

*Accelerometer // Proc. of the 1st Inter. Conf. of Young Scientists “Perspective Technologies and Methods in MEMS Design” (MEMSTECH 2005). – Lviv–Polyana, Ukraine, 2005. – P. 52–53. 10. Teslyuk V., Zaharyuk R. Model of Capacitive Microaccelerometer // Proc. of the 2nd Inter. Conf. of Young Scientists “Perspective Technologies and Methods in MEMS Design” (MEMSTECH 2006). – Lviv–Polyana, Ukraine, 2006. – P. 86.*

УДК 681.518:681.327.8

В.А. Висоцька

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

## УМОВНА ЕНТРОПІЯ ТА ЕНТРОПІЯ КОНТЕНТУ В СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

О Висоцька В.А., 2008

Проаналізовано основні проблеми електронної комерції контенту та запропоновано методи вирішення цих проблем. Інтернет-система електронної комерції – цілеспрямована множина об’єктів довільної природи з набором зв’язків між ними та між їхніми властивостями. Інакше кажучи, Інтернет-система електронної комерції – все, що складається з взаємопов’язаних частин. Проаналізовано основні моделі систем електронної комерції. Запропоновано узагальнену модель таких систем.

In the given article main problems of electronically commercial of content are analyzed. New methods for solution of discussed problems are proposed. The Internet-system of electronic commerce – purposeful set of objects of the any nature with a set of connections between them and between their properties. In other words, the Internet-system of electronic commerce – everything, that consists of the interconnected parts. Some problems of electronic commerce are analyzed in proposed paper. Generalized model of such kind systems are proposed.

### Вступ. Загальна постановка проблеми

Більшість Internet- і Intranet-додатків систем електронної комерції визначаються як “статистичні” або “динамічні” залежно від змісту і доступу до даних (рис. 1). Системи публікації баз даних систем електронної комерції забезпечують простий статичний доступ до динамічних даних (контроль складських залишків через Internet або перегляд статусу замовлення через Intranet). Обидва класи додатків (нарівні з простими додатками накопичення даних) розвинені в транзакційні бізнес-додатки. Клас WebOLTP – кероване діалогове опрацювання запитів (Online Transaction Processing – OLTP), а Web – як засіб доступу. Такі додатки є не простими програмами для перегляду даних, а додатками для обробки в реальному режимі часу важливої ділової інформації, наприклад, операції в банку, приймання замовлень, роботи з клієнтами [1–3].



Рис. 1. “Статичні” та “динамічні” додатки

Сучасну інфраструктуру Інтернет-систем подано на рис. 2, а. Вона об'єднує:

- Web-браузер, що призначений для відображення сторінок у форматі HTML (Hypertext Mark-up Language);
- Web-сервер, який забезпечує зберігання й управління HTML-сторінок.

Стандартні засоби зв'язку між браузером і сервером – це протокол HTTP (Hypertext Transfer Protocol). Базова інфраструктура систем електронної комерції була розроблена і досі цілком придатна для публікації статичної інформації, наприклад, даних маркетингових досліджень. Як показано на рис. 2, б, базова інфраструктура Інтернет-систем останнім часом була розширена в розумінні більшої динамічності додатків (інтерактивних можливостей користувачів) за рахунок [4]:

- простих форм запитів і форматування даних на основі JavaScript для браузера;
- API для Web-сервера, таких, як, наприклад, NSAPI і ISAPI, що дають змогу браузерам виконувати додатки на Web-сервері;
- серверів динамічної обробки, які перетворюють дані з БД на сторінки у форматі HTML.

