

Тернопільський національний економічний університет

На правах рукопису

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

Дисертація на здобуття наукового ступеня кандидата
технічних наук

Науковий керівник:

Дивак Микола Петрович,
доктор технічних наук, професор

Тернопіль – 2014

ЗМІСТ

| | |
|---|----|
| ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ | 4 |
| ВСТУП | 5 |
| РОЗДІЛ 1. ОСОБЛИВОСТІ МЕТОДІВ ТА ЗАСОБІВ ПІДТРИМКИ ВІДВІДУВАНOSTІ ВЕБ-САЙТІВ..... | 11 |
| 1.1. Особливості забезпечення відвідуваності Веб-сайтів та фактори впливу на неї | 11 |
| 1.2. Засоби підтримки відвідуваності Веб-сайту..... | 19 |
| 1.3 Особливості математичного моделювання процесів відвідуваності Веб-сайтів | 28 |
| 1.4. Постановка задачі дослідження | 35 |
| Висновки до розділу 1 | 39 |
| РОЗДІЛ 2. МОДЕЛІ ТА МЕТОДИ АВТОМАТИЗОВАНОЇ ПОБУДОВИ СТРУКТУРИ ВЕБ-САЙТІВ | 41 |
| 2.1. Обґрунтування методів, засобів та математичних моделей відвідуваності Веб-сайтів..... | 41 |
| 2.2. Метод побудови структур тематичних Веб-сайтів на основі частотного аналізу існуючих структур..... | 52 |
| 2.3. Моделі контенту Веб-сайту на основі слабо структурованої інформації | 56 |
| Висновки до розділу 2..... | 63 |
| РОЗДІЛ 3. МЕТОДИ ІДЕНТИФІКАЦІЇ МАТЕМАТИЧНИХ МОДЕЛЕЙ ПРОГНОЗУ ВІДВІДУВАНOSTІ ВЕБ-САЙТІВ. | 65 |
| 3.1. Обґрунтування вибору структури математичної моделі відвідуваності Веб-сайтів..... | 65 |
| 3.1.1. Модель відвідуваності із дискретною функцією якості контенту | 65 |
| 3.1.2. Модель відвідуваності із функцією якості контенту, побудованою на основі системи Моно | 78 |
| 3.2. Метод ідентифікації математичних моделей короткотермінового прогнозу відвідуваності Веб-сайтів..... | 84 |

| | |
|--|------------|
| 3.3. Метод ідентифікації математичних моделей довготермінового прогнозу відвідуваності Веб-сайтів..... | 91 |
| Висновки до розділу 3..... | 96 |
| РОЗДІЛ 4. ПРИКЛАДНІ АСПЕКТИ МОДЕЛЮВАННЯ ВІДВІДУВАНOSTІ ВЕБ-САЙТІВ..... | 98 |
| 4.1. Архітектура програмного комплексу для математичного моделювання процесів відвідуваності Веб-сайтів. | 98 |
| 4.2. Приклад застосування автоматизованої процедури побудови структури Веб-сайту для цільової аудиторії навчального підрозділу | 105 |
| 4.3. Моделювання процесів підвищення відвідуваності Веб-сайтів..... | 114 |
| 4.3.1. Математична модель відвідуваності Веб-сайту навчального підрозділу | 114 |
| 4.3.2. Математична модель відвідуваності промо сайту Api2Cart | 126 |
| Висновки до розділу 4..... | 135 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... | 138 |
| Додаток А. | 156 |
| Додаток Б. | 158 |
| Додаток В. | 164 |
| Додаток Г. | 170 |
| Додаток Д. | 175 |
| Додаток Е. | 177 |

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

| | |
|-------|---|
| ТНЕУ | Тернопільський національний економічний університет |
| ФКІТ | Факультет комп'ютерних інформаційних технологій |
| AJAX | asynchronous JavaScript and XML |
| ExtJS | Extended JavaScript |
| HTTP | Hypertext Transfer Protocol |
| SEO | Search engine optimization |
| WWW | World Wide Web |

ВСТУП

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 сформувати.

В дослідженні F. Menczer та J. Ratkiewicz 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 A. Stibe, J. Visevskis 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. Науково-дослідна робота за темою дисертації проводилася 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, 2012-2013рр.), 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. У 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. У 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 (Додаток Е).

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 досліджень: процеси функціонування Веб-орієнтованих програмних систем.

Предмет досліджень: 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 Web Mining та структурування Веб-онтологій забезпечило розвиток методу побудови структур тематичних Веб-сайтів.

Наукова новизна отриманих результатів.

Вперше:

- 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;

- побудовано та обґрунтовано метод ідентифікації моделей 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;

- 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 отриманих 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» (додаток Д);
- для прогнозування динаміки відвідуваності Веб-сайту проекту API2Cart ПП «МагнетікВан» (додаток Д).

Особистий внесок здобувача. Усі наукові результати, викладені у дисертаційній роботі, отримані автором особисто.

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 – Славсько, 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 ТНЕУ (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 [122], 6 статей у фахових наукових виданнях України [27, 28, 30-32, 35], 10 публікацій у матеріалах конференцій [26, 29, 33, 34, 118-121, 123,124].

Структура дисертації. Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Робота містить 178 сторінок, 5 таблиць, 57 рисунків, 6 додатків, 143 найменування використаних джерел. Додатки на 22 сторінках.

РОЗДІЛ 1. ОСОБЛИВОСТІ МЕТОДІВ ТА ЗАСОБІВ ПІДТРИМКИ ВІДВІДУВАНOSTІ ВЕБ-САЙТІВ

1.1. Особливості забезпечення відвідуваності Веб-сайтів та фактори впливу на неї

Термін “Веб-сайт” (від англ. Website: web - «павутина, мережа» і site - «місце», буквально «місце, сегмент, частина в мережі») означає сукупність електронних документів (файлів) приватної особи або організації в комп'ютерній мережі, об'єднаних під однією адресою (доменним ім'ям або IP-адресою) [17]. Всесвітню павутину утворюють мільйони Веб-серверів. Для перегляду інформації, отриманої від Веб-сервера, на клієнтському комп'ютері застосовуються спеціальні програми - Веб-браузери. HTML-файли (Веб-сторінки) є основним типом ресурсів Всесвітньої павутини. Можливість пошуку інформації в Інтернеті забезпечують пошукові системи. Програмною частиною пошукової системи є пошукова машина (пошуковий двигун) - комплекс програм, що забезпечує функціональність пошукової системи і зазвичай є комерційною таємницею компанії-розробника. За даними компанії Net Applications, у листопаді 2011 року використання пошукових систем розподілялося таким чином: Google - 83,87%; Yahoo! - 6,20%; Baidu - 4,22%; Bing - 3,69%; Yandex - 1,7%; Ask - 0,57%; AOL - 0,36% [41]. Окремі складові Інтернету - мережі різної архітектури, пов'язані між собою маршрутизаторами.

Виділяють декілька неформальних етапів розвитку Вебу. Першим етапом вважають період класичного Вебу або Web 1.0. Він характеризувався статичним наповненням сторінок, слабкістю засобів діалогу з користувачами, відсутністю Веб-сервісів, обмеженим використанням стилів при оформленні Веб-сторінок. Поточний стан глобальної мережі характеризують як Web 2.0. Цей термін позначає проекти та сервіси, які активно розвиваються та покращуються самими користувачами, тобто блоги, wiki, соціальні мережі. Окрім того сучасний стан Вебу характеризується розвитком Веб-служб,

технологій Ajax, Web mash-up, соціалізації, збагаченням графічного оформлення.

До небезпек, що супроводжують розвиток підходів Web 2.0, відносять надмірну залежність Вебу від наповнення користувачами, що призводить до розповсюдження неякісної інформації, залежність Веб-сервісів від базових служб підтримки, можливість маніпулювання суспільною свідомістю [16]. Web 3.0 - це принципово новий підхід до обробки інформації, представленої у Всесвітній павутині. Якщо Web 1.0 передбачає Веб-майстра в якості постачальника контенту, а Web 2.0 - спільноту рівноправних користувачів, що генерують контент в рамках тематичного проекту, то Web 3.0 вже дозволяє цим рівноправним користувачам "вибирати" експертів в певній області покладатися на їх рекомендації.

Семантична павутина (англ. Semantic Web) - частина глобальної концепції розвитку мережі Інтернет Web 3.0. Дана концепція ґрунтується на структуруванні документів не за їх формальним представленням, а за віднесенням елементів наповнення до певних змістових структур, які описуються метаданими. Це дає великі переваги в автоматизованому аналізі змісту документів. Перешкодою для впровадження семантичного структурування документів слугує як неохочість людей виконувати додаткові роботи з підтримки документів з метаданими, так і складність побудови узгоджених структур метаданих. Окрім того, реалізація семантичних пошукових систем, які будуть відбирати і надавати потрібний користувачеві фрагмент Веб-сторінки за семантичними ознаками, відміняє необхідність відвідувати Веб-сайт, що зменшить можливість використання інтернет-реклами, а отже і можливості фінансування інтернет-проектів [41].

Для класифікації Веб-сайтів застосовуються різні підходи та різна повнота охоплення. У нашому аналізі не розглядатимемо розважальних Веб-сайтів, тобто обмежимося прикладними. Як переосмислення класифікації, наведеної в [11, 45], нами запропоновано схему, наведену на рис. 1.1. Зокрема,

виділено окремі категорії бізнесових (представницькі та продуктові) та інформаційних Веб-сайтів, а також реферативних та ресурсних Веб-сервісів.

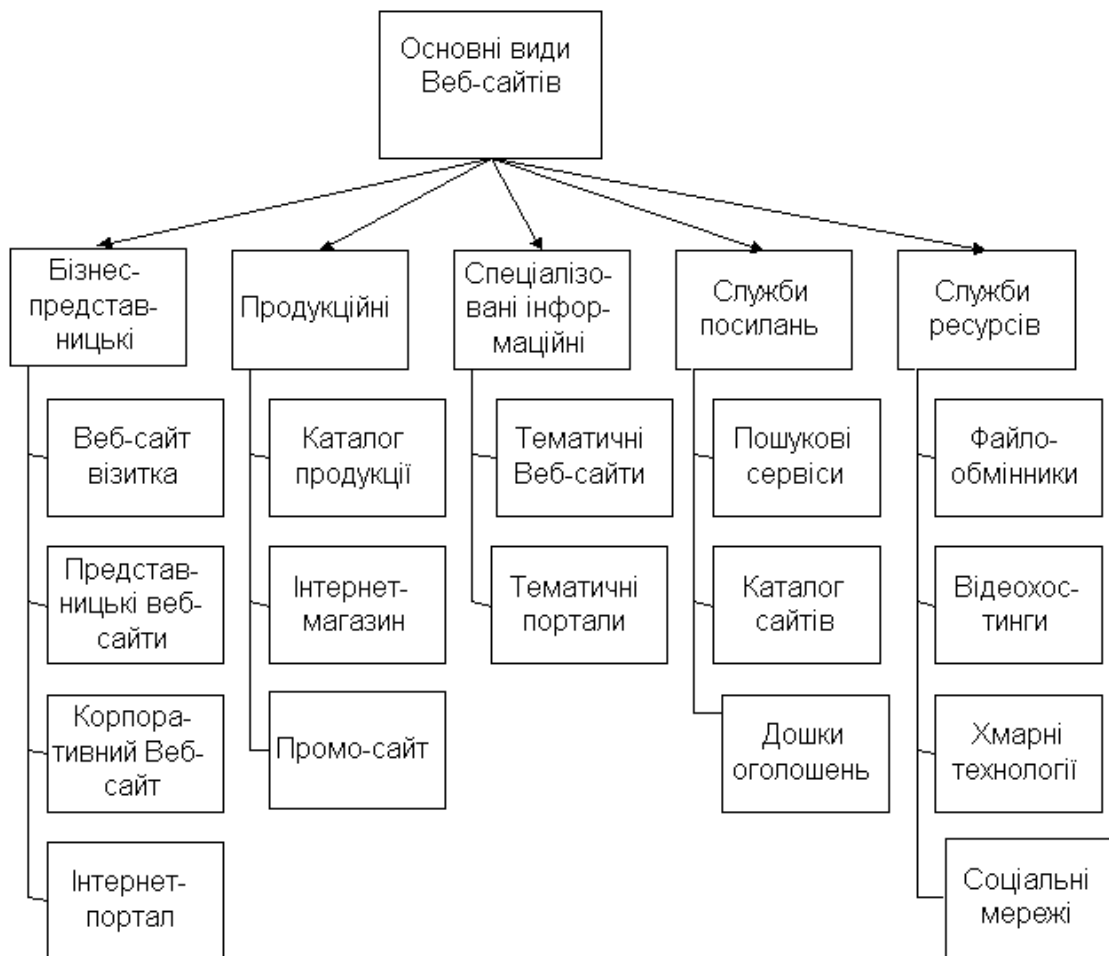


Рис. 1.1. Основні види Веб-сайтів

Інтернет-представництво власників бізнесу в найпростішій формі реалізується у Веб-сайті візитці. Такий ресурс містить лише загальну інформацію про власника, контактні дані, сферу діяльності, перелік продуктів та послуг. Якщо до Веб-сайту візитки додається детальний опис послуг та продукції, з портфелем виконаних та отриманих замовлень, формами діалогу із споживачами, то він вже відноситься до представницьких Веб-сайтів.

Наступним видом в градації даних ресурсів є корпоративний Веб-сайт, який відрізняється від представницького сайту тим, що містить численні сервіси для діалогу із користувачем, такі як пошукові сервіси, блоги, фотогалереї, форуми. Інтернет-портал формується об'єднанням окремих

самодостатніх Веб-сайтів підрозділів корпорації або навіть окремих споріднених організацій. До градації продуктових Веб-сайтів відносять каталоги продукції. Якщо до каталогу продукції додається сервіс для замовлення та оплати товарів, то отримують Інтернет-магазин. Промо-сайти присвячені конкретній торговій марці, яку рекламує і подає про неї всеможливу супутню інформацію а також підтримує контакти із потенційними покупцями.

В категорії інформаційних Веб-сайтів варто відзначити тематичний Веб-сайт та тематичний портал. Тематичний Веб-сайт надає спеціалізовану інформацію із вибраної теми. В свою чергу тематичний портал узагальнює та розвиває поняття тематичного Веб-сайту, фактично об'єднуючи кілька тематичних Веб-сайтів та надаючи вичерпну інформацію з певної тематики. Портали додатково містять засоби взаємодії з користувачами а також організовують форуми та чати.

Дві наступні категорії Веб-сайтів відведені для Веб-сервісів, тобто ресурсів, створених для надання послуг в мережі WWW. Сервіси посилань подають посилання на потрібні Веб-сайти. Це, зокрема, пошукові сервіси. В цю категорію також віднесено дошки оголошень, де розміщують повідомлення про умови продажу або купівлі товарів та послуг [11]. Наступна градація - це каталоги сайтів, тобто ресурси, на яких розміщують сайти і блоги, що сприяють їх просуванню. Це, зокрема, файлообмінний сервіс (напр. Bittorrent); хмарне сховище даних (Skydrive); сервіс редагування даних (Google Docs); відеохостинг (YouTube).

Кожен Веб-сайт має мету створення – спронукання доздійснення певних бажаних дій його відвідувачами. Обсяг таких дій визначає результативність Веб-сайту. Ця характеристика тісно пов'язана із відвідуваністю Веб-сайту, хоча і не визначається останньою однозначно. Відвідуваність – це один із найважливіших показників, які цікавлять власників Веб-сайтів, оскільки дозволяють цим ресурсам виконувати покладені на них завдання, зокрема отримання економічної вигоди, формуванні суспільної думки, інформування

суспільства про результати досліджень, зміцнення особистого авторитету або авторитету організації. Оскільки кількість сайтів в Вебі невинно зростає, а аудиторія збільшується значно повільнішими темпами, чітко відстежується нарощення конкуренції в боротьбі за аудиторію [18]. При цьому велика частка Веб-сайтів (до 50%) має вкрай низьку відвідуваність, а мінімальну відвідуваність мають 25% Веб-сайтів від загальної кількості [37, с.11]. Серед низько відвідуваних за критерієм суспільної значимості, виділимо Веб-сайти, що представляють колективи працівників, а не окремих суб'єктів. З метою конкретизації досліджень із згаданої множини виділимо категорії Веб-сайтів навчальних підрозділів та промо-сайтів нових видів продукції. Веб-сайти першого типу можуть володіти низькою відвідуваністю через відсутність системного підходу в її формуванні, а Веб-сайти другого типу – через необізнаність потенційної аудиторії про особливості нових видів продукції.

Загальну характеристику факторів впливу на відвідуваність наочно представлено на рис. 1.2.



Рис. 1.2. Основні фактори впливу на відвідуваність Веб-сайтів

Верхній рівень наведеної структури сформовано на основі загальних положень SEO, які, зокрема, наведені в роботі [42]. При цьому на першому етапі виходу на Веб-сайт дуже важливу обмежувальну функцію відіграє швидкість завантаження його Веб-сторінок. Якщо вона перевищує деяку

межову величину, то відповідний Веб-сайт втрачає переважну кількість своїх потенційних користувачів, оскільки пошукові системи пропонують численні альтернативи із прийнятними швидкостями завантаження.

На наступному етапі обмежувальну функцію на формування високої відвідуваності відіграє логічність та прозорість структури Веб-сайту, адже неоптимально структуровані ресурси значно втомлюють в процесі пошуку необхідної інформації. Дизайн Веб-сайту також є важливою характеристикою, оскільки привабливий вигляд Веб-сторінок та зручність роботи із ними сприяє збільшенню часу роботи із вибраним Веб-сайтом та росту ймовірності повторної його відвідуваності. Однак, на наш погляд, цей фактор допускає множинні авторські трактування і лише достатньо невдалі рішення можуть суттєво знизити потенційну аудиторію Веб-сайту. На цьому фоні надзвичайно важливим є фактор належного рекламування Веб-ресурсу, що значно збільшує кількість його потенційних користувачів.

Контент є найважливішою компонентою Веб-сайту. Якщо на ньому відсутня корисна інформація, здатна зацікавити певне коло споживачів, то ставиться під сумнів доцільність самого його функціонування. Контент є найважливішою компонентою Веб-сайту. Якщо на ньому відсутня корисна інформація, здатна зацікавити певне коло споживачів, то ставиться під сумнів доцільність самого його функціонування. В роботі [35] проведено експериментальний аналіз впливу структури контенту на популярність найуспішніших Веб-сайтів. В якості прикладу проаналізовано рейтинг Веб-сайтів українських навчальних закладів станом на 21.12.2010. Для вибору найвідвідуваніших українських Веб-сайтів можна використати електронні рейтинги популярності, що розташовані, зокрема, на Веб-сайтах [62, 132]. Проте статистика в них коливається з дня на день, тому проведемо свій аналіз на основі стабільніших характеристик, а саме позицій у високонкурентних запитах до пошукових систем. Оскільки аналізувалися тільки українські ВНЗ, то вибрано лише українські пошуковики та пошуковий запит „університет”. Адекватні списки посилань за цим запитом подали наступні пошукові системи

Мета [21] та Бігмір [39]. Для визначення найрейтинговіших сайтів відфільтровано 10 перших елементів із позицій кожного пошуковика та проранговано їх по шкалі від 10 до 1. Знайдено сумарний ранг для кожного із вибраних Веб-сайтів і знову відфільтровано перші 10 позицій. Відібрані Веб-сайти впорядковано за критерієм авторитетності. З цією метою врахуємо кількість зовнішніх посилань на дані сайти на основі даних пошукової системи Мета. Скоригований рейтинг наведено в таблиці 1.1.

Проаналізуємо ресурси перших чотирьох Веб-сайтів рейтингу. Розпочнемо із послуг для потенційних клієнтів даних організацій – абітурієнтів. Всі Веб-сайти подають інформацію про факультети та спеціальності підготовки. Однак лише КНУ ім. Т.Шевченка та НУ «Києво-Могилянська академія» подають детальну інформацію про можливості працевлаштування зі студентської лави, що має турбувати тих, хто при вступі одразу планує майбутню кар'єру.

КНУ ім. Т.Шевченка представляє окремий Веб-сайт по працевлаштуванню із тренінгами з пошуку першої роботи (аналітичний матеріал) та переліком поточних вакансій для працевлаштування, що можна трактувати як розміщення актуальних аналітичних матеріалів.

Вакансії представляє і НУ «Києво-Могилянська академія», але в дещо скромніших масштабах. Львівський національний університет ім. І. Франка надає електронні підручники для вступників, свого роду продукти, та сервіси пробного тестування із вступних дисциплін. Сайт студентів факультету прикладної інформатики та математики подає також невеличкий перелік вакансій для ІТ-спеціалістів. Далі проаналізуємо як представлено на Веб-сайтах „обслуговування клієнта”, тобто, яка допомога надається студентові в процесі навчання. КНУ ім. Т.Шевченка, Києво-Могилянська академія та Харківський національний університет надають доступ до колекції зарубіжних електронних публікацій (продукти). В то й же час ЛНУ ім. І.Франка подає літературу та тести (продукти та сервіси) для системи дистанційного навчання, що відображається у схемі, наведеній на рис. 1.2.

Таблиця 1.1.

Рейтинг сайтів українських вузів пропорційно кількості зовнішніх посилань на них

| N | Назва вузу | Сумарний рейтинг пошуковиків Мета та Бігмір | Зовнішні посилання в пошуковику Мета |
|----------|-----------------------------------|--|---|
| 1 | КНУ ім.Т.Шевченка | 19 | 70 |
| 2 | НУ «Києво-Могилянська академія» | 8 | 65 |
| 3 | ЛНУ ім. І.Франка | 9 | 39 |
| 4 | Харківський національний ун-т | 10 | 38 |
| 5 | НТУ «Київськ. Політехнічний ун-т» | 14 | 24 |
| 6 | НТУ «Харків. політехн ун-т» | 6 | 22 |
| 7 | Національний авіаційний ун-т | 10 | 20 |
| 8 | КНЕУ | 13 | 17 |
| 9 | КНТЕУ | 4 | 8 |
| 10 | Донецький національний. ун-т | 5 | 4 |

Таким чином аналіз кращих Веб-сайтів провідних університетів України доводить суттєвий вплив аналітичних матеріалів, продуктів та сервісів на популярність Веб-сайтів та на їх позиціонування в пошукових системах. Нарощення цих ресурсів сприятиме покращенню характеристик Веб-сайтів. Як приклад, можна навести відкриту колекцію електронних підручників кафедри інформатики та інтелектуальної власності факультету автоматики і приладобудування Харківського національного політехнічного університету [24]. На основі інформації пошукової системи <http://meta.ua> можна виявити 8 зовнішніх посилань на цей ресурс, що перевершує показники деяких університетських Веб-сайтів.

На основі публікації [35] а також класифікації типів Веб-сайтів, наведеної на рис. 1.1, структуровано основні функції контенту, які можуть суттєво вплинути на відвідуваність. Сюди віднесено наповненість Веб-сайту довідковими та аналітичними матеріалами, корисними сервісами та засобами залучення користувачами популярних продуктів. Із переліку видів прикладних Веб-сайтів (рис. 1.1), лише найпростіші за структурою ресурси, такі як Веб-сайти візитки, представницькі та каталоги продукції, можуть обмежуватися лише довідковим контентом. Корпоративні і тематичні Веб-сайти та портали

повинні також містити хоча б аналітичну інформацію. Продукційні та службові Веб-сайти відповідно базуються на наданні можливостей придбання певних видів продукції та наданні послуг.

Виділення факторів впливу на відвідуваність Веб-сайтів породжує завдання досягнення та підтримки її на належному рівні. Його розв'язання дозволяє концентруватися на підвищенні результативності Веб-сайту, що є кінцевою метою його функціонування. Самостійне і важливе значення має розробка наукових підходів у формуванні засобів підтримки відвідуваності. Аналізу цього питання приділений наступний підрозділ.

1.2. Засоби підтримки відвідуваності Веб-сайту

Наведений вище аналіз дозволяє виділити основні фактори впливу на відвідуваність. Зокрема на основі схеми, наведеної на рис. 1.2, до таких факторів віднесено логічність структури Веб-сайту а також оптимізацію та рекламування його контенту. Під логічністю структури Веб-сайту в даному контексті розуміють зручність роботи із нею користувача, що працює із відповідною предметною областю. Ця зручність досягається за допомогою максимальної адаптації структури цього Веб-сайту до типової структури Веб-сайтів предметної області. Проблема полягає в тому, що структура типового Веб-сайту певної предметної області є динамічною. Для її встановлення необхідні значні затрати часу розробника Веб-сайту, що також не виключає суб'єктивізму в побудованих оцінках. Уникнути цього можна із застосуванням автоматизованого аналізу структури Веб-сайтів певної предметної області, проте засоби такої автоматизації на сьогоднішній день невідомі.

Однак наявність зручної типової структури Веб-сайту можна розглядати лише як супутній фактор високої відвідуваності Веб-сайту. Основну роль звичайно відіграє його контент. Він повинен забезпечувати хорошу видимість Веб-сайту для початкового контакту із користувачем а також демонструвати

хорошу якість наповнення для стимулювання подальшої співпраці із Веб-сайтом, що наочно представлено на рис. 1.3.

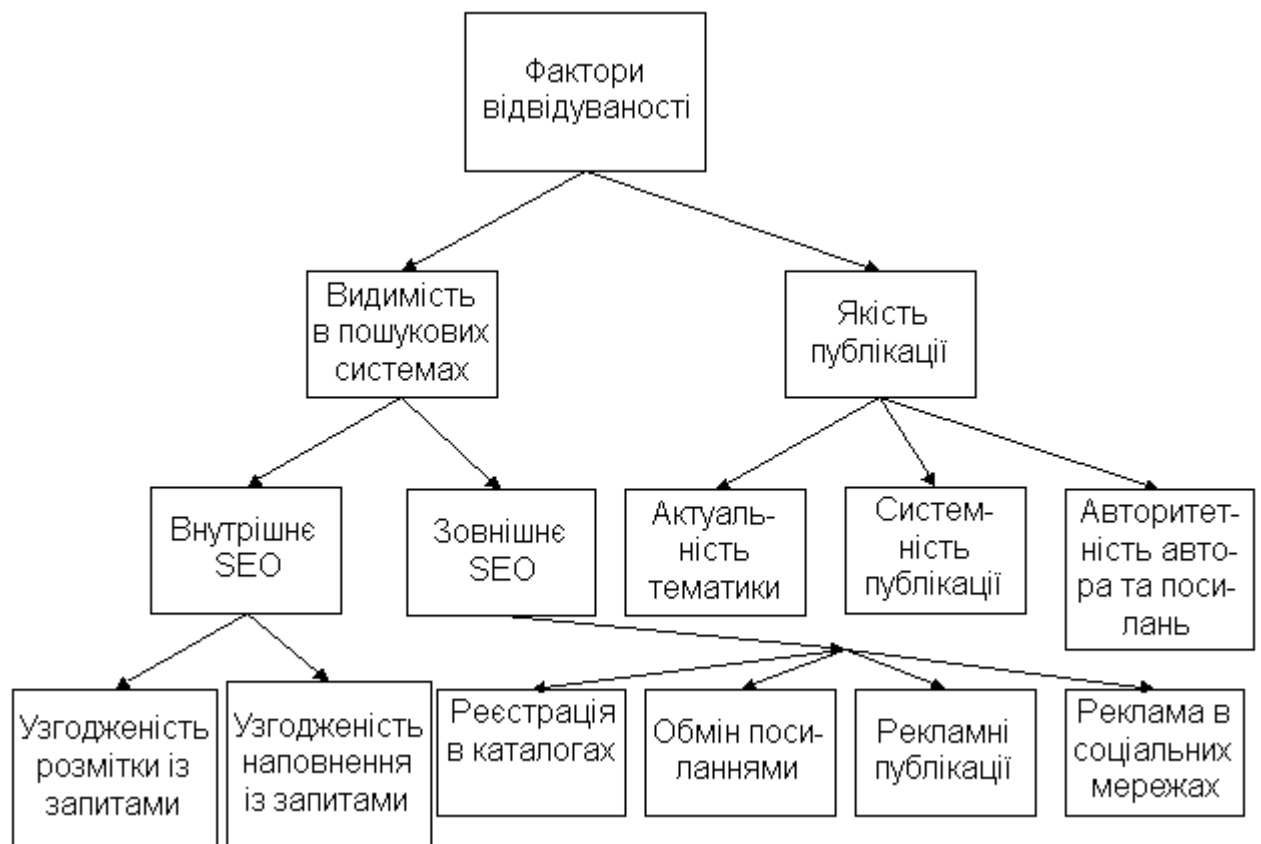


Рис. 1.3. Основні фактори впливу контенту на відвідуваність Веб-сайтів

Вплинути на користувачів мережі можуть лише ті ресурси, які мають найкращу видимість (бажано серед перших 10) по тематичних запитах до пошукових систем. Ранжування однотипних матеріалів пошуковими системами здійснюється автоматично, тому існує цілий спектр послуг та рекомендацій для підвищення видимості певних матеріалів в пошукових системах, об'єднаних спільним терміном SEO. Під пошуковою оптимізацією (англ. search engine optimization, SEO) розуміють заходи для покращення позицій Веб-сайту в пошукових системах по запитах відповідної цільової аудиторії.

Робота з оптимізації включає насичення тексту до певного рівня ключовими словами із бажаних запитів користувачів, покращення

стилістичного оформлення тексту із застосуванням засобів виділення основних елементів, застосування прозорої навігації, оптимізація внутрішніх та зовнішніх посилань. Окрім того необхідно забезпечити реєстрацію Веб-сайту в каталогах пошукових систем, обмін посиланнями, розміщення статей в соціальних мережах, блогах, спеціалізованих Веб-сайтах. Ці методи є відомими і широко вживаними. Вони також сприяють хорошій відвідуваності але не здатні її гарантувати, оскільки базуються на тимчасовій адаптації до пошукових алгоритмів конкретних пошукових систем.

Інколи в гонитві за високою видимістю Веб-дизайнери подають неякісний контент та прагнуть штучно підвищувати його видимість за рахунок врахування особливостей пошукових систем, тобто із використанням методів «сірого» або «чорного» SEO. На протидію цим підходам компанія Google в 2011 році увела до експлуатації новий алгоритм пошуку в Інтернет під назвою Panda [49]. У квітні 2012 вийшло ще одне оновлення алгоритму Панда під назвою Пінгвін. У цьому оновленні до вимог внутрішньої якості Веб-сайту додалася ще і вимога до якості зовнішніх посилань сайту. Ще в 2008 році пошукова система Яндекс запровадила фільтр АГС, із подібними функціями. Алгоритми таких фільтрів засекречені, щоб Веб-майстри його не обійшли [47].

Очевидно, що пошукові системи розпочали систематичну боротьбу із сірою та чорною SEO, залишаючи допустимим лише методи білої SEO, які закликають розробляти якісні сервіси для користувачів, а не для пошукових серверів. В цих умовах орієнтиром для розробників повинна стати концепція якісного наповнення Веб-сайту [3, 6, 45, 74, 114]. У цьому зв'язку численні SEO-оптимізатори для оцінки якості Веб-сторінок запропонували використовувати критерії лише дещо модифіковані від тих, які використовують для оцінки якості друкованої продукції, узагальнюючи при цьому публікації різних авторів [105, 131]. В роботі [78] відзначається, що найважливішою умовою, яка визначає відвідуваність Веб-сайту є клас потреб користувача, наприклад, професійні, аматорські, розваги.

В даному дослідженні аналізуються Веб-сайти, які задовольняють професійні та пізнавальні потреби користувачів. Для цього виду ресурсів доцільно використовувати критерії, що використовують в аналізі цінності наукових публікацій. Згідно дослідження [100], зокрема, одним із перших критеріїв є авторитетність автора публікації, або посилання його публікації на авторитетних авторів даної галузі досліджень. Окрім цього важливим є авторитетність сайту та його офіційний статус. Наступними факторами називається актуальність аналізованого питання, перевірюваність його положень, відповідність застосовуваних методик. Для Веб-ресурсу є важливим його системність, тобто підтримка множиною публікацій, близьких за тематикою. Також важливою є наявність бібліографії, яка підтверджує знайомство автора із базовими методиками наукових досліджень в даній галузі знань.

Наведені оцінки висувають контент Веб-сайту в ряд найважливіших факторів його популярності. При цьому надзвичайно важливе значення має системність охоплення тематик Веб-сайту, актуальність для його аудиторії та регулярне поновлення інформації. Це вимагає продуманої структуризації, як самого Веб-сайту, так і спрямованості його інформаційного оновлення. Такі критерії вимагають залучення висококваліфікованих та досвідчених спеціалістів для ведення відповідних рубрик.

Однак для основної маси непопулярних Веб-сайтів така вимога не є достатньо реальною, оскільки пов'язана із значним зростанням витрат на обслуговування цих сайтів. Виходом із даного протиріччя є залучення відповідальних та допитливих працівників, які оволодівають 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, як і виділення структури

тематичних Веб-сайтів, можлива із застосуванням методів та засобів Web Miningy.

Web Mining призначений для ефективного вирішення завдань пошуку, структуризації та аналізу, в основному хаотично організованої інформації в глобальній мережі. Основними класами методів Web Miningy вважаються: пошуковий Web Mining, структурний Web Mining, видобування знань із Вебу, персоналізація інформації у Вебі, пошук шаблонів в поведінці користувачів Вебу [9]. Вони представлені на рис. 1.4 та аналізуються нижче.

Методи Text Mining призначені для автоматичного видобування знань із текстових матеріалів, що дозволяє експерту обробляти величезні масиви інформації. При цьому для оцінки змісту та виявлення спільних тематик використовуються пошуки по ключових словах, кластеризація та класифікація документів, побудова семантичних мереж [51, 54, 69, 71, 114]. Окремим і важливим методом є формування онтологій засобами Text Mining, що детальніше буде розглянуто нижче.

Пошуковий Web Mining покликаний підвищити ефективність роботи пошукових систем, які ґрунтуються на підборі потрібних ключових слів в масиві аналізованих текстів. В ході цього аналізу результати кластеризуються в декілька однорідних груп, які характеризуються відповідними наборами ключових слів. Вибравши потрібний кластер, користувач переглядатиме результати видачі найближчі до контексту його запиту [93]. Структурний Web Mining полягає в аналізі структури посилань між різними Веб-сторінками та Веб-сайтами у виділеному мережевому сегменті. Такі задачі виникають при аналізі певних предметних областей за структурою взаємних посилань їх сторінок, аналізі соціальних мереж або при дослідженні системи цитувань окремого автора [92].

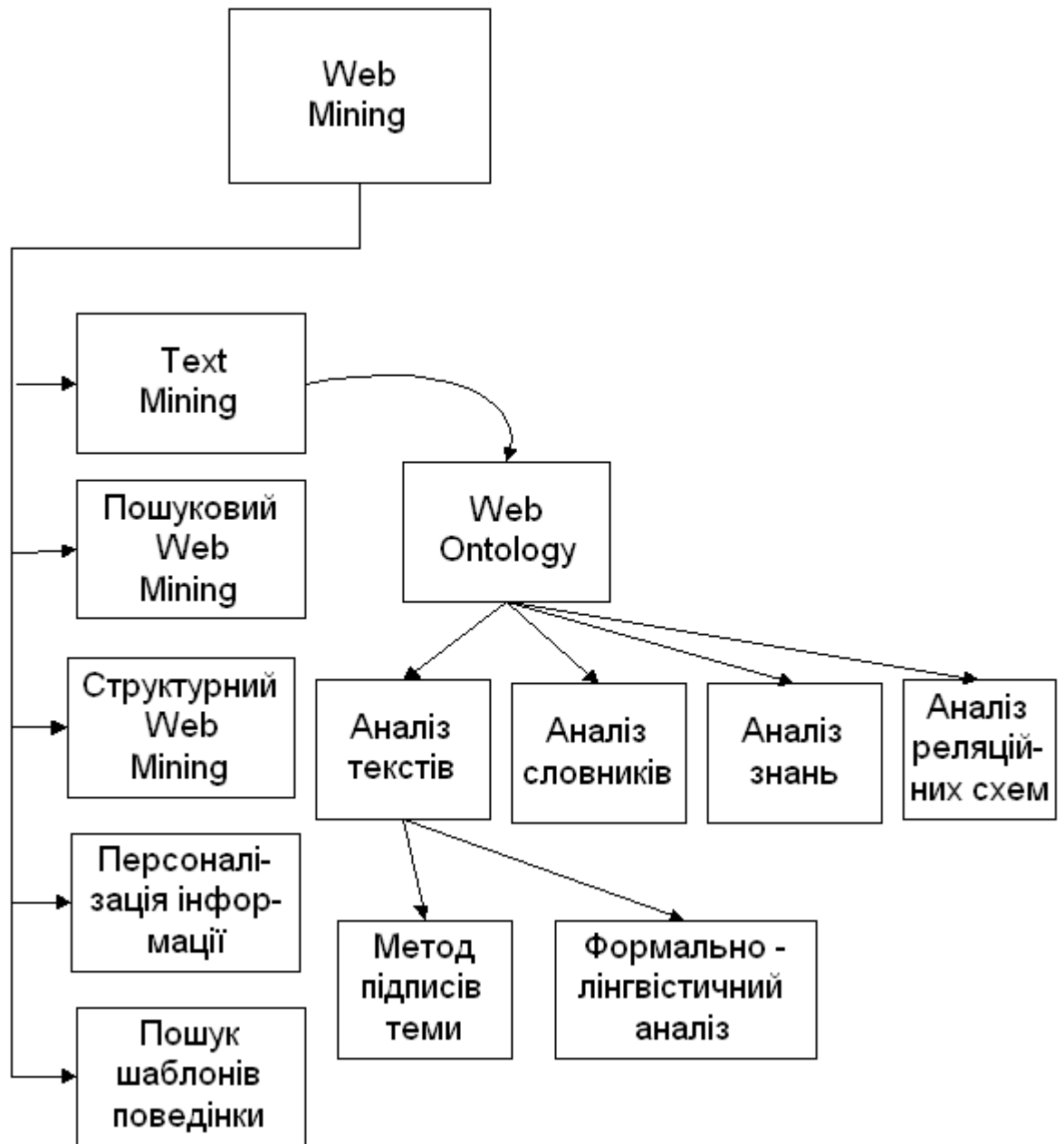


Рис. 1.3. Базові підходи Web Mining

Персоналізація Веб-простору — завдання по створенню Веб-систем, що адаптують свої можливості під експерта на підставі зібраної і проаналізованої інформації про виявлені переваги користувача. Для аналізу інформації про користувача орієнтуються на стійкі шаблони його "поведінки" в мережі: послідовності кліків усередині ресурсу, переходах на інші підресурси, періодах мережевої активності, здійснюваних покупках [70, 80, 107, 108, 134].

Пошук шаблонів в поведінці користувачів використовується для вивчення запитів цільової аудиторії із наступною модифікацією структури Веб-сайту [22, 55, 65-67, 76, 81, 95, 104]. Ці методи можуть принести відчутні вигоди при високій відвідуваності Веб-сайту.

Серед аналізованого напрямку досліджень найближчими до нашої роботи є методи видобування знань із Веб-текстів. Тому розглянемо детальніше окремі дослідження даної категорії. Онтологія явного подання знань є одним з ключових елементів у Semantic Web. Методи машинного навчання, такі як побудова асоціативних правил або кластеризація, використовуються для вилучення знання з Веб-документів з метою створення нової онтології або поліпшення існуючої [60, 79, 84, 90, 103, 104, 111, 142].

Однозначного трактування поняття онтології до цього часу не вироблено. Нам близьке трактування цього поняття, представлене в роботі [80], де під онтологією розуміється система понять деякої предметної області, яка представляється як набір сутностей, з'єднаних різними відношеннями. Перевагою онтологій в якості способу подання знань є їх формальна структура, яка спрощує комп'ютерну обробку.

Структура онтології являє собою набір елементів чотирьох категорій: поняття, відношення, аксіоми, окремі екземпляри. Класи (або поняття) описують групу індивідуальних сутностей, які об'єднані на підставі наявності спільних властивостей. Поняття можуть бути пов'язані різного роду відношеннями, які пов'язують воедино класи і описують їх. Найпоширенішим типом відношень, що використовується у всіх онтологіях, є таксономічне відношення (відношення типу is-a, або a-kind-of).

Складові онтології підпорядковуються своєрідній ієрархії. На нижньому рівні цієї ієрархії знаходяться екземпляри, конкретні індивіди, вище йдуть поняття, тобто категорії. На рівень вище розташовуються відношення між цими поняттями, а узагальнюючим і сполучним є рівень правил або аксіом. До онтології можна віднести ряд структур, що відрізняються різним ступенем формалізованості: глосарій, проста таксономія, тезаурус (таксономія з

термінами), понятійна структура з довільним набором відношень, повністю аксіоматизована теорія.

Детально засоби та методи побудови та використання Веб-онтологій проаналізовано в роботах [50, 55, 56, 63, 64, 79, 83, 86, 89, 96, 98, 102, 112, 115, 134]. В дослідженнях відзначається, що запропоновані методи дозволяють прискорити процес набуття знань, необхідних для створення онтології з нуля; зменшити час, необхідний для поповнення існуючих онтологій, прискорити побудову онтологій, які будуть використовуватися для різних цілей в семантичному Вебі. В дослідженнях [65, 85, 107, 111] пропонується наступна класифікація методів побудови онтологій за джерелами їх наповнення: на основі аналізу текстів, словників, знань, напівструктурованих схем та реляційних схем. Для наочнішого подання специфіки досліджень даної галузі проаналізуємо декілька методів, із наведених в даному огляді.

В дослідженнях [52, 57, 58, 68, 69, 88, 101, 110] розглядається задача опису предметної області Веб-документів на основі аналізу великих текстових колекцій. На етапі попередньої обробки запропоновано вилучення стоп-слів а також невеликої множини спеціальних малоінформативних слів, що можуть часто зустрічатися. На наступному етапі будуються фрази максимальної довжини, що часто зустрічаються в колекції документів (вище певного порогу). При цьому породжується велика кількість достатньо подібних фраз. Для їх узагальнення на наступному етапі будуються класи подібних фраз, відмінності складу яких не перевищують певного порогу (наприклад 90%). Після таких побудов документ може бути описаний фразами із відповідних класів, що характеризують його предметну область, хоча слабо характеризують його специфіку. Для характеристики останньої в опис документа пропонується включати терміни, що найчастіше зустрічаються в самому документі [54, 71, 87].

Метод Agirre [65] спрямований на поповнення понять в існуючих великих онтологіях за допомогою тексту з Вебу. Загальна мета цього підходу полягає в подоланні двох недоліків відомої онтології WordNet, таких як

відсутність тематичних зв'язків між поняттями, і наявність різних змістів для кожного поняття. Згідно методу спочатку необхідно отримати документи, пов'язані з поняттям, після чого відбираються слова в тексті, які найтісніше пов'язані з даним поняттям.

Цей підхід ґрунтується на використанні методу підписів теми, введеного в працях [88, 101]. Запропонована стратегія побудови таких списків полягає в тому, що інформація, яка міститься в онтологіях використовується для побудови запитів, на основі яких вибираються документи відповідно до змісту поняття. Тоді отримані тексти організуються в колекції, кожна за певним смислом слова. На наступному етапі для кожної колекції визначаються частоти слів і порівнюються з даними інших колекцій, які охоплюють інший зміст того ж поняття.

