

МЕТОДИ ТА ЗАСОБИ ОПРАЦЮВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ СИСТЕМ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

У статті проаналізовано основні проблеми електронної комерції та запропоновано методи вирішення цих проблем. Побудовано архітектури та моделі систем електронної контент-комерції. Сформульовано новий підхід застосування та впровадження бізнес-процесів для побудови систем електронної контент-комерції. Введено нова класифікація систем електронної комерції та систем електронної контент-комерції. Викладено теорію побудови, проектування та реалізації систем електронної контент-комерції. Визначена математична модель систем електронної контент-комерції та математична модель керування контентним потоком. Узагальнено поняття системи електронної контент-комерції. Розроблено методи проектування та реалізації системи електронної контент-комерції на прикладі Інтернет-газети.

Ключові слова: контент, контент-комерція, системи електронної контент-комерції.

В статье проанализированы основные проблемы электронной коммерции и предложены методы решения этих проблем. Построены архитектуры и модели систем электронной контент-комерции. Сформулирован новый подход применения и внедрения бизнес процессов для построения систем электронной контент комерции. Построена математическая модель систем электронной контент комерции и математическая модель управления контентным потоком. Обобщено понятие системы электронной контент-комерции. Разработаны методы проектирования и реализации системы контент комерции на базе Интернет-газеты.

Ключевые слова: контент, контент-комерция, системы электронной контент-комерции.

In the given article main problems of electronically commercial are analyzed. New methods for solution of discussed problems are proposed. It is built architecture and model of the systems of electronic content-commerce. New approach of application and introduction of business-process is formulated for the construction of the systems of electronic content-commerce. It is entered new classification of the systems of electronic commerce and systems of electronic content-commerce. The theory of construction, planning and realization of the systems of electronic content-commerce is expounded. The mathematical model of the systems of electronic content-commerce and mathematical case a content-stream frame is certain. Generalized concept of the system of electronic content-commerce. The methods of planning and realization of the system of electronic content-commerce are developed on the example of Internetnewspaper.

Key words: content, content commerce, electronical content-commerce systems.

ВСТУП. ЗАГАЛЬНА ПОСТАНОВКА ПРОБЛЕМИ

Особливості інформаційного Інтернет-ринку полягають у наступному: відкритість – доступ як для всіх компаній і для споживачів; глобальність – доступ із будь-якої точки земної кулі; відвертість ринку – досить низький бар'єр для входу на нього. Інтернет надає можливість скорочення каналів поширення контенту та ліквідації проміжних ланок (дистриб'ютор, оптовик), на зміну яким приходить прямий зв'язок виробник – покупець. Причиною такого скорочення є спроможність фірм взяти на себе функції, що традиційно виконуються фахівцями проміжних ланок, оскільки Інтернет володіє більш ефективною можливістю взаємодії зі споживачами й одночасно дозволяє відстежувати інформацію про споживачів. З технічного боку це зумовлено розвитком ІТ побудови і ведення БД та автоматичним опрацюванням запитів, які надходять. Особливості Інтернет як засобу торгівлі контентом: споживач взаємодіє з Інтернет і здійснює запити про купівлю безпосередньо через середовище Інтернет; споживач контролює інформацію про контент та одержує її в будь-який час доби; організація роздрібною контент-торгівлі в Інтернет вимагає менших вкладень порівняно з організацією торгівлі традиційними способами (видавництво, книгарні, бібліотеки, ринок консалтингових послуг тощо); порівняно з торгівлею по телебаченню або торгівлею по каталогах споживачі можуть запитувати додаткову інформацію, достатню для здійснення купівлі, в тому числі в інтерактивному режимі [5, 7, 8, 11, 12].

Актуальність впровадження системи електронної контент-комерції (СЕКК) зумовлена наступними чинниками: глобалізація бізнесу призводить до зростання потреб контенту та швидкого отримання доступу до контенту для успішного ведення бізнесу; нерівномірність функціонування бізнес-процесів відповідно до регіонів (країн, областей тощо) призводить до зростання потреб щодо регулярності та періодичності отримання необхідного контенту; економія часу в отриманні необхідного контенту; персоналізація у наданні послуг у СЕКК; інтегрованість СЕКК. Переваги впровадження СЕКК: збільшення оперативності одержання контенту, особливо при міжнародних операціях; скорочення циклу виробництва і продажу; зниження витрат, пов'язаних з обміном інформацією за рахунок використання сучасних ІТ та ПЗ; використання Інтернет-технологій електронної контент-комерції дозволяє компанії стати більш відкритою стосовно клієнтів; легке і швидке інформування партнерів та клієнтів про контент; створення альтернативних каналів продажів, наприклад, через електронні газети на корпоративному сайті. Розвинуті у роботі наукові положення базуються на доробку багатьох учених, зокрема А.М. Берези [1], Д.В. Ланде [2], А.Ю. Берка [3, 4], А.Ю. Пелещишина, С.М. Брайчевського, А.Н. Григор'єва, В.Н. Фурашева.

ЗВ'ЯЗОК ВИСВІТЛЕНОЇ ПРОБЛЕМИ ІЗ ВАЖЛИВИМИ НАУКОВИМИ ТА ПРАКТИЧНИМИ ЗАВДАННЯМИ

Основу роботи складають результати теоретичних та практичних досліджень, виконаних авторами в рамках планових робіт кафедри «Інформаційних систем та мереж» Національного університету «Львівська політехніка». Дослідження, висвітлені у роботі, здійснювались згідно з планом науково-дослідних робіт Національного університету «Львівська політехніка» у рамках держбюджетних тем ДБ/ВЕБ «Розробка методів, алгоритмів і програмних засобів моделювання, проектування та оптимізації інтелектуальних інформаційних систем на основі Web-технологій «ВЕБ» (номер держ. реєстр. 0102U001171) та ДБ/Ізоморф «Розроблення методів і засобів побудови інтелектуальних інформаційних систем на основі часових реляційних баз даних» (номер держ. реєстр. 0104U002299). Автори здійснили проектування узагальненої архітектури СЕКК, а також виконали аналіз методів керування потоком контенту в СЕКК.

АНАЛІЗ СУЧАСНИХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

З точки зору використання СЕКК контент – структурована множина, логічно завершена інформація, яка є об'єктом взаємовідносин між суб'єктом-користувачем СЕКК та самою СЕКК.

З точки зору моделювання СЕКК контент – набір інформації, об’єкт СЕКК, який неподільний у часі, є основним чинником функціонування СЕКК та існує лише в електронному вигляді. Множина контенту формує область діяльності, функціонування та призначення СЕКК. Фахівці із проектування й експлуатації складних СЕКК мають справу із підсистемами керування різних рівнів, що мають загальну властивість – прагнення досягти деякої мети (продаж контенту кінцевому користувачу). Система електронної контент-комерції S – цілеспрямована множина взаємозалежних елементів будь-якої природи, серед яких обов’язкова наявність множини контенту визначеного характеру (стаття, ПЗ, книга тощо). Зовнішнє середовище E – множина наявних поза СЕКК елементів будь-якої природи та контенту, що впливають на систему або перебувають під її впливом.

Як видно із схеми на рис. 1, кількість контентних потоків помітно більша, ніж шляхів переміщення товарів на промислових підприємствах [1].

