

Thermal simulation of heterogeneous structural components in microelectronic devices / V.I.Gavrysh, D.V. Fedasyuk // Semiconductor Physics, Quantum Electronics & Optoelectronics. – 2010. – 13, No. 4. – P. 439-443. 5. Подстригач Я.С. Термоупругость тел неоднородной структуры / Я.С. Подстригач, В.А. Ломакин, Ю.М. Коляно. – М.: Наука, 1984. – 386 с. 6. Коляно Ю.М. Методы теплопроводности и термоупругости неоднородного тела. – Киев: Наук. думка, 1992. – 280 с. 7. Корн Г. Справочник по математике для научных работников и инженеров / Корн Г., Корн Т. – М.: Наука, 1977. – 720 с. 8. Коздоба Л.А. Методы решения нелинейных задач теплопроводности. – М.: Наука, 1975. – 229 с. 9. Ломакин В.А. Теория упругости неоднородных тел. – М.: Изд-во Моск. ун-та, 1976. – 376 с. 10. Берман Р. Теплопроводность твердых тел. – М.: Мир, 1979. – 288 с.

УДК 004.738.5

В. Висоцька, Л. Чирун, Л. Чирун

Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

МОДЕЛЮВАННЯ ПРОЦЕСІВ ОПРАЦЮВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

© Висоцька В., Чирун Л., Чирун Л., 2011

Запропоновано модель життєвого циклу 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 та запропоновано методи вирішення цих проблем.

Ключові слова: інформаційний ресурс, контент, система управління контентом, життєвий цикл контенту, система електронної контент-комерції.

The article proposed to model the 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.

Keywords: information resources, content, content management system, content lifecycle, electronic content commerce system.

Вступ. Загальна постановка проблеми

Інформаційний ресурс (англ. Information resources) – це сукупність документів або масивів документів в інформаційних системах (бібліотеках, архівах, фондах, банках даних, депозитаріях, музейних сховищах тощо), яка є об'єктом дії системи електронної контент-комерції (СЕКК). *Контент* (англ. content) – це важлива рушійна сила економічного зростання та соціальних змін. *Контент* – це множина інформаційних ресурсів та продуктів, які збережені в середовищі комп'ютерної ІС і доступні для застосування користувачам цієї системи [1]. *Контент* – це дані без наперед визначеної структури наочного подання в протилежність структурованим даним, що зазвичай знаходяться під управлінням СУБД. Розвиток ІТ привів до того, що в сучасній світовій економіці контент став ключовим поняттям. Успішний розвиток Інтернету та зростання електронної контент-комерції в черговий раз довели, що інформаційний сектор економіки є найдинамічнішим і найприбутковішим. Існує три основні визначення поняття контенту [1]:

- 1) будь-яке інформаційне значуще наповнення ІС (зокрема Web-вузла), наприклад, тексти, графіка, мультимедіа;
- 2) частина повідомлення, яка не опрацьовується і не змінюється в процесі передавання;
- 3) змістовна частина даних документа (на відмінну від атрибутів), яке містить текст, зображення, відео, звук, сценарії, програми або будь-який інший матеріал аналогічно вмісту твердого носія.

Різновидом контенту є *комерційний контент* – це інформаційний товар або інформаційне наповнення Web-сайту, Інтернет-видавництва, маркетингових досліджень, консалтингових послуг, який є об'єктом бізнес-процесів СЕКК.

Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями

Актуальність теми зумовлена тим, що сучасний етап розвитку Інтернет-економіки обумовив розвиток нових форм інформаційного обслуговування користувачів ІС за допомогою технологій Інтернет-маркетингу і зростання потреб в актуальній інформації, яка відіграє роль виробничого фактора та стратегічного ресурсу. Документована інформація, яка підготовлена відповідно до потреб користувачів Інтернет-послуг і призначена для їх задоволення, відіграє роль виробничого фактора та стратегічного ресурсу, є інформаційним продуктом. Інформаційна послуга – це інформаційне обслуговування користувачів ІС, тобто дії суб'єктів щодо забезпечення споживачів інформаційними продуктами. Ринок інформаційних послуг – це сукупність економічних, правових, організаційних і програмних відносин з продажу і купівлі інформаційних продуктів та послуг (ІПП), що складаються між постачальниками і споживачами. Мета дослідження – розроблення моделі СЕКК та моделей життєвого циклу контенту як складової частини СЕКК. Мета дослідження визначає необхідність розв'язання таких задач: аналіз функціонування СЕКК; розроблення формальної моделі СЕКК; розроблення моделі опрацювання інформаційних ресурсів у СЕКК.

