

МОДЕЛЮВАННЯ ПРОЦЕСІВ І СИСТЕМ

УДК 004.738.5

В. Висоцька, М. Сороковський
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

МОДЕЛЮВАННЯ ПРОЦЕСІВ РОЗПОДІЛЕННЯ КОНТЕНТУ В ІНТЕРНЕТ-ЖУРНАЛАХ

© Висоцька В., Сороковський М., 2012

Запропоновано модель життєвого циклу комерційного Web-контенту в системах електронної комерції. Модель описує процеси опрацювання інформаційних ресурсів в системах електронної контент-комерції та спрощує технологію автоматизації формування, управління та реалізації контенту. Проаналізовано основні проблеми електронної комерції та функціональних сервісів управління контентом. Запропоновано методи вирішення цих проблем.

Ключові слова: контент, контекст, Web-контент, життєвий цикл Web-контенту, Інтернет-маркетинг.

The article proposed to model the commercial Web content lifecycle in electronic commerce systems. The model describes the processes of information processing resources in the electronic content commerce systems and automation technology simplifies the formation, management and implementation of content. The paper analyzes the main problems of e-commerce and content management services function. The methods of solving these problems.

Key words: content, context, Web content, Web content lifecycle, Internet Marketing.

Вступ. Постановка проблеми

Інтернет-журнал є множиною структурованого контенту (наприклад, електронні видання у вигляді статей, анонсів, дайджестів, книг, репортажів, блогів, коментарів тощо) в інформаційному ресурсі (наприклад, на Web-сайті), призначена для задоволення інформаційних потреб цільової аудиторії та сприйняття людиною за допомогою відповідних програмних і апаратних засобів через Інтернет. Переваги впровадження Інтернет-журналу: компактний; економічно вигідний; миттєво доставляє будь-який контент у будь-яку точку світу в будь-який час; не прив'язаний до проблеми тиражування; зменшена до мінімуму відстань між автором і читачем; нові шляхи донесення інформації до читача. Автоматичне опрацювання електронних видань зменшує час пошуку необхідного контенту лише за введення відповідної адреси або ключових слів у пошукову систему. Інтернет-журнал порівняно з друкованим виданням економічно вигідніший через відсутність витрат на видання, мінімальну вартість зберігання і донесення готового інформаційного продукту. Публікація електронних видань в Інтернет-журналах надає додаткових можливостей навіть маловідомим авторам контенту, який актуальний для вузького кола цільової аудиторії, але не публікуються через збитковість традиційними видавництвами. Автори контенту мають змогу публікуватися самостійно, формуючи тексти в редакторах і розміщуючи їх на Web-сайтах для поширення. Однак, постає питання про порушення авторського права. Виникають проблеми у встановленні юридичного порядку авторства опублікованого контенту.

Аналіз останніх досліджень та публікацій

Електронне видання як різновид публікації поширюється on-line або off-line. Off-line публікація є контентом, який бібліографічно ідентифікований, здатний до розпізнавання, збережений на фізичних носіях та доступ до якого обмежений як в часі, так і в просторі. On-line публікація (або ресурс) є контентом, який бібліографічно ідентифікований, здатний до розпізнавання, записаний на електронних носіях і доступний в даний момент, наприклад, on-line монографія (словник в Інтернет), on-line серійне видання (Інтернет-журнал), on-line ресурс (Web-сайт організації). Більшість публікацій пишуть, редагують і форматують з використанням звичайних текстових редакторів і персональних комп'ютерів (табл. 1).

Таблиця 1

Огляд сучасних Інтернет-журналів про автомобільний світ

Назва	Особливості системи електронної контент-комерції
5 КОЛЕСО	Розрахований на широке коло автолюбителів, автомобілістів-професіоналів, представників автомобільного бізнесу. Особливістю є безпосередня участь читачів у тестуваннях автомобілів та повний каталог автомобілів на російському ринку. Адреса Інтернет-журналу: http://5koleso.ru/ . У кожному номері журналу: <ul style="list-style-type: none"> • закордонні і російські автоновини, новини офіційних дилерів, інтерв'ю; • новинки світового і вітчизняного автомобілебудування, перспективні розробки; • статті власних кореспондентів і читачів про результати тестування автомобілів; • практична інформація для автомобілістів (тести і огляди олій, шин, аксесуарів, матеріали про будову сучасного автомобіля, поради з експлуатації, відповіді на запитання читачів); • звіти власних кореспондентів з чемпіонатів, авто- і мотозмагань; • статті про нестандартні автомобілі, мотоцикли, снігоходи, квадро- і гідроцикли; • матеріали про історію автомобілебудування; • щоквартально оновлювана інформація про новинки закордонного і вітчизняного, комерційного транспорту; • новинки car audio, огляди аксесуарів, інформація про продажі автомобілів в Росії (комплектація, ціни).
Off-road Drive	Присвячений повнопривідним автомобілям. Подано матеріали про позашляховики та техніку, яка дає змогу дістатися до будь-якої точки на планеті, про подорожі світом і про людей, які живуть пригодами. Співпрацює з інженерами, які спеціалізуються на створенні екстремальних автомобілів, учасниками і переможцями off-road змагань і експертами полігону, де проходить сертифікація автомобілів. Адреса: http://www.off-road-drive.ru .
КОЛЕСА	Орієнтований на широку цільову аудиторію. Основним напрямком діяльності журналу є висвітлення актуальних автомобілів, які користуються найбільшою популярністю. Кожен автомобіль проходить редакційний тест-драйв, тому огляди містять якісні ексклюзивні фотоматеріали, а текст – це суб'єктивне відчуття журналіста-водія. В кожному номері обов'язково присутні статті про абсолютно нові, тільки но представлені автовиробниками, автомобілі в рубриці «Прем'єра». В журналі кожен автовласник знайде корисну для себе інформацію з використання автохімії для догляду за своїм «чотириколісним другом». Адреса: www.kolesa.ru . Портал містить не тільки он-лайн версію журналу, але й одну з найкращих автомобільних збірок новин, фотоархів з можливістю оцінювання роботи фотографа, форум, на якому кожен бажаючий може задати запитання журналістам і отримати на нього відповідь. Ще одним інформаційним напрямком є юридична допомога автовласникам, які потрапили в проблемну ситуацію.
МОТОР	В кожному номері журналу розповідається про світ автомобілів: порівняльні тести на швидкість, потужність і керованість, історії народження класичних авто і новітніх концепт-карів, тенденції світового автомобілебудування, дизайн інтер'єру і ергономіка автомобіля, сучасні мотоцикли і екзотичні автомобілі. Адреса: www.motormedia.ru .

