

УДК 004.9

Коробчинський М. В.<sup>1</sup>, Чирун Л. Б.<sup>2</sup>, Висоцька В. А.<sup>3</sup>, Кондратьєв Є. О.<sup>4</sup>

<sup>1</sup>Д-р. техн. наук, доцент, професор Військово-дипломатичної академії імені Євгена Березняка, Київ, Україна

<sup>2</sup>Провідний спеціаліст інституту комп'ютерних наук та інформаційних технологій Національного університету «Львівська політехніка», Львів, Україна

<sup>3</sup>Канд. техн. наук, доцент, доцент кафедри «Інформаційні системи та мережі» Національного університету «Львівська політехніка», Львів, Україна

<sup>4</sup>Магістр кафедри «Інформаційні системи та мережі» Національного університету «Львівська політехніка», Львів, Україна

## ОСОБЛИВОСТІ ФОРМУВАННЯ ТА АНАЛІЗУ КОНТЕНТУ ІНТЕРНЕТ-ГАЗЕТИ МУЗИЧНИХ НОВИН

**Актуальність.** Негативні чинники у формуванні контенту ускладнюють процес пошуку необхідних даних при скануванні різних джерел інформації. Збільшення фізичного обсягу та зміна актуальності/динаміки контентних потоків (систематичне та нерегулярне оновлення) призводить до виникнення дублювання, інформаційного шуму та надмірності результатів пошуку контенту. Охоплення та узагальнення великих динамічних потоків контенту, які безперервно генерують в Інтернет-джерелах, вимагає якісно нових методів/підходів пошуку як контент-моніторинг.

**Мета.** Метою виконання роботи є розроблення методу формування та аналізу контенту Інтернет-газети музичних новин для активного залучення постійних користувачів інформаційного ресурсу новин та збільшення цільової аудиторії. Завданням є розроблення системи для розповсюдження музичних новин з використанням методу контент-моніторингу для формування та аналізу текстового контенту інформаційного ресурсу Інтернет-газети.

**Метод.** Вхідною інформацією для контент-моніторингу є текст на природній мові як послідовність символів, вихідна інформація – це таблиці розділів, речень і лексем аналізованого тексту. Контент-моніторинг є програмним засобом автоматизації знаходження важливих складових в потоках контенту. Це змістовний аналіз потоків контенту з метою постійного отримання необхідних якісних/кількісних зрізів на протязі не визначеного наперед проміжку часу. Складовою контент-моніторингу є контентний пошук та контент-аналіз тексту. Контент-аналіз призначений для пошуку контенту в масиві даних за змістовими лінгвістичними одиницями. Одиниця рахунку є кількісною мірою одиниці аналізу, що дозволяє реєструвати частоту (регулярність) появи ознаки категорії аналізу в тексті (кількість певних слів або їх поєднань, рядків, друкованих знаків, сторінок, абзаців, авторських аркушів, площа тексту тощо).

**Результати.** Розв'язано актуальну задачу дослідження і розроблення методів та засобів опрацювання інформаційних ресурсів інтелектуальної інформаційної систем формування контенту із використання розробленої класифікації, математичного та програмного забезпечення та узагальненої архітектури системи. Обґрунтовано необхідність розроблення методів і засобів опрацювання інформаційних ресурсів в інтелектуальних інформаційних систем формування контенту шляхом удосконалення архітектури системи з метою автоматизації процесів формування, управління та реалізації контенту.

**Висновки.** Використання контент-аналізу в моніторингу джерел даних в Інтернет автоматизує процес пошуку найбільш важливих компонент в потоці змісту, у виборі даних з цих джерел. Це усуває дублікати контенту, інформаційний шум, паразитні результати резервування контенту та багато іншого. Цей метод використовується на наступних етапах аналізу змісту для більш точних результатів для створення унікального контенту, який користується попитом.

**Ключові слова:** аналіз, контент-аналіз змісту інформаційних ресурсів, контент-аналіз, рейтингова оцінка, система управління контентом.

### НОМЕНКЛАТУРА

IS – інформаційна система;

OIP – опрацювання інформаційних ресурсів;

SEKK – система електронної контент-комерції;

X – множина вхідної інформації з Web-сайтів або від модераторів;

Formation – функція формування контенту;

C – множина контенту;

Management – функція управління контентом;

Realization – функція реалізації контенту;

Y – множина вихідної інформації;

Gathering – функція збирання інформації з джерел;

Formatting – форматування інформації, перетворення в множину контенту;

KeyWords – функція функцій виявлення ключових слів, понять;

Categorization – функція автоматичної рубрикації;

Backup – функція виявлення дублювання змісту контенту;

Dissemination – функція вибіркового поширення контенту;

Q – множина запитів;

R – функція формування та подання сторінки;

Edit – функція редагування та модифікації контенту;

Caching – функція формування кешу;

BuInfPortr – множина функцій формування інформаційних портретів;

BuDigest – множина функцій формування дайджестів;

IdThemTop – множина функцій виявлення тематичних сюжетів;

ConCorrTablConc – множина функцій побудови таблиць взаємозв'язку понять;

CalRankConc – множина функцій розрахунку рейтингів понять;

C<sub>p</sub> – скінченна множина контенту інформаційних портретів;

$C_D$  – скінченна множина контенту дайджестів;

$C_T$  – скінченна множина контенту тематичних сюжетів;

$C_C$  – скінченна множина контенту таблиць взаємозв'язку понять;

$C_R$  – скінченна множина контенту рейтингів понять.

### ВСТУП

Бурхливий розвиток електронної комерції сприяв виникненню потреби створення шаблону інформаційного ресурсу відповідно до потреб користувача для його задоволення у повній мірі [1]. Наприклад, більшість інформаційних ресурсів з музичних новин світу чи країни є не адаптивними та не зрозумілими для використання кінцевим користувачем, часто нагромаджені надлишковим контентом та рекламою. Більшість з них не володіють адаптивною версткою для різних електронних носіїв. Існуючі та розповсюджені моделі систем або їх шаблони для реалізації СЕКК не вирішують низку проблем ОІР згідно потреб кінцевого користувача або не враховує специфіки діяльності підприємства чи людини, специфіки галузі науки чи промисловості тощо [2]. Тому з'являється попит на вдосконалення процесу ОІР та розроблення загальної архітектури СЕКК з використанням математичної лінгвістики та контент-аналізу текстових масивів даних [3]. Проблема полягає у відсутності загального стандартизованого підходу до проектування і розроблення архітектури СЕКК та процесу ОІР в даній системі [4]. Відсутність загальної класифікації приводить до проблем визначення та формування уніфікованих методів ОІР та реалізації відповідних програмних засобів [5]. Це обґрунтовує мету, актуальність, доцільність та напрям дослідження.

### 1 ПОСТАНОВКА ЗАДАЧІ

Метою роботи є розроблення шаблону інформаційного ресурсу з автоматичним наповненням контенту відповідно до вимог користувача для вирішення таких завдань: автоматичне формування комерційного контенту; автоматичне збирання та створення контенту; автоматичне форматування контенту; виявлення ключових слів та понять; рубрикація контенту; виявлення дублювання контенту; формування дайджесту контенту; вибіркоче поширення контенту [5]. Формальна модель СЕКК – це шістка

$$S = \langle X, Formation, C, Management, Realization, Y \rangle.$$

Формальна модель формування інформаційних ресурсів це

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

Моделі управління інформаційними ресурсами.

1. Генерація сторінок за запитом відбувається на основі зв'язки «Модуль редагування → База даних → Модуль представлення». Формальна модель генерації сторінок за запитом –

$$Management_Q = \langle X, C, Q, R, Edit, Y \rangle.$$

2. Генерація сторінок при редагуванні. При внесенні змін до змісту сайту створюють набір статичних сторінок.

При такому способі не враховується інтерактивність між відвідувачем і вмістом сайту. Формальна модель системи генерації сторінок при редагуванні –  $Management_E = \langle C, Edit, Y \rangle$ . Процес формування сторінок описується функцією  $Edit$  вигляду:  $\bar{y}(t) = Edit(\bar{c}, Weight, t)$ .

