

## МАТЕМАТИЧНІ МОДЕЛІ ІНФОРМАЦІЙНИХ ПОТОКІВ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

© Висоцька В., 2010

**Проаналізовано основні проблеми електронної комерції та функціональних сервісів керування контентом. Запропоновано методи вирішення цих проблем.**

**Ключові слова:** контент, функціональні сервіси керування контентом.

**In the given article main problems of electronically commercial and content management interoperability services are analyzed. New methods for solution of discussed problems are proposed.**

**Keywords:** content, content management interoperability services.

### **Вступ. Загальна постановка проблеми**

Сьогодні Web-сайти є найважливішими складовими бізнесу, а інструментальні засоби для створення і розгортання Web-сайтів систем електронної контент-комерції (СЕКК) стають все більш гнучкими і простими у використанні. Проте створення складних Web-застосувань, що використовують не зовсім стандартні методи взаємодії (наприклад, блоги), не є тривіальним завданням. Дуже часто кожне застосування в організації може вимагати адаптації. Оскільки це захищений Web-сайт, необхідна система управління сесіями для підтримки завершення сесій і механізм підтвердження документа про умови використання сайту, які виконують свої функції до того, як стане можливим доступ до сайту [1–5]. Необхідно також підтримувати принцип безпосереднього управління в тому, що стосується редагування контенту – якщо є дія з додавання фрагмента контенту, необхідно, аби елемент управління взаємодією був поміщений поряд із відповідним контентом. Інші проблеми Web-сайту полягають у непослідовності візуальної мови, навігації та інформаційної архітектури. Контент часто захований усередині документів, що розміщуються в репозиторій, – не зазначено, що знаходиться в документі, тому його необхідно завантажити і проглянути. Із вищесказаного випливає, що модель репозиторія документів є помилковою моделлю для впровадження аналогічних систем. Необхідна система управління контентом із набагато ширшими можливостями.

### **Аналіз сучасних досліджень і публікацій**

Система управління контентом (англ. Content management system, CMS) — це система, що використовується для управління контентом чого-небудь (тобто даних без наперед визначеної структури наочного задання на протигагу структурованим даним, що зазвичай знаходяться під управлінням СУБД). Зазвичай такі системи використовуються для зберігання і публікації великої кількості документів, зображень, музики або відео. Окремим випадком такого роду систем є СЕКК. Подібні CMS дають змогу управляти текстовим і графічним наповненням, надаючи користувачеві зручні інструменти зберігання і публікації інформації.

Формальна модель СЕКК  $S = \langle X, C, V, H, Y \rangle$  – це множини величин, що описують процес функціонування 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  $y_j \in Y$ ,  $j = \overline{1, n_y}$ . У загальному випадку  $x_i$ ,  $v_l$ ,  $h_k$ ,  $y_j \in$  елементами непересічних підмножин і містять детерміновані і стохастичні складові [4–5]. Вхідні впливи  $x_i$ , впливи зовнішнього середовища  $E$  та внутрішні параметри системи є незалежними змінними, які у векторній формі мають вигляд

$$\vec{x}(t) = (x_1(t), x_2(t), \dots, x_{n_x}(t)); \quad \vec{c}(t) = (c_1(t), c_2(t), \dots, c_{n_c}(t));$$

$$\vec{v}(t) = (v_1(t), v_2(t), \dots, v_{n_v}(t)); \quad \vec{h}(t) = (h_1(t), h_2(t), \dots, h_{n_H}(t)),$$

а вихідні характеристики системи є залежними змінними й у векторній формі мають вигляд  $\vec{y}(t) = (y_1(t), y_2(t), \dots, y_{n_y}(t))$  [1–5].

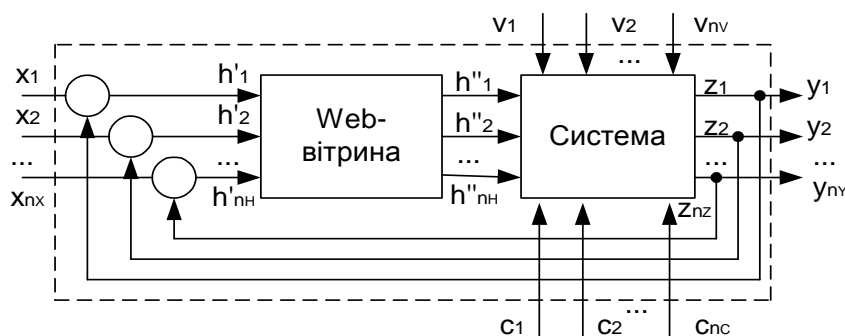


Рис 1. Структура системи автоматичного керування електронної контент-комерції

### Виділення проблем

Велика частина сучасних систем управління контентом реалізується за допомогою візуального (WYSIWYG) редактора — програми, яка створює HTML-код із спеціальної спрощеної розмітки, що дозволяє користувачеві простіше форматувати текст [1].