Аналіз останніх досліджень та публікацій

Система управління контентом (англ. Content management system, CMS) – це система, яка відповідає певному набору вимог (рис. 1, а) та призначена для управління контентом. Такі системи використовуються для зберігання і публікації великої кількості документів, зображень, музики або відео. Окремим випадком такого роду систем є СЕКК [1]. Такі CMS дозволяють управляти текстовим і графічним наповненням, надаючи користувачеві зручні інструменти зберігання і публікації інформації. На рис. 1, б подана класифікація CMS. Системи управління Web-контентом розробляються для генерації контенту всередині порталів (рис. 2), які мають ті самі проблеми (динамічне збирання, кешування контенту, його безпека тощо), що й інші різновиди Web-застосувань.

Рівень цінності контенту визначає його привабливість для споживача. Інтеграція контенту робить портали привабливими, а інтеграція застосувань – корисними, оскільки користувачі все частіше тяжіють до виконання на порталі щораз більшої кількості застосувань, які повинні виконуватися на самому порталі. Адміністративна панель CMS системи управління Web-сайтом дає змогу змінювати або додавати нову інформацію для різних мовних версій Web-сайту. Зміни на сайті відображаються відразу після внесення і збереження.

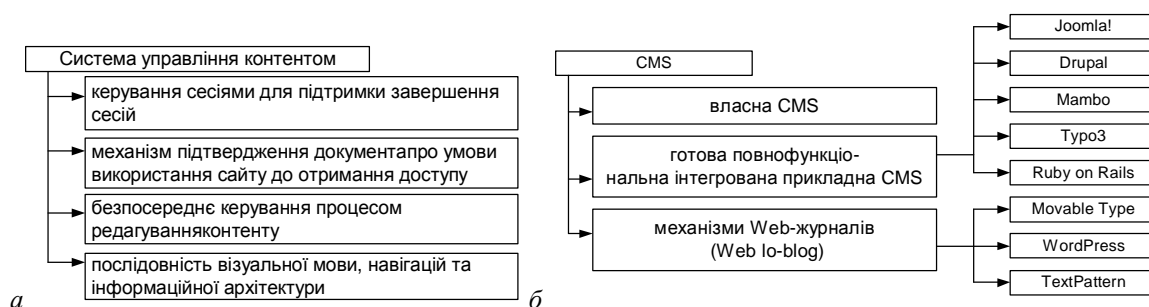


Рис. 1. Вимоги (а) та класифікація (б) систем управління контентом



Рис. 2. Призначення (а) та стандартна верстка CMS (б)

Системи управління Web-сайтом – це інструменти, що дають змогу моделювати розгалужені структури сайтів і керувати їхнім інформаційним наповненням. Використання таких системи управління контентом сайту не вимагає спеціальних технічних навичок, таких як програмування або html-верстка. Впровадження системи управління Web-сайтом можна контролювати доступ до сайту і внесенням змін. Система управління контентом покликана максимально спростити управління Web-сайтом із збереженням гнучкості налаштувань і контролю. Базова система управління сайтом передбачає такі можливості: швидке поновлення інформації на Web-сайті; пошук інформації на сайті; збирання даних про клієнтів та потенційних клієнтів; формування і редагування опитувань; аналіз відвідування Web-сайту. Основні модулі CMS сайту подано в табл. 1. Використання системи управління контентом сайту не вимагає встановлення спеціального програмного забезпечення на робочому місці. Для редагування й адміністрування використовується звичайний браузер (Internet Explorer або аналогічний). Простота інтерфейсу і роботи із системою дає змогу полегшити управління сайтом і знизити подальші витрати на підтримку Web-ресурсу. Для роботи із системою необхідно мати лише початкові навички роботи в Інтернеті. Найпопулярнішими CMS є Drupal і Joomla! (табл. 2).

Початковою інформацією при побудові математичних моделей процесів функціонування СЕКК є дані про призначення й умови роботи системи CMS [2]. Ця інформація визначає основну мету моделювання СЕКК і дає змогу сформулювати вимоги до формальної моделі системи S (рис. 3, а) та моделей управління контентом (рис. 3, б).

Формальна модель СЕКК $S = \langle X, C, V, H, Y \rangle$ – це множини величин, що описують процес функціонування системи (табл. 3). Величини x_i, c_r, v_l, h_k, y_j є елементами підмножин, що не перетинаються, і містять детерміновані і стохастичні складові [1, 2]. Вхідні впливи x_i , впливи потоку комерційного контенту c_r , впливи зовнішнього середовища v_l та внутрішні параметри системи h_k є незалежними змінними, а вихідні характеристики y_j системи є залежними змінними [2]. Процес функціонування S СЕКК описується функцією $y(t) = f(x, c, v, h, t)$.

