

Каценеленбаум Б.З. Расчет диаграммы направленности плоской резонаторной антенны с переменными прозрачностью и высотой. // Радиотехника и электроника. 1993, т. 38, № 7, – с. 1247–1255.
6. Войтович Н. Н., Германюк Р. И., Заморская О. Ф. Оптимизация параметров плоской резонаторной антенны с переменными высотой и прозрачностью. // Радиотехника и электроника. 1994, т. 39, № 8–9, – с. 1321–1328.
7. M. I. Andriychuk, O. F. Zamorska. Some generalizations for the synthesis problem of the waveguide antenna array. // Proc. of 10th Intern. Conf. on Mathematical Methods in Electromagnetic Theory. Dnipropetrovsk, Ukraine, Sept. 14-17, 2004, – p. 487–489.

УДК 681.3

А.М. Пелецишин

Національний університет “Львівська політехніка”,
кафедра ІСМ

МОДЕЛЮВАННЯ АУДИТОРІЇ ВЕБ-САЙТА

© Пелецишин А.М., 2004

Розглянуто проблеми формального опису та моделювання аудиторії сайту. При побудові моделі використано апарат нечітких множин.

This paper considers of formal descriptions and modelling of Web site auditory. Theory of fuzzy logic are used in proposed model.

Постановка проблеми та її актуальність. Проблема моделювання аудиторії World Wide Web є задачею, що зумовлена низкою факторів, серед яких найважливішими є:

- високий користувацький попит на сервіси пошуку, класифікації та аналізу інформаційних ресурсів WWW;
- потреба власників сайтів у точному відображенні тематики сайту в сервіси пошуку, класифікації та аналізу інформаційних ресурсів WWW.

Задачу визначення моделювання аудиторії сайту неодноразово розглядали як з теоретичного погляду, так і зі спробами реального впровадження. Проте дослідження у цій сфері мають односторонній характер – це автоматизоване (частково чи повністю) визначення підмножини користувачів Інтернету, які відвідали сайт протягом контрольного проміжку часу.

Аналіз досліджень. Існуючі підходи до структурування аудиторії сайту. Традиційно структурування аудиторії сайту здійснюється виділенням спеціальних груп користувачів, які і становлять аудиторію сайту. Таке групування базується на певних групових ознаках.

Найпоширенішими груповими ознаками є ознаки одного з таких класів:

- технічні ознаки – ознаки, що характеризують обладнання користувача, програмне забезпечення, методи мережевого підключення тощо ;
- демографічні ознаки;
- поведінкові ознаки – ознаки характеризують різні аспекти поведінки користувача в глобальному інформаційному середовищі.

Виділення невирішених раніше частин загальної проблеми. Існуючі моделі (з чітким фіксуванням приналежності користувача WWW до аудиторії сайту) не достатньою мірою задовольняють такі вимоги:

- модель аудиторії сайту повинна відображати різний ступінь приналежності користувача до аудиторії сайту;

– модель аудиторії сайта повинна враховувати методи отримання прибутку власниками сайта від факту приналежності користувача до аудиторії;

– модель аудиторії сайта повинна забезпечувати можливість порівняння реальної ємності аудиторій різних сайтів.

Формулювання цілей статті. Головними цілями цієї статті є

– дослідження та формалізація де-факто існуючих методів опису аудиторії сайта;

– побудова формальної моделі аудиторії сайта та моделі користувача WWW.

Модель користувача WWW. Користувачем WWW вважатимемо людей та програмних агентів, які розв'язують певні задачі, здійснивши доступ до сайтів через мережу Інтернет.

Формально користувач описується таким відношенням:

(основні характеристики, цілі користувача, історія користувача)

де

основні характеристики – базова інформація, що ідентифікує користувача;

цілі користувача – набір цілей, досягнення яких прагне користувач під час роботи в WWW;

історія користувача – історія змістовних дій користувача в WWW.

