

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ "ЛЬВІВСЬКА ПОЛІТЕХНІКА"

Пасічник Наталія Романівна

УДК 004.738.5:519.673

**МАТЕМАТИЧНІ МОДЕЛІ ВІДВІДУВАНOSTІ
ВЕБ-САЙТІВ ТА МЕТОДИ ЇХ ІДЕНТИФІКАЦІЇ**

01.05.03 – математичне та програмне забезпечення
обчислювальних машин і систем

Автореферат

дисертації на здобуття наукового ступеня
кандидата технічних наук

Львів - 2014

Дисертацією є рукопис.

Робота виконана у Тернопільському національному економічному університеті Міністерства освіти і науки України.

Науковий керівник:

доктор технічних наук, професор
Дивак Микола Петрович,
Тернопільський національний економічний університет,
декан факультету комп'ютерних інформаційних технологій,
завідувач кафедри комп'ютерних наук.

Офіційні опоненти:

доктор технічних наук, професор
Пелешишин Андрій Миколайович,
Національний університет "Львівська політехніка",
завідувач кафедри соціальних комунікацій та
інформаційної діяльності;


кандидат технічних наук, доцент
Шпортко Олександр Володимирович,
Рівненський державний гуманітарний університет,
доцент кафедри економічної кібернетики.

Захист відбудеться 09 жовтня 2014 о 16 годині на засіданні спеціалізованої вченої ради Д 35.052.05 у Національному університеті "Львівська політехніка" за адресою: 79013, м. Львів, вул. С.Бандери, 12.

З дисертацією можна ознайомитися в науково-технічній бібліотеці Національного університету "Львівська політехніка" (79013, м. Львів, вул. Професорська, 1).

Автореферат розіслано 8 вересня 2014 р.

Т.в.о. вченого секретаря
спеціалізованої вченої ради
доктор технічних наук, доцент



Н.Б.Шаховська

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. В умовах сучасного інформаційного суспільства роль Всесвітньої павутини, тобто Вебу, неухильно зростає. Кожен із Веб-сайтів, які складають одну із основних компонентів Вебу, створюється з певною метою, рівень досягнення якої виражається в його результативності. Важливим засобом досягнення високої результативності є достатня відвідуваність. В той же час дослідники відзначають низьку та вкрай низьку відвідуваність переважної більшості Веб-сайтів, на розробку та підтримку яких витрачаються певні кошти.

Варто відзначити, що характеристики якості Веб-сайтів у значній мірі формуються за рахунок рекомендацій пошукових систем, тому існує цілий спектр послуг та рекомендацій для підвищення видимості певних матеріалів у згаданих системах, об'єднаних спільним терміном «пошукова оптимізація» або SEO. Ці рекомендації мають емпіричний, а часто навіть напівлегальний характер (т.зв. «сіра» та «чорна» оптимізація). Внаслідок етичних міркувань, а також у зв'язку із посиленням боротьби провідними пошуковими системами із переоптимізацією Веб-сайтів, зростає потреба у науковому підході до дослідження проблеми підвищення відвідуваності. Ефективним засобом побудови обґрунтованих рекомендацій є математичне моделювання.

Основою формування наповнення Веб-сайту є його контент. Різні аспекти аналізу контенту Веб-сайту та шляхів його формування досліджувалася в працях Пелещишина А. М., Дубового В. М., Москвіна О. М., Smyth P., Cades I., Khalil F., Stibe A., Vicevskis J., Menczer F., Ratkievich J. Вимірювати ефективність модифікації Веб-сайту можна обсягом дій, здійснюваних користувачами Веб-сайту, бажаних для його власника, тобто результативністю. Моделювання такого показника вимагає виділення шаблонів поведінки користувачів, які розглядаються в науковій літературі та узагальнені в дослідженнях F. Khalil, P. Smyth, I. Cades на основі ланцюгів Маркова, методів кластеризації та асоціативних правил. В роботах В. М. Дубового та О. М. Москвіна запропоновано використати мультиагентні системи для структурування Веб-сайтів. Згадані підходи ефективні для Веб-сайтів, які мають високу відвідуваність. У дослідженнях А. М. Пелещишина відвідуваність Веб-сайту запропоновано підвищувати в рамках Веб-холдингу за рахунок високовідвідуваних сайтів-донорів, які також потрібно певним чином сформувані.

В дослідженні F. Menczer та J. Ratkievich на основі аналізу мільйонів Веб-сторінок показано, що їх кількісний показник відвідуваності відповідає степеневому закону розподілу із “важким хвостом”. Такий розподіл передбачає значущу вірогідність появи рівнів відвідуваності, які значно відрізняються від середнього рівня. Разом з тим, не досліджено фактори впливу, що спричиняють суттєві зміни відвідуваності. В дослідженнях А. Stibe, J. Vicevskis пропонується мова моделювання структури Веб-сайтів, яка дає можливість їх точнішого прототипування, ніж засіб загального призначення UML. Ці розробки стосуються покращення формальної реалізації Веб-сайту, залишаючи поза увагою питання структурування його змісту.

Таким чином, питання підвищення відвідуваності Веб-сайтів шляхом формування відповідного контенту залишається недостатньо дослідженим. Такі

дослідження повинні давати як рекомендації по реструктуризації Веб-сайту, так і по часовому періоду та інтенсивності, з якою її необхідно здійснювати. Самі дослідження повинні мати певну теоретичну базу, а також давати конкретні практичні рекомендації. Тому актуальним науковим завданням є підвищення відвідуваності Веб-сайтів шляхом розроблення процедур структурування контенту із використанням програмних засобів та інструментів математичного моделювання.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалася у відповідності з планами навчальної та науково-дослідної роботи Тернопільського національного економічного університету, зокрема в рамках теми «Інформаційна технологія для ідентифікації та візуалізації зворотного гортанного нерва в процесі хірургічної операції на щитовидній залозі» (номер державної реєстрації 0112U000078, 2012-2013 рр.), у якій автором розроблено критерії ефективності Веб-сайтів, а також метод побудови структур тематичних Веб-сайтів на основі онтологічного підходу. При дослідженнях по темі «Теорія побудови та методи реалізації в реальному часі міждисциплінарних математичних моделей зміни стану складних об'єктів» (номер державної реєстрації 0114U000569), автором розроблено дискретну динамічну модель відвідуваності Веб-сайтів, а також метод її ідентифікації. При виконанні госпдоговірної науково-дослідної роботи на тему „Макромодельовання складних систем та процесів в умовах структурної невизначеності на основі неточних даних” (номер державної реєстрації 0111U010356) автором розроблено метод ідентифікації моделі короткотермінової динаміки відвідуваності Веб-сайту. При виконанні госпдоговірної науково-дослідної роботи на тему „Математичні моделі процесів навантаження та релаксації в складних системах” (номер державної реєстрації 0113U000848) автором розроблено метод ідентифікації моделі довготермінової динаміки відвідуваності Веб-сайту.

Мета і задачі дослідження. Метою дослідження є розроблення математичних та програмних засобів, що використовують структурування контенту та базуються на математичному моделюванні процесів функціонування Веб-сайтів для підвищення їх відвідуваності.

Для досягнення цієї мети у дисертаційній роботі вирішено такі задачі:

- 1) дослідження особливостей методів та засобів підтримки відвідуваності Веб-сайтів із використанням математичних моделей;
- 2) розроблення моделей та методів автоматизованої побудови структур Веб-сайтів;
- 3) побудова методу ідентифікації математичних моделей короткотермінового прогнозу відвідуваності Веб-сайтів;
- 4) створення методу ідентифікації математичних моделей довготермінового прогнозу відвідуваності Веб-сайтів;
- 5) застосування розроблених методів ідентифікації моделей та структурування контенту для підвищення відвідуваностей Веб-сайтів на прикладі Веб-сайтів навчальних підрозділів, а також промо-сайтів програмних продуктів.

Об'єкт дослідження: процеси функціонування Веб-орієнтованих програмних систем.

Предмет дослідження: математичні моделі відвідуваності Веб-сайтів і процесів структурування контенту та методи їх ідентифікації.