Таблиця 1

Основні модулі систем управління контентом

Назва модуля	Характеристика модуля CMS
Пункти меню	додавання, редагування, управління пунктами меню сайту будь-якого рівня
Статті	додавання, редагування, планування, публікація статей (сторінок сайту) у Word-подібному інтерфейсі
Новини	додавання, редагування та публікування новин
Фотогалерея	можливість робити фотогалереї з підгалереями, автоматичне масштабування фото
Дошка оголошень	додавання оголошень з фотографіями, описом та контактними даними
Налаштування	зберігаються всі налаштування сайту та системи управління сайтом
Користувачі	управління правами зареєстрованих користувачів
Каталог фірм	додавання, редагування, публікування фірм у підгрупах будь-якої вкладеності
Опитування	додавання, редагування опитувань на сайті, результати у вигляді графіків, кількість питань: 2–10

Рейтинг CMS на основі вимог

Вимоги	Ruby on Rails	Drupal	Mambo	Типо3	Movable Type	Word Press	Text Pattern	Joomla!
Простота встановлення	+/-	+/-	+	-	+	+	+	+
Крива навчання	+/-	+/-	+/-	-	+/-	-	-	+
Управління сесіями	+/-	+	+/-	+	-	-	-	+
Управління користувачами	+/-	+	+	+	+/-	+/-	+/-	+
Розширюваність	+	+	+/-	+	+/-	+/-	-	+
Маштабованість	+	+	+	+	+	-	-	+
Здібність використання тем	+/-	+	+/-	+/-	+/-	+/-	+/-	+
xHTML/CSS	+	+	+	-	+	+	+	+

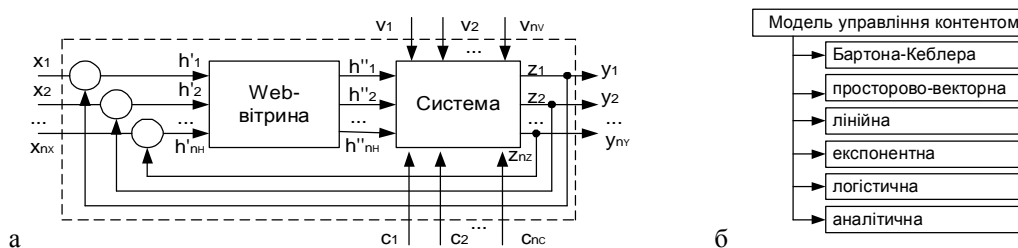


Рис. 3. Структура SECC (а) та класифікація (б) моделей управління контентом

Складові системи електронної контент-комерції

Множини	Позначення	Діапазон	Векторна форма
вхідні впливи на систему	$x_i \in X$	$i = \overline{1, n_x}$	$\mathbf{x}(t) = (x_1(t), x_2(t), \mathbf{K}, x_{n_x}(t))$
впливи потоку контенту на систему	$c_r \in C$	$r = \overline{1, n_c}$	$\mathbf{c}(t) = (c_1(t), c_2(t), \mathbf{K}, c_{n_c}(t))$
впливи зовнішнього середовища	$v_l \in V$	$l = \overline{1, n_v}$	$\mathbf{v}(t) = (v_1(t), v_2(t), \mathbf{K}, v_{n_v}(t))$
внутрішні (власні) параметри системи	$h_k \in H$	$k = \overline{1, n_h}$	$\mathbf{h}(t) = (h_1(t), h_2(t), \mathbf{K}, h_{n_h}(t))$
вихідні характеристики системи	$y_j \in Y$	$j = \overline{1, n_y}$	$\mathbf{y}(t) = (y_1(t), y_2(t), \mathbf{K}, y_{n_y}(t))$

Виділення проблем

Вивчення динаміки потоку та побудова моделей управління потоком комерційного контенту є важливим і залишається недослідженим. При розгляді в роботі [3] динаміки тематичних потоків контенту виявлено обмеженість моделей, що відкриває шлях для подальших досліджень. у табл. 4 подано порівняльні характеристики моделей управління контентом, які описують процес визначення та дослідження старіння/(актуальності) контенту, а не визначають етапи життєвого циклу (формування, управління, реалізація) контенту. Існуючі CMS не підтримують весь життєвий цикл контентного потоку та не вирішують основних проблем опрацювання інформаційних ресурсів – формування та реалізація контенту. Формальна модель SECC не розкриває механізмів управління контентом. Моделі CMS призначенні лише для визначення процесів старіння контенту, а деякі із них (логістична, аналітична) – і для тематичного контенту. Вони не вирішують проблеми формування, реалізації контенту, і вирішують далеко не всі проблеми управління контентом, наприклад, подання множини контенту кінцевому користувачу відповідно до його запиту, історії

або інформаційного портфеля, автоматичне формування дайджестів, інформаційних портретів, автоматичне виявлення тематичних сюжетів, побудова таблиць взаємозв'язку та розрахунок рейтингів понять, збирання та форматування інформації з різних джерел, виявлення ключових слів, понять та дублювання змісту контенту, автоматична рубрикація контенту, вибіркоче поширення контенту. Основний недолік формальних моделей СЕКК – це відсутність зв'язів між вхідною інформацією, контентом та вихідною інформацією.