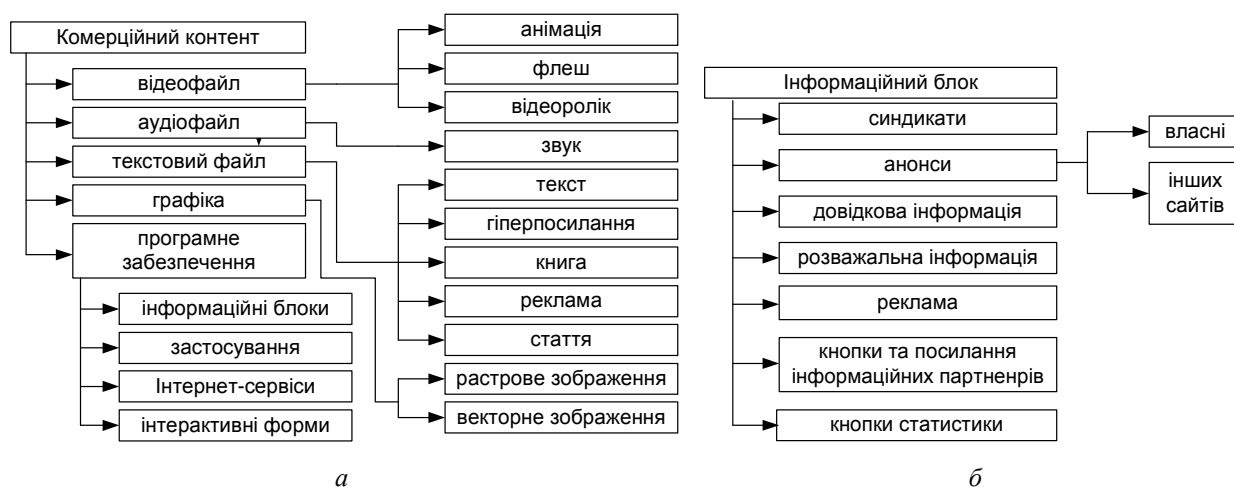


Рис. 2. Класифікація комерційного контенту (а) та інформаційних блоків (б)

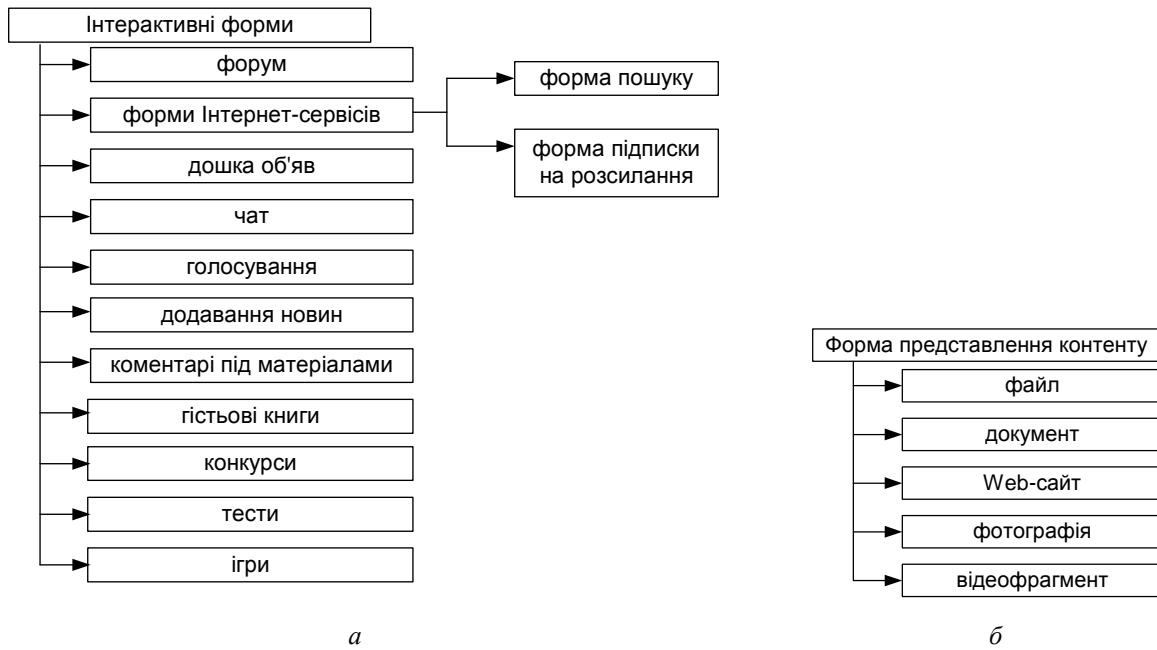


Рис. 3. Класифікація інтерактивних форм (а) та форми представлення контенту (б)

Інформаційні блоки як різновид контенту поділяють на синдикати (наприклад, блок «Погода»), анонси матеріалів інших розділів вашого сайту (з посиланням), анонси матеріалів, що знаходяться на інших сайтах (з посиланням), довідкова інформація, наприклад, сьогоднішнє свято, анонс заходу або розклад руху поїздів, розважальна інформація, наприклад, «анекдот дня», реклама, кнопки і посилання інформаційних партнерів, кнопки статистики [1].