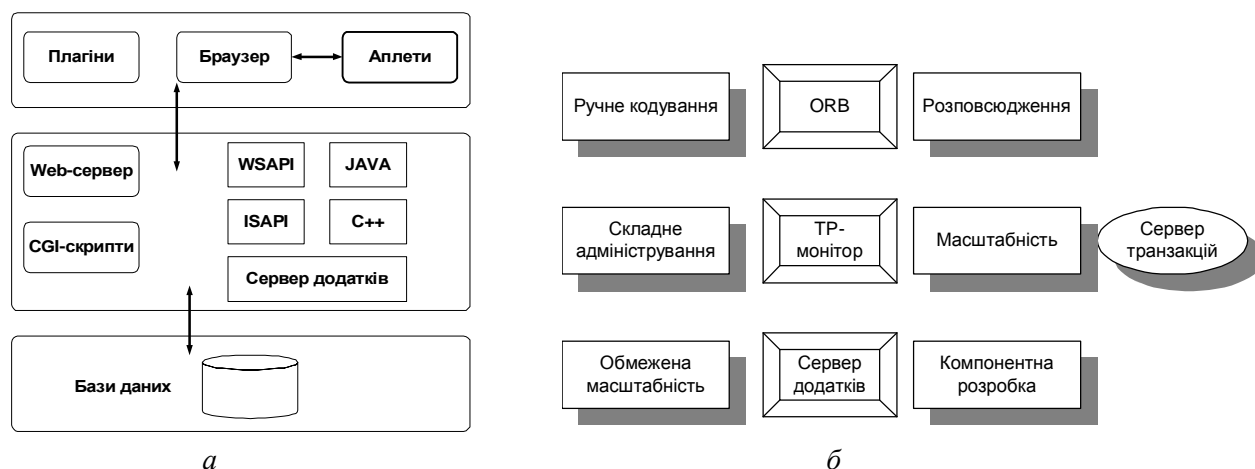


Рис. 2. Архітектура систем електронної комерції:  
а – базова/розширена; б – WebOLTP

### Аналіз сучасних досліджень і публікацій

Розширена інфраструктура Web-систем електронної комерції за рахунок динамічного опрацювання даних (здатність сервера повертати дані браузеру відповідно до запиту користувача або в іншій інтерактивній формі) надає можливість створення важливого класу додатків – від систем підтримки прийняття рішень через Інтернет до персональних Інтернет-новин. WebOLTP – назва класу додатків, що виконують транзакції в Internet/Intranet/Extranet або традиційних корпоративних мережах. WebOLTP-архітектура і модель покликані задовольнити вимоги до WebOLTP-додатків. Виникнення трирівневої/багаторівневої архітектури (рис. 3) відповідає потребам WebOLTP стосовно масштабованості (scalability) і динамічного доступу за умови збереження усіх переваг базової архітектури. У традиційних системах клієнт/сервер для запуску додатків необхідно, щоб клієнтське ПЗ було встановлено заздалегідь. В архітектурі з так званим “тонким” клієнтом спеціалізоване ПЗ не обов’язково встановлювати на боці клієнта, оскільки компоненти, що виконуються, можуть завантажуватися з Web site для подальшої взаємодії з клієнтом. Отже, “тонкий” клієнт, або клієнт з “нульовою інсталяцією”, одержує дві ключові переваги під час роботи в мережі: універсальний доступ і зменшення витрат на інсталяцію та управління. Однак через наявність браузерів і HTML “тонким” клієнтам для динамічного управління бізнес-додатками необхідне використання додаткових засобів, таких, як Java-аплети. На відміну від додатків клієнт/сервер і їхніх попередників, користувачі WebOLTP розширяють межі відділу або компанії, використовуючи Extranet/Internet. З цими додатками доступ більше не обмежується невеликою кількістю працівників, які реєструють замовлення, а стає відкритим для тисяч користувачів, що одночасно виконують транзакції. Це потребує добре масштабованої архітектури сервера для побудови WebOLTP-додатків.

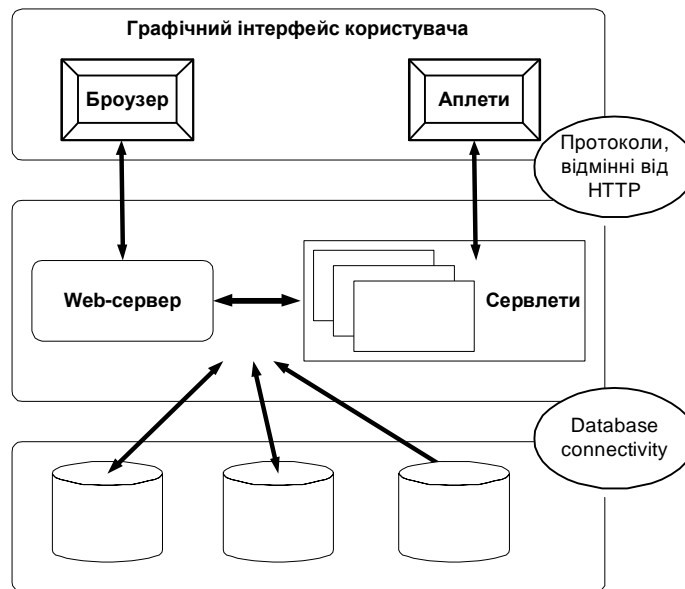


Рис. 3. Тривірнева архітектура систем електронної комерції

Додатки з чітко визначеним колом споживачів працюють з певними наборами дій і навантажень. Використання WebOLTP-додатків завдяки відкритому колу користувачів призводить до непередбачуваних навантажень. У міру того, як будь-який Web-додаток стає доступним, менеджерам Web-вузлів важко передбачити, коли і скільки користувачів намагатимуться під'єнатися. Щоб справитися з несподіваними стрибками навантаження, зберігши прийнятний час відгуку, потрібна наявність повністю регульованих й адаптованих систем. Кожна наступна генерація додатків має коротший “життєвий цикл”, ніж попередня. Хоч “життєві цикли” трохи подовжаться з розвитком технологій Інтернет, сучасні додатки для мережі вимагають лише декількох тижнів/місяців на розроблення. Web-менеджери намагаються коректувати вміст вузла щодня й оновлювати його графіку принаймні кожні 9–12 місяців. Отже, і технологія для побудови цих систем повинна бути дуже легкою для використання та розгортання.