Таблиця 4

Порівняльні характеристики моделей управління контентом

Модель	Перевага	Недолік
Бартона-Кеблера $C = \langle A, B, T \rangle$	Описує процес старіння контенту, втрати його актуальності. Рівняння моделі має точний розв'язок у вигляді дуже простої і зручної функції – експоненти. Можливість визначення швидкості розвитку окремих тематичних напрямів і всього контентного простору.	Є доволі сумнівною з погляду інтерпретації результатів. Монотонно зростає, не описує процеси, що за своєю природою повинні мати локальні екстремуми.
Класична просторово-векторна $C = \langle TF, IDF \rangle$	Визначення значущого терму в усьому потоці контенту. Можливість визначення найактуальнішого контенту із множини наявних.	Обов'язкове ранжування контенту, використання параметричних множників, що залежать від часу
Лінійна $C = \langle Y, T, T_0, V \rangle$	Визначення інтенсивності потоку тематичного контенту в часі (наприклад, у результаті старіння інформації)	Застосовують при лінійній динаміці управління тематичного контенту.
Експонентна $C = \langle N, T_0, T, I \rangle$	Описує процес старіння контенту, втрати його актуальності.	Кореляція між окремим контентом несуттєва.
Логістична $C = \langle N, T, M \rangle$	Поєднує відносну простоту формулювання задачі з можливістю варіювати розв'язок за допомогою набору параметрів, що можуть мати більш-менш прозорий фізичний зміст.	Вивчення динаміки лише окремого тематичного потоку. Розмірність параметрів та їхній вимір не враховуються.
Аналітична $C = \langle W, D \rangle$	Описує процес старіння контенту, втрати його актуальності.	Обов'язкова наявність словника ключових слів.

Формулювання мети

Формальну модель систем електронної контент-комерції подано у вигляді

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

де $X = \{x_1, x_2, \mathbf{K}, x_{n_x}\}$ – множина вхідної інформації, *Formation* – функція формування контенту, $C = \{c_1, c_2, \mathbf{K}, c_{n_c}\}$ – множина контенту, *Management* – функція управління контентом, *Realization* – функція реалізації контенту та $Y = \{y_1, y_2, \mathbf{K}, y_{n_y}\}$ – множина вихідної інформації (рис. 4). Вхідна інформація x_i є незалежними змінними і у векторній формі має вигляд $\overset{\mathbf{1}}{x}(t) = (x_1(t), x_2(t), \mathbf{K}, x_{n_x}(t))$, а контент c_j та вихідна інформація y_k є залежними змінними й у векторній формі мають відповідно вигляд $\overset{\mathbf{1}}{c}(t) = (c_1(x_i, t), c_2(x_i, t), \mathbf{K}, c_{n_c}(x_i, t))$ та $\overset{\mathbf{2}}{y}(t) = (y_1(t + \Delta t), y_2(t + \Delta t), \mathbf{K}, y_{n_y}(t + \Delta t))$.



Рис. 4. Модель життєвого циклу контенту в СЕКК

Процес формування контенту описується функцією вигляду $c(x_i, t) = Formation(u_F, x_i, t)$, де $u_F(x_i)$ – множина умов формування контенту, тобто $u_F(x_i) = (uf_1(x_i), uf_2(x_i), \mathbf{K}, uf_{n_{uf}}(x_i))$. Контент формується так: $c_j = \{ \mathbf{U} uf_k \mid (x_i \in X) \wedge (\exists uf_k \in U_F), U_F = U_{F_x} \vee U_{F_x}, i = \overline{1, m}, k = \overline{1, n} \}$.

Процес реалізації контенту описується функцією вигляду $y(t + \Delta t) = Realization(u_R, c, t, \Delta t)$, де $u_R(x_i)$ – множина умов реалізації контенту, тобто $u_R(x_i) = (ur_1(x_i), ur_2(x_i), \mathbf{K}, ur_{n_{ur}}(x_i))$. Вихідна інформація реалізується так $y_j = \{ \mathbf{U} ur_k \mid (c_i \in C) \wedge (\exists ur_k \in U_R), U_R = U_{R_c} \vee U_{R_c}, i = \overline{1, m}, k = \overline{1, n} \}$.

Тепер змоделюємо відповідні процеси опрацювання інформаційних ресурсів у СЕКК.

Аналіз отриманих наукових результатів

Модель формування інформаційних ресурсів. Відповідно до основного призначення, модуль формування контенту забезпечує збирання інформації з різноманітних Web-сайтів та її форматування; виявлення ключових слів, понять контенту; автоматичну рубрикацію контенту; виявлення дублювання змісту контенту; вибіркоче поширення контенту.

Формальна модель процесу формування інформаційних ресурсів подана у вигляді

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

де X – множина вхідної інформації з Web-сайтів або від модераторів (табл. 5); C – множина контенту.