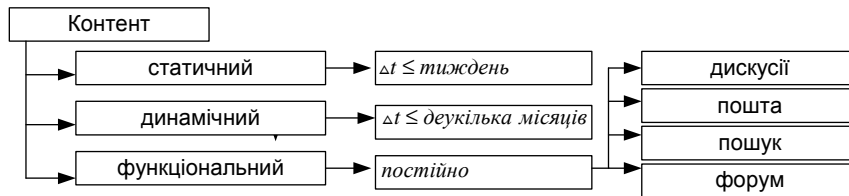


Рис. 4. Види контенту



Рис. 5. Категорії контенту

## Формування мети

Аналіз керування контентом, їхнє моделювання сьогодні стає одним з найінформативніших методів кількісного вивчення динаміки окремих тематичних напрямів. За зміною величин керування контентом судять про швидкість розвитку як окремих тематичних напрямів, так і всього контентного простору [5]. Стійкі статистичні зв'язки між окремими повідомленнями свідчать про кореляції окремих тематик, про ефективність посилань на публікації попередників, більш ранні роботи, цитування, републікації тощо. Механізми, що ґрунтуються на узагальнених методах кластерного аналізу, виявляють повідомлення в контентних потоках, що формують навколо себе нові тематичні напрями. Кластерний аналіз, теорія фракталів і автотельних процесів при їхньому коректному застосуванні кількісно оцінюють ступінь зв'язку в тематичних контентних потоках [1–5].

**Лінійна модель керування контентом** –  $C = \langle Y, T, T_0, V \rangle$ . Динаміка тематичного керування контентом (актуальність/старіння контенту) є лінійною [5]. Кількість повідомлень у момент часу  $t$  подамо формулою  $y(t) = y(t_0) \pm v(t - t_0)$ , де  $y(t)$  – кількість повідомлень на час  $t$ ,  $v$  – середня швидкість збільшення (зменшення) інтенсивності тематичного інформаційного потоку в часі (наприклад, у результаті старіння). Змістовну складову інформаційного потоку кількісно оцінено як флуктуацію (відхилення від норми) інформаційного потоку – зміна стандартного

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

**Експонентна модель керування контентом** –  $C = \langle N, T_0, T, \lambda \rangle$ . Процес збільшення (росту) актуальності/старіння контенту описується експонентною залежністю  $N(t) = N(t_0) e^{\lambda(t-t_0)}$ , де  $\lambda$  – середня відносна зміна інтенсивності інформаційного потоку.

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 неістотна [5].

**Логістична модель керування контентом.** Організації-генератори інформації у вигляді новин в абсолютній більшості працюють у стаціонарному режимі, що може характеризуватися максимальною ємністю інформаційного простору  $N$  (питання про розмірність параметрів та їхній вимір не розглядаємо). Це означає, що кожна організація-генератор робить потік інформації, у середньому постійний за кількістю як знаків, так і повідомлень. Змінюються в часі лише обсяги повідомлень, що відповідають тій або іншій темі. Іншими словами, ріст кількості публікацій з однієї теми супроводжується зменшенням публікацій з інших тем, так що для кожного проміжку

часу  $T$  маємо:  $C = \langle N, T, M \rangle$ ,  $\int_0^T \sum_{i=1}^M n_i(t) dt = NT$ , де  $n_i(t)$  – кількість публікацій за одиницю

часу, а  $M$  – загальна кількість усіх можливих тем. Передбачається, що частина  $n_i(t)$  завжди дорівнює нулеві [5].