Відомим методом опрацювання текстової інформації є контент-аналіз – стандартна методика дослідження в галузі суспільних наук, предметом якої є аналіз змісту текстових масивів і продуктів комунікативної кореспонденції (наприклад, коментарі, форуми, електронне листування, статті тощо). Суть методики контент-аналізу полягає в пошуку інформації за змістовими одиницями аналізу (словосполучення речення, тема, ідея, автор, персонаж, соціальна ситуація, частина тексту, кластеризована за змістом категорії аналізу) [1]. У табл. 2 подано основні характеристики формалізованого методу контент-аналізу.

Основні характеристики формалізованого методу контент-аналізу

Назва	Перелік характеристик формалізованого методу контент-аналізу
1	2
Завдання	<ul style="list-style-type: none"> • дослідження наповненості середовища; • опрацювання контенту (рекламні матеріали, публікації, стенограми групових дискусій), заснованого на їх формалізованому якісно-кількісному дослідженні; • формальна фіксація, числове опрацювання, оцінення та аналіз змісту джерела в контексті проблеми; • визначення насиченості рекламоносія рекламними зверненнями при оціненні ефективності реклами або аналізі конкурентного середовища; • формування ставлення носія до проблеми (позитивне, негативне, нейтральне); • визначення соціально-демографічного та психологічного портрету цільової аудиторії; • аналіз глибинних положень респондентів; • вимірювання та аналіз політичної, економічної орієнтації ЗМІ; • знаходження логічних протиріч у тексті, наприклад, у анонсах компаній; • виявлення рекламної стратегії фірми.
Особливості	<ul style="list-style-type: none"> • кількісно-якісний аналіз текстової інформації та масивів тексту з метою подальшої змістовної інтерпретації отриманих кількісно-якісних закономірностей; • спрямований на отримання важливої інформації про досліджуване явище; • застосовують у комплексі з допоміжними методами/процедурами опрацювання даних; • якісно-кількісний аналіз змісту текстового контенту з метою виявлення або вимірювання різних фактів і тенденцій, відображених у контентному потоці; • використовують при дослідженні джерел, інваріантних за структурою/змістом та існуючих як несистематизований, безладно організований текстовий матеріал; • формування з різноманіття текстового матеріалу абстрактної моделі змісту тексту; • дослідження текстової та графічної інформації, що полягає в перекладі аналізованної інформації в кількісні показники та її статистичне опрацювання; • характеризується великою строгістю, систематичністю; • існує два напрями контент-аналізу: кількісний і якісний; • не скасовує необхідності звичайного змістовного аналізу контенту, а доповнює його, їх поєднання поглиблює розуміння сенсу будь-якого тексту; • дає змогу виявити в контенті те, що ховається під час традиційного аналізу, але що має важливий сенс згідно з метою дослідження.
Сутність	<ul style="list-style-type: none"> • фіксація певних одиниць змісту; • аналіз контенту в його контексті; • використовують як основний метод дослідження (наприклад, для аналізу спрямованості Інтернет-журналу), паралельний з іншими методами (наприклад, при дослідженні ефективності функціонування Інтернет-журналу), допоміжний або контрольний (наприклад, для класифікації відповідей на анкетне опитування).
Об'єкт	контент (наприклад, повідомлення преси, протоколи зборів, листи, накази, розпорядження, дані вільних інтерв'ю та анкетне опитування тощо), аналіз змісту якого дає змогу задати однозначне правило для надійного фіксування потрібних характеристик (принцип формалізації), а також елементи змісту, що цікавлять дослідника і зустрічаються з достатньою частотою (принцип статистичної значущості).
Напрями	<ul style="list-style-type: none"> • виявлення множини змістовних одиниць, що існували до тексту та отримали в ньому відображення (індикатор досліджуваного об'єкта як реальності, автора або адресата); • визначення множини змістовних одиниць, які існують тільки в тексті (такі характеристики форми, як мова, структура, жанр повідомлення, ритм і тон мови); • формування множини змістовних одиниць, що існуватимуть після сприйняття (оцінення ефектів впливу).
Стадії	<ul style="list-style-type: none"> • формулюють тему, завдання і гіпотези дослідження; • визначають категорії аналізу (ключові поняття), відповідні дослідницьким завданням; • обирають відповідну одиницю аналізу (лінгвістичну одиницю мови або елемент змісту), який слугує в тексті індикатором явища, що цікавить дослідника; • встановлюють одиницю рахунку (кількісну міру взаємозв'язку текстових і позатекстових явищ): <ul style="list-style-type: none"> а) час-простір (число рядків, площа в квадратних сантиметрах, хвилини, час мовлення тощо); б) поява ознак в тексті та їх частота (інтенсивність); • вибір необхідних джерел для контент-аналізу, кількості та дати повідомлень, досліджуваного змісту.

1	2
Категорії	<ul style="list-style-type: none"> • доречні (відповідають поставленому завданню, досліджуваному змісту і рішенням); • вичерпні (охоплюють усі частини змісту, що визначаються завданнями даного дослідження, та доволі повно відображають сенс основних понять дослідження); • взаємовиключними (одні і ті самі частини не повинні належати різним категоріям); • надійними (між дослідниками не повинно бути розбіжностей з приводу того, які частини змісту слід відносити до тієї чи іншої категорії).
Інструменти	<ul style="list-style-type: none"> • лінгвістична одиниця аналізу; • класифікатор контент-аналізу; • протокол підсумків або бланк контент-аналізу; • реєстраційна картка або кодувальна матриця; • інструкція досліднику, який займається реєстрацією і кодуванням одиниць рахунку; • каталог або список проаналізованого контенту.

Одиниці аналізу не завжди правильно витлумачуються, тому їх розглядають на тлі ширших лінгвістичних/змістовних структур, що вказують на характер членування тексту, в межах якого ідентифікується присутність/відсутність одиниць аналізу – контекстуальних одиниць. Наприклад, для одиниці аналізу *слово* контекстуальна одиниця – *речення*. Одиниця рахунку є кількісною мірою одиниці аналізу, що дає змогу реєструвати частоту (регулярність) появи ознаки категорії аналізу в тексті (кількість певних слів або їх поєднань, рядків, друкованих знаків, сторінок, абзаців, авторських аркушів, площа тексту тощо). Параметри вибірки визначаються завданнями і масштабами дослідження, наприклад, для аналізу протоколів зборів достатньо 12 протоколів за кількістю місяців, для аналізу повідомлень Інтернет-журналів – 12–16 номерів або 200–600 текстів. У табл. 3 подано основні процедури формалізованого методу контент-аналізу [2].