3. Змішаний тип поєднує в собі переваги перших двох типів. Може бути реалізований шляхом кешування – модуль представлення генерує сторінку один раз, надалі вона в декілька разів швидше підвантажується з кешу. Кеш може оновлюватися як автоматично, по закінченню деякого терміну часу або при внесенні змін до певних розділів сайту, так і вручну по команді адміністратора. Інший підхід – збереження певних інформаційних блоків на етапі редагування сайту і збірка сторінки з цих блоків при запиті відповідної сторінки користувачем. Формальна модель системи змішаного типу –  $Management_M = \langle X, C, Q, R, Edit, Caching, Y \rangle$ .

Формальна модель реалізації інформаційних ресурсів – це

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

Скінченна множина контенту

$$C = \langle C_P, C_D, C_T, C_C, C_R \rangle$$
 є адекватною та повною.

### 2 ОГЛЯД ЛІТЕРАТУРИ

Однією з головних особливостей нашого часу є постійне зростання темпів виробництва інформації [3]. Цей процес об'єктивний і в цілому, безумовно, позитивний. Однак на сьогоднішній день людство зустрілося з парадоксальною, на перший погляд, ситуацією: прогрес в області виробництва інформації веде до зниження загальної рівня інформованості [3–5]. Окрім збільшення обсягів інформації до масштабів, які унеможливають її безпосереднє опрацювання, виник цілий ряд специфічних проблем, пов'язаних з швидким розвитком інформаційних технологій [6]. Ситуація різкого зростання темпів виробництва інформації породила ряд проблем: непропорційне зростання інформаційного шуму через слабку структурованість інформації; поява паразитної інформації (одержуваної в якості додатків); невідповідність формально релевантної інформації (тематично відповідної) дійсним потребам її споживачів; багаторазове дублювання інформації (типовий приклад – публікація одного повідомлення в різних виданнях) [7]. Внаслідок перелічених обставин, традиційні інформаційно-пошукові системи поступово стали втрачати свою актуальність. Причина цього криється не стільки в фізичних обсягах інформаційних потоків, скільки в їх динаміці, тобто в постійному систематичному відновленні інформації, яка далеко не завжди має очевидну регулярність. Охоплення та узагальнення великих динамічних інформаційних потоків, які безперервно генеруються в засобах масової інформації, вимагає якісно нових підходів [8]. Вихід може бути знайдений тільки в засобах автоматизації виявлення

найбільш важливих складових в інформаційних потоках [9]. Саме тому в останні роки все частіше стали використовувати системи моніторингу ресурсів [10], які тісно пов'язані з контент-аналізом [11]. Цей перспективний напрямок отримав назву контент-моніторинг [12]. Його поява була викликана, насамперед, завданнями систематичного відстеження тенденцій і процесів в інформаційному середовищі, яке постійно оновлюється. Під контент-моніторингом найчастіше розуміють змістовний аналіз інформаційних потоків з метою отримання необхідних якісних і кількісних зрізів, який ведеться постійно протягом невизначеного заздалегідь проміжку часу [12]. Найважливішою методологічною складовою контент-моніторингу є контент-аналіз та життєвий цикл контенту [13–16].

### 3 МАТЕРІАЛИ І МЕТОДИ

Складовою контент-моніторингу є контентний пошук та контент-аналіз тексту [17]. Контент-аналіз призначений для пошуку контенту в масиві даних за змістовими лінгвістичними одиницями [18–20]. Одиниця рахунку є кількісною мірою одиниці аналізу, що дозволяє реєструвати частоту (регулярність) появи ознаки категорії аналізу в тексті (кількість певних слів або їх поєднань, рядків, друкованих знаків, сторінок, абзаців, авторських аркушів, площа тексту тощо).

Алгоритм 1. Контент-аналіз текстового комерційного контенту.

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

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

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

Крок 3. Фільтрування згідно сформованого набору критеріїв контентного потоку та зберігання ідентифікованого релевантного контенту.

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

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

Етап. 4. Виділення одиниць рахунку аналізу текстового контенту.

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

Крок 2. Модератор на основі аналізованого контенту висуває одиниці рахунку, наприклад, протяжність текстів;

площа тексту, заповнена змістовними одиницями; кількість рядків (абзаців, знаків тощо); розмір/вид файлу; кількість рисунків з певним змістом тощо.

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

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

Крок 2. Статистичні розрахунки зрозумілості та атрактивності контенту.

Етап. 6. Розроблення інструменту контент-аналізу.

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

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

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

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

Етап. 8. Розроблення кодувальної матриці контент-аналізу.

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

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

Етап. 9. Проведення аналізу тексту згідно створених кодувальних матриць.

Етап. 10. Інтерпретація результатів. Виявляють і оцінюють характеристики контенту з врахуванням підрахованих коефіцієнтів за певний період часу на визначену категорію.

Застосування контент-аналізу для моніторингу Інтернет-джерел даних дозволяє автоматизувати процес знаходження найбільш важливих складових в потоках контенту при відборі даних з цих джерел шляхом використання частотно-лінгвістичних та евристичних методів, технології Text Mining, а також ідеології Web-сервісів, базуючись при цьому на інформаційних ресурсах, накопичених в WWW першого покоління [10–20]. Це усуває дублювання контенту, інформаційний шум, паразитичний контент, надмірність [21–23].

### 4 ЕКСПЕРИМЕНТИ

Модуль формування контенту забезпечує збирання інформації з різноманітних Web-сайтів та її форматування; виявлення ключових слів і понять контенту; автоматичну рубрикацію контенту; виявлення дублювання змісту контенту; вибіркоче поширення контенту. Головне завдання модуля управління контентом: формування,

ротація баз даних і забезпечення доступу до неї; формування оперативних і ретроспективних баз даних; персоналізація роботи користувачів, збереження персональних запитів користувачів і джерел, ведення статистики роботи; забезпечення пошуку в бази даних; генерація вихідних форм; інформаційна взаємодія з іншими базами даними. Існує три моделі управління контентом: генерація сторінок за запитом, генерація сторінок при редагуванні; змішаний тип. Генерація сторінок за запитом відбувається на основі зв'язки: Модуль редагування ⇒ База даних ⇒ Модуль подання. Генерація сторінок при редагуванні полягає у створенні набору статичних сторінок при внесенні змін до сайту (враховується інтерактивність між відвідувачем і вмістом сайту). Модель реалізації контенту забезпечує формування інформаційних портретів; формування дайджестів; побудова таблиць взаємозв'язку понять та розрахунок їх рейтингів, виявлення нових подій та тематичних сюжетів, їхнє відстеження кластеризація. Процес проектування та створення ІС формування контенту за допомогою Інтернет-маркетингу є інтерактивним і протікає від аналізу, проектування, розроблення плану до створення прототипу і експериментальних випробувань, починаючи з формування специфікацій, верстки, створення шаблону контенту, формування контенту та його подальше розміщення згідно структури сайту (рис. 1).

Завдяки моніторингу продуктивності та успіху користувачів визначають одну із вимог – доступ до даних про конвергенції або цільову аудиторію. За зворотною реакцією користувачів формують інформаційну архітектуру сайту. Основні класи користувачів сайту визначають дизайн сайту і процес ухвалення рішень. Визначивши важливу інформацію та її зв'язок із основними класами ко-

ристувачів, створюють архітектуру контенту сайту, його ієрархію, способи подання і взаємодії кожного класу користувачів з контентом. На рис. 2 подана типова схема взаємодії модулів, а на рис. 3 – схема модуля управління контентом.

Модуль формування контенту реалізується у вигляді комплексів контент-моніторингу зі збирання контенту з джерел за наперед визначеними методами (рис. 4), і забезпечують створення баз даних відповідно до потреб споживачів. Основні етапи формування контенту: збирання/створення контенту, систематизація контенту та поширення контенту. Аналіз та управління контентом сайту, їхнє моделювання є одним з найбільш інформативних методів кількісного вивчення динаміки окремих тематичних напрямів та проведення технічного аналізу сайту. За зміною величин управління контентом визначають швидкість розвитку окремих тематичних напрямів або всього контентного простору.