### Мета роботи

**Мета роботи** – розроблення методів формування системою електронної контент-комерції списку пропозицій або кошика контенту для визначеного постійного клієнта відповідно до визначеної архітектури систем електронної комерції.

У моделі системи електронної контент-комерції користувачі знаходять і запускають застосування, які використовують традиційні HTML-сторінки і Web-сервери. Але замість простого завантаження статичної сторінки динамічний “апет” завантажується в індивідуальний браузер. До того ж він підтримує високошвидкісні протоколи, які дають йому змогу сполучатися безпосередньо із сервлетами (“servlets”) – ПЗ, що працює на проміжному рівні. Звичайно, сервлети забезпечують доступ до однієї або більше баз даних, реалізують бізнес-логіку і повертають результати аплетові для відображення. Аплети – це програми, що динамічно завантажуються та управляють логікою подання даних. У архітектурі WebOLTP вони містять частину коду, що відповідає за комунікацію та дає їм змогу напряму сполучатися з сервлетами, які працюють на проміжному рівні. Аплети мають такі переваги: незначний об'єм, що забезпечує їхнє швидке завантаження; велику інтерактивність і дружній інтерфейс порівняно із звичайними HTML-сторінками; відносну легкість у розробленні та супроводі.

Сьогодні найпоширенішим протоколом у мережі є Hypertext Transfer Protocol – HTTP. Будучи частиною інфраструктури WWW, HTTP чудово підходить для роботи зі статичними HTML-сторінками. Однак він не зовсім зручний для інтерактивного опрацюванні бізнес-транзакцій, оскільки є сторінково орієнтованим і не підтримує інформацію про стани і з'єднання. Все це –

істотні недоліки, коли йдеться про застосування WebOLTP. Автором пропонуються альтернативні протоколи для зв'язку броузера і ПЗ проміжного рівня. Ключові вимоги для цих протоколів передбачають здатність:

- підтримувати інформацію про користувачів і транзакції;
- ефективно формувати підсумкову вибірку й управляти нею;
- підтримувати транзакції;
- надавати надійні механізми шифрування даних.

### **Проектування архітектури системи електронної комерції**

*З появою багаторівневої архітектури на проміжний рівень (або рівні) перенесений основний тягар прикладного опрацювання. Це робить проміжний рівень одним з найкритичніших компонентів WebOLTP архітектури.*

Системи управління базами даних залишаються важливими компонентами всіх OLTP-систем, зокрема WebOLTP. Ефективне використання СУБД в архітектурі WebOLTP потребує: підтримки нестабільних навантажень з відстеженням таких властивостей, як запит черг і пріоритетів; високої швидкості з'єднання додатків із СУБД; черги додатків і управління ресурсами як засобів скорочення загального об'єму ресурсів у системі і досягнення стабільної продуктивності в межах Інтернет-транзакцій; забезпечення безпеки, як наприклад, уповноважена авторизація (відповідність) для певних WebOLTP-додатків; розподілене опрацювання запитів, яке дало б змогу опрацьовувати різноманітні типи даних, що використовуються в середовищі WebOLTP.

*Для WebOLTP-архітектури поки що невирішеним залишається питання, яка технологія найкраща для реалізації та управління бізнес-логікою на проміжному рівні. Наведемо основні вимоги до ПЗ проміжного рівня:*

- масштабність і продуктивність під час роботи з великою кількістю користувачів, сесій, транзакцій і з'єднань з БД;
- високопродуктивне з'єднання броузера і back-end сховища даних;
- підтримка швидкого розроблення і розгортання WebOLTP-додатків на проміжному рівні;
- підтримка як синхронного, так і асинхронного управління транзакціями.

Відомі три найпоширеніші варіанти технологій побудови ПЗ проміжного рівня:

1. CORBA на основі брокерів об'єктних запитів (Object Request Brokers – ORBs).
2. Монітори обробки транзакцій (Transaction Processing Monitors – TP-Monitors).
3. Сервери Web-додатків.