Таблиця 3

Основні процедури формалізованого методу контент-аналізу

№	Назва	Особливості процедури формалізованого методу контент-аналізу
1	Виявлення одиниць	Змістовні одиниці контент-аналізу виділяються залежно від змісту, цілей, завдань і гіпотез конкретного дослідження.
2	Виділення одиниць рахунку	Одиниці рахунку можуть збігатися (процедура зводиться до підрахунку частоти згадки виділеної смислової одиниці) або не збігатися (дослідник на основі аналізованого матеріалу сам висуває одиниці рахунку) з одиницями аналізу.
3	Процедура підрахунку	Використовують стандартні прийоми класифікації за виділеними угрупованнями із застосуванням спеціальних формул (наприклад, оцінення питомої ваги змістовних категорій в загальному обсязі тексту), статистичних розрахунків зрозумілості і атрактивності тексту.
4	Розроблення інструментів	Класифікатором є загальна таблиця, в яку зведено всі категорії аналізу і одиниці аналізу. Основне призначення – гранично чітко зафіксувати те, в яких одиницях виражається кожна категорія наприклад, категорії аналізу відіграють роль запитань в анкеті згідно з протоколом, а одиниці аналізу – відповідей. Є основним інструментом контент-аналізу, який передбачає зміст інших інструментів цього методу. Протокол контент-аналізу містить: відомості про документ (автор, час видання, обсяг тощо); підсумки його аналізу (кількість вживання в ньому певних одиниць аналізу і висновки щодо категорій аналізу). Протоколи заповнюються в закодованому вигляді для компактності подання інформації та швидкого порівняння результатів аналізу різного контенту. Протокол кожного контенту заповнюється на основі підрахунку даних всіх його реєстраційних карток (кодувальні матриці, де зазначається кількість одиниць рахунку та характеризуються одиниці аналізу).
5	Розроблення таблиці	Тип таблиці визначається етапом дослідження, наприклад, у вигляді системи скоординованих і субординованих категорій аналізу, яка зовні нагадує анкету: кожна категорія (питання) передбачає ряд ознак (відповідей), за якими квантифікується зміст тексту.
6	Розроблення кодувальної матриці	Якщо обсяг вибірки ≥ 100 одиниць, то аналізується набір матричних листів. Якщо вибірка < 100 одиниць, то проводиться двовимірний/багатомірний аналіз. У цьому випадку для кожного тексту повинна бути своя кодувальна матриця.

Формулювання цілі статті

Комерційний контент – це об’єкт купівлі/продажу між учасниками електронної комерції, наприклад, інформаційні блоки, які поділяють на синдикати (наприклад, блок погоди), анонси матеріалів інших розділів сайту (з посиланням) або з інших сайтів (з посиланням), довідкова інформація (наприклад, святкові дати, анонс заходу, розклад руху поїздів), розважальна інформація (наприклад, анекдот дня), реклама, кнопки і посилання інформаційних партнерів, кнопки статистики. Формальна модель життєвого циклу комерційного контенту – це шістька

$$S = \langle X, \text{Creation}, C, \text{Processing}, \text{Distribution}, Y \rangle,$$

де $X = \{x_1, x_2, \dots, x_{n_x}\}$ – множина вхідної інформації, *Creation* – функція створення контенту, $C = \{c_1, c_2, \dots, c_{n_c}\}$ – множина контенту, *Processing* – функція опрацювання контенту, *Distribution* – функція поширення контенту та $Y = \{y_1, y_2, \dots, y_{n_y}\}$ – множина вихідної інформації (рис. 1). Вхідна інформація x_i є незалежними змінними і у векторній формі має вигляд $\vec{x}(t) = (x_1(t), x_2(t), \dots, x_{n_x}(t))$, а множина контенту c_j та вихідна інформація y_k є залежними змінними й у векторній формі мають вигляд $\vec{c}(t) = (c_1(x_i, t), c_2(x_i, t), \dots, c_{n_c}(x_i, t))$ та $\vec{y}(t) = (y_1(t + \Delta t), y_2(t + \Delta t), \dots, y_{n_y}(t + \Delta t))$.

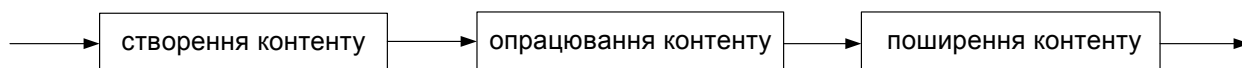


Рис. 1. Модель системи розподілення контенту в Інтернет-журналах

Аналіз отриманих наукових результатів

На рис. 2 подана типова схема взаємодії модулів СЕКК.