Цілі користувача в системі WWW поділятимемо на такі класи:

інформаційні цілі – отримання необхідної інформації зі системи WWW;

операційні цілі – здійснення в системі WWW певних операцій (купівля чи продаж товару, відсилення листа електронною поштою тощо);

комунікативні цілі – спілкування в системі WWW з її іншими користувачами.

Важливим підкласом інформаційних цілей користувача є **навігаційні (пошукові) цілі** – знаходження необхідного ресурсу в системі WWW.

Історію користувача (UH) в базовому варіанті можна подавати як історію навігації користувача по WWW (UNH)- історія доступу користувачем до інформаційних ресурсів WWW. У такому разі

$$UNH = \{Q_i\}_{i=1}^{N_{UNH}} \quad (1)$$

$$Q_i = (t_Q, Host, URI, Ref, RM, QS, SA, Page) \quad (2)$$

де t_Q – момент запиту на отримання ресурсу; $Host$ – унікальна адреса клієнта; URI – унікальний ідентифікатор ресурсу; Ref – вказівник на попередній запис у історії навігації; RM – метод запиту; QS – апараметри запиту; SA – відповідь сервера; $Page$ – отримана інформація.

Ця модель історії користувача фактично є відображенням журналу доступу до ресурсів WWW. Основним недоліком цієї моделі є недостатній ступінь відображення семантики дій користувача в системі WWW та занадто високий ступінь деталізації моделі. У цій моделі поведінка користувача описується послідовністю залежних записів, отже, під час проведення різних видів аналізу необхідно враховувати не лише записи в історії, а й взаємозв'язки між ними.

Іншим способом опису історії користувача WWW є використання історії транзакцій користувача у системі WWW.

$$UTH = \{Tr_i\}_{i=1}^{N_{UTH}} \quad (3)$$

$$Tr = (t_Q, \bar{t}_Q, Id_{Tr}, UNH(Id_{Tr})) \quad (4)$$

де t_Q – момент початку транзакції; \bar{t}_Q – момент завершення транзакції; Id_{Tr} – унікальний ідентифікатор транзакції; $UNH(Id_{Tr})$ – записи в історії навігації користувача, що входять до даної транзакції.

Транзакція – послідовність взаємозв'язаних запитів користувача до сервісів WWW та результатів їхнього опрацювання. Одна транзакція може складатися з багатьох запитів користувача у WWW. Проте один запит може одночасно входити до кількох транзакцій.

Основними відмінностями транзакційної моделі історії користувача від навігаційної є

– відображення завершеної взаємодії користувача та глобального середовища;

– незалежність кожної транзакції;

- наявність у транзакції двох часових міток (початку та кінця транзакції);
- наявність у транзакції окремих запитів користувачів до різних сайтів, які логічно пов'язані.

Транзакційна модель має принципові переваги над навігаційною моделлю користувача. Це зокрема

- можливість поглибленого аналізу поведінки користувача та його реальних потреб;
- можливість точнішої оцінки досягнутих результатів взаємодії користувача та WWW.

Модель аудиторії сайта

Вимоги до моделі аудиторії сайта. Існуючі підходи до моделювання аудиторії сайта базуються на визначенні певної характеристики груп користувачів, які відвідують чи можуть відвідати сайт. Така модель передбачає опис аудиторії сайта як підмножини користувачів WWW.

$$Aud = \{U_i\}_{i=1}^{N_u} \quad (5)$$

де U_i – користувач WWW, який можна зарахувати до аудиторії сайта.

Конструктивнішою, хоча і не такою точною, є модель, яка передбачає опис аудиторії сайта як сукупності підмножин користувачів WWW, які згруповані на основі певної ознаки.

$$Aud = Aud_1 \cup \dots \cup Aud_{N_{Aud}} \quad (6)$$

$$Aud_i = \{U_i\}_{i=1}^{N_u^{(i)}}$$

де Aud_i – підмножина користувачів WWW, яку можна зарахувати до аудиторії сайта за однією спільною ознакою.

