

О. В. Стець, доцент кафедри
математичного моделювання економічних систем,
Т. І. Медвідь, студентка групи УК-81,
Факультет менеджменту та маркетингу,
Національний технічний університет України «КПІ»

ЕКОНОМІКО-МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ АУДИТОРІЇ САЙТУ

АНОТАЦІЯ. Структурування та моделювання аудиторії сайту є одним з найважливіших завдань, що розв'язуються у процесах аналізу, проектування і моделювання сайту та його функціонування у системі World Wide Web.

Стаття присвячена побудові формальної моделі аудиторії сайту за допомогою економіко-математичних моделей та методів.

Використання множин для опису аудиторії сайту та її структури в недостатній мірі задовольняє вимогам, що встановлюються для моделі сайту, зокрема врахування міри приналежності користувача до аудиторії сайту. Можливим альтернативним підходом до її моделювання є опис сайту як нечіткої множини.

КЛЮЧОВІ СЛОВА: модель аудиторії сайту, Web 2.0, нечіткі множини, міра приналежності користувача.

АННОТАЦИЯ. Структурирование и моделирование аудитории сайта является одним из самых важных заданий, которые решаются в процессе анализа, проектирования и моделирования сайта и его функционирования в системе World Wide Web.

Стаття посвящена созданию формальной модели сайта с помощью экономико-математических моделей и методов.

Использование множеств для описания аудитории сайта и ее структуры не в полной мере удовлетворяет требованиям, которые устанавливаются для модели сайта, в частности учета степени принадлежности пользователя к аудитории сайта. Возможным альтернативным подходом к ее моделированию является описание сайта как нечеткого множества.

КЛЮЧЕВЫЕ СЛОВА: модель аудитории сайта, Web 2.0, нечеткие множества, степень принадлежности пользователя.

ANNOTATION. Structuring and Modeling audience of the site is one of the most important tasks which are resolved by a process of analysis, design and modeling of the site and its functioning system in World Wide Web

This article is devoted to construct a formal model of site's audience.

Using sets to describe the audience of the site and its structure in an insufficient measure satisfies the requirements set for the model site, including consideration of measures belonging to the user audience site. A possible alternative approach to its modeling is the description of the site as a fuzzy set.

Вступ. Розвиток Інтернету за останні 3—6 років у значній мірі здійснюється шляхом активного впровадження ряду принципів і технологій, які отримали спільну назву «Web 2.0».

За своєю суттю Web 2.0 є логічним розвитком існуючих веб-технологій за рахунок зміни пріоритетів та акцентів у використанні технологій і задоволенні потреб користувачів.

На сьогодні Web 2.0 розглядається як головний напрям розвитку Інтернету на найближче десятиліття. Основними причинами до його переходу є:

- критична недостача якісного інформаційного наповнення;
- потреба активного самовираження користувача в WWW;
- розвиток технологій пошуку та агрегації інформації в WWW.

Особливістю моделювання аудиторії сайту є те, що саме поняття аудиторії сайту є досить складним і суперечливим. Найчастіше модель аудиторії сайту має конструктивний характер — вона визначається методом виділення серед множини користувачів WWW тих, які з певних причин можуть бути зарахованими до аудиторії сайту.

Такий підхід до моделювання аудиторії сайту є цілком виправданим, проте необхідно враховувати, що поняття «аудиторія сайту» та «користувач сайту» у такому разі є суб'єктивним, на відміну від об'єктивного поняття «користувач WWW». Суб'єктивність поняття «аудиторії сайту», очевидно, є неминучою через наявність у власників кожного із сайтів особистого бачення форм і результатів його взаємодії з користувачами WWW. Враховуючи специфіку технології Web 2.0 та особливості менталітету сучасної Інтернет-спільноти, метою власника сайту першочергово виступає привернення уваги все більшої аудиторії та формування свідомих прихильників на тривалий проміжок часу.

