застосування ресурсів, скорочує час створення контенту та покращує характеристики контенту завдяки об'єктивності оцінювання якості виконаних

завдань. Із відсотком унікальності тексту оцінюють якість роботи автора цього контенту та заносять отриману оцінку в таблицю рейтингів. Збільшення обсягу контенту призводить до більш точного оцінювання якості та продуктивності кожного автора контенту. Збільшення кількості критеріїв оцінки дозволяє охопити ширший спектр аспектів роботи автора/модератора. На рис. 9а подано діаграму діяльності модуля, тобто специфікацію виконуваної поведінки у вигляді координованого послідовного і паралельного виконання підлеглих елементів. Значення на

рис. 9б показують пріоритети елементів ієрархії з точки зору мети. Найбільше значення надається критерію «Унікальність», який об'єктивно відображає якість роботи модератора. Наступним за значенням є «Час читання», який відображає

ступінь зацікавленості користувачів у контенті. Критерії «Користувацька оцінка» та «Кількість звернень» мають меншу важливість, так як не користуються довірою через легкість фальсифікації їх показників.

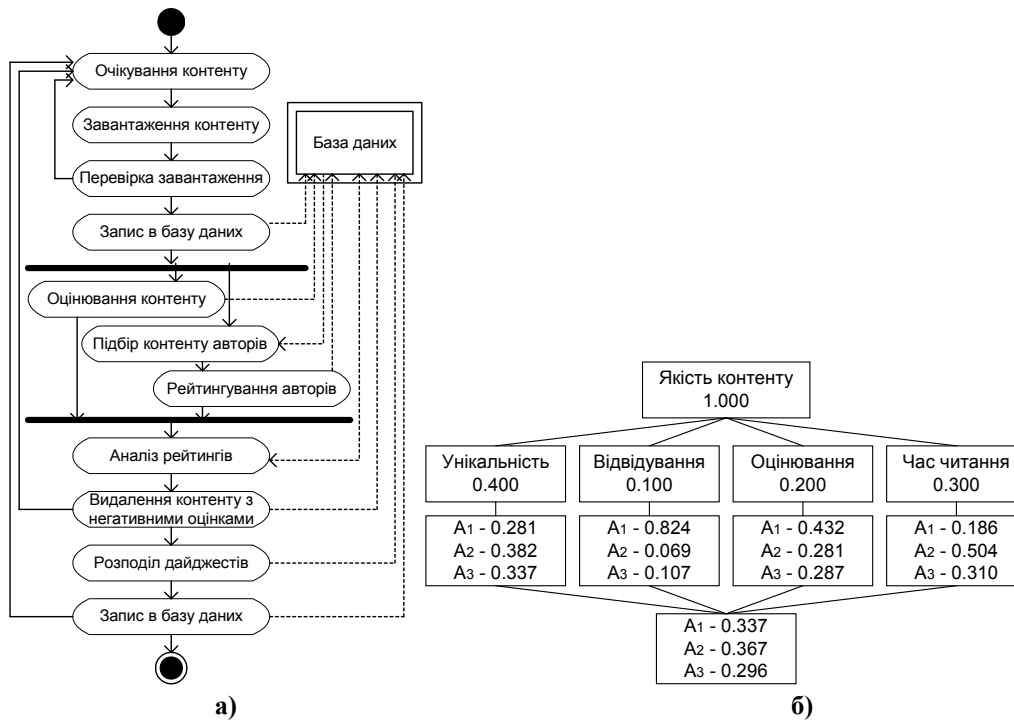


Рис. 9. Діаграма а) діяльності для процесу розподілу інформаційного продукту та б) ієрархії критеріїв для оцінювання рейтингу модераторів

Основні результати реалізовані в інтернет-проектах «Уголок» (Львів, vgoles.com.ua), «Victana» (Львів, victana.lviv.ua), «Татьяна» (Херсон, tatjana.in.ua), «Прес-Тайм» (Львів, presstime.com.ua), «AutoChip» (Вінниця, autochip.vn.ua), «Фотогалерея Висоцьких» (Львів, fotoghalereja-vysocjkykh.com),

«Курси валют» (Львів, kursyvalyut.com), «Добрий ранок» (Львів, dobryranok.com). У табл. 5 подано порівняльні характеристики результатів роботи розроблених систем за період часу 10-11.2012 р., отримані з Google Analytics (accounts.google.com), який забезпечує розширені можливості аналізу даних.