Таблиця 5

Складові моделі формування інформаційних ресурсів

Множини	Позначення	Діапазон	Векторне подання
вхідна інформація	$x_i \in X$	$i = \overline{1, n_x}$	$x(t) = (x_1(t), x_2(t), \mathbf{K}, x_{n_x}(t))$
контент	$c_r \in C$	$r = \overline{1, n_c}$	$c(t) = (c_1(t), c_2(t), \mathbf{K}, c_{n_c}(t))$

Gathering – множина функцій збирання інформації з джерел, *Formatting* – множина функцій форматування інформації, перетворення на множину контенту, *KeyWords* – множина функцій виявлення ключових слів, понять, *Categorization* – множина функцій автоматичної рубрикації, *Backup* – множина функцій виявлення дублювання змісту контенту, *Dissemination* – множина функцій вибіркового поширення контенту. Оптимальне рішення, яке здатне допомогти орієнтуватися в динамічній вхідній інформації та інформації з різноманітних Web-сайтів, надають процеси синдикації даних (збирання інформації з джерел і подальше поширення її фрагментів відповідно до потреб користувачів). Технологія синдикації контенту містить “навчання” програм збирання даних структурним особливостям окремих джерел (Web-сайтів), безпосереднє сканування контенту, приведення до загального формату (XML), класифікацію. Засоби класифікації та розподілу контенту являють собою інформаційно-пошукову систему вибіркового поширення контенту (контентний роутер). Контент, що надходить у систему, аналізується щодо відповідності тематичним запитам. Релевантний контент розсилається користувачам, а також завантажується в тематичні БД.

Множина контенту формується за допомогою такої залежності:

$$C = Formatting(Gathering(X, U_G), U_F),$$

де U_G – скінченна множина умов збирання інформації з різноманітних джерел, U_F – скінченна множина умов форматування інформації.

Функція виявлення дублювання змісту контенту $C = Backup(C, U_B)$, де U_B – скінченна множина умов виявлення дублювання змісту контенту. Дубльованого за змістом контент у СЕКК виявляють лінгвостатистичними методами, що полягають у виявленні загальних термів, ланцюжки яких утворюють словесні сигнатури контенту.

Для функції виявлення подібного контенту і дублікатів справедливе правило рефлексивності: $c_i \mathbf{p} c_j$, $c_i \equiv c_j$, де c_i, c_j – довільний контент, “ \mathbf{p} ” – оператор подібності, “ \equiv ” – оператор дублювання. Оператор подібності не має властивості симетричності. Із подібності контенту c_i контентові c_j не впливає протилежно, тобто: $c_i \mathbf{p} c_j \not\Rightarrow c_j \mathbf{p} c_i$. Також не виконується умова транзитивності: $c_i \mathbf{p} c_j, c_j \mathbf{p} c_k \not\Rightarrow c_i \mathbf{p} c_k$. Наприклад, окремий контент може бути подібний до тексту з добірки, що його містить, але сама добірка може не бути подібною до цього контенту. Або контент може бути подібний до двох частин контенту, з яких він скопійований, але самі оригінали можуть істотно відрізнятися. Для відношення дублювання симетричність і транзитивність виконуються: $c_i \equiv c_j \Rightarrow c_j \equiv c_i$ та $c_i \equiv c_j, c_j \equiv c_k \Rightarrow c_i \equiv c_k$. Відношення, яке має властивості рефлексивності, симетричності та транзитивності, є відношенням еквівалентності, тобто відношенням збігу змісту або дублювання. Властивість дублювання контенту є більш твердим критерієм подібності, наприклад, збіг 3, 4 або 5 термів свідчить про деяку близькість змісту. На практиці кожному контенту C_i за наведеним вище алгоритмом збігу термів у контексті ставиться у відповідність вектор з елементами:

$$a_{ij} = \begin{cases} 1, & c_i \equiv c_j, \\ 0, & \text{інакше.} \end{cases}, \text{ тобто } U_B = \begin{bmatrix} a_{11} \mathbf{K} a_{1n} \\ \dots\dots\dots \\ a_{n1} \mathbf{K} a_{nn} \end{bmatrix}.$$

Умова симетричності в цих позначеннях записується так: $\forall i, j: a_{ij} = a_{ji}$, а умова транзитивності: $\forall i, j, k: a_{ij} = 1, a_{jk} = 1 \Rightarrow a_{ik} = 1$. Критерії подібності при зміні кількості порівнюваних у контексті термів для досягнення коефіцієнтів асиметричності і транзитивності

$$\min \left(\left(\frac{\sum_i \sum_j |a_{ij} - a_{ji}|}{\sum_i \sum_j a_{ij}} \right) \right) \rightarrow 0, \max \left(\left(\frac{\sum_i \sum_j \sum_k a_{ij} a_{jk} a_{ik}}{\sum_i \sum_j a_{ij}} \right) \right) \rightarrow \max,$$

де N – кількість контенту. Коефіцієнт асиметричності асоціюється з визначенням дублікатів, а рівень транзитивності – з повнотою. Перевірка коефіцієнтів асиметричності і транзитивності використовується для формальної перевірки наближення відношення до властивостей еквівалентності.