При генерації сторінок модуль подання генерує сторінку із змістом при запиті на нього на основі інформації з бази даних (рис. 5a). У режимі діалогового доступу до бази даних забезпечується перегляд, пошук і відображення контенту, надається можливість звертання до оригіналів контенту в Інтернет (рис. 5b). Інформація в базі даних змінюється за допомогою модуля редагування (рис. 5c). Сторінки наново створюються сервером при кожному запиті, збільшує навантаження на ресурси системи (рис. 5d). Навантаження знижується при використанні засобів кешування в сучасних Web-серверах. Задача повнотекстового пошуку у великих масивах контенту є неефективною. Проблема точності вирішує пошук в анотованому контенті. Замість пошуку у повному контенті доцільно вести пошук за анотаціями – пошукових образах контенту.

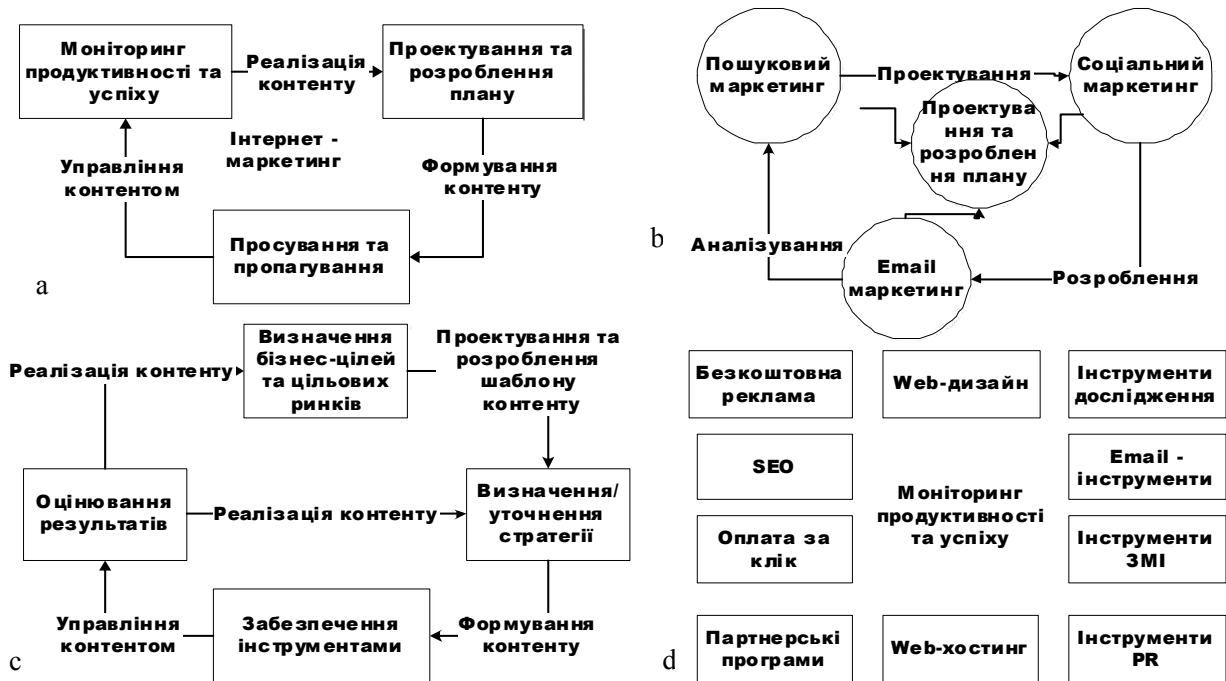


Рисунок 1 – Схема: а – Інтернет-маркетингу, б – проектування плану, с – реалізація контенту, d – моніторингу продуктивності та успіху



Рисунок 2 – Схема взаємодії модулів



Рисунок 3 – Загальна схема модуля управління контентом

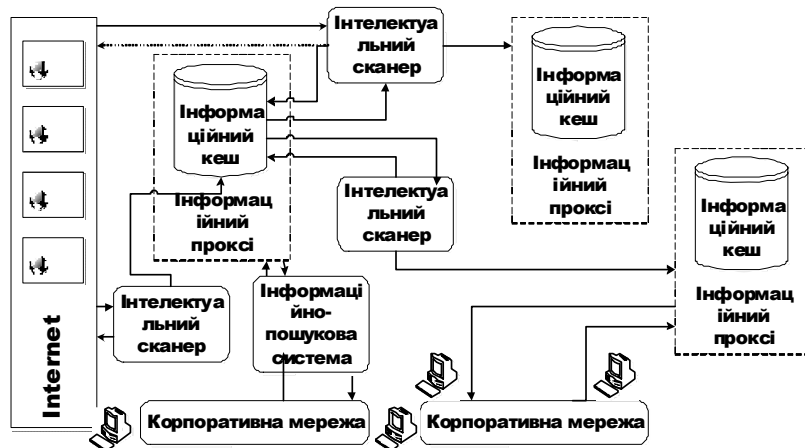


Рисунок 4 – Етап збирання та створення контенту

Модуль реалізації контенту – це комплексні процеси аналізу контенту. Актуальність розроблення загальної архітектури модуля реалізації контенту полягає у необхідності отримувати оперативні і об’єктивні оцінювання рівня конкуренції на сегменті фінансового ринку контенту; оцінювати рівень конкурентів та міри їх конкурентоспроможності на ринку. Отримані дані враховуються при створенні або оновленні Web-порталу. Модуль реалізації контенту

виконує процес підвищення рейтингу ІС формування контенту із інтерактивним інтерфейсом та вирішує наступні задачі: збирання маркетингової інформації з поширення контенту; підтримка діалогу з кінцевим споживачем контенту; формування каталогу контенту; інформаційна підтримка кінцевого споживача контенту; реєстрація кінцевого споживача контенту (рис. 6). Модуль реалізації контенту поданий на рис. 7–8 та в табл. 1–3.

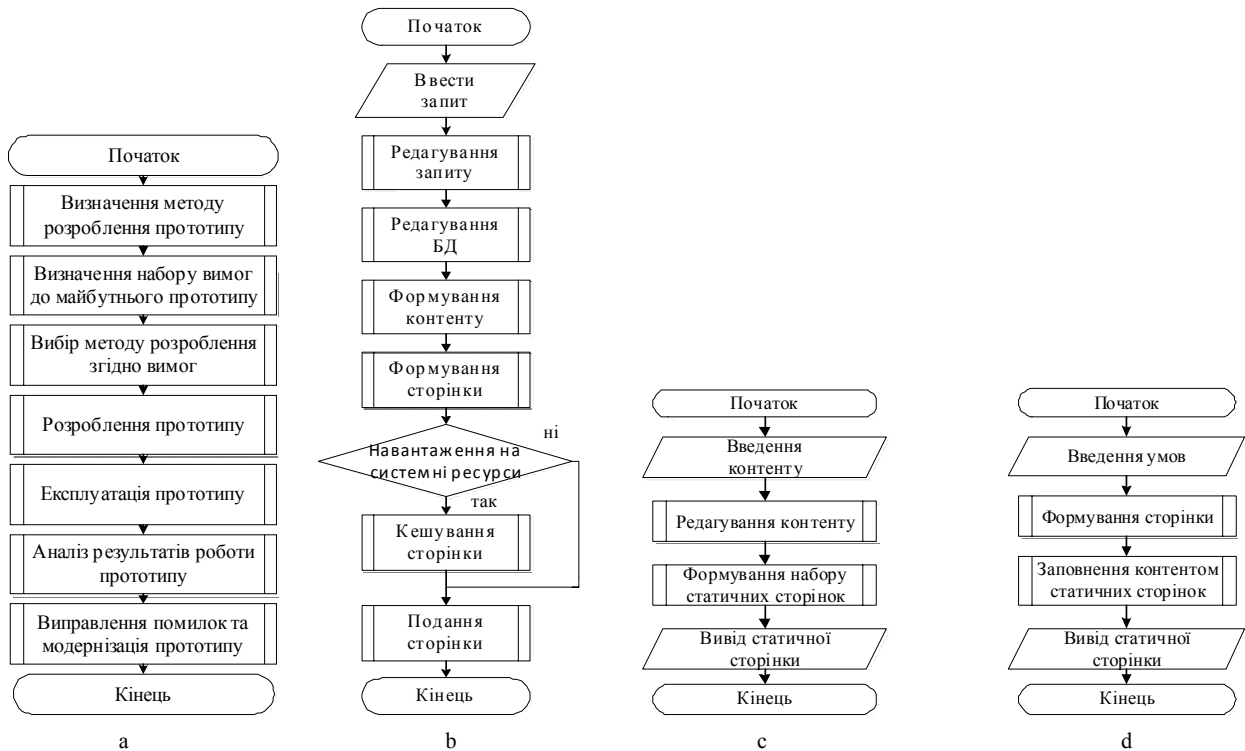


Рисунок 5 – Процес: а – створення прототипу, б – генерації сторінок за запитом, с – генерації та d – заповнення сторінок при редагуванні