Підхід Aussenac-Gilles здійснює наповнення онтологій, базуючись на видобуванні знань із технічних документів [57, 58, 97]. Метод формує доменну модель за допомогою аналізу інформаційного масиву, користуючись обробкою природної мови, лінгвістичними засобами і методами. Метод поєднує лінгвістичні інструменти набування знань з методами моделювання, що дозволяє відслідковувати зв'язки між моделями і текстами. Розробник онтології, якому допомагають інструментами, вибирає і комбінує результати, щоб створити онтологію. Згідно цього підходу були побудовані, наприклад, три онтології: інструменти інженерії знань (близько 80 концептів), методи волоконного скляного виробництва для приватної акціонерної компанії (близько 100 концептів), методи організації туристичних акцій для експерименту EON [61, 98, 100, 116, 128, 137, 138, 142]. Окрема група робіт приділена побудові онтологій на основі схем баз даних [59, 72, 75, 77, 91, 94, 109, 113, 117, 125, 130, 131, 135, 140].

На наш погляд проблеми суто синтаксичних підходів в побудові онтологій полягають в великій невизначеності структурування онтологій у встановленні її ядра. Призначення онтології може ґрунтуватися на аналізі розповсюдженої практики діяльності у відповідній галузі. Це дозволяє

використовувати онтології для структурування Веб-сайтів, а аналіз практики розвитку Веб-сайту для структурування онтології. Для забезпечення унікальності кожного Веб-сайту він повинен наповнюватися актуальними Веб-сервісами а також актуальним та систематизованим контентом. З метою систематизації контенту із актуальних предметних областей зручно використовувати онтології. Для автоматизації їх побудови доцільно використовувати методи Web Mining, що враховують специфіку відповідних предметних областей.

Ефект від оновлення контенту Веб-сайту можна посилити, застосувавши акції його рекламування засобами публікацій на спеціалізованих Веб-сайтах, анонсування на дружніх Веб-сайтах, направлених розсилах на електронну пошту. Важливою проблемою при плануванні та реалізації заходів розвитку контенту а також заходів рекламування контенту є оцінка їх ефективності. Така оцінка ускладнюється тим, що для отримання належного ефекту всі дії у рамках вказаних заходів повинні носити не разовий, а системний характер. Наступне ускладнення полягає у зміщеній в часі реакції на заходи щодо підвищення відвідуваності. Окрім того, на відвідуваність суттєво впливає людський фактор, що проявляється як у авторських підходах у реалізації запланованих заходів, так і у особливостях реакції цільової аудиторії Веб-сайту. Вказані ускладнення вимагають розробки принципово нових моделей відвідуваності на основі використання елементів існуючих підходів до моделювання. Аналізу останніх приділено наступний підрозділ.

1.3 Особливості математичного моделювання процесів відвідуваності Веб-сайтів

В результаті спілкування із керівниками компаній із розробки комерційного програмного забезпечення встановлено, що значимий вплив на результативність Веб-сайту можна здійснити лише при високій його відвідуваності. Висока відвідуваність потрібна також для ефективного пошуку

шаблонів в поведінці користувачів. Тому для Веб-сайтів із низькою відвідуваністю першочерговою є задача максимального її збільшення. Якщо при цьому охоплюється потенційна аудиторія даного Веб-сайту, то відвідуваність стабілізується і мало реагує на зміну структури. Таким чином, для Веб-сайтів із невисокою відвідуваністю основним критерієм ефективності є саме показник відвідуваності.

Далі розглянемо можливі підходи у моделюванні цього показника, вважаючи його цільовим. Ми трактуємо Веб-сайт як своєрідну динамічну інформаційну систему, ефективність розвитку якої можна моделювати [19]. Під динамічною системою розуміють будь-який об'єкт або процес, для якого однозначно визначено поняття стану як сукупності деяких величин в даний момент часу, і заданий закон, який описує зміну (еволюцію) початкового стану з плином часу. Закон еволюції динамічної системи в часі подається системою звичайних диференціальних рівнянь [4, с.13].

Якщо розглядати змінні стану x_1, x_2, \dots, x_n як координати точки \vec{x} в n -вимірному просторі, то отримуємо геометричне представлення стану динамічної системи у вигляді точки, яку називають фазовою точкою. При дослідженні систем, що характеризуються фазовими змінними різних видів, прийнято поділяти змінні на „вхідні”, „вихідні” і „змінні стану”. Самі системи записують у вигляді [20]:

$$\begin{cases} \frac{d\vec{x}}{dt} = F(\vec{x}, \vec{u}, t) \\ \vec{y} = G(\vec{x}, \vec{u}, t) \end{cases} \quad (1.1)$$

де $\vec{x}(t)$ – змінні стану системи; $\vec{u}(t)$ — вхідні дії, $\vec{y}(t)$ — вихідні сигнали; F, G – вектор-функції, що відображають вектори вхідних впливів $\vec{u}(t) \in R^p$ та внутрішніх змінних стану системи $\vec{x}(t) \in R^n$ на множину вихідних величин $\vec{y}(t) \in R^r$.

Аналогічно можна записують і різницеві рівняння для дискретних моделей:

$$\begin{cases} \vec{x}_{k+1} = F(\vec{x}_k, \vec{u}_k, t_k) \\ \vec{y}_k = G(\vec{x}_k, \vec{u}_k, t_k) \end{cases} \quad (1.2)$$

Згідно теорії Я. З. Ципкіна [48], ідентифікація динамічних об'єктів в загальному випадку складається з визначення їх структури і параметрів за даними спостережень вхідного впливу і вихідної величини. Для вирішення цієї задачі необхідно окреслити клас об'єктів, вибрати модель, вибрати показники якості ідентифікації, які б характеризували різницю між значеннями вихідних змінних об'єкту і моделлю, сформулювати алгоритм ідентифікації, який, використовуючи доступні для спостереження значення вхідних та вихідних величин, змінював би параметри моделі так, щоб показники якості ідентифікації з плином часу досягали оптимуму.

За способом побудови структури функціонального оператора, моделі можна поділити на дедуктивні, індуктивні та комбіновані дедуктивно-індуктивні. В дедуктивних моделях структура функціонального оператора встановлюється дослідником апріорі на основі тих чи інших фізичних міркувань (законів) стосовно елементів модельованої системи та їх взаємозв'язку. Побудова структури оператора індуктивної моделі здійснюється на основі таблиці експериментальних даних шляхом перебору (послідовного випробовування) великої кількості альтернативних структур за допомогою комп'ютерного алгоритму індуктивного моделювання. Індуктивне моделювання базується на принципі застосування зовнішнього доповнення [14], зокрема може використовуватися похибка прогнозування на даних, що не використовувалися при їх побудові.

За критерієм врахування невизначеності змінних стану і параметрів моделі можна поділити на детерміновані, інтервальні, нечіткі та стохастичні. Принципова відмінність детермінованих моделей від моделей з врахуванням невизначеності (інтервальних, нечітких та стохастичних) полягає в тому, що їх побудова ґрунтується на припущенні про те, що існує деяка ідеальна структура (апріорі невідома в індуктивних моделях), яка адекватно і з достатньою

точністю описує реальний процес. В той же час, моделі з врахуванням невизначеності такого припущення не вимагають і вважається, що жодна структура не спроможна відобразити динаміку процесу бездоганно [25]. При прогнозуванні динаміки процесу з використанням детермінованих моделей відхилення модельних оцінок від спостережень реального процесу вважаються шумом, рівень якого не досліджується, тоді як відхилення прогнозів в моделях з врахуванням невизначеності — предмет дослідження самих моделей, їхнє основне завдання.

За адаптивністю моделі можна поділити на неадаптивні, адаптивні по параметрах та адаптивні по структурі і параметрах. Неадаптивні моделі (звичайні) ігнорують спостереження динаміки поточної реалізації процесу, не використовуючи їх для уточнення свого функціонального оператора. В адаптивних моделях організовано замкнутий контур налаштування параметрів (і структури) функціонального оператора під поточну реалізацію процесу. Найчастіше адаптивні моделі належать до класу моделей з врахуванням невизначеності (інтервальних, нечітких або стохастичних), що використовують спостереження динаміки процесу для звуження інтервалів невизначеності.

Оскільки кількісних моделей відвідуваності в літературі ще не описано, обмежимося максимально можливими простими підходами. Зокрема, при моделюванні будемо дотримуватися дедуктивного підходу, використовуватимемо детерміновані моделі у вигляді диференціальних рівнянь. Завдяки присутності в модельованому процесі людського фактора використаємо адаптацію моделі за параметрами.

Будемо розглядати відвідуваність як випадкову величину, в якій може спостерігатися явище сезонності. Якщо такі факти буде виявлено, використовуючи коефіцієнти сезонності це явище може бути виключене із аналізу, суттєво не спотворивши його результати. Врахуємо особливості випадкової величини відвідуваності Веб-сторінок та Веб-сайтів. Розподіл такого роду величин описаний в дослідженнях Ratkievich та ін. [127]. На

основі аналізу мільйонів Веб-сторінок ними встановлено, що відвідуваність підлягає експоненційному закону розподілу із важкими хвостами. Це означає, що серед значень відвідуваностей достатньо часто спостерігаються значення, які аномально сильно відрізняються від середнього значення. Така особливість породжує необхідність відсіву аномально високих збурень, що може бути здійснене за допомогою методів ковзаючого середнього.

Після встановлення структури модель повинна бути ідентифікована. З цією метою необхідно отримати експериментальні дані із динаміки відвідуваності реального Веб-сайту. Нас буде цікавити в першу чергу не кількість відвідувачів, а інтенсивність їх роботи із Веб-сайтом. Тому мірилом відвідуваності ми вибираємо число переглядів сторінок всього Веб-сайту або окремої його тематики. Число переглядів можна аналізувати за допомогою аналізу журналу Веб-сервера Веб-сайту. Однак значно зручніше робити це за допомогою спеціалізованих Веб-сервісів, зокрема Веб-сервісу Google Analytics. Користувачам цього сервісу також пропонується підтримка рекомендацій щодо покращення відвідуваності Веб-сайту. Однак ці рекомендації здійснюються на основі статистики даного сайту, без використання прогностичних моделей. Внаслідок відтермінування дії заходів щодо підвищення відвідуваності такий підхід не можна вважати достатньо ефективним.

Після отримання експериментальних даних необхідно вибрати метод ідентифікації моделі. В даний час широко відомі класичні методи ідентифікації: найменших квадратів, максимальної правдоподібності, стохастичної апроксимації [20]. Розглянемо перший із них. Раніше згадувалося, що відвідуваність підлягає експоненціальному розподілу із важкими хвостами. Однак за рахунок відсіву екстремальних значень закон розподілу випадкової величини вдається наблизити до нормального і ідентифікація моделі методом найменших квадратів дає задовільні результати.

Більшість із згаданих методів для побудови коефіцієнтів моделі спирається на задачу нелінійної оптимізації із складною цільовою функцією,

компоненти якої є елементами розв'язку систем нелінійних диференціальних рівнянь. Такі складні цільові функції, як правило, є багатоекстремальними і вимагають пошуку точки не локального, а глобального екстремуму, або точки, близької до глобального екстремуму. При цьому можна використати один із альтернативних підходів [20]. Серед них варто згадати методи випадкового пошуку. Однак цей метод в процесі своєї реалізації вимагає складних налаштувань величини кроку та коридору випадкового пошуку. Дещо простішим є використання градієнтного методу оптимізації Левенберга-Марквардта, який, однак, потребує підбору початкового наближення в околі точки глобального екстремуму. Розглянемо особливості даного методу детальніше.

Нехай розглядається задача мінімізації функціоналу, який отримується внаслідок застосування методу найменших квадратів [46]:

$$\min_{x \in R^n} f(x) = \frac{1}{2} \sum_{i=1}^m F_i(x) \quad (1.3)$$

Згідно методу Левенберга-Марквардта напрямком \vec{d}^k пошуку екстремуму на k -му кроці знаходиться із наступної системи лінійних алгебраїчних рівнянь:

$$[J(\vec{x}_k)^T J(\vec{x}_k) + \lambda_k] \vec{d}^k = -J(\vec{x}_k) F(\vec{x}_k) \quad (1.4)$$

де $J(\vec{x}_k)$ - $m \times n$ матриця якобіану функцій F_1, \dots, F_m , λ_k - параметр, що підбирається на кожному кроці.

При прямуванні λ_k до 0 співвідношення методу наближаються до методу Ньютона, а при зростанні λ_k - до методу найшвидшого спуску. Для вибору напрямку зміни параметру λ_k прогноують значення функціоналу, які можуть бути отримані методами найшвидшого спуску або Ньютона. Прогноз згідно уточнень по методу найшвидшого спуску здійснюється за допомогою лінійно прогнозованого значення найменших квадратів:

$$f_p(\vec{x}_k) = J(\vec{x}_{k-1}) \vec{d}^{k-1} + f_p(\vec{x}_{k-1}). \quad (1.5)$$

Прогноз $f_k(\bar{x}_*)$ згідно уточнень по методу Ньютона здійснюється за допомогою методу кубічної апроксимації значень $f(\bar{x}_{k-1})$ та $f(\bar{x}_{k-2})$. Якщо прогноз за лінійним законом дає краще значення

$$f_p(\bar{x}_k) < f_k(\bar{x}_*), \quad (1.6)$$

то значення λ_k збільшується:

$$\lambda_k = \lambda_{k-1} + \frac{f_k(\bar{x}_*) - f_p(\bar{x}_k)}{\alpha^*}. \quad (1.7)$$

В протилежному випадку значення λ_k зменшується згідно співвідношення:

$$\lambda_k = \frac{\lambda_{k-1}}{1 + \alpha^*}. \quad (1.8)$$

Якщо досліджувана характеристика представляється системою диференціальних рівнянь, то для обчислення одного значення функції $F_i(\bar{x})$ необхідно розв'язати цю систему. Якщо ці системи моделюють перехідні процеси із високими та низькими значеннями показників, вона буде володіти властивістю жорсткості. Для розв'язання такого класу систем використовуються, зокрема, методи Гіра.

Методи Гіра застосовуються до розв'язання задачі Коші для систем жорстких диференціальних рівнянь виду [33] :

$$\begin{aligned} \bar{y}' &= \bar{f}(\bar{x}, \bar{y}), \\ \bar{y}(\bar{x}_0) &= \bar{y}_0. \end{aligned} \quad (1.9)$$

Для побудови методу Гіра m -го порядку необхідно попередньо побудувати розв'язок системи (1.9) в $m-1$ вузлі, що можна зробити за допомогою явних методів Рунге-Кути. Далі за допомогою інтерполяційного многочлена Ньютона m -го порядку будують наближення розв'язку в околі вузла \bar{x}_m . Для побудови многочлена Ньютона використовують вузли $\bar{x}_0 - \bar{x}_{m-1}$ та значення розв'язку $\bar{y}_0 - \bar{y}_{m-1}$. Продиференціювавши згаданий многочлен Ньютона, отримаємо наближення похідної розв'язку в околі вузла \bar{x}_m . Прирівнявши це представлення при $\bar{x} = \bar{x}_m$ із значенням $\bar{f}(\bar{x}_m, \bar{y}_m)$, отримаємо

систему нелінійних рівнянь для побудови значення розв'язку \bar{y}_m . Ця система розв'язується за допомогою методу Ньютона. Для отримання початкового наближення для цього методу розглядається раніше побудоване наближене представлення розв'язку рівняння системи (1.9) у вузлі \bar{x}_{m-1} , з якого явним чином визначається початкове значення розв'язку \bar{y}_m . Реалізований таким чином неявний метод розв'язання жорстких систем диференціальних рівнянь дозволяє будувати розв'язки без надмірного подрібнення кроків дискретизації по часу, як це спостерігається для явних методів.

З'ясувавши один із можливих підходів до ідентифікації моделі відвідуваності, зупинимося на особливості організації ідентифікації із врахуванням дії людського фактора в людинно-машинних системах. Стосовно відвідуваності Веб-сторінок це означає, що реалізація впливів на відвідуваності в різні періоди часу здійснюється різними командами працівників, або одними і тими ж командами, досвід яких, як і досвід цільової аудиторії, невпинно збагачується. Це приводить до того, що параметри моделей факторів впливів на відвідуваність будуть змінюватися в залежності від періоду їх реалізації. Тому ідентифікацію таких моделей слід здійснювати в процесі їх спостереження по етапах. При цьому спостережені значення на попередньому етапі використовуватимуться для ідентифікації параметрів моделі на наступному етапі.

Проаналізовані особливості процесу моделювання відвідуваності Веб-сайту дають можливість сформулювати постановку задачі на побудову моделі відвідуваності, що буде здійснено в наступному підрозділі.

1.4. Постановка задачі дослідження

Проведений аналіз дозволив виявити значну роль відвідуваності в забезпеченні результативності низько відвідуваних Веб-сайтів а також ключову роль контенту у формуванні такої відвідуваності. Діяльність із підвищення відвідуваності Веб-сайту віднесено до компетенції служби його

підтримки. Проведений аналіз літературних джерел та досвіду практичних розробок дозволив побудувати схему підтримки функціонування Веб-сайтів, представлену на рисунку 1.4.

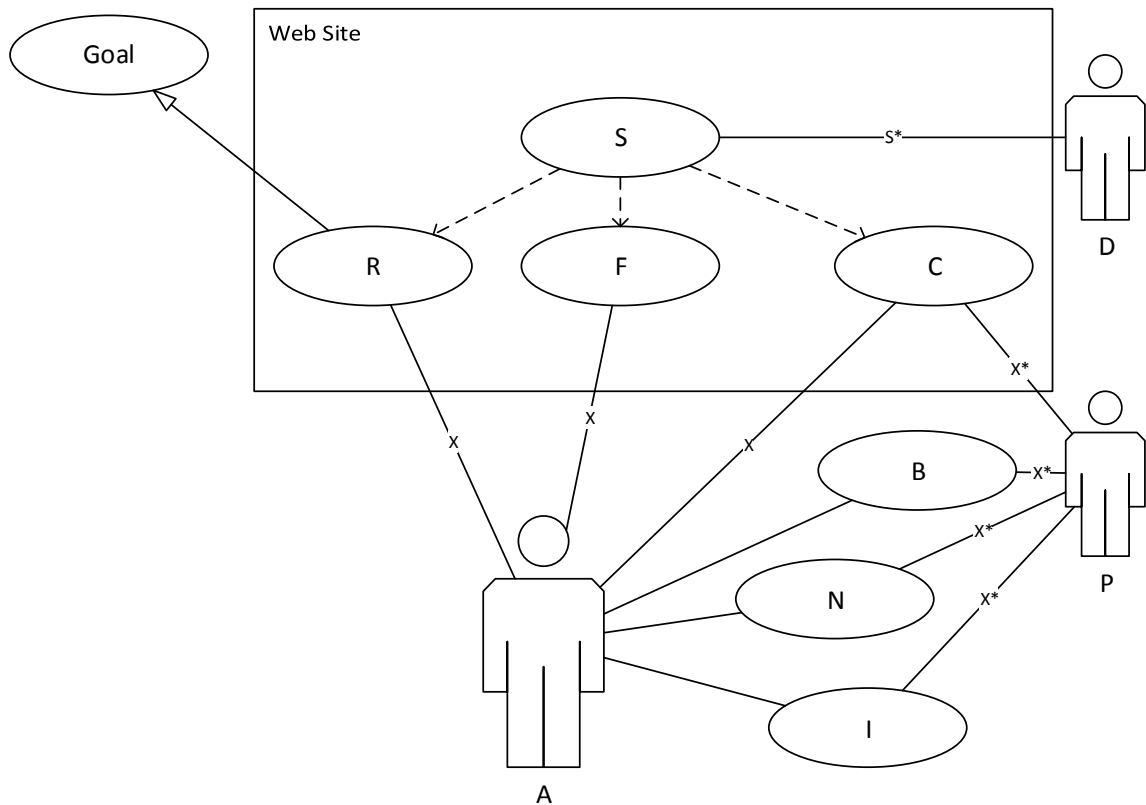


Рис. 1.4. Схема підтримки функціонування Веб-сайтів

Служба розробки D створює сайт S, з певним контентом C, продуктами і послугами F та цільовими сторінками R для досягнення мети Goal. Служба підтримки поновлює наповнення Веб-сайту, реструктуризуючи його структуру, та здійснює рекламування, з метою підняття відвідуваності та результативності. Засобами реклами як правило виступають публікації B в спеціалізованих ресурсах Вебу, анонсування N на дружніх Веб-сайтах а також направлені e-mail розсилання I. При цьому ефективність описаних заходів часто прогнозується на основі суб'єктивних оцінок X^* відповідних приростів відвідуваностей.

Зазначений суб'єктивізм в розвитку Веб-сайту може завадити йому досягти потенційних показників відвідуваності. Для цього необхідно

підкріпити суб'єктивну інтуїцію розробників об'єктивними рекомендаціями, сформованими на основі відповідних інформаційних та математичних моделей. Побудова напрямків поповнення вмісту та розробки нових структурних елементів Веб-сайту може бути значно спрощена із використанням Веб-онтологій T , які формуються на основі закономірностей, виявлених в предметній області функціонування Веб-сайту. Щоб бути затребуваною, подана інформація повинна бути актуальною. Однак із врахуванням різного рівня інформованості користувачів Веб-сайту, вона повинна також бути і повною, тобто спиратися на концептуалізацію даної предметної області. Ці характеристики обумовлюють використання онтологій для підтримки консультаційних функцій Веб-сайтів. Щоб оновлення інформації по обраних рубриках спричиняло підвищення загальної відвідуваності, варто забезпечити хорошу навігабельність Веб-сайту та активність цільової аудиторії на згаданих цільових рубриках. Для підвищення зручності навігації по Веб-сайту його структура повинна бути добре зрозумілою представникам аудиторії відповідної тематики.

Щоб забезпечити достатню відвідуваність розроблюваних тематик, їх наповнення повинно задовольняти критеріям актуальності та повноти. Систематизувати зусилля по формуванню наповнення тематик відповідно до згаданих критеріїв можна на основі спеціальних тезаурусів.

Заходи для підвищення відвідуваності вимагають певних затрат. Перед реалізацією доцільно оцінити їх ефективність, зокрема за допомогою математичних моделей. Прогнозування відвідуваності Веб-сайтів вже частково досліджено в літературі, де аналізуються питання навігації користувачів по сторінках Веб-сайту з метою виявлення найчастіше відвідуваних маршрутів, а також питанням вибору ефективних Веб-сайтів та Веб-сторінок для розміщення реклами. В той же час розробники низьковідвідуваних Веб-сайтів зацікавлені в рекомендаціях іншого плану, зокрема щодо напрямів та темпів розвитку Веб-сайту, які б забезпечили максимізацію його відвідуваності.

Служба підтримки поновлює наповнення Веб-сайту та здійснює його рекламу, з метою підняття відвідуваності. Це спричиняє підвищення відвідуваності інших тематик Веб-сайту в цілому і, зокрема, відвідуваності цільових сторінок, які безпосередньо направлені на виконання Веб-сайтом своїх завдань. Одним із факторів, що сприяють такій відвідуваності є регулярне поповнення тематик Веб-сайту актуальним контентом. Ефективність такого поповнення можна оцінювати за допомогою формалізованих математичних моделей, що дозволить також прогнозувати ефективність подібних операцій в майбутньому.

Якщо вдасться побудувати залежність (модель) M відвідуваностей X , то з'явиться можливість прогнозування результатів варіантів розвитку Веб-сайту ще до їх реалізації. Це дасть можливість вибрати ефективні шляхи такого розвитку. Відповідна удосконалена схема підтримки функціонування ділових Веб-сайтів представлена на рис. 1.5.

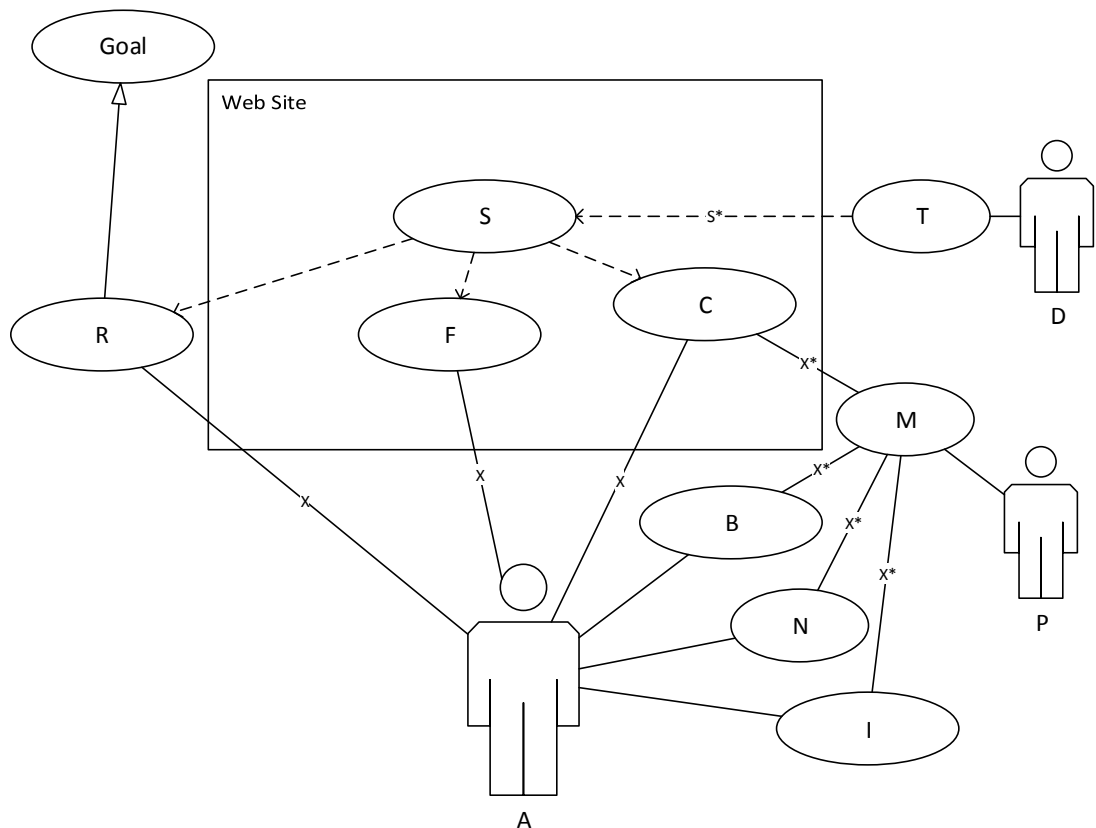


Рис. 1.5. Пропонована схема підтримки функціонування Веб-сайтів

Підсумовуючи, відзначимо, що проведений аналіз дозволяє сформулювати підвищення відвідуваності Веб-сайтів шляхом розроблення процедур структурування контенту із використанням програмних засобів та інструментів математичного моделювання. Це породжує мету дисертаційного дослідження — розроблення математичних та програмних засобів, що використовують структурування контенту та базуються на математичному моделюванні процесів функціонування Веб-сайтів для підвищення їх відвідуваності. Досягнення поставленої мети забезпечується шляхом розв'язання наступних задач дослідження:

- дослідження особливостей методів та засобів підтримки відвідуваності Веб-сайтів із використанням математичних моделей;
- розроблення моделей та методів автоматизованої побудови структур Веб-сайтів;
- побудова методу ідентифікації математичних моделей короткотермінового прогнозу відвідуваності Веб-сайтів;
- створення методу ідентифікації математичних моделей довготермінового прогнозу відвідуваності Веб-сайтів;
- застосування розроблених методів ідентифікації моделей та структурування контенту для підвищення відвідуваностей Веб-сайтів на прикладі Веб-сайтів.

Висновки до розділу 1

1. Проаналізовані фактори, що впливають на ефективність Веб-сайтів. Відзначено важливу роль відвідуваності в системі згаданих показників, а також роль пошукових систем в її формуванні. Відзначено провідну роль контенту Веб-сайту як одного із ключових факторів його високої відвідуваності.

2. Проаналізовано відомі методи та засоби процесу формування контенту Веб-сайту. Описано класифікацію методів Web Mining, як

основного напрямку автоматизації пошуку, структуризації і аналізу інформації в Вебі. Проаналізовані класи методів OntoWeb та засоби їх підтримки. Встановлено відсутність методів автоматизації процесів структурування контенту Веб-сайтів на основі Веб-онтологій.

3. Проаналізовано відомі методи моделювання систем, як засіб підвищення ефективності поповнення змісту Веб-сайтів. Виділено класи моделей процесів підвищення відвідуваності Веб-сайтів та базові підходи до побудови методів їх ідентифікації.

4. На основі проведеного аналізу побудовано існуючу та запропоновано нову схему підтримки функціонування Веб-сайтів, яка передбачає формування онтологій для формалізацію процесу структурування контенту, а також динамічних математичних моделей для оптимізації напрямків та темпів оновлення контенту. Сформульовано задачу дослідження, що полягає у підвищенні відвідуваності Веб-сайту за рахунок розробки процедур структурування контенту із використанням Веб-онтологій а також засобів математичного моделювання у вигляді детермінованих методів ідентифікації систем диференціальних рівнянь.

РОЗДІЛ 2. МОДЕЛІ ТА МЕТОДИ АВТОМАТИЗОВАНОЇ ПОБУДОВИ СТРУКТУРИ ВЕБ-САЙТІВ

Структура Веб-сайту є базою для формування його контенту, що є основним фактором підтримки його відвідуваності. Критерії хорошої навігабельності вимагають стандартизації структури Веб-сайтів для відповідних предметних областей а критерій корисності контенту вимагає її оригінальності та системності покриття певних тематик. Задоволення цих критеріїв за допомогою неформалізованих емпіричних оцінок вимагає значних затрат робочого часу експертів на виконання значного обсягу рутинних операцій.

В даному розділі пропонується метод частотного аналізу структурних елементів тематичних Веб-сторінок та формування онтологій шляхом аналізу зашумленої слабо структурованої інформації тематичних Веб-сторінок спеціалізованих Веб-сайтів, що уможлиблює автоматизацію процесів формування оновлення контенту Веб-сайтів. Особливості даного методу описані в роботах [29-32; 118,119, 124] та аналізуються в ході подальшого викладу.

2.1. Обґрунтування методів, засобів та математичних моделей відвідуваності Веб-сайтів

Для забезпечення привабливості Веб-сайту, його сторінки повинні містити актуальну інформацію з точки зору цільової аудиторії. Представники цієї аудиторії мають різну поінформованість про функціонування об'єкту, представленого Веб-сайтом та різні актуальні інформаційні потреби. Тому інформація Веб-сайту повинна бути достатньо різноманітною. Для спрощення доступу до неї Веб-сайт повинен мати певну структуру. Вона є базою для побудови контенту. Елементи контенту Веб-сайту відрізняються як за тематикою, так і за ступенем та способом впливу на цільову аудиторію. Їх аналіз уможливив виокремлення основних функцій сайту: представницька,

надання інформаційних послуг, надання он-лайн послуг. Представницька функція забезпечується звичайним вмістом сторінок Веб-сайту і відображає особливості функціонування об'єкту. Інформаційні послуги надають за рахунок тематичних колекцій інформаційних сторінок або інформаційних ресурсів. Он-лайн послуги надають за допомогою спеціальних он-лайн сервісів. Якщо представницька функція, як правило, не може збільшити аудиторію Веб-сайту, то дві останніх можуть суттєво її розширити. Кожен із засобів реалізації функцій Веб-сайту має певну вартість створення та ефективність дії. Вони можуть доповнювати або виключати одна одну і по різному діяти на різних представників цільової аудиторії.

Незалежно від функціонального призначення тематичного Веб-сайту, першим етапом його розробки є створення узагальненої моделі структури цільового сегменту як основи для формування його інформаційного наповнення, а також побудови процедури ідентифікації інтенсивності представницької функції сайту на певній цільовій аудиторії. Фактично вказана модель є онтологічним деревом, яке відповідає вимогам цільової аудиторії. Об'єктивно для побудови такої моделі можливо застосувати два шляхи:

- аналіз вмісту численних Веб-сторінок — результатів запитів до пошукових серверів, тобто за аналогією до найпоширеніших Веб-сайтів, які є привабливими для даної цільової аудиторії;
- опитування експертів.

Обидва шляхи вимагають розробки у певній мірі формального методу синтезу онтологічного дерева. В основу методу синтезу структури Веб-сайту покладемо багатокроковий пошук та упорядкування концептів онтологічного дерева.

Представники аудиторії цільового Веб-сегменту аналізують численні Веб-сайти, відзначаючи подібності та відмінності їхньої структури. Користувачі шукають потрібну для них інформацію, але прагнуть, щоб шлях цього пошуку був максимально простим та звичним. Тому доцільно основу структури Веб-сайту будувати із найпоширеніших структурних елементів Веб-

сайтів цільового сегмента. Формалізуємо метод побудови такої основи. При інтуїтивній специфікації структури розроблюваного сегмента дослідник переглядає структури кількох конкуруючих Веб-сайтів. Недолік такого аналізу полягає у великих витратах та можливості упущення за масивами повторюваної інформації унікальних структурних особливостей. Використання інформаційних технологій дозволяє забезпечити повноту аналізу великої множини структур Веб-сайтів з уникненням перегляду великого обсягу однотипних сторінок.

При цьому необхідно вирішити проблеми: автоматичного відбору сторінок Веб-сайтів із певної предметної області, виділення елементів головного меню, відбір найчастіше згадуваних елементів, впорядкування відібраних елементів. Принцип пропонованого розв'язання цих проблем представлено на рис. 2.1. Для вибору Веб-сайтів із певної предметної області використовуються засоби автоматизованого звертання до існуючих пошукових систем та автоматизованої обробки результатів запитів. Належність Веб-сторінки до предметної області підтверджується наявністю ключових слів, що маркують предметну область в титульному списку сторінки.



Рис. 2.1. Діаграма використання процедури побудови типового меню

Зміст пунктів меню встановлюється за анкерами у спискових структурах. Серед пунктів меню вибираються ті, частота появи яких перевищує деякий пороговий рівень. Для відібраних пунктів меню обчислюються їхні взаємні ранги на сторінках появи. За критерієм коефіцієнту конкордації Кендалла відбираються ті, позиції яких у сукупності вибраних сторінок є узгодженими. Таким чином формується типове меню Веб-сайту певної предметної області.

Для зацікавлення Веб-спільноти до окремим Веб-сайтом його інформаційне наповнення повинно відповідати критеріям актуальності та унікальності. Саме вимога унікальності ускладнює формалізацію методів побудови його наповнення. Очевидно, що вона може бути застосована лише до окремих аспектів такої побудови. Для виділення цих аспектів аналізуємо підходи до задоволення запитів інформаційних потреб Веб-спільноти [31, 32].

В момент усвідомлення інформаційної потреби користувач прагне отримати інформацію, яка сприятиме задоволенню цієї потреби. В процесі заглиблення в дане питання виявляється множинність джерел інформації, і на передній план виступає її достовірність та повнота. Для розв'язання складних інформаційних проблем огляду, оцінки та синтезу потрібної інформації із множини доступних ресурсів необхідні значні часові затрати. З цими проблемами постійно зустрічаються користувачі неспеціалізованих інформаційних ресурсів, тобто користувачі середовища Веб 1.0.

Із появою спеціалізованих Веб-сайтів та співдружностей користувачів в соціальних мережах, тобто елементів технології Веб 2.0, гострота цієї проблеми значно зменшилася. Однак, на рівні останнього підходу, все-таки доводиться вести пошук експертних суджень, аналізувати їх та концентруватися на окремих спеціалізованих сайтах, якість наповнення яких може бути нерівномірною. Тому перед засобами технології Веб 3.0 стоїть завдання виключити прив'язку до окремих Веб-сайтів для отримання доступу до достовірної та повної інформації. Цю прив'язку повинні взяти на себе

служби Семантичного Вебу із застосуванням різного роду онтологій [1, 8]. На основі описаного аналізу побудовано схему варіантів формування контенту Веб-сайту, представлену на рис. 2.2.

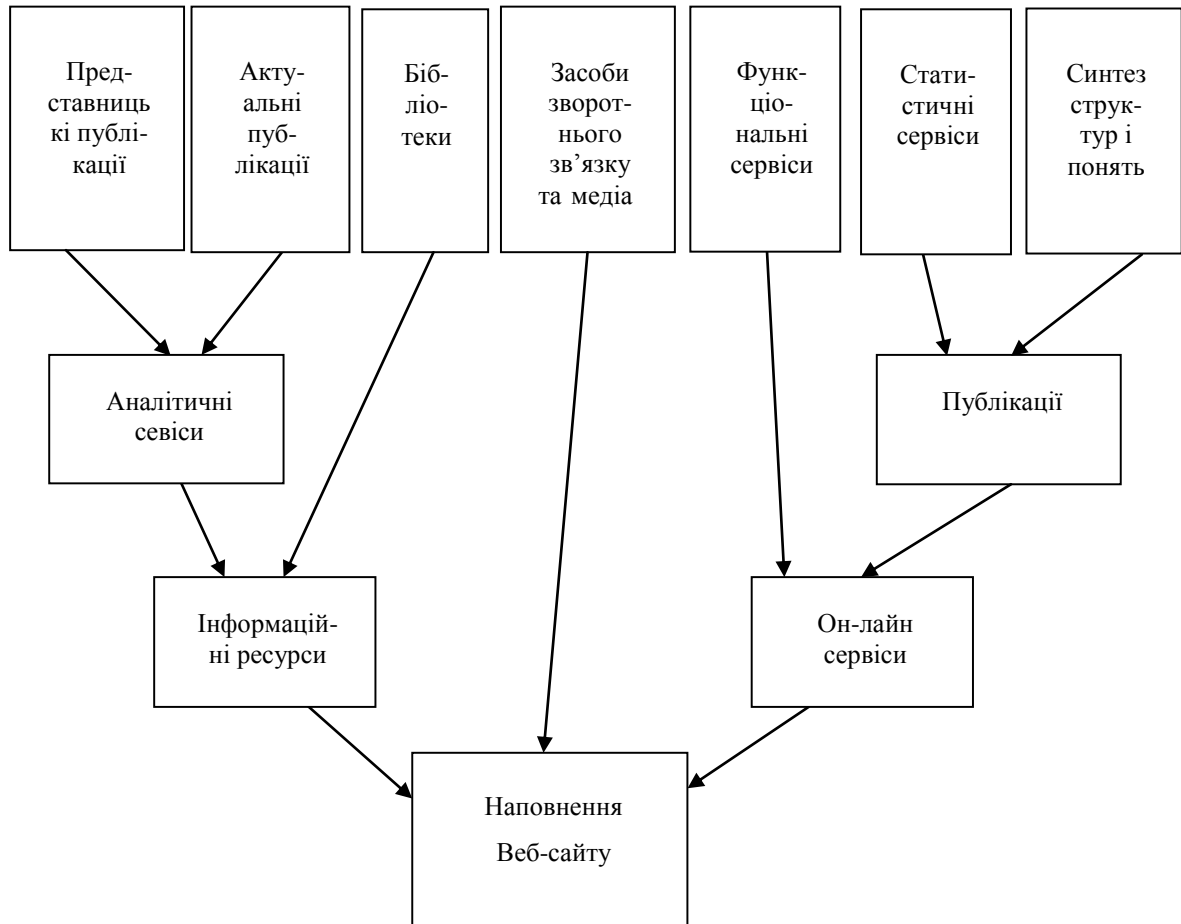


Рис. 2.2. Схема варіантів формування контенту Веб-сайту

Серед інформаційних ресурсів, що складають основу технології Веб 1.0 виділяємо публікації, які подають актуальну інформацію, як про об'єкт, представлений Веб-сайтом, так і про пов'язані з ним питання, що цікавлять певний сегмент Веб-спільноти. В процесі формування наповнення Веб-сайту редактори використовують різні Веб-ресурси, наводячи посилання на найважливіші та найцінніші з них. Тим самим вони формують бібліотеку Веб-сайту, або її каталог. Частина ресурсів сайту відводиться для реалізації засобів зворотнього зв'язку із користувачами та засобів медіа-впливу на них.

Створення такого контенту вимагає копiтких зусиль розробників, актуальність роботи яких швидко знижується.

Бiльш продуктивним є пiдхiд iз застосуванням он-лайн сервiсiв, якi надають користувачам рiзноманiтнi iнформацiйнi послуги, як у здiйсненнi певних функцiй при роботi iз Веб, так i при аналiзi тематичного наповнення Веб-простору. Такий пiдхiд є одним iз базових технологiй Веб 2.0.

Статистичнi сервiси, спираючись на вiдповiдне програмне забезпечення, збирають тематичну iнформацiю iз доступних Веб-сайтiв i заносять iх у власну базу даних та перiодично оновлюють її. При потребi вони подають узагальнену iнформацiю на основi ранiше нагромаджених даних. При цьому виникають проблеми вибору структури бази даних, побудови запитiв для отримання iнформацiї iз доступних тематичних Веб-сайтiв, встановлення вiдповiдностей мiж атрибутами Веб-сайтiв та полями бази даних, а також проблеми максимальної незалежностi коду сервiсу вiд особливостей структур даних тематичних бази та Веб-сайтiв.

В наступну групу, яка ще не набула широкого розповсюдження, видiлимо синтезуючi сервiси, якi формують певнi тематичнi онтологiї, тим самим розбудовуючи Семантичний Веб. В залежностi вiд того, яку iнформацiю вони обробляють (структурну iз формалiзованих меню або видiлену iншими методами iнформацiю Веб-сторiнок), ми подiляємо iх на сервiси синтезу структур та понять. Понятiйнi онтологiї ефективно використовують для аналізу формалiзованих характеристик понять, що зазнають динамiчних структурних змiн. Структуру загальних понять визначають через систему пов'язаних часткових понять iз урахуванням багатомовностi та синонiмiчностi можливих мовних iх реалiзацiй. Тому необхідно спроектувати структури даних, якi дозволяли б створювати, зберiгати та модифiковувати онтологiї. Окрiм пiдтримки технологiй Web 3.0, описанi сервiси можуть використовуватися у традицiйних Веб-сайтах для пiдтримки структурування iхнього контенту.

Після розробки онтологічних структур можна здійснювати їх наповнення та використання. Однак створення онтологій на основі лише експертних суджень є достатньо трудомістким із важко контрольованою прийнятністю для окремих користувачьких спільнот та суб'єктивізмом у проведенні їх реструктуризації. Поряд із цим зовсім виключити втручання експертів можна лише у випадку, коли варіанти онтологічних структур уже реалізовані і потребують лише систематизації та узагальнення. Така ситуація зустрічається при аналізі структур Веб-сайтів, але для формалізації змісту понять на основі Веб-контенту вона не актуальна. Тому необхідно розробити метод формування онтологічного наповнення із мінімізацією зусиль залучених експертів.

Пропонований метод ґрунтується на наступних основних положеннях. Отримати актуальні соціально-значимі характеристики предметних областей, що динамічно розвиваються, можна за допомогою аналізу інформації, представленої у Вебі. Ця інформація отримується зі спеціалізованих Веб-сайтів, які здійснюють моніторинг запитів на відповідні види продукції або послуги. Згадані запити містять слабо структуровану та неструктуровану інформацію. Слабко структурована інформація міститься у списках, елементи яких структуруються довільним чином. Характеристики предметних областей, що містяться у слабо структурованих списках, можуть бути виявлені на основі частотного аналізу і використані для формування наповнення предметних онтологій. Базою для формування структури онтологій, а також метаонтологічних понять, є Веб-анотації характеристик предметних областей. Структурування онтології відбувається експертним шляхом, із мінімізацією завантаженості та вимог до кваліфікації експерта.