*Кожна з цих технологій має певні переваги, але жодна з них ідеально не задовольняє вимоги WebOLTP на проміжному рівні. Об'єкти CORBA мають чудові можливості побудови багаторівневої архітектури з викликом розподілених об'єктів та іншими сервісами. Складність загального вирішення і нестача надійних засобів підтримки обмежує їхнє застосування тільки кваліфікованими розробниками. До того ж більшість ORB мають примітивні механізми виконання на боці сервера, що також обмежує ефективність і масштабність. TP-монітори, з іншого боку, мають стійкі та відпрацьовані механізми виконання, які забезпечують достатню ефективність і масштабність. Однак подібно до об'єктів ORB їхня загальна складність і власний інтерфейс API часто робить TP-монітори важкими у використанні і дорогими стосовно встановлення, управління і підтримки. Сервери Web-додатків використовують новітні рішення в галузі програмного забезпечення проміжного рівня. Технологія сервера Web-додатків з'явилася внаслідок спроби трансформувати Netscape і Web-сервери Microsoft у сервери додатків. Спонукало до цього останнє покоління відповідних API – NSAPI і ISAPI. Сервери Web-додатків є спеціалізованими розробленнями на основі одного з інструментальних засобів створення Web-вузла. Це – потужний набір інструментальних засобів, що сприяє підвищенню продуктивності розробника. З іншого боку, масштабність істотно обмежена прямим звертанням сервера додатків до Web-серверів і відсутністю підтримки відмінних від HTTP протоколів зв'язку. Для того щоб подолати недоліки наявних систем і задовольнити вимоги WebOLTP на проміжному рівні, що забезпечують масштабність і простоту використання, запропонований новий клас системного ПЗ – сервери транзакцій. Сервери транзакцій*

об'єднують найкращі властивості ORB і TP-моніторів з компонент-базованою розробкою: це уможливує швидке створення масштабних WebOLTP-додатків. Першими доступними серверами транзакцій стали Sybase Jaguar CTS (Компонентний сервер транзакцій) від Sybase, Inc. і Microsoft Transaction Server (раніше відомий як Viper). Оскільки попит на WebOLTP зростає, незабаром повинні з'явитися інші продукти цього класу. Сервери транзакцій характеризуються такими властивостями: пропонують вбудовані можливості управління транзакціями; забезпечують механізм запуску й управління сервлетами; підтримують виклики розподілених об'єктів для забезпечення зв'язку в багаторівневих додатках; підтримують засоби швидкого розроблення ПЗ для проміжного рівня, зокрема компонентне розроблення.

Порівняно з традиційними системами клієнт/сервер або універсальною СУБД WebOLTP-додатки – особливо Internet- або Extranet-додатки – обслуговуватимуть значно більші групи користувачів. “Управління сеансами” означає управління зв'язками між броузерами і сервером транзакцій, а “управління з'єднаннями” – управління зв'язками між сервером транзакцій і СУБД. Працюючи разом, диспетчери сеансів і з'єднань перетворюють велику кількість сеансів броузера на набагато меншу кількість з'єднань із СУБД, поліпшуючи у такий спосіб загальну масштабність системи з забезпеченням стабільнішого часу відгуку за змінних робочих навантажень.

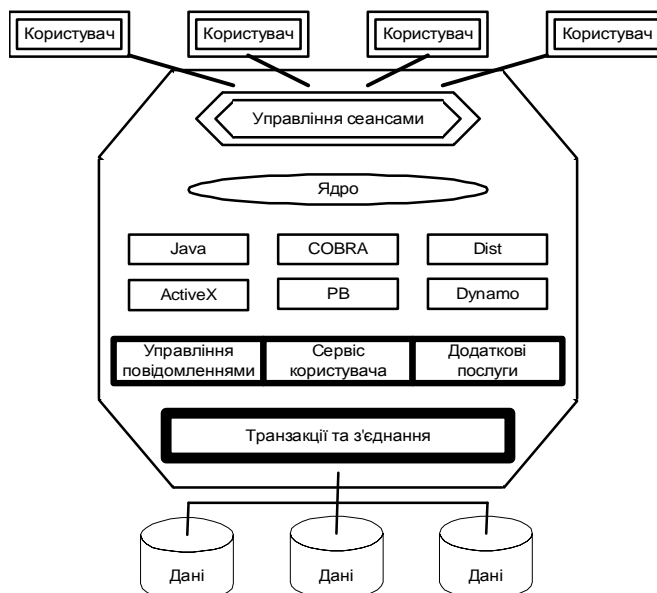


Рис. 4. Сервер транзакцій систем електронної комерції

У багаторівневому середовищі швидкодія з'єднання визначається часом відповіді кінцевому користувачеві. Основні показники з'єднуваності передбачають: узагальнену багатопротокольну підтримку броузерів та інших клієнтів мережі. Підтримувані протоколи включають HTTP, TDS (для швидкодіючого опрацювання потоків даних) і протокол CORBA IIOP; з'єднуваність з основою СУБД, зокрема Sybase SQL Server і SQL Anywhere, Oracle і MS SQL Server через стандарти ODBC, JDBC і Sybase Open Client; з'єднуваність з mainframe й іншими джерелами даних через продукти доступу до даних; швидкодіюча з'єднуваність з Java як для аплетів, так і для сервлетів; високошвидкісний обмін підсумковими вибірками між усіма рівнями; ефективне HTTP-тунелювання для забезпечення сумісності з firewall.