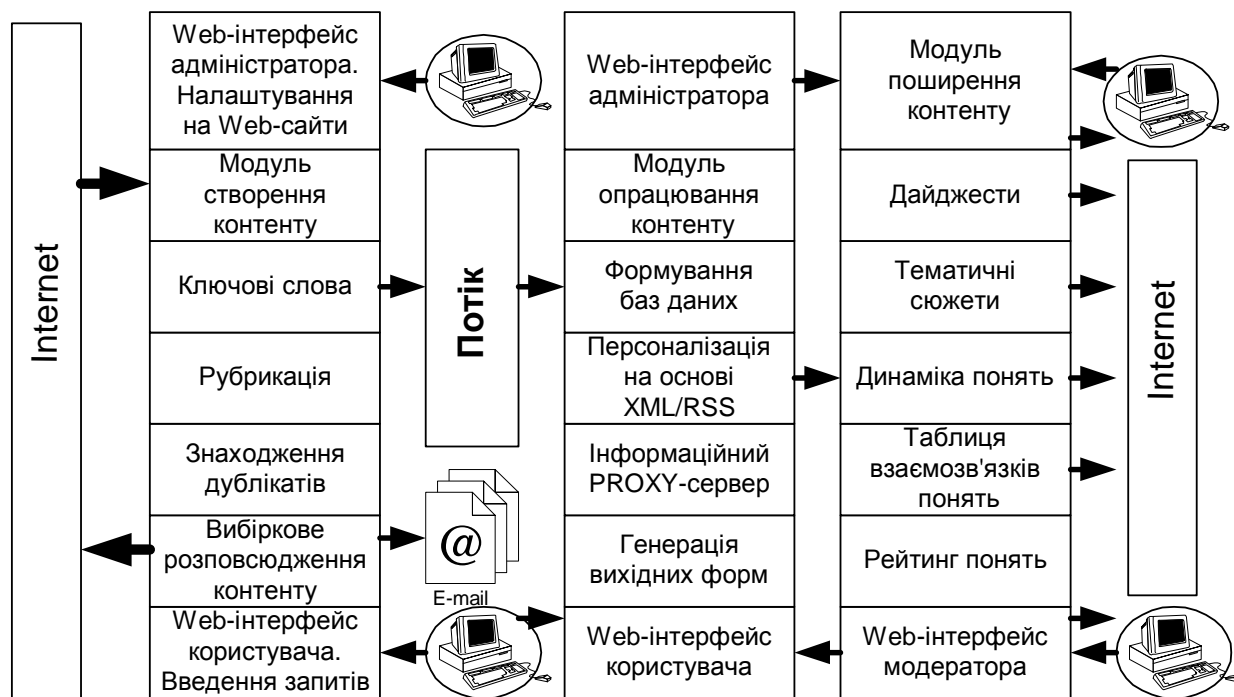
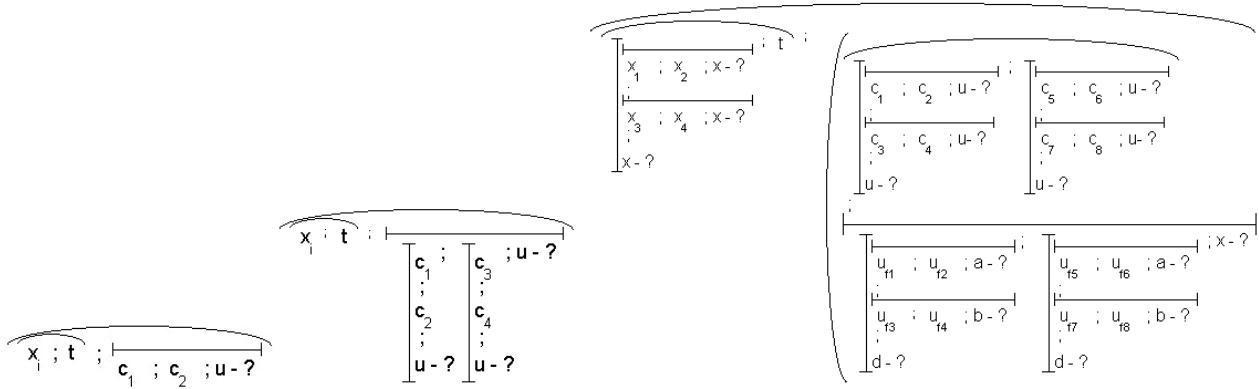
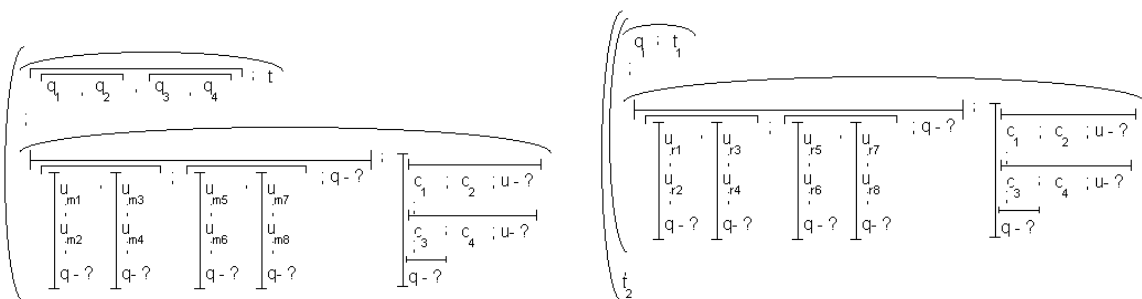


Рис. 2. Схема взаємодії модулів розподілення контенту в Інтернет-журналах

Процес створення контенту описується функцією вигляду $\bar{c}(x_i, t) = Creation(\bar{u}_C, x_i, t)$, де $\bar{u}_C(x_i)$ – множина умов створення контенту, тобто $\bar{u}_C(x_i) = (u_{c_1}(x_i), u_{c_2}(x_i), \dots, u_{c_{w_c}}(x_i))$. Контент створюють як $c_j = \left\{ \bigcup u_{c_k} \mid (x_i \in X) \wedge (\exists u_{c_k} \in U_F), U_C = U_{C_x} \vee U_{C_x}, i = \overline{1, m}, k = \overline{1, n} \right\}$, тобто



Процес опрацювання контенту описується як $\bar{c}(q_i, t) = Processing(\bar{u}_P, q_i, t)$, де $Q = \{q_1, q_2, \dots, q_{n_Q}\}$ – множина запитів користувачів, $\bar{u}_P(q_i)$ – множина умов опрацювання контенту, тобто $\bar{u}_P(q_i) = (u_{p_1}(q_i), u_{p_2}(q_i), \dots, u_{p_{m_p}}(q_i))$. Опрацювання контенту відбувається як $c_j = \left\{ \bigcup u_{p_k} \mid (q_i \in Q) \wedge (\exists u_{p_k} \in U_P), U_P = U_{Mq} \vee U_{Pq}, i = \overline{1, m}, k = \overline{1, n} \right\}$, тобто



Процес поширення контенту описується як $\bar{y}(t + \Delta t) = Distribution(\bar{u}_D, \bar{c}, q_i, t, \Delta t)$, де $\bar{u}_D(q_i, \bar{c})$ – множина умов поширення контенту $\bar{u}_D(q_i, \bar{c}) = (u_{d_1}(q_i, \bar{c}), u_{d_2}(q_i, \bar{c}), \dots, u_{d_{m_d}}(q_i, \bar{c}))$ при $y_j = \left\{ \bigcup u_{d_k} \mid (q_i \in Q) \wedge (\bar{c} \in C) \wedge (\exists u_{d_k} \in U_R), U_D = U_{Dc} \vee U_{D\bar{c}}, i = \overline{1, m}, k = \overline{1, n} \right\}$.