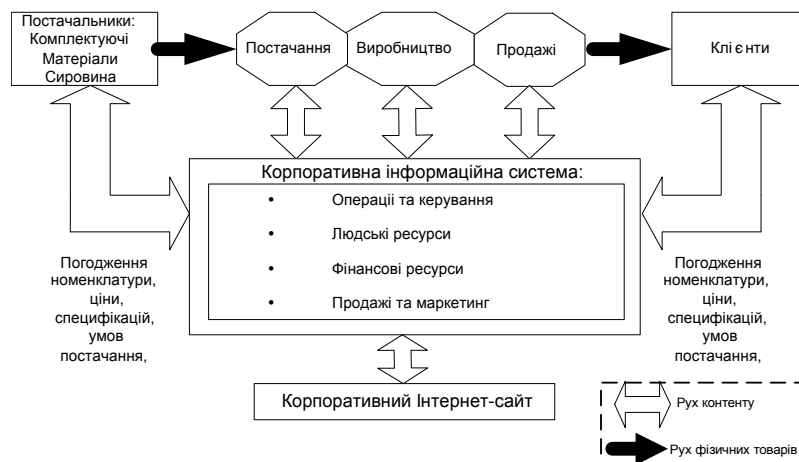


Рис. 1. Основні товарні й контентні потоки в СЕКК

Значна частина цих контентних потоків складається з легко сформалізованих [3] і автоматизованих [4] процедур (рис. 2).

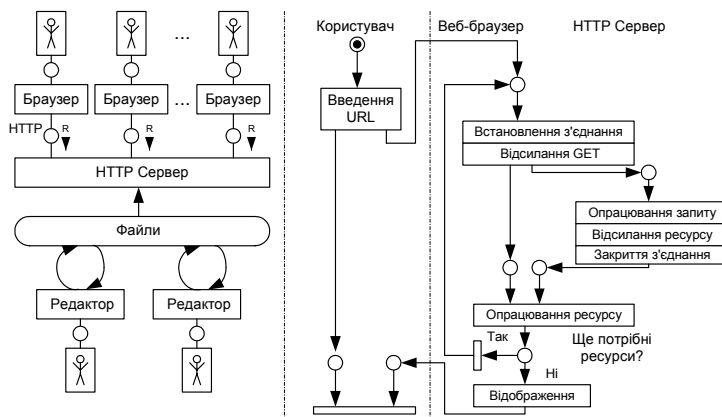


Рис. 2. Проста структура функціонування СЕКК

ВИДІЛЕННЯ ПРОБЛЕМ

Основна проблема проектування СЕКК – відсутність загального підходу до процесу моделювання, проектування та розроблення СЕКК. Відсутність загальної та детальної класифікації СЕК та СЕКК, що приводить до проблеми визначення та формування загальних методів проектування та розроблення архітектури та алгоритмів функціонування СЕК і СЕКК [3-4]. Це обґрунтовує мету, актуальність, доцільність та напрями дослідження. Ще складнішою є інтеграція внутрішніх ІС [1] з онлайн-системами інших учасників ринку (рис. 3).

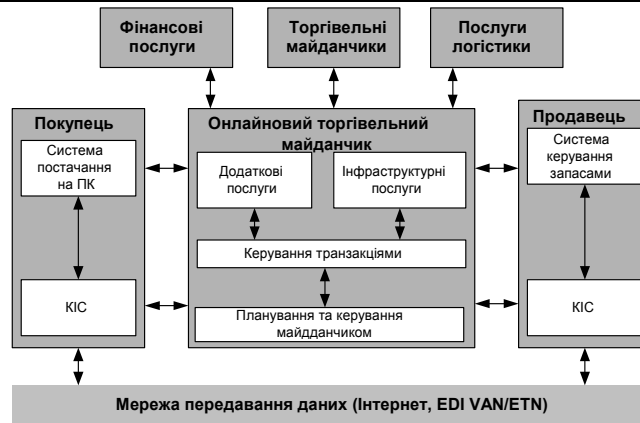


Рис. 3. Точки інтеграції онлайн-торгівельних СЕКК

СЕК класифікують за видом основного потоку бізнес-процесів: матеріальний, інформаційний, фінансовий [1]. Процеси, які становлять цикл функціонування СЕК: доступ до інформації, оформлення замовлення, оплата, виконання замовлення, післяпродажне обслуговування і підтримка [1, 3, 4]. Типи СЕК представлені на рис. 4, типи СЕКК – на рис. 5.

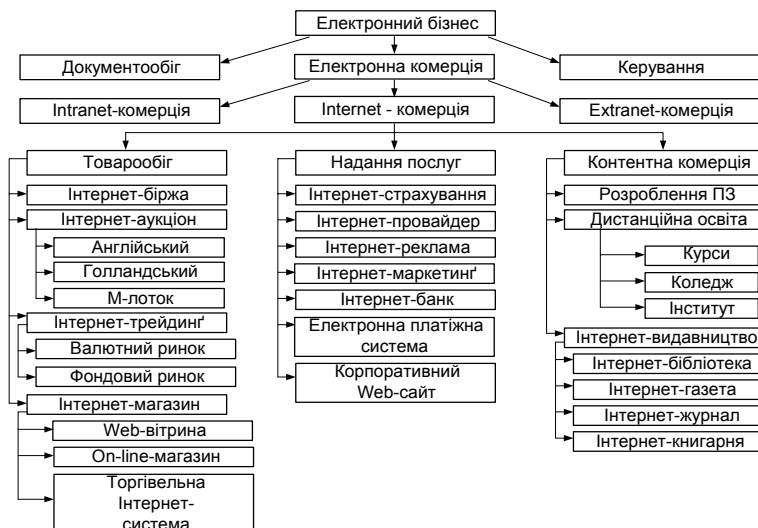


Рис. 4. Типологія систем електронної комерції

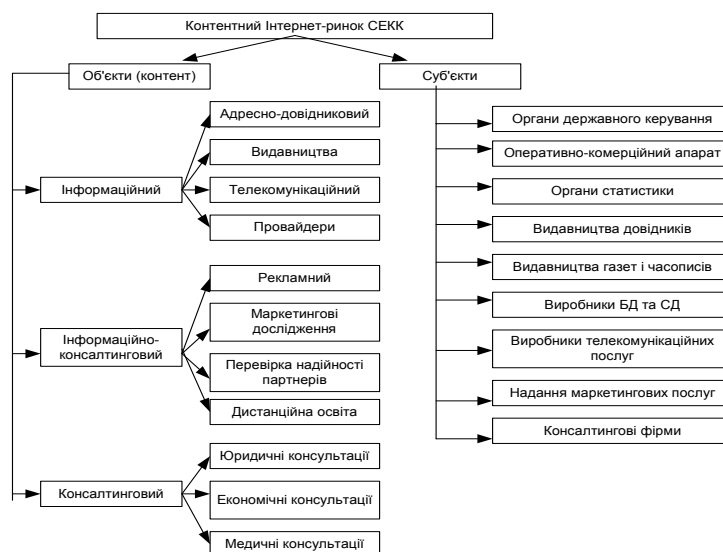


Рис. 5. Інтернет-ринок та типологія СЕКК

ФОРМУВАННЯ МЕТИ

Мета дослідження – розроблення моделей та архітектурних принципів побудови СЕКК і моделей керування контентом як складової частини СЕКК.

Мета роботи визначила необхідність вирішення таких задач:

- провести системний аналіз об'єктів та процесів електронної контент-комерції;
- визначити пріоритетні задачі, що виникають в процесі розроблення та функціонування СЕКК;
- здійснити аналіз функціонування і класифікувати СЕК та СЕКК;
- розробити моделі керування контентом як складової частини СЕКК;
- розробити архітектурні принципи побудови СЕКК;
- розробити моделі побудови СЕКК.

Об'єкт дослідження – процеси та системи електронної контент-комерції.

Предмет дослідження – методи та засоби опрацювання інформаційних ресурсів СЕКК.