Функція автоматичної рубрикації $C_{Ct} = Categorization(KeyWords(C, U_K), U_{Ct})$, де $KeyWords(C, U_K)$ – функція виявлення ключових слів, понять, U_K – множина умов виявлення ключових слів, понять, U_{Ct} – множина умов автоматичної рубрикації, C_{Ct} – множина класифікованого контенту. Опрацювання множини контенту C для виявлення значущих ключових слів побудоване на принципі виявлення ключових слів за змістом (термів), тобто ґрунтується на законі Зіпфа і зводиться до вибору слів із середньою частотою появи (найвживаніші слова ігноруються за допомогою “стоп-словника”, а рідкісні слова з текстів повідомлень не враховуються).

Функція вибіркового поширення контенту $C_D = Dissemination(C_{Ct}, U_{Ds})$, де C_D – множина вибірково поширеного контенту, U_{Ds} – множина умов вибіркового поширення контенту. Вибіркове поширення списку контенту $C_{Ds} = \text{sup}(C_{Ct})$ залежить від рівня попиту на цей контент. Асоціативне правило формування списку контенту складається зі списку контенту C_{Ct} (первісний список) і списку контенту, вибраного із первісного – похідний список C_{Ds} , тобто $C_{Ct} \rightarrow C_{Ds}$. Формування

асоціативного правила – формування контенту, утвореного внаслідок об'єднання первісного і похідного списків. Ймовірність утворення асоціативного правила є ймовірність, з якою контент із первісного списку C_{Ct} з'являється разом із контентом із похідного списку C_{Ds} в БД, тобто

$$cf(C_{Ct} \rightarrow C_{Ds}) = \frac{\sup(C_{Ct} \cup C_{Ds})}{\sup(C_{Ct})}$$

До інших показників належать прибутковість (lf) та зростання (ct). Правила зростання та прибутковості формуються через функції відповідно

$$lf(C_{Ct} \rightarrow C_{Ds}) = \frac{\sup(\emptyset)}{\sup(C_{Ds})} \cdot cf(C_{Ct} \rightarrow C_{Ds}), \quad ct(C_{Ct} \rightarrow C_{Ds}) = \frac{\sup(\emptyset) - \sup(C_{Ds})}{\sup(\emptyset)[1 - cf(C_{Ct} \rightarrow C_{Ds})]}$$

Похідний список вибраних правил визначає користувач. Часто вживаним списком контенту є список, формування якого перевищує мінімальний рівень. Навіть при встановленні цієї умови списків часто вживаного контенту в запитах буде так само багато, тому обмежується вибірка: $ip(C_{Ct} \rightarrow C_{Ds}) = \min(\forall C_{Ct} \subset C_{Ct}, cf(C_{Ct} \rightarrow C_{Ds}) - cf(C_{Ct} \rightarrow C_{Ds}))$.

Моделі управління інформаційними ресурсами. Головне завдання модуля управління контентом (рис. 5 а): формування, ротація БД і забезпечення доступу до неї; формування оперативних і ретроспективних БД; персоналізація роботи користувачів, збереження персональних запитів користувачів і джерел, ведення статистики роботи; забезпечення пошуку в БД; генерація вихідних форм; інформаційна взаємодія з іншими БД.

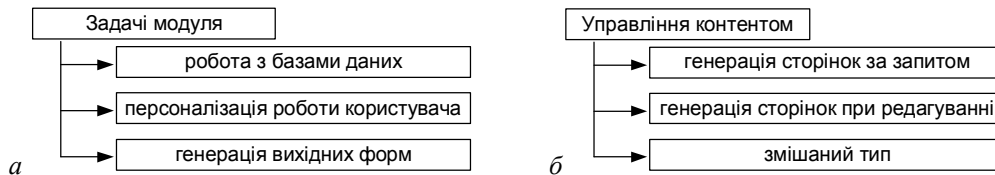


Рис. 5. Основні задачі (а)модуля та моделі управління контентом (б)

Класифікація моделей управління контентом (рис. 5, б).

1. Генерація сторінок за запитом відбувається на основі зв'язки, поданої на рис. 6.

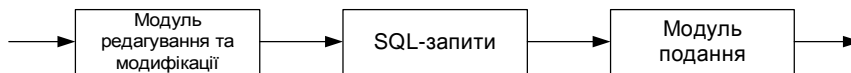


Рис. 6. Модель генерації сторінок за запитом

Формальна модель генерації сторінок за запитом – $Management_Q = \langle X, C, Q, R, Edit, Y \rangle$, де $X = \{x_1, x_2, \mathbf{K}, x_{n_x}\}$ – множина вхідної інформації; $C = \{c_1, c_2, \mathbf{K}, c_{n_c}\}$ – множина контенту; $Y = \{y_1, y_2, \mathbf{K}, y_{n_y}\}$ – множина сформованих сторінок; $Q = \{q_1, q_2, \mathbf{K}, q_{n_q}\}$ – множина запитів; R – функція формування та подання сторінки; $Edit$ – функція редагування та модифікації контенту.