Методи дослідження. Дослідження, виконані під час роботи над дисертацією, ґрунтуються на методах системного аналізу, методах прогнозування, а також теорії динамічних систем. Використання методів оптимізації та чисельних методів дало можливість сформулювати методи ідентифікації моделей довго- та короткотермінової відвідуваності Веб-сайтів. Використання методів Веб-майнінгу та структурування Веб-онтологій забезпечило розвиток методу побудови структур тематичних Веб-сайтів.

Наукова новизна одержаних результатів.

Вперше:

1) запропоновано та обґрунтовано метод ідентифікації математичних моделей короткотермінового прогнозу відвідуваності Веб-сайтів на основі диференціальних рівнянь Моно, який, на відміну від існуючих, враховує отримані аналітичні вирази для обчислення початкових наближень параметрів моделі та другої похідної функції якості контенту Веб-сайту, що забезпечує адекватне відображення короткотермінової динаміки відвідуваності і підвищення якості структурування контенту;

2) побудовано та обґрунтовано метод ідентифікації моделі довготермінового прогнозу відвідуваності Веб-сайтів, який, на відміну від існуючих, враховує лінійні екстраполяції значень функції якості контенту Веб-сайту, а також її похідної, що у сукупності забезпечує адекватне відображення довготермінової динаміки відвідуваності і підвищення якості структурування контенту;

3) створено моделі довго- та короткотермінової відвідуваності Веб-сайтів на основі модифікованої системи звичайних диференціальних рівнянь Моно, які, на відміну від існуючих, враховують функції впливу на якість контенту Веб-сайту, що забезпечило адекватність відображення процесу зміни відвідуваності Веб-сайту і можливість створення формальних процедур його підтримки.

Набув подальшого розвитку метод побудови структур тематичних Веб-сайтів на основі частотного аналізу та ранжування, який, на відміну від існуючих, використовує аналіз узгодженості структурних елементів Веб-сайту з предметною областю, а також структуру базових концептів експертних запитів, що уможливорює автоматизацію побудови типових структур тематичних Веб-сайтів, а також дозволяє автоматизувати процеси оновлення контенту.

Практичне значення одержаних результатів полягає в тому, що на основі розроблених методів та засобів реалізовано систему засобів підтримки процесу підвищення якості тематичних Веб-сайтів. На базі системи розроблено програмне забезпечення, яке використовується у навчальному процесі в Тернопільському національному економічному університеті та в ПП «МагнетікВан». Зокрема, основні практичні результати дисертаційної роботи використано:

- для підвищення відвідуваності Веб-сайту факультету комп'ютерних інформаційних технологій Тернопільського національного економічного університету;
- у навчальному процесі на кафедрі комп'ютерних наук Тернопільського національного економічного університету при викладанні дисциплін

- «Web дизайн» та «Управління продажами в ІТ»;
- для прогнозування динаміки відвідуваності Веб-сайту проекту API2Cart ПП «МагнетікВан».

Особистий внесок здобувача. Усі наукові результати, викладені у дисертаційній роботі, отримані автором особисто. У друкованих працях, опублікованих у співавторстві, автору належать такі результати: [1] – запропоновано та обгрунтовано набір дискретних динамічних моделей для прогнозування відвідуваності Веб-сайту; [2] – запропоновано та обгрунтовано методи ідентифікації моделі процесу короткотермінового росту відвідуваності Веб-сайтів та процесу довготермінового росту відвідуваності Веб-сайтів; [3] – удосконалено метод формування онтологій шляхом аналізу зашумленої слабо структурованої інформації тематичних Веб-сторінок спеціалізованих Веб-сайтів; [4] – запропоновано та обгрунтовано метод формування онтологій тематичних Веб-сторінок спеціалізованих Веб-сайтів; [7] – запропоновано та обгрунтовано формалізовану систему показників якості Веб-сайту та факторів, що сприяють їх покращенню; [8] – покращено набір неавтономних дискретних динамічних моделей прогнозування відвідуваності Веб-сайту; [9] – запропоновано та обгрунтовано динамічну модель згладжених тижневих відвідуваностей Веб-сайту; [10] – розроблено метод контролю відвідуваності Веб-сайту на основі прогнозування параметрів відвідуваності розроблюваних Веб-сторінок; [12] – запропоновано та обгрунтовано інформаційну технологію підтримки аналізу структури Веб-сайту, що зменшує зусилля по її реалізації; [17] – покращено метод адаптивного управління структурою навчального процесу.

Апробація результатів дисертації. Основні положення дисертації доповідалися та обговорювалися на міжнародних та всеукраїнських конференціях, семінарах, симпозіумах: X, XI, XII Міжнародних науково-технічних конференціях “Сучасні проблеми радіоелектроніки, телекомунікацій та комп’ютерної інженерії: TCSET’2010, 2012, 2014 (Львів – Славсьько, 2010, 2012, 2014 рр.); XII Міжнародній науково-технічній конференції “Досвід розробки й застосування САПР в мікроелектроніці CADSM’2013” (Львів – Поляна, 2013 р.); I, II, III Всеукраїнських школах-семінарах молодих вчених і студентів «Сучасні комп’ютерні інформаційні технології», 2011, 2012, 2013 рр. (ТНЕУ, Тернопіль); VI Міжнародному науково-технічному симпозіумі “Інтелектуальні засоби збору даних і сучасні обчислювальні системи: розробка та застосування IDAACS’2011” (Чехія, Прага, 2011 р.); II міжнародній конференції «Інформаційні технології та комп’ютерна інженерія» (Харків, 2011р); V міжнародній школі–семінарі для молодих науковців «Індуктивне моделювання: теорія і застосування» (с. Жукін, Київська обл., 2012 р.); IV міжнародній конференції з Індуктивного моделювання ICIM’2013, (Київ, 2013р); науково-методичних семінарах кафедри комп’ютерних наук Тернопільського національного економічного університету (Тернопіль, 2011-2014 рр.).

Публікації. Основні результати дисертаційної роботи висвітлено в 17 друкованих працях, загальним обсягом 52 сторінки, у тому числі одна стаття в іноземному періодичному науковому виданні [1], 6 статей у фахових наукових виданнях України [2-7], 10 публікацій у матеріалах конференцій [8-17].

Структура дисертації. Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Робота містить 133 сторінки основного тексту, 5 таблиць, 57 рисунків, 6 додатків, 143 найменування використаних джерел. Загальний обсяг дисертації – 178 сторінок. Додатки викладено на 21 сторінці.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** наведено загальну характеристику роботи, обґрунтовано актуальність теми, визначено об'єкт та предмет дослідження, сформульовано мету та задачі роботи, розкрито наукову та практичну цінність отриманих результатів, а також подано відомості про апробацію дисертації.

У **першому розділі** досліджено особливості відомих методів та засобів структурування контенту Веб-сайтів, виділено фактори, що впливають на їх ефективність, проаналізовано методи підтримки процесу формування контенту, розглянуто моделі оцінювання ефективності реструктуризації Веб-сайтів.

Зокрема, наведено загальну характеристику Веб-середовища та основних етапів його розвитку, а також загальну класифікацію прикладних Веб-сайтів. Відзначено ключову роль відвідуваності Веб-сайтів в оцінці їх ефективності, а також наведено загальну класифікацію факторів впливу на відвідуваність. Ці фактори поділено на групу, що визначає видимість по тематичних запитах у пошукових системах, а також групу, яка обумовлює якість публікації. Серед основних факторів, що визначають видимість публікації, відзначено узгодженість із запитами як структурних елементів, так і наповнення Веб-сторінки, а також достатнє рекламування сторінки в спеціалізованих ресурсах. Серед факторів, що впливають на якість публікації, виділено актуальність публікації та повноту тематики, яка включає цю сторінку. Сюди також віднесено авторитетність автора публікації та посилань, наведених у ній.