### Аналіз отриманих наукових результатів

Процес проектування системи керування контентом є ітеративним і відбувається від аналізу до створення прототипу і пробних випробувань. На найраніших етапах необхідно підключити до процесу кінцевих користувачів за допомогою опитувальних листів, альтернатив проектування і прототипів різного ступеня готовності. Необхідно сконцентрувати рішення як на бізнес-цілях, так і на потребах кінцевих користувачів. Ще до визначення функціональних вимог і початку процесу розроблення необхідно підключити користувачів. Чим раніше підключити користувачів, тим швидше можна зрозуміти, що було б корисним і які рішення є правильними. Не прикладаючи значних зусиль, збирають доволі цінну інформацію, одночасно викликаючи у користувачів відчуття безпосередньої участі в процесі проектування. Крім того, завойовується довіра співтовариства користувачів. Ще одним компонентом осмислення проекту є розуміння інформаційної архітектури. Наприклад, завдяки користувачам можна з'ясувати, що одними з найважливіших вимог є доступ до трьох основних областей: інформація про робочі групи; інформація про минулі і майбутні конференції; доступ до інформації про всіх членів співтовариства. За зворотною реакцією користувачів стає також зрозуміло, що необхідна чітка, ясна і проста архітектура. Є три класи користувачів (або персонажів), що приходять на сайт: клієнти, керівники робочих груп і адміністратори. Ці персонажі визначають дизайн сайту і процес ухвалення рішень. Визначивши, яка інформація є важливою, і як вона пов'язана з основними класами користувачів, можна створити архітектуру контенту сайту, його ієрархію, способи представлення і способи взаємодії кожного класу користувачів з цією інформацією. Наприклад, інформація про конференції може містити питання порядку денного або сесії, планування цих питань, теми, що зачіпають на конференції, і питання, вирішені між конференціями. Під час аналізу з'являються додаткові функціональні можливості. Наприклад, для підтримки активного співтовариства приймається рішення дозволити дискусії і коментарі за вмістом, підтримку контекстної зворотної реакції та інтерактивну взаємодію. Новою важливою вимогою користувачів є використання унікального, але при цьому нейтрального, бренду або візуального ідентифікатора. Оскільки Web-сайт надає нейтральне місце, куди для взаємодії приходять самі різні користувачі, яскраво виражений візуальний зв'язок із якою-небудь компанією або оточенням міг би викликати небажану реакцію. Є декілька вимог до середовища розроблення системи керування контентом, наприклад, можливість вносити зміни до коду і тестувати внесені зміни автономно. Після тестування змін необхідно зробити цей код доступним для групи розробників. Такий ітеративний цикл розроблення змушує використовувати віддалену систему управління версіями CVS, яка давала б змогу синхронізуватися із членами групи розроблення і управляти базою вихідного коду, яку можна було би спільно використовувати.

Для підтримки проекту і використання всіх технологій в інтегрованому середовищі розроблення (IDE) можна обрати Eclipse. Eclipse надає велику кількість розширень і дуже тісно інтегрується із Concurrent Versions System (CVS). Eclipse-перспективи надають декілька видів і редакторів, що підтримують поточну діяльність. У нашому випадку такою діяльністю є редагування PHP-модулів і HTML-фрагментів. Eclipse відстежує локальні зміни в коді. Тобто, навіть якщо не виконується вхідний контроль файлів в CVS, то все одно можна відновити попередні версії файлу на локальній машині. Ця функціональна можливість Eclipse гарантує, що ніколи не загубиться вихідний код. Створивши централізоване середовище розроблення і тестування, можна оптимізувати роботу із кодом та іншими членами групи – час повинен витратитися на написання і тестування коду, а не на управління файлами та іншим ресурсами системи. Вибір моделі системи управління контентом впливає на необхідність використання інших інструментальних засобів. У випадку з Drupal це означає використання PHP, HTML і Cascading Style Sheets (CSS) для розроблення сторінок, а також MySQL для серверного сховища.

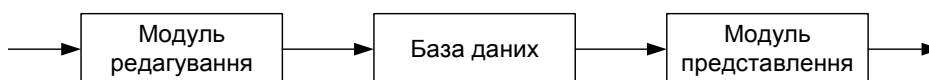


Рис. 6. Модель системи генерації сторінок за запитом

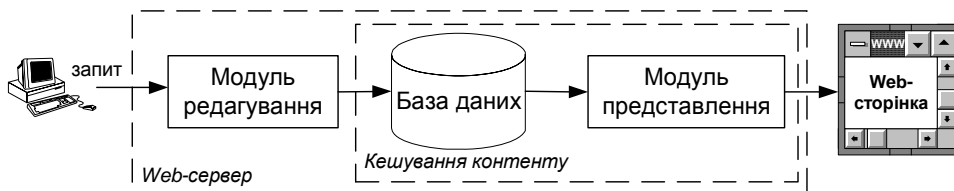


Рис. 7. Схема системи генерації сторінок за запитом

Переваги кешування сторінок з контентом:

1. Модуль представлення генерує сторінку один раз;
2. Сторінка в кеші існує деякий період часу  $\Delta t$  – поки актуальний контент;
3. Готова сторінка швидше завантажується з кешу;
4. Кеш оновлюється періодично вручну/автоматично: після деякого терміну часу  $\Delta t$ ; при внесенні змін у визначені розділи сайту або модифікації самого контенту.