Постановка задачі. Проблема моделювання аудиторії WWW є задачею, яка зумовлена рядом факторів, серед яких найважливішими є:

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

Задача моделювання аудиторії сайту неодноразово розглядалася як з теоретичної точки зору, так і зі спробами реального впровадження. Проте дослідження у даній сфері носять односторонній характер — це автоматизоване (частково чи повністю) визначення підмножини користувачів Інтернету, які відвідали сайт протягом досліджуваного періоду.

Традиційно структурування аудиторії сайту здійснюється шляхом виділення спеціальних груп користувачів, які і складають аудиторію сайту. Таке групування базується на певних групових ознаках.

Найпоширенішими груповими ознаками є ознаки одного з таких класів:

1) технічні ознаки — ознаки, що характеризують обладнання користувача, програмне забезпечення, методи мережного підключення тощо;

2) демографічні ознаки;

3) поведінкові ознаки — ознаки, що характеризують різні аспекти поведінки користувача в глобальному інформаційному середовищі.

Існуючі моделі (з чітким фіксуванням приналежності користувача WWW до аудиторії сайту) не в достатній мірі задовольняють таким вимогам:

1. Модель аудиторії сайту повинна відображати різний ступінь приналежності користувача до аудиторії сайту.

2. Модель аудиторії сайту повинна враховувати методи отримання прибутку власниками сайту від факту приналежності користувача до аудиторії.

3. Модель аудиторії сайту повинна забезпечувати можливість порівняння реальної ємності аудиторій різних сайтів.

Головною ціллю роботи є побудова формальної моделі аудиторії сайту.

Методологія. Використання множин для опису аудиторії сайту та її структури в недостатній мірі задовольняє наведеним вище вимогам до моделі сайту. Можливим альтернативним підходом до моделювання сайту є його опис за допомогою економіко-математичних методів та апаратів. У даній статті сайт буде описано як нечітку множину:

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

де $Aud(Site)$ — аудиторія сайту;

U_i — користувач WWW;

$B(U_i, Site)$, — міра приналежності користувача U_i до аудиторії сайту $Site$.

Таким чином, кожному користувачу WWW ставиться у відповідність величина, яка характеризує міру приналежності кори-

стувача до аудиторії сайту. Природними обмеженнями на величину приналежності є:

$$0 \leq B(U_i, Site) \leq 1. \quad (2)$$

Для користувачів, які взагалі не мають жодного відношення до сайту, приналежність дорівнює нулю.

Визначення реального змісту функції приналежності — це задача, яка носить суб'єктивний для власників сайту характер. Проте доцільним є побудова даної функції відповідно до цілей, що ставляться власниками щодо свого сайту.

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

$$B(U_i, Site) = \Pr(T, U_i, Site). \quad (3)$$

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

$$B(U_i, Site) = \sum_{C=1}^{N_c} C \times \Pr(T, U_i, C, Site), \quad (4)$$

де C — число заходів користувача на сайт протягом контрольного періоду T ;

N_c — максимальне цінне число заходів користувача протягом контрольного періоду T .

Отже, використання апарату нечітких множин дозволяє будувати модель аудиторії сайту, яка задовольняє наведеним нижче вимогам:

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

2. Суб'єктивні інтереси власника сайту враховуються у структурі функції приналежності $B(U_i, Site)$.

3. Порівняння аудиторій сайтів стає можливим завдяки наявності числових оцінок, поданих як функція приналежності $B(U_i, Site)$.

Результати дослідження. Для ілюстрації моделювання аудиторії сайту використано статичні дані Інтернет-студії «Anate Studio».

Моделювання аудиторії Веб-сайту «Anate Studio» зумовлене потребою розробників студії визначення імовірності замовлення інформаційного продукту відвідувачем даної сторінки.

Нашою ціллю є моделювання аудиторії сайту «Anate Studio» з відображенням різного ступеня приналежності користувача до аудиторії та імовірності придбання продукту, що пропонується власниками сайту.

Моделювання аудиторії сайту базується на визначенні певної характеристики груп користувачів, які відвідують чи можуть відвідати сайт.