На рис. 3 діаграма зображує функціонування системи опрацювання контенту у вигляді статей, яка відображає бізнес-процеси Інтернет-журналу з погляду керівника проекту. Система реалізує основну мету – розподілення статей. На вхід подаються неопрацьовані статті, які подаються на блок опрацювання статей (рис. 4). За правилами, які подані у вигляді параметрів пошуку, дати та ключових слів переглянутих статей, виконується розподіл статей на категорії (популярні, останні, схожі або відібрані в результаті пошуку) та розміщення їх в БД. Процес контролює адміністратор.

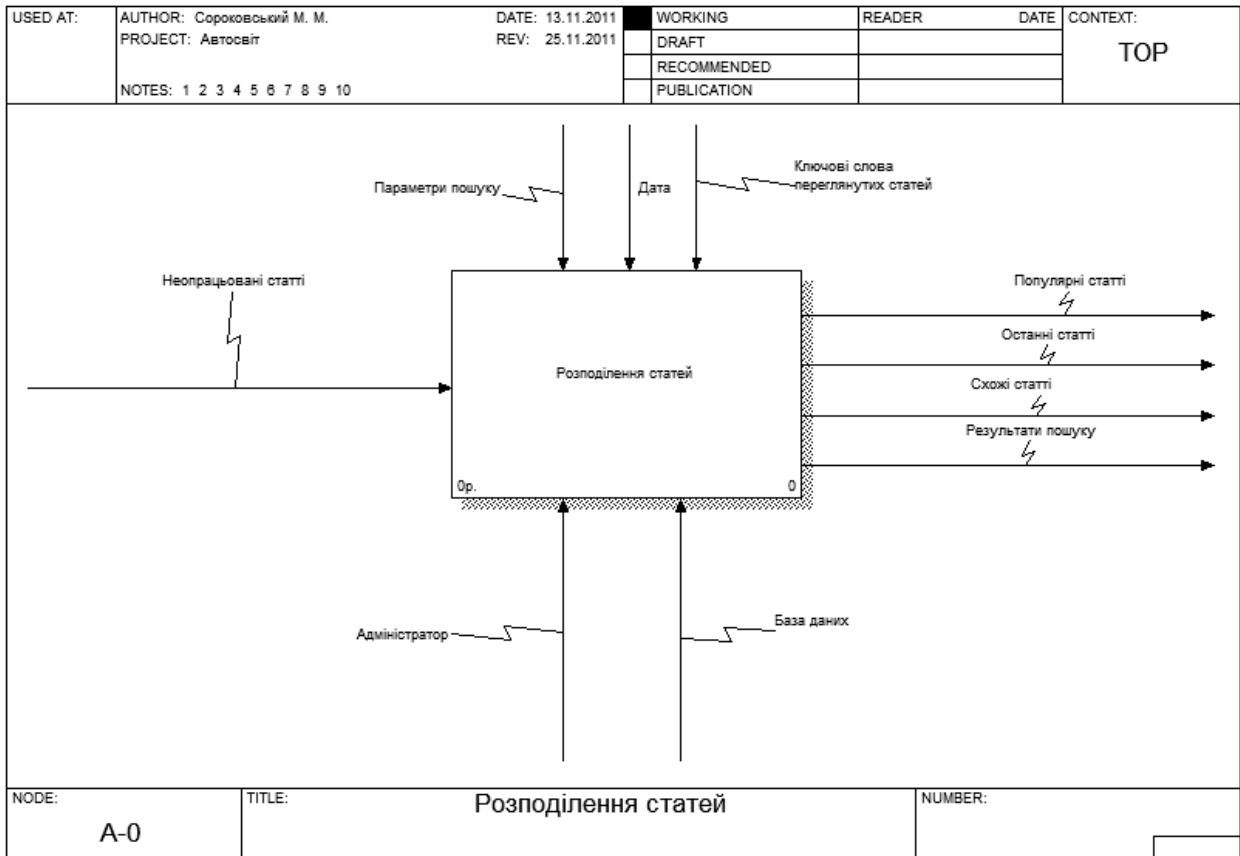


Рис. 3. Діаграма розподілення статей в Інтернет-журналі «Автосвіт»