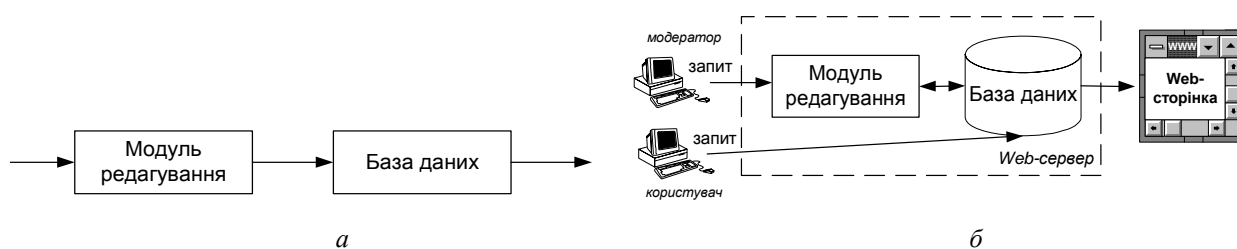


Рис. 8. Модель (а) та схема (б) системи генерації сторінок під час редагування

Недоліки системи генерації сторінок під час редагування: набір статичних сторінок, відсутність інтерактивності між відвідувачем та контентом сайту. Підходи до генерації сторінок змішаного типу: реалізація шляхом кешування; реалізація через формування інформаційних блоків.

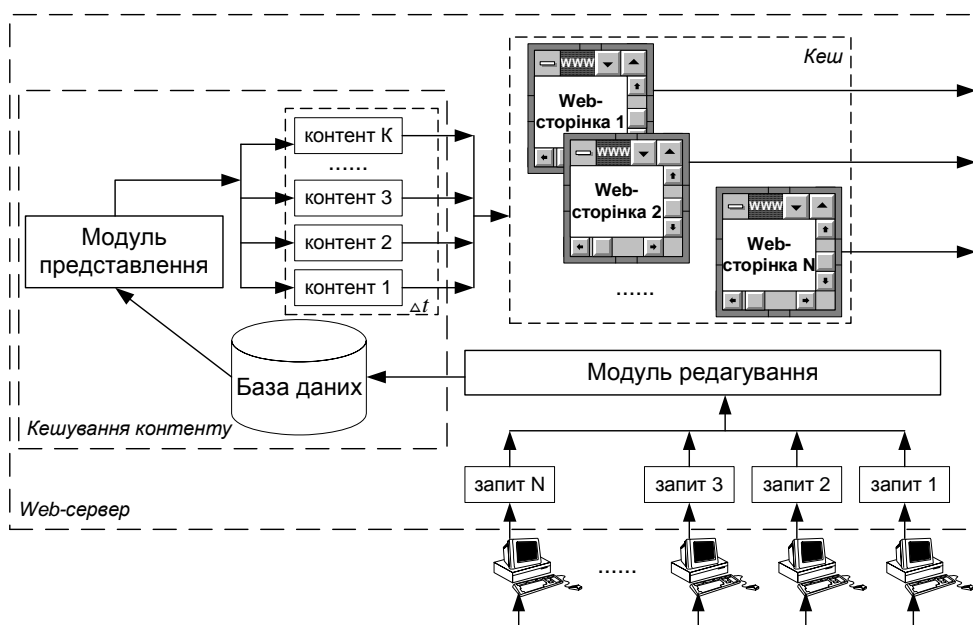


Рис. 9. Схема кешування згенерованих сторінок змішаного типу

Етапи генерації сторінок змішаного типу через формування інформаційних блоків: редагування сайту – збереження визначених інформаційних блоків; формування сторінок – збирання сторінки із інформаційних блоків при запиті відповідного контенту користувачем.

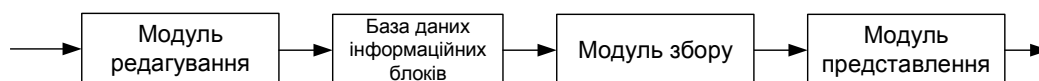


Рис. 10. Модуль системи генерації сторінок змішаного типу

Вимоги до СЕКК: розподіл контенту і представлення; вбудоване коментування вмісту; вбудоване редагування вмісту; тематичні дискусійні групи; управління правами доступу; пошук вмісту; аутентифікація перед прогляданням якого-небудь вмісту; управління сесіями, включаючи завершення, а також угоду з умовами використання; підтримка взаємодії співтовариства через дискусії; простий курс навчання роботі із системою управління вмістом; простий інтерфейс адміністрування системи управління вмістом (для передачі клієнтам)

На рис. 11 зображено детальну схему CMS.