Таблиця 5

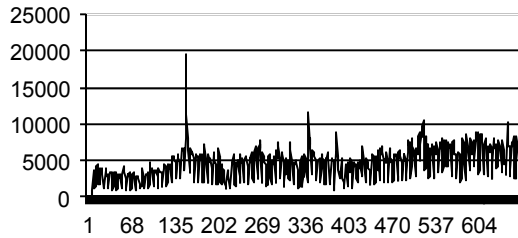
Порівняльні характеристики результатів роботи систем за період часу з 10.2012 р. до 11.2012 р.

Характеристика систем	Фотогалерея	Уголок	Татьяна	ПресТайм	AutoChip	Курси валют	Добрий ранок
Підсистема формування контенту	+/-	+	-	+/-	+	+	+/-
Підсистема управління контентом	+	+	+/-	+	+	+	+/-
Підсистема супроводу контенту	+/-	+	+/-	+	+/-	-	+/-
Відвідування	73	326 940	49	167 856	406	103	58
Унікальні відвідувачі	62	217719	21	123 756	326	42	7
Перегляд сторінок	136	562455	142	345 234	863	237	226
Сторінок/відвідування	1,86	1,72	2,90	1,45	2,13	1,67	3,90
Середній час відвідування (хв:с)	00:47	01:45	04:38	01:09	01:08	00:37	09:35
Показник відмовлень (%)	71,23	76,92	46,94	79,56	56,90	61,23	48,28
Нові відвідування (%)	80,82	51,83	36,43	45,65	77,59	90,87	12,07
Returning Visitor (%)	82,19	48,15	63,27	54,35	77,59	62,79	87,93
New Visitor (%)	17,81	51,85	36,43	45,65	22,41	37,21	12,07
Відвідувачі з України (%)	87,67	89,81	71,43	92,33	73,89	97,07	55,17
Відвідувачі з Росії (%)	2,74	2,55	24,49	6,27	17,00	1,05	43,10
Відвідувачі з США (%)	1,37	0,58	0,07	0,06	0,05	0,61	1,72
Пошуковий трафік (%)	69,86	36,03	73,47	60,05	88,67	59,03	43,10
Трафік переходів (%)	12,33	54,62	0	34,65	3,45	35,65	6,90
Прямий трафік (%)	17,82	9,21	26,53	5,25	7,88	5,32	50,00

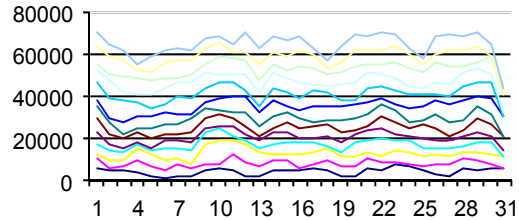
У табл. 5 подано порівняльні характеристики результатів роботи розроблених систем за період часу 10-11.2012 р., отримані з Google Analytics, який забезпечує розширені можливості аналізу даних. Google Analytics є безкоштовним сервісом ведення статистики відвідувань Web-сайтів, який

дозволяє оцінити трафік Web-сайту та ефективність маркетингових заходів, наприклад для газети «Уголок» (рис. 10).

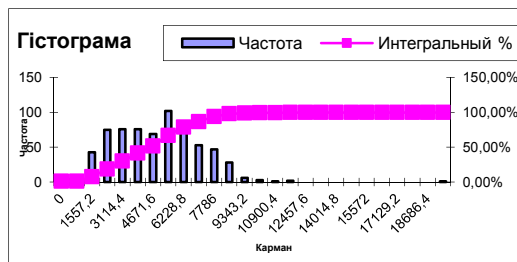
На рис. 11-12 подано результати роботи розробленої системи електронної контент-комерції «Доброго ранку, бухгалтер!» (<http://dobryjranok.com/>).



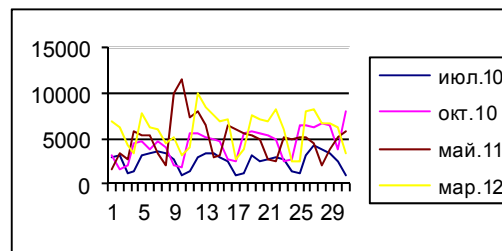
а)



б)

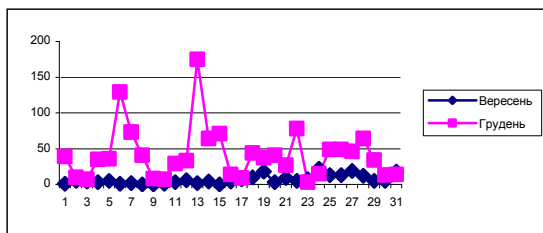


а)