Реклама «Anate Studio» розміщена на сайтах з відповідною тематикою, для привернення уваги того кола користувачів, які можуть бути зацікавлені у продукті, що пропонується на даній сторінці.

Для дослідження обрано 10 основних Веб-сайтів, на яких розміщена тема «Anate Studio» з прямим посиланням на наш сайт. Такий список отриманий з Google Analytics шляхом сортування по числу переходів за період 1 місяць.

Опис аудиторії сайту як нечіткої множини має вигляд:

$$Aud(Site) = \{(U_i, B(U_i, Site))\}_{i=1}^{N_U}, \quad (5)$$

де $Aud(Site)$ — аудиторія сайту;

U_i — користувачі сайту, на якому розміщена реклама студії;

$B(U_i, Site)$ — міра приналежності користувачів U_i до аудиторії сайту ($Site$).

Міра приналежності не може визначатись за показниками лише однієї категорії (наприклад, кількість переходів за прямим посиланням). Тому для кожного користувача U_i необхідно виділити набір характеристик $(Ch_1^{(i)} \dots Ch)_{N_{Ch}}^{(i)}$, які служать базовими для визначення міри приналежності до аудиторії сайту.

$$B(U_i, Site) = B(Ch_1^{(i)} \dots Ch)_{N_{Ch}}^{(i)}, Site) \quad (6)$$

В якості першої вхідної змінної використано кількість переходів по розміщеному на сайті «лінку». Чим більше користувачів

досліджуваних сайтів переходять на сайт «Anate Studio», тим більше приналежність цих користувачів до аудиторії нашого сайту. У якості другої вхідної змінної використано показник відмови, який показує відсоток користувачів, які переглянули не більше однієї сторінки, або які покинули сайт зі сторінки входу. Високе значення цього показника свідчить про те, що сторінка не відповідає потребам користувача. Третя вхідна змінна — середня кількість переглянутих сторінок за одне відвідування. Високе значення цього показника свідчить про зацікавленість поданою інформацією загалом і пропонованим продуктом зокрема, що безпосередньо впливає на імовірність покупки. Четверта вхідна змінна — це середній час перебування на сайті. Високий даний показник свідчить про якісне відвідування нашого сайту. Тобто, чим більше користувач виділив часу на ознайомлення з інформацією сайту, тим з більшою впевненістю можна стверджувати, що вона для нього була корисною, що в свою чергу прямо впливає на імовірність покупки. В якості вихідної змінної взято оцінку ймовірності замовлення інформаційного продукту представленого на сайті «Anate Studio». Побудовано нечітку модель системи в формі відповідної системи нечіткого виведення.

При побудові нечіткої моделі для оцінки ймовірності покупки товару будемо використовувати шкалу від 0 до 1. Найнижча ймовірність покупки — 0, а найвища — 1. Проведемо фазифікацію змінних.

В якості терм-множини вхідної змінної «Кількість переходів» використано множину $T1 = \{\text{«Низька»}, \text{«Середня»}, \text{«Висока»}\}$ з функціями приналежності зображеними на рис. 1.

Вважаємо, що високий рівень кількості користувачів досягається вже при 300, оскільки за досліджуваний період взято 1 місяць і тип сайту не вимагає періодичного перегляду (як блоги чи соціальні мережі).

Аналогічні припущення взято і для решти терм-множин вхідних характеристик.

Для вихідної змінної «Ймовірність купівлі» використано терм-множину $T5 = \{\text{«Низька»}, \text{«Нижче середнього»}, \text{«Середня»}, \text{«Вище середнього»}, \text{«Висока»}\}$ з функціями приналежності термів, зображеними на рис. 2.

Сформовано базу правил нечіткого виведення в словесній формі з урахування вибору назв і значень наших лінгвістичних змінних та отримано набір правил, які призначені для використання в нашій системі.