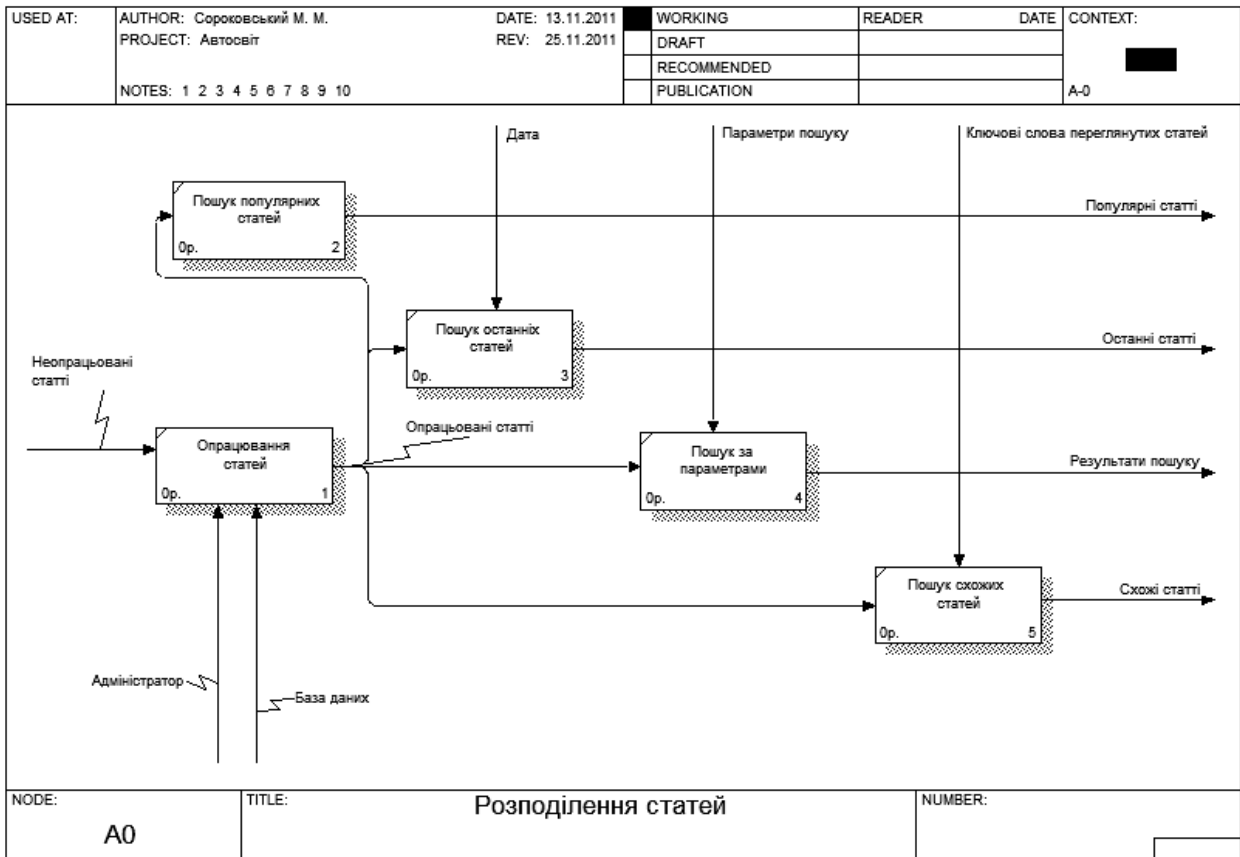


Рис. 4. Діаграма процесів розподілення статей

Опрацьовані статті подаються на блоки:

- пошук популярних статей – процедура відбору статей з найбільшими кількостями переглядів та виведення їх для перегляду користувачем;
- пошук останніх статей – процедура відбору статей за датою публікації у межах періоду від заданої дати до поточної дати і виведення відібраних статей для перегляду користувачем;
- пошук статей за параметрами – процедура відбору статей за введеними користувачем параметрами і виведення результату;
- пошук схожих статей – процедура відбору статей, використовуючи як параметри пошуку ключові слова з вже переглянутих користувачем статей, і виведення результату.

На рис. 5 опрацьовані статті потрапляють на блок пошуку збігів з ключовими словами, переглянутих користувачем раніше. Сформований список статей сортується за кількістю збігів для подальшого відбирання статей, які найбільш схожі на переглянуті автором. Відбираються перші статті, кількість яких задається в налаштуваннях системи. Далі роботи автора статті, яка переглядається в цей час, переміщуються на початок списку і виводяться для перегляду користувачем.

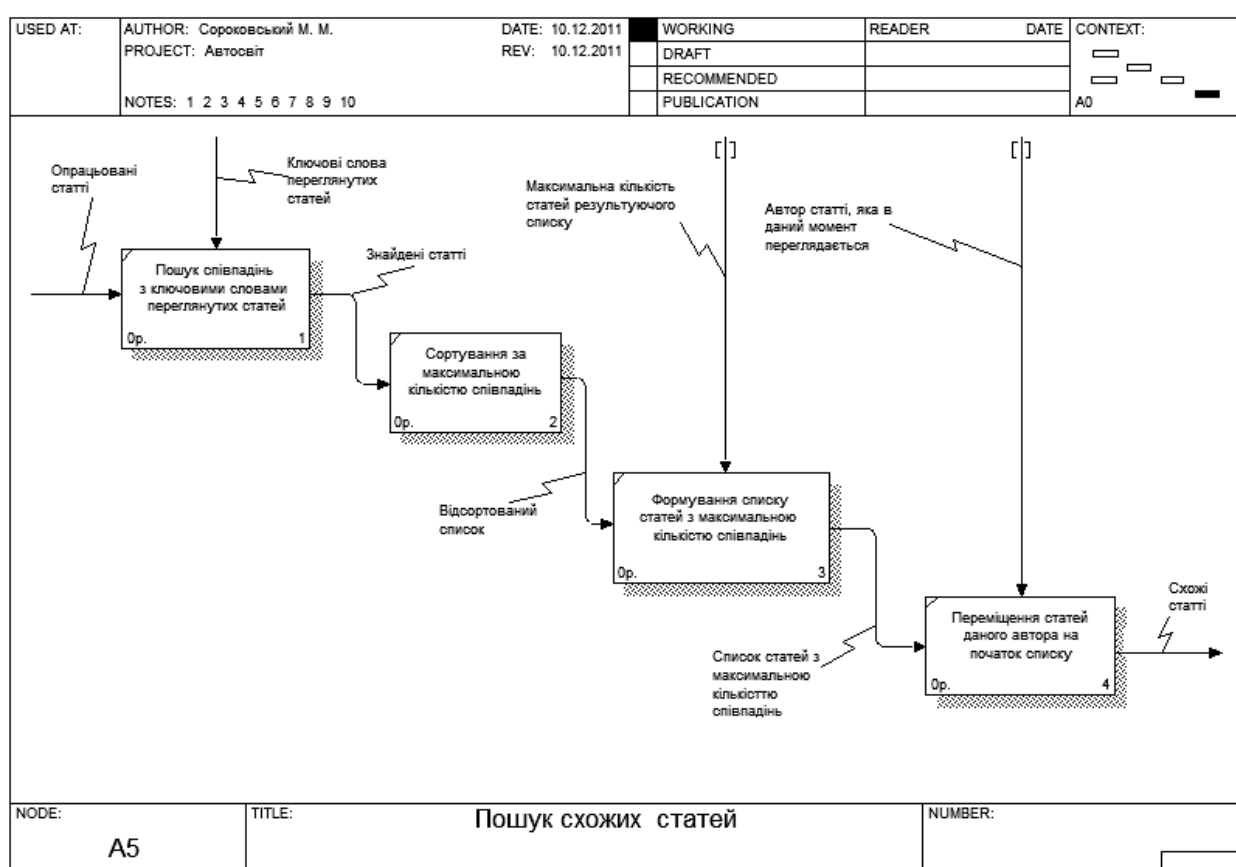


Рис. 5. Діаграма пошуку схожих статей

На рис. 6 подано сценарій покрокового виконання завдань Інтернет-журналу. Спочатку автор відправляє статтю на опрацювання, далі опрацьована стаття зберігається в БД.

Паралельно з цим користувач відправляє параметри пошуку в блок пошуку статей за параметрами. Далі блок пошуку статей за параметрами робить запит в базу даних і формує список статей, які відповідають параметрам пошуку і передає їх користувачеві, а також в блок формування списку статей переглянутих користувачем, який відправляє сформований список у блок формування списку схожих статей. Останній видає результат користувачеві.