Наведені критерії вказують, що наповнення Веб-сайту – один із найважливіших факторів його популярності. Це вимагає залучення відповідальних та допитливих працівників, які мають оволодівати новими для себе знаннями, поступово покращуючи якість матеріалів, розміщених на Веб-сайті та поетапно готуючи його аудиторію до сприйняття дедалі складнішої актуальної інформації. Для правильної організації процесу засвоєння базових та виробництва нових знань необхідно володіти засобами пошуку потрібної інформації та засобами її структурування адекватно поставленим задачам. Автоматизація цього трудомісткого процесу можлива із застосуванням методів та засобів Web Mining.

Основними класами методів Web Mining вважаються: пошуковий Web Mining, структурний Web Mining, видобування знань із Вебу, персоналізація інформації у Вебі, пошук шаблонів у поведінці користувачів Вебу. Серед аналізованого напрямку досліджень найближчими до роботи автора дисертації є методи видобування знань із Веб-текстів. Для систематизації знань, вилучених з Веб-документів методами Web Mining, зручно використовувати онтології, які спрощують їх формальну обробку. Одним із ефективних методів побудови Веб-онтологій є метод підписів теми, розроблений Lin С.-У. та Нову Е.Н.

На погляд автора дисертації, проблеми суто синтаксичних підходів у побудові онтологій полягають у суттєвій невизначеності структурування онтологій у встановленні її ядра. Призначення онтології має ґрунтуватися на аналізі розповсюдженої практики діяльності у відповідній галузі. В цьому дослідженні автор обмежився двома класами Веб-сайтів – навчальними сайтами та промо-сайтами програмних продуктів, які часто страждають від низької відвідуваності.

Для формалізації подальшого наповнення онтології необхідно встановити критерій ефективності розвитку структури і контенту Веб-сайтів, а також методи прогнозування ефективності різних його варіантів. У цьому контексті під ефективністю розуміємо здатність деякої обставини чи дії спричинити певний кількісно вимірюваний ефект. В результаті досліджень встановлено, що для Веб-сайтів із невисокою відвідуваністю основним критерієм такої ефективності служить саме показник відвідуваності.

Окрім класичних методів моделювання необхідно врахувати і специфіку роботи із людино-машинними системи, якими є Веб-сайти, а саме: необхідність відсіву аномально високих збурень відвідуваності, необхідність врахування ефекту, коли параметри моделей фактора впливу на відвідуваність змінюються у залежності від періоду його реалізації. Тому ідентифікацію моделей відвідуваності слід здійснювати в процесі їх спостереження по етапах. При цьому спостережені значення на попередньому етапі служать для ідентифікації параметрів моделі на наступному етапі.

Проведений аналіз дав змогу сформулювати наукове завдання – підвищення відвідуваності Веб-сайту шляхом розроблення процедур структурування контенту із використанням програмних засобів та інструментів математичного моделювання.

У **другому розділі** запропоновано та обґрунтовано ряд підходів до зниження трудомісткості процесів структурування Веб-сайту та його інформаційного наповнення. Зокрема розроблено концептуальну модель якості Веб-сайтів, методи побудови структур тематичних Веб-сайтів на основі онтологічного підходу та формування шляхів оновлення контенту Веб-сайту на основі аналізу слабо структурованої інформації. Проведений аналіз кращих Веб-сайтів провідних університетів України підтвердив суттєвий вплив аналітичних матеріалів, продуктів та сервісів на популярність сайтів та на їх позиціонування в пошукових системах. Аналіз аргументів запропонованої моделі якості сайту показав першочергову необхідність планування та контролю структури його контенту для збільшення відвідуваності.

З метою підвищення комфорту користувачів загальна структура тематичного Веб-сайту повинна бути легко впізнаваною для цільової аудиторії. Для автоматизації такого структурування розроблено метод побудови типового головного меню тематичного Веб-сайту. При цьому тематика задається текстовим маркером *SAM* предметної області та уточнюється специфікатором *SS*. За допомогою пошукового сервера та операції конкатенації символічних стрічок, за тематичним запитом $QS = SAM \ \& \ SS$ отримуємо множину *SP* потужності *P*, що складається з HTML-кодів HP_i Веб-сторінок:

$$SP(QS, P) = \{HP_i\}_{i=1}^P, \quad (1)$$

з тегу *title* якої виділяємо підмножину кодів предметної області:

$$SPR = \{HP_{i^*}^* | HP_{i^*} \in SP, HP_{i^*}.title \cap SAM \neq \emptyset\}_{i=1}^P. \quad (2)$$

Зміст пунктів меню встановлюємо на основі анкорів тегів "*< a... href ... >*", які групуються в ієрархічні списки. Вибрані елементи верхнього рівня утворюють впорядкований список $LRA(HP_i^*)$ анкорів сторінки:

$$LRA(HP_i^*) = \langle A_{ij}(HP_i^*) \rangle_{j=1}^{AP_i}. \quad (3)$$

На основі списків анкорів та виділених елементів сформуємо множину SCF частот NC концептів C :

$$SCF = \left\{ (C_l, NC_l) / C_l \in \bigcup_i LRA(HP_i^*) \right\}. \quad (4)$$

Для виявлення концептів, релевантних до предметної області, впорядкуємо елементи множини SCF у порядку спадання частот елементів та формуємо концептуальну множину SCN :

$$SCN = \left\{ (C_l, NC_l) / FCL_l = \frac{NC_l}{P^*} > F_0 \right\}, \quad (5)$$

де пороговий рівень частоти F_0 вибирають виходячи зі специфіки дослідження.

Для впорядкування концептів, проводимо їх ранжування на кожній сторінці, список анкорів якої містить концепти предметної області. При цьому рангові номери виставляємо лише концептам. Так, R_{ik} позначає номер (ранг) k -го концепту на i -тій сторінці.

Узгодженість рангів оцінюється за коефіцієнтом конкордації Кендалла:

$$W = \frac{12 \sum_{k=1}^K \left(\sum_{i=1}^I R_{ik} - \bar{R} \right)}{I^2 (K^3 - K)}, \quad (6)$$

де $\bar{R} = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^I R_{ik}$. За умови, що $I(K-1)W > \chi_{K-1, \alpha}^2$, ранжування вважається значущим. Якщо ранжування по повному списку концептів не можна вважати значущим, то усуваємо їх по одному із списку в порядку зростання їхніх частот аж до отримання значущого ранжування. У такому разі відібрану множину концептів розбиваємо на дві частини: множину RG із погодженими рангами та NR , ранжування по якій встановити не вдалося.

Типова структура тематичного Веб-сайту спрощує вибір тематики, цікавої для користувача. Однак привабити користувача може тільки актуальна для нього інформація. Із врахуванням різного рівня інформованості користувачів Веб-сайту, вона повинна бути також і повною, спираючись на концептуалізацію даної предметної області, тобто на її онтологію. Відомі загальні онтології (наприклад, WordNet) рідко описують актуальні предметні області. Отримати актуальні

соціально-значимі характеристики предметних областей, що динамічно розвиваються, можна за допомогою аналізу інформації, представленої у Вебі. Цю інформацію отримують із запитів споживачів продукції певної предметної області, які розташовуються на дошках оголошень. Згадані запити містять слабо структуровану та неструктуровану інформацію. Тому автором дисертації розроблено метод, що передбачає побудову ряду відношень із наступними структурами. Це деревоподібна AS структура зберігання основ концептів, отримана з аналізу Веб-сторінок, яка дає змогу зберігати багатоконпонентні концепти предметної області :

$$AS = \langle IdPg, IdLst, IdIt, IdBs, IdFm, IdPrBs \rangle, \quad (7)$$

де $IdPg$ – ідентифікатор аналізованої Веб-сторінки, $IdLst$ – ідентифікатор списку в тексті сторінки, $IdIt$ – ідентифікатор елемента списку, $IdBs$ – ідентифікатор основи слова, $IdFm$ – ідентифікатор форми слова, $IdPrBs$ – ідентифікатор батьківської основи концепту.

У відношенні BF основи впорядковуються по частотах $BsFr$ їх появи на різних сторінках:

$$BF = \langle IdBs, BsFr, IdLPg, Phn \rangle, \quad (8)$$

де $IdLPg$ – ідентифікатор останньої із Веб-сторінок, де зустрічалася основа, Phn – маркер фоновості базової основи поняття.