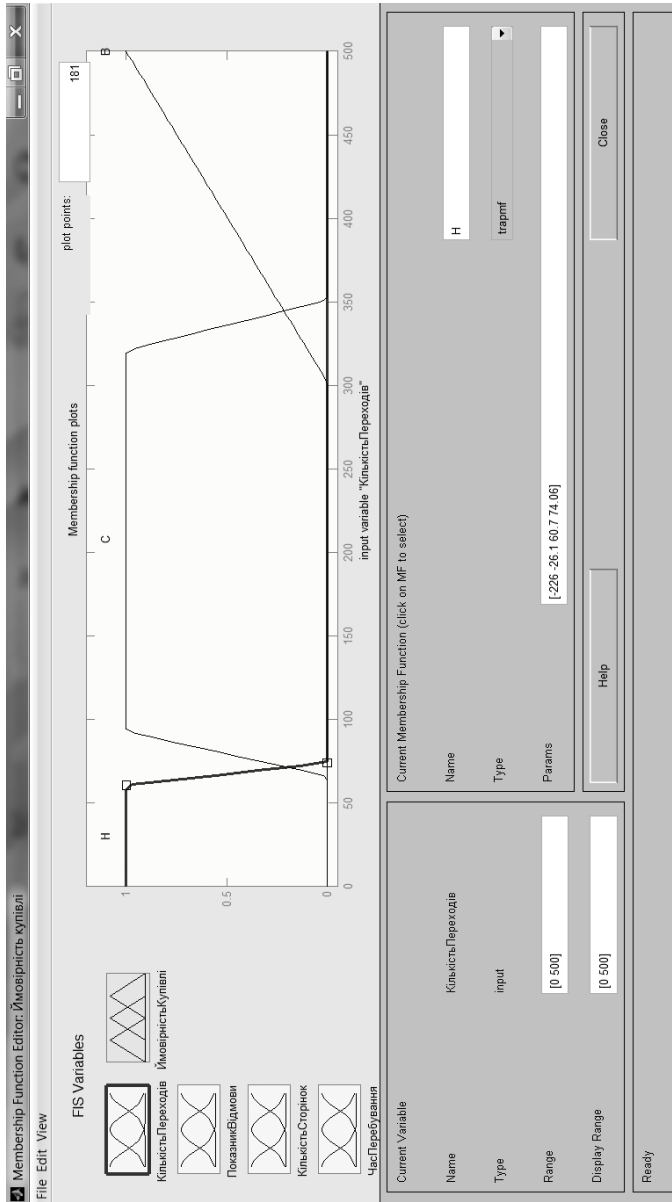


Рис. 1. Терм-множина кількості переходів користувачів з різних сайтів на сайт «Anate Studio» по прямих «ліпках» (переходах)

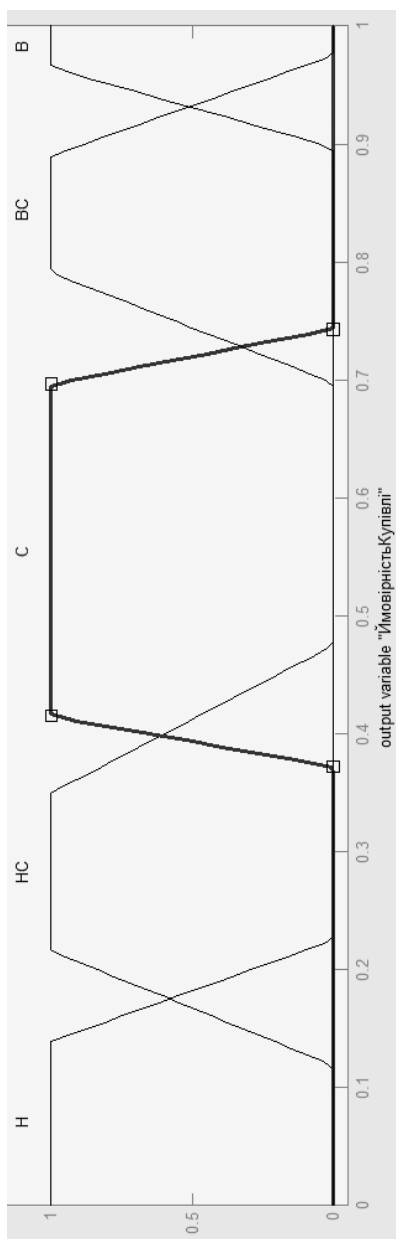


Рис. 2. Терм-множина вихідної змінної — ймовірності покупки інформаційного продукту.

