

## ДИСКРЕТНА ДИНАМІЧНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ ТА ПРОГНОЗУВАННЯ ВІДВІДУВАНОСТІ ВЕБ-САЙТУ

Пасічник Н.Р.

*Тернопільський національний економічний університет, аспірант*

### I. Постановка проблеми

Відвідуваність тематичних сторінок Веб-сайту підвищує його рейтинг у відповідних тематичних доменах. Це сприяє підвищенню відвідуваності інших тематик сайту в цілому і зокрема відвідуваності цільових сторінок, які безпосередньо направлені на виконання Веб-сайтом своїх завдань. Одним із факторів, що забезпечують таку відвідуваність є регулярне поповнення тематик сайту актуальним контентом. Ефективність такого поповнення можна оцінювати за допомогою формалізованих математичних моделей, що дозволить також прогнозувати ефективність подібних операцій в майбутньому.

### II. Мета роботи

Прогнозування відвідуваності Веб-сторінок частково висвітлене в літературі, зокрема в роботах [1-2]. Згадані роботи досліджують питання навігації користувачів по сторінках сайту з метою виявлення найчастіше відвідуваних маршрутів а також досліджують питання вибору ефективних сайтів та Веб-сторінок для розміщення реклами. В той же час розробники Веб-сайтів зацікавлені в рекомендаціях щодо напрямів та темпів розвитку даного продукту, які б забезпечили максимізацію його результативності. Одному із підходів до вирішення цієї задачі присвячена дана робота.

### III. Структура динамічної моделі та експерименти

Для підвищення зручності навігації по Веб-сайту, його структура повинна бути добре зрозумілою представникам аудиторії відповідної тематики. Вона формується на основі формального аналізу Веб-контенту, як описано в роботі [3]. Щоб забезпечити достатню відвідуваність розроблюваних тематик, їх необхідно узгоджувати із ключовими бізнес-процесами об'єкта, який представляється Веб-сайтом. Наповнення тематик повинно задовольняти критеріям актуальності та повноти. Систематизувати зусилля по формуванню наповнення тематик відповідно до згаданих критеріїв можна на основі спеціальних тезаурусів, які формуються в автоматизованому режимі на основі методу, описаному в роботі [4].

Побудову моделей відвідуваності розпочнемо із визначення результуючих та вхідних параметрів. Результуючими параметрами обираємо  $y_1$  загальну відвідуваність сторінок Веб-сайту, як фактору його представництва в рейтингах по пошукових запитах, а також відвідуваність  $y_2$  результуючих сторінок Веб-сайту як засіб поширення цільової інформації. Параметрами управління моделі оберемо інтенсивність наповнення затребуваною інформацією тематик сайту, які користуються найвищою популярністю. Цю інтенсивність оцінюємо за передбачуваними відвідуваностями  $x_j$  згаданих тематик.

Щоденна відвідуваність характеризується великою кількістю випадкових факторів, що робить проблематичним достатньо точне її прогнозування. В той же час на характеристики сайту впливають не окремі екстремальні, а усереднені показники. Тому в нашій моделі аналізуватимемо середньотижневі відвідуваності. Для врахування інерційних характеристик відвідуваності включимо в модель лагове значення прогнозованої величини, яке просумуємо із зваженими відповідними значеннями відвідуваностей актуальних тематик. Для координації динаміки найвідвідуваніших тематик із результуючими відвідуваностями до першої застосуємо степеневу функцію. Еволюцію Веб-сайту враховуємо за рахунок впровадження в моменти часу  $T_j$  нових підструктур, додаткова відвідуваність яких  $x_j$  суттєво впливатимуть на загальну відвідуваність. В результаті отримаємо наступні співвідношення

$$y_{i,k} = g_{i,1}y_{i,k-1} + q_{i,1}(x_{1,k})^{q_{i,2}} + \sum_{j=2}^m H(t - T_j)[g_{ij}z_{jk-1}^i + q_{i,2j-1}(x_{j,k})^{q_{i,2j}}] \quad , \quad (1)$$

де

$$z_{j,k}^i = y_{i,k} - g_{i,1}y_{i,k-1} - q_{i,1}(x_{1,k})^{q_{i,2}} - \sum_{l=2}^{j-1} H(t - T_l)[g_{i,l}z_{l,k-1}^i + q_{i,2l-1}(x_{l,k})^{q_{i,2l}}] \quad (2)$$

похибки прогнозу  $i$ -ої результуючої змінної в момент часу  $k$  внаслідок впливу  $j$ -го фактора, який виник момент часу  $T_j$ ;  $H$  - функція Хевісайда,  $g, q$  - параметри моделі, які необхідно ідентифікувати.