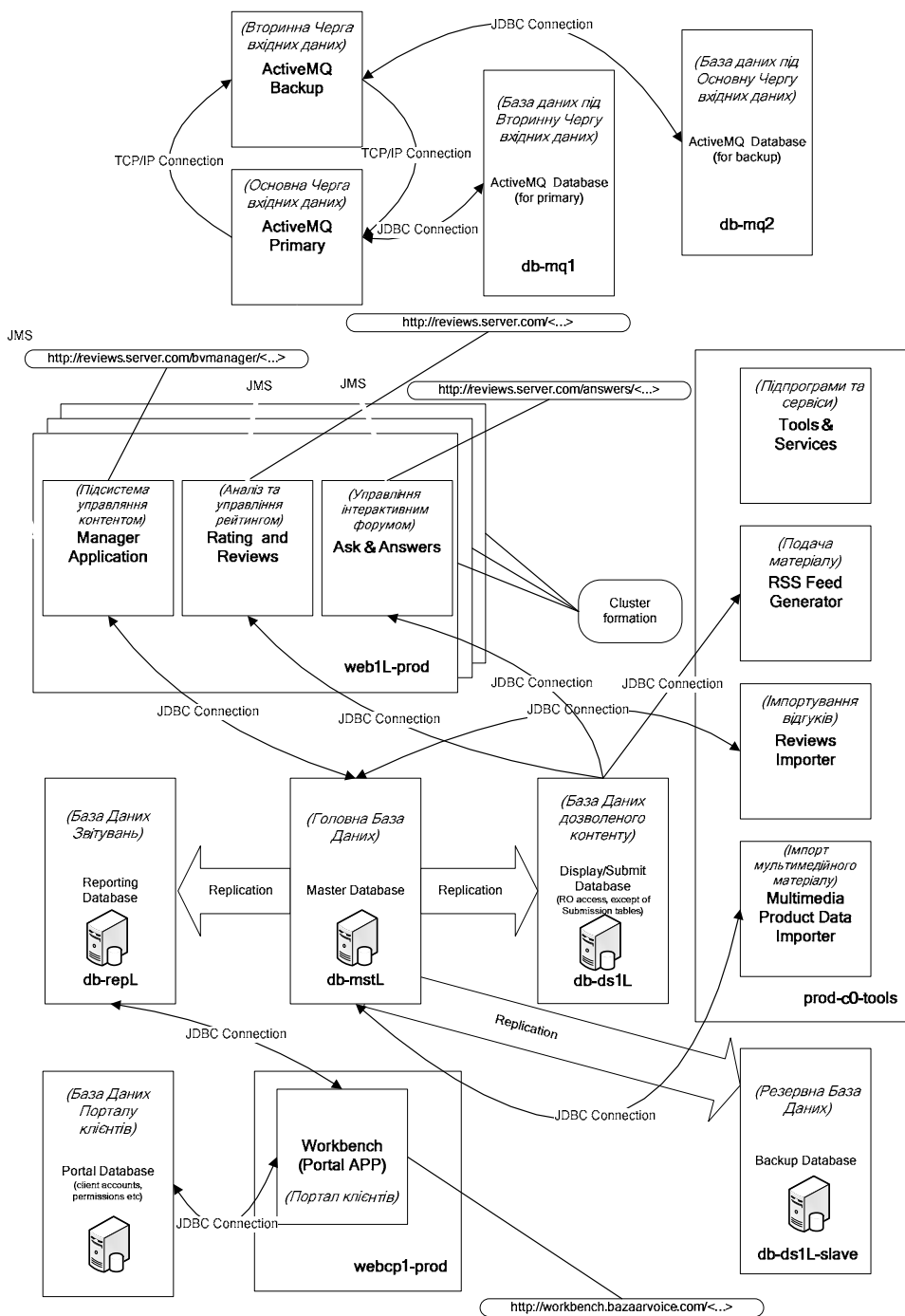


Рис. 11. Схема CMS

**Бази даних CMS (рис. 12).**

- *Master* – головна БД, яка є основною ланкою системи і для розподілу навантаження реплікує дані на дочірні бази.
- *Display* – дочірня БД від *Master*. З нею працює сайт клієнта на зчитування того контенту, який має дозвіл на публікацію.
- *Submit* – БД, яка 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 **Workpanch** аплікацією.