Щоб зробити експертний вибір соціально-значимим та мінімізувати суб'єктивізм оцінювання, автором роботи використовується механізм автоматизованого анування понять, який підтримується інформаційною структурою $ConcAnot$:

$$ConcAnot = \langle IdCn, AnURL, AnTitle, Anot, Desc \rangle, \quad (9)$$

де $IdCn$ – ідентифікатор концепта, $AnURL$ – адреса Веб-сторінки анотації, $AnTitle$ – тег *title* Веб-сторінки анотації, $Desc$ – опис концепта.

Відібрані концепти структуруються у відношеннях деревоподібної онтологічної структури O_Str та фразових описів онтологічних концептів O_Phr :

$$O_Str = \langle IdCn, PrCn, Meta \rangle, \quad (10)$$

$$O_Phr = \langle IdCn, IdLg, IdPh, IdBs, IdFm, IdPrBs \rangle, \quad (11)$$

де $PrCn$ – ідентифікатор батьківського концепта, $Meta$ – атрибут, який дозволяє відділити робочі поняття предметної області від її мета-понять, $IdLg$ – ідентифікатор мови реалізації, $IdPh$ – ідентифікатор фрази.

Метод передбачає встановлення початкового значення частоти основи при її реєстрації (12), а також збільшення частоти при появі основи на новій сторінці (13) :

$$\begin{aligned} & (Count(\pi_{IdBs}(\sigma_{IdBs=IdBsWrd}(BF))) = 0) \Rightarrow \\ & \Rightarrow (BF.IdBs := IdBsWrd) \wedge (BF.BsFr := 1) \wedge (BF.IdLPg := CurPgId), \quad (12) \end{aligned}$$

$$\begin{aligned} & (\text{Count}(\pi_{IdBs}(\sigma_{IdBs=IdBsWrd \text{ AND } IdLPg \langle \rangle CurPgId}(BF))) \neq 0) \Rightarrow \\ & \Rightarrow (BF.BsFr := BF.BsFr + 1) \wedge (BF.IdLPg := CurPgId), \end{aligned} \quad (13)$$

де $IdBsWrd = \pi_{IdBs}(\sigma_{WBase=BsC(Wrd_k(It_{Pg,Lst}))}(O_Bs))$ – ідентифікатор основи терміна, виділеного із списку, $Wrd_k(It_{Pg,Lst})$ – k -те слово, виділене із елемента списку $It_{Pg,Lst}$, $BsC()$ – процедура побудови основ слів.

Основи $CBase = rand(\pi_{IdBs}(\sigma_{BsFr=MFR}))$ із максимальними частотами $MFR = max(\pi_{BsFr}(\sigma_{Phn=NULL}))$, якщо вони ще не включені в онтологію або не відкинуті на попередніх кроках, разом із контекстами:

$$\begin{aligned} ACont[i] = & (\pi_{IdFm}(\sigma_{IdIt=AphN[i] \wedge IdPrBs=NULL}(AS)) AS HL) \cup \\ & \cup (\pi_{IdFm}(\sigma_{IdIt=AphN[i] \wedge (IdPrBs=HL.IdBsORIdPrBs=HL.IdBs)}(AS)) AS HL1) \end{aligned} \quad (14)$$

пропонуються експерту для включення в онтологію. При цьому критерієм завантаженості експерта:

$$EE = \frac{NOI}{NSE}, \quad (15)$$

приймаємо відношення кількості відібраних термінів NOI до кількості NSE переглянутих для цього стрічок тексту. За критерієм результативності відбору важливої інформації ефективність роботи експерта із використанням цього методу зростає у середньому в 9 разів.

У **третьому розділі** розроблено методи ідентифікації моделі динаміки приростів відвідуваності Веб-сайту із врахуванням бізнес-процесів його об'єкта, а також моделі динаміки приростів відвідуваності Веб-сайту під дією коротко- та довготермінових факторів впливу.

Відвідуваність тематичних сторінок Веб-сайту сприяє високому рейтингу у відповідних тематичних доменах. Це спричиняє підвищення відвідуваності інших тематик Веб-сайту в цілому. При цьому відвідуваностям тематик властива суттєво простіша динаміка, що дає змогу використовувати їх в якості факторів впливу для моделювання загальної відвідуваності.

Динаміка відвідуваностей тематик залежить від інтенсивності їх поповнення. Тому для численних Веб-сайтів, для яких таке оновлення здійснюється епізодично, в рамках моделі динаміки приростів відвідуваності Веб-сайту із врахуванням бізнес-процесів його об'єкта, динаміка відвідуваностей з достатньої точністю може бути описана за допомогою множини дискретних станів $D = d_1, \dots, d_p$ невеликого обсягу p , яким відповідають певні усереднені рівні відвідуваностей тематик, що оцінюються як "високий", "значний", "середній", "незначний", "малий", "низький". Зміна цих станів описується певною послідовністю V таких послідовностей відвідуваності переходять із високих значень на момент активізації фактора до помірних та низьких, причому активність фактору може періодично поновлюватися. При поновленні активності фактора спостерігається нелінійність у співвідношеннях між факторами впливу та загальними і результируючими відвідуваностями. Ця

нелінійність моделюється за допомогою врахування відвідуваностей у попередні моменти часу, а також степеневій трансформації фактора впливу. Однак апробація такого підходу показала обмеженість сфери його застосування. При інтенсифікації процесу поповнення тематик Веб-сайту та рекламування його ресурсів кількість станів швидко зростає, що вимагає застосування моделей динаміки самих факторів впливу.

Щоденна відвідуваність формується під дією великої кількості випадкових факторів, що робить проблематичним достатньо точне її прогнозування. В той же час на характеристики Веб-сайту впливають не окремі екстремальні, а усереднені показники. Тому в розробленій автором дисертації моделі аналізувалися середньотижневі згладжені загальні відвідуваності Y та відвідуваності субсайтів X . При цьому використано кратне згладжування методом ковзаючого середнього із мінімальним шаблоном, яке, на відміну від інших відомих методів, суттєво нівелює аномальні викиди, а не лише частково згладжує їх.

Помічено, що для згладженої відвідуваності характерні інтервали відносної стабільності, а також інтервали помітного її росту. Для формального виділення інтервалів активності росту відвідуваності використовуємо порогове значення D_{min} модуля похідної:

$$Y'(t) > D_{min}. \quad (16)$$

На основі максимальних відвідуваностей тематик Веб-сайту, що активно розвиваються, визначаємо функцію якості контенту X для даного інтервалу. Згладжена динаміка усередненої функції якості контенту має достатньо регулярний характер, що дає можливість застосувати для її прогнозування диференціальні рівняння. Динаміка зміни концентрації мікроорганізмів нагадує динаміку функції якості контенту та самої відвідуваності. Це дозволяє використати систему Моно для моделювання функції якості контенту із степеневою трансформацією її у показник відвідуваності. Модель відвідуваності Веб-сайту включає модель динаміки системи на основі модифікованої системи Моно:

$$\begin{cases} X'(t) = \left(a_1 \frac{A_e(t)}{a_4 + A_e(t)} - a_2 \right) X(t), \\ A'(t) = -a_3 \frac{A_e(t)}{a_4 + A_e(t)} X(t), \end{cases} \quad (17)$$

при умові $\left(A > \frac{A_0}{2} \right) \vee (X \geq d_{X0})$:

$$\begin{cases} X'(t) = 0, \\ A'(t) = 0, \end{cases} \quad (18)$$

при умові $\left(A \leq \frac{A_0}{2} \right) \wedge (X < d_{X0})$:

$$Y(t) = Y(t_0) + q_1(X(t) - X_0)^{q_2} e^{-q_3 t}, \quad (19)$$

$$X(0) = X_0, \quad (20)$$

$$A(0) = A_0 = X_{max}. \quad (21)$$

Модель функції якості повинна бути адаптивною по параметрах, тобто параметри шуканої функції повинні уточнюватися в процесі спостереження за процесом. Для цього період активності функції якості розділяємо на етапи, на кожному із яких коефіцієнти моделі уточнюються на основі початкових спостережених значень етапу. Етапи виділяються за критерієм порогового рівня відносної похибки моделі попереднього етапу.