Модель сайта на основі нечітких множин. Використання множин для опису аудиторії сайта та її структури недостатньою мірою задовольняє наведені вище вимоги до моделі сайта. Можливим альтернативним підходом до моделювання сайта є опис сайта як нечіткої множини

$$Aud(Site) = \{(U_i, B(U_i, Site))\}_{i=1}^{N_u} \quad (7)$$

де $Aud(Site)$ – аудиторія сайта; U_i – користувач WWW; $B(U_i, Site)$ – міра приналежності користувача U_i до аудиторії сайта $Site$.

Отже, кожному користувачу WWW ставиться у відповідність величина, яка характеризує міру приналежності користувача до аудиторії сайта. Природними обмеженнями на величину приналежності є

$$0 \leq B(U_i, Site) \leq 1 \quad (8)$$

Для користувачів, які взагалі не мають жодного стосунку до сайта, приналежність дорівнює нулю.

Визначення реального змісту функції приналежності – це задача, яка має суб'єктивний для власників сайта характер. Проте доцільним є побудова цієї функції відповідно до цілей, що ставляться власниками щодо свого сайта.

У такому разі реальний зміст функції B визначається природою міри цінності користувача для сайта. У структурі функції відображається основний механізм отримання прибутку від відвідування сайта користувачами WWW.

Так, для тих мір цінності, які пропорційно залежать від числа відвідувачів сайта (наприклад, прибуток від пропаганди ідей, розміщених на сайті), функція приналежності може визначатися як імовірність відвідування користувачем WWW сайта протягом контрольного періоду

$$B(U_i, Site) = \Pr(T, U_i, Site)$$

У деяких випадках може бути важливим не тільки факт відвідування користувачем сайта, а й частота відвідування сайта. У такому разі можливе визначення функції приналежності, наприклад, так:

$$B(U_i, Site) = \sum_{C=1}^{N_C} C * \Pr(T, U_i, C, Site) \quad (9)$$

де C – число заходів користувача на сайт протягом контрольного періоду T ; N_C – максимальне цінне число заходів користувача протягом контрольного періоду T .

Отже, використання апарата нечітких множин дозволяє будувати модель аудиторії сайта, яка задовольняє наведені вище вимоги.

– Вірогідність попадання користувача WWW в аудиторію сайта описується за допомогою функції приналежності $B(U_i, Site)$.

– Суб'єктивні інтереси власника сайта враховуються в структурі функції приналежності $B(U_i, Site)$.

– Порівняння аудиторій сайтів стає можливим завдяки наявності числових оцінок, поданих як функція приналежності $B(U_i, Site)$. Детальніше проблему порівняння обсягів аудиторії сайтів розглянемо пізніше.

На практиці окреме визначення міри приналежності для кожного користувача WWW є неможливим через велику розмірність множини та відсутності аналітичного подання функції приналежності.

Проте достатньо ефективним методом опису аудиторії сайта як нечіткої множини є використання додаткових характеристик чи ознак користувача як аргументів функції приналежності.

У такому разі для кожного користувача U_i виділяються набір характеристик $(Ch_1^{(i)} \dots Ch_{n_{ch}}^{(i)})$, які слугують базовими для визначення міри приналежності до аудиторії сайта.

Тоді

$$B(U_i, Site) = B(Ch_1^{(i)} \dots Ch_{n_{ch}}^{(i)}, Site) \quad (10)$$

Такий підхід дозволяє побудувати достатньо компактний та ефективний опис аудиторії сайта як нечіткої множини на основі базових характеристик користувача WWW,

Базовими характеристиками можна використовувати наведені вище ознаки користувачів – технічні, демографічні, поведінкові.

Функцію приналежності можна задавати як в аналітичному, так і в табличному вигляді (при невеликій скінченній множині можливих значень базових ознак).