Методи дослідження. Дослідження, що виконані під час роботи над статтею, ґрунтуються на теорії проектування за допомогою CASE-засобів, елементів загальної теорії систем, апараті теорії реляційних баз даних та теорії імітаційного моделювання. При побудові математичних моделей процесів функціонування СЕКК застосовані такі основні підходи: неперервно-детермінований (диференціальні рівняння); дискретно-детермінований (кінцеві автомати); дискретно-стохастичний (імовірнісні автомати); неперервно-стохастичний (системи масового обслуговування); узагальнений, або універсальний (агрегативні системи). Математичні схеми допомагають оперувати різними підходами у практичній роботі при моделюванні конкретних СЕКК. Для проектування та структурування контенту застосовано елементи теорії графів. Для реалізації програмного забезпечення СЕКК використано апарат розподілених інформаційних систем класу «клієнт-сервер» і технології обміну інформацією у відкритих системах, а також методи структурного й об'єктно-орієнтованого проектування.

Вихідною інформацією при побудові математичних моделей процесів функціонування СЕКК є дані про призначення й умови роботи системи S . Ця інформація визначає основну мету моделювання системи S (рис. 6) і дозволяє сформулювати вимоги до математичної моделі M [10].

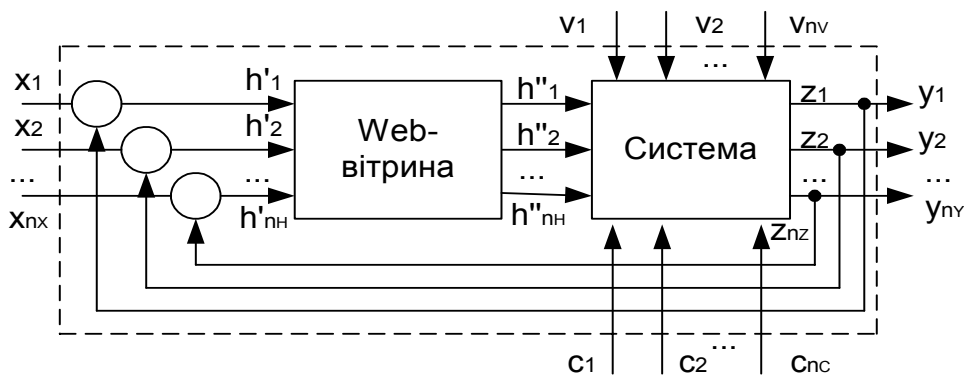


Рис. 6. Структура системи електронної контент-комерції

Формальну модель об'єкта моделювання СЕКК S відобразимо у вигляді множини величин, що описують процес функціонування реальної системи і утворюють такі підмножини: *вхідні впливи на систему* $x_i \in X, i = \overline{1, n_x}$; *впливи контенту на систему* $c_i \in C, i = \overline{1, n_c}$; *впливи зовнішнього середовища* $v_l \in V, l = \overline{1, n_v}$; *внутрішні (власні) параметри системи* $h_k \in H, k = \overline{1, n_h}$; *вихідні характеристики системи* $y_j \in Y, j = \overline{1, n_y}$ [10]. У перерахованих підмножинах x_i, v_l, h_k, y_j є елементами непересічних підмножин і містять детерміновані і стохастичні складові. Вхідні впливи, впливи зовнішнього середовища E та внутрішні параметри системи є

незалежними змінними, які у векторній формі мають вигляд $\vec{x}(t) = (x_1(t), x_2(t), \dots, x_{n_x}(t))$; $\vec{c}(t) = (c_1(t), c_2(t), \dots, c_{n_c}(t))$; $\vec{v}(t) = (v_1(t), v_2(t), \dots, v_{n_v}(t))$; $\vec{h}(t) = (h_1(t), h_2(t), \dots, h_{n_h}(t))$, а вихідні характеристики системи є залежними змінними й у векторній формі мають вигляд $\vec{y}(t) = (y_1(t), y_2(t), \dots, y_{n_y}(t))$. Процес функціонування СЕКК S описується в часі F_S вигляду:

$$\vec{y}(t) = F_S(\vec{x}, \vec{c}, \vec{v}, \vec{h}, t). \quad (1)$$

Функція (1) є математичним описом поведінки системи в часі t , тобто відбиває його динамічні властивості [19]. Тому математичні моделі такого виду є *динамічними моделями (системами) СЕКК*. Для статичних моделей математична модель (1) являє собою відображення між двома підмножинами властивостей Y і $\{X, C, V, H\}$, що у векторній формі записується як

$$\vec{y} = f(\vec{x}, \vec{c}, \vec{v}, \vec{h}). \quad (2)$$

Співвідношення (1) і (2) отримуються через властивості системи S у конкретний момент часу [10], *стан СЕКК S* , що характеризується векторами: $\vec{z}' = (z'_1, z'_2, \dots, z'_k)$ та $\vec{z}'' = (z''_1, z''_2, \dots, z''_k)$, де $z'_1 = z_1(t')$, $z'_2 = z_2(t')$, \dots , $z'_k = z_k(t')$ у момент часу $t' \in (t_0, T)$; $z''_1 = z_1(t'')$, $z''_2 = z_2(t'')$, \dots , $z''_k = z_k(t'')$ у момент $t'' \in (t_0, T)$ і т.д., $k = \overline{1, n_z}$. Стани СЕКК S у момент часу $t_0 < t^* \leq T$ повністю визначаються [10] початковими умовами $\vec{z}^0 = (z_1^0, z_2^0, \dots, z_k^0)$ [$z_1^0 = z_1(t_0)$, $z_2^0 = z_2(t_0)$, \dots , $z_k^0 = z_k(t_0)$], вхідними впливами $\vec{x}(t)$, впливами контенту $\vec{c}(t)$, внутрішніми параметрами $\vec{h}(t)$ і впливами зовнішнього середовища $\vec{v}(t)$, які мали місце за проміжок часу $t^* - t_0$, за допомогою двох векторних рівнянь:

$$z(t) = \Phi(\vec{z}^0, \vec{x}, \vec{c}, \vec{v}, \vec{h}, t); \quad \vec{y}(t) = F(\vec{z}, t) \quad (3)$$

Рівняння (3) за початковим станом \vec{z}^0 і екзогенними змінними $\vec{x}, \vec{c}, \vec{v}, \vec{h}$ визначає вектор-функцію $\vec{z}(t)$, а за отриманим значенням станів $\vec{z}(t)$ – ендогенні змінні на виході системи $\vec{y}(t)$ [10]. Ланцюжок рівнянь об'єкта «вхід – стани – вихід» визначає характеристики СЕКК: $\vec{y}(t) = F[\Phi(\vec{z}^0, \vec{x}, \vec{c}, \vec{v}, \vec{h}, t)]$.

Час у моделі СЕКК S розглядається на інтервалі моделювання $(0, T)$ як безперервний і дискретний, тобто квантований на відрізки довжиною $\square t$ тимчасових одиниць кожний, коли $T = m \square t$, де $m = \overline{1, m_T}$ – число інтервалів дискретизації [10]. *Математична модель системи* – це кінцева підмножина змінних $\{\vec{x}(t), \vec{c}(t), \vec{v}(t), \vec{h}(t)\}$ з математичними зв'язками між ними і характеристиками $\vec{y}(t)$. Якщо математичний опис СЕКК не містить елементів випадковості або вони не враховуються, тобто якщо стохастичні впливи зовнішнього середовища $\vec{v}(t)$ і стохастичні внутрішні параметри $\vec{h}(t)$ відсутні, то модель є *детермінованою*, тобто характеристики однозначно визначаються детермінованими вхідними впливами $\vec{y}(t) = f(\vec{x}, \vec{c}, t)$.