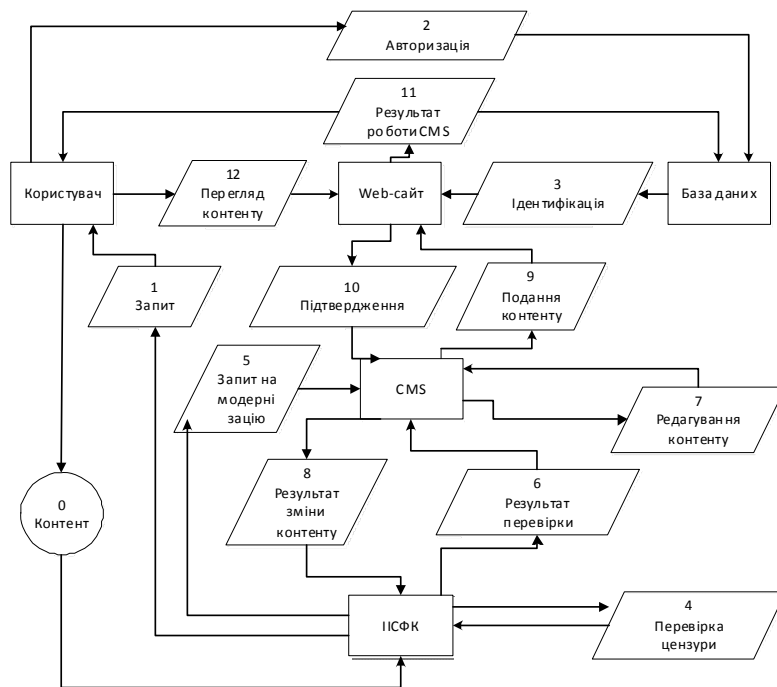


Рисунок 6 – Схема модерзації контенту

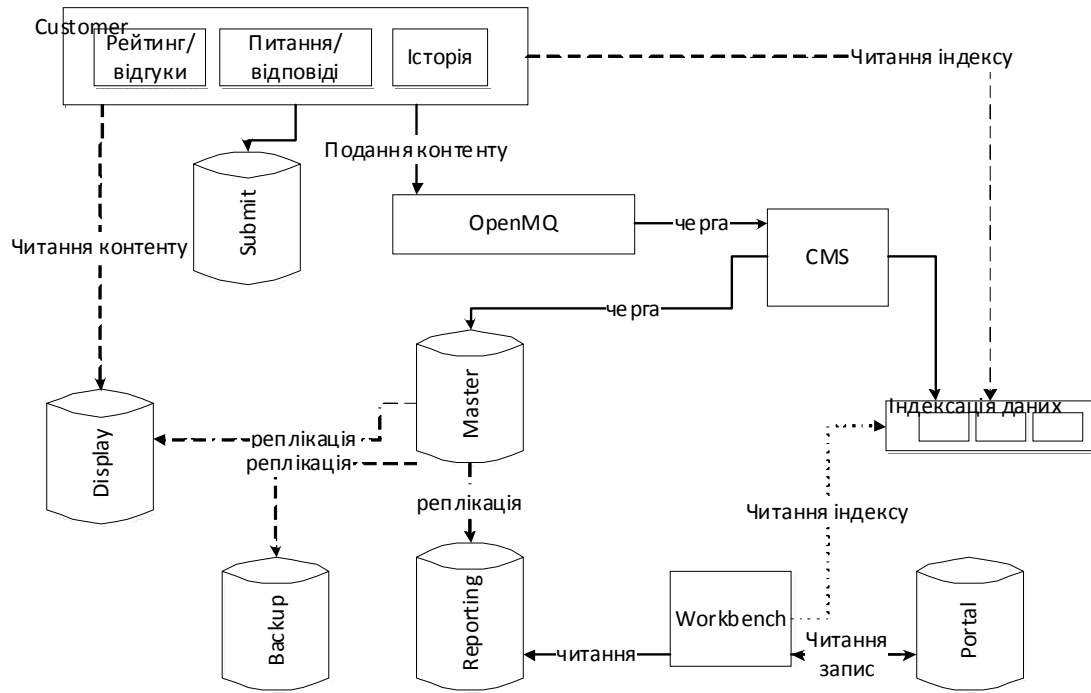


Рисунок 7 – Схема потоків даних в модулі реалізації контенту

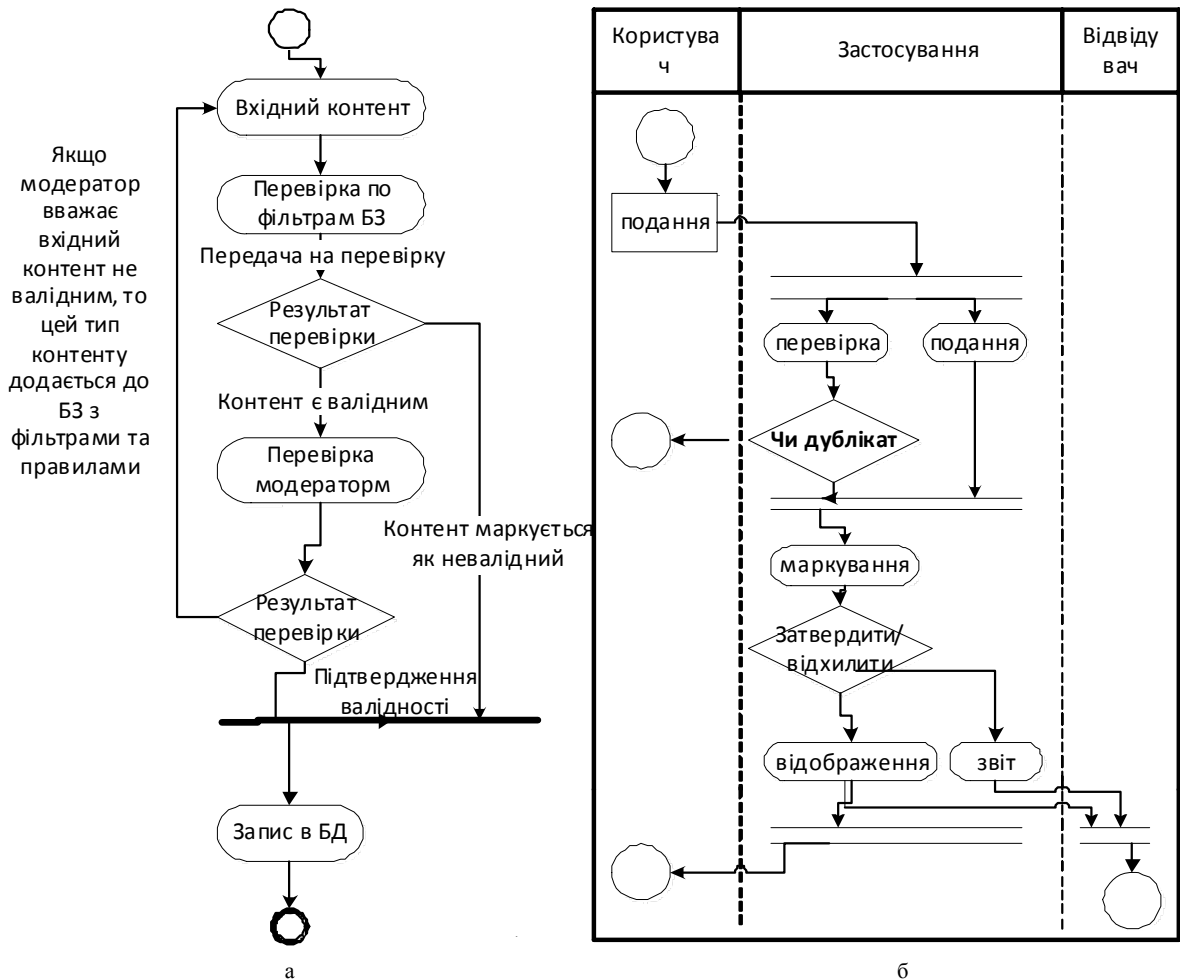


Рисунок 8 – Узагальнення контенту: а – алгоритм та б – процес

Таблиця 1 – Основні компоненти модуля реалізації контенту

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

Таблиця 2 – Бази даних модуля реалізації контенту

Назва	Характеристика
Master	Головна БД, яка є основною ланкою системи і для розподілу навантаження реплікує дані на дочірні бази.
Display	Дочірня БД від Master, з якою працює сайт клієнта на зчитування дозволеного на публікацію контенту.
Submit	Попереджує дублювання інформації на першому етапі.
Reporting	На основі цієї БД формуються статистичні звіти для клієнтів
Backup	Додаткове СД
Portal	БД призначена для роботи з аплікацією Workbench.

Таблиця 3 – Правила модерзації контенту із самонавчанням

Фільтр	Характеристика	Режим роботи
Нецензурної лексики	Перевірка за словником та заборона при TRUE	Напівавтоматичний, при FALSE перевіряє модератор, словник періодично поповнюється модератором.
За кількістю символів	Порівняння кількості символів із лімітом та заборона при FALSE	Автоматичний, ліміт визначає адміністратор або власник сайту.
За URL-посиланнями	Визначення наявності URL, перевірка за списком та заборона при TRUE	Напівавтоматичний, у разі FALSE остаточно перевіряє модератор, список періодично поповнюється модератором.
За «blacklist» списком	Перевірка за списком та заборона при TRUE	Напівавтоматичний, у разі FALSE остаточно перевіряє модератор, список періодично поповнюється автоматично у разі перевищення ліміту.
За посиланням на конкурентів	Перевірка за списком та заборона при TRUE	Напівавтоматичний, у разі FALSE остаточно перевіряє модератор, список періодично поповнюється модератором.
За IP адресами	Перевірка за списком та заборона у разі TRUE	Автоматичний, список періодично поповнюється автоматично при перевищенні ліміту, ліміт визначає адміністратор/власник сайту.