Модель функції якості містить чотири невідомих параметри і тому вимагає хоча б чотирьох точок ідентифікації. Дві точки отримуємо під час встановлення початку інтервалу суттєвого приросту відвідуваності, оскільки при цьому необхідно оцінювати значення похідної відвідуваності за її різницевою апроксимацією. До отримання ще 2-4 точок спостереження функції якості необхідно їх прогнозувати на основі апріорних оцінок.

Це дає можливість виділити функції якості, які містять не більше чотирьох спостережень до досягнення максимального значення, в клас короткотермінових. Всі інші прояви функцій якості відносимо в клас довготермінових. Класифікувати тип функцій якості для конкретного Веб-сайту можна, проаналізувавши динаміку її першого прояву. Наступні прояви будуть того ж типу, протягом періоду, поки не відбудеться якісних змін в діяльності по розробці та підтримці Веб-сайту або зовнішнього середовища.

Для ідентифікації параметрів моделі функції якості та відвідуваності використовуємо середньоквадратичний критерій, а також градієнтний метод оптимізації Левенберга Марквардта. Ключовим питанням успішної ідентифікації служить задовільне початкове наближення параметрів моделі та наявність достатньої кількості точок ідентифікації.

Специфіка короткотермінових функцій якості полягає у їх суттєвій нелінійності, яка проявляється вже із перших спостережених значень. Експериментальні дослідження показують, що друга похідна q короткотермінової функції якості на початковому етапі має наближено лінійний характер $q_i^0 = b_1 + b_2 i$. Помічена властивість дає можливість побудувати рекурентні співвідношення $\tilde{x}_i = \tilde{x}_{i-1} + p_i$, $p_i = p_{i-1} + q_{i-1}$, $q_i = q_{i-1} + b_2$. Обчислення за рекурентними співвідношеннями продовжуємо поки відносна похибка прогнозу буде лежати в допустимих межах δ_0 .

Побудовані співвідношення дозволяють прогнозувати динаміку функції якості на початковому етапі та сприяють отриманню інформації про дійсні її значення в обсягах, достатніх для застосування процедури ідентифікації моделі функції якості. При побудові початкових значень для параметрів моделі функції якості встановлено, що коефіцієнти a_2 і a_4 є відносно незалежними. Тому їх значення підбираємо на вузлах рівномірних сіток, які покривають деякі діапазони:

$$a_4 \in W_4 = \{a_4^0, a_4^0 + h_4, a_4^0 + 2h_4, \dots, a_4^N\}, \quad (22)$$

$$a_2 \in W_2 = \{a_2^0, a_2^0 + h_2, a_2^0 + 2h_2, \dots, a_2^N\}. \quad (23)$$

Після вибору значень коефіцієнтів a_2 та a_4 із першого рівняння системи (17) визначаємо початкове значення коефіцієнта a_1 відносного значень функції якості:

$$a_1 = \left(\frac{p_0}{x_1} + a_2 \right) \frac{A_0}{A_0 + a_4} = \left(\frac{p_0}{x_1} + a_2 \right) \frac{X_{max}}{X_{max} + a_4}. \quad (24)$$

На протязі певного часового інтервалу відвідуваність досягає свого максимуму. За цей же проміжок часу потенційна аудиторія зменшиться приблизно на половину. Тому природно покласти коефіцієнт відносного зменшення потенційної аудиторії вдвічі меншим, ніж коефіцієнт відносного приросту відвідуваності: $a_3 = a_1 / 2$.

Для побудови початкових значень коефіцієнтів моделі відвідуваності прологарифмуємо співвідношення (21), отримуючи систему лінійних рівнянь:

$$\ln(q_1) + q_2 \ln(X(t_i) - X_0) - q_3 = \ln(Y(t_i) - Y(t_0)), \quad i=1,2,3. \quad (25)$$

Ідентифікація системи Моно, а також моделі відвідуваності, здійснюється на основі мінімізації середньоквадратичних критеріїв якості:

$$\bar{a} = \arg \min_c \sum_{j=1}^I (\tilde{x}(\bar{c}, t_j) - x_j)^2, \quad \bar{q} = \arg \min_r \sum_{j=1}^I (\tilde{y}(\bar{r}, t_j) - y_j)^2. \quad (26)$$

Короткотермінові фактори відвідуваності характерні для Веб-сайтів, які не є життєво важливими для об'єкта, який він представляє. В іншому випадку команда, що підтримує відповідний Веб-сайт, докладає максимальних зусиль до постійного росту його відвідуваності аж до виходу на рівень, який забезпечує його належну результативність. Для моделювання приросту відвідуваності цього типу використаємо раніше запропоновану модель (17)-(21), однак ідентифікація цієї моделі в даному випадку буде мати свої особливості.

На першому етапі ідентифікації моделі для підвищення її точності штучно збільшуємо кількість точок ідентифікації за допомогою лінійної інтерполяції функції якості по двох останніх точках спостереження:

$$\bar{a} = \arg \min_{\bar{c}} \left\{ \begin{array}{l} \sum_{k=1}^m (\tilde{x}(\bar{c}, t_k) - x_k)^2 + \\ + \sum_{k=m-1}^{m+m_1} (x_{m-1} + (x_m - x_{m-1})(k - m + 1) - \tilde{x}(\bar{c}, t_k))^2 \end{array} \right\}, \quad (27)$$

де m – кількість спостережених точок, m_1 – кількість додаткових точок ідентифікації. Ідентифікація за лінійним наближенням P_m її похідної використовується після завершення інтервалу лінійного росту відвідуваності:

$$\vec{a} = \arg \min_{\vec{c}} \left\{ \begin{array}{l} \sum_{k=1}^{m_2} (\tilde{x}(\vec{c}, t_k) - x_k)^2 + \\ + \sum_{k=m-3}^{l+m_1} (p_{m-3}(m-2-k) + p_{m-2}(k-m+3) - \tilde{x}'(\vec{c}, t_k))^2 \end{array} \right\}. \quad (28)$$

Таким чином розроблена та обґрунтована в даному розділі модель відвідуваності Веб-сайтів дозволила встановити причинно-наслідковий зв'язок між відвідуваністю Веб-сайту та окремих його елементів, а запропоновані та обґрунтовані методи ідентифікації математичних моделей відвідуваності – врахувати неоднорідність динаміки функції якості для прогнозування коротко- та довготермінової відвідуваності під її впливом.

У **четвертому розділі** описано розроблений та практично реалізований програмний комплекс підвищення відвідуваності Веб-сайтів. Цей комплекс складається із трьох основних підсистем: оцінка типової структури тематичного Веб-сайту, формування шляхів наповнення контенту Веб-сайту за допомогою створення онтологій предметних областей на основі дощок оголошень та прогнозування відвідуваності Веб-сайтів. При програмній реалізації підсистем використано HTML та CSS кодування, а також програмні одиниці, згенеровані на алгоритмічних мовах JavaScript та PHP. Для програмної реалізації підсистеми моделювання відвідуваності Веб-сайту використано сервіс Google Analytics та програмне середовище MatLab.

Для побудови типової структури тематичного Веб-сайту в першу чергу здійснюється пошук самих Веб-сайтів певної тематичної спрямованості за допомогою запитів відповідного програмного агента до пошукової системи, аналізу результатів запиту і запису отриманих адрес у відповідний архів.