Вивчення динаміки, побудова моделей керування контентом є важливим і залишається недослідженим [2]. При розгляді динаміки тематичних потоків контенту в рамках логістичної моделі виявлена обмеженість розглянутої моделі, що відкриває шлях для подальших досліджень. Інтернет-простір із достатньою часткою умовності поділяється на дві складові – стабільну і динамічну з різними характеристиками з точки зору інтеграції керування контентом. Стабільна складова Інтернет містить інформацію «довгострокового» плану, а динамічна складова містить постійно обновлювані ресурси. Деяка частина цієї складової згодом вливається в стабільну, інша частина «зникає» або потрапляє у сегмент «схованого» Web-простору, не доступного користувачам за допомогою публічних інформаційно-пошукових

систем. Найбільш динамічним є сегмент контенту у вигляді новин. Він має найвищий рівень оновлення та – у ньому генеруються і поширюються великі обсяги даних. Тому він є найкращим об'єктом досліджень. Процеси старіння контенту, втрата його актуальності у відомій *моделі Бартон-Кейлера* [2] описуються рівняннями, що складається з двох компонентів: $m(t) = 1 - ae^{-t} - be^{-2t}$, де $m(t)$ – частка корисної інформації в загальному потоці через час T , перший від'ємник відповідає стабільним ресурсам, а другий – динамічним – новинам. Інформаційна динаміка в мережі обумовлена багатьма факторами, більшість з яких взагалі не піддаються точному аналізу. У рамках задачі моделювання як розумне допущення є загальний характер тимчасової залежності кількості тематичних публікацій у мережі, що визначається простими закономірностями. Для одержання більш адекватної залежності варто звернутися до складніших моделей [2].

Класична просторово-векторна модель керування контентом – модель $TF * IDF$ [2], де TF – це локальна частота терму (Term Frequency), а IDF – величина, зворотна частоті появи повідомлень у всьому потоці контенту, що містять цей терм (Inverse Document Frequency). У той час, як локальна частота терму в документі говорить про значущість терму в межах документа, то зворотна частота появи свідчить про унікальність терму у всьому потоці документів. Тому добуток цих величин – досить вдалий критерій визначення значущості терму – ваги. Повідомлення-новини старіють, втрачаючи свою актуальність з інтенсивністю, що визначається деяким емпіричним законом. Для ілюстрації припустимо, що це експонентний закон [2]. Один із запропонованих підходів до такої частини узагальнення, як рангування повідомлень, полягає у використанні параметричних множників, що залежать від часу, наприклад, визначимо вагу повідомлення як добуток елементів типу $TF * IDF * e^{-\alpha t}$, де α – деяка константа, t – інтервал часу, що минув з моменту появи повідомлення в потоці контенту (значення α – це коефіцієнт напіврозпаду актуальності контенту, тобто, якщо передбачається застосування експонентної моделі, це $e^{-\alpha t} = 1/2$, де t – період часу, що визначається експертним шляхом, протягом якого повідомлення через старіння втрачає свою актуальність наполовину). Облік старіння контенту (втрати частини актуальності) має велике значення при аналітичних дослідженнях, створенні інформаційних продуктів типу інформаційних портретів, основних сюжетів подій, рангуванні результатів роботи інформаційно-пошукових систем. Приблизна оцінка швидкості старіння контенту має величезну практичну цінність, так як допомагає тримати в полі зору тільки найактуальніший контент [2].

Лінійна модель керування контентом – $S = \langle Y, T, T_0, V \rangle$. Динаміка керування тематичного контенту (актуальність/старіння контенту) відбувається лінійно. Кількість повідомлень у момент часу t представимо формулою $y(t) = y(t_0) \pm v(t - t_0)$ [2], де $y(t)$ – кількість повідомлень на час t , v – середня швидкість збільшення/зменшення інтенсивності потоку тематичного контенту в часі (наприклад, у результаті старіння). Змістова складова контенту кількісно оцінена як флуктуація (відхилення від норми) потоку контенту – зміна стандартного відхилення

$$\sigma(t_i) = \sqrt{\frac{1}{i} \sum_{k=0}^i \{y(t_k) - (y(t_0) \pm v(t_i - t_0))\}^2}$$

Якщо ці величини змінюються як корінь квадратний з

часу, то процес зміни публікацій за темою вважаємо процесом з незалежними збільшеннями. Зв'язками з попередніми публікаціями можна знехтувати. У випадку поводження стандартного відхилення за часом як $\sigma(t) \propto t^\mu$, чим більше значення μ (ступінь зв'язку між випадковими подіями і приймає значення від $1/2$ до 1), тим вище кореляція між поточними і попередніми публікаціями [2].

Експонентна модель керування контентом – $S = \langle N, T_0, T, \lambda \rangle$. Процес збільшення (росту) актуальності/старіння контенту описується експонентною залежністю $N(t) = N(t_0) e^{\lambda(t-t_0)}$ [2], де λ – середня відносна зміна інтенсивності потоку контенту. Відносна зміна інтенсивності у визначений момент часу $\lambda(t_i) = (N(t_i) - N(t_{i-1})) / N(t_{i-1})$. Зміна флуктуацій величини $\lambda(t_i)$ щодо

середнього значення $\sigma(t_i) = \sqrt{\frac{1}{i} \sum_{k=0}^i \{\lambda(t_k) - \lambda\}^2}$. Якщо $\sigma(t)$ змінюється як корінь квадратний від часу, то процес є з незалежними збільшеннями, кореляція між окремими повідомленнями несуттєва. При наявності значної кількості залежних повідомлень справедливо $\sigma(t) \propto t^\mu$, $\mu > 1/2$, але обмежено 1. Значення μ , що перевищує $1/2$, говорить про наявність довгострокової пам'яті системи. Такі системи породжують клас процесів (автомодельних), для яких передбачається кореляція між кількістю повідомлень керування контентом у різні моменти часу [2].

Логістична модель керування контентом. Організації-генератори контенту у вигляді новин в абсолютній більшості працюють у стаціонарному режимі, що може характеризуватися максимальною ємністю простору контенту N (питання про розмірність параметрів та їхній вимір не розглядаємо). Це означає, що кожна організація-генератор робить потік контенту в середньому постійний за кількістю знаків і повідомлень [2]. Змінюються в часі лише обсяги повідомлень, що відповідають тій або іншій темі. Іншими словами, ріст кількості публікацій за однією темою супроводжується зменшенням публікацій на інші теми, так що для кожного проміжку часу T маємо: $S = \langle N, T, M \rangle$, $\int_0^T \sum_{i=1}^M n_i(t) dt = NT$, де $n_i(t)$ – кількість публікацій за одиницю часу, а M – загальна кількість усіх можливих тем. Частина $n_i(t)$ завжди дорівнює нулеві [2].

Підхід до аналізу потоків контенту як дискретних сигналів – $S = \langle W, D \rangle$. Вага повідомлення визначається за формулою: $W_D = \frac{\sum_{w \in D} w}{|D|}$ [2], де W_D – вага повідомлення, w – вага ключового слова з повідомлення, $|D|$ – кількість ключових слів у документі (у розглянутій моделі $1 \leq |D| \leq 12$). При значеннях β у зазначеному вище діапазоні $w \in$ монотонно зростаючою функцією від n . Середня вага унікального ключового слова дорівнює загальній кількості слів із потоку, розділений на кількість унікальних слів: $w(n) = n / v(n) = n^{1-\beta} / K$ [2].

Однією з найперспективніших є *логістична модель* [2]. Перевагою цієї моделі є те, що вона поєднує відносну простоту формулювання задачі з можливістю варіювати розв'язок за допомогою набору параметрів, що можуть мати прозорий фізичний зміст. Аналіз керування контентом, їхнє моделювання сьогодні стає одним із найбільш інформативних методів кількісного вивчення динаміки окремих тематичних напрямів. За зміною величин керування контентом судять про швидкість розвитку як окремих тематичних напрямів, так і всього інформаційного простору. Механізми, що базуються на узагальнених методах кластерного аналізу, виявляють повідомлення у множині контенту, що формують навколо себе нові тематичні напрями. Кластерний аналіз, теорія фракталів і автоматичних процесів при їхньому коректному застосуванні кількісно оцінюють ступінь зв'язку в тематичному контенті.