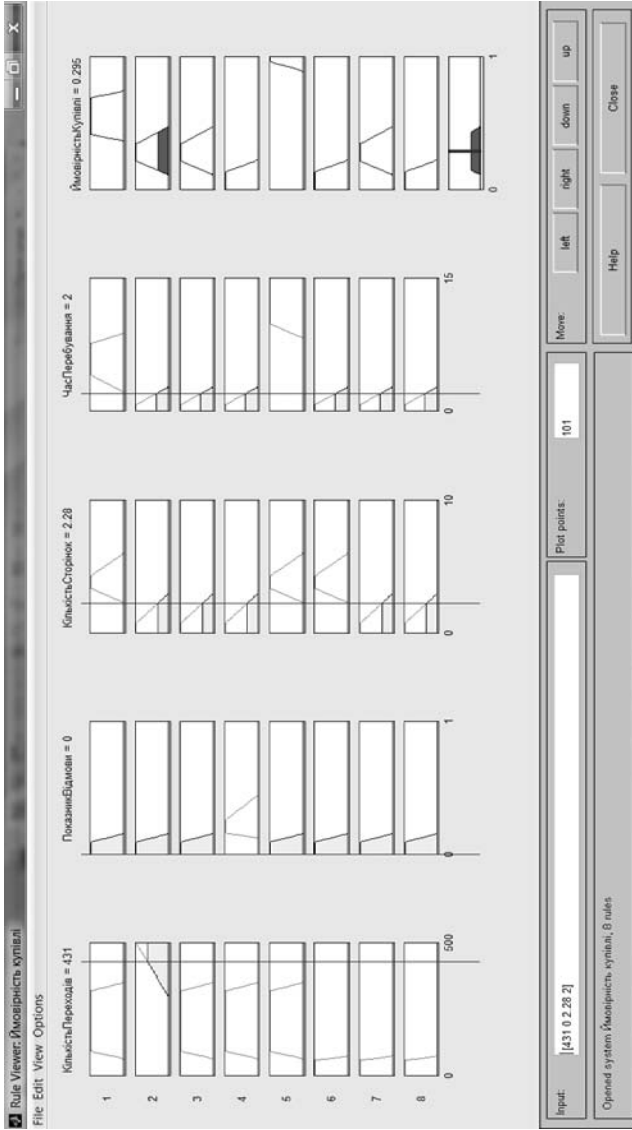


Рис. 3. Вікно введення вхідних даних про сайт forum.dead-code.org

В якості схеми нечіткого виведення використано спрощений алгоритм нечіткого виведення Мамдані. У всіх правилах нечіткого виведення використовується тільки нечітка кон'юнкція, тому в якості методу агрегування використано операцію *min*-кон'юнкції. Для акумуляції висновків правил взято «стандартний» для алгоритму Мамдані метод *max*-диз'юнкції.

Задавши вхідні змінні для описаної моделі (рис. 3), що відповідають статистичним даним сайту forum.dead-code.org, отримано ймовірність купівлі користувачами цього сайту товару, який представлений на нашому сайті рівна 0,29. Ввівши відповідні вхідні змінні по інших сайтах отримано ймовірності наведені у таблиці 1.