На наступному етапі відбувається звертання до відповідних Веб-сайтів із архівуванням кодів головних сторінок. Після цього HTML код сторінок аналізується, з нього виділяються пункти меню із яких методом частотного аналізу виділяються базові поняття. Далі за допомогою рангування визначається порядок пунктів меню та перевіряється узгодженість побудованих оцінок. У результаті підсистема будує типове меню тематичного Веб-сайту та множину невпорядкованих понять, що часто зустрічаються в типових меню. Результат встановлення типової структури Веб-сайту факультету українського вузу за критерієм наявності в заголовку сторінки ключового терміну “*факультет*” наведено у таблиці. Побудова коефіцієнта конкордації по повному набору параметрів показала незначущість такого ранжування $W = 0.05$, $\chi_{emp}^2 = 6.31$, $\chi_{10,0.05}^2 = 18.31$. В процесі послідовного редукування списку концептів значущі значення коефіцієнта конкордації отримано для сукупності із 9 елементів. При цьому отримано наступні значення $W = 0.18$, $\chi_{emp}^2 = 18.60$, $\chi_{8,0.05}^2 = 15.51$. Оскільки емпіричне значення критерію χ^2 перевищує критичне значення, то гіпотезу про значимість отриманого впорядкування концептів слід прийняти із рівнем значимості в 5%. При цьому впорядкований список концептів склав $RG = \langle \text{головна, кафедри, факультет, наука, абітурієнту, студенту, форум, контакти, бібліотека} \rangle$, а невпорядкована множина

отримала наступне представлення $NR = \{ \text{новини, фотогалерея} \}$.

Таблиця

Результати встановлення типової структури Веб-сайту факультету українського вузу

Концепти	Порядкові ранги	Ранги частот	Матсподівання рангів
Головна	1	1	1
Кафедри	2	3	3
Факультет	3	4	3.22
Наука	4	6	3.63
Новини	5	11	3.8
Абітурієнту	6	2	4.55
Студенту	7	5	4.71
Форум	8	9	4.8
Контакти	9	8	5
Фотогалерея	10	10	6
Бібліотека	11	7	6.4

Робота підсистеми формування шляхів наповнення контенту Веб-сайту за допомогою видобування інформації з дощок оголошень розпочинається із вибору тематики дошки оголошень та вибору переліку адрес Веб-сторінок цієї тематики і запису їх в архів. В подальшому за цими адресами HTML-коди сторінок записуються в архів. На основі аналізу збережених HTML-кодів вибираються елементи тих списків, які мають безпосередній стосунок до вибраної тематики. За допомогою додаткового аналізу із елементів списків виділяються базові концепти предметної області із відповідним контекстом.

На основі частотного аналізу вибираються найтипівіші елементи концептів. Разом із відповідними контекстами вони подаються для експертної оцінки. Аналізуючи подані контексти, експерт вибирає базові поняття для включення в онтологію. Для впорядкування понять засобами пошукових систем подаються анотації вибраних понять. Приклад сформованої онтології наведено на рис. 1.

Для побудови онтології, значимої для софтверних українських компаній, вибрано множину сайтів, що спеціалізуються на пропозиціях вакантних посад на підприємствах України, зокрема "rabota.ua", "jobs.ua", "work.ua" і містять спеціальні розділи вакансій у сфері ІТ. Серед цих сайтів для проведення експериментів з другою підсистемою програмного комплексу вибрано сайт "rabota.ua" та множину ключових слів, яка складається з єдиного елемента $KWS = \{ "PHP" \}$. За критерієм результативності відбору інформації ефективність роботи експерта (15) зросла від значення 0.045 до 0.391, тобто майже на порядок.

Для апробації третьої підсистеми автором дисертаційної роботи досліджено ефективність запропонованих методів прогнозування приростів відвідуваності Веб-сайту факультету комп'ютерних інформаційних технологій (ФКІТ- <http://tanet.tneu.org/>) Тернопільського національного економічного університету (ТНЕУ) майже за річний період. Аналіз динаміки першого із виділених проявів функції якості контенту свідчить, що свій максимум він досягає у третій точці від моменту фіксації початку суттєвого росту відвідуваності. Це дає можливість віднести такий ріст відвідуваності до короткотермінових. Встановлено, що всі наступні прирости

відвідуваності для цього Веб-сайту теж відносяться до такого класу. Результати моделювання наведено на рис. 2. При цьому похибки моделювання склали 13.6% по фактору впливу і 11% по результуючій відвідуваності.

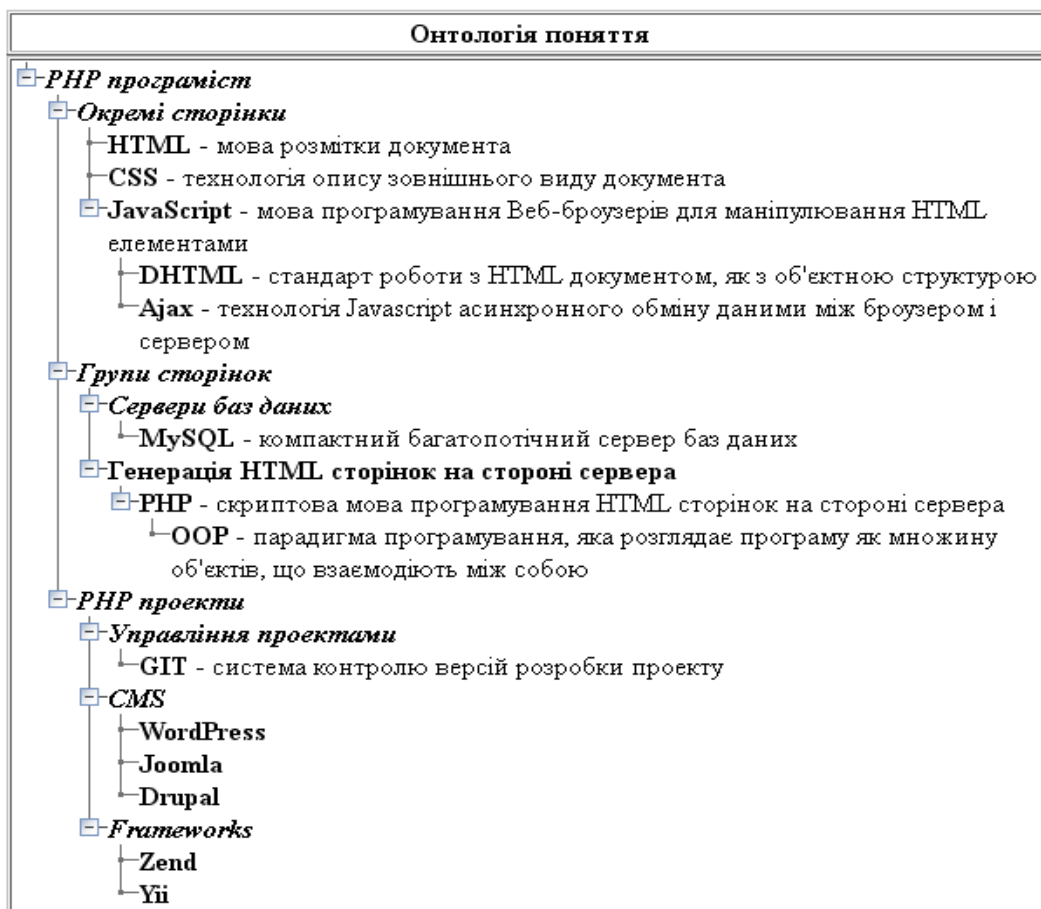
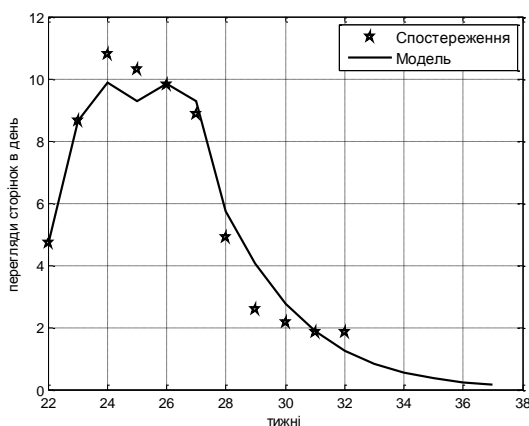
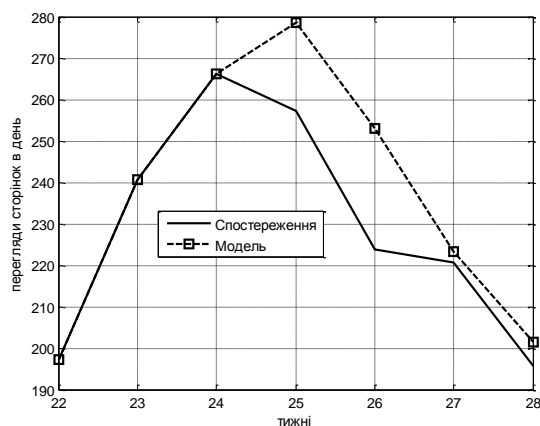