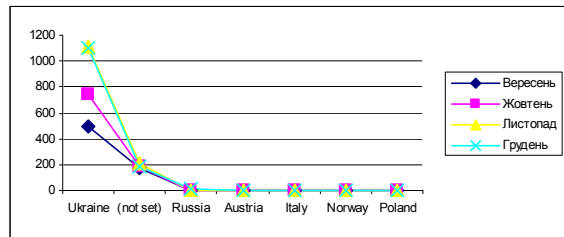


б)

Рис. 10. Розподіл кількості щомісячних відвідувачів за 2010-2012 рр.

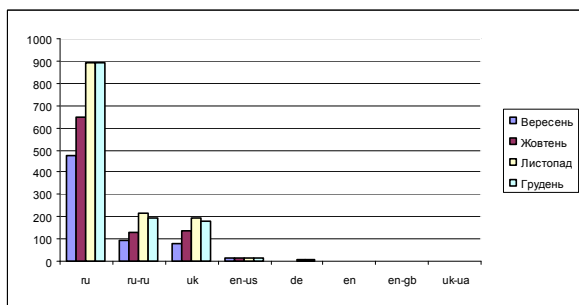


а)

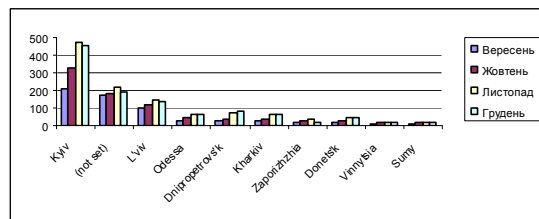


б)

Рис. 11. Розподіл а) відвідування ресурсу та б) відвідування з різних країн світу



а)



б)

Рис. 12. Розподіл а) мови трафіку та б) відвідування ресурсу з різних міст

Результати отримані з Google Analytics у вигляді графіків та діаграм, з яких випливає, що за наявності всіх етапів життєвого циклу контенту на ресурсі суттєво збільшується обсяг відвідувань та унікальних користувачів. Причини надмірності результатів стандартного контентного пошуку поділені на дві різні категорії: дублювання й

невідповідність. Істотним є те, що належність контенту до дублів має цілком об'єктивний характер і визначається автоматично на підставі формальних критеріїв.

Висновки та перспективи подальших наукових розвідок. У статті пояснено терміни, поняття, категорії, сутність, зміст та науково-

теоретичні засади систем контент-комерції. Досліджено нові теоретичні та практичні напрями та методологічні підходи організації електронної комерції. Проаналізовано закономірності, тенденції, явища, характерні для організації контент-комерції. Досліджено організаційні засади, структури, форми і методи організаційної діяльності систем електронної контент-комерції. Обґрунтовано принципи, правила, засоби організації, перспективні моделі, пріоритети організаційного розвитку проектування аналогічних систем. Визначено зміст організаційних дій, способи посилення організаційного впливу для підвищення ефективності комерційної діяльності і реалізації функцій електронної контент-комерції. Описано основні інструменти систем електронної контент-комерції.

Із позиції системного підходу проведено аналіз принципів та розроблено моделі опрацювання інформаційних ресурсів у системах електронної контент-комерції, що дало змогу розробити методи формування, управління та поширення комерційного контенту. Розроблено комплексний метод формування контенту, що дає можливість створити засоби опрацювання інформаційних ресурсів та реалізувати модуль автоматичного формування контенту. Створено оперативний метод управління контентом, що дає можливість реалізувати модуль управління комерційним контентом. Реалізовано комплексний метод поширення контенту, що дає можливість розробити модуль поширення комерційного контенту.