Вхідна інформація $\mathbf{x}(t) = (x_1(t), x_2(t), \mathbf{K}, x_{n_x}(t))$ та запити $\mathbf{q}(t) = (q_1(t), q_2(t), \mathbf{K}, q_{n_q}(t))$ є незалежними змінними. Контент $\mathbf{c}(t) = (c_1(x_i, t), c_2(x_i, t), \mathbf{K}, c_{n_c}(x_i, t))$ та сформовані сторінки $\mathbf{y}(t) = (y_1(t + \Delta t), y_2(t + \Delta t), \mathbf{K}, y_{n_y}(t + \Delta t))$ є залежними змінними. Процес редагування та модифікації контенту описується функцією $\mathbf{c}(x_i, t) = Edit(\mathbf{c}, x_i, t)$. Процес формування сторінок описується функцією $\mathbf{y}(t + \Delta t) = R(\mathbf{q}, \mathbf{c}, w, t, \Delta t)$, тобто

$$y_i(t + \Delta t) = \left\{ \mathbf{U}(c_j, w) \mid (\forall c_j \in C_q) \wedge (\exists q_i(t) \in C_q), C = C_q \vee C_{\bar{q}}, j = \overline{1, m}, i = \overline{1, n} \right\}$$

Загальна вага контентного блоку визначається формулою: $\mathbf{w}(t) = Weight(Locat, KeyPhr, StatTr)$, де $Weight$ – функція визначення загальної ваги контентного

блоку, $Locat$ – коефіцієнт розташування блоку у вихідному контенті, $KeyPhr$ – ключові фрази, $StatTr$ – статистична вага контентного блоку. $Locat$ залежить від того, де з'являється такий фрагмент – на початку, у середині або наприкінці блоку, а також чи використовується він у ключових розділах контенту, наприклад, у висновку. $KeyPhr$ – це конструкції-маркери, що резюмують/оцінюють. $StatTr$ обчислюється як нормована за довжиною блоку сума ваг термінів, що входять до нього. Вага текстових фрагментів з урахуванням переваг користувачів описується уточненою формулою $w(t) = Weight(Locat, KeyPhr, StatTr, UsrPrf)$, де $UsrPrf$ – користувачькі переваги – залежить від частоти вживання слів/словосполучень, які наведені в запиті користувача та присутні в такому фрагменті.

2. Генерація сторінок при редагуванні. При внесенні змін до змісту сайту створюють набір статичних сторінок (рис. 7), тоді не враховується інтерактивність між відвідувачем і контентом. Формальна модель системи генерації сторінок при редагуванні – $Management_E = \langle C, Edit, Y \rangle$, де $C = \{c_1, c_2, \mathbf{K}, c_{n_c}\}$ – множина контенту; $Y = \{y_1, y_2, \mathbf{K}, y_{n_y}\}$ – множина статичних сторінок; $Edit$ – функція редагування та модифікації контенту. Процес формування статичних сторінок описується функцією $Edit$ вигляду: $y(t) = Edit(c, w, t)$.

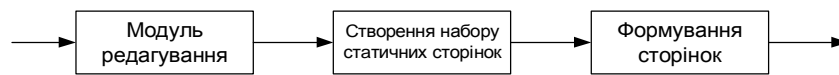


Рис. 7. Модель генерації статичних сторінок при редагуванні

3. Змішаний тип поєднує в собі переваги перших двох типів (рис. 8). Може бути реалізований шляхом кешування — модуль подання генерує сторінку один раз, надалі вона в декілька разів швидше підвантажується з кешу. Кеш може оновлюватися як автоматично, по закінченні деякого терміну або при внесенні змін до певних розділів сайту, так і за командою адміністратора. Інший підхід — збереження певних інформаційних блоків на етапі редагування сайту і складання сторінки з цих блоків при запиті відповідної сторінки користувачем. Формальна модель системи змішаного типу – $Management_M = \langle X, C, Q, R, Edit, Caching, Y \rangle$, де $X = \{x_1, x_2, \mathbf{K}, x_{n_x}\}$ – множина вхідної інформації; $C = \{c_1, c_2, \mathbf{K}, c_{n_c}\}$ – множина контенту; $Y = \{y_1, y_2, \mathbf{K}, y_{n_y}\}$ – множина сформованих сторінок; $Q = \{q_1, q_2, \mathbf{K}, q_{n_q}\}$ – множина запитів; R – функція формування та подання сторінки; $Edit$ – функція редагування та модифікації контенту, $Caching$ – функція формування кешу. Процес формування кешу сторінок $Cache = Caching(y, w, t, \Delta t)$, тобто $Cache = \{U y_i \mid y_i \in Y, t + \Delta t, i = \overline{1, n}\}$.