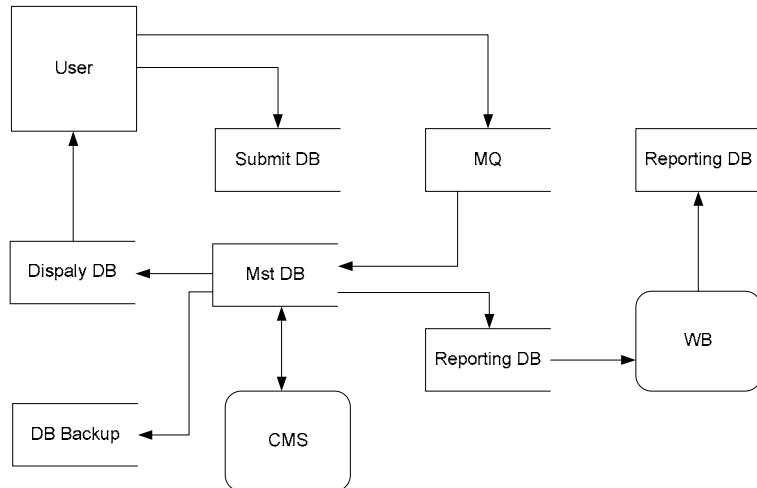
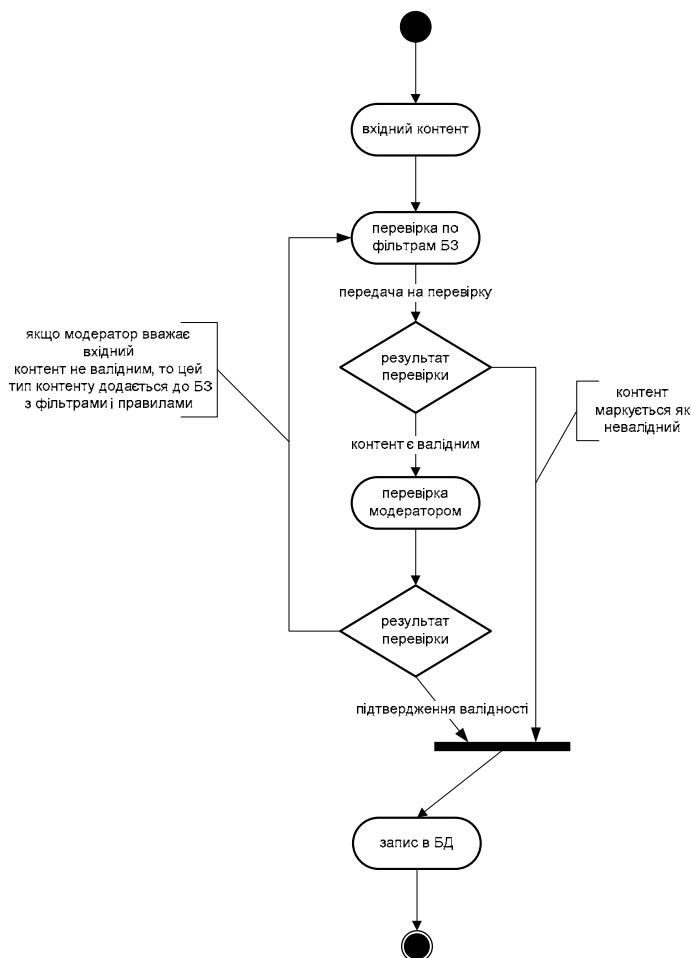


Рис. 12. Система прийняття рішень з поширення електронних видань



Запропонована CMS дає змогу:

- будувати направлені моделі спілкування з кінцевим споживачем;
- керувати динамічним матеріалом, генерованим кінцевими споживачами з боку клієнта як такого;
- надавати можливості кінцевим споживачам “будувати бізнес-клієнта”;
- реалізовувати новий тип сприйняття світу як нової альтернативної реальності.

Рис. 13. Алгоритм інтелектуальної компоненти CMS



Правила роботи фільтру із самонавчанням.

- Фільтр нецензурної лексики.
- Фільтр за кількістю символів.
- Фільтр за наявністю URL-посилань.
- Фільтр за “blacklist” списком.
- Фільтр за рейтингом.
- Фільтр за наявністю посилань на конкурентів.
- Фільтр за IP адресами.
- Фільтр за ID користувачами.