Вибір бази даних (СУБД) залежить від тих задач, які планується розв'язувати через Web-сервер, специфічних рис конкретної БД – таких, як технологічні основи, тип СУБД, вид інтерфейсів, зв'язки між таблицями, обмеження цілісності та організаційні рішення, пов'язані з підтримкою актуальності бази даних і забезпеченням доступу до неї. При забезпеченні WWW-доступу до БД можливі кілька шляхів – комплексів технологічних та організаційних рішень, перевірених практикою використання WWW-технологій для доступу до БД, по-різному пов'язаних

між собою. Вибір конкретних рішень при забезпеченні доступу залежить від специфіки конкретної СУБД; наявності фахівців, здатних з мінімальними витратами освоїти певні технологічні рішення; існування інших БД, WWW-доступ до яких повинен здійснюватися з мінімальними додатковими витратами тощо. WWW-доступ до баз даних здійснюється за одним із трьох основних сценаріїв: одноразове або періодичне перетворення даних з БД на статичні документи; динамічне створення гіпертекстових документів на основі вмісту БД; створення інформаційного сховища на основі високопродуктивної СУБД з мовою запитів SQL, періодичне завантаження даних у сховище з основної СУБД.

Важливою характеристикою інформації Web-сайту (контента) є її адекватність, тобто певний рівень відповідності створюваного за допомогою отриманого контенту образу реальному об'єкту, процесу, явищу тощо. Адекватність контенту може виражатися в трьох формах: прагматичній, семантичній та синтаксичній. *Прагматична (споживацька) адекватність* відображає відповідність контенту меті управління, яка на її основі реалізується. Ця форма адекватності безпосередньо пов'язана з практичним використанням контенту. *Семантична (смісловова) адекватність* визначає ступінь відповідності образу об'єкта і самого об'єкта. Семантичний аспект припускає облік смислового змісту контенту. На цьому рівні аналізуються ті відомості, які відображає контент, розглядаються смислові зв'язки. Ця форма призначена для формування понять і уявлень, виявлення значення, змісту контенту та його узагальнення. *Синтаксична адекватність* відображає формально-структурні характеристики контенту і не торкається її смислового змісту. На синтаксичному рівні враховуються тип носія і спосіб передавання контенту, швидкість передавання та опрацювання, розміри кодів подання контенту, надійність і точність перетворення цих кодів тощо. Ця форма сприяє сприйняттю структурних зовнішніх характеристик, тобто синтаксичного аспекту контенту. Якщо розглядається дослід, що може закінчитись одним із  $N$  можливих результатів, то необхідно оцінити такий результат. Такою оцінкою може стати міра контенту (або події). Міра загального об'єднання подій дорівнює сумі мір кожної події.

Введення поняття ентропії уможливорює кількісне вимірювання контенту. Справді, в результаті здійснення дослідів  $A$  ми отримаємо нові відомості, тобто деяку інформацію. Одночасно знання результату дослідів знімає повністю або частково ту невизначеність, яка була до його здійснення. Тому правильно припустити, що знята в результаті дослідів  $A$  ентропія дорівнює кількості одержаної інформації.

### Аналіз отриманих результатів

Поняття умовної ентропії в теорії інформації використовується при визначенні взаємозалежності<sup>1</sup> між символами умовного алфавіту, що кодується, для визначення втрат під час передавання інформації по каналах зв'язку, розрахунку ентропії поєднання, а також для оцінки формування кошика контенту або списку пропозицій для постійного відвідувача Web-сайту. У всіх випадках при розрахунку умовної ентропії в тому чи іншому вигляді використовуються умовні ймовірності. Якщо при формуванні  $n$  повідомлень (наприклад, новин) контент  $A$  з'явився  $m$  разів, контент  $B$  з'явився  $l$  разів, а контент  $A$  разом із контентом  $B$  –  $k$  разів, то ймовірність появи контенту  $A$   $p(A) = \frac{m}{n}$ ; ймовірність появи контенту  $B$  –  $p(B) = \frac{l}{n}$ ; ймовірність сумісної появи контенту  $A$  та  $B$   $p(AB) = \frac{k}{n}$ ; умовна ймовірність появи контенту  $A$  відносно контенту  $B$  та умовна ймовірність появи контенту  $B$  відносно контенту  $A$ :

$p(A|B) = \frac{p(AB)}{p(B)} = \frac{k}{l}$ ; умовна ймовірність появи контенту  $B$  відносно контенту  $A$   $p(B|A) = \frac{p(AB)}{p(A)} = \frac{k}{m}$ .