За ID користувачів	Перевірка за списком та заборона у разі TRUE	Автоматичний, список періодично поповнюється автоматично при перевищенні ліміту, ліміт визначає адміністратор/власник сайту.

## 5 РЕЗУЛЬТАТИ

Для повного функціонування інтелектуальної ІС формування контенту інтернет-газет характерна складна система взаємозв'язаних операцій (рис. 9–10), методів, прийомів, які формують технологічний процес формування контенту, який складається із п'яти фаз: ознайомлення з ІС; пошук контенту; отримання інформації про послугу/товар; вибір категорії; доставка контенту е-поштою або в режимі on-line.

Система поширення контенту поділяється на дві підсистеми: з вільним доступом; адміністративна підсистема. Для користувачів існує частина з вільним доступом. В ній реалізовані такі сервіси як вибір контенту за визначений період часу з початку контентного наповнення за допомогою календаря. Зручний рубрикатор дозволяє вибирати контент за певною категорією. За допомогою пошуку здійснюють пошук в бази даних за ключовими словами. Можна в зручному вигляді переглянути анонси на події, які мають відбутись. Адміністрування системи здійснюється в адміністративному модулі інтелектуальної ІС формування контенту інтернет-газет, доступ до якого обмежений і реалізований за допомогою логіна і пароля. Тут вносяться коригування в структуру, додається, редагується або видаляється контент, встановлюється

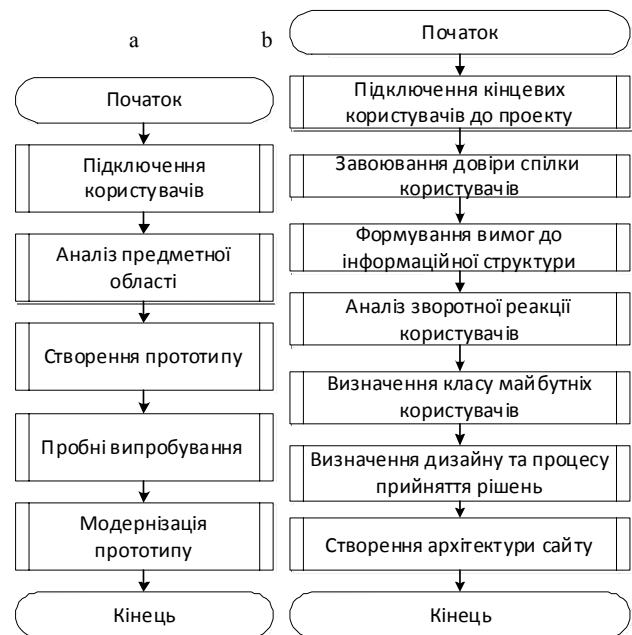


Рисунок 9 – Процес: а – проектування СЕКК; б – підключення користувачів



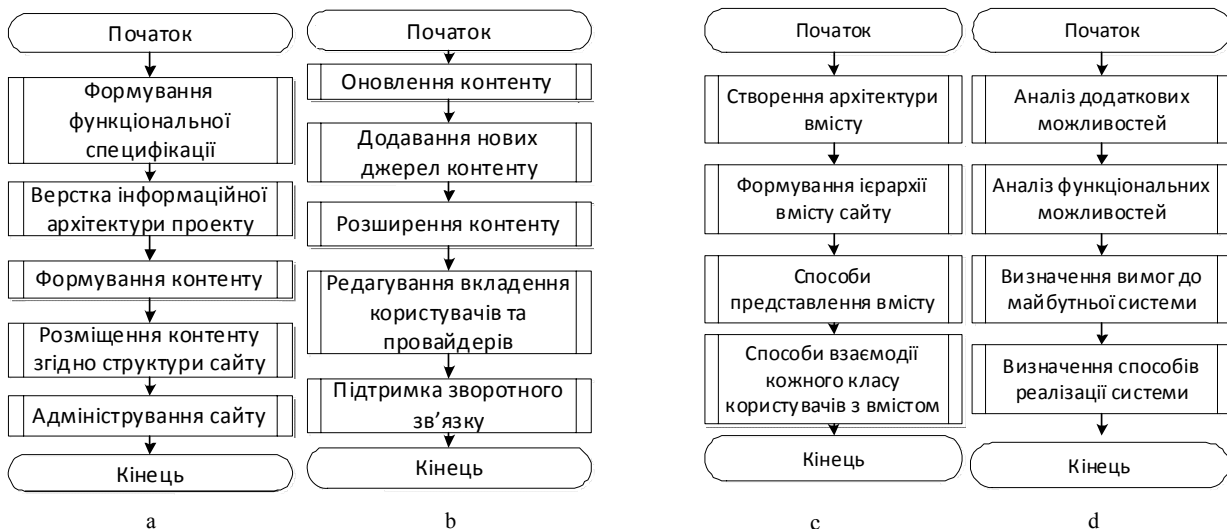


Рисунок 10 – Процес: а – організації контенту; б – адміністрування сайту; с – створення; д – аналізу архітектури сайту

контент дня. Змінюється зміст статичних сторінок, відбувається розсилка листів з контентом. Реалізується зручна система для додавання нових клієнтів з вибором на певні категорії та встановлення часових має дії доступу, який блокується автоматично після закінчення періоду дії [3]. Структура бази даних з описом полів подана в табл. 4 та наведена на рис. 11.