Рис. 1. – Онтологія, що використовується для побудови шляхів наповнення Веб-сайту навчального підрозділу



а)

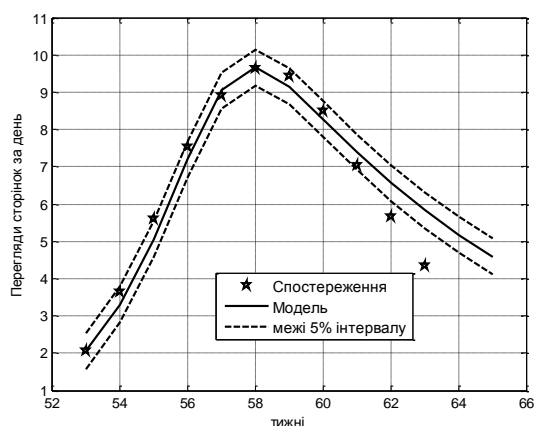


б)

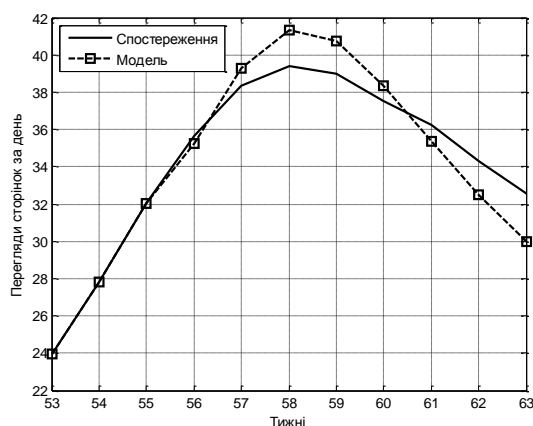
Рис. 2 – Результат ідентифікації моделі: а) фактора впливу та б) загальної відвідуваності Веб-сайту ФКІТ

Довготерміновими функціями якості характеризуються Веб-сайти, підвищенням відвідуваності яких займаються команди служб підтримки на регулярній основі. В роботі, як приклад таких сайтів, проаналізовано відвідуваність

промо-сайту сервісу API2Cart компанії MagneticOne (<http://www.api2cart.com/>) більш ніж за річний період. Результати моделювання наведено на рис. 3.



а)



б)

Рис. 3 – Результат ідентифікації моделі: а) фактора впливу та б) загальної відвідуваності Веб-сайту API2Cart

При цьому похибки моделі склали 13.6% по фактору та 10% по результуючій відвідуваності. Тобто багатоетапне прогнозування фактору впливу дає можливість будувати його прогноз із задовільною точністю.

ВИСНОВКИ

В дисертаційній роботі розв'язано актуальне наукове завдання підвищення відвідуваності Веб-сайту за рахунок розробки процедур структурування контенту з використанням програмних засобів та інструментів математичного моделювання.

При цьому отримано наведені нижче наукові та практичні результати.

1. Обґрунтовано використання методів Веб-майнінгу та математичних моделей відвідуваності Веб-сайтів в автоматизації процесів підвищення їх відвідуваності.

2. Розроблено метод побудови структур тематичних Веб-сайтів на основі частотного аналізу та формалізованого структурування Веб-онтологій, який, на відміну від відомих, дає можливість генерувати типові структури тематичних Веб-сайтів, узгоджених із Веб-сайтами відповідної тематики та автоматизувати структурування напрямків поповнення контенту Веб-сайту. За критерієм результативності відбору важливої інформації ефективність роботи експерта із використанням такого методу зростає у середньому в 9 разів.

3. На основі згладжування агрегованих відвідуваностей Веб-сайту та окремих його елементів і застосування системи звичайних диференціальних рівнянь Моно побудовано модель відвідуваності Веб-сайтів, що дозволило встановити причинно-наслідковий зв'язок між ними. Для формування математичних моделей короткотермінових прогнозів відвідуваності запропоновано метод ідентифікації, який включає етапи побудови апріорних оцінок та початкових значень коефіцієнтів моделі із подальшим їх уточненням методом Левенберга-Марквардта, що дало можливість врахувати стрімку динаміку короткотермінового фактора впливу.

4. Для довготермінових прогнозів відвідуваності запропоновано та

обґрунтовано метод ідентифікації, який включає етапи побудови початкових значень коефіцієнтів моделі та її ідентифікації за критеріями, що містять лінійні екстраполяції значень функції якості оновлення контенту та його першої похідної. Це дає можливість врахувати неоднорідність динаміки функції якості для прогнозування довготермінової відвідуваності під її впливом.

5. З метою автоматизації процедур оцінювання ефективності структурування Веб-сайтів розроблено комплекс прикладних програм із застосуванням технологій PHP, JavaScript та MatLab. Ефективність сформованих прогнозних моделей відвідуваності Веб-сайту навчальних підрозділів та промо сайтів програмних продуктів оцінена експериментально. Зокрема, максимальна відносна похибка при прогнозуванні короткотермінового росту відвідуваності склала 12%, а при прогнозуванні довготермінового росту – 7%. При цьому ширина прогнозного коридору складала від 20 до 35 днів, що дає менеджменту можливість заздалегідь коригувати ефективність заходів із підвищення відвідуваності Веб-сайту. Застосування процедур структурування контенту дало можливість збільшувати відвідуваність у середньому на 28% за період її суттєвого росту.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Pasichnyk N. Mathematical modeling of the Website quality characteristics in dynamics / Pasichnyk N., Dyvak M., Pasichnyk R. // Journal of Applied Computer Science / Technical University Press – Lodz, Poland. – 2014. – V. 22, № 1. – P. 171-183.

2. Пасічник Н. Математична модель динаміки відвідуваності тематичних веб-сайтів та методи її ідентифікації / Пасічник Н., Дивак М., Пасічник Р. // Індуктивне моделювання складних систем. Збірник наукових праць // Відп. редактор В. С. Степашко. – Київ : Міжнар. наук.-навч. центр інформ. технологій та систем НАН та МОН України, 2013. – Вип. №5. – С. 236-246.

3. Пасічник Н. Метод формування онтологічного наповнення на основі аналізу зашумленої слабо структурованої інформації спеціалізованих Веб-сайтів / Пасічник Н., Дивак М. // Індуктивне моделювання складних систем. Збірник наукових праць // Відп. редактор В. С. Степашко. – Київ : Міжнар. наук.-навч. центр інформ. технологій та систем НАН та МОН України, 2012. – Вип. №4. – С. 158-168.

4. Пасічник Н. Метод та алгоритм побудови структур тематичних Веб-сайтів на основі онтологічного підходу / Пасічник Н., Дивак М. // Наукові праці Донецького національного технічного університету, серія «Інформатика, кібернетика та обчислювальна техніка». – 2012. – Вип. 15 (203). – С. 184-189.

5. Пасічник Н. Метод формування онтологічного контенту на основі аналізу інформації спеціалізованих Веб-сайтів / Пасічник Н. // Вісник Хмельницького національного університету : Технічні науки. – 2012. – Вип. № 5. – С. 241-244.

6. Пасічник Н. Інформаційна технологія формування аналітичних ресурсів сайту / Пасічник Н. // Системи обробки інформації : тематичний збірник наукових праць. – Харків : ХУПС, 2011. – Вип. 3 (93). – С. 224-227.

7. Пасічник Н. Формалізм в постановці задачі створення якісного сайту. / Пасічник Н., Дивак М. // Наукові праці Донецького національного технічного університету, серія «Інформатика, кібернетика та обчислювальна техніка». –

Донецьк, 2011. – Вип. 14 (188). – С. 325-329.

8. Pasichnyk N. Parametric identification methods of the model of Websites' attendance / N. Pasichnyk, A. Melnyk, L. Honchar // Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), 25 Feb.-1 Mar. 2014. – Lviv-Slavsko, 2014. – P. 93-94.