<sup>1</sup> Зміст взаємозалежності символів контенту списку новин полягає в тому, що ймовірність появи  $i$ -го контенту в будь-якому місці повідомлення залежить від того, які контенти стоять перед і після, і буде відрізнятися від безумовної ймовірності  $p_i$ , відомої із статичних властивостей цього списку новин.

$$p(A/B) = \frac{p(AB)}{p(B)} = \frac{k}{l}; p(B/A) = \frac{p(AB)}{p(A)} = \frac{k/n}{m/n} = \frac{k}{m}. \quad (1)$$

Якщо відома умовна ймовірність, то можна легко визначити і ймовірність сумісної появи контенту  $A$  та  $B$ , використовуючи вираз (1):  $p(AB) = p(B)p(A/B) = p(A)p(B/A)$

Від класичного виразу  $H = \sum_{i=1}^m p_i \cdot \log_2 \frac{1}{p_i} = -\sum_{i=1}^m p_i \cdot \log_2 p_i$  формула умовної ентропії відрізняється тим, що в ній ймовірності – умовні:

$$H(b_j/a_i) = -\sum_j p(b_j/a_i) \log p(b_j/a_i), H(a_i/b_j) = -\sum_i p(a_i/b_j) \log p(a_i/b_j), \quad (2)$$

де індекс  $i$  вибраний для характеристики довільного стану джерела повідомлень  $A$ , індекс  $j$  вибраний для характеристики довільного стану адресату  $B$ . Відрізняють поняття часткової та загальної ентропії. Вирази (2) являють собою часткові умовні ентропії. Загальна умовна ентропія повідомлення  $B$  відносно повідомлення  $A$  характеризує кількість інформації, яка міститься в будь-якому загальному списку контенту, і визначається усередненням по всіх контентах, тобто по усіх станах з урахуванням ймовірності появи кожного із станів, і дорівнює сумі ймовірностей появи контенту на невизначеність, яка залишається після того, як адресат прийняв повідомлення:

$$H(B/A) = -\sum_i p(a_i) H(b_j/a_i) = -\sum_i \sum_j p(a_i) p(b_j/a_i) \log p(b_j/a_i). \quad (3)$$

Вираз (3) є загальним виразом для визначення кількості інформації на один контент повідомлення для нерівномірних та взаємозалежних контентів. Оскільки  $p(a_i)p(b_j/a_i)$  являє собою ймовірність сумісної появи двох подій  $p(a_i, b_j)$ , то (3) можна записати так:

$$H(B/A) = -\sum_i \sum_j p(a_i, b_j) \log p(b_j/a_i).$$

Поняття загальної та часткової умовної ентропії широко застосовується для розрахунку інформаційних втрат при формуванні кошика контенту або списку пропозицій для постійного відвідувача Web-сайту. Загалом, якщо система електронної контент-комерції формує  $m$  контентів  $A$  та клієнти отримують  $m$  контент  $B$ , формування кошика контенту для постійного відвідувача Web-сайту в певні визначені інтервали часу повністю описується каналною матрицею (табл. 1).

Таблиця 1

**Канальна матриця формування кошика контенту для постійного відвідувача Web-сайту у певні визначені інтервали часу**

$A \backslash B$	$b_1$	$b_2$	...	$b_j$	...	$b_m$
$a_1$	$p(b_1/a_1)$	$p(b_2/a_1)$	...	$p(b_j/a_1)$	...	$p(b_m/a_1)$
$a_2$	$p(b_1/a_2)$	$p(b_2/a_2)$	...	$p(b_j/a_2)$	...	$p(b_m/a_2)$
...	...	...	...	...	...	...
$a_i$	$p(b_1/a_i)$	$p(b_2/a_i)$	...	$p(b_j/a_i)$	...	$p(b_m/a_i)$
...	...	...	...	...	...	...
$a_m$	$p(b_1/a_m)$	$p(b_2/a_m)$	...	$p(b_j/a_m)$	...	$p(b_m/a_m)$

Ймовірності, розташовані по діагоналі, визначають адекватне формування списку пропозицій або кошика товару для постійного відвідувача Web-сайту в певні визначені інтервали часу, інші –