Аналіз аргументів запропонованої моделі якості Веб-сайту показує необхідність чіткого планування та контролю структури його контенту. Наповнення окремих сторінок Веб-сайту повинно супроводжуватися нагромадженням певних допоміжних інформаційних ресурсів та їх аналітичним переосмисленням. В іншому випадку матеріали сторінки не

нестимуть нової інформації та не концентруватимуть її. Тоді потреби у звертанні до такої сторінки не буде. Окрім корисної інформації, ресурс на якому розміщена сторінка, повинен бути відомим широкому колу відвідувачів. Досягти цього можна за допомогою реклами, реалізація якої вимагає затрат певних ресурсів. Реалізації дій по наповненню Веб-сайту та його рекламуванню багатоваріантні, їх віддача обмежена та розподілена. Для оцінки наслідків реалізації застосованої тактики створення контенту сторінок або їх рекламування необхідно вміти прогнозувати розподіл її наслідків в часі. Це дозволить вибрати місце та інтенсивність відповідної діяльності, раціонально застосовуючи наявні трудові ресурси.

Заходи по підвищенню відвідуваності реалізуються протягом тривалого часу згідно методик, які можуть бути відпрацьованими або перебувати на стадії формування. Як в реалізації методики, так і в реакції на неї аудиторії Веб-сайту, значну роль відіграє суб'єктивний фактор. Оцінити результативність цієї діяльності зазвичай можна лише після завершення її активної стадії. Маючи прогноз динаміки процесу на початковій його стадії, можна скоректувати тактику реалізації даного виду діяльності, або вчасно спланувати нову посилюючу дію. Це породжує необхідність побудови прогнозної моделі відвідуваності Веб-сайту.

Як відзначалося раніше, основним фактором росту відвідуваності Веб-сайту є суттєве удосконалення його контенту, яке отримує резонанс серед цільової аудиторії. Ступінь цього резонансу, що характеризує якість контенту локальної тематики, вимірюємо її відвідуваністю, тобто числом переглядів відповідних Веб-сторінок на протязі дня. Щоденна відвідуваність характеризується великою кількістю випадкових факторів, що робить проблематичним достатньо точно її прогнозування. В той же час, на характеристики Веб-сайту впливають не окремі екстремальні, а усереднені показники. Тому в нашій моделі аналізуватимемо середньотижневі відвідуваності.

При реалізації даного підходу виникає ряд проблем. Перша із них полягає у зборі та зручному доступі до інформації щодо відвідуваності Веб-сайту. Наступні проблеми полягають у виборі структури математичної моделі відвідуваності Веб-сайту та методів її параметричної ідентифікації. З метою їх розв'язання запропоновано процедуру побудови моделей відвідуваності, загальну взаємодію процесів якої представлено на рис. 2.3.

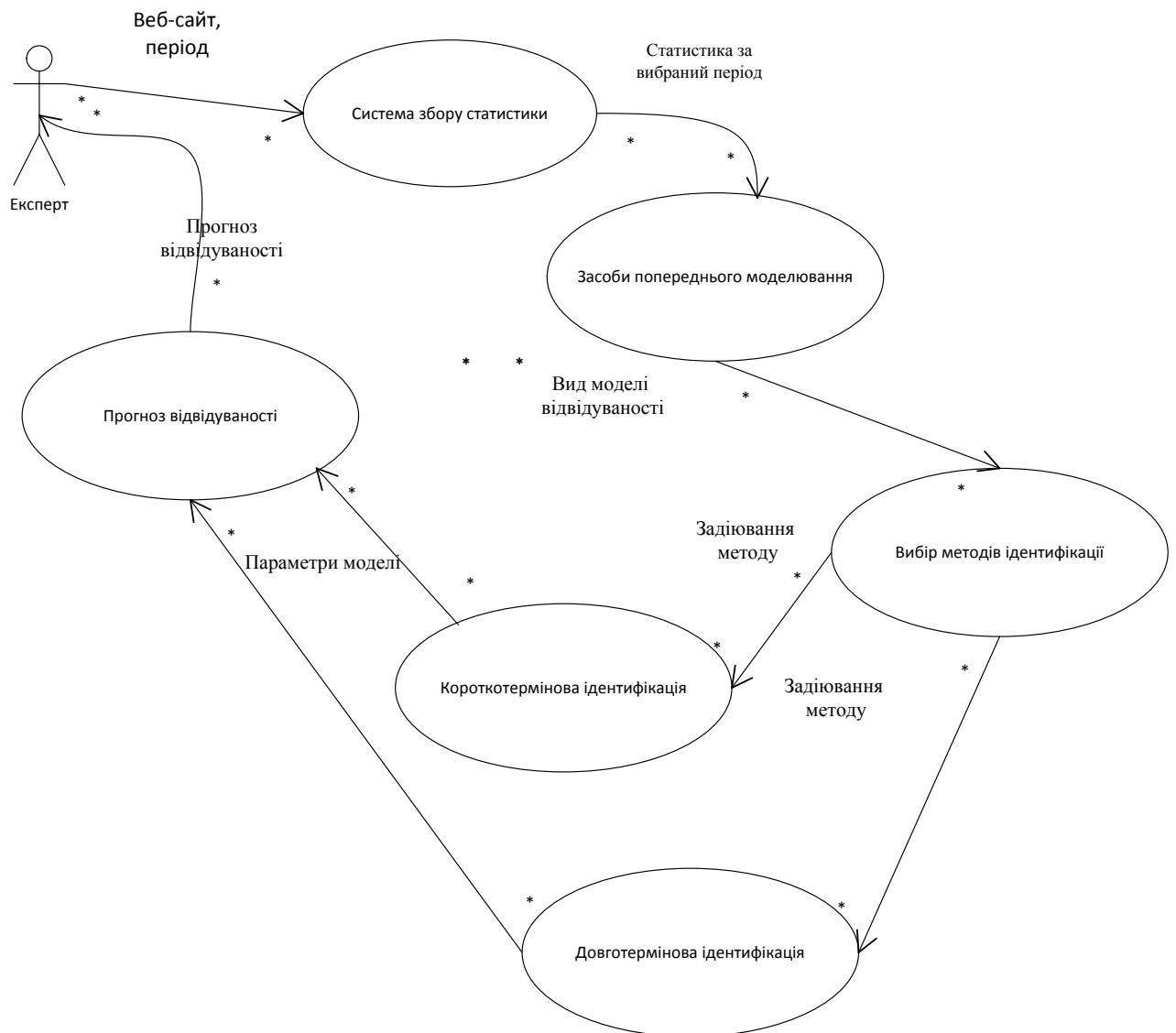


Рис. 2.3. Діаграма використання процедури побудови моделей відвідуваності

Ця процедура передбачає використання спеціальної системи збору статистичної інформації про відвідуваність. В якості такої системи може бути використаний безкоштовний Веб-сервіс Google Analytics. Після збору необхідної статистики можна приступити до вирішення проблеми побудови структури моделі відвідуваності, використовуючи з цією метою засоби попереднього моделювання. В даному дослідженні в якості таких засобів використана модель відвідуваності Веб-сайтів із нетривалими приростами.

В цій моделі якість реструктурування контенту тематик Веб-сайту оцінювалася за допомогою простої дискретної моделі прив'язаної до бізнес-процесів об'єкта, який представляє Веб-сайт. В результаті попередніх оцінок вдалося встановити адекватність степеневого зв'язку між відвідуваністю Веб-сайту та функцією якості оновлення контенту, який використано у всіх наступних моделях.

Функцією якості оновлення контенту побудовано на основі відвідуваності тематик Веб-сайту, що корелюють із ростом відвідуваності всього Веб-сайту, а значить її обумовлюють. Коли на відвідуваність Веб-сайту впливає багато різних факторів, то ріст відвідуваності окремих тематик обумовлюється, в значній мірі, оновленням її контенту. Це дозволяє ріст відвідуваності тематик Веб-сайту трактувати як функцію підвищення якості його контенту.

Аналіз літературних джерел виявив закон розподілу випадкової величини відвідуваності Веб-сайтів. В цій якості виступає експоненціальний закон із важким хвостом. Це обумовлює використання методів згладжування відвідуваності для ліквідації впливу аномальних викидів внаслідок присутності в розподілі важкого хвоста. Після згладжування статистичних даних відвідуваності виявилася подібність профілю функції якості оновлення контенту до профілю росту системи мікроорганізмів, що моделюється системою диференціальних рівнянь Моно. В поєднання із степеневою залежністю між функцією якості оновлення контенту та приростом

відвідуваності, згадана модифікована система утворює структуру запропонованої моделі відвідуваності Веб-сайтів.

Після вибору моделі відвідуваності необхідно розробити метод її параметричної ідентифікації. Складність ідентифікації моделі даного виду полягає у врахування впливу людського фактора. Це обумовлює необхідність адаптації моделі до умов перебігу її конкретної реалізації. Така адаптація здійснюється розбиттям інтервалу росту відвідуваності на певні етапи. На початку кожного етапу по спостережених раніше значеннях функції якості зміни контенту прогнозуються параметри її моделі на весь період. У випадку коли тривалість етапу співмірна із кількістю параметрів моделі, опиняємося в ситуації короткотермінової ідентифікації, в іншому випадку - в ситуації довготермінової ідентифікації.

В кожній із виділених ситуацій розробляється свій метод ідентифікації. Справа в тому, що у випадку довготермінового приросту відвідуваності вдається виділити етапи лінійної поведінки модельованої величини або її похідної, а у випадку короткотермінового приросту таких ділянок встановити не вдається. В останньому випадку спостерігається подібність профілю функції якості для різних її реалізацій. Така подібність проявляється в поведінці другої похідної функції якості оновлення контенту, що використано у формуванні процедури ідентифікації в даному випадку.

У випадку довготермінового приросту відвідуваності спостережень на початку етапу ідентифікації недостатньо для забезпечення прийнятної точності. Тому у функціоналі якості ідентифікації використовуються не лише ці спостережені значення, а і їх екстраполяції. Сукупність запропонованих заходів повинна приводити до підвищення відвідуваності підтримуваних Веб-сайтів. Детальніше особливості запропонованих підходів аналізуються в ході подальшого викладу.

Перш ніж поповнювати вміст Веб-сайту, необхідно проаналізувати та оптимізувати його структуру. Питання такого аналізу розглядаються в наступному підрозділі.

2.2. Метод побудови структур тематичних Веб-сайтів на основі частотного аналізу існуючих структур

Формалізуємо метод побудови типового головного меню тематичного Веб-сайту. Нехай тематику таких сайтів задано загальним текстовим маркером SAM предметної області та маркером SS її специфікатора. За допомогою пошукового сервера та операції конкатенації символічних стрічок, за тематичним запитом $QS = SAM \& SS$ отримаємо множину SP Веб-сторінок, представлених своїми $HTML$ -кодами

$$SP(QS, P) = \{HP_i\}_{i=1}^P, \quad (2.1)$$

де P — потужність множини SP ; HP_i — елемент множини, що визначає $HTML$ -код i -тої сторінки.

Для подальшого аналізу доцільно використовувати лише сторінки підмножини SPR , в яких предметна область є об'єктом аналізу, а не елементом ширшого контексту. Критерієм виконання цієї вимоги є наявність маркера предметної області в заголовку сторінки:

$$SPR = \left\{ HP_{i^*}^* \mid HP_{i^*} \in SP, HP_{i^*}.title \cap SAM \neq \emptyset \right\}_{i=1}^{P^*}. \quad (2.2)$$

Із $HTML$ -коду сторінки необхідно вибрати інформацію, яка характеризує її структуру. У першу чергу інформація про структуру представлена в меню даної сторінки. Зміст пунктів меню встановлюємо на основі анкорів тегів у вигляді " $\langle a \dots href \dots \rangle$ ", які групуються в ієрархічні списки. Вибрані елементи верхнього рівня утворюють впорядкований список $LRA(HP_i^*)$ анкорів сторінки

$$LRA(HP_i^*) = \langle A_{ij}(HP_i^*) \rangle_{j=1}^{AP_i}. \quad (2.3)$$

Якщо список анкорів відсутній, то інформацію про структуру вибираємо з виділених елементів сторінки. Інформація такого роду є впорядкованим

списком $LRB(HP_i^*)$ виділених елементів сторінки :

$$LRB(HP_i^*) = \langle B_{ik}(HP_i^*) \rangle_{k=1}^{BP_i} . \quad (2.4)$$

Звичайно, такі списки можуть містити і випадкову інформацію. Однак елементи таких списків, які повторюються, дають інформацію про засоби структурування інформації цільового сегменту. Тому на основі списків анкорів та виділених елементів сформуємо звичайну SCF та узагальнену GCF множини пар концепт - частота концепту, які визначаємо у такий спосіб:

$$SCF = \left\{ (C_l, NC_l) \mid C_l \in \bigcup_i LRA(HP_i^*) \right\}, \quad (2.5)$$

$$GCF = \left\{ (C_m, NC_m) \mid C_m \in \left(\bigcup_i LRA(HP_i^*) \right) \cup \left(\bigcup_i LRB(HP_i^*) \right) \right\}. \quad (2.6)$$

Для виявлення концептів релевантних до предметної області, впорядкуємо елементи множини SCF у порядку спадання частот елементів. При цьому частину понять одразу включаємо в концептуальну множину SCN

$$SCN = \left\{ (C_l, NC_l) \mid FCL_l = \frac{NC_l}{P^*} > F_0 \right\}, \quad (2.7)$$

де величина F_0 , як правило, належить інтервалу $[0.2; 0.5]$, а конкретне значення вибирають виходячи зі специфіки дослідження. Якщо таким чином поповнити множину концептів не вдалося, то здійснюємо відбір серед концептів із низькою частотою. Розглянемо впорядковану вибірку частот SSF

$$SSF = \{ NC_l \mid (C_l, NC_l) \in SCF \}. \quad (2.8)$$

Найвищі частоти цієї вибірки перевіряємо на аномальність за критерієм 4σ [44]. Концепти, що відповідають виявленим аномальним значенням, включаємо в множину SCN . Якщо і далі ця множина залишається пустою, то збільшуємо потужність множини сторінок для аналізу, виданих пошуковим сервером. Коли при цьому множина SCN не поповниться, то робимо висновок,

що базові концепти даної предметної області не розпізнано.

Якщо вдається поповнити концептуальну множину SCN , то її концепти C_l включаємо в такий запит до пошукового сервера:

$$QS = SAM \ \& \ \{C_l \mid (C_l, NC_l) \in SCN\} \ \& \ SS. \quad (2.9)$$

Після цього повторюємо спроби поповнення множини концептів на основі нового уточненого запиту та вище описаної процедури. При остаточному уточненні набору K концептів, що специфікують дану предметну область, їх необхідно впорядкувати. Для цього, для кожної сторінки, список анкорів якої містить концепти предметної області, проводимо їхнє ранжування. При цьому рангові номери виставляємо лише концептам. Так R_{ik} позначає номер (ранг) k -го концепту на i -тій сторінці. Необхідно також враховувати нерівносильність впливу $HTML$ сторінок на аудиторію предметної області, оскільки багато користувачів переглядають перші 10 – 20 елементів видачі пошукових серверів, і лише в дуже нечисленних випадках — сторінки із другої сотні списку. Для врахування цього факту потрібно використати деяку монотонно спадну вагову функцію $w(i)$.

Оскільки для перших 10-20 сторінок їх важливість плавно спадає, а далі це спадання значно прискорюється, в якості такої функції зручно використовувати кубічний сплайн. Для її однозначного визначення необхідно накласти хоча б 4 умови. Зокрема покладаємо важливість першої сторінки рівною 1 при нульовій похідній для цього аргументу, а також деякі характерні ваги наприклад для 20-ї та 100-ї сторінок, які вибираються експертним шляхом. На основі описаного підходу вводимо нормовану систему ваг

$$G(i) = \frac{g(i)}{\sum_i g(i)}.$$

Після цього усереднений ранг k -го концепту по предметній області обчислюємо за допомогою наступного співвідношення:

$$RA_k = \sum_i R_{ik} G(i) \quad . \quad (2.10)$$

На основі усереднених рангів і відбувається ранжування концептів. Узгодженість рангів оцінюється за коефіцієнтом конкордації Кендалла [10, 38]:

$$W = \frac{12 \sum_{k=1}^K \left(\sum_{i=1}^I R_{ik} G(i) - \bar{R} \right)}{I^2 (K^3 - K)}, \quad (2.11)$$

де $\bar{R} = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^I R_{ik} G(i)$. За умови, що

$$I(K-1)W > \chi_{K-1, \alpha}^2 \quad , \quad (2.12)$$

ранжування вважається значущим [44].

Якщо ранжування по повному списку концептів не можна вважати значущим, то усуваємо по одному із списку в порядку зростання їхніх ваг, тобто сумарних ваг сторінок, де вони зустрічаються. Усунення починаємо від концепта із найменшою вагою і продовжуємо аж до отримання значущого ранжування. У такому разі відібрану множину концептів розбиваємо на дві частини: множину RG із погодженими рангами та NR , рангування по якій встановити не вдалося.

На основі наведених теоретичних положень сформуємо алгоритм синтезу структури Веб-сайту:

1. Встановлюємо лічильник ітерацій $CI = 1$ та формуємо підмножини SPR множини SP HTML-кодів Веб-сторінок на основі запиту $QS = SAM \ \& \ SS$ для яких маркер SAM предметної області міститься в назві коду.
2. Формуємо множину SCF частот концептів на основі анкорів тегів, а якщо вона пуста, то узагальнену GCF множину частот концептів на основі виділених елементів Веб-сторінок.
3. Формуємо концептуальну множину SCN предметної області за критерієм перевищення мінімальної частоти F_0 та оцінюємо її

потужність PSC .

4. Якщо концептуальна множина порожня, поповнюємо її елементами із аномальними частотами.
5. Якщо $PSC = 0$ та $CI = 1$, то завершення алгоритму. Якщо потужність концептуальної множини зменшилась в порівнянні з попереднім кроком, то перехід на пункт 6, інакше збільшуємо лічильник ітерацій $CI = 1$ та формуємо підмножини SPR множини $SP HTML$ -кодів Веб-сторінок на основі початкового запиту QS , поповненого елементами концептуальної множини даного кроку і переходимо на пункт 2.
6. Ранжуємо елементи найпотужнішої із побудованих множин SCN та аналізуємо значущість коефіцієнту конкордації цього ранжування. При незначимості коефіцієнта конкордації послідовно вилучаємо із множини концепти із найменшими вагами аж до отримання значущості згаданого коефіцієнта.

Запропоновані метод та алгоритм дозволяють виявляти базову структуру тематичних Веб-сайтів, яку доцільно врахувати при проектуванні структури конкретних ресурсів [30]. Однак ефективність Веб-сайту забезпечує його якісне наповнення, яке необхідно постійно оновлювати. Метод формалізованої структуризації напрямків такого наповнення описано в наступному підрозділі.

2.3. Моделі контенту Веб-сайту на основі слабо структурованої інформації

Під онтологіями розуміємо деревоподібну систему соціально-значимих понять (концептів) певної предметної області. Сюди включаються як концепти, що використовуються в робочих документах предметної області, так і поняття, що їх узагальнюють, тобто мета-поняття. Тому онтологію моделюємо наступною деревоподібною структурою

$$O_Str = \langle IdCn, Pr Cn, Meta \rangle, \quad (2.13)$$

яка включає ідентифікатори $IdCn$ концептів, ідентифікатор $PrCn$ батьківського концепта та атрибут $Meta$, який дозволяє відділити робочі поняття предметної області від її мета-понять. При цьому загальне поняття відрізняється від часткових відсутністю батьківського, для якого $PrCn = NULL$. Кожне із понять допускає різні лінгвістичні представлення у вигляді словосполучень O_Phr . Слова, що визначають онтологічні поняття представляються в конкретних словоформах O_Frm а також своїми основами O_Bs . Словоформи використовуються для представлення понять користувачам, а основи – для автоматичної ідентифікації еквівалентності мовних представлень. Атрибути введених понять згрупуємо в наступні структури:

$$O_Bs = \langle IdLg, IdBs, WBase \rangle, \quad (2.14)$$

$$O_Frm = \langle IdLg, IdFm, IdBs, WForm \rangle, \quad (2.15)$$

$$O_Phr = \langle IdCn, IdLg, IdPh, IdBs, IdFm, IdPrBs \rangle, \quad (2.16)$$

де $IdLg$ ідентифікатор мови реалізації, $IdBs$ ідентифікатор основи слова, $WBase$ основа слова, $IdFm$ ідентифікатор форми слова, $WForm$ форма слова, $IdCn$ ідентифікатор концепта, $IdPh$ ідентифікатор фрази, $IdPrBs$ ідентифікатор батьківської основи поняття.

Для формування онтологічного наповнення, значимого для певного сегменту Веб-аудиторії, зручно використати описи, представлені на відповідних Веб-сторінках. З метою підвищення значимості такої інформації для аналізу необхідно експертним шляхом відбирати лише певні спеціалізовані Веб-сайти. Наповнення сторінок таких Веб-сайтів формується для сприйняття користувачами, а тому не є строго структурованим за певними жорсткими правилами. Окрім того, на цих сторінках розташовано багато додаткової інформації, яка з точки зору онтологічного наповнення може розглядатися як шум. Варто вимагати також, щоб інформація на Веб-сторінках була структурована, а не просто розбита на параграфи чи абзаци. Така вимога

дозволяє значно звужувати сферу пошуку, тим самим піднімаючи його ефективність. В даному випадку під структурованістю мається на увазі оформлення інформації у вигляді спискових структур.

Для відбору структурованої інформації, що стосується даної предметної області, формуємо множину ключових термінів KWS , що її характеризує. Для підтримки аналізу вмістимого Веб-сторінок створено наступну допоміжну структуру AS :

$$AS = \langle IdPg, IdLst, IdIt, IdBs, IdFm, IdPrBs \rangle, \quad (2.17)$$

де $IdPg$ — ідентифікатор аналізованої Веб-сторінки, $IdLst$ — ідентифікатор списку сторінки, $IdIt$ — ідентифікатор елемента списку.

Повторюваність мовного представлення є його важливою характеристикою, що дозволяє відділити значимі представлення від несуттєвої інформації даної предметної області. Для її контролю вводиться наступна структура BF частот основ:

$$BF = \langle IdBs, BsFr, IdLPg, Phn \rangle \quad (2.18)$$

де $IdBs$ — ідентифікатор основи, $BsFr$ — частота появи основи на різних Веб-сторінках, $IdLPg$ — ідентифікатор останньої із Веб-сторінок, де зустрічалася основа, Phn — маркер фоновості поняття, що приймає значення невизначеності $NULL$ за замовчуванням.

При аналізі $HTML$ -коду чергової Веб-сторінки спеціалізованого Веб-сайту встановлюємо її ідентифікатор:

$$CurPgId := \max(\pi_{IdLPg}(BF)) + 1. \quad (2.19)$$

Далі виділяємо елементи $LSTIt$ її спискових структур, що наповнюють теги $\langle li \rangle$. Елементи списку розбиваються на елементарні поняття It із використанням роздільників, які утворюють спеціальну множину сепараторів:

Якщо основа слова $Wrd_k(I_{t_{Pg,Lst}})$ не розпізнана, вона вноситься в список основ, саме слово вноситься в список словоформ, а поповнення відношення частот основ здійснюється згідно співвідношення (2.24).

Для включення в онтологію експерту пропонуються лише основи, частота яких буде перевищує деяке мінімальне значення $BF0 \geq 2$, яке вибирається користувачем. Експертові пропонується список основ для включення в онтологію, коли вищезгаданій умові задовольняють не менше $BC0$ основ:

$$\text{Список основ} \quad (2.25)$$

Для прийняття адекватного рішення, основи пропонуються в тому контексті, в якому вони зустрічаються на Веб-сторінках. Це дає змогу виділяти поняття, які складаються із кількох слів, а також не пропонувати повторно основи, які не вибрані експертом для включення в онтологію при аналізі попередніх контекстів. Основу для формування контексту $CBase$ вибираємо за критерієм її максимальної частоти MFR :

$$MFR \quad (2.26)$$

$$CBase \quad (2.27)$$

Для формування контексту вибраної основи використовується двовимірний масив $ACont$, що містить ідентифікатори словоформ. Його перший індекс позначає номер фрази контексту, а другий – номер слова в цій фразі. Кількість фраз контексту $NPhC$ встановлюється наступним чином:

$$NPhC \quad (2.28)$$

Ідентифікатори елементів аналізованої структури заносяться в допоміжний одновимірний масив $APhId$:

$$APhId = \pi_{IdIt}(\sigma_{IdBS=CBase}(AS)) \quad (2.29)$$

розмірності $NPhC$. І-та стрічка контекстного масиву формується за правилом:

$$ACont[i] = (\pi_{IdFm}(\sigma_{IdIt=APhN[i] \wedge IdPrBs=NULL}(AS)) AS HL) \cup (\pi_{IdFm}(\sigma_{IdIt=APhN[i] \wedge (IdPrBs=HL.IdBs \vee IdPrBs=HL1.IdBs)}(AS)) AS HL1) \quad (2.30)$$

Масив словоформ подається експертові для вибору елементів онтологічного наповнення. Для вибору однієї словоформи потрібно просто вибрати її представлення у поданому елементі інтерфейсу (рис. 4.8) програмної реалізації. Основи, які ввійшли в онтологію отримують фоновий індекс $Phn := 2$, щоб повторно не подаватися для аналізу. Основи, які не були вибрані жодного разу не можуть складати основи контекстів, вони помічаються як фонові $Phn := 1$.

Після вибору елементів онтологічного наповнення, вони будуть включені в онтологічну ієрархію. Таке включення робиться експертом, який закладає свої знання в онтологію. Щоб зробити його вибір соціально-значимим та мінімізувати суб'єктивізм оцінювання, використовується механізм автоматизованого анотування понять, який підтримується інформаційною структурою *ConcA*:

$$\langle \langle IdCn, AnURI, AnTitl, Desc \rangle \rangle, \quad (2.31)$$

де $IdCn$ ідентифікатор концепта, $AnURI$ адреса Веб-сторінки анотації, $AnTitl$ тег title Веб-сторінки анотації, $Desc$ опис концепта.

Пошук анотації здійснюється на основі запиту до пошукового сервера, який включає перелік основ даного концепта. Із списку анотацій, впорядкованого пошуковою системою, вибирається сторінка, тег title якої містить слова із переліку словоформ. В анотацію включаються не більше 3-х абзаців із відібраної сторінки, які містять елементи запиту. Експерт здійснює перегляд множини відібраних концептів із їх анотаціями. Він редагує анотації а також отримувати нову анотацію, або переглядати всю сторінку, що містить анотацію.

На основі анотацій експерт формує опис поняття, відбирає підмножини вкладених понять за принципом "частина-ціле", впорядковує однорівневі поняття за принципом "від загального до конкретного" та "від простого до складного". На основі термінів анотації експерт групує поняття та вводить мета-поняття, що їх об'єднують. При великій кількості мета-понять формуються мета-поняття вищих порядків.

Представлені вище метод демонструє достатньо громіздкі підходи до переробки інформації, яка проте не виключає і активної роботи експерта. Викликає зацікавлення оцінка впливу згаданих підходів на ефективність роботи експерта. В якості критерію ефективності таких оцінок виберемо відношення кількості відібраних термінів NOI до кількості переглянутих стрічок NSE :

$$EE = \frac{NC}{NS}. \quad (2.32)$$

На основі наведених теоретичних положень сформуємо алгоритм автоматизованого формування онтологічного наповнення:

1. Встановлюємо перелік релевантних спеціалізованих сайтів StL а також множину KWS ключових слів мінімальної потужності, які характеризують найважливішу особливість предметної області.
2. Будуємо запит, що включає слова із множини KWS до кожного Веб-сайту із множини StL та формуємо множини $HTML$ -кодів Веб-сторінок із верхньої частини видачі пошукової системи.
3. Якщо сторінка містить список, тобто концентровану сукупність вимог користувача, хоча б один елемент якого містить слова із множини KWS , що описують один із термінів побудованої онтології, то елементи списку заносяться у відношення AS . При цьому вони розбиваються на елементарні терміни за допомогою роздільників із множини SS , а елементарні терміни розбиваються на слова. Основи відібраних слів заносяться у відношення BF , а якщо вони вже там зареєстровані із

сторінки, що не співпадає із поточною і також не належать до фону, їх кратність збільшується на 1 та оновлюється ідентифікатор сторінки реєстрації.

4. Якщо при зміні кратність основи перевищить деяке наперед задане значення $BF0$, кількість кандидатів на включення в онтологію OMC збільшується на 1. Якщо $OMC \geq BC0$, то контекст кандидатів на включення в онтологію подається експертові.
5. Для кожного входження основи-кандидата в відношення AS вибираються словоформи, що формують елемент списку, який містить дану основу. Атрибут Phn основи-кандидата у відношенні BF повинен бути невизначеним.
6. Після сформування контексту він подається експертові для аналізу. Після відбору експертом елементів для онтології, елементи списку, жодна компонента якого не була відібрана, помічаються значенням атрибуту $Phn := 1$ у відношенні BF для виключення їх повторної подачі в контексті іншого терміна.
7. Відібрані для онтології концепти поміщаються в її структуру на основі згенерованих анотацій.

Висновки до розділу 2

1. Обґрунтовано загальні підходи до розробки методів автоматизованого структурування та реструктурування Веб-сайтів, побудови математичних моделей їх відвідуваності а також методів ідентифікації цих моделей.

2. Запропоновано метод частотного аналізу структурних елементів тематичних Веб-сторінок, у якому використано формальний аналіз елементів структур Веб-сайтів, описаних у HTML-кодах цільових сторінок а також узгодженість виділеної структури типового Веб-сайту із структурами Веб-сайті певної предметної області за допомогою коефіцієнту конкордації. Це

дозволяє автоматизувати процес побудови типових структур Веб-сайту певної предметної області.

3. Запропоновано систему структур підтримки процесу формування онтології предметних областей, яка окрім базових понять включає контексти та анотації останніх. Це дозволяє формалізувати процес структурування онтологій, які використовуються для формування шляхів оновлення контенту Веб-сайту.

4. Запропоновано метод автоматизованого формування наповнення онтологій шляхом аналізу зашумленої слабо структурованої інформації спеціалізованих Веб-сайтів. В його основу покладено структурний аналіз тематичних сторінок спеціалізованих Веб-сайтів, відсів фонових термінів за частотним критерієм із формуванням контекстів найбільш вживаних понять. Застосування цього методу дозволяє зменшити навантаження на експерта при формуванні онтологій, що використовуються для шляхів оновлення контенту.

РОЗДІЛ 3. МЕТОДИ ІДЕНТИФІКАЦІЇ МАТЕМАТИЧНИХ МОДЕЛЕЙ ПРОГНОЗУ ВІДВІДУВАНOSTІ ВЕБ-САЙТІВ.

Суттєве підвищення відвідуваності Веб-сайту вимагає системних дій служби підтримки щодо оновлення контенту або рекламування його в мережі., що вимагає значних затрат робочого часу кваліфікованого персоналу. Результативність таких дій проявляється із значним часовим запізненням. Прогнозування динаміки покращення відвідуваності після таких системних акцій дозволило б раціонально використовувати трудові ресурси служби підтримки.

В даному розділі пропонується модель відвідуваності Веб-сайтів, побудованої на основі системи звичайних диференціальних рівнянь Моно. Для її ідентифікації запропоновані методи, що залежать від засобів, які закладаються службою підтримки в забезпечення росту відвідуваності. При цьому виділяються короткотермінові та довготермінові прирости відвідуваності. Особливості запропонованих методів розглядаються в роботах [26, 28, 120-123] а також в ході подальшого викладу.

3.1. Обґрунтування вибору структури математичної моделі відвідуваності Веб-сайтів.

3.1.1. Модель відвідуваності із дискретною функцією якості контенту

З метою спрощення аналізу розглянемо Веб-сайти, системне оновлення контенту яких здійснюється епізодично. Побудову моделей відвідуваності розпочнемо із визначення результуючих та вхідних параметрів. Змінними стану системи обираємо x_1 загальну відвідуваність сторінок Веб-сайту WS, як фактору його представництва в рейтингах по пошукових запитах, а також відвідуваність x_2 результуючих сторінок Веб-сайту як засобу поширення цільової інформації, що наочно представлено на рис. 3.1.

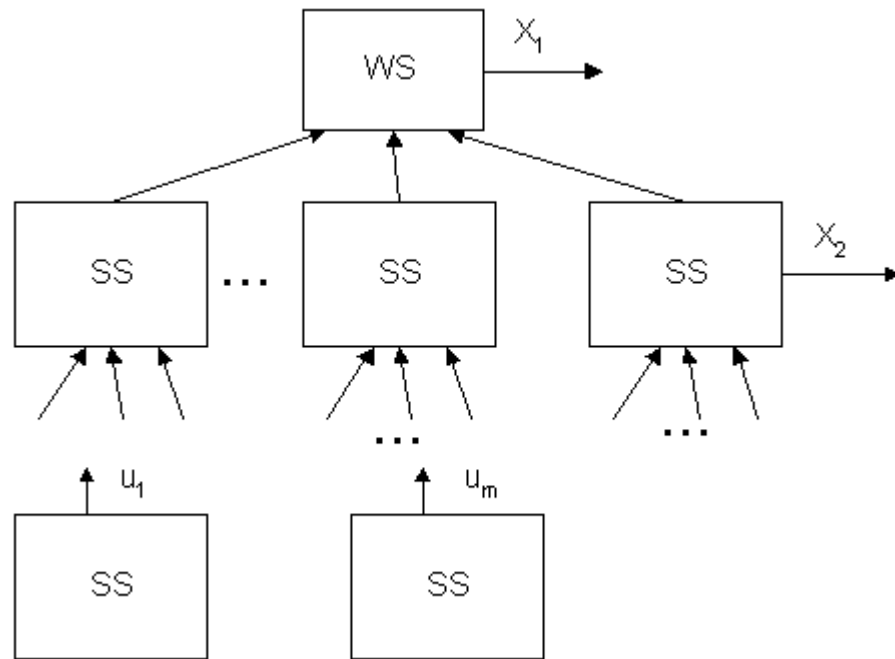


Рис.3.1. Роль параметрів моделі відвідуваності Веб-сайту

В ролі вхідних параметрів моделі виберемо функції якості контенту тематик Веб-сайту [26]. Значення цих функцій прирівнюємо до відвідуваності згаданих тематик, позначаючи їх через u_j . Динаміка відвідуваностей тематик залежить від інтенсивності їх поповнення. З метою спрощення моделювання таких відвідуваностей для Веб-сайтів, для яких оновлення здійснюється епізодично, динаміка відвідуваностей описується 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 послідовністю $S_{j,K} = \{s_{j,0}, \dots, s_{j,K}\}$, $s_{j,k} \in D$, де j - номер функції якості контенту, K - тривалість аналізованого періоду. Їм відповідають відвідуваності локальних тематик $v_j(s_{j,k}) = \hat{u}_{j,k}$ $k = \overline{1, K}$, значення яких встановлюються експериментально.

В таких послідовностях відвідуваності переходять із високих значень на момент помітного покращення якості контенту до помірних так низьких,

причому якість контенту може періодично посилюватися. При посиленні якості контенту спостерігається нелінійність у співвідношеннях між функціями якості та загальними і результуючими відвідуваностями. Пропонується моделювати цю нелінійність за допомогою врахування якості контенту в попередні моменти часу а також степеневій трансформації значень функції якості контенту. При цьому природно допустити, що інтенсивність дії підвищення якості контенту на відвідуваність знижується із бігом часу. Найпростіше такий ефект моделюється за допомогою експоненціальної функції з від'ємним показником.

Еволюцію Веб-сайту враховуємо за рахунок впровадження в моменти часу T_j нових елементів контенту, додаткова відвідуваність яких u_j суттєво впливатиме на загальну відвідуваність. Зокрема, на основі спостережень по першому часовому інтервалі до моменту появи другої модифікації контенту, модель відвідуваності включає лише запізнені значення відвідуваності $x_{i,k-1}$ та степеневу функцію від функції якості контенту $u_{1,k}$ в поточний момент часу

$$\tilde{x}_{i,k} = g_{i,1} x_{i,k-1} + q_{i,1} (u_{1,k})^{q_{i,2}}, \quad t < T_2. \quad (3.1)$$

Похибка моделі, побудованої по першому оновленню контенту при появі другого оновлення суттєво зростає і позначається як спричинена ним додаткова відвідуваність

$$z_{2,k}^i = x_{i,k} - g_{i,1} x_{i,k-1} - q_{i,1} (u_{1,k})^{q_{i,2}}. \quad (3.2)$$

При цьому модель відвідуваності дещо ускладнюється. Зокрема, окрім, побудованих на попередньому кроці компонент додасться нова компонента. Вона враховує лише значення відвідуваності, спричиненої другим оновленням контенту та степеневу трансформацію функції якості повторного оновлення контенту:

$$\begin{aligned} \tilde{x}_{i,k} = & g_{i,1}x_{i,k-1} + q_{i,1}(u_{1,k})^{q_{i,2}} + 1 \\ & + H(k - T_2)(g_{i,2}z_{2,k-1}^i + q_{i,3}(u_{2,k})^{q_{i,4}}), \quad k < T_3. \end{aligned} \quad (3.3)$$

При появі третього оновлення контенту спричинена ним відвідуваність оцінюється наступним чином

$$\begin{aligned} z_{3,k}^i = & x_{i,k} - g_{i,1}x_{i,k-1} - q_{i,1}(u_{1,k})^{q_{i,2}} - \\ & - H(k - T_2)(g_{i,2}z_{2,k-1}^i + q_{i,3}(u_{2,k})^{q_{i,4}}) \end{aligned} \quad (3.4)$$

Якщо відвідуваність формується під дією m оновлень контенту, отримаємо наступну узагальнену модель

$$\begin{aligned} \tilde{x}_{i,k} = & g_{i,1}x_{i,k-1} + q_{i,1}(u_{1,k})^{q_{i,2}} + \\ & + \sum_{j=2}^m H(k - T_j)(g_{i,j}z_{j,k-1}^i + q_{i,2j-1}(u_{j,k})^{q_{i,2j}}), \end{aligned} \quad (3.5)$$

$$\begin{aligned} z_{j,k}^i = & x_{i,k} - g_{i,1}x_{i,k-1} - q_{i,1}(u_{1,k})^{q_{i,2}} - \\ & - \sum_{l=2}^{j-1} H(k - T_l)(g_{i,l}z_{l,k-1}^i + q_{i,2l-1}(u_{l,k})^{q_{i,2l}}), \end{aligned} \quad (3.6)$$

де $z_{j,k}^i$ - прогноз i -ої результуючої змінної в момент часу k внаслідок впливу j - го оновлення контенту, який виник момент часу T_j ; H - функція Хевісайда, g, q - параметри моделі, які необхідно ідентифікувати. При оцінці відвідуваності $z_{j,k}^i$ спричиненої j -ою зміною контенту враховується дія $(j - 1)$ попередніх факторів за допомогою відвідуваностей виду $z_{l,k}^i$.

З метою вибору вірного напрямку подальших досліджень будемо ускладнювати попередню модель за допомогою окремих елементів, аналізуючи отримані значення похибок. На першому етапі доповнимо модель (3.5)-(3.6) взаємовпливом змінних стану. При цьому індексом i^* позначається альтернативна змінна, яка чинить вплив на поточну

$$\begin{aligned} \tilde{x}_{i,k} = & g_{i,1}x_{i,k-1} + g_{i,2}x_{i^*,k} + q_{i,1}(u_{1,k})^{q_{i,2}} + \\ & + \sum_{j=2}^m H(k - T_j) \left(g_{i,2j-1}z_{j,k-1}^i + g_{i,2j}z_{j,k-1}^{i^*} + q_{i,2j-1}(u_{j,k})^{q_{i,2j}} \right), \end{aligned} \quad (3.7)$$

$$\text{де } i^* = \begin{cases} 2 & \text{їдє } i=1, \\ 1 & \text{їдє } i=2, \end{cases}$$

$$\begin{aligned} z_{j,k}^i = & x_{i,k} - g_{i,1}x_{i,k-1} - g_{i,2}x_{i^*,k-1} - q_{i,1}(u_{1,k})^{q_{i,2}} - \\ & - \sum_{l=2}^{j-1} H(k - T_l) \left(g_{i,2l-1}z_{l,k-1}^i + g_{i,2l}z_{l,k-1}^{i^*} + q_{i,2l-1}(u_{l,k})^{q_{i,2l}} \right). \end{aligned} \quad (3.8)$$

На наступному етапі модель (3.5)-(3.6) доповнимо врахуванням двох попередніх значень змінних стану, отримуючи співвідношення

$$\begin{aligned} \tilde{x}_{i,k} = & g_{i,1}x_{i,k-1} + g_{i,2}x_{i,k-2} + q_{i,1}(u_{1,k})^{q_{i,2}} + \\ & + \sum_{j=2}^m H(k - T_j) \left(g_{i,2j-1}z_{j,k-1}^i + g_{i,2j}z_{j,k-2}^i + q_{i,2j-1}(u_{j,k})^{q_{i,2j}} \right), \end{aligned} \quad (3.9)$$

де

$$\begin{aligned} z_{j,k}^i = & x_{i,k} - g_{i,1}x_{i,k-1} - g_{i,2}x_{i,k-2} - q_{i,1}(u_{1,k})^{q_{i,2}} - \\ & - \sum_{l=2}^{j-1} H(k - T_l) \left(g_{i,2l-1}z_{l,k-1}^i + g_{i,2l}z_{l,k-2}^i + q_{i,2l-1}(u_{l,k})^{q_{i,2l}} \right). \end{aligned} \quad (3.10)$$

На завершальному етапі модель (3.5)-(3.6) доповнимо врахуванням змінності інтенсивності впливу параметрів управління за експоненційним законом, оскільки інтенсивність фактора впливу на змінних стану знижується з часом

$$\begin{aligned} \tilde{x}_{i,k} = & g_{i,1}x_{i,k-1} + q_{i,1}e^{-k \cdot q_{i,2}}(u_{1,k})^{q_{i,3}} + \\ & + \sum_{j=2}^m H(k - T_j) \left(g_{i,2j-1}z_{j,k-1}^i + q_{i,3j-2}e^{(k-T_j) \cdot q_{i,3j-1}}(u_{j,k})^{q_{i,3j}} \right), \end{aligned} \quad (3.11)$$

$$\begin{aligned} z_{j,k}^i = & x_{i,k} - g_{i,1}x_{i,k-1} - q_{i,1}e^{-k \cdot q_{i,1}}(u_{1,k})^{q_{i,3}} - \\ \text{де } & - \sum_{l=2}^{j-1} H(k - T_l) \left(g_{i,2l-1}z_{l,k-1}^i + q_{i,3l-2}e^{-(l-T_l) \cdot q_{i,3l-1}}(u_{l,k})^{q_{i,3l}} \right). \end{aligned} \quad (3.12)$$

Для ідентифікації моделі необхідно напрацювати відповідний статистичний матеріал. Його формуємо за допомогою інструменту Google Analytics, в якому аналізуємо засіб Content Drill Down. При цьому вибираємо тижневі характеристики відвідуваностей відповідних рубрик. Оскільки модель повинна володіти прогностичними властивостями, то управляючі параметри формуються як модельовані характеристики. Ці моделі будуємо на основі статистики відвідуваностей найактивніших рубрик та аналізу процесів, що їх супроводжують. При аналізі зміни станів в процесах, які супроводжують помітні зміни в динаміці показників u_j будуємо модель їх динаміки у вигляді дискретної функції

$$v_j(s_{j,l}) = \hat{u}_{j,l} \quad l = \overline{1, p_j}, \quad (3.13)$$

де $\{s_{j,l}\}_{l=1}^{p_j}$ - множина ідентифікованих станів показника u_j , p_j - кількість ідентифікованих станів показника u_j , $\hat{u}_{j,l}$ - усереднена оцінка показника u_j , коли він перебуває в стані $s_{j,l}$.