9. Pasichnyk N. Mathematical model of the Website's Pages Visibility Dynamics / N. Pasichnyk, M. Dyvak // Proceedings of the 4th International Conference on Inductive Modelling (ICIM'2013), 16-20 Sept. 2013. – Kyiv, 2013. – P. 298-301.

10. Pasichnyk N. Management the Website Attendance Based on the Projected Traffic / N. Pasichnyk, A. Melnyk, N. Dobrovolska // Proceedings of the international conference "The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)", 19-23 Feb. 2013. – Polyana-Svalyava (Zakarpattya), 2013. – P. 277.

11. Пасічник Н. Р. Дискретна динамічна модель для оцінювання та прогнозування відвідуваності Веб-сайту / Н. Р. Пасічник // Матеріали школи-семінару молодих вчених і студентів «Сучасні комп'ютерні інформаційні технології (АСІТ-2013)», 17-18 травня 2013 р. – Тернопіль, 2013. – С. 49-50.

12. Pasichnyk N. Formalized analysis of the Web-site structure / N. Pasichnyk, R. Shevchuk, I. Spivak // Conference on Modern Problems of Radio Engineering Telecommunications and Computer Science (TCSET), 21-24 Feb. 2012. – Lviv-Slavsko, 2012. – P. 391.

13. Пасічник Н. Р. Метод структурування тематичних Веб-сайтів на основі направленою інформаційного пошуку / Н. Р. Пасічник // Матеріали школи-семінару молодих вчених і студентів «Сучасні комп'ютерні інформаційні технології (АСІТ-2012)», 4-5 травня 2012 р. – Тернопіль, 2012. – С. 47-49.

14. Пасічник Н. Р. Система підтримки інформаційного наповнення сайту / Н. Р. Пасічник // Матеріали школи-семінару молодих вчених і студентів «Сучасні комп'ютерні інформаційні технології (АСІТ-2011)», 20-21 травня 2011 р. – Тернопіль, 2011. – С. 130.

15. Пасічник Н. Р. Основи автоматизації формування аналітичних ресурсів сайту / Н. Р. Пасічник // Тези доповідей ІІ МНПК «Інформаційні технології та комп'ютерна інженерія». 26-27 травня 2011 р. – Харків, 2011. – С. 126-127.

16. Pasichnyk N. Modeling of problematic aspects in the Web-site development process / N. Pasichnyk // Modern Problems of Radio Engineering, Telecommunications and Computer Science. Proceedings of the International Conference. TCSET'2010. – Lviv-Slavsko, 2010. – P. 207.

17. Pasichnyk R. Method of adaptive control structure learning based on model of test's complexity / R. Pasichnyk, A. Melnyk, N. Pasichnyk, I. Turchenko // Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications : Proceedings of the 6th IEEE International Conference IDAACS' 2011, Volume 2 (Prague, Czech Republic, 15–17 September 2011) / IEEE Ukraine I&M / CI Joint Societies chapter, Research Institute for Intelligent Computer Systems, 2011. – P. 692–695.

АНОТАЦІЇ

Пасічник Н. Р. Математичні моделі відвідуваності Веб-сайтів та методи їх ідентифікації. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.03 – математичне і програмне забезпечення обчислювальних машин та систем. – Національний університет "Львівська політехніка" Міністерства освіти і науки України, Львів, 2014.

Дисертація присвячена вирішенню актуального наукового завдання – підвищення відвідуваності Веб-сайту за рахунок розробки процедур структурування контенту з використанням програмних засобів та інструментів математичного моделювання. В ході дослідження було виявлено недоліки сучасних моделей відвідуваності Веб-сайтів, розроблено концептуальну модель якості Веб-сайтів.

Автором розроблено методи побудови структур тематичних Веб-сайтів, які забезпечують узгодженість структурних елементів Веб-сайту із предметною областю та автоматизацію побудови шляхів оновлення його контенту. Запропоновано та обґрунтовано нові методи ідентифікації математичних моделей коротко- та довготермінового прогнозу відвідуваності Веб-сайтів, що забезпечує адекватне відображення динаміки відвідуваності і підвищення якості структурування контенту. Удосконалено автономні моделі довго- та короткотермінової відвідуваності Веб-сайтів, що забезпечило адекватність відображення процесу зміни відвідуваності Веб-сайту.

Розроблені методи та засоби практично реалізовано у програмному комплексі планування діяльності із підвищення відвідуваності Веб-сайтів. Розроблені методи та засоби впроваджено на приватному підприємстві «Magnetic One», а також в Тернопільському національному економічному університеті.

Ключові слова: моделі відвідуваності, методи ідентифікації, Веб майнінг, Веб онтології, управління якістю структурування контенту.

Пасичник Н. Р. Математические модели посещаемости Веб-сайтов и методы их идентификации. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 01.05.03 – математическое и программное обеспечение вычислительных машин и систем. – Национальный университет "Львовская политехника" Министерства образования и науки Украины, Львов, 2014.

Диссертация посвящена решению актуального научного задания – повышения посещаемости Веб-сайтов за счет разработки процедур структурирования контента с использованием программных средств и инструментария математического моделирования. В ходе исследования были выявлены недостатки современных моделей структурирования контента, разработано концептуальную модель качества Веб-сайтов.

Автором разработаны методы построения структур тематических Веб-сайтов, которые обеспечивают анализ согласованности структурных элементов Веб-сайта с предметной областью а также автоматизацию создания путей обновления его контента. Предложены и обоснованы новые методы идентификации математических

моделей краткосрочного и долгосрочного прогноза посещаемости Веб-сайтов, который, что обеспечивает адекватное представление динамики посещаемости и повышения качества структурирования контента. Усовершенствованы автономные модели долго и коротко срочной посещаемости Веб-сайтов, что обеспечило адекватность отражения процесса изменения посещаемости Веб-сайта.

Разработаны методы и средства практически реализованы в программном комплексе планирования деятельности по повышению посещаемости Веб-сайтов. Разработанные методы и средства внедрены на частном предприятии «Magnetic One», а также в Тернопольском национальном экономическом университете.

Ключевые слова: модели посещаемости, методы идентификации, Веб майнинг, Веб онтологии, управления качеством структурирования контента.

Pasichnyk N. R. Mathematical models of Websites attendance and methods for their identification. – Manuscript.

Thesis for a Ph.D. degree of engineering science in specialty 01.05.03 - mathematical and software support of computer machines and systems. Lviv Polytechnic National University, Ministry of Education and Science of Ukraine, Lviv, 2014.

The thesis is dedicated to solving important scientific and technical problem which consists in increasing Website attendance by developing the procedures for structuring the Website content, with the help of software practices and tools of mathematical modelling. The study revealed the shortcomings of current models of Website attendance and a conceptual model of quality websites was developed.

Author introduced methods for construction of the thematic Website content that are based on frequency analysis and ranking, which, unlike existing methods, use analysis of the coherence of structural elements of a Website and that of the selected thematic domain; consider the structure of the basic concepts of expert queries. It provided for automatic construction of the typical content structure for a Website and automation of creating content update ways. A method to identify a mathematical model of short-term estimates of Websites' traffic, which, unlike existing ones, considers the second derivative of the quality function of Website content as well as the procedure for constructing the initial approximation of the model parameters, providing an adequate reflection of short-term dynamics of Website traffic and of improvements in the quality of structured content. A method of identifying a model of long-term estimates of Web sites traffic was constructed. Unlike existing ones, it considers linear extrapolation of the content quality function and of its derivative, which provides adequate reflection of the long-term dynamics of traffic and of improvements in the content structure quality. Autonomous models of long- and short-term Website traffic dynamics were improved basing on a modified system of ordinary differential equations by Mono. It takes into account the impact of the Website content quality functions, ensuring adequate reflection of Website traffic.

The methods and means were practically implemented in software complex for planning activities to increase Websites' traffic. The methods and tools were applied by private enterprise «Magnetic One» and Ternopil National Economic University.

Keywords: attendance models, methods of identification, Web Mining, Web ontology, quality management of content structuring.