Простим прикладом опису аудиторії сайта як нечіткої множини може бути нечітка множина, де функція приналежності залежить від віку користувача:

$$B(U_i, Site) = \Pr(T, Age(U_i), Site) \quad (11)$$

де $Age(U_i)$ – вік користувача.

Складнішим прикладом опису аудиторії сайта можуть бути визначення функції приналежності до аудиторії сайта як функції від географічних координат користувача WWW:

$$B(U_i, Site) = \Pr(T, Geo(U_i), Site) \quad (12)$$

де $Geo(U_i)$ – координати користувача WWW.

Окремим важливим випадком опису аудиторії сайта є визначення міри приналежності користувача до аудиторії сайта як функції від шляху від історії транзакцій користувача WWW:

$$B(U_i, Site) = \Pr(T, UTH(U_i), Site) \quad (13)$$

На практиці замість наведеного часто використовується дуже спрощений варіант функції

$$B(U_i, Site) = \Pr(T, Ref(U_i), Site)$$

де Ref – WWW-ресурс, з якого користувач потрапив на сайт.

Підхід до моделювання аудиторії сайта, де міра приналежності визначається транзакційною історією користувача (або її навігаційним замінником) є одним з найважливіших при розв'язанні практичних задач оптимізації сайта та просування сайта в Інтернет. Тому він вимагає окремого детального розгляду.

Групування користувачів сайта. На практиці рідко є можливість побудувати функцію приналежності користувача WWW до аудиторії сайта в аналітичному вигляді. У такому разі опис функції здійснюється в табличному вигляді. Для зменшення розмірності такого опису користувачів WWW доцільно групувати.

Групою користувачів сайта (j -ю групою користувачів) вважатимемо множину користувачів WWW, об'єднаних однаковим значенням ознаки Ch :

$$Aud_j = \{U_i\}, \forall U_i : Ch(U_i) = Ch_j \quad (14)$$

З погляду структурування аудиторії сайта групування має сенс, якщо всі користувачі, що віднесені до однієї групи за спільною ознакою, мають однакову міру приналежності до аудиторії сайта. Тобто групою ознакою повинні вибиратися такі характеристики користувача, які визначають міру його приналежності до аудиторії сайта.

Вважатимемо, що для всіх користувачів певної групи міра приналежності до аудиторії сайта є константою.

$$B(U_i, Site) = B_j(Site) \quad \forall U_i \in Aud_j, j = 0 \dots N_{Aud} \quad (15)$$

Аудиторія сайта є об'єднанням усіх груп користувачів

$$Aud(Site) = \bigcup_{j=0}^{N_{Aud}} Aud_j(Site) \quad (16)$$

У разі, якщо для групування по окремих групах використовується єдина групова ознака Ch , то користувач може належати лише до однієї групи.

Відмітимо, що для кожного сайта існує спеціальна група Aud_0 користувачів WWW, для яких міра приналежності до аудиторії сайта дорівнює

$$B_0(Site) < \varepsilon \quad (17)$$

де ε – мала контрольна величина.

До цієї групи належать користувачі WWW, які практично ніколи не скористаються послугами сайта. Можна говорити, що імовірність потрапити на сайт для них дорівнює нулю у сенсі геометричної імовірності.

При практичному описі аудиторії сайта група Aud_0 з розгляду, як правило, усувається. Крім того, слід відмітити, що часто під аудиторією сайта розуміють множину користувачів WWW, які не потрапляють до групи Aud_0 .

Ця група для більшості сайтів (за винятком кількох найпопулярніших сайтів WWW) значно переважає за кількістю членів інші групи.

Кількісне порівняння та оцінка обсягів аудиторій сайтів. Для сайтів близької чи суміжної тематики часто існує задача порівняння обсягів аудиторії. Зокрема ця задача спостерігається в плануванні рекламних заходів в Інтернет, оцінюванні ефективності сайтів, визначення актуальності та якості матеріалів, що розміщені на сайтах.