Прогнозуючи послідовність зміни станів $S_{j,K} = \{s_{j,0}, \dots, s_{j,K}\}$ показника u_j , отримуємо дискретну множину його прогнозованих значень

$$U_j = \{v_j(s_{j,0}), \dots, v_j(s_{j,K})\}. \quad (3.14)$$

Підставляючи множину прогнозованих значень замість спостережених в представлення (3.5)-(3.6), отримуємо прогностичну динамічну модель наступного виду:

$$\begin{aligned} \tilde{x}_{i,k} = & g_{i,1} x_{i,k-1} + q_{i,1} (v_1(s_{1,k}))^{q_{i,2}} + \\ & + \sum_{j=2}^m H(k - T_j) (g_{ij} z_{jk-1}^i + q_{i,2j-1} (v_j(s_{j,k}))^{q_{i,2j}}), \end{aligned} \quad (3.15)$$

$$z_{j,k}^i = x_{i,k} - g_{i,1}x_{i,k-1} - q_{i,1}(v_1(s_{1,k}))^{q_{i,2}} - \sum_{l=2}^{j-1} H(k - T_l) (g_{i,l}z_{l,k-1}^i + q_{i,2l-1}(v_l(s_{l,k}))^{q_{i,2l}}). \quad (3.16)$$

Аналогічні перетворення застосуємо до модельних співвідношень (3.7)-(3.12), подаючи результати моделювання у наступному згорнутому вигляді:

$$x_{i,k} = f_{i,k}^n(\bar{g}, \bar{q}, x_{i,k-1}, x_{i,k-2}, x_{i,2,k}, W), n = \overline{1,4}, \quad (3.17)$$

де n - порядковий номер запропонованої моделі, $W = \bigcup_j U_j$ - матриця прогнозованих значень відвідуваностей тематик, що розвиваються.

Узагальнимо дане співвідношення, ввівши вектор змінних стану \vec{x}_k для часової дискрети k , множину векторів змінних стану X_k для всіх часових дискрет, що передують $k+1$ дискреті а також множину векторів відвідуваностей тематик, що розвиваються для всього часового інтервалу аналізу. Тоді представлення (3.17) спроститься наступним чином

$$\vec{x}_k = F_k^n(\bar{g}, \bar{q}, X_k, W), n = \overline{1,4}. \quad (3.18)$$

Для ідентифікації згаданих моделей використаємо метод найменших квадратів. Нехай значення змінних стану в часових дискретах з номерами, що є елементами множин $K_A = \{k_1^A, \dots, k_{A_1}^A\}$ та $K_B = \{k_1^B, \dots, k_{B_1}^B\}$ утворюють відповідно навчальну та контрольну вибірки. Тоді критеріальна функція методу найменших квадратів матиме вигляд

$$V(\bar{g}, \bar{q}, \vec{X}_k | k \in K_A) = \sum_{k \in K_A} [\vec{x}_k - F_k^n(\bar{g}, \bar{q}, X_k, W)]^2. \quad (3.19)$$

Абсолютна та відносна похибка побудованої моделі оцінюється наступним чином:

$$E_k(n, K_A) = |\bar{x}_k - F_k^n(\bar{g}, \bar{q}, X_k, W)|, \quad k \in K_A \cup K_B, \quad (3.20)$$

$$\varepsilon_k(n, K_A) = E_k(n, K_A) / \Delta(n, K_A), \quad k \in K_A \cup K_B, \quad (3.21)$$

де $\Delta(n, K_A) = \max_k E_k(n, K_A)$. Основним критерієм точності моделі вважаємо її найбільшу відносну похибку.

Модель відвідуваності ідентифікувалася на даних, отриманих для сайту факультету комп'ютерних інформаційних технологій Тернопільського національного економічного університету (ФКІТ, <http://www.fcit.tneu.org/>) на протязі 16 тижневого періоду. Під час спостережень було виявлено одне суттєве оновлення контенту та задіяно одне наступне оновлення, тому даний період аналізу розбито на два півперіоди тривалістю по 8 тижнів.

На рис. 3.2 представлено профіль функцій якості оновлення контенту деякого Веб-сайту.

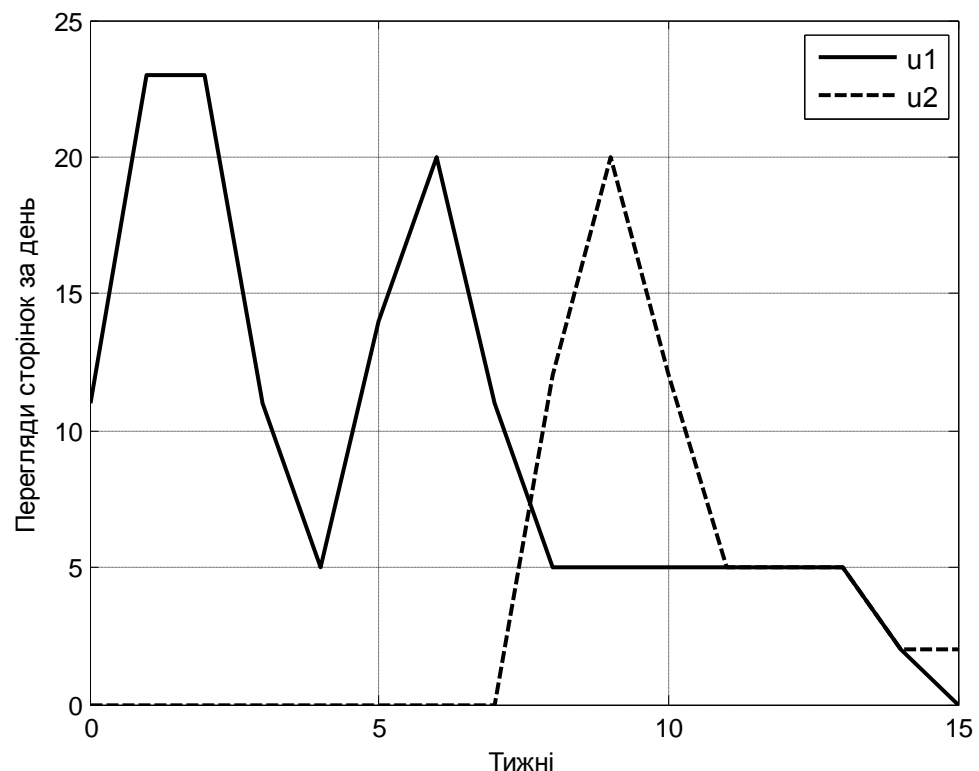


Рис.3.2. Профіль функцій якості оновлення контенту по двох періодах

Перший із яких відповідає оновленню інформації із структурами дисциплін та розкладом занять на початку семестру, термін якого для різних

курсів може відрізнятись. Друге оновлення контенту передбачає використання навчальних матеріалів, розміщених на Веб-сайті, для самостійної роботи студентів. Активність дії цього оновлення залежить від графіку проведення модульного контролю.

За критерієм мінімальної абсолютної похибки по першому періоду відібрано просту інерційну модель із врахуванням зміни інтенсивності впливу параметра управління за допомогою експоненціального закону. Результат моделювання представлено на рис. 3.3, а похибка ідентифікації відповідної моделі по півперіоду аналізу представлена на рис. 3.4. На останньому із згаданих рисунків показано розподіл похибки по тій частині аналізованого інтервалу, де вона приймає допустимі значення. Подальший значний ріст похибки свідчить про необхідність поповнення моделі.

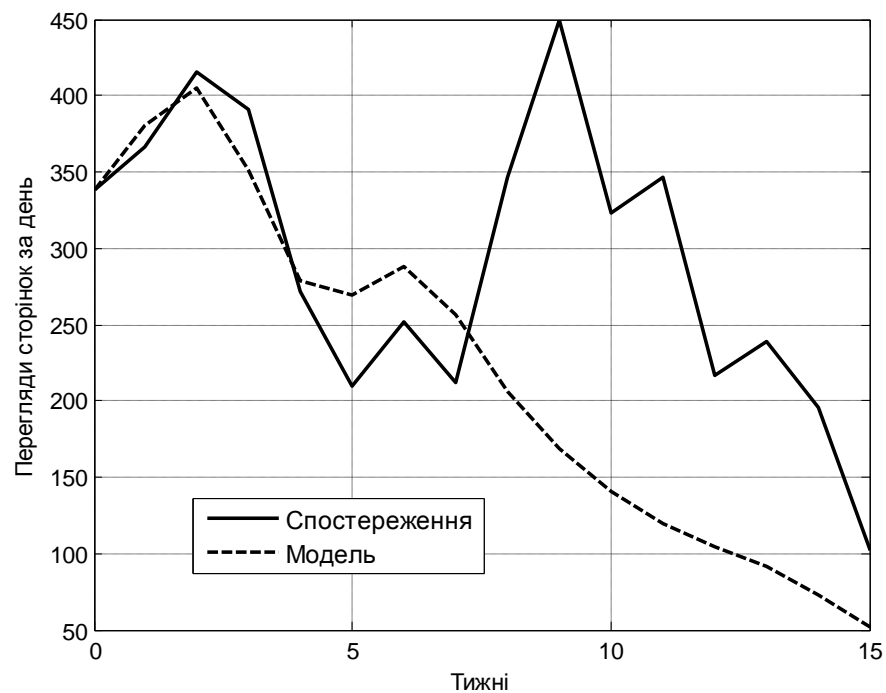


Рис.3.3. Моделювання загальної відвідуваності по першому оновленні контенту

Аналіз розподілу похибок засвідчує, що майже всі відносні похибки не перевищують 10%. Лише на протязі п'ятого тижня похибка прогнозу сягала

максимального значення більшого 10% межі. При виході за семитижневий інтервал похибка моделі стрімко зростає, делеко виходячи за межу 15%.

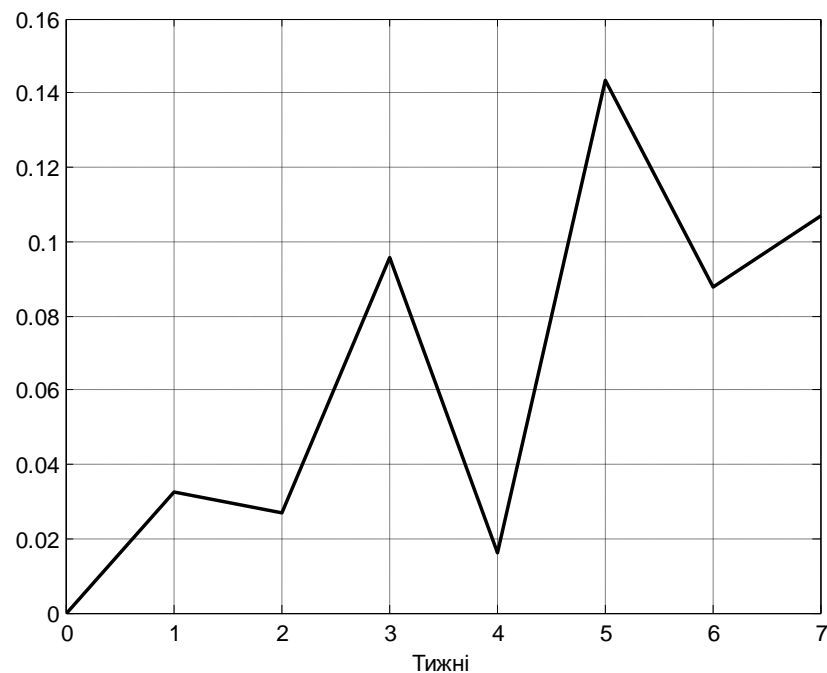


Рис.3.4. Розподіл відносних похибок ідентифікації моделі загальної відвідуваності по першому оновленню контенту

Це вказує на те, що модель повинна бути поповнена новим врахуванням оновлення контенту. В якості цього чинника вибираємо друге оновлення, яке полягає в активному використанні навчального контенту Веб-сайту. Активність останнього розпочинається якраз на сьомий тиждень аналізованого періоду. Тому можна сподіватися, що врахування другого оновлення дозволить адекватно прогнозувати відвідуваність вже по повному інтервалу аналізу. На рис. 3.5 та 3.6 представлено результати моделювання загальної відвідуваності для інтервалу аналізу вже по двох оновленнях контенту.

Максимальна відносна похибка моделі склала 13.06%. При цьому структура моделі для другого оновлення контенту аналогічна структурі моделі для першого оновлення. Загальний вигляд побудованої моделі подано співвідношенням (3.22).

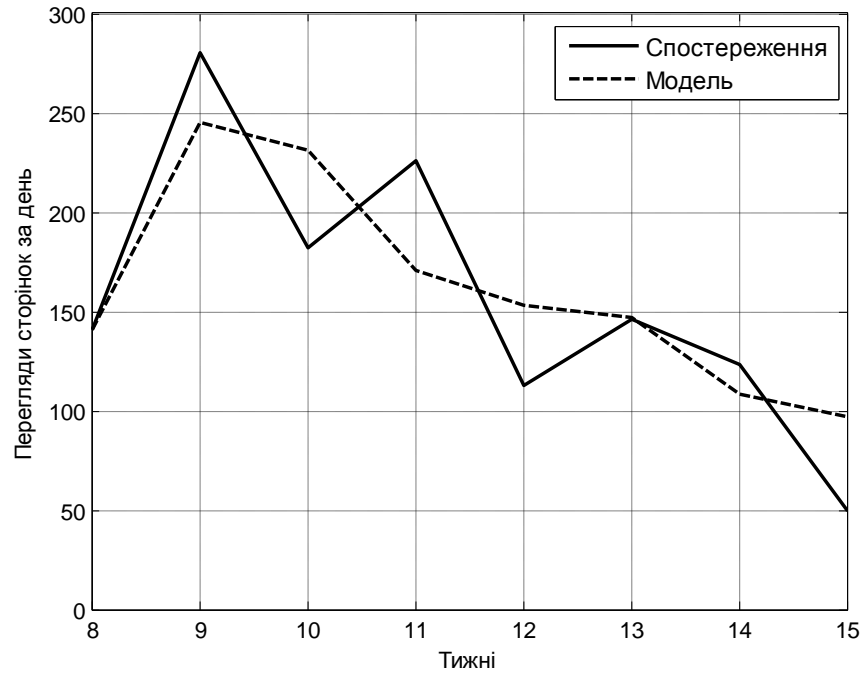


Рис.3.5. Моделювання загальної відвідуваності по двох оновленнях контенту

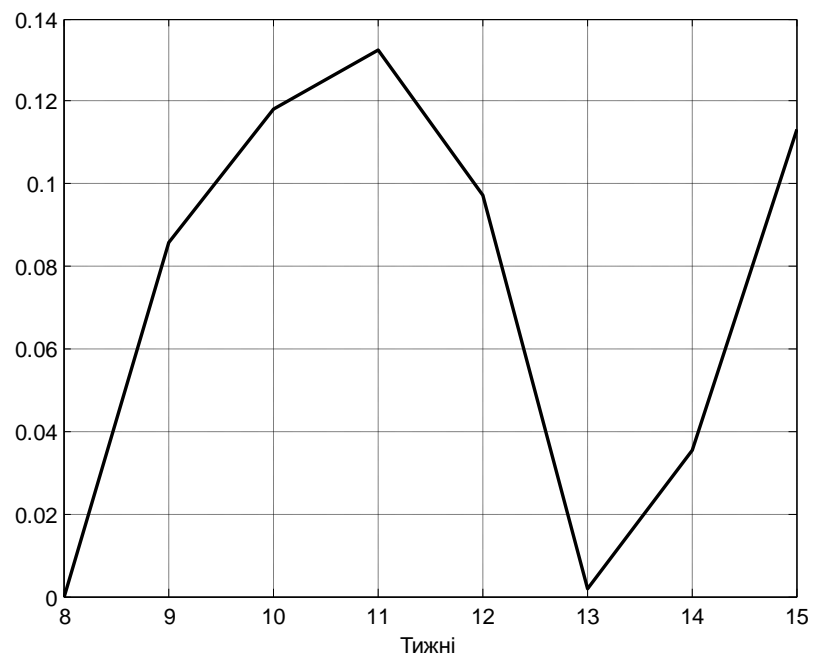


Рис.3.6. Розподіл відносних похибок ідентифікації моделі загальної відвідуваності по другому оновленню контенту

$$\begin{aligned}
 x_{1,k} = & g_{1,1}x_{1,k-1} + q_{1,1}e^{-k \cdot q_{1,2}} (v_1(s_{1,k}))^{q_{1,3}} \\
 & + H(t - T_2) \left(g_{1,2}z_{2,k-1}^1 + q_{1,4}e^{(k-T_2) \cdot q_{1,5}} (v_2(s_{2,k}))^{q_{1,6}} \right) (1 \pm \varepsilon_1) ,
 \end{aligned} \tag{3.22}$$

$$\text{де } z_{2,k}^1 = x_{1,k} - g_{1,1}x_{1,k-1} - q_{1,1}e^{-k \cdot q_{1,2}} (v_1(s_{1,k}))^{q_{1,2}},$$

$$g_{1,1} = 0.7168, \quad q_{1,1} = 6.9958, \quad q_{1,2} = 0.0482, \quad q_{1,3} = 0.9824, \quad T_2 = 7,$$

$$g_{1,2} = 0.7168, \quad q_{1,4} = 45.0815, \quad q_{1,5} = 0.3578, \quad q_{1,6} = 0.6321,$$

$$\varepsilon_1 = 13.06\%.$$

Аналіз похибки засвідчує, що майже всі відносні похибки не перевищують 10%. Лише на протязі одинадцятого та п'ятнадцятого тижнів похибка прогнозу сягала значень більших 10% межі.

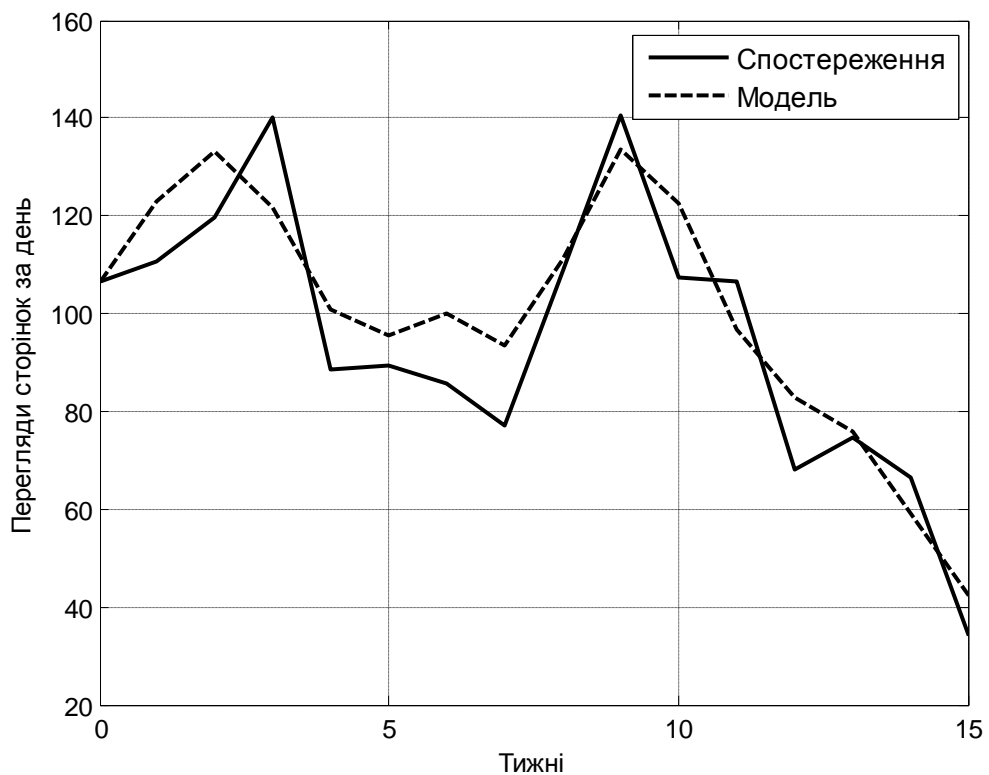


Рис.3.7. Моделювання результуючої відвідуваності по двох оновленнях контенту

При ідентифікації моделі результуючої відвідуваності для досягнення задовільної точності довелося врахувати вплив загальної відвідуваності. Ця ж модель без додаткових уточнень дала непогані результати для двох періодів аналізу (похибка склала 13.28%). Результати моделювання подані на рис. 3.7, а структура ідентифікованої моделі представлена в наступному співвідношенні

$$x_{2,k} = \left(g_{2,1}x_{2,k-1} + g_{2,2}x_{1,k} + q_{2,1}e^{-k \cdot q_{2,2}} \left(v_1(s_{1,k}) \right)^{q_{2,3}} \right) (1 \pm \varepsilon_2), \quad (3.23)$$

де $g_{2,1} = 0.2$, $g_{2,2} = 0.25$, $q_{2,1} = 5.0$, $q_{2,2} = 0.$, $q_{2,3} = 0.06$, $T_2 = 7$,
 $\varepsilon_2 = 13.28\%$.

При ідентифікації цієї моделі достатньо було використати лише перших 2 спостереження. Подальша динаміка результуючої відвідуваності забезпечилася використанням прогнозованих обсягів загальної відвідуваності, тобто змінною X_1 .

Розподіл похибок ідентифікації даної моделі наводиться на рис. 3.8. Варто відзначити, що максимальні значення похибок спостерігаються в першій частині спостережень, а на другому підінтервалі спостережень відносна похибка майже вкладається в 10% межі. Таким чином, проведені обчислення підтверджують можливість застосування дискретних динамічних моделей для прогнозування відвідуваностей Веб-сайтів при використанні в якості фактору впливу функції якості оновлення контенту.

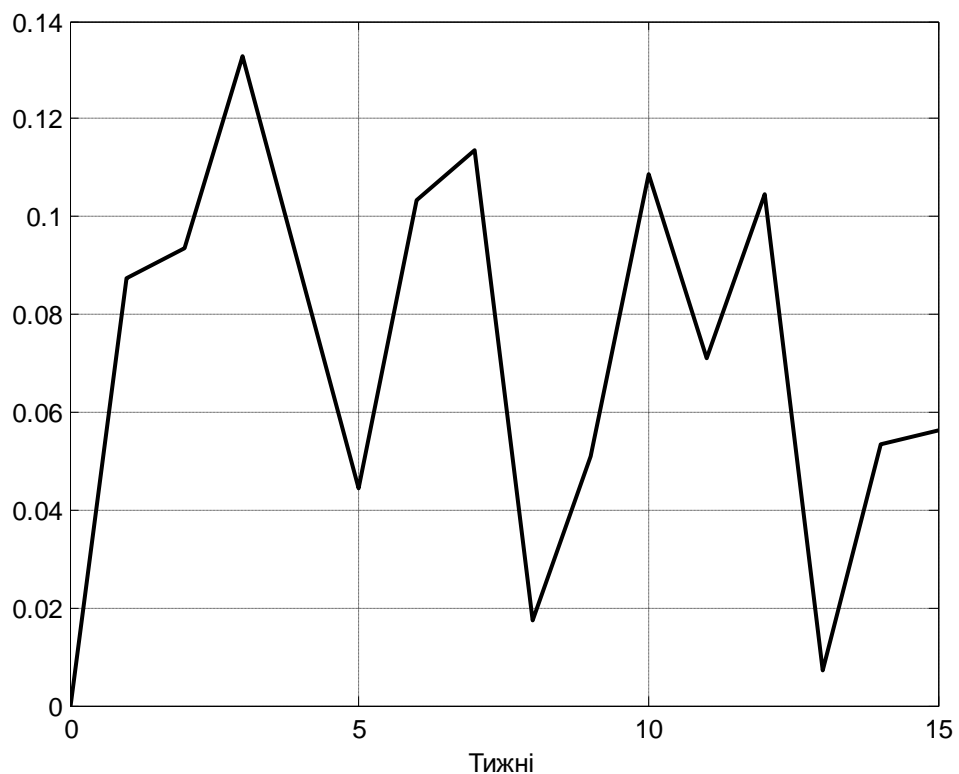


Рис.3.8. Розподіл відносних похибок ідентифікації моделі загальної відвідуваності по оновленню контенту

При цьому для побудови відповідності між фактором впливу та результируючим фактором може бути використана степенева залежність. Варто відзначити, що даний підхід до моделювання має свої обмеження. Найбільший недолік полягає в необхідності прогнозування динаміки функції якості контенту на основі характеристик бізнес-процесів об'єкта, для підтримки якого використовується даний Веб-сайт. Далеко не завжди таке моделювання доцільне. При системному суттєвому оновленні контенту Веб-сайту число можливих станів зростає настільки, що дискретна модель втрачає свою ефективність. Однак, щоб моделювати функцію якості оновлення контенту аналітичним співвідношенням, необхідно регуляризувати її спостережені значення, в яких, як відзначалося, присутні випадкові викиди. Аналіз цих питань здійснюється в наступному підрозділі.

3.1.2. Модель відвідуваності із функцією якості контенту, побудованою на основі системи Моно

Незважаючи на тижневе усереднення відвідуваності та факторів впливу для реальних Веб-сайтів, їх динаміка характеризується значною нерегулярністю. Це послаблює точність побудованих моделей та ускладнює прогнозування функції якості контенту. Тому перейдемо до згладжування вказаних показників. При цьому використаємо кратне згладжування методом ковзаючого середнього із мінімальним шаблоном, яке, на відміну від інших методів, суттєво нівелює аномальні викиди, а не лише частково згладжує їх. Приклад згладженої динаміки Веб-сайту та факторів впливу на її формування наведено на рис. 3.9.

Для демонстрації динаміки відвідуваності вибрано Веб-сайт навчального підрозділу, а саме Веб-сайт факультету комп'ютерних інформаційних технологій (ФКІТ) Тернопільського національного економічного університету (ТНЕУ). Наведений рисунок демонструє усунення аномальних викидів, за рахунок чого і отримано максимальне відхилення згладжених значень від

спостережених на рівні 31%. При цьому однак середнє відхилення склало практично лише 11%. При проведенні обчислень використовувався триточковий шаблон ковзаючого середнього, який еквівалентний простому середньому по трьох сусідніх точках. Згаданий шаблон застосовувався тричі до одного набору даних.

Аналіз згладженої відвідуваності дозволяє розбити її на інтервали відносної постійності та інтервали суттєвих приростів. З практичної точки зору найбільший інтерес становлять ділянки суттєвого приросту відвідуваності. Моделювання відвідуваностей тільки для цих часових інтервалів дозволяє суттєво зменшити складність моделі.

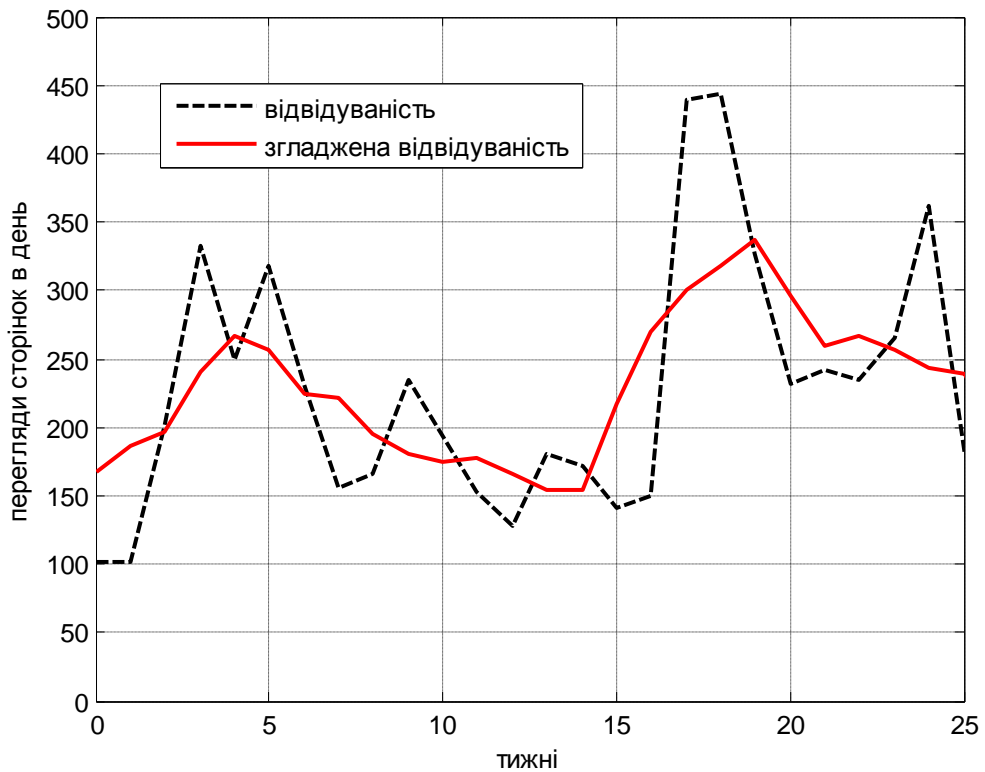


Рис. 3.9. Згладжена загальна відвідуваність Веб-сайту ФКІТ за піврічний період

Проблема полягає в тому, щоб виділити такі інтервали за кількома початковими значеннями. З цією метою використовується формальний критерій перевищення похідної відвідуваності деякого мінімального значення

$$Y'(t) > D_{min} \cdot \quad (3.24)$$

Величина D_{min} підбирається для кожного Веб-сайту його дослідником індивідуально. На наступному рис. 3.10 подана згладжена відвідуваність, яка за критерієм (3.24), формально розбита на 4 інтервали приросту відвідуваності, початок яких відображено пунктирними лініями. В даному випадку використано значення $D_{min} = 0.04$. Закінчення інтервалу приросту фіксується по моменту зменшення відвідуваності нижче початкового рівня в разі вичерпання впливу поточного оновлення контенту або по початку стадії зростання після попереднього спадання у випадку дії нового оновлення до завершення дії попереднього.

При фіксації початку інтервалу зростання відвідуваності можна виявити тематику, зміна контенту якої її підтримує. З цією метою аналізуємо згладжені відвідуваності основних тематик Веб-сайту, які ще можна назвати субсайтами. Серед них вибираємо ту, яка характеризується максимальним приростом, вважаючи її функцією якості оновлення контенту для даного інтервалу росту відвідуваності. Відвідуваності тематик згладжуємо аналогічно згладжуванню загальної відвідуваності. На рис. 3.10 представлені графіки згладжених відвідуваностей деякого Веб-сайту та можливих функцій якості.

Для більшої наочності значення функцій якості пропорційно збільшені. Рисунок демонструє синхронізацію загальної відвідуваності та відповідних функцій якості. Аналогічна картина спостерігалася для відвідуваностей всіх аналізованих Веб-сайтів. Наявність факторів впливу на відвідуваність можна пояснити ростом популярності деякої розроблюваної або рекламованої тематики. Цей ріст популярності сприяє покращенню видимості сторінок тематики в пошукових системах по різних пошукових запитах. Покращення видимості сприяє підвищенню відвідуваності всіх тематик Веб-сайту, тим самим сприяючи підвищенню загальної відвідуваності.

Аналіз згладженої динаміки функцій якості оновлення контенту на стадії активації показує їх подібність, що дозволяє використовувати єдиний математичний апарат для їх моделювання. В якості такого апарату нами використано модифіковану систему диференціальних рівнянь Моно, які

моделюють динаміку функції якості оновлення контенту X та його потенційної аудиторії A .

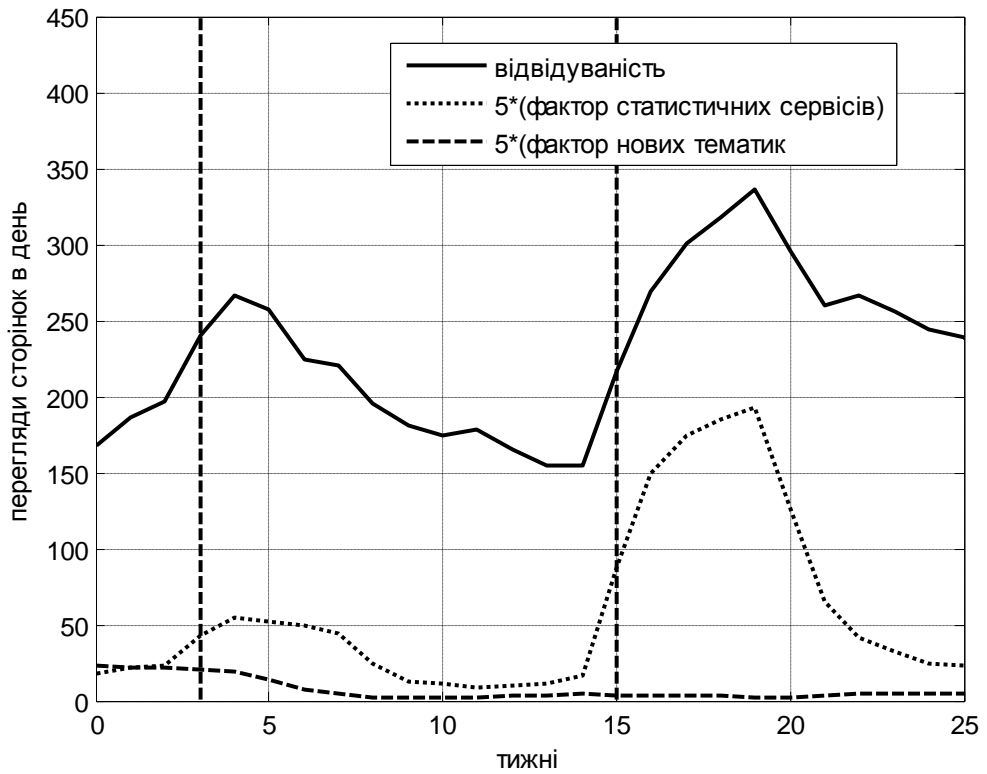


Рис. 3.10. Згладжена відвідуваність та функції якості оновлення контенту

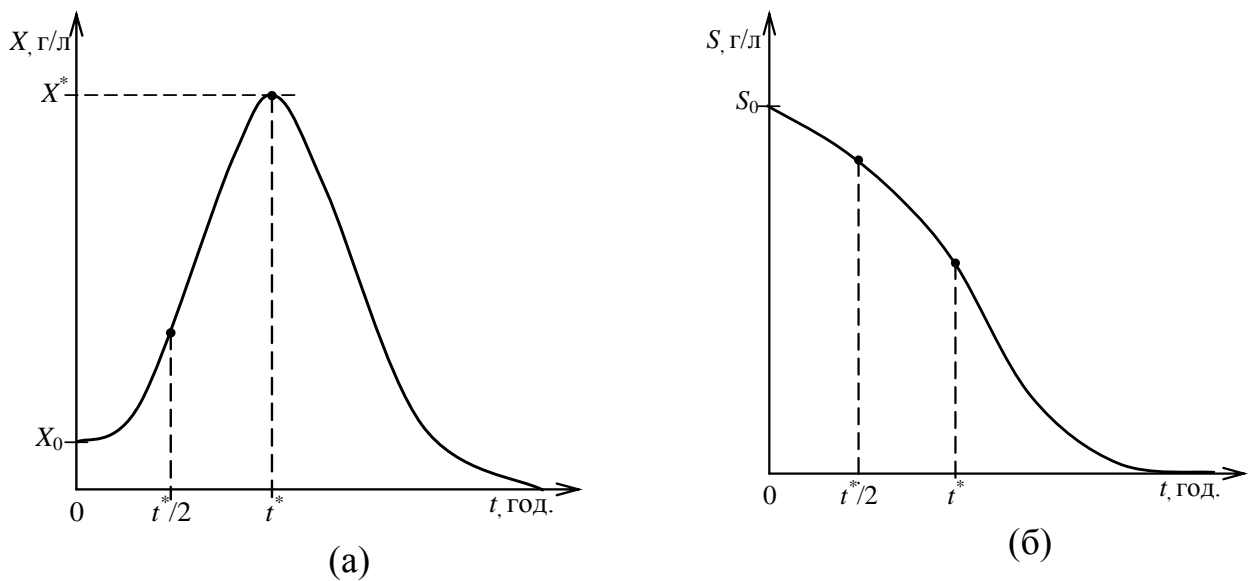


Рис. 3. 11 Типові траєкторії концентрації (а) мікроорганізмів $X(t)$ та (б) субстрату $S(t)$ в періодичному реакторі

Система диференціальних рівнянь Моно подає динаміку концентрації системи мікроорганізмів X , при скінченному обсязі концентрації субстрату S , що сприяє їх розмноженню, і використовується при моделюванні біотехнологічних процесів. Типові динаміки згаданих величин наведені на рис. 3.11.

Загальний вигляд системи Моно наводиться нижче – співвідношеннями (3.25). Співставлення графіків, поданих на рис. 3.10 та 3.11, показує подібність динаміки факторів впливу на відвідуваність із динамікою концентрації мікроорганізмів, яка моделюється системою Моно. Відмінність між ними полягає в тому, що на завершальній стадії процесу чисельність системи мікроорганізмів прямує до 0, а функція якості оновлення контенту на завершальній стадії приймає деяке значення d_{X0} , що, як правило, відрізняється від 0.

$$\begin{cases} X'(t) = \left(a_1 \frac{S(t)}{a_4 + S(t)} - a_2 \right) X(t), \\ S'(t) = -a_3 \frac{S(t)}{a_4 + S(t)} X(t), \end{cases} \quad (3.25)$$

$$X(0) = X_0, \quad S(0) = S_0.$$

Окрім того, необхідно відшукати аналог субстрату, який використовується в класичній моделі Моно. З цією метою вводиться умовна допоміжна величина A — обсяг потенційної аудиторії. Її початковий рівень прирівнюємо до максимальної відвідуваності Веб-сайту. Як було показано в попередньому підрозділі, пов'язати функцію якості контенту X із обсягом згладженої відвідуваності Y достатньо точно можна за допомогою степеневі функції із відповідно підібраними параметрами.

Щоб забезпечити вихід відвідуваності на завершальній стадії на рівень d_{X0} , в правій частині диференціальних рівнянь переходимо від звичайного $A(t)$ до скоригованого обсягу аудиторії $A_e(t)$, що визначається

співвідношенням (3.26). Згідно цього співвідношення скоригована функція якості приймає значення з інтервалу $[d_{X0}, A_0]$, на відміну від звичайної, яка теоретично може приймати значення з інтервалу $[0, A_0]$.

$$A_e(t) = \begin{cases} A(t) - d_{X0} \frac{A_0 - A(t)}{A_0 - d_{X0}} & A(t) > d_{X0}, \\ 0 & A(t) \leq d_{X0}. \end{cases} \quad (3.26)$$

В результаті проведених перетворень отримаємо структуру моделі відвідуваності наступного виду:

$$\begin{cases} X'(t) = \left(a_1 \frac{A_e(t)}{a_4 + A_e(t)} - a_2 \right) X(t), \\ A'(t) = -a_3 \frac{A_e(t)}{a_4 + A_e(t)} X(t), \end{cases} \quad (3.27)$$

$$\text{при умові } \left(A > \frac{A_0}{2} \right) \vee (X \geq d_{X0}), \quad (3.28)$$

$$\begin{cases} X'(t) = 0, \\ A'(t) = 0, \end{cases} \quad (3.29)$$

$$\text{при умові } \left(A \leq \frac{A_0}{2} \right) \wedge (X < d_{X0}), \quad (3.30)$$

$$Y(t) = Y(t_0) + q_1 (X(t) - X_0)^{q_2} e^{-q_3 t}, \quad (3.31)$$

$$X(0) = X_0, \quad (3.32)$$

$$A(0) = A_0 = X_{max}. \quad (3.33)$$

Після встановлення структури моделі відвідуваності необхідно вказати метод її ідентифікації. Ці питання аналізуються в наступних підрозділах.

3.2. Метод ідентифікації математичних моделей короткотермінового прогнозу відвідуваності Веб-сайтів

Ідентифікацію моделі відвідуваності здійснюємо в 2 етапи. На першому етапі на основі спостереженої функція якості оновлення контенту X ідентифікуємо параметри $a_1 - a_4$ динаміки системи (3.27)-(3.30), а на другому етапі ідентифікуємо параметри $q_1 - q_3$ моделі загальної відвідуваності (3.31) [28]. Ідентифікація здійснюється за одним із найпростіших критеріїв мінімізації середньоквадратичної похибки активності фактора X на точках навчальної вибірки. Критерій якості мінімізується за допомогою модифікованого градієнтного методу Левенберга-Марквардта, який забезпечує пошук лише локального мінімуму і тому вимагає хорошого початкового наближення. Таке початкове наближення можна побудувати за допомогою аналізу особливостей розв'язків системи диференціальних рівнянь (3.26).

Подібні задачі ідентифікації розв'язувалися в дослідженнях Ю.Р. Піговського [36] для моделювання біотехнологічних процесів. Однак аналізовані ним задачі передбачали ідентифікацію моделі по системі точок, які покривали весь допустимий діапазон значень. Внаслідок дії людського фактора кожна конкретна реалізація моделі є унікальною. Тому таким чином ідентифікована модель функції якості оновлення контенту не може бути використана для нового її прояву. Вона повинна бути адаптивною по параметрах, тобто параметри шуканої функції повинні уточнюватися в процесі спостереження за процесом. Для кожного прояву цього процесу адаптивна ідентифікація повинна здійснюватися заново.

Модель (3.26) 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. Для доотримання ще 2-4 значень функції якості оновлення контенту необхідно їх прогнозувати на основі апріорних оцінок.

Побудова оцінок такого роду вимагає хорошої інформованості про властивості шуканих функцій. Аналізуючи різні реалізації функції якості оновлення контенту, можна помітити, що всі вони містять по одному інтервалу зростання та спадання, однак за тривалістю ці інтервали можуть суттєво відрізнятись. Зокрема, деякі із них містять лише 3-4 точки до досягнення максимуму функції якості, тобто є суттєво нелінійною.

Це дозволяє виділити процеси суттєвого росту відвідуваності, які містять не більше чотирьох спостережень до досягнення максимального її значення, в клас процесів короткотермінового росту відвідуваності. Всі інші процеси росту відвідуваності віднесемо в клас довготермінових. Класифікувати тип процесів росту відвідуваності для конкретного Веб-сайту можна, проаналізувавши динаміку його першого прояву. Наступні прояви росту відвідуваності протягом тривалого періоду будуть того ж типу.