Одна з актуальних задач – побудова чіткої моделі сучасного інформаційного простору, що базується на досягненнях в області лінгвістики й інформатики, на методах, близьких до методів теоретичної фізики, строгому математичному інструментарії [2]. Розвиток алгоритмів, які навчають, забезпечує можливість побудови рекурентних процедур з інтерактивною участю людського інтелекту. Дослідження потоків контенту становить чималий інтерес для лінгвістів, математиків, фізиків, наприклад, у плані аналогового моделювання статистичних процесів, складних нелінійних систем з елементами самоорганізації. Семантика інформаційного простору обумовлює і розвиток нових методів кодування і стиски контенту, включаючи засоби забезпечення однозначності дешифрування повідомлень. Ступінь розвитку Web-простору визначається технологіями роботи з великим обсягом контенту. Перспективи охоплення простору контенту залежать від створення і розвитку ефективної інфраструктури, у рамках якої працюють програмні продукти з боку Web-серверів і користувачів. Часткове рішення названих

задач при наявності великої і дешевої експериментальної бази дозволяє реалізувати корисні й ефективні інструменти роботи і серфінгу в потоках контенту.

АНАЛІЗ ОТРИМАНИХ НАУКОВИХ РЕЗУЛЬТАТІВ

Web-сайт містить Web-вітрину, на якій відображається каталог контенту (з можливістю пошуку) і необхідні інтерфейсні елементи для введення реєстраційної інформації, формування замовлення, здійснення платежів через Інтернет, оформлення доставки (e-mail або on-line), одержання інформації про компанію-продавця і on-line допомоги. Реєстрація покупця відбувається при оформленні замовлення або при входженні в систему (рис. 7, а). Після вибору контенту від покупця вимагається заповнити форму, в якій вказується, як буде здійснена оплата/доставка. Для захисту персональної інформації взаємодія здійснюється по захищеному каналу (наприклад, за протоколом SSL). Після закінчення формування замовлення і реєстрації вся зібрана інформація про покупця надходить з Web-вітрини в торговельну СЕКК, де ініціюється запит до платіжної системи. Структура комплексу керування СЕКК реалізується у вигляді триланкової архітектури клієнт/сервер [1]. Процес опрацювання даних відбувається за схемою «клієнт-сервер застосувань – база даних». Запит, що надійшов, опрацьовується сервером застосувань, який зв'язується зі СД і платіжною системою, а при наявності підключення до бізнес-процесу організації, обмінюється даними з відповідними системами (рис. 7, б) [1-4].

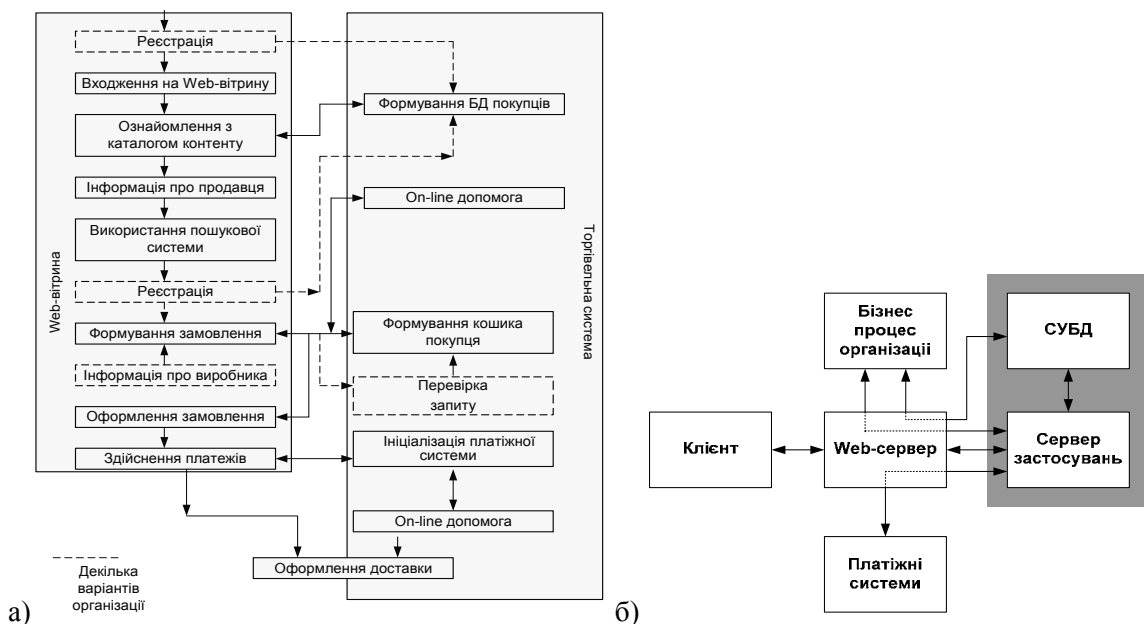


Рис. 7. а) Процес функціонування та б) типова структура СЕКК

СЕКК забезпечує ознайомлення, вибір категорії контенту, оформлення замовлення, здійснення взаєморозрахунків, відстеження виконання замовлення (рис. 8, а) [1]. Для функціонування СЕКК необхідно мати програмно-апаратні компоненти: Інтернет-вітрину на Web-сервері; е-каталоги; платіжну систему; ІС, інтегровану у фронт-офіс, тобто бек-офіс (бухгалтерія, відділ маркетингу та ін.). Для повнофункціонального СЕКК характерна складна система взаємозв'язаних операцій, методів, прийомів, який складається із шести фаз (рис. 8, б).



Рис. 8. а) Структурна схема та б) фази технологічного процесу роботи СЕКК

Збирання, систематизація й узагальнення контенту реалізується у вигляді комплексів контент-моніторингу, що виконують основну роботу зі збирання контенту з БД, СД, Інтернет тощо, і забезпечують створення документальних сховищ, що відповідають потребам споживачів [2]. Такі комплекси забезпечують постійне поповнення сховища контентом, ефективний одночасний доступ до БД багатьма користувачами, зручні засоби пошуку контенту (рис. 9).