Структурна схема Інтернет-газети розподілена на три частини: mods, libraries та templates. Опис призначення

скриптів з даних груп подано в табл. 5. Група скриптів libraries (бібліотеки) – це класи призначені для роботи з певними об'єктами. Вони поділяються на три категорії: створені автором, створені іншими розробниками та скрипти на мові JavaScript. Звернення до цих скриптів відбувається з групи скриптів mods.

## 6 ОГОВОРЕННЯ

В табл. 6 наведені приклади найпопулярніших інтернет-газет. Кожна з яких відрізняється технологіями та підходом побудови ІС. Вони наповнюються контентом через модератора, що збільшує витрати та час на формування нового контенту. В них відсутня ідентифікація постійного користувача, що дозволило б виконувати підбір контенту за вподобанням користувача. Кожен з вище зазначених сайтів не є вузько спеціалізованим, що зменшує можливість пошуку потрібної інформації. На сайтах № 1–4, 7–9 реалізовано систему пошуку інформації по сайту, що значно зменшує витрати часу, на сайтах № 5–6 така система відсутня, що зменшує зручність його використання. На сайтах № 2, 8 реалізовано адаптивну верстку, що є дуже важливим підходом побудови таких сайтів. Сайти № 1–9 наповнюються контентом за допомогою модератора, тобто інформація відображається не вчасно, а значить може бути не актуальною на період перегляду. Автоматичне наповнення контентом сайту значно покращить його інформаційну наповненість. Сайти № 1–6 перенасичені рекламою, що не стосується музичної теми.

## ВИСНОВКИ

Розроблена формальна модель ІС формування контенту, що дало змогу розробити життєвий цикл контенту, узагальнену типову архітектуру системи та уніфіковані методи ОІР. Вдосконалено загальну архітектуру ІС формування контенту, відмінну від існуючих наявністю модулів ОІР. Розроблено комплексний метод формування контенту, оперативний метод управління контентом та комплексний метод реалізації контенту для досягнення ефекту від роботи на рівні розробника системи. Розроблено загальні рекомендації з проектування архітектури системи, відмінні від існуючих більшою деталізацією етапів та наявністю модулів ОІР, що дають змогу ефек-

Таблиця 4 – Структура бази даних

Назва	Призначення
clients	зберігаються відомості про клієнта;
deliver	містить коди останніх відправлених новин та номера розсилок;
news	в цій таблиці розташовані всі новини;
option	розміщені опції;
pages	інформація зі статичних сторінок;
statistics	зберігається статична інформація про відвідуваність;
tree	містить структуру у вигляді ієрархічного дерева;
users	інформація про користувачів адміністративної частини;

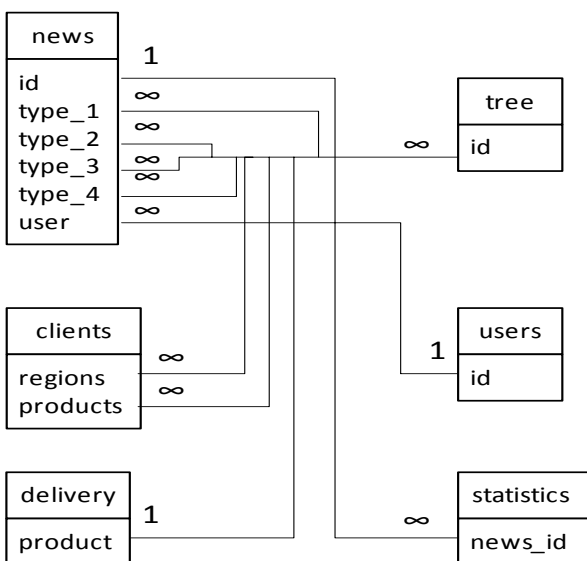


Рисунок 11 – Структура бази даних

Таблиця 5 – Опис призначення з групи

скриптів	Назва	Призначення
mods	mod_index	взаємодія користувача з загальнодоступною інформацією;
	mod_admin	адміністрування;
	mod_admin_news_edit	забезпечує маніпуляції з контентом;
	mod_admin_news_search	забезпечує пошук контенту в архіві за критерієм відбору;
	mod_admin_news_day	дозволяє вибрати контент за період часу;
	mod_admin_delivery	забезпечує розсилку контенту;
	mod_admin_users	забезпечує маніпуляції з даними про адміністрування;
	mod_admin_pages	забезпечує редагування статичних даних;
libraries	mod_admin_test	в цьому модулі відбувається відлагодження нових скриптів.
	class_lib	набір різноманітних допоміжних функцій;
	class_news	клас для маніпуляції над об'єктом «контент»;
	class_clients	клас для маніпуляції над об'єктом «клієнт»;
	class_users	клас для маніпуляції над об'єктом «адміністратор»;
	class_pages	клас для маніпуляції над об'єктом «статичні сторінки»;
	class_tree	клас для маніпуляції над об'єктом «дерево»;
	class_mail	клас для маніпуляції над об'єктом «пошта»;
	class_message	клас для маніпуляції над об'єктом «повідомлення»;
	class_mysql	клас для маніпуляції над об'єктом «БД»;
	class_debug	клас для маніпуляції над об'єктом «відлагодження»;
	class_error	клас для маніпуляції над об'єктом «помилки»;
	jsalendar	календар.
templates	mod_index	шаблон для роботи з загальною частиною;
	mod_admin	шаблон для роботи з адміністративною частиною;
	html	різноманітні шаблони спільні складові.

Таблиця 6 – Інтернет газети музичних новин

№	Назва	Посилання
1.	Mojo	<a href="http://www.mojo4music.com">http://www.mojo4music.com</a>
2.	QMAGAZINE	<a href="http://www.qthemusic.com">http://www.qthemusic.com</a>
3.	Pitchfork	<a href="http://pitchfork.com">http://pitchfork.com</a>
4.	SPIN	<a href="http://www.spin.com">http://www.spin.com</a>
5.	FACT	<a href="http://www.factmag.com">http://www.factmag.com</a>
6.	Navsi100.com	<a href="http://beta.navsi100.com">http://beta.navsi100.com</a>
7.	POK-OKO	<a href="http://www.rock-oko.com">http://www.rock-oko.com</a>
8.	MUZNEWS	<a href="http://muzsweet.com">http://muzsweet.com</a>
9.	13NOT	<a href="http://13notes.info">http://13notes.info</a>

тивно просто реалізувати OIP на рівні розробника системи. Розроблено архітектури модулів системи для реалізації життєвого циклу контенту. Розроблено та впроваджено прикладне програмне забезпечення формування, управління та реалізації контенту для досягнення ефекту від роботи на рівні власника та користувача інтелектуальної інформаційної систем формування контенту.

**ПОДЯКИ**

У статті розв'язана науково-практична формування та аналізу текстового контенту Інтернет-шазети музичних нових для збільшення цільової аудиторії. Роботу виконано в рамках спільних наукових досліджень кафедри інформаційних систем та мереж Національного університету «Львівська політехніка» на тему «Дослідження, розроблення і впровадження інтелектуальних розподілених інформаційних технологій та систем на основі ресурсів баз даних, сховищ даних, просторів даних та знань з метою прискорення процесів формування сучасного інформаційного суспільства», а також Військово-дипломатичної академії імені Євгена Березняка. Наукові дослідження провадилися також в рамках ініціативної тематики досліджень кафедри ІСМ Національного університету «Львівська політехніка» на тему «Розроблення інтелектуальних розподілених систем на основі онтологічного підходу з метою інтеграції інформаційних ресурсів».