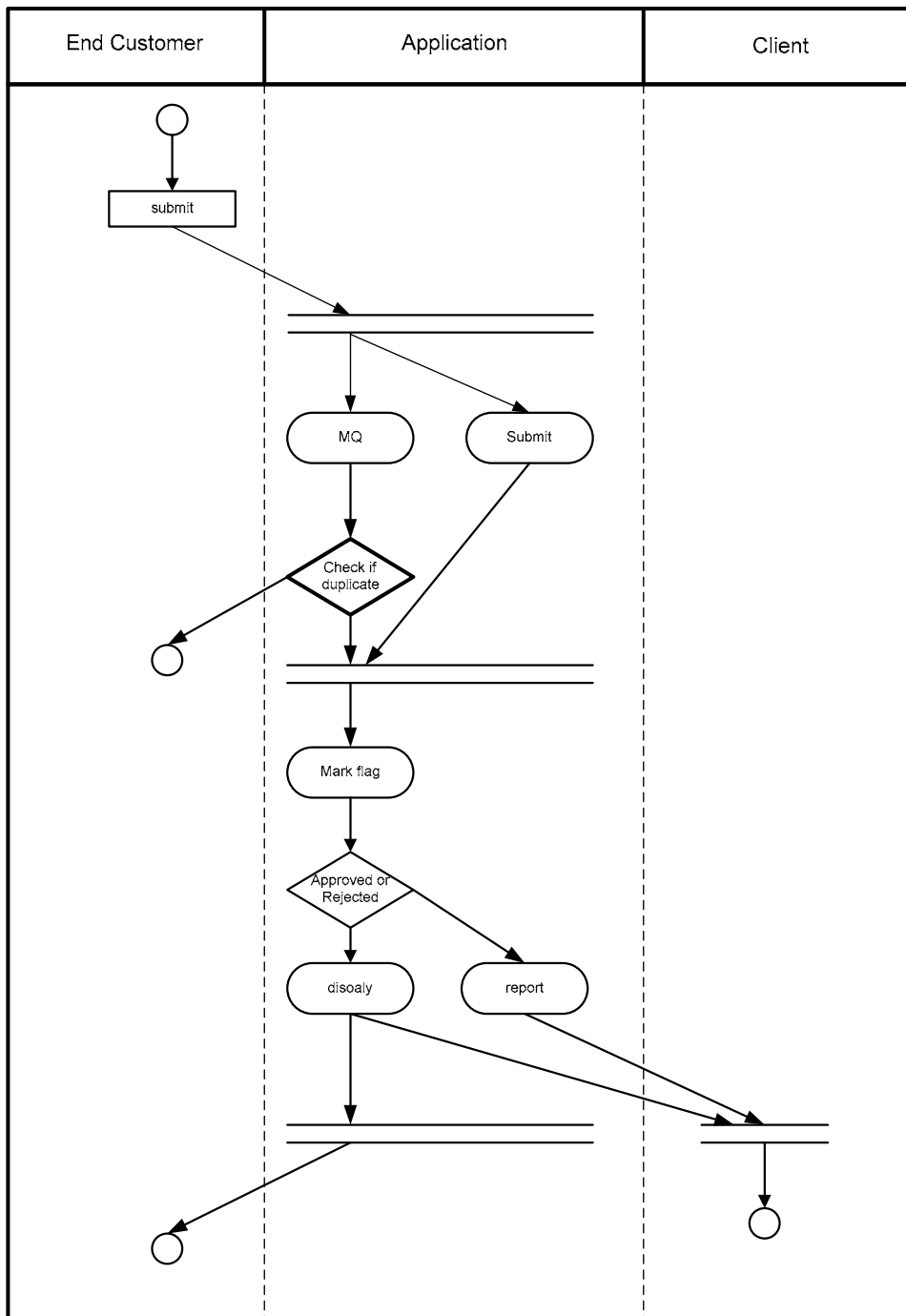


Рис. 14. Алгоритм роботи CMS

Отже, інформаційна насиченість за принципом “мультимедіа” дає змогу сприймати світ в цілісному комплексі відчуттів. Звідси випливають нові вимоги до вирішення завдань маркетингу в комп'ютерному інформаційному середовищі – концепції Інтернет-маркетингу.

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

Розглянуто проектування, розроблення і розгортання Web-сайту для спільної роботи з використанням програмного забезпечення із відкритими вихідними кодами. У статті був представлений огляд проекту, вимоги і порівняння декількох систем управління контентом, які автори проаналізували. У статті автори розглянули гнучку методологію проектування в процесі розроблення застосувань керування контентом. Цей процес може використовуватися для проектування призначеної для користувача взаємодії із Web-сайтами або застосуваннями.

1. Берко А.Ю. Системи електронної контент-комерції / А.Ю. Берко, В.А. Висоцька, В.В. Пасічник. – Львів: Вид-во Нац. ун-ту “Львівська політехніка”, 2009. – 612 с. 2. Ландэ Д.В. Сканер системи контент-мониторинга *InfoStream* / Д.В. Ландэ // *Открытые информационные и компьютерные интегрированные технологии: Сб. науч. трудов.* – Харьков: НАКУ «ХАИ», 2005. – Вып. 28 – С. 53–58. 3. Ландэ Д.В. Определение тематической направленности запросов путем анализа набора рейтинговых источников / Д.В. Ландэ, С.М. Брайчевский // *Открытые информационные и компьютерные интегрированные технологии: Сб. научн. трудов.* – Харьков: Нац. аэрокосмический ун-т «Хай», 2005. – Вып. 29. – С. 169–174. 4. Советов Б.Я. Моделирование систем (2-е изд.) / Б.Я. Советов, С.А. Яковлев. – М.: Высшая школа, 1998 р. 5. 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. – 348 с.