Періодично блоки формування списку нових статей та формування списку популярних статей роблять запити в базу даних та виводять сформовані списки для перегляду.

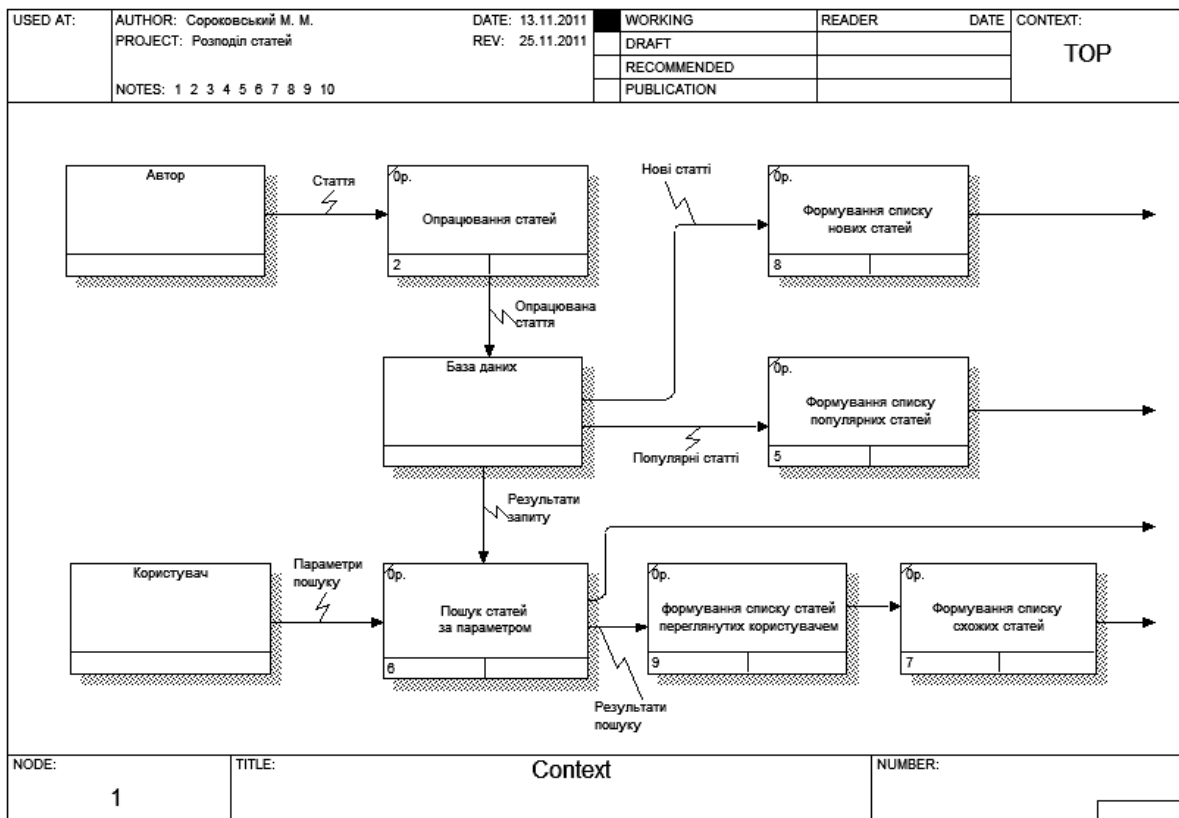


Рис. 6. Сценарій розподілення статей в Інтернет-журналі

На рис. 7 подано блок опрацювання статей, який складається з чотирьох операцій. Після виконання операцій неопрацьована стаття автора перетворюється в опрацьовану статтю і зберігається в базі даних.

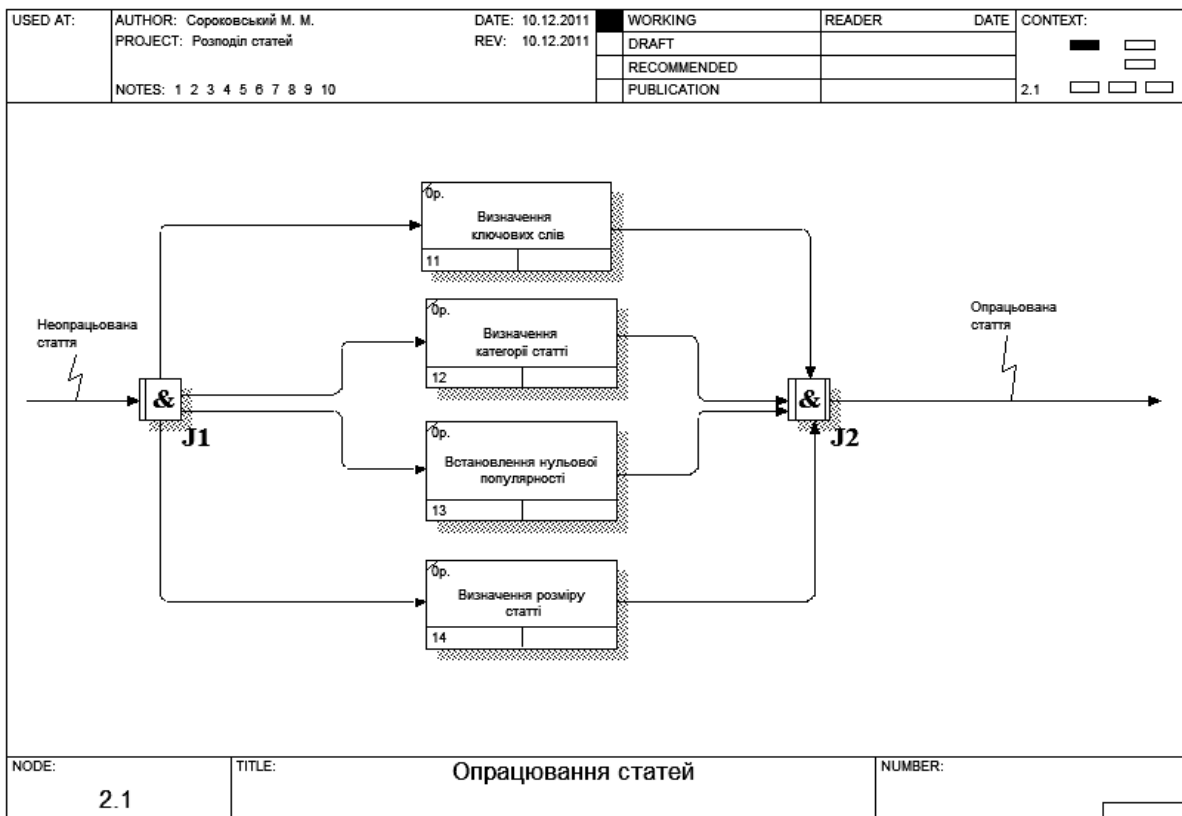


Рис. 7. Процес опрацювання статей Інтернет-журналу

Збережена в базі даних стаття використовується для пошуку за розміром, за ключовими словами, за автором, а також за популярністю.