**СПИСОК ЛІТЕРАТУРИ**

1. Джерк Н. Разработка приложений для электронной коммерции / Н. Джерк, Санкт-Петербург : Питер. – 2001 р.
2. Пелецишин А.М. Методи побудови Ефективних Web-систем / А. М. Пелецишин // Інформаційні системи та мережі. Вісник НУЛП. – 2012. – № 464. –С. 240–254.
3. Берко А. Системи електронної контент-комерції : монографія / А. Берко, В. Висоцька, В. Пасічник. – Львів : Видавництво НУ «Львівська політехніка», 2009. – 612 с.
4. Methods based on ontologies for information resources processing / [V. Lytvyn, V. Vysotska, L. Chyrun, D. Dosyn] // LAP Lambert Academic Publishing, Germany. – 324 p.
5. Висоцька В. А. Методи і засоби опрацювання інформаційних ресурсів в системах електронної контент-комерції : автореферат дисертації на здобуття наукового ступеня к.т.н. : 05.13.06 / В. А. Висоцька; НУ «Львівська політехніка». – Львів, 2014. – 27 с.
6. Ландэ Д. Основы моделирования и оценки электронных информационных потоков / Д. Ландэ, В. Фурашев, С. Брайчевский, О. Григорьев. – К. : Инжиниринг, 2006. – 348 с.
7. Ландэ Д. В. Некоторые методы анализа новостных информационных потоков / Д.В. Ландэ // ИКВТ-2005. – Вып. 93. – Донецк : ДонНТУ, 2005. – С. 277–287.
8. Ландэ Д. В. Сканер системы контент-мониторинга / Д. В. Ландэ // Открытые информационные и компьютерные интегрированные технологии. – Харьков : НАКУ «ХАИ», 2005. – Вып. 28. – С. 53–58.
9. Большакова Е. И. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / [Е. И. Большакова, Э. С. Клышинский, Д. В. Ландэ и др.] – М. : МИЭМ, 2011. – 272 с.
10. Григорьев А. Н. Адаптивный интерфейс уточнения запросов к системе контент-мониторинга / А. Н. Григорьев, Д. В. Ландэ // XI международная научно-практическая конференция. – К. : УкрИНТЭИ, 2005. – С. 17–20.
11. Иванов В. Ф. Контент-анализ / В. Ф. Иванов. – К., 1994. – 112 с.
12. Математична лінгвістика / [В. Висоцька, В. Пасічник, Ю. Щербина, Т. Шестакевич]. – Львів : Новий Світ-2000, 2012. – 359 с.
13. McGovern G. Content Criticsl / G. McGovern, R. Norton. – NJ : FT Press, 2001. – 256 p.

14. McKeever S. Understanding Web content management systems: evolution, life cycle and market. / S. McKeever // *Industrial Management & Data Systems*, 2003. – 103 (9) – P. 686–692.
  15. Rockley A. Managing Enterprise Content: A Unified Content Strategy. / A. Rockley. – Reading, Mass : New Riders Press, 2002. – 592 p.
  16. Висоцька В. А. Моделювання етапів життєвого циклу комерційного web-контенту / В. А. Висоцька, Л. Б. Чирун, Л. В. Чирун // *Вісник НУЛП*. – 2011. – № 715. – С. 69–87.
  17. Берко А. Ю. Алгоритми опрацювання інформаційних ресурсів в системах електронної комерції / А. Ю. Берко, В. А. Висоцька, Л. В. Чирун // *Комп'ютерні науки та інформаційні технології*. *Вісник НУЛП*. – 2008. – № 616. – С. 128–136.
  18. Бісікало О. В. Застосування методу синтаксичного аналізу речень для визначення ключових слів україномовного тексту / О. В. Бісікало, В. А. Висоцька // *Радіоелектроніка, інформатика, управління*. – 2016. – № 3(38). – С. 54–65.
  19. Бісікало О. В. Виявлення ключових слів на основі методу контент-моніторингу україномовних текстів / О. В. Бісікало, В. А. Висоцька // *Радіоелектроніка, інформатика, управління*. – 2016. – № 1(36). – С. 74–83.
  20. Литвин В. Застосування системи алгоритмічних алгебр для граматичного аналізу символічних обчислень виразів логіки висловлювань / В. Литвин, І. Бобик, В. Висоцька // *Радіоелектроніка, інформатика, управління*. – 2016. – № 39. – С. 54–67.
  21. Висоцька В. А. Особливості проектування та впровадження систем електронної комерції / В. А. Висоцька // *Вісник НУЛП*. – 2008. – № 626. – С. 116–125.
  22. Berko A. Features of information resources processing in electronic content commerce / Andriy Berko, Victoria Vysotska, Lyubomyr Chyrun // *Applied Computer Science*. ACS journal. – Poland, 2014. – Volume 10, Number 2. –P. 5–19.
  23. Vysotska Victoria. Web Content Processing Method for Electronic Business Systems / Victoria Vysotska, Lyubomyr Chyrun // *International Journal of Computers & Technology*. – December 2013. – Vol 12, No. 2. – P. 3211–3220.
- Стаття надійшла до редакції 05.04.2017.  
Після доробки 15.05.2017.

Коробчинский М. В.<sup>1</sup>, Чирун Л. Б.<sup>2</sup>, Висоцкая В. А.<sup>3</sup>, Кондратьев Е. А.<sup>4</sup>

<sup>1</sup>Д-р. техн. наук, старший научный сотрудник Военно-дипломатической академии имени Евгения Березняка, Киев, Украина

<sup>2</sup>Ведущий специалист института компьютерных наук и информационных технологий Национального университета «Львовская политехника», Украина

<sup>3</sup>Канд. техн. наук, доцент, доцент кафедры «Информационные системы и сети» Национального университета «Львовская политехника», Украина

<sup>4</sup>Магистр кафедры «Информационные системы и сети» Национального университета «Львовская политехника», Украина

#### **ОСОБЕННОСТИ ФОРМИРОВАНИЯ И АНАЛИЗА КОНТЕНТА ИНТЕРНЕТ-ГАЗЕТЫ МУЗЫКАЛЬНЫХ НОВОСТЕЙ**

**Актуальность.** Негативные факторы в формировании контента усложняют процесс поиска необходимых данных при сканировании различных источников информации. Увеличение физического объема и изменение актуальности / динамики контентных потоков (систематическое и нерегулярное обновление) приводит к возникновению дублирования, информационного шума и избыточности результатов поиска контента. Охват и обобщение больших динамических потоков контента, непрерывно генерируют в Интернет-источниках, требует качественно новых методов / подходов поиска как контент-мониторинг.

**Цель.** Целью выполнения работы является разработка метода формирования и анализ контента Интернет-газеты музыкальных новостей для активного привлечения постоянных пользователей информационного ресурса новостей и увеличение целевой аудитории. Задачей является разработка системы для распространения музыкальных новостей с использованием метода контент-мониторинга для формирования и анализ текстового контента информационного ресурса Интернет-газеты.

**Метод.** Входной информацией для контент-мониторинга является текст на естественном языке как последовательность символов, исходная информация – это таблицы разделов, предложений и лексем анализируемого текста. Контент-мониторинг является программным средством автоматизации нахождения важных составляющих в потоках контента. Это содержательный анализ потоков контента с целью постоянного получения необходимых качественных / количественных срезов на протяжении не определенного заранее периода времени. Составной контент-мониторинга является контентный поиск и контент-анализ текста. Контент-анализ предназначен для поиска контента в массиве данных по смысловыми лингвистическими единицами. Единица счета является количественной мерой единицы анализа, позволяющая регистрировать частоту (регулярность) появления признака категории анализа в тексте (количество определенных слов или их сочетаний, строк, печатных знаков, страниц, абзацев, авторских листов, площадь текста и т. д.).

**Результаты.** Решено актуальную задачу исследования и разработки методов и средств обработки информационных ресурсов интеллектуальной информационной системы формирования контента по использованию разработанной классификации, математического и программного обеспечения обобщенной архитектуры системы. Обоснована необходимость разработки методов и средств обработки информационных ресурсов в информационных системах формирования контента путем усовершенствования архитектуры системы с целью автоматизации процессов формирования, управления и реализации контента.