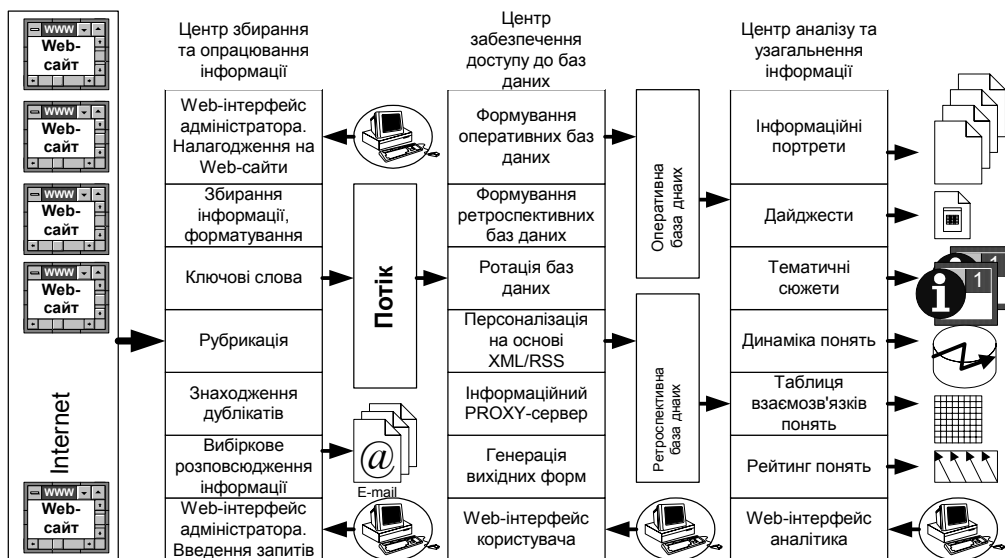


Рис. 9. Процеси комплексу контент-моніторингу

Створення анотованої БД [2] – це створення бази даних пошукових образів первинного контенту і їх кластеризація, тобто автоматичне формування груп контенту із близькими за деякими критеріями пошуковими образами (ПОД). При формуванні анотованої БД найважливіший аспект – формування бази даних анотацій, тобто пошукових образів кластерів (ПОК), що і використовуються у процесі пошуку. Ця база даних зв'язана з БД кластерів, кожен запис якої відповідає визначеному кластерові і містить його опис, виконаний методами автоматичного реферування). Методи автоматичного реферування використовуються для створення ПОД та описів, доступних користувачам (рис. 12, а). Задача виявлення нових подій з потоку контенту припускає, що на вхід відповідного програмно-технологічного комплексу послідовно надходять новий контент. Він надходить безпосередньо від засобів сканування (політематичний потік) або від контентного роутера, системи вибірного поширення контенту,

відібрані за тематичним запитом (рис. 12, б). Далі виявляються нові події, що описуються в контенті, для яких за допомогою окремих програмних модулів у тимчасовій ретроспективі формуються ланцюжки подібного контенту (сюжетні ланцюжки). Контент, що відображає різні нові події, є основою нових груп взаємозалежного контенту (кластерів), що приблизно заповнюються надалі. У цьому припущенні і полягає прогностичний момент технології виявлення нових подій. Кожен із кластерів може стати основою формування повноцінного сюжетного ланцюжка [2].

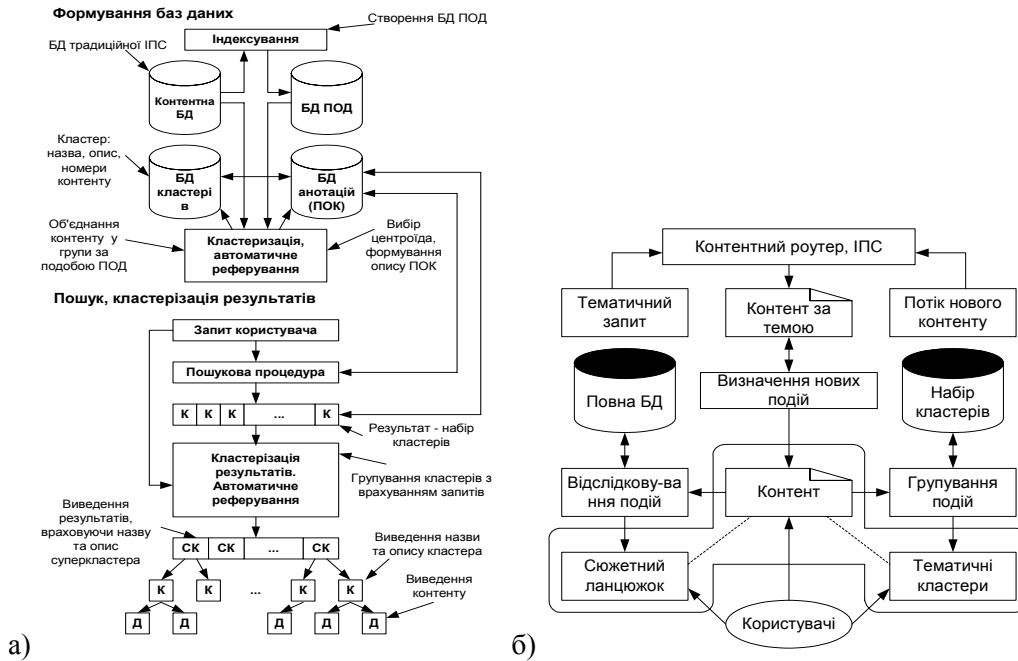


Рис. 12. а) Архітектура БД та б) процес визначення нових подій в СЕKK

Технологія синдикації контенту (рис. 13) містить у собі «навчання» програм збору даних структурним особливостям окремих джерел (Web-сайтів), безпосереднє сканування контенту, її приведення до загального формату (як правило, XML), а також класифікацію [2].

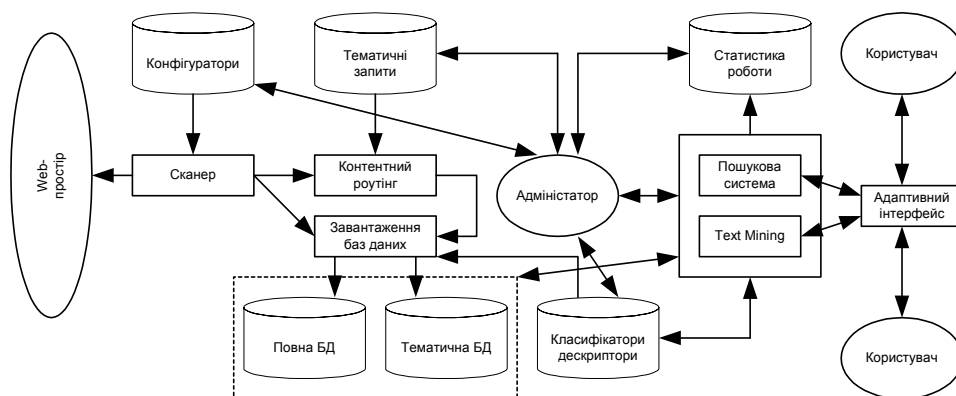


Рис. 13. Функціональна схема системи контент-моніторингу

Засоби класифікації та розподілу контенту являють собою інформаційно-пошукову систему виборчого поширення контенту (Контентний роутер). Контент, що надходить у систему, аналізується на відповідність тематичним запитам. У режимі діалогового доступу до БД забезпечується перегляд, пошук і відображення контенту [2]. На рис. 14 наведений ієрархічний принцип організації та функціонування мережі проксі-серверів СЕKK [2]. Проксі-сервер першого рівня забезпечує доступ до кешу, заповнюваного інтелектуальним сканером. Проксі-

сервери другого рівня завантажують контент з кешу проксі-серверу 1-го рівня та можуть доповнювати своє сховище даними, що скануються безпосередньо з Інтернету.

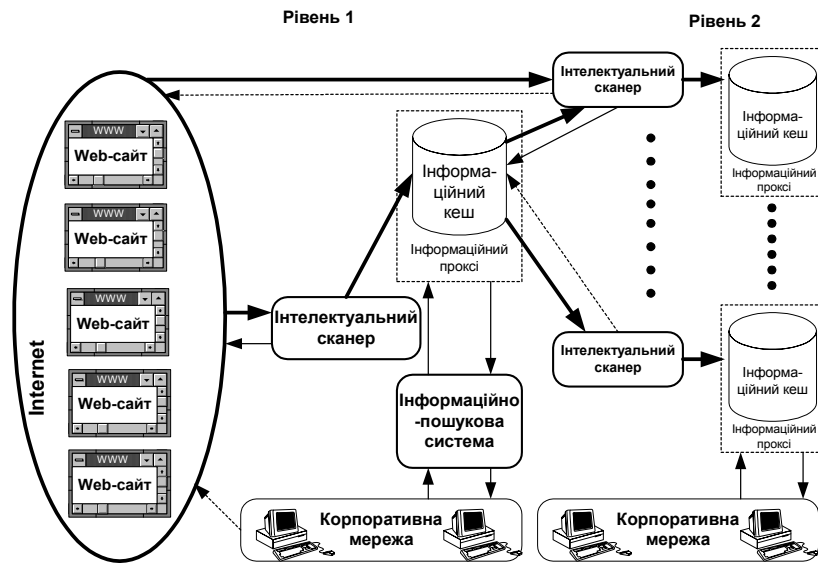


Рис. 14. Принцип організації мережі проксі-серверів

Інтернет-газета «Прес-Тайм» розташована за адресою www.presstime.com.ua (рис. 15, а) та має три основні підсистеми, які між собою взаємодіють: для редагування газети з боку журналістів, перегляду новин з боку клієнта, підписки клієнта на отримання газети. На рис. 15, б зображене ПЗ, необхідне для функціонування СЕКК та їх взаємодія між собою [3; 4].