На рис. 8 подано процеси розподілення статей в Інтернет-журналі. Після авторизації автор відправляє статтю на опрацювання. Періодично з базою даних відбувається обмін інформацією для формування списку нових статей, популярних статей, статей, які відповідають параметрам пошуку користувача, та списку статей, схожих на ті, які переглядав користувач. Всі ці чотири списки статей відправляються для перегляду користувачу. На основі отриманих списків користувач робить новий запит, і механізм повторюється.

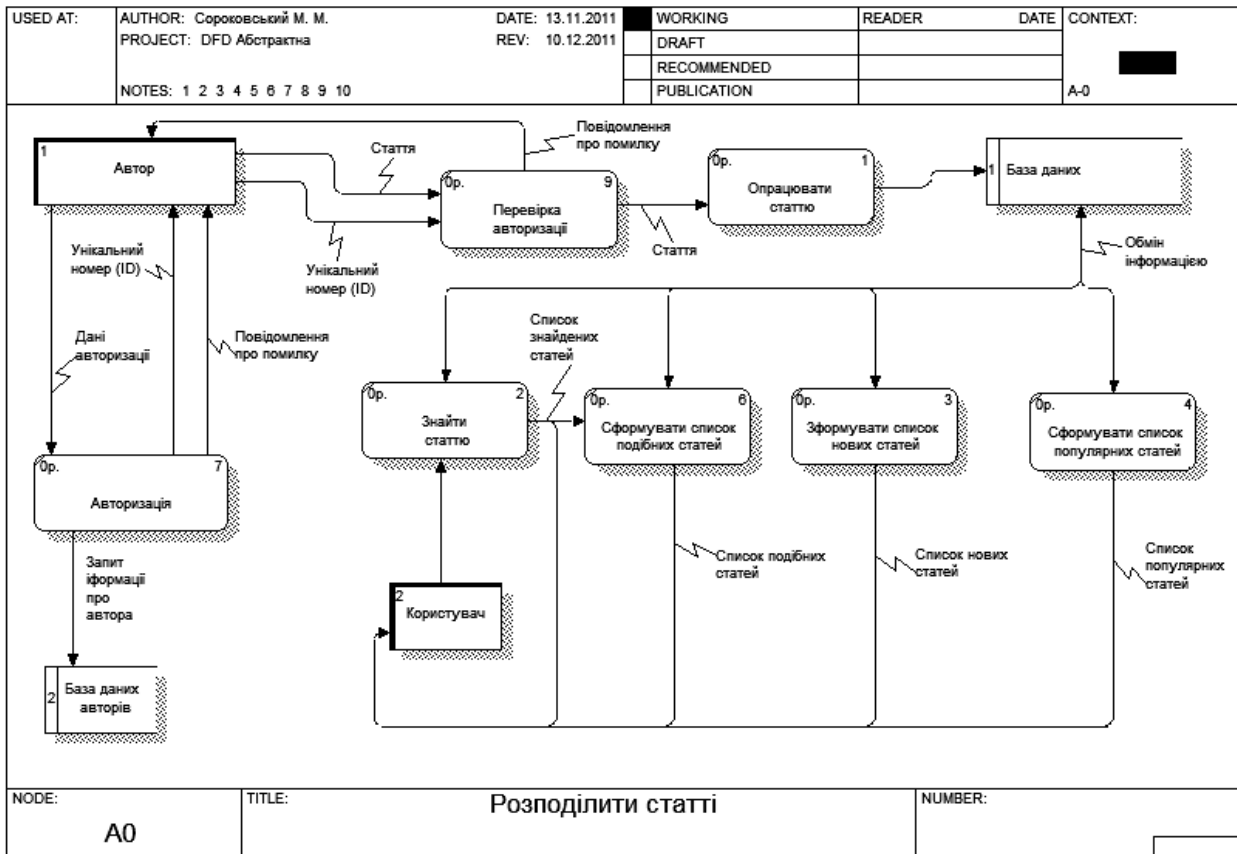


Рис. 8. DFD системи розподілення статей в Інтернет-журналі

Висновки і перспективи подальших наукових розвідок

Контент – це основа Інтернет-журналу. За ним користувач здійснює пошук для отримання необхідної інформації. Але текстів, які переповнені ключовими словами, не завжди достатньо для того, щоб користувач отримав потрібну інформацію. До того ж виділення ключових слів до кожної статті є довготривалим і трудомістким процесом. За допомогою формалізованого методу контент-аналізу процес є повністю автоматизованим і відбувається при додаванні автором нової статті. Також за допомогою контент-аналізу визначають схожі статті до тих, які переглядав користувач та виводять їх для перегляду. З використанням ключових слів користувач отримує не зовсім схожу або зовсім не схожу на ту, яка його цікавить. Це пов'язано з тим, що автор може додати ключові слова до матеріалу, які стосуються до різних тем з метою підвищення популярності його статті на сайті. Ще однією перевагою використання контент-аналізу контенту Інтернет-журналу є можливість визначення наявності контенту під конкретний запит, наприклад, визначення відсутності контенту з певною тематикою і спрямування роботи авторів на розвиток цього питання.

1. Зміст методики «контент-аналіз» Т. Хорошилова // Прикладна лінгвістика. – Режим доступу: http://studentstpl.ucoz.ru/publ/teorija_vozdejstvija/metodika_kontent_analiza/soderzhание_metodik_i_kontent_analiz/12-1-0-116. 2. Контент-аналіз як метод дослідження О.Т.Манасєв // Псі-фактор. – Режим доступу: <http://psyfactor.org/lib/content-analysis3.htm>.