ЛІТЕРАТУРА

1. Берко А. Ю. Системи електронної контент-комерції. Монографія / А. Ю. Берко, В. А. Висоцька, В. В. Пасічник. – Л.: НУЛП, 2009. – 612 с.
2. Ермаков А. Е. Лингвистическая модель для компьютерного анализа тональности публикаций СМИ / А. Е. Ермаков, С. Л. Киселев // Компьютерная лингвистика и интеллектуальные технологии. – М.: Наука, 2005. – С. 282–285.
3. Иванов В. Ф. Контент-анализ: Методология и методика дослідження ЗМК / В. Ф. Иванов; наук. ред. А. З. Москаленко. – Київ, 1994. – 112 с.
4. Клифтон Брайан. Google Analytics: профессиональный анализ посещаемости веб-сайтов / Брайан Клифтон. – М.: ООО «И. Д. Вильямс», 2009. – 400 с.
5. Пасічник В. В. Математична лінгвістика. Книга 1. Квантитативна лінгвістика / В. А. Висоцька, В. В. Пасічник, Ю. М. Щербина, Т. В. Шестакевич. – Львів: «Новий Світ – 2000», 2012. – 359 с.
6. Сорока М. Б. Використання методу контент-аналізу при створенні автоматизованих інформаційних систем / М. Б. Сорока, Н. В. Танатар // Бібліотека. Наука. Культура. Інформація: Наукові праці НБУВ. – 1998. – № 1. – С. 318–322.
7. Федорчук А. Г. Контент-мониторинг информационных потоков / А. Г. Федорчук // Б-ки нац. акад. наук: пробл. функционирования, тенденции развития. – Киев, 2005. – № 3. [Електронний ресурс]. – Режим доступу: <http://www.nbu.gov.ua/articles/2005/05fagmip.html>.
8. Boiko Bob. Content Management Bible / Boiko Bob. – Hoboken, 2004. – 1176 p.
9. CM Lifecycle Poster / Content Management Professionals [Електронний ресурс]. – Retrieved 20 July 2010. – Режим доступу: <http://www.cmprosold.org/resources/poster/>.
10. Doyle Bob. Seven Stages of the CM Lifecycle [Електронний ресурс] / Bob Doyle // EcontentMag.com. – 2010. – Режим доступу: <http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=13554&AuthorID=155>.
11. Hackos JoAnn. Content Management for Dynamic Web Delivery / Hackos JoAnn. – Hoboken, NJ: Wiley, 2002. – 432 p.
12. Halvorson Kristina. Content Strategy for the Web / Halvorson Kristina. – Reading, Mass: New Riders Press, 2009. – 192 p.
13. McGovern G. Content Critical / G. McGovern, R. Norton. – Upper Saddle River, NJ: FT Press, 2001. – 256 p.
14. McKeever S. Understanding Web content management systems: evolution, lifecycle and market / S. McKeever // Industrial Management & Data Systems (MCB UP), 2003. – 103 (9). – P. 686–692.
15. Nakano Russell. Web content management: a collaborative approach / Nakano Russell. – Boston: Addison Wesley Professional, 2002. – 222 p.
16. Randy Woods. Defining a Model for Content Governance [Електронний ресурс] / Randy Woods. – 2010. – Режим доступу: www.contentmanager.net/magazine/article_785_defining_a_model_for_content_governance.html.
17. Rockley Ann. Managing Enterprise Content: A Unified Content Strategy / Rockley Ann. – Reading, Mass: New Riders Press, 2002. – 592 p.
18. Stone W. R. Plagiarism, Duplicate Publication and Duplicate Submission: They Are All Wrong! / W. R. Stone // IEEE Antennas and Propagation, Aug. 2003. – Vol. 45. – № 4.
19. The Content Management Possibilities Poster [Електронний ресурс] / Metatorial Services, Inc. – Retrieved 20 July 2010. – Режим доступу: <http://metatorial.com/pagea.asp?id=poster>.
20. Vysotska V. Commercial Web Content Lifecycle Model / V. Vysotska, L. Chyrun, L. Chyrun // CSIT'2011, LPNU. – Lviv, 2011. – P. 160–163.
21. Web Content Lifecycle [Електронний ресурс] / State Government, Victoria, Australia, 2005. – Режим доступу: <http://www.egov.vic.gov.au/victorian-government-resources/reports-victoria/web-content-lifecycle-and-content-management-roles/1-web-content-lifecycle.html>.

© Висоцька В. А., Чирун Л. Б., 2013

Дата надходження статті до редколегії 02.02.2013 р.

ВИСОЦЬКА Вікторія Анатоліївна – асистент кафедри інформаційних систем та мереж Національного університету «Львівська політехніка», м. Львів.

Коло наукових інтересів: Internet-технології, математична лінгвістика, електронна комерція, системи підтримки прийняття рішень, бази даних, інформаційні технології.

ЧИРУН Лілія Борисівна – старший викладач кафедри інформаційних систем та мереж Національного університету «Львівська політехніка», м. Львів.

Коло наукових інтересів: системи підтримки прийняття рішень, бази знань, сховища даних, простори даних, інформаційні технології в освіті.