хибне. Значення цифр, які заповнюють колонки каналної матриці, звичайно зменшуються у міру відхилення від головної діагоналі та за повної відсутності перешкод всі, крім цифр, розташованих на головній діагоналі, дорівнюють нулю. Якщо описувати формування повідомлень системою електронної контент-комерції, то проходження цього виду контенту у цьому кошику (списку пропозицій) описується розподілом умовних ймовірностей вигляду  $p(b_j/a_i)$ , так, для контенту  $a_1$  розподілом вигляду:

$$p(b_1/a_1) + p(b_2/a_1) + \dots + p(b_j/a_1) + \dots + p(b_m/a_1). \quad (4)$$

Сума ймовірностей розподілу (4) завжди повинна дорівнювати 1. Втрати інформації (не внесення у список пропозицій або кошик для визначеного клієнта), що припадають на частку контенту  $a_1$  описуються за допомогою часткової умовної ентропії вигляду:

$$H(b_j/a_1) = -\sum_{j=1} p(b_j/a_1) \log p(b_j/a_1). \quad (5)$$

Підсумовування відбувається за  $j$ , оскільки  $i$ -й стан (у цьому випадку перший) залишається постійним. Щоби врахувати втрати всіх контентів при формуванні повідомлень системою електронної контент-комерції, необхідно підсумувати всі часткові умовні ентропії (подвійне підсумовування за  $i$  та  $j$ ). При цьому у разі рівноймовірнісних появ контенту у списку пропозицій повідомлень або кошику для визначеного клієнта (на  $m$  ділимо, оскільки ентропія – невизначеність на один контент):  $H(B/A) = -\frac{1}{m} \sum_i \sum_j p(b_j/a_i) \log p(b_j/a_i)$

У разі нерівноймовірної появи контенту джерела повідомлень враховують ймовірність появи кожного контенту, помноживши на її відповідну часткову умовну ентропію. Загальна умовна ентропія дорівнює  $H(B/A) = -\sum_i p(a_i) \sum_j p(b_j/a_i) \log p(b_j/a_i)$ .

Якщо ми досліджуємо список пропозицій або кошик з боку користувача повідомлень (тестова або рейтингова система оцінки), то з отриманням контенту  $b_j$  припускаємо, що був відісланий якийсь із контентів  $a_1, a_2, \dots, a_i, \dots, a_m$  (табл. 2).

У цьому разі одиниці повинні дорівнювати сумі умовних ймовірностей по стовпцях каналної матриці:  $p(a_1/b_j) + p(a_2/b_j) + \dots + p(a_i/b_j) + \dots + p(a_m/b_j) = 1$ .

Часткова умовна ентропія

$$H(a_i/b_j) = -\sum_{i=1}^m p(a_i/b_j) \log p(a_i/b_j), \quad (6)$$

а загальна умовна ентропія  $H(A/B) = -\sum_j p(b_j) \sum_i p(a_i/b_j) \log p(a_i/b_j)$ .

Таблиця 2

Канальна матриця формування кошика контенту з боку користувача

$A \backslash B$	$b_1$	$b_2$	...	$b_j$	...	$b_m$
$a_1$	$p(a_1/b_1)$	$p(a_1/b_2)$	...	$p(a_1/b_j)$	...	$p(a_1/b_m)$
$a_2$	$p(a_2/b_1)$	$p(a_2/b_2)$	...	$p(a_2/b_j)$	...	$p(a_2/b_m)$
...	...	...	...	...	...	...
$a_i$	$p(a_i/b_1)$	$p(a_i/b_2)$	...	$p(a_i/b_j)$	...	$p(a_i/b_m)$
...	...	...	...	...	...	...
$a_m$	$p(a_m/b_1)$	$p(a_m/b_2)$	...	$p(a_m/b_j)$	...	$p(a_m/b_m)$



Оскільки  $p(a_i)p(b_j/a_i) = p(a_i, b_j)$ , для розрахунку загальної умовної ентропії використовується такий вираз:  $H(B/A) = -\sum_i \sum_j p(a_i, b_j) \log p(b_j/a_i)$ . Якщо задана канална матриця вигляду  $p(b_j/a_i)$  (часткова умовна ентропія відповідає (5)) та безумовні ймовірності вигляду  $p(a_i)$ , то безумовні ймовірності приймача  $p(b_j)$  знаходимо як  $\sum_i p(a_i)p(b_j/a_i)$ , тобто якщо задані безумовні ймовірності джерела та каналної матриці, то ентропія приймача  $H(B) = -\sum_j p(b_j) \log p(b_j)$ , і навпаки, якщо задані ймовірності вигляду  $p(b_j)$  та канална матриця, яка описує канал зв'язку з боку приймача повідомлень (часткова умовна ентропія відповідає виразу (6)), то  $p(a_i) = \sum_j p(b_j)p(a_i/b_j)$ , а, отже, визначається ентропія джерела повідомлень  $H(A) = -\sum_i p(a_i) \log p(a_i)$ . При взаємозалежності трьох елементів  $a_i, a_j, a_k$  умовна ентропія розраховується як  $H(A/B, K) = -\sum_i \sum_j \sum_k p(a_i, a_j, a_k) \log p(a_i, a_j, a_k)$ ,