В даному підрозділі розглянемо метод побудови апріорних оцінок та метод ідентифікації модифікованої системи Моно для короткотермінових процесів росту відвідуваності. Апріорні оцінки параметрів відповідних функцій якості оновлення контенту повинні враховувати їх суттєву нелінійність. Це виключає лінійну апроксимацію функції якості як ефективний інструмент прогнозування її поведінки. В той же час експериментальні дослідження показують, що вже друга похідна функції якості оновлення контенту має прогнозований лінійний характер. Помічена властивість дозволяє побудувати рекурентні співвідношення для покрокового прогнозу значень функції якості в точках, які необхідно спостерігати для ефективною ідентифікації моделі.

З метою побудови прогнозних значень \tilde{x}_i фактору впливу x_i використовуються прогнози його першої p_i та другої q_i похідної. Для побудови прогнозів другої похідної використовується її лінійна модель для попередньо спостереженого інтервалу росту $[t_0, t_1]$ функції якості контенту даного Веб-сайту. Ця модель будується на основі моделі відвідуваності

спостереженої функції якості без побудови апіорних співвідношень, лише за допомогою розробленого методу ідентифікації. Це можливо тому, що в цьому першому випадку до початку ідентифікації вже спостережено всі значення даної функції якості. Тут модель використовується не для прогнозу, а для згладжування значень експериментальних даних. Наближаємо значення другої похідної функції якості наступним співвідношенням :

$$q_i^0 = b_1 + b_2 i, \quad (3.34)$$

$$b_1 = \min_{t \in [t_0, t_1]} (q(t)), \quad b_3 = \max_{t \in [t_0, t_1]} (q(t)), \quad (3.35)$$

$$b_2 = \frac{b_3 - b_1}{t_1 - t_0}, \quad (3.36)$$

де q_i^0 — значення другої похідної моделі початкового фактора впливу похідної.

Маючи оцінку параметрів наближення другої похідної, будуємо багатокрокові апіорні співвідношення прогнозування значень функції якості. На першій стадії ініціалізуємо початкові значення:

$$\tilde{x}_0 = x_0, \quad (3.37)$$

$$\tilde{x}_1 = x_1, \quad (3.38)$$

$$p_0 = x_1 - x_0, \quad (3.39)$$

$$q_0 = b_1. \quad (3.40)$$

На наступній стадії будуємо багатокроковий прогноз, який включає послідовні оцінки першої та другої похідних а також саме значення прогнозованого фактора:

$$p_i = p_{i-1} + q_{i-1}, \quad (3.41)$$

$$q_i = q_{i-1} + b_2, \quad (3.42)$$

$$\tilde{x}_i = \tilde{x}_{i-1} + p_i. \quad (3.43)$$

Обчислення за співвідношеннями (3.41) – (3.43) продовжуємо, поки похибка прогнозу буде лежати в допустимих межах δ_0 :

$$\frac{|\tilde{x}_i - x_i|}{X_{prev}} \leq \delta_0, \quad (3.44)$$

де X_{prev} — максимальне значення попередньої реалізації функції якості.

Апріорні співвідношення дозволяють прогнозувати динаміку функції якості на етапі її зростання а також сприяють отриманню інформації про дійсні значення функції якості в обсягах, достатніх для застосування процедури ідентифікації її моделі. Ідентифікація системи Моно методом Левенберга-Марквардта вимагає побудови процедури задання початкових значень коефіцієнтів системи. Розглянемо її детальніше. Роль коефіцієнта a_4 полягає у балансуванні коефіцієнтів моделі фактору впливу. Тому значення a_4 підбираємо на вузлах рівномірної сітки W_4 , яка покриває деякий діапазон:

$$a_4 \in W_4 = \{ a_4^0, a_4^0 + h_4, a_4^0 + 2h_4, \dots, a_4^N \}. \quad (3.45)$$

Параметри сітки a_4^0 , h_4 , a_4^N підбираються експериментально.

Коефіцієнт a_2 визначає відносну швидкість падіння інтересу аудиторії, його ідентифікація можлива лише після проходження точки максимуму відвідуваності. Аналогічно попередньому значення a_2 підбираємо на вузлах рівномірної сітки W_2 :

$$a_2 \in W_2 = \{ a_2^0, a_2^0 + h_2, a_2^0 + 2h_2, \dots, a_2^N \}, \quad (3.46)$$

Для вибору початкового значення сітки a_2^0 використаємо наближену оцінку швидкості падіння інтересу аудиторії на основі принципу симетрії, як аналога швидкості p_0 наростання інтересу аудиторії. Для побудови

мінімального значення коефіцієнта відносимо цю швидкість до максимального спостереженого значення функції якості X_{max} :

$$a_2^0 = \frac{P_0}{X_{max}} \quad . \quad (3.47)$$

Максимальне початкове значення відносної відвідуваності a_2^N вибираємо кратним мінімальному, в найпростішому випадку просто подвоюючи його.

Після вибору значень коефіцієнтів a_2 та a_4 , із першого рівняння системи (3.27) можна визначити початкове значення коефіцієнта a_1 відносного росту функції якості:

$$a_1 = \left(\frac{p_0}{x_1} + a_2 \right) \frac{A_0}{A_0 + a_4} = \left(\frac{p_0}{x_1} + a_2 \right) \frac{X_{max}}{X_{max} + a_4} \quad . \quad (3.48)$$

Початкове значення коефіцієнта a_3 будується на основі значення коефіцієнта a_1 за допомогою простого співвідношення (3.49), яке можна обґрунтувати наступним чином. На протязі певного часового інтервалу відвідуваність досягає свого максимуму. За цей же проміжок часу потенційна аудиторія зменшиться приблизно до свого половинного значення. Оскільки початкове значення обсягу потенційної аудиторії прирівняно до максимальної відвідуваності, то природньо покласти коефіцієнт відносного зменшення потенційної аудиторії вдвічі меншим, ніж коефіцієнт відносного приросту відвідуваності:

$$a_3 = \frac{a_1}{2} \quad . \quad (3.49)$$

Окрім методу побудови апріорних оцінок значень функції якості оновлення контенту та початкових значень коефіцієнтів $a_1 - a_4$ системи Моно, необхідно запропонувати також метод побудови апріорних оцінок відвідуваностей та початкових значень коефіцієнтів $q_1 - q_3$ моделі

відвідуваності. Для побудови апріорних оцінок відвідуваностей, використовуючи подібність конфігурацій функції якості оновлення контенту та відвідуваностей, застосуємо формули аналогічні співвідношенням (3.34)-(3.43) з тією різницею, що ці співвідношення стосуються значень відвідуваностей, а не значень функції якості. На основі попередньо спостереженої відвідуваності будуємо лінійну оцінку її другої похідної:

$$Q_i^0 = B_1 + B_2 i, \quad (3.50)$$

$$\text{де } B_1 = \min_{t \in [t_0, t_1]} (Q(t)), \quad B_3 = \max_{t \in [t_0, t_1]} (Q(t)), \quad (3.51)$$

$$B_2 = \frac{B_3 - B_1}{t_1 - t_0}. \quad (3.52)$$

На першій стадії ініціюємо початкові значення моделі відвідуваності:

$$\tilde{y}_0 = y_0, \quad (3.53)$$

$$\tilde{y}_1 = y_1, \quad (3.54)$$

$$P_0 = y_1 - y_0, \quad (3.55)$$

$$Q_0 = b_1. \quad (3.56)$$

На наступній стадії будуємо багатокроковий прогноз, який включає послідовні оцінки першої та другої похідних а також саме значення прогнозованого фактора:

$$P_i = P_{i-1} + Q_{i-1}, \quad (3.57)$$

$$Q_i = Q_{i-1} + B_2, \quad (3.58)$$

$$\tilde{y}'_i = \tilde{y}'_{i-1} + P_i. \quad (3.59)$$

Обчислення за співвідношеннями (3.53) – (3.59) продовжуємо поки похибка прогнозу буде лежати в допустимих межах δ_0 :

$$\frac{|\tilde{y}_i - y_i|}{Y_{prev}} \leq \delta_0, \quad (3.60)$$

де Y_{prev} — максимальне значення попередньої реалізації росту відвідуваності.

Для побудови початкових значень коефіцієнтів моделі відвідуваності прологарифмуємо співвідношення (3.31), отримуючи систему лінійних рівнянь:

$$\ln(q_1) + q_2 \ln(X(t_i) - X_0) - q_3 = \ln(Y(t_i) - Y(t_0)), \quad i = 1, 2, 3. \quad (3.61)$$

Після встановлення початкових значень параметрів функції \tilde{X} якості оновлення контенту, а також функції \tilde{Y} відвідуваності здійснюється їх ідентифікація на основі мінімізації середньоквадратичних критеріїв якості:

$$\bar{a} = \arg \min_c \sum_{j=1}^I (\tilde{x}(\bar{c}, t_j) - x_j)^2. \quad (3.62)$$

$$\bar{q} = \arg \min_r \sum_{j=1}^I (\tilde{y}(\bar{r}, t_j) - y_j)^2. \quad (3.63)$$

Цю мінімізацію здійснюємо градієнтним методом --Левенберга-Марквардта, пристосованим до мінімізації функціоналів вказаного виду.

На основі наведених теоретичних положень сформуємо алгоритм ідентифікації короткотермінових приростів відвідуваності:

1. Аналізуємо динаміку спостереженого першого короткотермінового приросту відвідуваності після його повного завершення. За допомогою співвідношень (3.45)-(3.49), (3.61) будуємо початкове наближення коефіцієнтів моделі короткотермінових приростів відвідуваності, та уточнюємо їх методом Левенберга-Марквардта за критеріями (3.62), (3.63).
2. Будуємо лінійне наближення (3.34), (3.50) других похідних фактора впливу та відвідуваності за ідентифікованою моделлю першого спостереженого короткотермінового приросту відвідуваності.
3. Встановлюємо появу наступного приросту відвідуваності за критерієм (3.24).

4. За допомогою наближеного значення других похідних та спостережень, які ідентифікують приріст відвідуваності, будуємо наближення функцій якості оновлення контенту та відвідуваності. Точність прогнозу контролюємо за критеріями (3.44), (3.60). При порушенні точності прогнозу він перераховується від спостережуваного, а не прогнозованого значення.
5. Після отримання 5-6 спостережень проводиться побудова початкових значень коефіцієнтів моделі функції якості оновлення контенту та відвідуваності а також їх уточнення методом Левенберга-Марквардта.

В загальному така ж процедура застосовувалася до ідентифікації динаміки довготривалих приростів відвідуваностей, однак вона має ряд суттєвих відмінностей, які будуть розглянуті в наступному підрозділі.

3.3. Метод ідентифікації математичних моделей довготермінового прогнозу відвідуваності Веб-сайтів

Короткотермінові прирости 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, докладает максимальних зусиль до постійного росту його відвідуваності, аж до виходу на рівень, який забезпечує його належну результативність. При цьому відпрацьовуються спеціальні методики, які забезпечують досягнення максимуму відповідного приросту відвідуваності на протязі не менше шести тижнів. Для прикладу динаміки відвідуваності вибрано промо-сайт, а саме Веб-сайт сервісу API2Cart підприємства із розробки програмного забезпечення Magnetic One, графік згладженої загальної відвідуваності якого наведено на рис. 3.12.

Наведений рисунок демонструє усунення аномальних викидів, за рахунок чого і отримано максимальне відхилення на рівні 53%. При цьому середнє відхилення склало лише 7%. Для виділення інтервалів суттєвого приросту відвідуваності використовується формальний критерій (3.24).

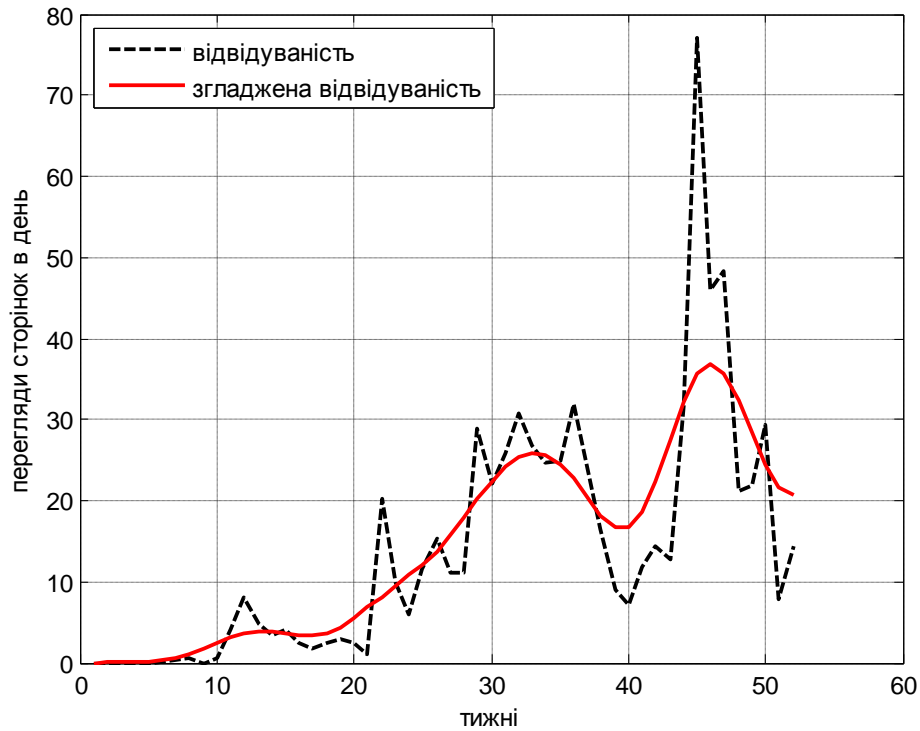


Рис. 3.12. Згладжена загальна відвідуваність Веб-сайту API2Cart за річний період

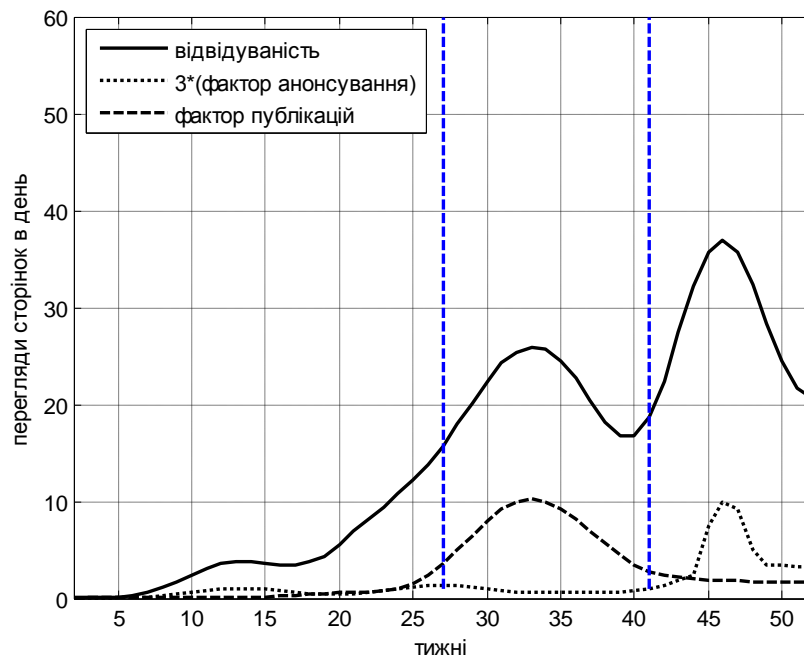


Рис. 3.13. Згладжена відвідуваність та фактори впливу на неї для Веб-сайту API2Cart

Величина D_{min} склала 0.04. На рис. 3.13 представлені графіки згладжених відвідуваностей згаданого Веб-сайту та можливих факторів

впливу на цю відвідуваність. Для більшої наочності значення факторів впливу пропорційно збільшені. Рисунок демонструє синхронізацію загальної відвідуваності та сукупності відповідних факторів впливу.

Для моделювання приросту відвідуваності даного типу використаємо раніше запропоновану модель (3.27)-(3.33). Однак ідентифікація цієї моделі в даному випадку буде мати свої особливості. Зокрема не спостерігається лінійної поведінки другої похідної фактора впливу, як у випадку короткотермінових інтервалів підвищення відвідуваності. Справа в тому, що тривалість процесу збільшення обсягу фактору впливу спричиняє зміну швидкості його росту. Зокрема на першій стадії процесу росту швидкість росту відносно невелика, згодом вона суттєво зростає. Зразок подібної динаміки наведено на рис. 3.14.

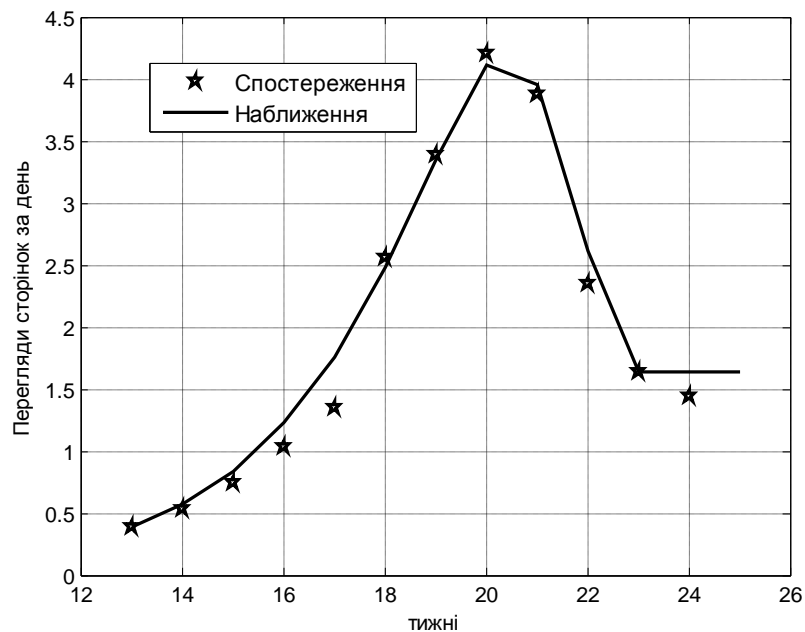


Рис. 3.14. Динаміка типової функції якості оновлення контенту для довготривалого приросту відвідуваності

Така мінливість не дозволяє застосовувати лінійне наближення для прогнозування динаміки фактору впливу на значному інтервалі. В той же час, на початковому етапі кількість точок недостатня для застосування процедури

ідентифікації. Тому на початковому етапі використовуємо співвідношення для побудови початкових значень (3.45)-(3.49) параметрів системи Моно.

При цьому підбір кращих значень коефіцієнтів a_2 та a_4 (значення яких належать відповідним множинам) здійснюється за критерієм мінімізації відхилення модельного значення від спостереженого в останній на даний момент, тобто другій точці:

$$\bar{a} = \arg \min_c (\tilde{x}(\bar{c}, t_2) - x_2)^2, \quad (3.64)$$

а не в результаті застосування процедури ідентифікації. Підібрані таким чином параметри функції якості дозволяють будувати прогноз для періоду, отримавши спостереження по якому, можна застосувати процедуру ідентифікації функції якості. Однак точність такої ідентифікації достатньо невисока. Тому штучно збільшуємо кількість точок ідентифікації за допомогою лінійної інтерполяції значень функції якості оновлення контенту Веб-сайту по двох останніх точках спостереження. Наближення будемо за рівнянням прямої з кутовим коефіцієнтом, яка проходить через точки $(t_{m-1}, x_{m-1}), (t_m, x_m)$:

$$X_m(t) = x_{m-1} + (x_m - x_{m-1})(t - m + 1). \quad (3.65)$$

Це дозволяє модифікувати критерій побудови коефіцієнтів моделі фактору наступним чином:

$$\begin{aligned} \bar{a} = \arg \min_{\bar{c}} \{ & \sum_{k=1}^m (\tilde{x}(\bar{c}, t_k) - x_k)^2 + \\ & + \sum_{k=m-1}^{m+m_1} (x_{m-1} + (x_m - x_{m-1})(k - m + 1) - \tilde{x}(\bar{c}, t_k))^2 \}, \end{aligned} \quad (3.66)$$

де m — кількість спостережених точок, m_1 — кількість додаткових точок ідентифікації.

Таким чином ідентифікована функція якості є достатньо точною на етапі її лінійного росту. Однак на наступному етапі поведінка функції якості наближається до параболічного закону, оскільки фаза зростання відвідуваності переходить у фазу її зменшення. В цьому випадку вже похідна функції якості еволюціонує за законом, наближеним до лінійного. Отже після завершення інтервалу лінійного росту відвідуваності використовується ідентифікація за лінійним наближенням P_m її похідної. Для наближення використовується многочлен Лагранжа першого порядку

$$P_m(t) = p_{m-3}(m-2-t) + p_{m-2}(t-m+3) . \quad (3.67)$$

На цьому етапі функціонал похибки моделі включає як відхилення від спостережених значень, так і відхилення від лінійного наближення похідної в ще не спостережених точках:

$$\begin{aligned} \bar{a} = \arg \min_{\bar{c}} \{ & \sum_{k=1}^{m_2} (\tilde{x}(\bar{c}, t_k) - x_k)^2 + \\ & + \sum_{k=m-3}^{l+m_1} (p_{m-3}(m-2-k) + p_{m-2}(k-m+3) - \tilde{x}'(\bar{c}, t_k))^2 \}. \end{aligned} \quad (3.68)$$

Якщо ідентифікація за критерієм (3.63) не дозволяє досягати достатньої точності на всьому інтервалі приросту відвідуваності, то на наступному етапі вже можна використовувати звичайний середньоквадратичний критерій

$$\bar{a} = \arg \min_{\bar{c}} \{ \sum_{k=1}^{m_2} (\tilde{x}(\bar{c}, t_k) - x_k)^2 \}. \quad (3.69)$$

На цей момент нагромаджується достатньо спостережень для ідентифікації моделі фактору впливу по всьому інтервалу приросту відвідуваності із достатньою точністю.

Для моделювання відвідуваності необхідно будувати перш за все апіорні оцінки. З цією метою використовуємо співвідношення (3.61).

Значення відвідуваності в точці t_3 на початку процесу ідентифікації моделюємо за лінійною екстраполяцією виду:

$$\tilde{y}_3 = y_2 + (y_2 - y_1) = 2y_2 - y_1 \quad (3.70)$$

На основі таких попередніх оцінок моделюємо ще 2-3 значення відвідуваності, після чого можна проводити процедуру ідентифікації моделі відвідуваності Веб-сайту на основі моделі функції якості.

На основі наведених теоретичних положень сформуємо алгоритм ідентифікації довготермінових приростів відвідуваності:

1. За допомогою співвідношень (3.45)-(3.49), (3.65), (3.61) за критерієм (3.64) будуємо початкове наближення коефіцієнтів моделі відвідуваності, точність яких контролюємо за критеріями (3.44), (3.60).
2. Після отримання 4-5 спостережень відвідуваності та функції якості оновлення контенту здійснюємо перший етап ідентифікації моделі функції якості за критерієм (3.66), який включає лінійну екстраполяцію значень згаданого фактора. Модель відвідуваності ідентифікується за звичайним критерієм (3.63).
3. Після втрати точності моделі функції якості контенту, ідентифікованої на першому етапі, продовжуємо ідентифікацію вже за критерієм (3.68), в який включено лінійну екстраполяцію першої похідної функції якості контенту.
4. У випадку втрати точності моделі функції якості контенту, яка ідентифікована на попередньому етапі, кількість точок спостереження вже буде достатньою для ідентифікації цієї моделі за звичайним критерієм.

Висновки до розділу 3

1. В результаті використання теорії систем побудовано набір дискретних динамічних моделей прогнозування відвідуваності Веб-сайту, у яких функції якості оновлення контенту узгоджуються із бізнес-процесами об'єкта,

представленого цим Веб-сайтом, що уможливило встановлення залежності між відвідуваністю та функцією якості оновлення контенту.

2. На основі згладжування агрегованих відвідуваностей Веб-сайту та функції якості оновлення контенту і застосування системи звичайних диференціальних рівнянь Моно побудована модель відвідуваності Веб-сайтів, що дозволило встановити причинно-наслідковий зв'язок між ними для широкого класу Веб-сайтів. Запропоновано формальний критерій виділення функції якості оновлення контенту.

3. Здійснено класифікацію приростів відвідуваності Веб-сайтів на короткотермінові та довготермінові. Для 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 дозволяє врахувати стрімку динаміку короткотермінового приросту відвідуваності.

4. Для довготермінових прогнозів відвідуваності запропоновано 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 дозволяє врахувати неоднорідність динаміки функції якості для прогнозування довготермінової відвідуваності під її впливом.

РОЗДІЛ 4. ПРИКЛАДНІ АСПЕКТИ МОДЕЛЮВАННЯ ВІДВІДУВАНOSTІ ВЕБ-САЙТІВ.

Розроблені математичні методи вимагають експериментального дослідження ефективності. З цією метою створено програмний комплекс для математичного моделювання процесів відвідуваності Веб-сайтів. Для експериментів відібрані Веб-сайти, відвідуваність яких вимагає суттєвого покращення. Зокрема це Веб-сайт навчального підрозділу а також промо-сайт програмного продукту. Особливості реалізації програмного комплексу а також проведення експериментальних досліджень розглядалися в роботах [28, 30-32] а також в ході подальшого викладу у цьому розділу.

4.1. Архітектура програмного комплексу для математичного моделювання процесів відвідуваності Веб-сайтів.

Підвищення відвідуваності Веб-сайтів можна забезпечити шляхом покращення його структури та постійним оновленням контенту як безпосередньо на Веб-сайті, так і на сторонніх спеціалізованих Веб-сайтах або у цілеспрямованих розсиланнях в ході рекламних компаній. На даному етапі ці процеси плануються та моніторяться розробниками та службами підтримки в ручному режимі без підтримки за допомогою програмної реалізації математичних методів та моделей. Це приводить до неефективних витрат робочого часу спеціалістів, які могли б інтенсивніше займатися розв'язанням творчих завдань.

Розроблені математичні методи та моделі, які описані в попередніх підрозділах, дозволяють зменшити завантаженість спеціалістів рутинними управлінськими задачами та підвищують достовірність прогнозів, необхідних при плануванні підвищення відвідуваності. Однак ефективним інструментом для практичного використання вони можуть стати лише за умови програмної реалізації. Ця реалізація здійснена у вигляді програмного комплексу планування діяльності із підвищення відвідуваності Веб-сайтів.

Даний 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, прогнозування відвідуваності Веб-сайтів. Загальна структура взаємодії процесів у двох підсистемах комплексу представлені на рис. 4.1 у вигляді схем використання.

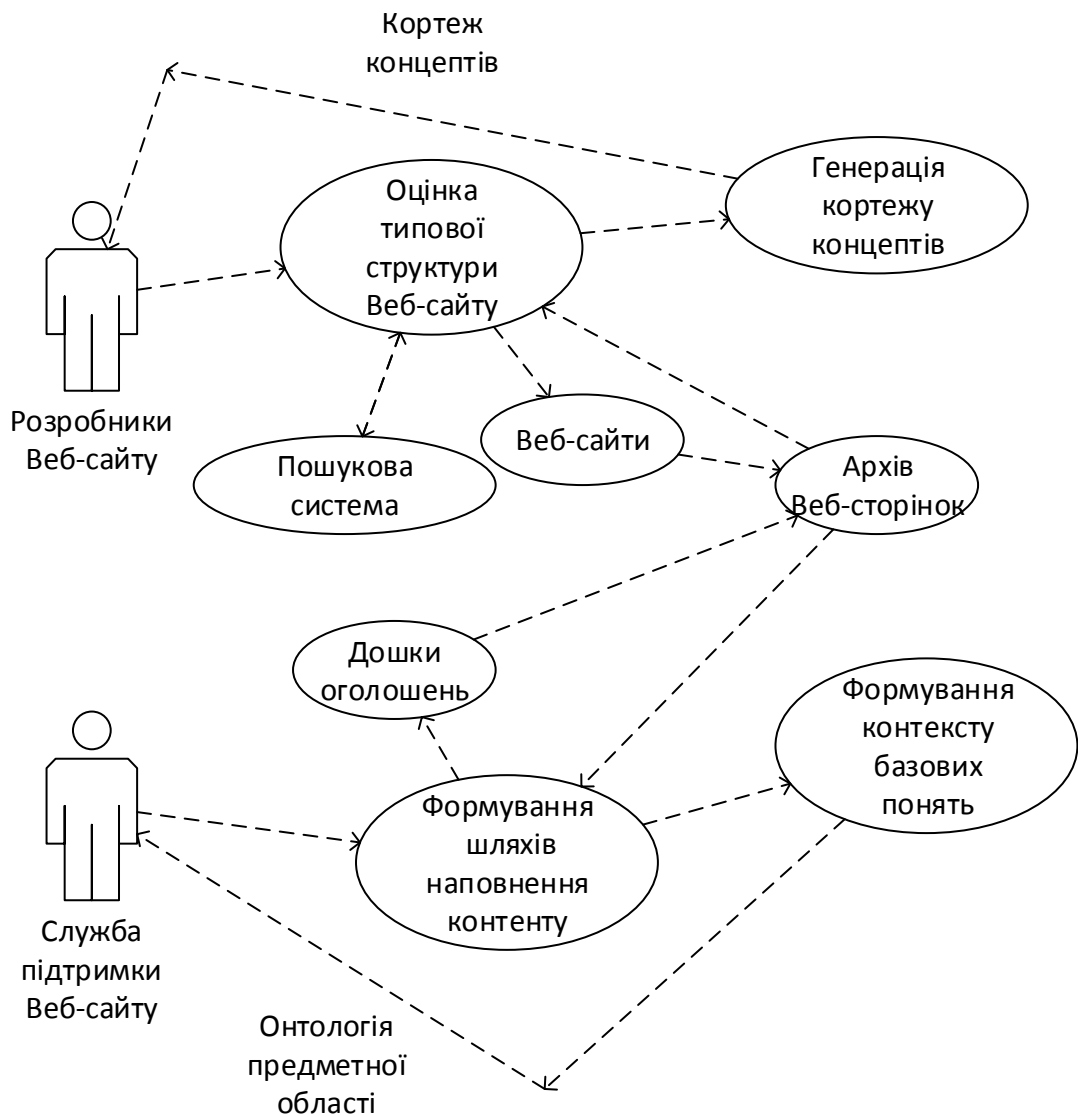


Рис. 4.1. Діаграма використання підсистем підтримки процесу структурування Веб-сайтів та формування шляхів наповнення контенту

Для програмної реалізації підсистем використано HTML та CSS кодування а також програмні одиниці, згенеровані на алгоритмічних мовах

JavaScript та PHP. Окрім того, для спрощення програмної реалізації використовується система фреймворків. Серед них Twitter Bootstrap, як набір інструментів від Twitter, створений для полегшення розробки Веб-застосунків та Веб-сайтів. Він, зокрема, включає класи оформлення для вкладок, сторінок, меню, панелі інструментів, діалогових вікон, підказок і спливаючих вікон.

Для роботи з кодом Javascript використано фреймворк jQuery з метою використання елементів DOM, обробки подій, і технології AJAX. Окрім того, з метою організації зручного інтерфейсу роботи з таблицями використовується бібліотека ExtJS. Для побудови типової структури тематичного Веб-сайту здійснюється, в першу чергу, пошук самих Веб-сайтів певної тематичної спрямованості за допомогою запитів із використанням пакету CURL в середовищі PHP до пошукової системи, аналізу результатів запиту і запису отриманих адрес в відповідний архів.

На наступному етапі відбувається звертання до відповідних Веб-сайтів із архівуванням кодів головних сторінок в архів. В подальшому HTML-код сторінок аналізується, з нього виділяються пункти меню із яких методом частотного аналізу виділяються базові поняття. В подальшому за допомогою ранжування визначається порядок пунктів меню та перевіряється узгодженість побудованих оцінок. В результаті підсистема будує типове меню тематичного Веб-сайту, а також множину неупорядкованих понять, що часто зустрічаються в типових меню.

Робота підсистеми формування шляхів наповнення контенту Веб-сайту за допомогою видобування інформації з дошок 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 Веб-сторінок. 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. Таким чином збільшується обґрунтованість впорядкування понять в онтології та зменшується обсяг інформації, який необхідно проглянути для прийняття відповідних рішень.

Реалізується описана система за допомогою діаграми класів, представленої на рис. 4.2.

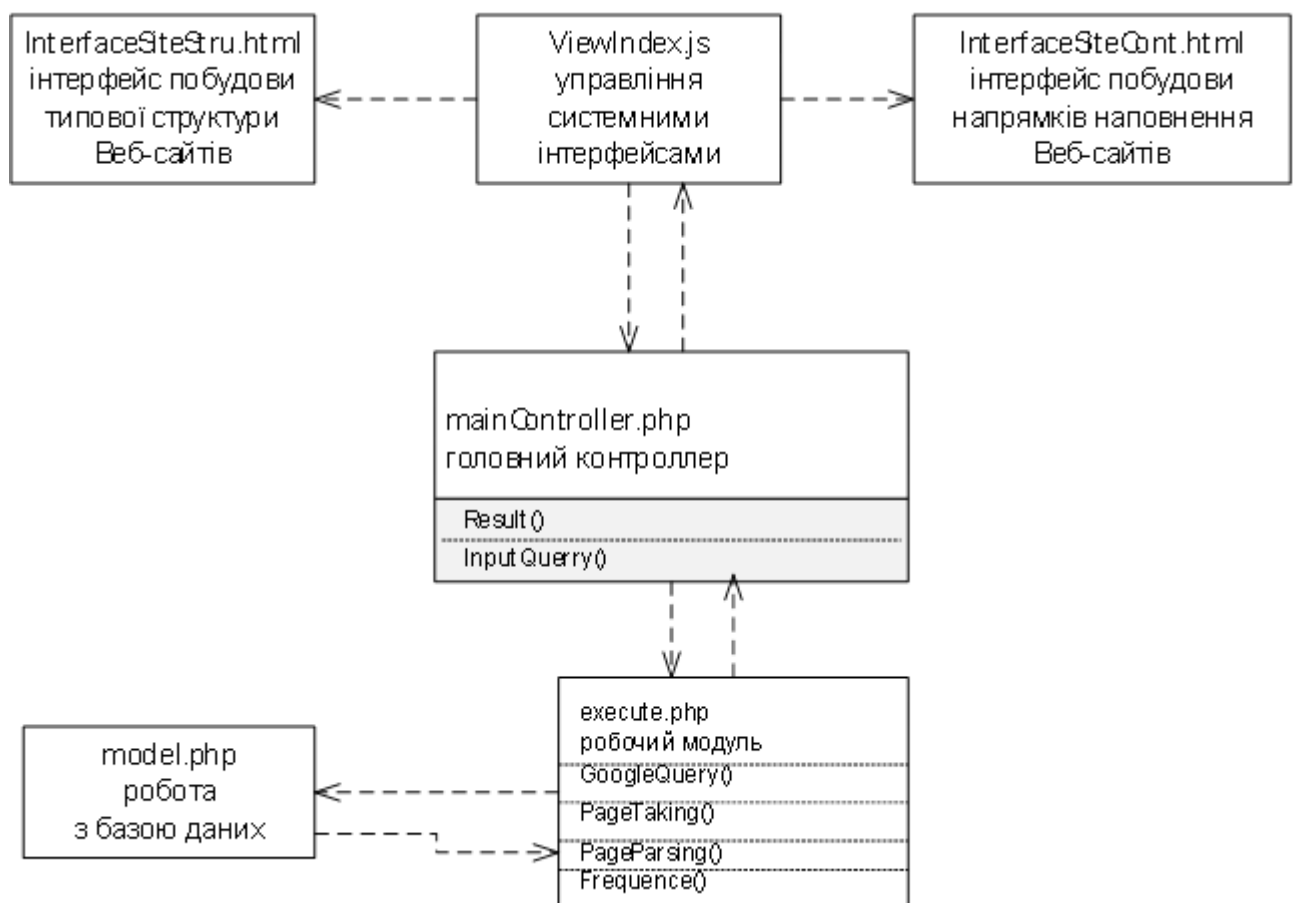


Рис. 4.2. Діаграма класів підсистеми підтримки процесу структурування Веб-сайтів

Система спроектована з використанням технології Model-View—Controller (MVC). До шару View віднесено класи інтерфейсу діалогу побудови

типової структури та напрямків наповнення Веб-сайту, а саме класи `InterfaceSiteStru` (додаток В) та `InterfaceSiteCont`. Сюди ж віднесено клас `ViewIndex` засобів управління діалогом.

Запити користувачів обробляються головним контролером `mainController` (див. Додаток А), який і подає результати виконання запитів. Основні операції обробки із структурування контенту здійснює робочий модуль `execute.php` (див. Додаток Б). Для ведення архіву підсистем та вибору даних для обробки використовується клієнтський модуль бази даних `model.php`. Таким чином програмно реалізована підсистема дозволяє поетапно будувати типові структури Веб-сайтів та формувати шляхи наповнення його контенту.

Для програмної реалізації підсистеми моделювання відвідуваності Веб-сайту використано сервіс `Google Analytics` та програмне середовище `MatLab`. Загальна структура взаємодії процесів у підсистемі моделювання відвідуваності Веб-сайту представлені на рис. 4.3 у вигляді схеми використання.

Для отримання статистики відвідуваностей Веб-сайту та його субсайтів зручно використати сервіс ведення статистики відвідувань. Із безкоштовного сервісу `Google Analytics` вибрана статистика експортується в текстовий файл, інформація з якого передається в середовище `MatLab`, де відбувається основна обробка інформації. В першу чергу дані агрегуються та згладжуються. На основі формалізованих критеріїв виділяються інтервали помітного росту активності та кандидат із експертним шляхом сформованого набору на роль фактора впливу за критерієм синхронності поведінки із загальною відвідуваністю. По кожному Веб-сайту експертним шляхом визначається тип характерних для нього приростів відвідуваності: короткотерміновий або довготерміновий.

У випадку короткотермінового приросту будуються попередні наближення функції якості оновлення контенту та відвідуваності на основі

лінійного наближення їх других похідних, отриманих на попередньому інтервалі активності.

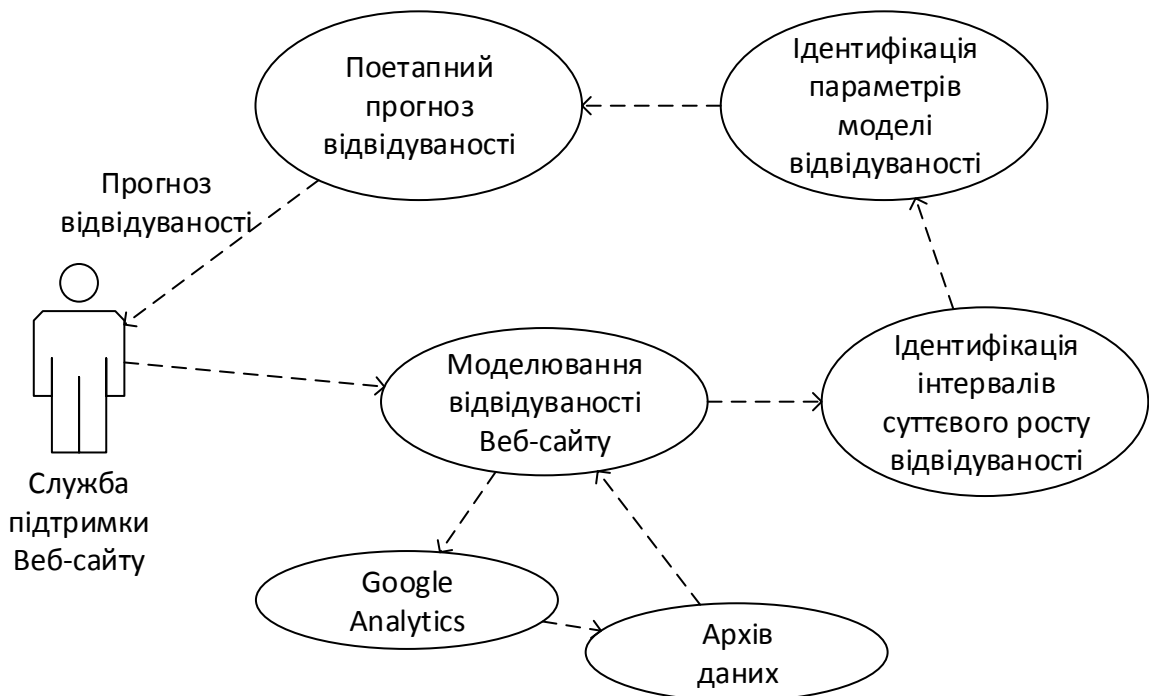


Рис. 4.3. Діаграма використання підсистеми моделювання відвідуваності Веб-сайтів

Після отримання достатньої кількості точок спостереження здійснюється побудова наближених значень параметрів функції якості та відвідуваності. В подальшому ці параметри уточнюються методом Левенберга-Марквардта. На будь-якому етапі аналізу будується прогноз відвідуваності, який в подальшому уточнюється.

У випадку довготермінового приросту будуються початкові наближення, точність яких контролюється по відхиленню в останній із спостережених точок. Таким чином вдається побудувати прогноз для часового інтервалу, спостереження якого дозволять почати процедуру ідентифікації із лінійною екстраполяцією значень функції якості оновлення контенту та звичайним середньоквадратичним функціоналом якості для відвідуваності. Після виходу за межі заданої точності ідентифікованої моделі функції якості мінімізується функціонал із лінійною екстраполяцією похідної функції якості. Реалізується описана система за допомогою набору модулів, представленого на рис. 4.4.

Зокрема окремо виділено модуль управління підсистеми. Він забезпечує вибір статистики конкретного Веб-сайту, яка на етапі попередньої обробки агрегується та згладжується за допомогою модуля AZ (додаток Г). Якщо по даному Веб-сайту не виділено інтервалу суттєвого приросту відвідуваності та активної тематики, то робиться спроба їх виділення.



Рис. 4.4. Схема підпорядкування модулів підсистеми моделювання відвідуваності Веб-сайтів

Якщо немає попередніх спостережень активності факторів короткотермінового періоду суттєвого росту відвідуваності, відбувається збір статистики по ньому. Після її нагромадження запускається модуль AD2, яким здійснюється апроксимація другої похідної аналізованих відвідуваності та функції оновлення контенту. Якщо другі похідні для функції якості оновлення контенту вже побудовано, то вони використовуються для попереднього прогнозування за допомогою модуля PD2.

У випадку отримання 5 спостережень запускається ідентифікація моделей функції якості та відвідуваності за середньоквадратичним критерієм

із застосуванням модуля ISK. Після ідентифікації моделей прогнозування значень на майбутні інтервали здійснюється за допомогою модуля PFV.

Якщо для аналізованого Веб-сайту характерні довготермінові прирости відвідуваності, то для нього будуються попередні наближення параметрів функції якості оновлення контенту та відвідуваності, а після отримання чотирьох спостережень – ідентифікація моделі функції якості та відвідуваності із лінійною екстраполяцією значень функції якості. Ці дії реалізує модуль ILED. Таким чином, програмно реалізована підсистема дозволяє поетапно моделювати та прогнозувати відвідуваність Веб-сайтів.

4.2. Приклад застосування автоматизованої процедури побудови структури Веб-сайту для цільової аудиторії навчального підрозділу

За допомогою запропонованого методу побудови структур тематичних Веб-сайтів на основі онтологічного підходу досліджено процес побудови типової структури сайту факультету українського вузу. З цією метою вибрано наступні значення параметрів методу $SAM="факультет"$, $SS="ua"$ (див (2.1)). Процес проходження першого етапу експерименту наведені на рис. 4.5.