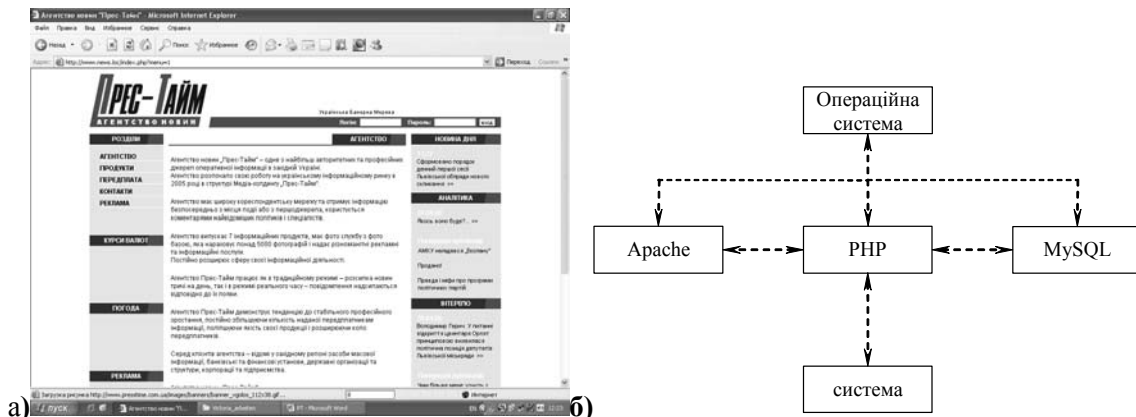


Рис. 15. а) Інтернет-газета та б) структурна схема середовища СЕКК

Проведення інформаційного аналізу функціонування СЕКК дає можливість оцінити якість контентного потоку для подальшого прийняття рішення відповідною особою. Для цього необхідно оцінити контент в межах від 0 до 1 відповідно до категорії (рис. 18, а) та скористатися відповідною формулою $f(S_1 \cdot S_2) = f(S_1) + f(S_2)$. Існує тільки одна функція аргументу S , яка задовольняє три наступні умови: $f(1) = 0$; якщо $S_1 > S_2$, то $f(S_1) > f(S_2)$; $f(S_1 \cdot S_2) = f(S_1) + f(S_2)$. Цією функцією є логарифмічна залежність $H = \log S$, за допомогою якої аналізуємо контентний потік. Визначення приналежності статті до тематики та регіону нерівно ймовірні, і кожне визначення є суб'єктивним, тому має свою ймовірність p_i (рис. 18, б-в). Логарифмічна залежність розподілу тематичного контенту $\tilde{I}_i = -p_i \cdot \log_2 p_i$.

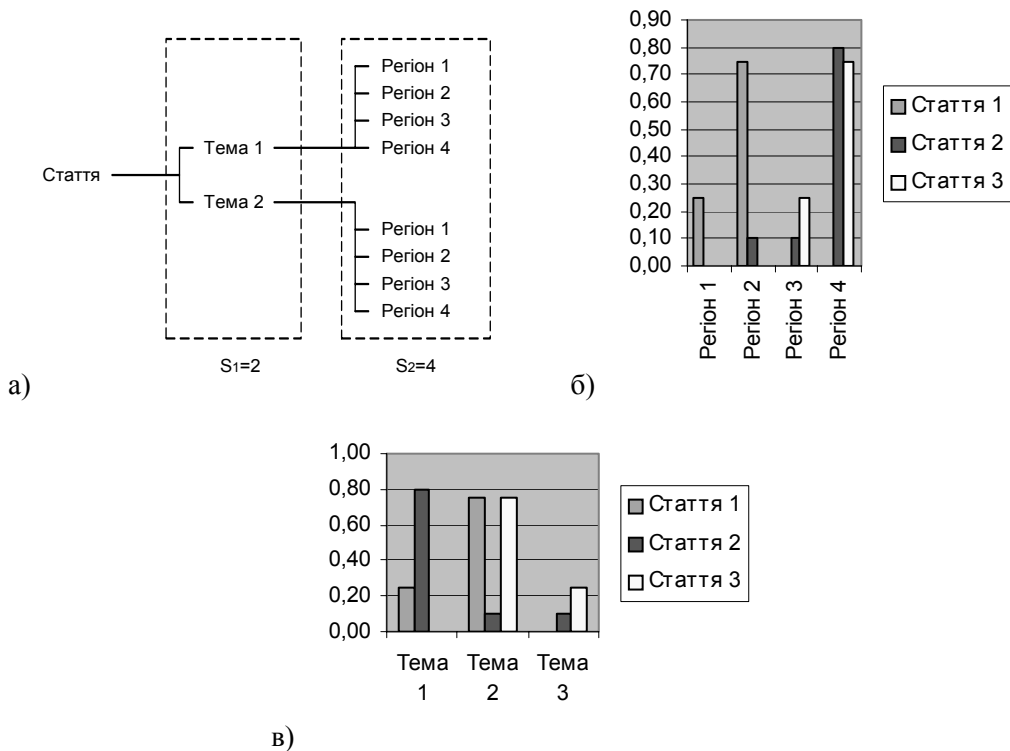


Рис. 18. Приклад а) категоризації контентного потоку та розподіл контенту відносно відповідної категорії: б) регіону, в) тематики

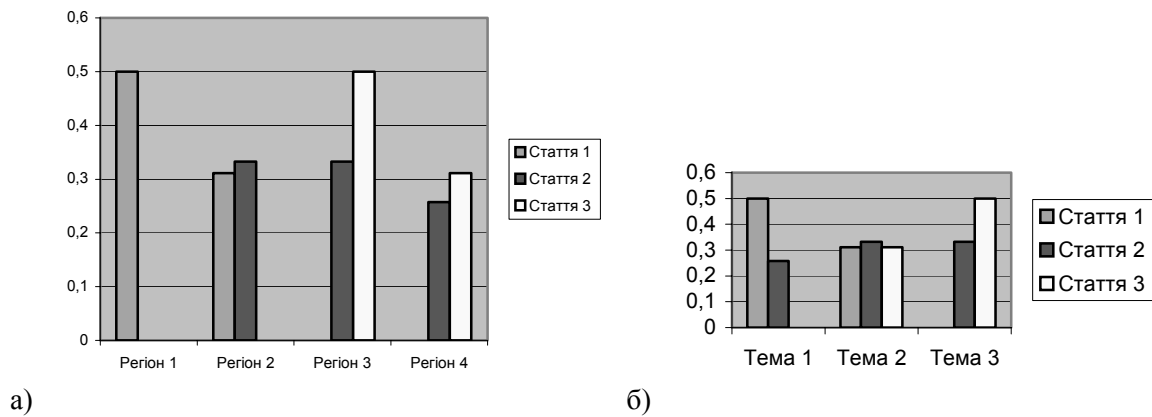
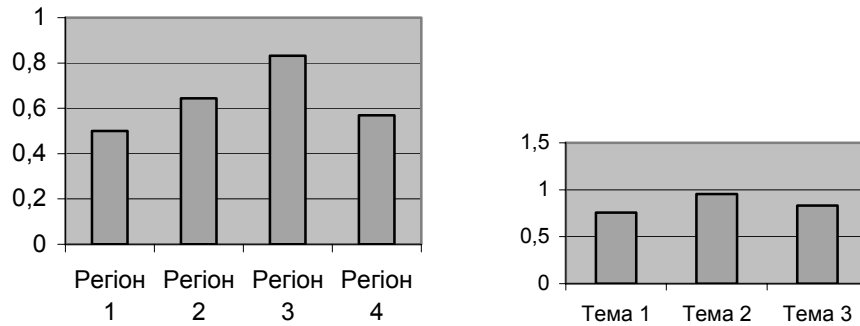