аналогічно і для 4, 5, ..., n елементів. Ентропія поєднання використовується для розрахунку ентропії сумісної появи статично залежних повідомлень. Наприклад, якщо передати сто разів статтю визначеному списку клієнтів через систему електронної газети, та сама стаття визначеної тематики буде сприйнята адекватно 90 користувачами, неадекватно – 8 разів та невпевнено – 2 рази. Невизначеність виникнення комбінацій вигляду “адекватно” – “неадекватно”, “адекватно” – “неадекватно”, “адекватно” – “невпевнено” при передаванні статті описується за допомогою ентропії поєднання.  $H(A, B)$  – невизначеність того, що буде відіслано A, а прийнято B. Для асамблів переданих повідомлень A та прийнятих повідомлень B ентропія поєднання являє собою суму  $H(A, B) = -\sum_i \sum_j p(a_i, b_j) \log_2 p(a_i, b_j)$  біт/два контенти.

Ентропія поєднання та умовна ентропія пов'язані між собою співвідношеннями:  $H(A, B) = H(A) + H(B/A) = H(B) + H(A/B)$ ,  $H(B/A) = H(A, B) - H(A)$ ,  $H(A/B) = H(A, B) - H(B)$ . Ентропія поєднання розраховується за допомогою матриці:

$$p(a_i, b_j) = \begin{vmatrix} p(a_1, b_1) & p(a_1, b_2) & \mathbf{K} & p(a_1, b_m) \\ p(a_2, b_1) & p(a_2, b_2) & \mathbf{K} & p(a_2, b_m) \\ \dots & \dots & \dots & \dots \\ p(a_m, b_1) & p(a_m, b_2) & \mathbf{K} & p(a_m, b_m) \end{vmatrix}.$$

Матриця володіє такою властивістю:  $\sum_i p(a_i, b_j) = p(b_j)$ ,  $\sum_j p(a_i, b_j) = p(a_i)$ ,  $\sum_i p(a_i) = \sum_j p(b_j) = 1$ . Ця властивість дає змогу розраховувати ентропію як джерела, так і приймача повідомлень безпосередньо по каналній матриці:  $H(A) = -\sum_i \sum_j p(a_i, b_j) \log \sum_j p(a_i, b_j)$ ,  $H(B) = -\sum_i \sum_j p(b_j, a_i) \log \sum_i p(b_j, a_i)$ . Підсумовування здійснюємо за i та j, оскільки для того, щоби знайти безумовні ймовірності, необхідно підсумувати їх за однією координатою (матричне подання ймовірностей), а для знаходження H підсумовування здійснюється по іншій

координаті. Умовні ймовірності вигляду  $p(a_i/b_j)$  та  $p(b_j/a_i)$  розраховуються як:

$$p(a_i/b_j) = \frac{p(a_i, b_j)}{\sum_i p(a_i, b_j)}, \quad p(b_j/a_i) = \frac{p(a_i, b_j)}{\sum_j p(a_i, b_j)}.$$

переданого системою електронної контент-комерції, в якому адекватність описується за допомогою ентропії поєднання, підраховується так:  $I(A, B) = H(A) + H(B) - H(B, A)$ .

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

У цій роботі розглянуто розвиток систем електронної комерції, їхня характерна архітектура, недоліки, переваги, процеси функціонування. Ключовою частиною роботи є розроблення методів формування системою електронної контент-комерції списку пропозицій або кошика контенту для визначеного постійного клієнта. Наведений математичний опис функціонування системи електронної комерції та методів формування системою електронної контент-комерції списку пропозицій або кошика контенту, а також схеми побудови аналогічних систем електронної комерції.

1. Берко А.Ю., Висоцька В.А. *Intranet архітектура інтелектуальних систем електронного навчання* // Вісн. Нац. ун-ту "Львівська політехніка". – 2001. – № 438: Інформаційні системи та мережі. – С. 3–10. 2. Берко А.Ю., Висоцька В.А., Чирун Л.В. *Алгоритми опрацювання інформаційних ресурсів в системах електронної комерції* // Вісн. Нац. ун-ту "Львівська політехніка". – 2004. – № 519: Інформаційні системи та мережі. – С. 10–20. 3. Берко А.Ю., Висоцька В.А. *Проектування навігаційного графу web-сторінок бази даних систем електронної комерції* // Вісн. Нац. ун-ту "Львівська політехніка". – 2004. – № 521: Комп'ютерні науки та інформаційні технології. – С. 48–57. 4. Береза А.М. *Електронна комерція*. – К., 2002.