Сьогодні традиційним і найпоширенішим методом порівняння обсягів аудиторії конкуруючих сайтів є порівняння простих технічних показників – кількості відвідувачів протягом певного періоду часу (“хостів”) та кількості поданих ними запитів до сайту (“хітів”). Такий підхід є основою різноманітних Інтернет-рейтингів сайтів. Відповідно на цьому підході базуються основні підходи до оцінки ефективності функціонування сайта, доцільності розміщення на ньому реклами тощо.

Проте цей підхід не дає достатньо якісної картини щодо реального співвідношення аудиторій різних сайтів та активно критикується власниками сайтів та дослідниками WWW.

Основними недоліками такого підходу є ігнорування структури сайтів та ігнорування різноманітних форм взаємодії сайтів з користувачами. Об'єднуювальним фактором у цих недоліках є те, що не враховуються реальні результати взаємодії користувачів зі сайтами.

Точнішою оцінкою обсягів аудиторії може слугувати величина, базована на мірі приналежності користувача до аудиторії:

$$\|Aud(Site)\| = \sum_{i=1}^{N_U} B(U_i, Site) \quad (18)$$

де $\|Aud(Site)\|$ – обсяг аудиторії сайта

У випадку, коли аудиторія сайта структурована за групами

$$\|Aud(Site)\| = \sum_{j=1}^{N_{Aud}} B_j(Site) \|Aud_j(Site)\| \quad (19)$$

Висновки. Моделювання аудиторії Веб-сайта є ключовим етапом для багатьох важливих задач проектування та управління сайта. Серед них – проектування структури та наповнення сайта, визначення та оптимізація тематики сайта, організація та проведення рекламних кампаній в Інтернеті.

Існуючі на сьогодні моделі аудиторії Веб-сайтів передбачають жорсткий опис приналежності користувача WWW до множини користувачів сайта. Аудиторія сайта визначається як підмножина користувачів World Wide Web. Проте, такі моделі недостатньою мірою :

- відображають різний ступінь приналежності користувача до аудиторії сайта;
- враховують методи отримання прибутку власниками сайта від факту приналежності користувача до аудиторії;
- забезпечують можливість порівняння реальної ємності аудиторій різних сайтів.

Наукова новизна статті. У статті пропонується нова модель аудиторії сайта як нечіткої множини, що описує міру приналежності користувачів WWW до аудиторії сайта. Пропонуються нові підходи до визначення такої міри. На основі отриманих результатів сформульовано нові підходи до оцінки розмірів аудиторії сайта,

Практична цінність. Наукові результати, отримані в цій статті, дозволяють провадити подальші практичні дослідження з моделювання та оптимізації Веб-сайтів та побудови формальних методів проектування та організації успішної діяльності сайтів у глобальному середовищі. Зокрема можливе точніше позиціонування тематики сайта щодо потреб користувачів та проведення робіт з покращання позиціонування сайта в пошукових системах.

1. B. Mobasher, H. Dai, T. Luo, M. Nakagawa, Y. Sun, and J. Wiltshire. *Discovery of aggregate usage profiles for Web personalization.* // *Proceedings of the WebKDD 2000 Workshop at the ACM SIGKDD 2000, Boston, August 2000.* 2. Flake G., Lawrence S., Giles C. *Efficient identification of web communities.* // *Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD-2000), Boston, MA, 2000. ACM Press.* 3. Flake G., Lawrence S., Giles C., Coetzee F. *Self-Organization of the Web and Identification of Communities.* *IEEE Computer*, 35(3), 66–71, 2002 <http://webselforganization.com/>. 4. J. Srivastava, R. Cooley, M. Deshpande, P-T. Tan. *Web usage mining: discovery and applications of usage patterns from Web data.* *SIGKDD Explorations*, (1) 2, 2000. 5. Пелецишин А.М. *Методи та алгоритми моделювання Web-систем.* // *Вісник ДУ "Львівська політехніка"*, – Львів, 2000. №406. – С. 199–211