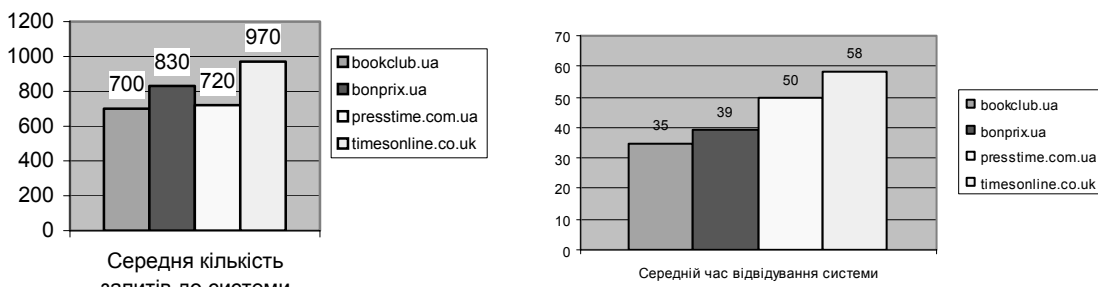
Рис. 19. Логарифмічна залежність розподілу контенту

Визначення проблемних областей контентного потоку $I = -\sum p_i \cdot \log_2 p_i$.

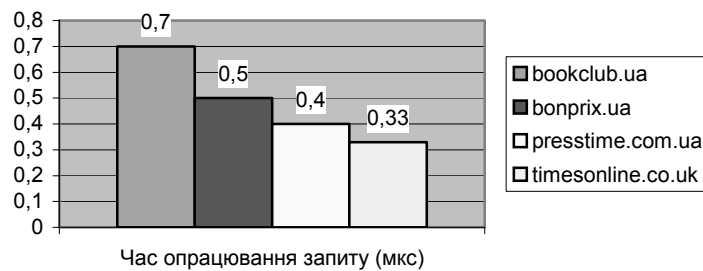
Інформаційний аналіз розподілу контенту дає можливість якісно оцінити контентний потік в СЕКК для подальшого прийняття рішень відповідною особою. Результати порівняння, подані на рис. 20, показують, що є необхідність приділити увагу для написання відповідного контенту типу «регіон 1», «тема 1». На рис. 21, а відображена порівняльна характеристика середньої кількості запитів до систем електронної комерції bookclub.ua, bonprix.ua, presstime.com.ua, timesonline.co.uk. На рис. 21, б відображена порівняльна характеристика середнього часу відвідування відповідних систем користувачами. На рис. 21, а відображена порівняльна характеристика у вигляді часу опрацювання запиту в досліджуваних системах.



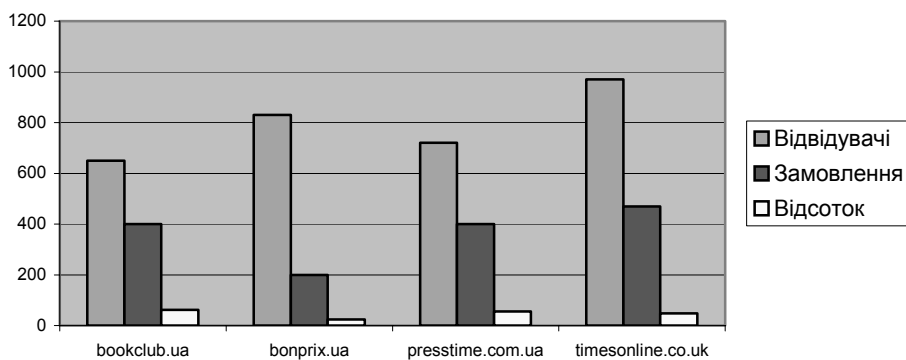
а) б)
Рис. 20. Проблемні області контентного потоку відповідної категорії



а) б)



в)



г)

Рис. 21. Порівняння систем електронної системи

ВИСНОВКИ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ НАУКОВИХ РОЗВІДОК

Стаття присвячена питанням моделювання та проектування інформаційних систем електронної контент-комерції. Побудовано архітектури та моделі систем електронної контент-комерції. Сформульовано новий підхід застосування та впровадження бізнес-процесів для побудови систем електронної контент-комерції. Введена нова класифікація систем електронної

комерції та систем електронної контент-комерції. Викладено теорію побудови, проектування та реалізації систем електронної контент-комерції. Визначена математична модель систем електронної контент-комерції та математична модель керування контентним потоком. Узагальнено поняття системи електронної контент-комерції. Розроблено методи проектування та реалізації системи електронної контент-комерції на прикладі Інтернет-газети. Розроблено інформаційну систему електронної контент-комерції – Інтернет-газету «Прес-тайм», яка на практиці відображає результати теоретичних досліджень.

ЛІТЕРАТУРА

1. Береза А.М. Електронна комерція. – К., 2002.
2. Ландэ Д.В. Основы моделирования и оценки электронных информационных потоков: монография / Д.В. Ландэ, В.М. Фурашев, С.М. Брайчевский, О.М. Григорьев – К.: ТОВ «Інжиніринг», 2006. – 176 с.
3. Берко А.Ю., Висоцька В.А., Чирун Л.В. Алгоритми опрацювання інформаційних ресурсів в системах електронної комерції // Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі. – 2004. – № 519. – С. 10-20.
4. Берко А.Ю., Висоцька В.А. Проектування навігаційного графу web-сторінок бази даних систем електронної комерції // Вісник Національного університету «Львівська політехніка». Комп'ютерні науки та інформаційні технології. – 2004. – № 521. – С. 48-57.
5. Джерк Н. Разработка приложений для электронной коммерции. – СПб.: Питер, 2001.
6. Катренко А.В. Системний аналіз об'єктів та процесів комп'ютеризації. – Львів: «Новий світ – 2000», 2003. – С. 286-322.
7. Козье Девид, Электронная коммерция. – М.: Русская Редакция, 1999.
8. Крупник А. Бизнес в Интернет. – М.: Микроарт, 2002.
9. Питерсон Дж. Теория сетей Петри и моделирование систем. – М.: Мир, 1984.
10. Советов Б.Я. Яковлев С.А. Моделирование систем. – М.: ВШ, 1985.
11. Успенский И. Энциклопедия Интернет-бизнеса. – СПб.: Питер, 2001.
12. Холмогоров В. Интернет-маркетинг. – СПб.: Питер, 2001.
13. Эймор Дэниел, Электронный бизнес. Эволюция и/или революция. – М.: Вильямс, 1999.

Рецензенти: д.т.н., проф. Фісун М.Т.,
д.т.н., проф. Казарєзов А.Я.

© Чирун Л.В., Висоцька В.А. 2009

Стаття надійшла до редакції 20.09.09