У верхній частині форми наведені значення текстового маркера предметної області а також її специфікатора, які утворюють словосполучення "факультет ua". При натисненні кнопки "пошук" реалізується звертання до пошукової системи із видачі якої вибирається деяка множина перших M результатів, що заносяться в базу даних. В умовах даного експерименту покладено $M=100$. В подальшому із результатів запиту виділяються посилання на відповідні Веб-сторінки та відбувається вибір їхнього вмісту. Якщо тег "title" вибраної сторінки містить текстовий маркер, то код цієї сторінки теж заносяться в базу даних. В подальшому цей код опрацьовується на предмет виявлення концептів, що включені в меню відповідної Веб-сторінки. Для відслідковування динаміки процесу список опрацьованих сторінок виводиться разом із посиланнями на них. Це дозволяє переходити на них та при

необхідності здійснювати візуальний контроль наповнення. Ця опція не є необхідною при штатному функціонуванні програми і використовується лише при її відлагодженні.

| Концепт | Кількість | Частота % | Ігнорувати |
|-------------|-----------|-----------|------------|
| головна | 34 | 2.129 | ● |
| кафедри | 32 | 2.004 | ● |
| факультет | 31 | 1.941 | ● |
| наука | 30 | 1.879 | ● |
| новини | 28 | 1.753 | ● |
| абітурієнту | 28 | 1.753 | ● |
| студенту | 25 | 1.565 | ● |
| форум | 24 | 1.503 | ● |
| контакти | 22 | 1.378 | ● |
| фотогалерея | 22 | 1.378 | ● |
| бібліотека | 20 | 1.252 | ● |

Рис.4.5. Відбір базових концептів типової структури Веб-сайту факультету

На основі опрацювання текстів відібраних Веб-сторінок формуються списки відібраних концептів (табл. 4.1) в порядку спадання їхніх частот в процентному представленні. Концепти із частотами нижчими порогової не аналізуються.

Таблиця 4.1

Результати процесу відбору базових концептів

| Кількість відібраних сторінок | Кількість відібраних концептів | Перелік концептів |
|-------------------------------|--------------------------------|---|
| 13 | 11 | головна, кафедри, факультет, наука, новини, абітурієнту, студенту, форум, контакти, фотогалерея, новини, бібліотека |

У таблиці 4.2 наведені адреси сторінок зі 100 опрацьованих, за допомогою аналізу яких відбиралися концепти.

Таблиця 4.2

Характеристики Веб-сторінок остаточно відібраних концептів

| Номер сторінки | Позиція сторінки | Адреса сторінки |
|----------------|------------------|---|
| 1 | 5 | http://istfak.org.ua/ |
| 2 | 12 | http://www.history.univ.kiev.ua/ |
| 3 | 14 | http://istorikznu.at.ua/index/nash_fakultet/0-6 |
| 4 | 18 | http://fknit.com.ua/ |
| 5 | 28 | http://www-psychology.univer.kharkov.ua/ |
| 6 | 31 | http://www.university.kherson.ua/About/Faculty/Faculty_of_Law.aspx |
| 7 | 43 | http://istfak.lg.ua/ |
| 8 | 56 | http://fizmat.chnpu.edu.ua/ |
| 9 | 57 | http://fizmatsspu.sumy.ua/ |
| 10 | 58 | http://www.znu.edu.ua/ukr/university/departments/fizvosp |
| 11 | 61 | http://istorikznu.at.ua/index/nash_fakultet/0-6 |
| 12 | 70 | http://www.fitis.ck.ua/ |
| 13 | 75 | http://forlan.org.ua/index.php?option=com_content&task=section&id=15&Itemid=115 |

Рис. 4.6 наочно ілюструє принцип відбору релевантних сторінок для аналізу за критерієм наявності в заголовку сторінки ключового терміну “факультет”.

Весь Интернет

Объявление - Почему мне показано это объявление?

Фото высокого разрешения | Shutterstock.com
www.shutterstock.com/
Роялти-фри сток фотографии и иллюстрации высокого разрешения

Головна
www.tneu.edu.ua/ - Перевести эту страницу +1
Головна новина. Творчий конкурс для **студентів** "Україна-Німеччина: міжкультурний діалог". 12 березня 2012 р. Автор: Відділ інформації та зв'язків з ...
Ви посетили эту страницу несколько раз (11). Дата последнего посещения: 12.03.12

Історичний факультет Київського національного університету ...
www.history.univ.kiev.ua/ - Перевести эту страницу
Головна, Факультет, Кафедри, Наука, Студентам, Аспірантура · Докторантура, Абітурієнтам, Контакти ... (044)234-09-71 або mail: history@univ.net.ua ...
Ви посетили эту страницу несколько раз (4). Дата последнего посещения: 12.03.12

Чернівецький торговельно-економічний інститут
chtei-knteu.cv.ua/ - Перевести эту страницу
http://www.soc.chtei-knteu.cv.ua/ - місце для спілкування **студентів**, випускників та викладачів інституту. **Новини**, події, оголошення, **фотогалерея** та багато ...

istorikZNU - сайт історичного факультету ЗНУ - Наш факультет
istorikznu.at.ua/index/nash_fakultet/0-6
Розповіджені питання **абітурієнтів** · Гостьова ... BBC World News · Веб-сайт ...
На **факультете** учится 484 **студента**, в т.ч. на дневном отделении - 288.
Ви посетили эту страницу 12.03.12.

Рис.4.6. Фрагмент другої сторінки списку посилань пошукового сервера

У подальшому реалізовувалася процедура ранжування відібраних концептів. У табл. 4.3 наведено результати ранжування відібраних концептів, а табл. 4.4 ілюструє процес побудови згаданих рангів.

Концепти множини *SCN* впорядковуються по частоті згадування на відібраних сторінках. Текстові описи концептів наведені в таблиці 4.3. В рядках табл. 4.4 наводяться ранги концептів за порядком їх появи на відповідній сторінці. Будуються середні значення рангів по всіх аналізованих сторінках як основа для їх впорядкування. Для зважування значущості сторінок використані наступні параметри $g(20) = 0.8$, $g(100) = 0.1$. Вага концепту визначається як сума ваг сторінок на яких він зустрічається.

Побудова коефіцієнту конкордації за повним набором параметрів показала незначущість такого ранжування $W = 0.05$, $\chi_{emp}^2 = 6.31$, $\chi_{10,0.05}^2 = 18.31$. При послідовному редукуванні списку концептів значущі значення коефіцієнта конкордації отримано для сукупності із 9 елементів.

Таблиця 4.3

Результати встановлення рангів відібраних концептів

| 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 |
| Фотагалерея | 10 | 10 | 6 |
| Бібліотека | 11 | 7 | 6.4 |

В цьому випадку 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 $NR = \{ \text{новини, фотогалерея} \}$. Таким чином, розглянуто один із можливих шляхів зниження витрат на розробку та підтримку тематичних Веб-сайтів шляхом синтезу онтологічного дерева як загальної моделі структури цільового сегменту.

Таблиця 4.4

Середні значення рангів відібраних концептів

| № стор. | Вага стор. | Номери концептів, впорядкованих по частоті | | | | | | | | | | |
|--------------------------------|------------|--|------|------|------|------|------|------|------|------|------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | 1 | 1 | | | | | 2 | 4 | | 5 | | 3 |
| 2 | 0.8 | 1 | 6 | 3 | 2 | 5 | 4 | | 7 | | | |
| 3 | 0.78 | 1 | 5 | 2 | 4 | 6 | | 8 | | | 7 | |
| 4 | 0.73 | 1 | 8 | 4 | 3 | | | | 7 | 5 | 6 | 2 |
| 5 | 0.5 | 1 | 5 | 2 | | 4 | 3 | 6 | | | | |
| 6 | 0.49 | | 1 | 6 | 5 | 2 | | | 4 | 3 | | 7 |
| 7 | 0.48 | 1 | 3 | | 2 | | 5 | | 4 | | | |
| 8 | 0.3 | 1 | 6 | 3 | | 5 | 4 | | | 8 | 7 | 2 |
| 9 | 0.29 | 1 | | 2 | | | 3 | | | | | |
| 10 | 0.28 | 1 | 3 | | 2 | | 4 | | | | | |
| 11 | 0.27 | 1 | 4 | 2 | 3 | 5 | | 7 | | | 6 | |
| 12 | 0.2 | 1 | 7 | | 5 | 6 | | | 2 | 3 | 4 | |
| 13 | 0.19 | 1 | 2 | | 3 | | 4 | 7 | 6 | | | 5 |
| Середнє значення рангів | | 1 | 4.55 | 3 | 3.22 | 4.71 | 3.63 | 6.4 | 5 | 4.8 | 6 | 3.8 |
| Вага концепту | | 5.82 | 5.02 | 4.16 | 4.22 | 3.34 | 3.84 | 2.74 | 2.89 | 2.72 | 2.28 | 2.71 |

Підтримка високої відвідуваності Веб-сайту потребує постійного оновлення його контенту. Без попереднього оцінювання часто важко вибрати перспективні шляхи такого оновлення. Трудомісткість цього процесу можна знизити завдяки використанню спеціальних методів, зокрема запропонованого в підрозділі 2.3 методу формування шляхів оновлення контенту Веб-сайту на основі аналізу слабо структурованої інформації.

Цю діяльність можна розглядати як надання високо технологічних програмістських послуг, особливості виконання яких можна описати

онтологією поняття “PHP програміст” (“PHP programmer”). 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 в сфері ІТ. Серед цих сайтів для проведення перших експериментів вибрано сайт ”rabota.ua” та множину ключових слів, яка складається з єдиного елемента $KWS = \{ "PHP" \}$.

Серед 20 перших сторінок, що описують вакансії по даному запиту лише 10 містили спискові структури із входженням ключового слова ”PHP”. Фрагмент сторінки однієї із вакансій наведено на рис. 4.7. На даній сторінці вимоги до кандидата на заміщення посади сформовані у вигляді списку.

Крупный Восточно-Европейский интернет-холдинг открывает
вакансию

Middle php developer

График работы: полная занятость **Возраст:** 0 - 0 лет **Опыт работы:** от 1-го до 2-х лет

Успешный кандидат имеет:

Опыт работы с PHP - объектно-ориентированное программирование, понимание принципов проектирования и программирования в ООП (PHP4/PHP5).
Опыт использования одного из фреймворков (ZendFramework, CodeIgniter, Kohana и т.п.)

Рис. 4.7. Фрагмент сторінки однієї із вакансій

На рис. 4.8 представлено фрагмент моделі відношення AS , структуру якого представлено співвідношенням (2.17). В даній моделі для наочності замість посилання на батьківські елементи багатослівного терміну вони просто нумеруються, основи представлені своїми ідентифікаторами в стовпчику $IdWrd$, а словоформи подані безпосередньо в стовпчику Wrd . На рис. 4.9 наведено фрагмент списку термінів, який пропонувався для включення в онтологію разом із відповідними контекстами. Надається можливість включення в онтологію як окремого слова, натисканням кнопки біля представленого виразу, так і деякого його контексту, за допомогою кнопок в стовпчику «Обрати».

| Знайдені вирази | | | | |
|-----------------|-------|------------------|---|------|
| Оновити | | | | |
| Id | IdWrd | Wrd | | Інфо |
| 1 | 75 | Опыт | 1 | |
| | 76 | Работы | 3 | |
| | 4 | PHP | 3 | |
| 2 | 77 | обектно | 1 | |
| | 78 | ориентированое | 2 | |
| | 79 | программирование | 3 | |
| 3 | 80 | понимание | 1 | |
| | 81 | принципов | 2 | |
| | 82 | проектирования | 3 | |
| 4 | 83 | программирования | 1 | |
| | 84 | ООП | 2 | |

Рис.4.8. Фрагмент відношення аналізу вмістимого Веб-сторінок AS

| Контекст знаходження виразу | |
|-----------------------------|--------|
| Контекст | Обрати |
| strong knowledge PHP | |
| опыт работы с PHP | |
| PHP 4 | |
| опыт разработки на PHP 5 | |
| хорошее знание PHP | |
| их реализации в рnr 5.3 | |
| хороше знание PHP | |

Рис.4.9. Фрагмент списку термінів із контекстами для їх включення в онтологію

Зразок анотації першого онтологічного поняття подано на рис. 4.10. Сама онтологія, побудована на основі автоматизованого аналізу п'яти Веб-сторінок, подана на рис. 4.11, де метапоняття відображені курсивом. Як бачимо, навіть при аналізі незначного числа слабоформалізованих вимог онтологія включає базові напрямки аналізованої спеціалізації.

Для їх позначення експертом введено відповідні мета терміни. Окрім 9 термінів онтології також відібрано 8 фонових термінів, які марковані за допомогою атрибуту *Phn* відношення *BF*. Сюди були віднесені наступні терміни: “Понимание”, “Программирование”, “Проектирование”, “Knowledge”, “Опыт”, “Работа”, “Хорошее”, “Знание”.

Всього експертів для перегляду із врахуванням контексту було подано 23 стрічки, з яких відібрано 9 понять. При перегляді повних текстів 5 аналізованих сторінок експерт повинен був би переглянути біля 200 стрічок. Тобто вдалося принаймі на порядок зменшити завантаженість експерта і частково зняти інформаційну зашумленість даних. За критерієм результативності відбору важливої інформації ($\frac{\text{[кількість відібраних термінів]}}{\text{[кількість переглянутих стрічок]}}$) ефективність роботи експерта зросла від значення 0.045 до 0.391.

| HTML |
|--|
| <p>HTML (від Hypertext Markup Language — мова розмітки гіпертексту) — це стандартна мова розмітки документів у Всесвітній павутині. Всі веб-сторінки створюються за допомогою мови HTML (або XHTML). Мова HTML інтерпретується браузером і відображається у вигляді документа, зручному для людини. HTML є додатком SGML (стандартної узагальненої мови розмітки) і відповідає міжнародному стандарту ISO 8879.</p> <p>HTML-документ є текстовим файлом розмічений за допомогою спеціальних (текстових) команд. Текстовий формат представлення веб-документів був вибраний, виходячи з основних вимог до веб-документу: простота, можливість безпосередньої інтерпретації в будь-якій операційній системі, мінімальний розмір файлу, зручність редагування і інтерпретації.</p> <p>Мова розмітки гіпертекстових документів HTML дозволяє визначити різні типи елементів, що забезпечують функціональність документа: текстові фрагменти із заданими параметрами форматування, списки, таблиці, зображення, гіперпосилання і т.д. Елементи HTML оголошуються за допомогою команд розмітки, званих тегами (від англійського tag — ярлик). Всі HTML-теги, що зустрічаються в тексті документа інтерпретуються браузером при відображенні документа.</p> |

Рис.4.10. Зразок анотації онтологічного поняття

Для їх позначення експертом введено відповідні мета терміни. Окрім 9 термінів онтології також відібрано 8 фонових термінів, які марковані за допомогою атрибуту *Phn* відношення *BF*. Сюди були віднесені наступні терміни: “Понимание”, “Программирование”, “Проектирование”, “Knowledge”, “Опыт”, “Работа”, “Хорошее”, “Знание”.

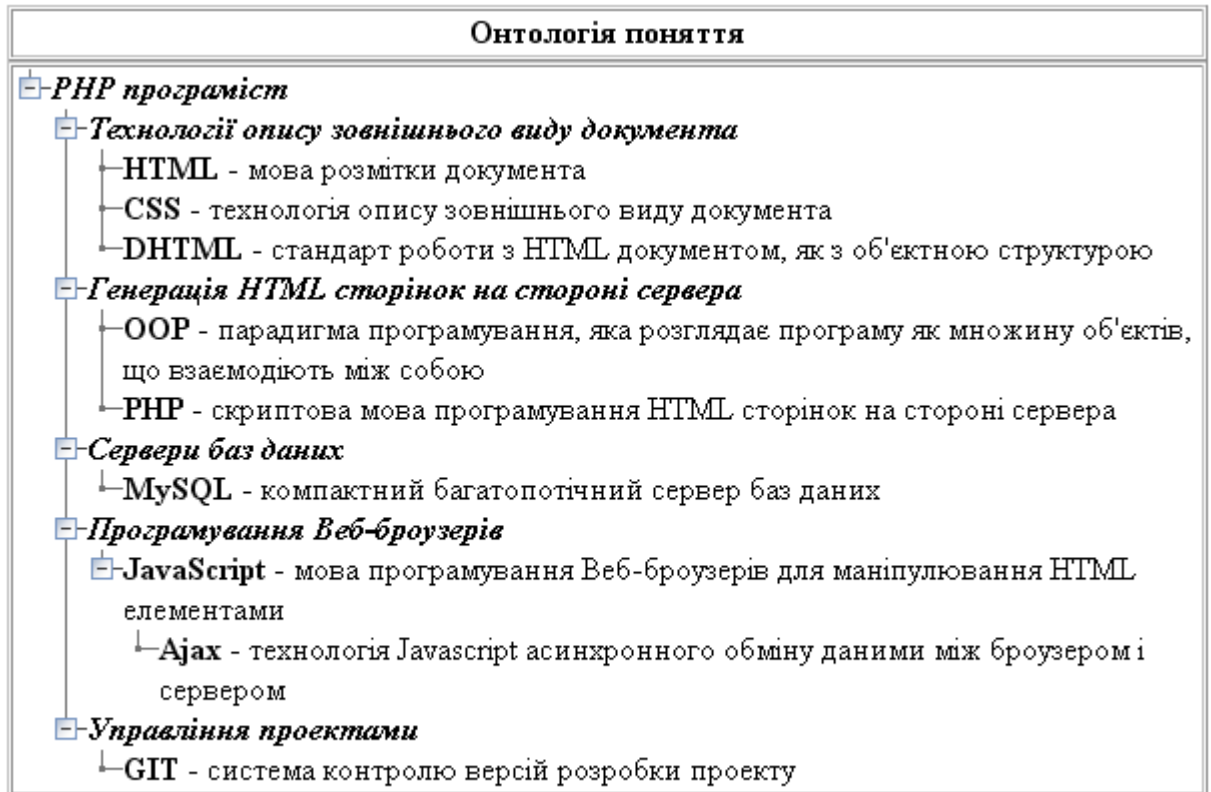


Рис.4.11. Онтологія поняття PHP програміст, побудована на основі автоматизованого аналізу п'яти Веб-сторінок

Всього експертові для перегляду із врахуванням контексту було подано 23 рядки, з яких відібрано 9 понять. При перегляді повних текстів 5 аналізованих сторінок експерт повинен був би переглянути біля 200 рядків.

Тобто вдалося принаймі на порядок зменшити завантаженість експерта і частково зняти інформаційну зашумленість даних. За критерієм результативності відбору важливої інформації ($\frac{\text{кількість відібраних термінів}}{\text{кількість переглянутих стрічок}}$) ефективність роботи експерта зросла від значення 0.045 до 0.391.

Ефективність запропонованого методу та алгоритму підтверджено при аналізі початкового етапу структурування онтології “РНР програміст”, значимої для працедавців софтверних компаній України.

4.3. Моделювання процесів підвищення відвідуваності Веб-сайтів.

В цьому підрозділі наведені результати з метою перевірки ефективності запропонованого методу моделювання короткотермінового росту відвідуваності Веб-сайту. Такий характер росту відвідуваності характерний для Веб-сайтів, відвідуваність яких не є ключовою у функціонуванні об’єкта, який цей Веб-сайт підтримує. Зокрема такий тип росту відвідуваності спостерігався у функціонуванні Веб-сайтів навчальних підрозділів.

4.3.1. Математична модель відвідуваності Веб-сайту навчального підрозділу

Проаналізуємо відвідуваність Веб-сайту факультету комп’ютерних інформаційних технологій (ФКІТ- <http://tanet.tneu.org/>) Тернопільського національного економічного університету (ТНЕУ) майже за річний період. Для отримання статистики відвідуваності необхідно увійти у систему збору статистики Веб-сайту.

На рисунках 4.12, 4.13 наведе-но фрагмент головної сторінки відповідного Веб-сайту та фрагмент головної сторінки сервісу збору його статистики. Зібрана статистика піддається процедурі агрегації та згладжування. Результат такого згладжування наведено на рис. 3.9. На рис. 3.10 наведено результат виділення інтервалів приросту відвідуваності, що дозволяє спостерігати два таких інтервали а також виділити фактори, які їх обумовлюють. Аналіз динаміки першого із виділених періодів росту відвідуваності свідчить про те, що максимум функція якості оновлення контенту досягає в третій точці від моменту фіксації початку суттєвого росту відвідуваності. Це дозволяє віднести даний ріст відвідуваності до

короткотермінових. Спостережено, що всі наступні прирости відвідуваності для даного Веб-сайту теж відносяться до даного класу.

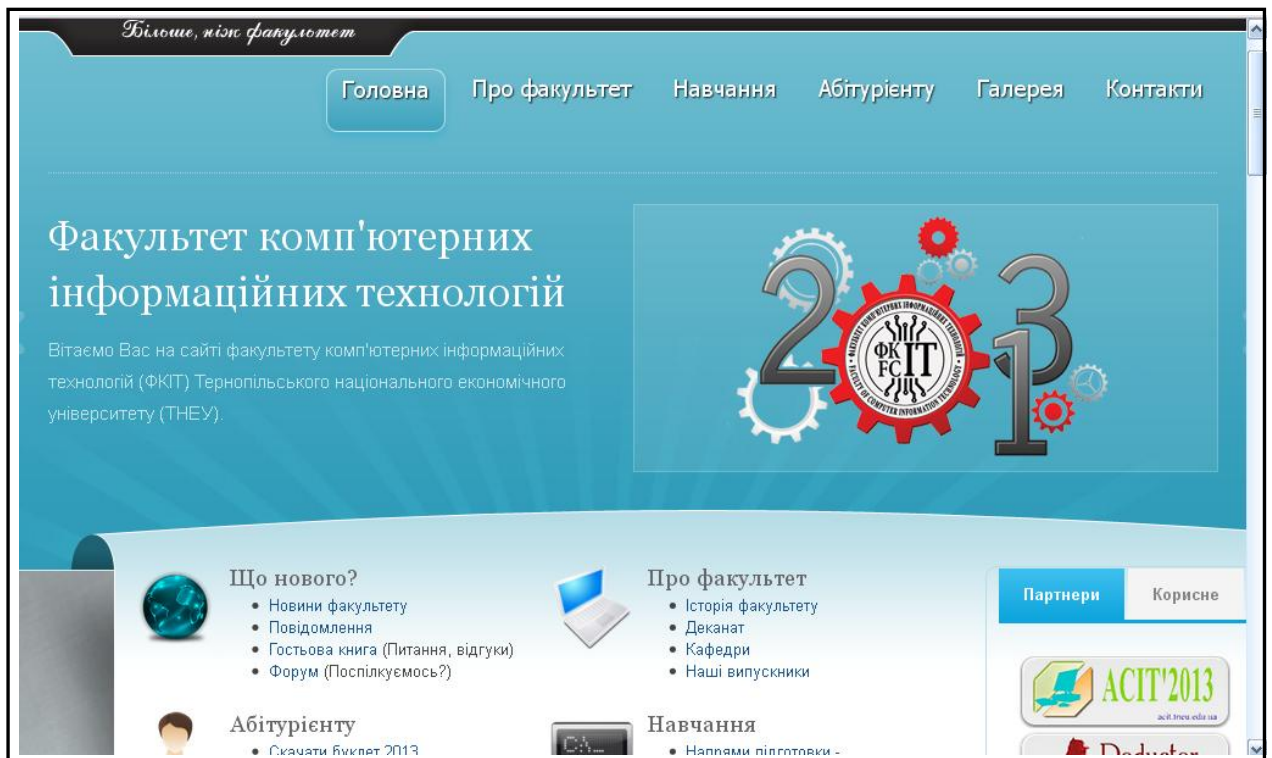


Рис. 4.12. Фрагмент головної сторінки Веб-сайту ФКІТ

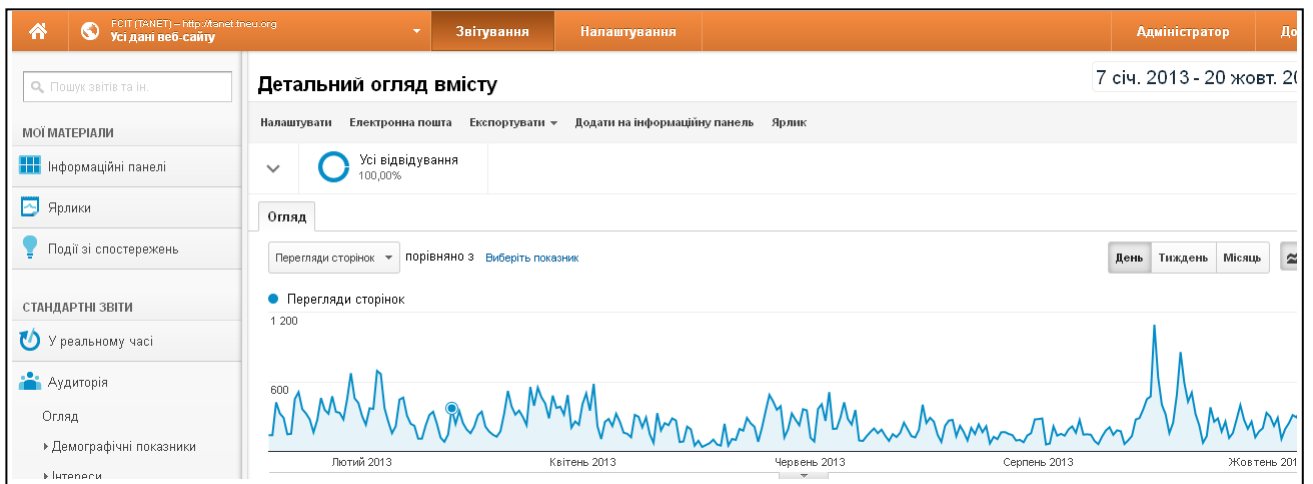


Рис. 4.13. Фрагмент головної сторінки системи збору статистики Веб-сайту із неагрегованою відвідуваністю

Це підтверджує ефективність підходу вибору типу приростів відвідуваності для конкретного Веб-сайту за аналізом типу першого із

спостережених для нього інтервалу відвідуваності. Для отримання засобів моделювання короткотермінових приростів відвідуваності, перший із них спостерігаємо від початку до завершення та будуємо лінійне наближення функції якості оновлення контенту.

Побудувати достатньо точне наближення функції якості вдається із застосуванням точок ідентифікації по всьому інтервалу активності. Графік змодельованого фактора наведено на рис. 4.14.

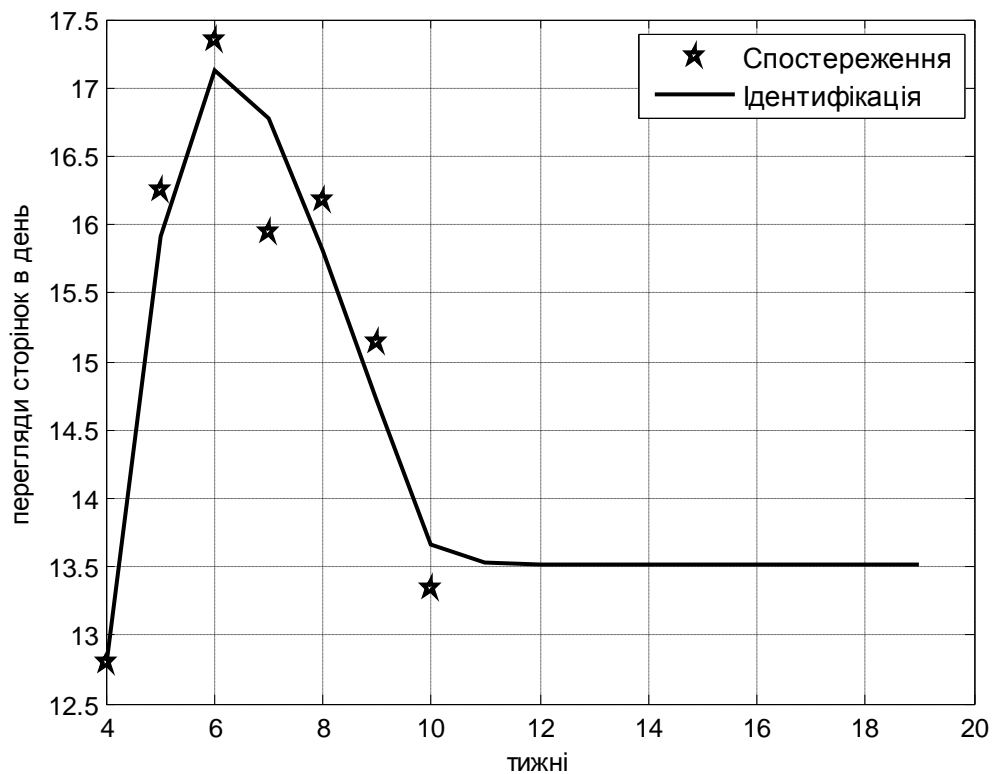


Рис. 4.14. Результат ідентифікації функції якості оновлення контенту для першого приросту відвідуваності Веб-сайту ФКІТ

Наведені результати свідчать про використання всіх семи спостережень для ідентифікації моделі. В цьому випадку вдалося побудувати модель відносна похибка якої не перевищує 4.6%. Розподіл похибки наведено на рис. 4.15.

Аналіз розподілу похибки свідчить, що її максимум досягається на четвертому спостереженні (тиждень 7), а похибки для решти спостережень не перевищують 2.5%.

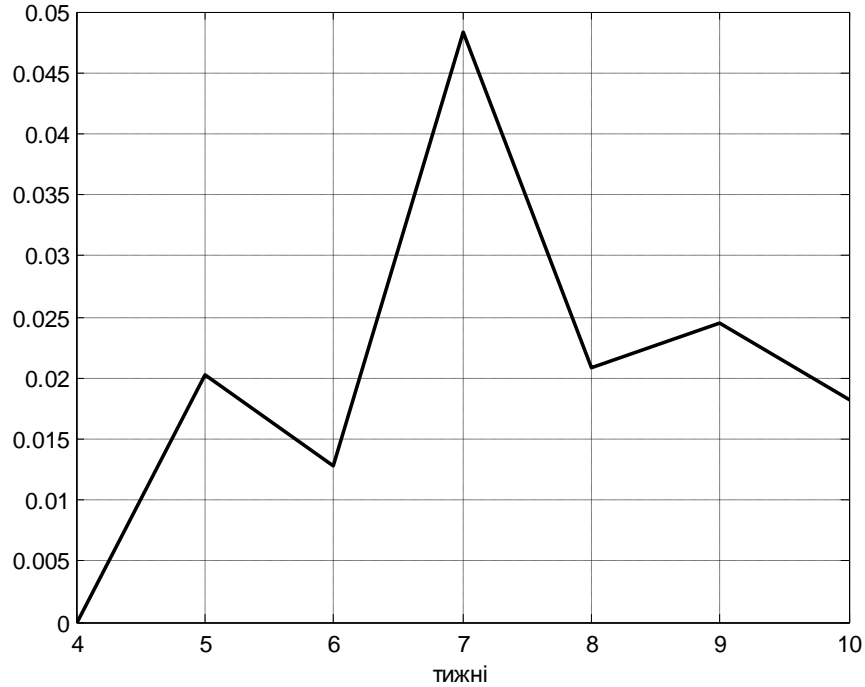


Рис. 4.15. Розподіл похибки ідентифікації функції якості оновлення контенту для першого приросту відвідуваності Веб-сайту ФКІТ

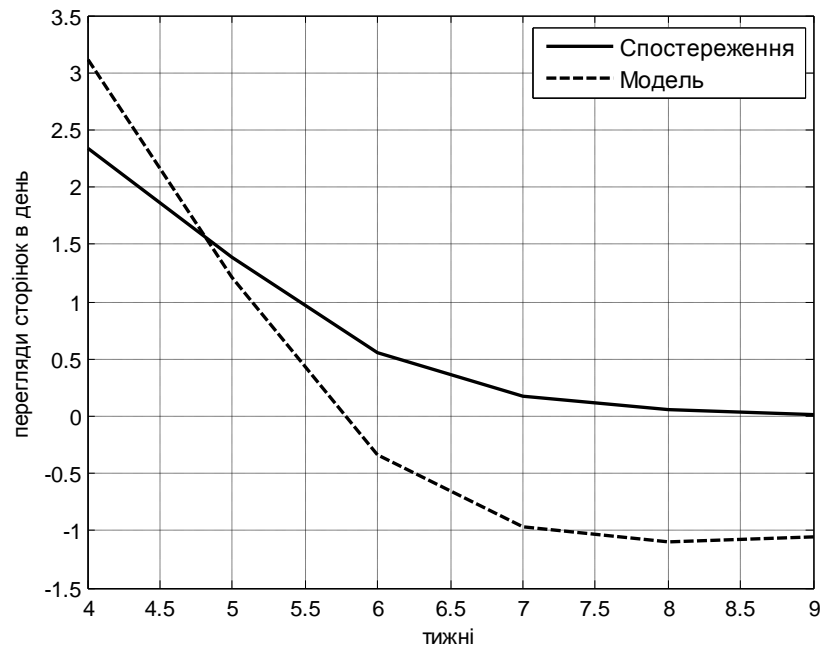


Рис.4.16. Похідна першого порядку, побудована за результатами спостережень та моделювання

На рис. 4.16 наводяться графіки першої похідної, побудованої на основі функції якості та побудованої за результатами спостережень. Порівняння цих варіантів побудови першої похідної свідчить про їх доволі суттєву відмінність. Тому переходимо до аналізу другої похідної за результатами спостережень. Відповідні графіки наводяться на рис. 4.17. Така залежність достатньо точно наближається лінійною моделлю, коефіцієнти якої будуть використані при моделюванні наступних факторів.

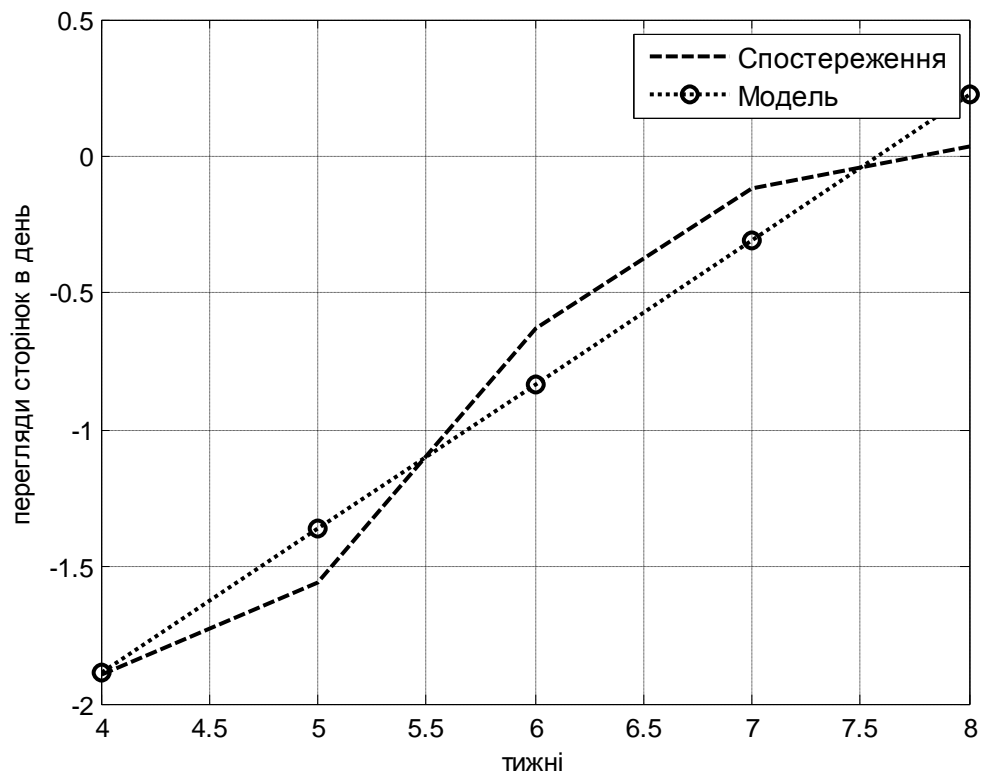


Рис. 4.17. Похідна другого порядку, побудована за результатами спостережень та її лінійне наближення

Після моделювання функції якості оновлення контенту перейдемо до моделювання відвідуваності по першому інтервалу її суттєвого приросту. Результат моделювання наведено на рис. 4.18. Рисунок свідчить про задовільну точність наближення, оскільки максимальна похибка прогнозу склала приблизно 6%.

На основі побудованої моделі другої похідної функції якості перейдемо до прогнозування наступних інтервалів суттєвого приросту відвідуваності.

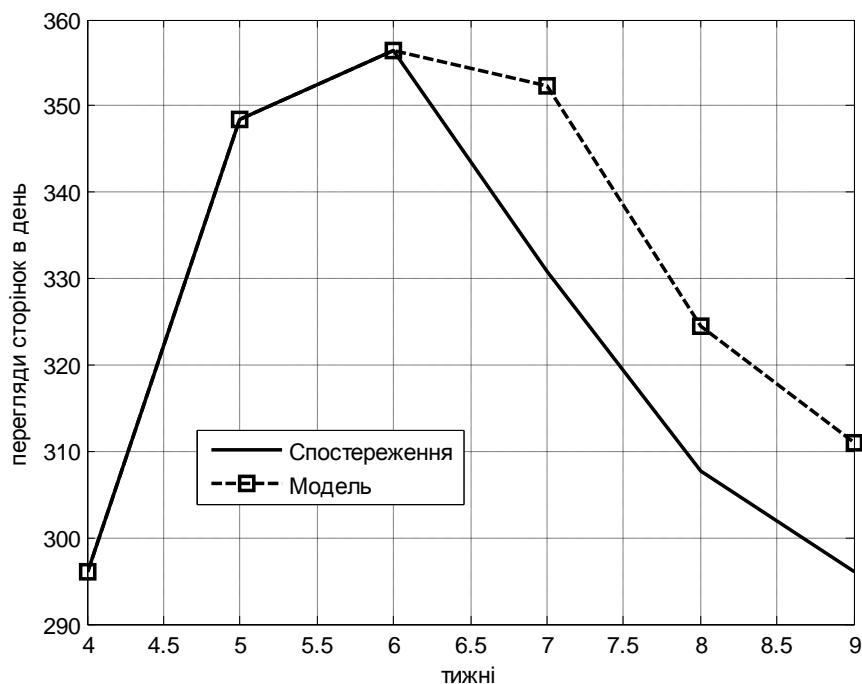


Рис.4.18. Модель відвідуваності Веб-сайту ФКІТ за функцією якості для першого етапу росту

Прогнозуючи, що другий прояв суттєвого росту активності також буде короткотерміновим, для прогнозування початкових обсягів функції якості використовуємо попередньо побудовану модель другої похідної згідно описаної вище методики.

Результат побудови початкового наближення по перших двох спостережених точках наведено на рис. 4.19. Після побудови початкового наближення моделювання функції якості оновлення контенту перейдемо до її ідентифікації. Результат моделювання із використанням семи точок ідентифікації наведений на рис. 4.20.

Рисунок свідчить, що отримана таким чином модель дозволяє здійснити прогноз по етапу із похибкою, що не перевищує 12%. Ця точність для даного виду моделі є задовільною, оскільки ідентифікація функції якості оновлення контенту Веб-сайту по всіх точках другого інтервалу суттєвого росту відвідуваності містить похибку аналогічної величини.

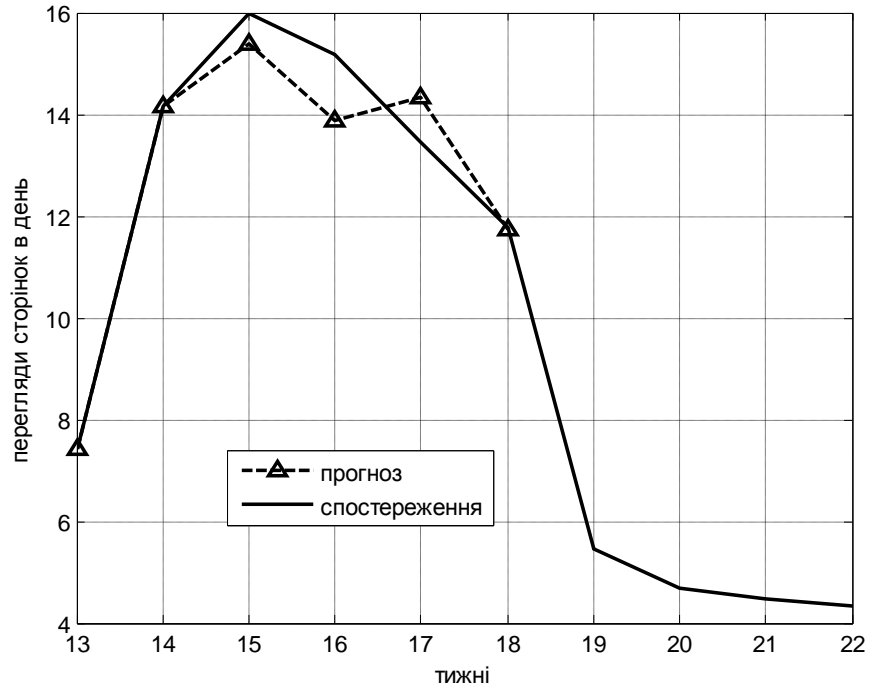


Рис. 4.19. Початкове наближення функції якості оновлення контенту для другого періоду суттєвого росту відвідуваності Веб-сайту ФКІТ

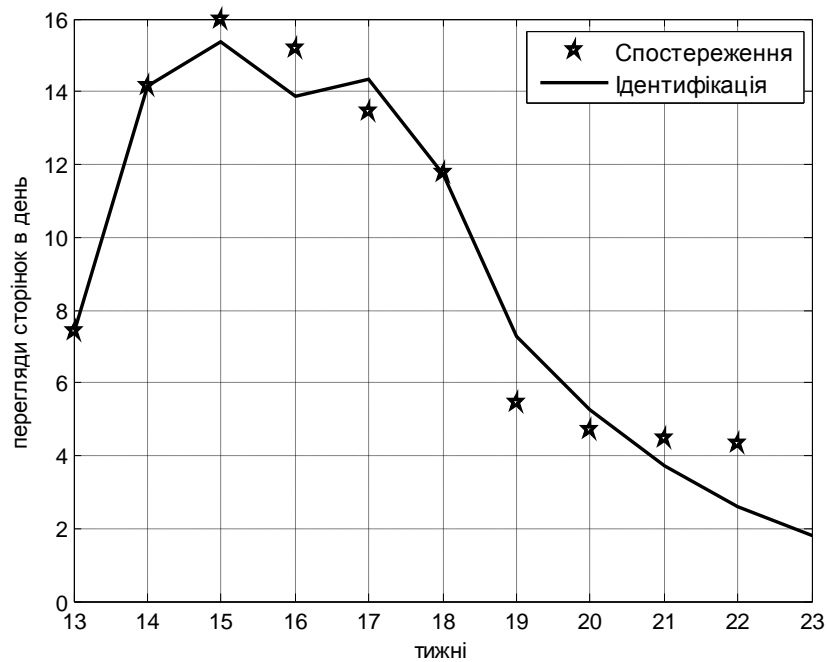


Рис.4.20. Моделювання функції якості оновлення контенту для другого періоду суттєвого росту відвідуваності Веб-сайту ФКІТ із застосуванн семи точок ідентифікації

Після моделювання функції якості оновлення контенту перейдемо до моделювання відвідуваності по другому інтервалу її суттєвого приросту. Результат моделювання наведено на рис. 4.21. Рисунок свідчить про допустиму точність наближення. При цьому максимальна похибка прогнозу склала приблизно 11%, що дещо нижче максимальної відносної похибки функції якості оновлення контенту. Результат моделювання для третього етапу росту відвідуваності із використанням семи точок ідентифікації наведено на рис. 4.22. Рисунок свідчить, що отримана таким чином модель дозволяє здійснити прогноз по етапу із похибкою, що не перевищує 12%. Ця точність для данного виду моделі є задовільною, оскільки ідентифікація функції якості оновлення контенту Веб-сайту по всіх точках третього інтервалу суттєвого росту відвідуваності містить похибку аналогічної величини.