**Выводы.** Использование контент-анализа в мониторинге источников данных в Интернет автоматизирует процесс поиска наиболее важных компонент в потоке содержания, в выборе данных из этих источников. Это устраняет дубликаты контента, информационный шум, паразитные результаты бронирования контента и другое. Этот метод используется на следующих этапах анализа содержания для более точных результатов создания уникального контента, который пользуется спросом.

**Ключевые слова:** анализ, контент-анализ содержания информационных ресурсов, контент-анализ, рейтинговая оценка, система управления контентом.

Korobchynskiy M. V.<sup>1</sup>, Chyrun L. B.<sup>2</sup>, Vysotska V. A.<sup>3</sup>, Kondratiev E. O.<sup>4</sup>

<sup>1</sup>F.d., Senior Research Fellow of Military-Diplomatic Academy named Eugene Bereznyak, Kyiv, Ukraine

<sup>2</sup>Leading Specialist of Information Systems and Networks Department of Lviv Polytechnic National University, Lviv, Ukraine

<sup>3</sup>PhD, Associate Professor, Associate Professor of Information Systems and Networks Department of Lviv Polytechnic National University, Lviv, Ukraine

<sup>4</sup>Master of Information Systems and Networks Department of Lviv Polytechnic National University, Lviv, Ukraine

#### **OF CONTENT FORMATION AND ANALYSIS FEATURES IN ONLINE NEWSPAPER OF MUSIC NEWS**

**Context.** Negative factors in shaping the content complicate necessary data finding process when scanning various sources. Increased volume and variable Relevance / content flow dynamics (regular and non regular updates) leads to duplication and excessive information noise content in search results. Coverage and synthesis of large dynamic flow of content are continuously generating Internet media requires

qualitatively new methods / approaches of search as a content monitoring.

**Objective.** The work purpose is to develop content forming and analyzing method in online newspaper for the active involvement of regular users of news and information resource increase audytoryii target. The objective is to develop a system to distribute music news using content monitoring method for creation and content analyzing in information resource of online newspaper.

**Method.** Input information for content-monitoring is a natural language into a sequence of symbols, background information – a partition table, sentences and tokens analyzed text. Content monitoring is a software tool for automation of the important components in the flow of content. This is meaningful analysis of streams of content for the purpose of obtaining the necessary permanent qualitative / quantitative sections within pre-defined period of time not. Content monitoring component is content search and content analysis of the text. Content analysis designed to search for content in an array of data for linguistic semantic units. The unit of account is a quantitative measure of the unit of analysis that allows you to record the frequency (regularity) symbol appears in the text analysis category (number of certain words or combinations of lines, characters, pages, paragraphs, copyright sheets of text area, etc.).

**Results.** The problem of current research and development of methods and means of processing information resources information systems intellectual formation content with the use of the classification, mathematical and software and the generalized system architecture are solved. The necessity of methods and means development of information resources processing in intelligent information systems forming content by improving the system architecture to automate processes of formation, management and implementation of content.

**Conclusions.** The content analysis use in monitoring Internet data sources automates the finding process of the most important components in content flow in data selection from these sources. This eliminates duplicate content information noise, parasitic content redundancy results and more. This method is used in subsequent stages of content for more accurate relevant results – create a unique content that is in demand.

**Keywords:** analysis, content analysis content, information resource, content-analysis, rating evaluation, content management system.

## REFERENCES

- Dzherk N. Razrabotka prilozheniy dlya elektronnoy kommertsii. Sankt-Peterburg, Piter, 2001 r.
- Peleshchyshyn A. M. Metody pobudovy Efektyvnykh Web-system, *Informatsiyeni systemy ta merezhi. LPNU Visnyk*, 2012, No. 464, pp. 240–254.
- Berko A. Y., Vysotska V. A., Pasichnyk V. V. Systemy elektronnoy kontent-komertsii: monohrafiya. Lviv, LPNU Publishing, 2009, 612 p.
- Lytvyn V., Vysotska V., Chyrun L., Dosyn D. Methods based on ontologies for information resources processing, *LAP Lambert Academic Publishing*, Germany, 324 p.
- Vysotska V. A. Metody i zasoby opratsyuvannya informatsiynykh resursiv v systemakh elektronnoy kontent-komertsii : avtoreferat dysertatsiyi na zdobuttya naukovooho stupenya k.t.n. : 05.13.06. NU «Lvivska politehnika». Lviv, 2014, 27 p.
- Lande D., Furashev V., Braychevskiy S., Grigoriyev O. Osnovy modelirovaniya i otsenki yelektronnykh informatsionnykh potokov. Kiev, Inzhiniring, 2006, 348 p.
- Lande D. V. Nekotorye metody analiza novostnykh informatsionnykh potokov, *IKVT-2005*, Vol. 93. Donetsk, DonNTU, 2005, pp. 277–287.
- Lande D. V. Skaner sistemy kontent-monitoringa, *Otkrytyye informatsionnyye i komp'yuterny integrirovannyye tekhnologii*. Khar'kov, NAKU «KHA», 2005, Vyp. 28, pp. 53–58.
- Bolshakova Y. I., Klyshinskiy E. S., Lande D. V., Noskov A. A., Peskova O. V., Yagunova Ye. V. Avtomaticheskaya obrabotka tekstov na yestestvennom yazyke i komp'yuternaya lingvistika: ucheb. Posobiye. Moscow, MIEM, 2011, 272 p.
- Grigoryev A. N., Lande D. V. Adaptivnyy interfeys utochneniya zaprosov k sisteme kontent-monitoringa, *KHH mezhdunarodnaya nauchno-prakticheskaya konferentsiya*. Kiev, UkrINTEI, 2005, pp. 17–20.
- Ivanov V. F. Kontent-analiz. Kiev, 1994, 112 p.
- Pasichnyk V., Shcherbyna Y., Vysotska V., Shestakevych T. Matematychna lnhvistyka. Lviv, Novyy svit-2000, 2012, 359 p.
- McGovern G., Norton R. Content Criticsl. NJ, FT Press, 2001, 256 p.
- McKeever S. Understanding Web content management systems: evolution, life cycle and market, *Industrial Managemant & Data Systems*, 2003, 103 (9), pp. 686–692.
- Rockley A. Managing Enterprise Content: A Unified Content Strategy. Reading, Mass, New Riders Press, 2002, 592 p.
- Vysotska V. Chyrun L., Chyrun L. Modelyuvannya etapiv zhyttyevoho tsyklu komertsiynoho web-kontentu, *LPNU Visnyk*, 2011, No. 715, pp. 69–87.
- Berko A. Y., Vysotska V. A., Chyrun L. V. Alhorytmy opratsyuvannya informatsiynykh resursiv v systemakh elektronnoy komertsii, *Kompyuterni nauky ta informatsiyeni tekhnolohiyi. LPNU Visnyk*, 2008, No. 616, pp. 128–136.
- Bisikalo O. V., Vysotska V. A. Sentence syntactic analysis application to keywords identification Ukrainian texts, *Radio Electronics, Computer Science, Control*, 2016, No. 3(38), pp. 54–65.
- Bisikalo O. V., Vysotska V. A. Identifying keywords on the basis of content monitoring method in Ukrainian texts, *Radio Electronics, Computer Science, Control*, 2016, No. 1(36), pp. 74–83.
- Lytvyn V. V., Bobyk I. O., Vysotska V. A. Application of algorithmic algebra system for grammatical analysis of symbolic computation expressions of propositional logic, *Radio Electronics, Computer Science, Control*, 2016, No. 4(39), pp. 54–67.
- Vysotska V. A. Osoblyvosti proektuvannya ta vprovadzhennya system elektronnoy komertsii, *LPNU Visnyk*, 2008, No. 626, pp. 116–125.
- Berko A., Vysotska Victoria, Chyrun Lyubomyr Features of information resources processing in electronic content commerce, *Applied Computer Science. ACS journal*, Poland, 2014, Volume 10, Number 2, pp. 5–19.
- Vysotska Victoria, Chyrun Lyubomyr Web Content Processing Method for Electronic Business Systems, *International Journal of Computers & Technology*, December 2013, Poland, 2014, Volume 10, Number 2, pp. 5–19.