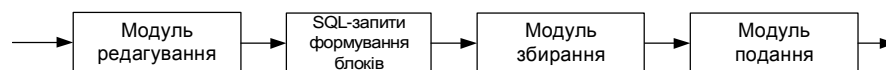


Рис. 8. Модель генерації сторінок змішаного типу

Модель реалізації інформаційних ресурсів. Модуль реалізації контенту забезпечує формування інформаційних портретів; формування дайджестів (короткий зміст публікацій); виявлення тематичних сюжетів; побудову таблиць взаємозв'язку понять; розрахунок рейтингів понять, виявлення, відстеження й кластеризацію нових подій. Формальна модель реалізації інформаційних ресурсів – це

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

де X – множина вхідної інформації, C – множина контенту, $BuInfPortr$ – 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, Y – множина вихідної інформації. Скінченна множина контенту $C = \langle C_P, C_D, C_T, C_C, C_R \rangle$, де C_P – множина контенту інформаційних портретів, C_D – множина контенту дайджестів, C_T – множина контенту тематичних сюжетів, C_C – множина контенту таблиць взаємозв'язку понять, C_R – множина контенту рейтингів понять. Множина контенту інформаційних портретів $C_P = BuInfPortr(X, U_P, C_P)$, де U_P – скінченна множина умов формування інформаційних портретів. Множина контенту дайджестів $C_D = BuDigest(X, U_D)$, де U_D – скінченна множина умов формування дайджестів.

Множина контенту тематичних сюжетів $C_T = IdThemTop(X, C_T, U_T)$, де U_T – скінченна множина умов виявлення тематичних сюжетів. Умови виявлення тематичних сюжетів $U_T = sim(c_i, PD) > a$, де n – величина спостереження потоку контенту; c_i – поточний контент спостереження; PD – словник; $sim(c_i, PD)$ – міра близькості контенту i до контенту словника, a – емпірично обумовлений параметр. Множина контенту таблиць взаємозв'язку понять: $C_C = ConCorrTablConc(C, U_C)$, де U_C – скінченна множина умов побудови таблиць взаємозв'язку понять. Множина контенту рейтингів понять $C_R = CalRankConc(C_C, Th, U_R, Sp)$, де U_R – скінченна множина параметрів розрахунку рейтингів понять, Th – критерій тональності контенту, Sp – функція визначення спаму. Множина вихідної інформації $Y = Realization(C_D, C_R)$. У випадку оцінювання тональності контенту простір гіпотез міститиме: $Th = H_{-1}$ – тональність негативна, $Th = H_0$ – тональність нейтральна і $Th = H_1$ – тональність позитивна. Із множини контенту з позитивною тональністю вибираються терми, характерні для цього контенту. Із них вибираються терми t зі значеннями ймовірності $p(t | H_1) \geq \frac{1}{2}$.

Висновки і перспективи подальших наукових розвідок

Враховуючи різні класи СЕKK і широкі можливості реалізації моделей, автори виділили основні закономірності переходу від процесів формування інформаційних ресурсів до їх реалізації. Відповідно запропоновано формальні моделі формування, управління та реалізації інформаційних ресурсів, які дають змогу оптимально реалізувати архітектуру СЕKK. Найістотнішим фактором, який враховувався при формалізації й алгоритмізації моделей, є використання інструментарію дослідження апаратно-програмних засобів ІТ. Виділення етапів моделювання СЕKK пов'язане з необхідністю залучення колективів розроблювачів різних спеціальностей (аналітиків, алгоритмістів, програмістів). Ступінь розвитку Web-простору визначається роботою з величезним обсягом інформації, що нагромадились в Інтернеті. Web 2.0 характеризується переходом від мережі документів до мережі даних, що за необхідності агрегуються в семантично зв'язаний контент за допомогою Web-сервісів. Передбачається існування єдиного інформаційного простору у вигляді множини одиниць даних, що можуть розміщуватися на численних сайтах в Інтернеті. Користувач одержуватиме контент шляхом агрегування на своєму робочому місці цих інформаційних одиниць. Перспективи охоплення інформаційного простору залежать від створення і розвитку ефективної інфраструктури, у межах якої працюють програмні продукти з боку Web-серверів і користувачів. Навіть часткове вирішення названих задач за наявності великої і дешевої експериментальної бази дасть змогу реалізувати корисні й ефективні інструменти роботи і навігації в контентних потоках, наприклад, за допомогою СЕKK.

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. Советов Б.Я. Моделирование систем (2-е изд.) / Б.Я. Советов, С.А. Яковлев. – М.: Высшая школа, 1998. 3. 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.