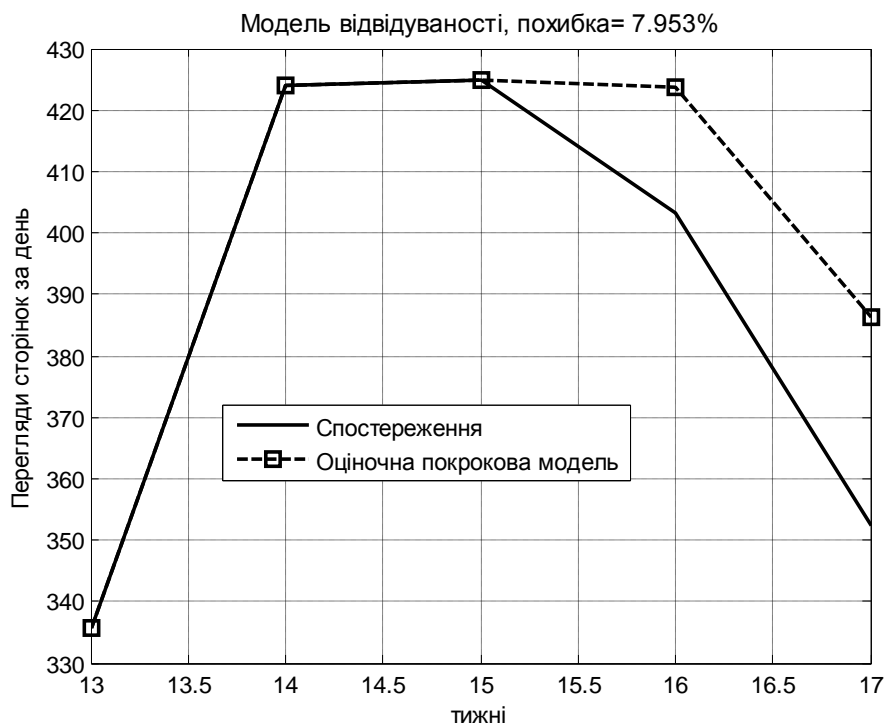


Рис.4.21. Модель відвідуваності Веб-сайту ФКІТ за функцією якості для другого етапу росту

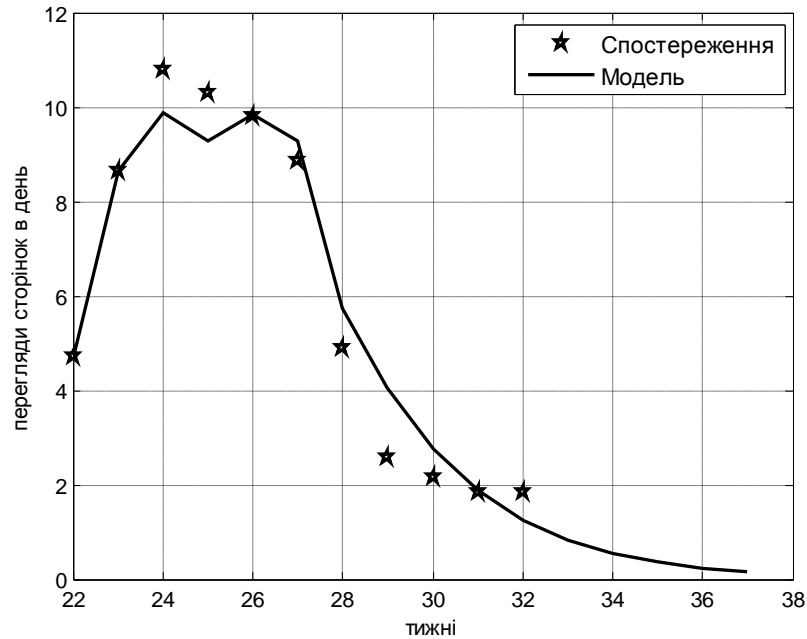


Рис.4.22 Моделювання функції якості оновлення контенту для третього етапу періоду суттєвого росту відвідуваності Веб-сайту ФКІТ із застосуванням семи точок ідентифікації

Після моделювання функції якості оновлення контенту перейдемо до моделювання відвідуваності по третьому інтервалу її суттєвого приросту. Результат моделювання наведено на рис. 4.23. Рисунок свідчить про допустиму точність наближення. При цьому максимальна похибка прогнозу склала приблизно 11%, що також нижче максимальної відносної похибки функції якості оновлення контенту.

Результат моделювання для четвертого етапу росту відвідуваності із використанням восьми точок ідентифікації наведено на рис. 4.24. Рисунок свідчить, що отримана таким чином модель дозволяє здійснити прогноз по етапу із похибкою, що не перевищує 8.5%. Ця точність для даного виду моделі є задовільною, оскільки ідентифікація функції якості оновлення контенту Веб-сайту по всіх точках четвертого інтервалу суттєвого росту відвідуваності містить похибку аналогічної величини.

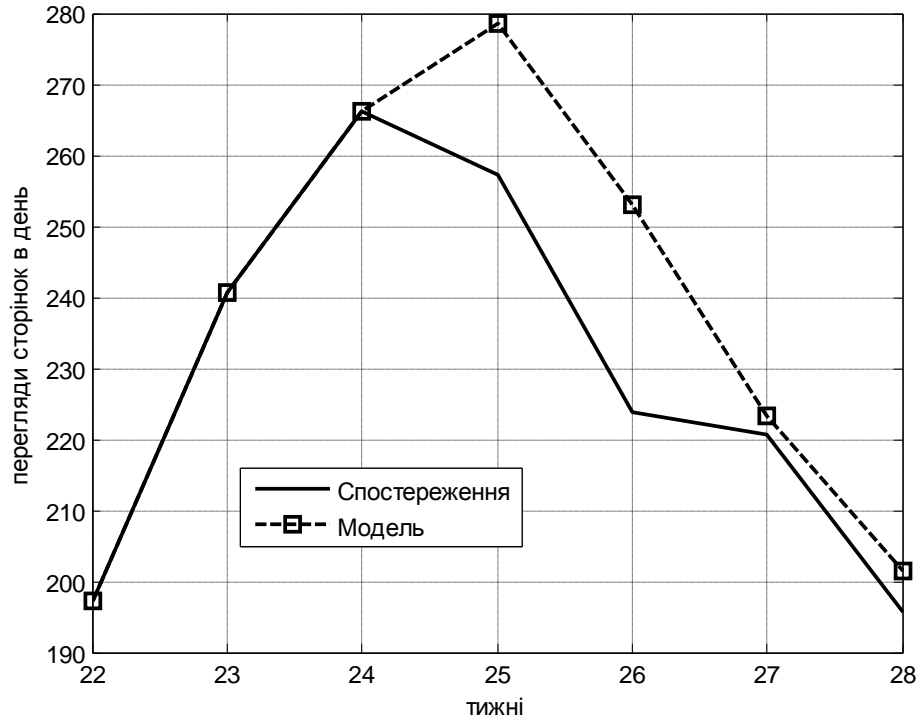


Рис. 4.23. Модель відвідуваності Веб-сайту ФКІТ за функцією якості для третього етапу росту

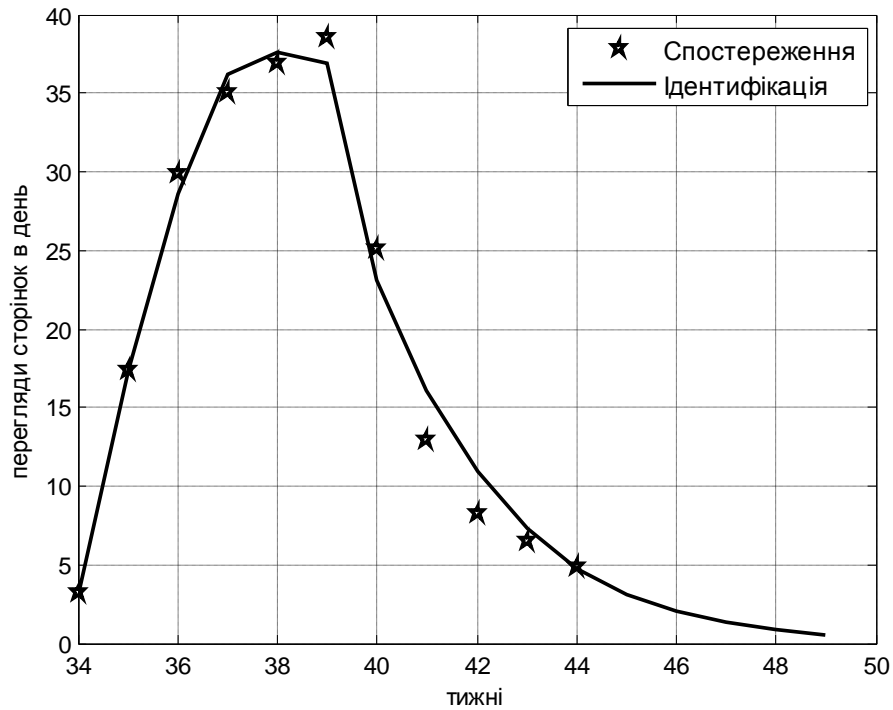


Рис. 4.24. Моделювання функції якості оновлення контенту для четвертого етапу періоду суттєвого росту відвідуваності Веб-сайту ФКІТ із застосуванн восьми точок ідентифікації

Модель функції якості оновлення контенту для четвертого етапу росту відвідуваності служить основою для моделювання обсягу росту цієї відвідуваності. Результат моделювання наведено на рис. 4.25. Рисунок свідчить про допустиму точність наближення. При цьому максимальна похибка прогнозу склала приблизно 13%, що дещо вище максимальної відносної похибки функції якості оновлення контенту.

Аналіз проведених експериментів підтверджує допустимість запропонованого підходу в моделюванні короткотермінового росту відвідуваності Веб-сайту на основі функції якості оновлення його контенту. При побудові функції якості оновлення контенту попередньо використовуються два спостереження по яких виділявся період суттєвого росту відвідуваності.

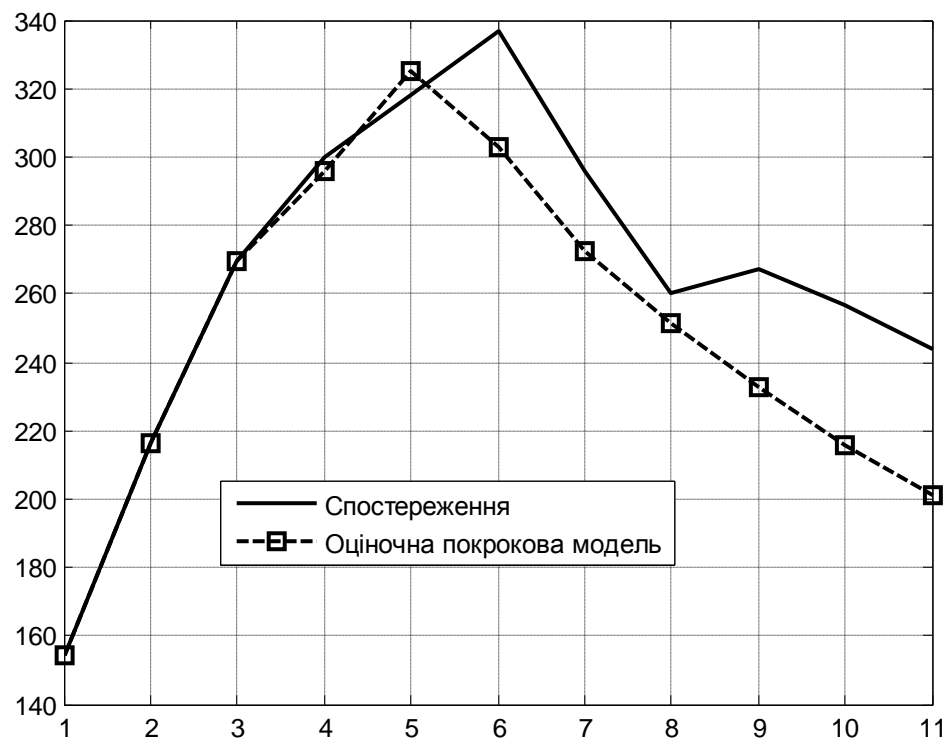


Рис. 4.25. Модель відвідуваності Веб-сайту ФКІТ за функцією якості для четвертого етапу росту

На основі моделі другої похідної функції якості першого спостереженого етапу підвищення відвідуваності генерувалися початкові значення функції

якості оновлення контенту для наступних другого, третього та четвертого етапів. При цьому задовільна точність такого генерування зберігалася для наступних 4-5 точок, тобто на інтервалі основного приросту значень функції якості.

Спостереження значень функції якості оновлення контенту в 7-8 точках, які попередньо оцінювалися на основі згенерованих значень дозволило використати процедуру ідентифікації параметрів згаданої функції. В результаті із задовільною точністю було наближено значення функції якості ще для 3-4 точок.

Модель відвідуваності будувалася на основі значень функції якості оновлення контенту. Розглядався лише той період, на протязі якого значення відвідуваності перевищувало значення на початку періоду суттєвого її росту. При цьому прогнозний інтервал склав 4-5 точок, що відповідає періоду у 30-35 днів. Це дає час адміністрації Веб-сайту для прийняття рішення про недостатню ефективність конкретної акції із підвищення відвідуваності Веб-сайту а також пошуку засобів та ресурсів для підвищення такої ефективності.

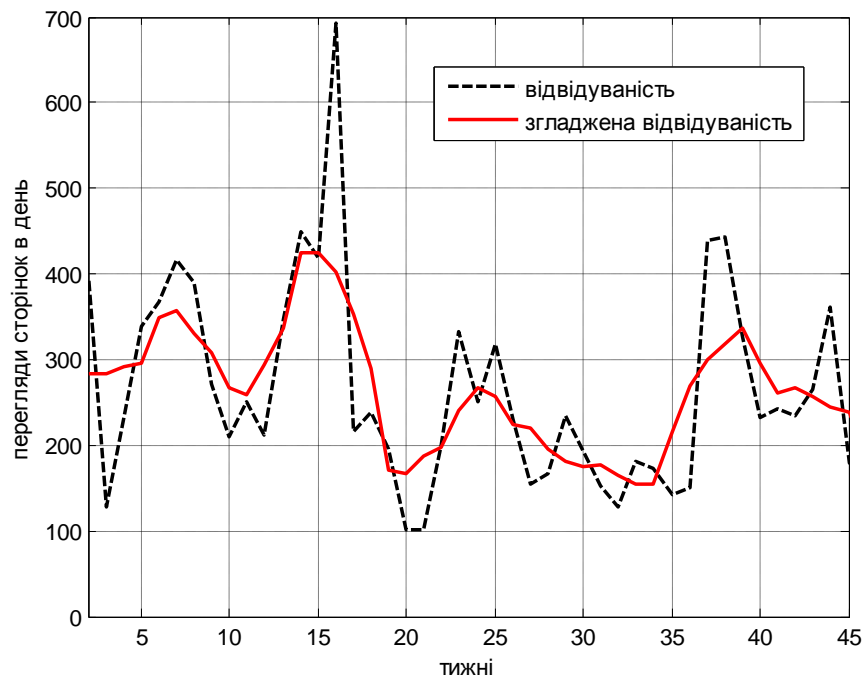


Рис. 4.26. Згладжена відвідуваність Веб-сайту ФКІТ за неповний річний період

Відвідуваність Веб-сайту ФКІТ за річний період, досягнута із використанням засобів моделювання, подана на рис. 4.26.

Таким чином на основі одного спостереженого періоду суттєвого росту відвідуваності змодельовано три наступні періоди суттєвого росту. Другий період суттєвого росту спричинений розробкою нової рубрики Веб-сайту на основі онтології «РНР програміст», побудову якої описано в попередньому підрозділі. Цей ріст відвідуваності має найвищу амплітуду, що не менш як на 20% перевершує інші. Відзначимо, що перший та четвертий ріст відвідуваності викликаний ознайомленням студентів із структурами курсів на початку семестрів, а третій викликаний активністю абітурієнтів під час вступної компанії. Розроблений засіб моделювання забезпечив найсуттєвіший ріст відвідуваності а також виявив перспективний напрямок її нарощення при системній роботі в даному напрямку. Представляє інтерес також моделювання відвідуваності Веб-сайту, над підвищенням якої працюють системно та професійно. Аналіз такого роду моделювання проводиться в наступному підрозділі.

4.3.2. Математична модель відвідуваності промо сайту Api2Cart

Після аналізу результатів моделювання короткотермінових періодів росту відвідуваності перейдемо до експериментів із довготерміновими періодами росту. Такий характер росту відвідуваності характерний для Веб-сайтів, підвищенням відвідуваності яких займаються команди служб підтримки на регулярній основі. Зокрема такий тип росту відвідуваності спостерігався у функціонуванні Веб-сайтів компаній із розробки програмного забезпечення. Тому проаналізуємо, наприклад, відвідуваність промо сайту сервісу API2Cart компанії Magnetic One (<http://www.api2cart.com/>) більш ніж за річний період.

На рис. 4.27 наведено фрагмент головної сторінки досліджуваного Веб-сайту. Зібрана статистика по відвідуваності Веб-сайту також піддається

процедурі агрегації та згладжування. Результат такого згладжування наведено на рис. 3.12.



Рис.4.27. Фрагмент головної сторінки Веб-сайту API2Cart

На рис. 3.13 наведено результат виділення інтервалів приросту відвідуваності, що дозволяє спостерігати два таких інтервали, а також виділити фактори впливу по кожному із них. Аналіз динаміки відвідуваності для першого із виділених інтервалів суттєвого її приросту свідчить, що максимум вона досягає в сьомій точці від моменту фіксації початку інтервалу. Це дозволяє віднести даний ріст відвідуваності до довготермінових. Всі наступні прирости відвідуваності для даного Веб-сайту теж відносяться до даного класу.

Побудувати достатньо точне наближення першого прояву функції якості оновлення контенту можна із застосуванням точок ідентифікації по всьому інтервалу активності. Графік змодельованого фактора наведено на рис. 4.28.

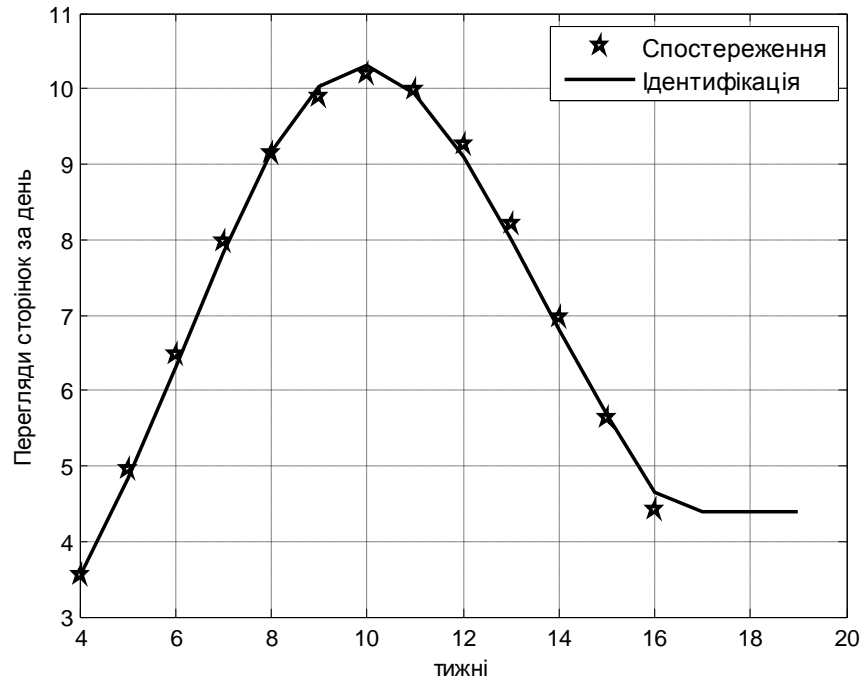


Рис. 4.28. Результат ідентифікації функції якості оновлення контенту для першого приросту відвідуваності Веб-сайту API2Cart

В цьому випадку вдалося побудувати модель, відносна похибка якої не перевищує 2.4%. Однак досягти такої високої точності, використовуючи невелику частину точок ідентифікації, надзвичайно важко. Тому використовується запропонований метод ідентифікації моделі довготривалого фактора впливу. Згідно першого підходу в реалізації методу використовується всього два спостереження модельованої величини. На їх основі будується значення параметрів моделі, які забезпечують достатню точність для кількох початкових точок. Результат реалізації першого підходу в моделюванні функції якості оновлення контенту для третього приросту відвідуваності Веб-сайту API2Cart наведено на рис. 4.29.

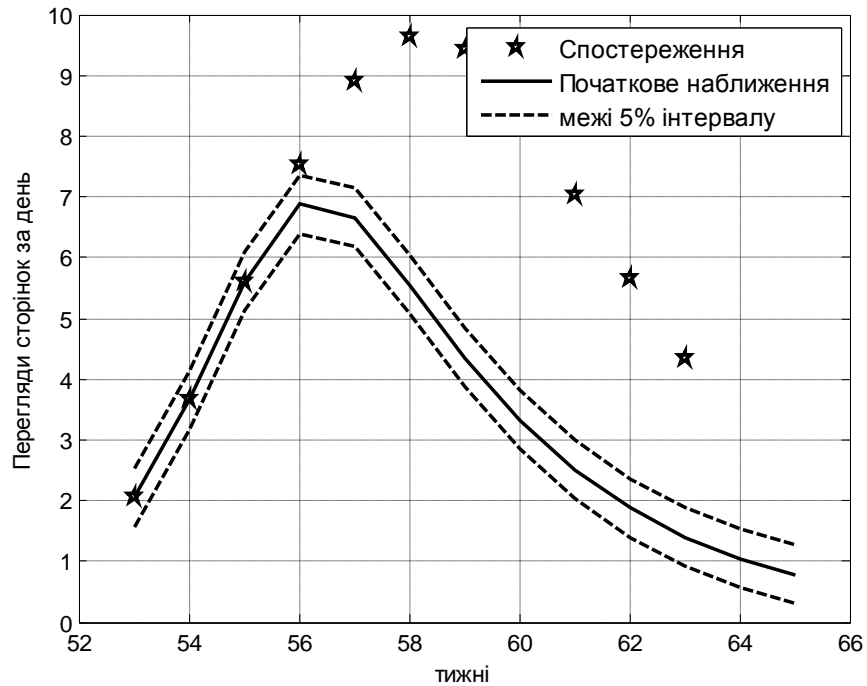


Рис. 4.29. Результат першого підходу в моделюванні функції якості оновлення контенту для третього приросту відвідуваності Веб-сайту API2Cart

Аналіз рисунку свідчить, що похибка даного наближення не перевищує 5% лише для однієї наступної точки, а вже для четвертої точки спостереження похибка моделі перевищує 5% межу. В цьому випадку реалізується наступний підхід в реалізації методу, згідно якого розпочинається ідентифікація параметрів моделі за функціоналом якості, який включає не лише відхилення від спостережених значень, а й від їхньої лінійної екстраполяції. Значення спостережених та екстрапольованих даних для наочності наводяться на рис. 4.30.

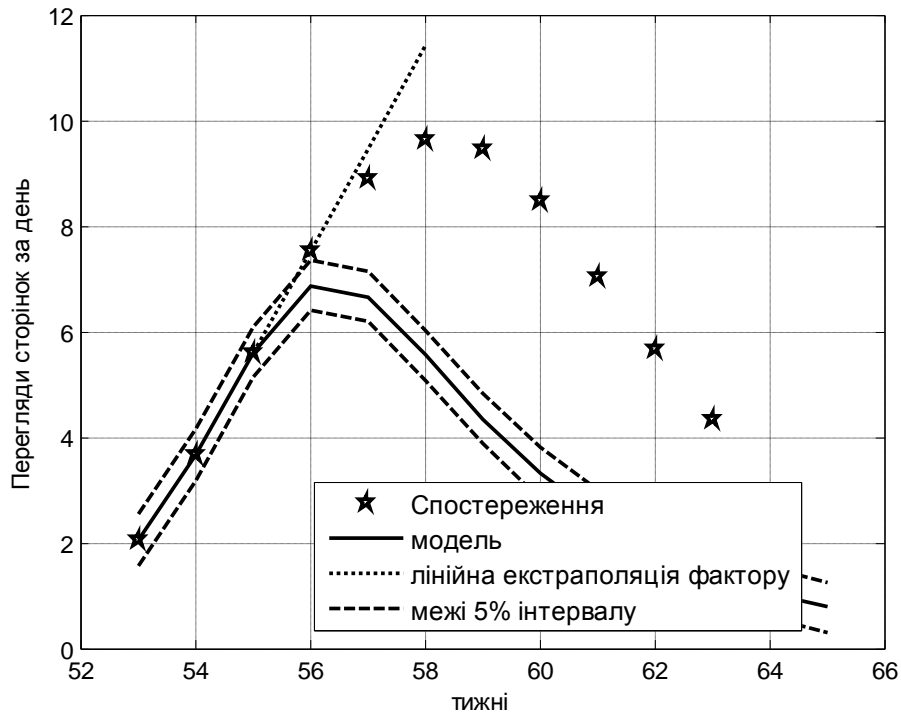


Рис. 4.30. Результат першого підходу в моделюванні функції якості оновлення контенту для третього приросту відвідуваності Веб-сайту API2Cart в сукупності із лінійною екстраполяцією функції якості

Аналіз результату моделювання, представленого на рис. 4.31, свідчить про задовільну точність моделі для семи спостережень. У восьмій точці спостереження точність моделі перевищує 5% рівень. Тому для наступного прогнозування ідентифікуємо модель по восьми точках із застосуванням функціоналу якості, який включає, окрім відхилення від спостережених значень модельованої величини, відхилення похідної моделі від лінійної екстраполяції похідної фактора впливу. Результат прогнозування по моделі наводяться на рис. 4.32. Аналіз рисунку свідчить про задовільну точність прогнозування для дев'яти спостережених значень. Десяте спостереження відхиляється від прогнозу більше ніж на 5%.

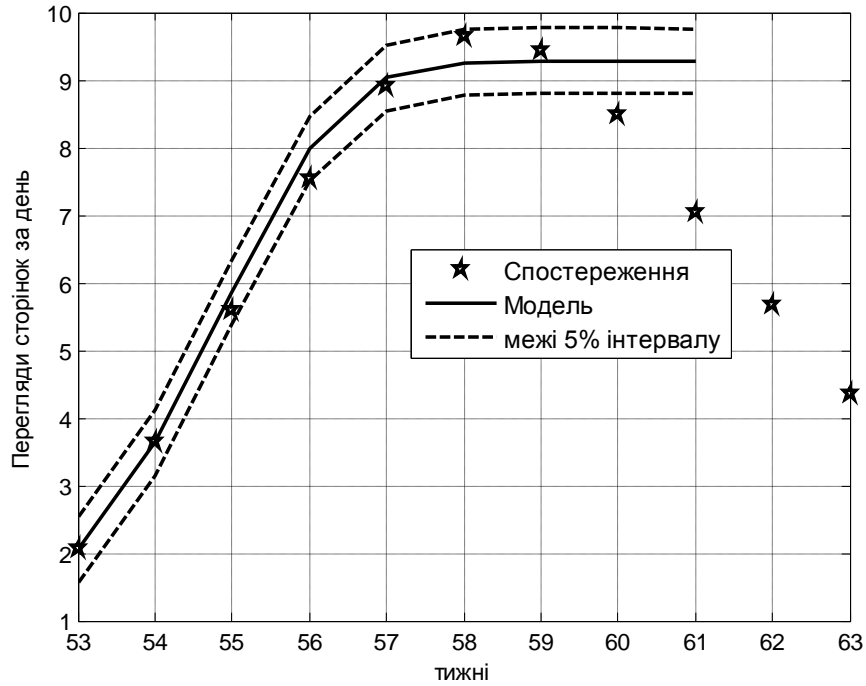


Рис. 4.31. Результат другого підходу в моделюванні функції якості оновлення контенту для третього приросту відвідуваності Веб-сайту API2Cart

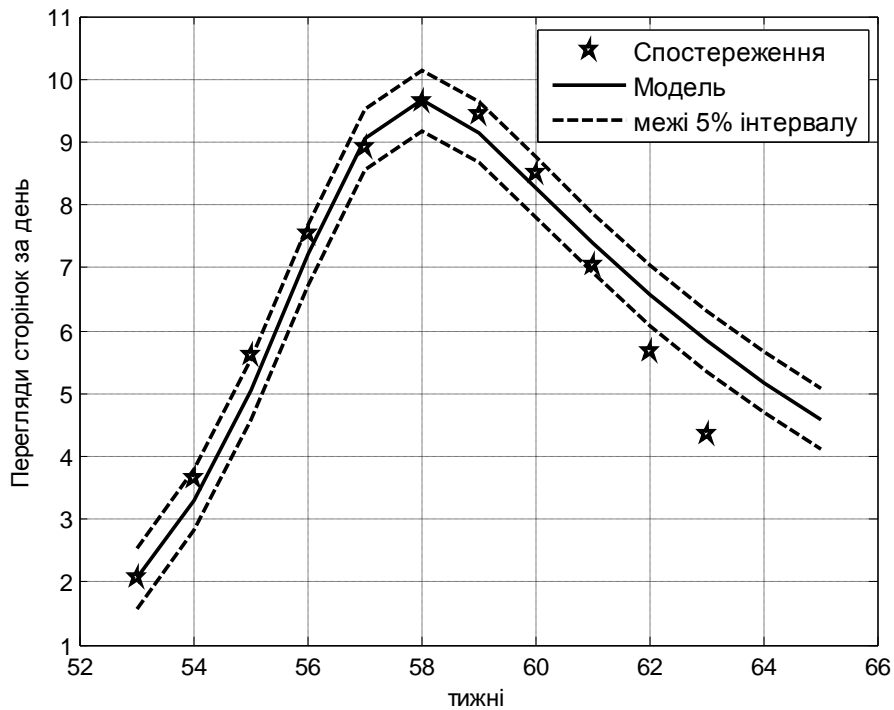


Рис. 4.32. Результат третього підходу в моделюванні функції якості оновлення контенту для третього приросту відвідуваності Веб-сайту API2Cart із використанням екстраполяції похідної спостережених значень

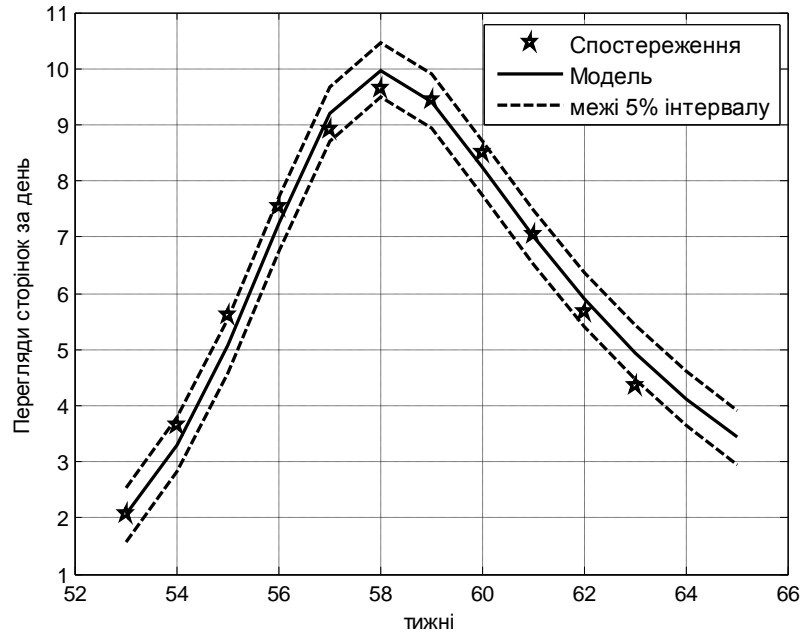


Рис. 4.33. Результат ідентифікації моделі фактору впливу електронних розсилок на відвідуваність Веб-сайту API2Cart із використанням екстраполяції похідної спостережених значень

Після моделювання функції якості оновлення контенту розглянемо моделі відвідуваності Веб-сайту API2Cart. На відміну від моделювання функції якості, модель відвідуваності вдається побудувати за початковим наближенням по двох спостережених значеннях і ця модель, побудована на основі моделі фактора впливу не потребує подальших уточнень. Відповідний графік наведено на рис. 4.34. В цьому випадку багатоетапність прогнозу відвідуваності пов'язана лише із багатоетапністю ідентифікації фактора впливу. Похибка моделі не перевищує 6.6%.

Після моделювання функції якості оновлення контенту розглянемо моделі відвідуваності Веб-сайту API2Cart. На відміну від моделювання функції якості, модель відвідуваності вдається побудувати за початковим наближенням по двох спостережених значеннях і ця модель, побудована на основі моделі фактора впливу не потребує подальших уточнень. Відповідний графік наведено на рис. 4.34. В цьому випадку багатоетапність прогнозу

відвідуваності пов'язана лише із багатоетапністю ідентифікації фактора впливу. Похибка моделі не перевищує 6.6%.

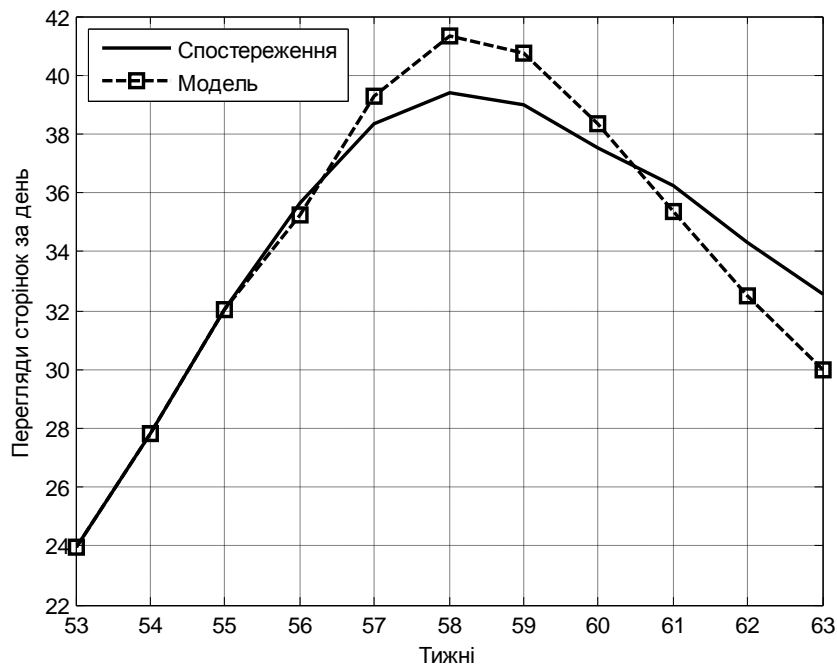


Рис. 4.34. Результат ідентифікації моделі третього інтервалу росту відвідуваності Веб-сайту API2Cart

Аналіз проведених експериментів підтверджує допустимість запропонованого підходу в моделюванні довготермінового росту відвідуваності Веб-сайту на основі функції якості оновлення його контенту. При побудові функції якості оновлення контенту попередньо використовуються два спостереження по яких виділяється період суттєвого росту відвідуваності. На основі першого підходу в методі ідентифікації генеруються початкові значення параметрів функції якості для наступних 1-2 точок.

Спостереження значень функції якості оновлення контенту в 3-4 точках із використанням екстраполяції цих значень в оптимізаційному функціоналі, дозволило застосувати процедуру ідентифікації параметрів згаданої функції. В результаті із задовільною точністю було наближено значення функції якості ще для 3-4 точок. В подальшому при модифікації критерія оптимізації

за рахунок включення лінійної екстраполяції похідної вдається ідентифікувати модель в наступних 3-4 точках.

Модель відвідуваності будувалася на основі значень функції якості оновлення контенту згідно співвідношення (3.31). При цьому прогнозний інтервал склав два періоди по 3-4 точок, що відповідає тривалості у 20-25 днів. Це дає час адміністрації Веб-сайту для прийняття рішення про необхідність зміни курсу конкретної акції із підвищення відвідуваності Веб-сайту а також пошуку засобів та ресурсів для підвищення такої ефективності.

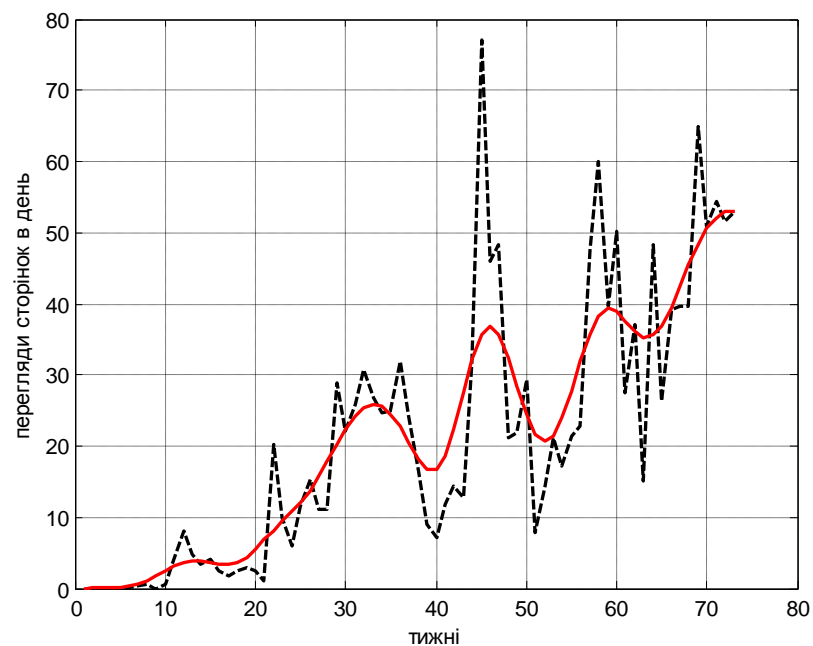


Рис. 4.35. Згладжена відвідуваність Веб-сайту API2Cart за півторарічний період

Аналогічно змодельовано ще два періоди росту відвідуваності Веб-сайту API2Cart. Таким чином всього змодельовано чотири періоди росту відвідуваності. Результати моделювання дозволили регулювати інтенсивність заходів щодо підвищення відвідуваності. Це забезпечило постійний ріст відвідуваності, що наочно представлено на рис. 4.35. Перший та четвертий ріст відвідуваності забезпечений рекламою на зовнішніх спеціалізованих Веб-сайтах, Другий ріст відвідуваності забезпечений анонсуванням рубрик даного Веб-сайту на інших Веб-сайтах компанії що виробляє програмне

забезпечення, а третій ріст відвідуваності спричинений направленими розсиленнями засобами електронної пошти.

В результаті трьох направлених акцій відвідуваність зросла на 108%, тобто в середньому на 36% за одну акцію. Систематична діяльність над підвищенням відвідуваності дала відчутний ефект, а використання засобів моделювання дозволило раціонально використати ресурси компанії із його забезпечення.

Висновки до розділу 4

1. З метою автоматизації процедур оцінки ефективності структурування Веб-сайтів розроблено комплекс прикладних програм із застосуванням технологій PHP, JavaScript та MatLab, який дозволяє автоматизувати процеси структурування Веб-сайту, формування шляхів оновлення контенту та прогнозувати результативність акцій щодо підвищення відвідуваності сайту.
2. Згідно запропонованого методу побудови структур тематичних Веб-сайтів на основі онтологічного підходу досліджено процес побудови типової структури сайту факультету українського вузу.
3. Проведено експериментальні дослідження управління процесом розміщення контенту Веб-сайту. Показано, що трудомісткість цього процесу можна знизити завдяки використанню запропонованого методу формування шляхів оновлення контенту Веб-сайту на основі аналізу слабо структурованої інформації. За критерієм результативності відбору важливої інформації ефективність роботи експерта зросла від значення 0.045 до 0.391.
4. Ефективність побудованих прогнозних моделей відвідуваності Веб-сайту 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 планувати заходи із підвищення відвідуваності Веб-сайту.

ВИСНОВКИ

В дисертаційній роботі розв'язане актуальне наукове завдання підвищення відвідуваності Веб-сайтів шляхом розроблення процедур структурування контенту із використанням програмних засобів та інструментів математичного моделювання.

При цьому отримано такі наукові та практичні результати:

1) Обгрунтовано використання методів Web Miningu а також 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, дозволяє 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 даного методу зростає в середньому в 9 разів.

3) На основі згладжування агрегованих відвідуваностей Веб-сайту та субсайтів і застосування системи звичайних диференціальних рівнянь Моно побудована 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 дозволяє врахувати стрімку динаміку короткотермінового фактора впливу.

4) Для довготермінових прогнозів відвідуваності запропоновано 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

дозволяє врахувати неоднорідність динаміки функції якості для прогнозування довготермінової відвідуваності під її впливом.