Таблиця 1

**ВІДПОВІДНІСТЬ СТАТИСТИЧНИХ ПОКАЗНИКІВ
І ЙМОВІРНОСТЕЙ ПОКУПКИ КОРИСТУВАЧАМИ
З ДЖЕРЕЛ ТРАФІКУ САЙТУ «ANATE STUDIO»**

Джерело або канал	Відвідуваність	Число сторінок за відвідування	Середній час відвідування (хв.)	Показник відмов	Ймовірність покупки
forum.dead-code.org	431	2.28	2	0	0.29
Google	184	6.89	5.28	0	0.5
gry-online.pl	157	2.2	0.08	0	0.295
IndieDB.com	146	4.36	9.31	0	0.961
questomafia.ru	116	4.18	12.42	0	0.965
tetelo.blox.pl	87	2.75	3.02	0.125	0.557
Desura.com	71	7.29	5.03	0	0.5
vk.com	73	2	0	0	0.27
gamedev.ru	32	2.67	0.07	0	0.09
questory.ru	39	3.33	0.1	0	0.091

Як бачимо з табл. 1, найбільшу ймовірність здійснення покупки користувачами мають сайти [indiedb.com](http://IndieDB.com) і questomafia.ru.

Висновки. У даній статі зроблено спробу проаналізувати та певною мірою змоделювати аудиторію сайту за допомогою математичного апарату нечітких множин. Отримані показники ймовірності покупки дають чітке розуміння того, що звичний аналіз окремих числових показників, які характеризують поведінку аудиторії сайту, не дозволяє об'єктивно оцінити наших користувачів. Також усі зміни політики та функціонування сайту, що основані на даних показниках, можуть бути хибними та не приводити до бажаного результату. Таку ситуацію можна спостерігати, на-

приклад, порівнюючи показник відвідуваності та ймовірності купівлі користувачами окремих джерел трафіку нашого сайту. Можна помилково припустити, оскільки з сайту forum.deadcode.org за досліджуваній місяць за посиланням на наш сайт перейшло 431 користувач, то відповідно найбільша кількість потенційних покупців належить саме цій аудиторії та, як мінімум, найактивнішу співпрацю слід проводити саме з цим сайтом. Проте, комплексно проаналізувавши чотири показники поведінки користувачів, а також врахувавши їх належність до сформованої бази правил нечіткого виведення, ми отримали результат, який свідчить про те, що сайт questomafia.ru є більш ефективним джерелом трафіку стосовно ймовірності здійснення покупки нашого продукту користувачами саме цієї аудиторії, яка становить 96 %. Отримавши такі дані, ми можемо спрямовувати певні заходи комерційного сприяння саме на аудиторію questomafia.ru, наша частка якої хоч і не є найбільш чисельною (116 осіб за місяць), але містить найбільш ймовірних потенційних покупців.

Відсутність належної аудиторії сайту, чи в ширшому значенні — позиції сайту, є однією з основних причин невдач у реалізації Інтернет-проекту. Крім того, саме це є головним критичним місцем у реалізації проекту. Саме тому, всі роботи зі створення та підтримки сайту повинні вестися з огляду на необхідність формування його позиції.

Література

1. *Пелещишин А. М.* Інтернет-технології опрацювання консолідованих даних інформаційних ресурсів: навчальний посібник / А. М. Пелещишин; за заг. ред. В. В. Пасічника. — Львів: Видавництво Львівської політехніки, 2010. — 248 с. (Серія «Консолідована інформація» вип.4)
2. *Гаркавенко С. С.* Маркетинг. Підручник. — 7-ме вид. — Київ: Лібра, 2010. — 720 с.
3. *Пелещишин А. М., Шаховська Н. Б.* Використання апарату нечітких множин для опису аудиторії веб-сайту // Штучний інтелект.- № 3/2005. — Донецьк. — 2005. — С. 521—530.
4. *Рассихина В. Є.* Використання інструментів інтернет-аналітики в маркетинговому сегментуванні // Економічні науки. — № 6/2011 [Електронний ресурс]. — Режим доступу - http://www.rusnauka.com/31_ONBG_2011/Economics/6_96557.doc.htm 11

Стаття надійшла до редакції 19.04.2013 р.