Для ідентифікації моделі використовується метод найменших квадратів в точках навчальної вибірки. Після ідентифікації параметрів, модель(1)-(2) використовується як прогнозна, значні похибки якої сигналізуватимуть про появу нових суттєвих неврахованих факторів. При цьому точність моделі оцінюється по контрольній вибірці. Результати експериментів підтвердили ефективність запропонованої методики. Максимальна похибка прогнозу при цьому не перевищувала 12%.

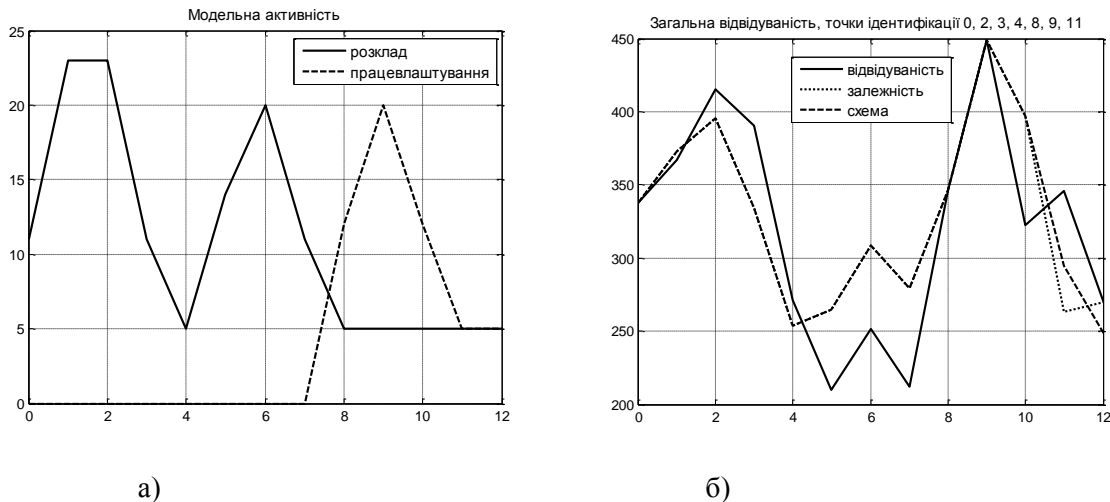


Рисунок 1 – Ілюстрація процесу ідентифікації моделі: а) передбачувана активність по тематиках Веб-сайту ФКІТ; б) залежність між вхідними та результуючою змінною а також прогнозована відвідуваність, побудована на основі різницевої схеми.

### Висновок

У статті розглянуто один із можливих шляхів побудови дискретної динамічної моделі відвідуваності Веб-сайтів. В основу моделі покладено нелінійні співвідношення між результуючими показниками загальної відвідуваності а також відвідуваності результуючих тематик із параметрами прогнозованої відвідуваності найчастіше оновлюваних тематик та лаговими результуючими показниками. Для ідентифікації запропонованої моделі використано критерій мінімізації середньоквадратичної похибки на множині контрольних точок. Прогнозні значення моделі будуються в режимі обчислень за відповідною різницевою схемою із відомими початковими значеннями. У результаті проведених експериментів підтверджено ефективність побудованої моделі. Це уможливило прогнозування ефективності заходів щодо розвитку наповнення Веб-сайтів а також дозволяє виявити появу нових факторів, що суттєво впливають на цільову відвідуваність.

### Список використаних джерел

1. Gorbunov A.L. Markov models for website traffic // IMAT -2007, Ural University Publ., p 65-73, 2007.
2. Khalil F. Combining. Web Data Mining Techniques for Web Page Access Prediction. // Queensland, 2008, pp.197. <http://eprints.usq.edu.au/4341/>
3. Пасічник Н.Р., Дивак М..П. Метод та алгоритм побудови структури контенту Веб-сайту на основі онтологічного підходу // Вісник Донецького НТУ, серія "Інформатика, кібернетика та комп'ютерні науки", т.15, 2012. – С.184-189.
4. Пасічник Н. Метод формування онтологічного контенту, базований на аналізі інформації спеціалізованих Веб-сайтів // Вісник ХНУ: Інженерія, т.5, 2012. – С.241-244.