

УДК 316.776:519.213

**МОДЕЛЮВАННЯ ТА ЕФЕКТИВНІСТЬ ДОСТУПУ ДО
ПОСЛІДОВНО ОРГАНІЗОВАНИХ СТОРІНОК WEB-
САЙТУ ДЛЯ РІЗНИХ ЗАКОНІВ РОЗПОДІЛУ
ЙМОВІРНОСТЕЙ ЗВЕРТАННЯ ДО СТОРІНОК**

М.І. Юськів – аспірант, ЛНУ ім.І.Франка,
markiyan_yuskiv@live.com

Г.Г. Цегелик – д.ф. –м.н., професор, ЛНУ ім.І.Франка

Розглянуто підхід до математичного моделювання оптимального доступу користувачів до послідовно організованих сторінок Web-сайту. Знайдено вираз для математичного сподівання загального часу, необхідного для пошуку сторінки, для різних законів розподілу ймовірності звертання до сторінок.

Yuskiv M.I., Tsehelyk H.H. Modelling and efficiency of access to adherently organized pages of Web-site for different distribution laws of probability of invocation to the page.

Ключові слова: МОДЕЛЮВАННЯ ДОСТУПУ, ПОСЛІДОВНО ОРГАНІЗОВАНІ СТОРІНКИ, WEB-САЙТ, ЗАКОНИ РОЗПОДІЛУ ЙМОВІРНОСТЕЙ.

Keywords: MODELLING OF ACCESS, ADHERENTLY ORGANIZED PAGES, WEB-SITE, DISTRIBUTION LAWS OF PROBABILITY

Мережа інтернет є унікальним інформаційним ресурсом, через який кожен день проходять великі об'єми інформації. Основними складовими цього ресурсу є сервери та Web-сайти. Кожен сайт відповідно складається з набору сторінок. Сторінки сайту – атомарні одиниці, які між собою пов'язані логічними зв'язками. Нами розглянуто організацію і перегляд сторінок сайту при заданих ймовірностях звертання до сторінок. За критерій ефективності пошуку сторінки приймається математичне сподівання загального часу, необхідного для пошуку потрібної сторінки користувачем. Вважатимемо, що пошук інформації на сайті відбувається послідовно по сторінках. Серед законів розподілу ймовірностей звертання до сторінок

розглянемо рівномірний, бінарний, узагальнений закони розподілу та Закон Зіпфа [1, 2].

Припустимо, що інформація, яка міститься на *Web*-сайті, розміщена на N сторінках, розбитих на n блоків по m сторінок в кожному. Нехай p_i – ймовірність звертання до i -ї сторінки; $a = b + dm$ – час читання блоку сторінок користувачем, де b, d – деякі сталі; t – середній час перегляду однієї сторінки; E – математичне сподівання загального часу, необхідного для пошуку веб-сторінки на сайті.

Тоді E виразиться наступною формулою

$$E = \sum_{i=1}^n \sum_{j=1}^m (ia + ((i-1)m + j)t) p_{(i-1)m+j}$$

Знайдемо явний вираз для E як для рівномірного закону розподілу ймовірностей звертання до сторінок, так і для “бінарного”, закону Зіпфа та узагальненого закону розподілу [1, 3].

1. У випадку рівномірного закону розподілу ймовірностей звертання до сторінок отримаємо наступний вираз для E

$$E = \frac{1}{2} ((n+1)a + (N+1)t)$$

або

$$E = \frac{1}{2} \left(\left(\frac{N}{m} + 1 \right) (b + dm) + (N+1)t \right)$$

2. У випадку, коли ймовірності звертання до сторінок задовольняють “бінарний” закон, отримаємо наступну формулу

$$E = \left(\frac{2^m}{2^m - 1} a + 2t \right) (1 - 2^{-N})$$

Якщо знехтувати величиною 2^{-N} , то з достатньо високою точністю можемо прийняти

$$E = \left(\frac{2^m}{2^m - 1} a + 2t \right).$$

3. Якщо ймовірності звертання до сторінок розподілені за законом Зіпфа, то

$$E = \frac{1}{H_N} \left(((n+1)H_N - S_m(n))(a+mt) + \left(\binom{1}{n} S_m(n) - H_{N+1} \right) N - mH_N \right) t,$$

де

$$S_m(n) = \sum_{k=1}^n H_{km}.$$

Використовуючи апроксимацію $S_m(n)$ функцією $\bar{S}_m(n)$,

$$\bar{S}_m(n) = n(H_N - 1) + \frac{1}{2} \ln n + C_1,$$

де $C_1 = \frac{1}{2} \ln 2\pi$, одержимо

$$E = \frac{1}{H_N} \left(\left(H_N + n - \frac{1}{2} \ln n - \frac{1}{2} \ln 2\pi \right) a + Nt \right).$$

4. У випадку, якщо розподіл ймовірностей звертання до сторінок задовольняє узагальнений закон розподілу, отримуємо

$$E = \frac{1}{H_N^{(c)}} \left(((n+1)H_N^{(c)} - S_m^{(c)}(n))(a+mt) + \right. \\ \left. (H_N^{(c-1)} + mS_N^{(c)}(n) - nH_N^{(c)} - mH_N^{(c)})t \right),$$

де

$$S_m^{(c)}(n) = \sum_{k=1}^n H_{km}^{(c)}.$$

Використовуючи апроксимацію $S_m^{(c)}(n)$ функцією $\bar{S}_m^{(c)}(n)$, де

$$\bar{S}_m^{(c)}(n) = nH_N^c + \frac{N^{1-c}}{1-c} \left(\frac{c-1}{2-c} n + \frac{\alpha^{(c)}(n)}{n^{1-c}} \right)$$

З достатньо високою точністю отримуємо

$$E = \frac{1}{H_N^{(c)}} \left(\left(H_N^{(c)} - \frac{N^{1-c}}{1-c} \right) \left(\frac{c-1}{2-c} n + \frac{\alpha^{(c)}(n)}{n^{1-c}} \right) \right) \left(b + \frac{dN}{n} \right) + H_N^{(c-1)} t$$

В результаті виконання роботи побудовано математичну модель доступу до сторінок інформаційного Web-сайту зі сторони користувача. За критерій ефективності прийнято математичне сподівання загального часу, необхідного для послідовного пошуку сторінки на Web-сайті. Математична модель враховує ймовірності звертання до сторінок, час читання

Інформатика та системні науки (ІСН-2016)

блоку сторінок та час перегляду сторінки користувачем. Знайдено вирази для математичного сподівання, залежні від різних законів розподілу ймовірності звертання до сторінок.

1. Кнут Д. Искусство программирования для ЭВМ: Сортировка и поиск. — М.: Издательский дом “Вильямс”, 2000. — Т.3. — 840 с.
2. Цегелик Г.Г. Организация и поиск информации данных. — Львов: Свит, 1990. — 186 с.
3. Baeza-Yates R., Castilio C. Relating Web Structure and User Search Behavior. — Center for Web Research // Department of Computer Science, University of Chile, 2002. — 24 p.
4. Hu W.-C., Chen Y., Smalz M., Ritter G. An Overview of World Wide Web Search Technologies // Department of Computer Science. Auburn University, 2000. — 6 p.