5) З метою автоматизації процедур оцінки ефективності структурування Веб-сайтів розроблено комплекс прикладних програм із застосуванням технологій PHP, JavaScript та MatLab, який реалізує підтримує математичних моделей. 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 дозволило збільшувати відвідуваність в середньому на 28% за період її суттєвого росту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Анатольев А. Г. Веб-программирование. Конспект лекций [Электронный ресурс] / А.Г. Анатольев. — 2003. — Режим доступа : www.4stud.info/web-programming/lecture1.html. — Назва з екрану.
2. Анищенко В.С. Динамические системы [Электронный ресурс] / В. С. Анищенко // Соросовский образовательный журнал — 1997. — № 11. — С. 77-84. — Режим доступа до журналу: http://sins.хаос.ru/pdf/articles/articles_r043.pdf. — Назва з екрану.
3. Ашихмин В.Н. Введение в математическое моделирование / [В. Н. Ашихмин, М. Б. Гитман, И. Э. Келлер и др.] ; под ред. П.В. Трусова. — Москва : Университетская книга, Логос, 2007. — 440 с.
4. Биркгоф Дж. Д. Динамические системы / Дж. Д. Биркгоф; пер. с англ. Е. М. Ливенсона. — Ижевск : Издательский дом. «Удмуртский университет», 1999. — 408 с.
5. Большая Советская Энциклопедия: в 30 т. Т. 11. Италия – Кваркуш [Электронный ресурс]/ [гл. ред. А. М. Прохоров]. — Москва : Сов. Энциклопедия, 1973. — Режим доступа : <http://bse.scilib.com/article060128.html>.
6. Веб-сайт: визначення й застосування [Електронний ресурс] // Інтернет-студія Webtec. — 2009. — Режим доступа : <http://www.webtec.com.ua/ukr/articles.php?title=web-site>. — Назва з екрану.
7. Востров А. Почему тематические сайты доходные [Электронный ресурс] / А. Востров. — 2008. — Режим доступа : <http://www.seoded.ru/webmaster/sozдание-saita/tematicheskie-sajty.html>. — Назва з екрану.
8. Глибовец Н. Н. Становление технологии WEB 3.0 / Н. Н. Глибовец, Л. О. Шыпович // Математичне та комп'ютерне моделювання. Серія: Технічні науки: зб. наук. пр. — Кам'янець-Подільський : Кам'янець-Подільськ. нац. ун-т, 2010. — Вип. 3. — С. 52-63.
9. Гончаров М. Web Mining — добыча знаній из World Wide Web

- [Електронний ресурс] / М. Гончаров. — 2013. — Режим доступу : http://masters.donntu.edu.ua/2013/fknt/linkin/library/web_mining.htm. — Назва з екрану.
10. Дивак М.П. Методичний посібник з дисципліни "Системний аналіз" / М.П. Дивак. — Тернопіль: Тернопільська академія народного господарства, 2004. — 136 с.
 11. Етапи створення веб-сайтів [Електронний ресурс] / Г.В. Ломаковська, Г.О. Проценко, Й.Я. Ривкінд. — 2008. — Режим доступу : http://school.Xvatit.com/index.php?title=Етапи_створення_веб-сайтів. — Назва з екрану.
 12. Живицкая Е. Н. Системный анализ и проектирование информационных систем : учебно-методическое пособие / Е. Н. Живицкая, О.П.Едемская. — Минск : БГУИР, 2005. — 60 с.
 13. Закс Л. Статистическое оценивание / Л. Закс. — Москва : Статистика, 1976. — 598 с.
 14. Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем / А. Ивахненко. — К. : Наукова думка, 1981. — 296 с.
 15. Ивахненко А. Г. Помехоустойчивость моделирования / А. Г. Ивахненко, В. С. Степашко. — Киев : Наук. думка, 1985. — 216 с.
 16. История развития компьютерных сетей [Електронний ресурс] — 2005. — Режим доступу : <https://sites.google.com/site/komputernyesetibalabana/global/global-story>. — Назва з екрану.
 17. История связи и перспективы развития телекоммуникаций : уч. пос. / Ю. Украинцев, М. Цветов. — Ульяновск : УлГТУ, 2009. — 128 с.
 18. Кальченко Д. Оптимизация сайтов для поисковых систем / Д. Кальченко // КомпьютерПресс. — Москва, 2006. — Вып 2. — С.46-50.
 19. Корилов А. М. Теория систем и системный анализ // А. М. Корилов, С. Н. Павлов. — Томск : ТУСУР, 2007. — 343 с.
 20. Льюнг Л. Идентификация систем. Теория для пользователя / Л. Льюнг. — Москва : Наука, 1991. — 432 с.

21. МЕТА – Україна : українська пошукова система [Електронний ресурс] — Режим доступу : <http://meta.ua>. — Назва з екрану.
22. Митрофанова О. А. Онтології як системи хранения знань [Електронний ресурс] / О. Митрофанова, Н. Константинова. — 2008. — Режим доступу : http://www.ict.edu.ru/lib/index.php?id_res=5706.
23. Мудров А. Е. Численные методы / А. Е. Мудров. — Томск : Раско, 1991. — 272с.
24. Національний технічний університет України ХПІ. Кафедра інформатики та інтелектуальної власності [Електронний ресурс] — 2003. — Режим доступу : <http://khti-iiр.mipk.kharkiv.edu/library/index.html>. — Назва з екрану.
25. Орловский С. А. Проблемы принятия решений при нечеткой исходной информации / С. Орловский. — М. : “Радио и связь”, 1981. — 286 с.
26. Пасічник Н. Р. Дискретна динамічна модель для оцінювання та прогнозування відвідуваності Веб-сайту / Н. Р. Пасічник // Матеріали школи-семінару молодих вчених і студентів АСІТ-2013. — Тернопіль : ТНЕУ, 2013. — С.49-50.
27. Пасічник Н. Р. Інформаційна технологія формування аналітичних ресурсів сайту / Н. Р. Пасічник // Тематичний збірник наукових праць «Системи обробки інформації». — Харків : ХУПС, 2011. — Вип 3 (93). — С. 224-227.
28. Пасічник Н. Р. Математична модель динаміки відвідуваності тематичних Веб-сайтів та методи її ідентифікації / Н. Пасічник, М. Дивак, Р. Пасічник // Індуктивне моделювання складних систем. Збірник наукових праць. — Київ : Міжнар.наук.-навч.центр інформ.технологій та систем НАН та МОН України, 2013. — Вип. №5. — С. 236-246.
29. Пасічник Н. Р. Метод структурування тематичних Веб-сайтів на основі направленою інформаційного пошуку / Н. Р. Пасічник // Матеріали школи-семінару молодих вчених і студентів АСІТ-2012 : 4-5 травня 2012р. — Тернопіль, 2012. — С. 47-49.

30. Пасічник Н. Р. Метод та алгоритм побудови структур тематичних Веб-сайтів на основі онтологічного підходу / Н. Р. Пасічник, М.П. Дивак // Наукові праці ДонНТУ, серія «Інформатика, кібернетика та обчислювальна техніка», 2012. — Вип 15 (203). — С. 184-189.
31. Пасічник Н. Р. Метод формування онтологічного наповнення на основі аналізу зашумленої слабо структурованої інформації спеціалізованих Веб-сайтів / Н. Р. Пасічник, М.П. Дивак // Індуктивне моделювання складних систем. Збірник наукових праць. — Київ : Міжнар.наук.-навч.центр інформ.технологій та систем НАН та МОН України, 2012. — Вип. №4. — С.158-168.
32. Пасічник Н. Р. Метод формування онтологічного контенту на основі аналізу інформації спеціалізованих Веб-сайтів / Н. Р. Пасічник// Вісник Хмельницького національного університету : Технічні науки, 2012. — Вип. №5. — С. 241-244.
33. Пасічник Н. Р. Основи автоматизації формування аналітичних ресурсів сайту / Н. Р. Пасічник // Тези доповідей II МНПК «Інформаційні технології та комп'ютерна інженерія» : 26-27 травня 2011р. — Харків, 2011. — С. 126-127.
34. Пасічник Н. Р. Система підтримки інформаційного наповнення сайту / Н. Р. Пасічник // Матеріали школи-семінару молодих вчених і студентів АСІТ-2011 : 20-21 травня 2011р. — Тернопіль, 2011. — С. 130.
35. Пасічник Н. Р. Формалізм в постановці задачі створення якісного сайту / Н. Р. Пасічник, М.П. Дивак // Наукові праці ДонНТУ, серія «Інформатика, кібернетика та обчислювальна техніка», — Донецьк, 2011. — Вип 14 (188). — С. 325-329.
36. Пасічник Р. М. Ідентифікація системи Моно-Ієрусалимського з керованим зворотнім зв'язком / Р. М. Пасічник, Ю. Р. Піговський // Комп'ютинг. — Тернопіль, 2008. — Т. 7, Вип. 1. — С. 146–152.
37. Пелецишин А. М. Позиціонування сайтів у глобальному інформаційному середовищі / А. М. Пелецишин. — Львів: Видавництво

- НУ “Львівська політехніка”, 2007. — 260с.
38. Петров Е. В. Видимый результат, или система сбалансированных показателей для службы персонала: [Электронный ресурс] / Е. В. Петров, А. А. Югов, О. В. Гурина. — Режим доступа : http://www.iteam.ru/publications/human/section_45/article_2708. — Назва з екрану.
 39. Поиск в Украине — bigmir.net [Электронный ресурс] — Режим доступа : <http://search.bigmir.net>. — Назва з екрану.
 40. Поисковая система Google – история, цифры, факты [Электронный ресурс] — Режим доступа : <http://seoprofy.ua/blog/poiskovye-sistemy/google>. — Назва з екрану.
 41. Развитие веб-технологий: тенденции и перспективы [Электронный ресурс] — 2013. — Режим доступа : <http://www.4stud.info/web-programming/lecture9.html>. — Назва з екрану.
 42. Разработка и продвижение сайта [Электронный ресурс] — 2003. — Режим доступа : http://www.web-sitio.ru/sozдание-i-prodvizhenie-sajta/razrabotka_i_prodvizhenie_sajta.html. — Назва з екрану.
 43. Растрингин Л. А. Введение в идентификацию объектов управления / Л. А. Растрингин, Н. Е. Маджаров. — М.: Энергия, 1987. — 216 с.
 44. Ромашкина Г. Ф. Коэффициент конкордации в анализе социологических систем / Г. Ф. Ромашкина, Г. Г. Татарова // Социология. 2005. — № 20. — С. 131-158.
 45. Створення веб сайту [Электронный ресурс] — 2008. — Режим доступа : <http://www.victoria.lviv.ua/html/km/6.html>. — Назва з екрану.
 46. Трифонов А. Г. Оптимизация методом наименьших квадратов. Optimization Toolbox 2.2 : Руководство пользователя [Электронный ресурс] / А. Г. Трифонов. — Режим доступа : http://matlab.exponenta.ru/optimiz/book_1/13.php. — Назва з екрану.
 47. Фильтр Гугл Пингвин, Панда и фильтры Яндекса [Электронный ресурс] — 2012. — Режим доступа : <http://www.raskruty.ru/poiskoviki/google->

- penguin-filtr-yandex.html. — Назва з екрану.
48. Цыпкин Я. З. Информационная теория идентификации / Я. З. Цыпкин — М.: Наука, 1995. — 336 с.
 49. Что такое Гугл Пингвин и Панда [Электронный ресурс] — 2012. — Режим доступа : <http://digitalforce.ua/chto-takoe-filtr-gugl-pingvin/>. — Назва з екрану.
 50. Abe N. Learning word association norms using tree cut pair models [Электронный ресурс] / N. Abe, H. Li // Proceedings. of 13th Int. Conf. on Machine Learning, 1996. — Режим доступа : arxiv.org/pdf/cmp-lg/9605029. — Назва з екрану.
 51. Adriaans P. Data Mining / P. Adriaans, D. Zantinge. — New York : Addison-Wesley Publ. Co., 1996.
 52. Agirre E. Enriching very large ontologies using the WWW [Электронный ресурс] / E. Agirre, O. Ansa, E. Hovy, D. Martinez // Proceedings of the Workshop on Ontology Construction of the European Conference of AI (ECAI-00). — Berlin, 2000. — Режим доступа : www.isi.edu/natural-language/people/hovy/.../00ECAI-ontol-enrich.pdf. — Назва з екрану.
 53. Agrawal R. Mining association rules between sets of items in large databases / R. Agrawal, T. Imielinski, A. Swami // Proc. of the ACM SIGMOD Conference on Management of Data. — Washington D. C., 1993. — PP. 207-216.
 54. Ahonen–Мука Н. Finding co-occurring text phrases by combining sequence and frequent set discovery [Электронный ресурс] / Н. Ahonen–Мука, О. Heinonen , М. Klemettinen // Proc. of the Workshop on Text Mining: Foundations, Techniques and Applications, IJCAI-99. — Stockholm, 1999. — Режим доступа : www.cs.helsinki.fi/u/hahonen/ham_ijcai99.ps.
 55. Alfonseca E. An unsupervised method for general named entity recognition and automated concept discovery [Электронный ресурс] / E. Alfonseca, S. Manandhar // Proceedings of the 1st International Conference on General WordNet. — Mysore, India, 2002. — Режим доступа :

- <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.1501>.
56. Alfonseca E. Improving an Ontology Refinement Method with Hyponymy Patterns / E. Alfonseca, S. Manandhar // Language Resources and Evaluation. — Las Palmas, Spain, 2002. — PP. 235–239.
 57. Aussenac-Gilles N. Corpus Analysis For Conceptual Modelling [Электронный ресурс] / N. Aussenac-Gilles, B. Biébow, S. Szulman // 12th International Conference EKAW' 2000. — Juan-les-pins, France, 2000. — Режим доступа : ceur-ws.org/Vol-51/AussenacBiebowSzulman.pdf.
 58. Aussenac-Gilles N. Revisiting Ontology Design: A Methodology Based on Corpus Analysis / N. Aussenac-Gilles, B. Biébow, S. Szulman // Lecture Notes in Artificial Intelligence. — Vol. 1937. — Berlin: Springer-Verlag, 2000. — PP. 172–188. — Назва з екрану.
 59. Behm A. On the Migration of Relational Schemas and Data to Object-Oriented Database Systems. / A. Behm , A. Geppert, K. Dittrich // Proceeding of the 5th Int. Conference on Re-Technologies for Information Systems. — Klagenfurt, 1997. — PP. 13-33.
 60. Berners-Lee T. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor / T. Berners-Lee. — New York : HarperCollins Publishers, 1999. — 256 p.
 61. Biébow B. TERMINAE: a linguistic-based tool for the building of a domain ontology. / B. Biébow B, S. Szulman // Proceedings of the 11th European Workshop EKAW'99. — Germany, Berlin, 1999. — PP. 49-66.
 62. Bigmir)net. Рейтинг сайтов bigmir)net. — Режим доступа : <http://top.Bigmir.net>. — Назва з екрану.
 63. Bisson G. Conceptual Clustering / G. Bisson // Proc. Mlnet Summer School on Machine Learning and Knowledge Acquisition. — Dourdan, France, 1994. — PP. 5-10.
 64. Bisson G. Learning in FOL with a similarity measure / G. Bisson // Proc. 10th National Conference of AI. — San José, CA, 1992. — PP. 12-16.
 65. Borges J. Data Mining of User Navigation Patterns [Электронный ресурс] /

- J. Borges, M. Levene // *Web Usage Analysis and User Profiling: LNCS.* — Berlin : Springer, 2000. — Vol. 1836. — PP. 92-112. — Режим доступа : http://link.springer.com/chapter/10.1007%2F3-540-44934-5_6.
66. Borges J. A Comparison of Scoring Metrics for Predicting the Next Navigation Step With Markov Model-Based Systems / J. Borges, M. Levene // *International Journal of Information Technology & Decision Making.* — 2010. — Vol. 9, No. 4. — PP. 547-573.
67. Buechner A. G. User-Driven Navigation Pattern Discovery from Internet Data [Электронный ресурс] / A. G. Büchner, S. S. Anand, M. D. Mulvenna, J. G. Hughes. // *Web Usage Analysis and User Profiling: Lecture Notes in Computer Science.* — Springer, 2000. — Vol. 1836. — PP. 74-91. Режим доступа : http://link.springer.com/chapter/10.1007/3-540-44934-5_5.
68. Burke R. Hybrid Web Recommender Systems / R. Burke // *The Adaptive Web: Methods and Strategies of Web Personalization.* — Berlin, Heidelberg : Springer-Verlag, 2007. — PP. 377-408.
69. Chakrabarti S. Data mining for hypertext / S. Chakrabarti // *ACM SIGKDD Explorations Newsletter.* — New York: ACM. — Vol. 1, No.2. — PP. 1 – 11.
70. Chen L. Using Web Structure and Summarization Techniques for Web Content Mining / L. Chen, W.L. Chue // *Journal Information Processing and Management.* — 2005 — Vol. 41, Issue 5. — PP. 1225-1242.
71. Compton P. A philosophical basis for knowledge acquisition / P. Compton, A. Jansen // *Knowledge acquisition Journal.* — London, UK : Academic Press Ltd., 1990. — Vol. 2, Issue 3. — PP. 241-257.
72. Cumbrowski C. 50 Questions to Evaluate the Quality of Your Website [Электронный ресурс] / C. Cumbrowski // *Search Engine Journal.* — 2008. — Режим доступа : <http://www.Searchenginejournal.com/50-questions-to-evaluate-the-quality-of-your-website/6400/>. — Назва з екрану.
73. DeFazio G. Types of websites: Basic website design types their purpose, scope and structure [Электронный ресурс] / G. DeFazio // *Rocketface Workshop – Webmaster’s Tutorial.* — 2010. — Режим доступа :

- http://www.rocketface.com/analyze_website/types_of_websites.html.
74. Deitel A. Learning ontologies from RDF annotations [Электронный ресурс] / A. Deitel, C. Faron, R. Dieng // Proceedings of the IJCAI Workshop in Ontology Learning. — Seattle, 2001. — Режим доступа : osm.cs.byu.edu/CS652s04/DFD01Learning.pdf. — Назва з екрану.
 75. Deshpande M. Selective Markov Models for Predicting Web-Page Accesses / M. Deshpande, G. Karypis // ACM Transactions on Internet Technology. — New York, USA, 2004. — Vol. 4(2). — PP. 163–184.
 76. Doan A. Learning Source Descriptions for Data Integration [Электронный ресурс] / A. Doan, P. Domingos, A. Levy // Proceedings of the 3d Intl. Workshop on the Web and Databases. — 2000. — PP. 81-86. — Режим доступа : <http://citeseerx.ist.psu.edu/viewdoc/similar?doi=10.1.1.42.9378>.
 77. Evaluating Web Sites Criteria and Tools [Электронный ресурс]. — Режим доступа : <http://olinuris.library/cornell/edu/ref/research/webeval/html>. — Назва з екрану.
 78. Faure D. First experiments of using semantic knowledge learned by ASIUM for information extraction task using INTEX [Электронный ресурс] / D. Faure, T. Poibeau // Proceedings of the Workshop on Ontology Learning, ECAI 2000. — Режим доступа : http://ceur-ws.org/Vol-31/DFaure_6.pdf
 79. Ghani R., Building Recommender Systems Using a Knowledge Base of Product Semantics [Электронный ресурс] / R. Ghani, A. Fano // Proceedings of the Workshop on Recommendation and Personalization in E-Commerce, 2nd Int'l Conf. on Adaptive Hypermedia and Adaptive Web Based Systems. — Malaga, Spain, 2002. — Режим доступа : <http://rayidghani.com/publications/recommenderws02.pdf>. — Назва з екрану.
 80. Gómez-Pérez A. OntoWeb: A survey of ontology learning methods and techniques [Электронный ресурс] / A. Gómez-Pérez, D. Manzano-Macho // Universidad Politécnica de Madrid. — Режим доступа : www.bibsonomy.org/bibtex/.../hotho. — Назва з екрану.
 81. Gorbunov A. L. Markov models for website traffic / A. L. Gorbunov // IMAT

- 2007. — Ural University Publ., 2007. — PP. 65-73.
82. Hahn U. Joint knowledge capture for grammars and ontologies / U. Hahn, K. Markó // Proceedings of the First International Conference on Knowledge Capture K-CAP. — New York, NY, USA, 2001. — PP. 68-75.
83. Hahn U. Towards Very Large Terminological Knowledge Bases: A Case Study from Medicine / U. Hahn, S. Schulz // Proc. of 13th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, Montréal. — Berlin : Springer, 2000. — PP. 176-186.
84. Harabagiu S. M. Enriching the WordNet Taxonomy with Contextual Knowledge acquired from text / S. M. Harabagiu, D. I. Moldovan // Natural Language Processing and Knowledge Representation. — Cambridge: AAAI/MIT Press, 2000. — PP. 301-334.
85. Hearst M. A. Automated Discovery of WordNet Relations / M. A. Hearst // WordNet: An Electronic Lexical Database. — Cambridge : MIT Press, 1998. — PP. 132-152.
86. Hearst M.A. Automatic acquisition of Hyponyms from large text corpora [Электронный ресурс] / M. A. Hearst // Proceedings of the Fourteenth International Conference on Computational Linguistic. — France, Nantes, 1992. — Режим доступа : <http://people.ischool.berkeley.edu/~hearst/papers/coling92.pdf>. — Назва з екрану.
87. Hovy E. H. Automated Text Summarization in SUMMARIST / E. H. Hovy, C.-Y. Lin // Advances in Automatic Text Summarization. — Cambridge : MIT Press, 1999. — PP 1-14.
88. Hwang C. H. Incompletely and imprecisely speaking: using dynamic ontologies for representing and retrieving information. / C. H. Hwang // Proceedings of the 6th Intl. Workshop on Knowledge Representation meets Databases. — Sweden, Linköping: CEUR-WS, 1999. — PP 14-20.
89. Joachims T. A. Probabilistic analysis of the Rocchio Algorithm with TFIDF for text categorization. Logic J. of the IGPL [Электронный ресурс]. — 1998. — Режим доступа : www.cs.cornell.edu/people/tj/publications/

- joachims_97a.pdf. — Назва з екрану.
90. Johannesson P. A. Method for Transforming Relational Schemas into Conceptual Schemas / P. A. Johannesson // 10th Intl. Conference on Data Engineering. — Houston: IEEE Press, 1994. — PP. 115-122.
 91. Jon M. Kleinberg. Authoritative sources in hyperlink environment [Електронний ресурс]. — 1999. — Режим доступу : www.cs.cornell.edu/home/kleinber/auth.pdf. — Назва з екрану.
 92. Joshi A. Improving Web Search Engine Results Using Clustering [Електронний ресурс]. — Режим доступу : ebiquity.umbc.edu/_file_directory_/papers/44.pdf. — Назва з екрану.
 93. Kashyap V. Design and Creation of Ontologies for Environmental Information Retrieval [Електронний ресурс] / V. Kashyap // Proc. 12th Workshop on Knowledge Acquisition, Modelling and Management, Banff, Canada, October, 1999. — Режим доступу : citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.8959. — Назва з екрану.
 94. Khalil F. Combining Web Data Mining Techniques for Web Page Access Prediction : PhD Thesis : 080109 – Pattern Recognition and Data Mining, 080704 – Information Retrieval and Web Search / F. Khalil. — Australia, Quinsland, 2008. — 197 p.
 95. Khan L., Luo F. Ontology Construction for Information Selection / L. Khan , F. Luo // Proc. of 14th IEEE International Conference on Tools with Artificial Intelligence, Washington DC, November 2002. — Los Alamitos, CA : IEEE Computer Society, 2002. — pp. 122-127.
 96. Khribi, M.K. Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval / M. Khribi, M Jemni, O. Nasraoui // Proc. Of 8th IEEE International Conference on Advanced Learning Technologies ICALT '08, Santander, Spain, 2008. — USA, Los Calamitos : IEEE Computer Society, 2008. — pp. 241-245.
 97. Kietz J.U. A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet [Електронний ресурс] / J. Kietz, A. Maedche, R. Volz //

- Proc. Of EKAW'00 Workshop "Ontologies and Texts", Juan-Les-Pins, France. — 2000. — Режим доступа : sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-51/Maedche.pdf. — Назва з екрану.
98. Kirk E. E. Evaluating Information Found on the Internet [Електронний ресурс] / E. E. Kirk // The Sheridan Libraries, Johns Hopkins University. — 2010. — Режим доступа : <http://guides.library.jhu.edu/evaluatinginformation-19>. — Назва з екрану.
99. Lee C. Automatic WordNet Mapping Using Word Sense Disambiguation [Електронний ресурс] / C. Lee, G. Lee, S. J. Yun // Proc. of 38th Annual Meeting of the Association for Computational Linguistics (ACL'2000). — 2000. — Режим доступа : acl.ldc.upenn.edu/W/W00/W00-1318.pdf.
100. Lin C.-Y. The Automated Acquisition of Topic Signatures for Text Summarization [Електронний ресурс] / C.-Y. Lin, E. H. Hovy // Proc. Of the COLING Conference, Strasbourg, France, August 2000. — Режим доступа : research.microsoft.com/en-us/people/cyl/col2000-final.pdf
101. Lonsdale D. Peppering Knowledge Sources with SALT: Boosting Conceptual Content for Ontology Generation / D. Lonsdale, Y. Ding, D. W. Embley, A. Melby // Papers of the AAAI Workshop: Semantic Web Meets Language Resources, Edmonton, Alberta, Canada, July 2002. — USA, Stanford: AAAI Press, 2002. — PP. 30-36.
102. Maedche A. Discovering Conceptual Relations from Text / A. Maedche, S. Staab // ECAI 2000: Proceedings of the 14th European Conference on Artificial Intelligence, Berlin, August 21-25, 2000. — Amsterdam: IOS Press, 2000. — PP. 321-325.
103. Maedche A. Ontology Learning for the Semantic Web [Електронний ресурс] / A. Maedche, S. Staab // IEEE Intelligent Systems: Special Issue on the Semantic Web. — Vol. 16 (2). — 2001. — Режим доступа : ceur-ws.org/Vol40/maedche%2Bstaab.pdf. — Назва з екрану.
104. Martinez M. How to Identify High Quality Websites [Електронний ресурс] / M. Martinez. — 2011. — Режим доступа : <http://>

- www.searchenginepeople.com/blog/identifying-quality-websites.html. — Назва з екрану.
105. Masand B. Web Usage Analysis and User Profiling [Електронний ресурс] / B. Masand, M. Spiliopoulou. — 1999. — Режим доступу : <http://www.springer.com/computer/security+and+cryptology/book/978-3-540-67818-2>. — Назва з екрану.
106. Missikoff M. The Usable Ontology: An Environment for Building and Assessing a Domain Ontology [Електронний ресурс] / M. Missikoff, R. Navigli, P. Velardi // Research paper at International Semantic Web Conference (ISWC) 2002, June 9-12th, 2002 Sardinia, Italia. — 2002. — Режим доступу : link.springer.com/chapter/10.1007%2F3-540-48005-6_6. — Назва з екрану.
107. Mobasher B. Data Mining for Web Personalization // The Adaptive Web : Methods and Strategies of Web Personalization (Lecture Notes in Computer Science / Information Systems and Applications, incl. Internet/Web, and HCI). — Springer-Verlag Berlin, Heidelberg, 2007. — Book 4321. — PP. 90-135.
108. Modica G. The Use of Machine-Generated Ontologies in Dynamic Information Seeking / Modica G., Gal A., Jamil H. M. // Proceedings of the Sixth International Conference on Cooperative Information Systems (CoopIS 2001), September 5-7, 2001, Trento, Italy. — Berlin, Heidelberg: Springer-Verlag, 2001. — PP. 433-447.
109. Mohomed I. Community-Driven Adaptation: Automatic Content Adaptation in Pervasive Environments / I. Mohomed, A. Chin, J. Cai, E. Lara // Proceedings of the Workshop on Mobile Computing Systems and Applications WMCSA'04 2-3 December, 2004, English Lake District, UK. — Los Alamitos, CA : IEEE Computer Society, 2004. — PP. 124-133.
110. Moldovan D. I. Domain-Specific Knowledge Acquisition from Text / D. Moldovan, R. Girju, V. Rus // Proceedings of the Applied Natural Language Processing Conference, Seattle, WA., April-May, 2000. — Stroudsburg, PA,

- USA : Association for Computational Linguistics, 2000. — PP. 268-275.
111. Morik K. Balanced Cooperative Modelling. / K. Morik // Machine Learning. — Boston, USA : Kluwer Academic Publishers, 1993. — Vol. 11, Issue 2-3. — PP. 217-235.
 112. Murata M. Taxonomy of XML Schema Languages using Formal Language Theory [Электронный ресурс] / M. Murata, D. Lee, M. Mani // Extreme Markup Languages, Montreal, Canada, Aug. 2001. — Режим доступа : web.cs.wpi.edu/~mmani/toit/taxonomy/new/taxonomy.pdf. — Назва з екрану.
 113. Navadiya D. Web Content Mining Techniques — A Comprehensive Survey [Электронный ресурс] / D. Navadiya, R. Patel // — 2012. — Режим доступа : <http://www.ijert.org/view.php?id=1950&title=web-content-mining-techniques-a-comprehensive-survey>. — Назва з екрану.
 114. Navigli R. Ontology Learning and its application to automated terminology translation / R. Navigli, P. Velardi, A. Gangemi // IEEE Intelligent Systems. — 2003. — Vol. 18 (1). — PP. 22-31.
 115. Navigli R. Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites / R. Navigli, P. Velardi. // Computational Linguistics. — Cambridge: MIT Press, 2004. — Vol. 30(2). — PP. 151-179.
 116. Nobécourt J. A Method to Build Formal Ontologies from Text [Электронный ресурс] / J. Nobécourt // Workshop on Ontologies and Text, Knowledge Engineering and Knowledge Management: Methods, Models and Tools : 12th International Conference EKAW'2000. — Juan-Les-Pins, France, 2000. — Режим доступа : ceur-ws.org/Vol-51/Nobecourt.pdf. — Назва з екрану.
 117. Papatheodorou C. Discovery of Ontologies for Learning Resources Using Word-based Clustering / C. Papatheodorou, A. Vassiliou, B. Simon // Proc. ED-MEDIA 2002: World Conference on Educational Multimedia, Hypermedia and Telecommunications, August 2002, Denver, Colorado, USA. — Chesapeake, VA: Association for the Advancement of Computing in Education (AACE), 2002. — PP. 1523-1528.

118. 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, Ukraine, 2010. — PP. 207.
119. 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. – PP. 391.
120. 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. — PP. 277.
121. 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, Ukraine, 2013. — PP.298-301.
122. Pasichnyk N. Mathematical modeling of Website quality characteristics in dynamics / N. Pasichnyk, M. Dyvak, R.Pasichnyk // Journal of Applied Computer Science / Technical University Press – Lodz, Poland, 2014. – Vol. 22 №1. – PP. 171-183.
123. Pasichnyk N. Parametric identification methods of the model of Websites' attendance / N. Pasichnyk, A. Melnyk, L.Honchar // Proceedings of the International Conference on Modern Problems of Radio Engineering Telecommunications and Computer Science (TCSET), 25 Feb. – 1 March 2014. – Lviv-Slavsko, 2014. – PP. 93-94.
124. 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, Prague, Czech Republic, 15–17 September 2011. — 2011. — Vol. 2. — PP. 692–695.
125. Perkowitz M. Adaptive Web Sites: Automatically synthesizing Web pages [Электронный ресурс] / M. Perkowitz, O. Etzioni // Proceedings of the Fifteen National Conference in Artificial Intelligence (AAAI 98). — 1998. — Режим доступа : homes.cs.washington.edu/~etzioni/papers/aaai98.pdf.
 126. Quality Criteria for Website Excellence [Электронный ресурс] // World Best Website Awards — 2004. — Режим доступа : <http://www.worldbestwebsites.com/criteria.htm>. — Назва з екрану.
 127. Ratkiewicz J. Characterizing and Modeling the Dynamics of Online Popularity [Электронный ресурс] / J. Ratkiewicz, S. Fortunato, A. Flammini // Physical Review Letters, 2010. — Режим доступа : <http://arxiv.org/pdf/1005.2704.pdf>. — Назва з екрану.
 128. Rigau G. Building Accurate Semantic Taxonomies from Monolingual MRDs [Электронный ресурс] / G. Rigau, H. Rodríguez, E. Agirre // Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics, Montreal, Canada. — 1998. — Режим доступа до видання : <http://www.lsi.upc.es/~rigau/>.
 129. Roux C. An Ontology Enrichment Method for a Pragmatic Information Extraction System Gathering Data on Genetic Interactions [Электронный ресурс] / C. Roux, D. Proux, F. Rechemann, L. Julliard // Position paper in Proceedings of the ECAI2000 Workshop on Ontology Learning, Berlin, August 2000. — Режим доступа : http://ceur-ws.org/Vol-31/CRoux_3.pdf.
 130. Rubin D.L. Automatic data acquisition into ontologies from pharmacogenetics relational data sources using declarative object definitions and XML [Электронный ресурс] / D. Rubin, M. Hewett, D. Oliver, T. Klein // Proceedings of the Pacific Symposium on Biology , Lihue, HI, 2002. — Режим доступа : psb.stanford.edu/psb-online/proceedings/psb02/rubin.pdf.
 131. Sharma M. Types of web sites — a categorization based on content

- [Электронный ресурс] / М. Sharma // WDN online tutorials, tips, tricks, resources, articles and more. — 2010. — Режим доступа : http://www.webdevelopersnotes.com/basics/types_of_websites.php. — Назва з екрану.
132. Singhal A. More guidance on building high-quality sites [Электронный ресурс] / A. Singhal. — 2011. — Режим доступа : <http://googlewebmastercentral.blogspot.com/2011/05/more-guidance-on-building-high-quality.html>. — Назва з екрану.
133. Sowa J.F. Conceptual Structures: Information Processing in Mind and Machine / J. F. Sowa. — Reading, MA. : Addison-Wesley, 1984. — 481 p.
134. Srivastava J. Web Mining — Concepts, Applications & Research Directions [Электронный ресурс] / J. Srivastava, P. Desikan, V. Kumar. — Режим доступа : www.users.cs.umn.edu/~desikan/publications/wmo.pdf. — Назва з екрану.
135. Stojanovic L. Migrating Data-Intensive Web Sites into the Semantic Web. / L. Stojanovic, N. Stojanovic, R. Volz // Proceedings of the 17th ACM symposium on applied computing, Leganes, Spain. — ACM Press, 2002. — PP. 1100-1107.
136. Superfund Data: The Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS25). — 1999. — Режим доступа : <http://www.epa.gov/edr>. — Назва з екрану.
137. Suryanto H. Discovery of Ontologies from Knowledge Bases / H. Suryanto, P. Compton // Proceedings of the First International Conference on Knowledge Capture, Victoria, Canada, 21-23 Oct. 2001. — New York, USA : The Association for Computing Machinery. — PP. 171-178.
138. Suryanto H. Learning Classification Taxonomies From a Classification Knowledge Based System / H. Suryanto, P. Compton // Proceeding of the First Workshop on Ontology Learning in conjunction with ECAI- 2000, Berlin, Germany, 22 Aug. 2000. — Berlin, 2000. — PP. 1-6.
139. TopPing. Рейтинг украинских сайтов. [Электронный ресурс] — Режим доступа: <http://www.Topping.com.ua>. — Назва з екрану.

140. Volz, R. OntoLiFT Prototype [Электронный ресурс] / R. Volz, D. Oberle, S. Staab, R. Studer // IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web. — 2003. — Режим доступа: wonderweb.man.ac.uk/deliverables/documents/D11.pdf . — Назва з екрану.
141. Wagner A. Enriching a Lexical Semantic Net with Selectional Preferences by Means of Statistical Corpus Analysis / A. Wagner // Proceedings of the ECAI-2000 Workshop on Ontology Learning, Berlin, August 2000. — Berlin, Germany, 2000. — PP. 37-42.
142. Web Content Mining and Structured Data Extraction and Integration: An Implement of Vertical Search Engine System [Электронный ресурс] / J. Gou — 2012. — Режим доступа: <https://wiki.engr.illinois.edu/download/attachments/200017637/.ResearchReport.pdf?version=1&modificationDate=1336540363000>
143. Xu F. A Domain Adaptive Approach to Automatic Acquisition of Domain Relevant Terms and their Relations with Bootstrapping [Электронный ресурс] / F. Xu, D. Kurz, J. Piskorski, S. Schmeier // Proceedings of LREC, the third international conference on language resources and evaluation. — Las Palmas, Canary island, Spain, 2002. — Режим доступа : www.piskorski.waw.pl/papers/LREC_TermExtraction_final.pdf.

**Додаток А. Лістинг модуля головного контролера підсистем
структурування контенту.**

```

<?php
header('Content-Type: text/html; charset=utf-8');
ini_set("max_execution_time", 60);
define("FILTER", "0.8");
error_reporting(E_ALL);

if (isset($_REQUEST['action'])){
    $action = $_REQUEST['action'];
    if (isset($_REQUEST['q'])){
        $q = $_REQUEST['q'];
    }
    if (isset($_REQUEST['num'])){
        $num = intval($_REQUEST['num']);
    }
    if (isset($_REQUEST['view'])){
        $view = $_REQUEST['view'];
    }else{
        $view = 'json';
    }
    if (isset($_REQUEST['ignore'])){
        $ignore = $_REQUEST['ignore'];
    }
    $worker = new Worker();
    $worker->setKeyword($q);
    switch ($action){
        case 'get_search_result':
            echo $worker->getSearchResult();
            break;
        case 'process_page':
            echo $worker->processPage($q, $num);
            break;
        case 'get_final':
            if ($view == 'normal'){
                $out = json_decode(calculateFinal($q),true);
                foreach ($out as $row){
                    var_dump($row['concept']);
                    echo " " . $row['fr'] . "<br>";
                }
            }
            else{
                echo $worker->calculateFinal($q);
            }
        }
    }
}

```

```
    }
    break;
case 'set_ignore':
    echo $worker->setIgnored($q,$ignore);
    break;
case 'unset_ignore':
    echo $worker->unSetIgnored($q,$ignore);
    break;
case 'get_ignored':
    echo $worker->getIgnored($q);
    break;
case 'clear_ignored':
    echo $worker->clearIgnored($q);
    break;
case 'get_link':
    echo $worker->getLink($q,$num);
    break;
case 'get-search-string':
    echo $worker->getSearchString($q);
    break;
}
}
?>
```

**Додаток Б. Лістинг виконавчого модуля підсистем
структурування контенту.**

```

<?php
class Worker{
    public $db;
    private $config;
    private $keyword;
    private $searchResult;
    private $curlHandler;
    public function __construct(){
        $this->config = include 'config.php';
        include 'model.php';
        $this->db = new Model($this->config['db']);
    }

    public function setKeyword($keyword){
        $this->keyword = $keyword;
    }

    public function filterResult(){
        $result = array();
        foreach ($this->searchResult as $item){
            if (strpos(mb_strtolower($item['title'],'utf8'), $this->keyword) !==
FALSE){
                $result[] = array(
                    'link' => $item['link'],
                    'displayLink' => $item['displayLink']
                );
            }
        }
        return $result;
    }

    private function getPage($url,$ch){
        $header[] = "Cache-Control: max-age=0";
        $header[] = "Connection: keep-alive";
        $header[] = "Keep-Alive: 300";
        $header[] = "Accept-Charset: utf-8";
        $header[] = "Accept-Language: en-us,en;q=0.5";
        $header[] = "Pragma: "; // browsers keep this blank.
        curl_setopt($ch, CURLOPT_URL, $url);
        curl_setopt($ch, CURLOPT_HTTPHEADER, $header);
        curl_setopt($ch, CURLOPT_SSL_VERIFYPEER, false);
    }
}

```

```

    curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
    curl_setopt($ch, CURLOPT_USERAGENT, "Mozilla/5.0 (Windows NT
5.1; rv:23.0) Gecko/20100101 Firefox/23.0");
    $result = curl_exec($ch);
    return $result;
}

public function getSearchResult(){
    $count = 100;
    $searchResult = array();
    $result = $this->db->getSQuery($this->keyword,false);
    if (count($result) == 0){
        $ch = curl_init();
        $pages = ceil($count/10);
        if ($pages == 1){
            $num = $count;
        }else{
            $num = 10;
        }
        for ($i = 0;$i < $pages;$i++){
            $start = $i*10+1;
            $url =
            "https://www.googleapis.com/customsearch/v1?key=AIzaSyCNBXBSNaiUhL8A0I
j1sqpcQOoK0Ah_hww&cx=010973468806829165452:cl7skk0cg6a&q="
            .
            urlencode($this->keyword) . "&start=$start&num=$num";
            $searchResultPaged = $this->getPage($url,$ch);
            $decodedResult = json_decode($searchResultPaged,true);

            if (is_array($decodedResult) && isset($decodedResult['items'])){
                $items = $decodedResult['items'];
                $filteredResult = $items;
                $searchResult = array_merge((array)$searchResult,
                $filteredResult);
            }
        }
        $id = $this->db->setSQuery($this->keyword,serialize($searchResult));
        curl_close($ch);
    }else{
        $searchResult = unserialize($result[0]['search_result']);
        $id = $result[0]['id'];
    }

    return json_encode(
        array(

```

```

        'id' => $id,
        'query' => $this->keyword,
        'count' => count($searchResult),
    )
    );
}

public function processPage($id,$num){
    $searchResult = $this->db->getSQuery($id,true);
    $searchResult = unserialize($searchResult[0]['search_result']);
    if (isset($searchResult[$num])){
        $currentPage = $searchResult[$num];
        $$Result = $this->db->getSResult($id,$num);
        if (count($$Result)== 0){
            $ch = curl_init();
            $content = $this->getPage($currentPage['link'], $ch);
            $decodeFrom = FALSE;
            $content = mb_strtolower($content,'utf8');
            $parsed_content = = $this-
>parseContent($content,$currentPage['link'],$ch);
            curl_close($ch);
            $this->db-
>setSResult($id,$currentPage['displayLink'],$content,$parsed_content,$currentPage
['link'],$num);
        }
    }
    $result = array(
        'url' => $currentPage['displayLink'],
        'pageid'=> $num
    );
    return json_encode($result);
}

private function parseContent($content,$link,$ch){
    $result = array();
    $dom = new DOMDocument;
    @$dom->loadHTML($content);
    $frames = $dom->getElementsByTagName('frame');
    if ($frames){
        foreach ($frames as $frame) {
            $frameUrl = $link . '/' . $frame->getAttribute('src');
            $content.= getPage($frameUrl, $ch);
        }
        unset($dom);
    }
}

```



```

    $dom = new DOMDocument;
    @$dom->loadHTML($content);
}
$links = $dom->getElementsByTagName('a');
foreach ($links as $link) {
    $text = str_replace("\", ", $link->textContent);
    $text = trim($link->textContent);
    $text = preg_replace('/\d/', "", $text);
    if ($text != ""){
        $result[] = array(
            'text' => $text,
            'link' => $link->getAttribute('href'). '<br>'
        );
    }
}
return $result;
}

public function getIgnored($id){
    $ignoredConcepts = $this->db->getSQuery($id, true);
    $ignoredConcepts = unserialize($ignoredConcepts[0]['ignored']);
    $output = array();
    if (is_array($ignoredConcepts)){
        foreach ($ignoredConcepts as $key=>$value){
            $output[] = array('concept' => $value);
        }
    }
    return json_encode($output);
}

public function setIgnored($id,$concept){
    $ignoredConcepts = $this->db->getSQuery($id, true);
    $ignoredConcepts = unserialize($ignoredConcepts[0]['ignored']);
    if (!$ignoredConcepts){
        $ignoredConcepts = array();
    }
    if (is_array($ignoredConcepts) &&
!in_array($concept,$ignoredConcepts)){
        $ignoredConcepts = array_merge(array($concept), $ignoredConcepts);
        $this->db->setIgnored($id, serialize($ignoredConcepts));
        return 'yes';
    }
    return 'no';
}

```

```

public function unSetIgnored($id,$concept){
    $signoredConcepts = $this->db->getSQuery($id, true);
    $signoredConcepts = unserialize($signoredConcepts[0]['ignored']);
    if (is_array($signoredConcepts)){
        foreach ($signoredConcepts as $key=>$value){
            if ($value == $concept){
                unset($signoredConcepts[$key]);
                $this->db->setIgnored($id, serialize($signoredConcepts));
                return 'yes';
            }
        }
    }
    return 'no';
}

```

```

public function clearIgnored($id){
    $this->db->setIgnored($id, serialize(array()));
    return 'ok';
}

```

```

public function calculateFinal($id){
    $$SQuery = $this->db->getParsedResult($id);
    $signoredConcepts = $this->db->getSQuery($id, true);
    $signoredConcepts = unserialize($signoredConcepts[0]['ignored']);
    if (!$signoredConcepts){
        $signoredConcepts = array();
    }
    $linkTitles = array();
    foreach ($$SQuery as $value) {
        $parseResult = unserialize($value['parsed_content']);
        foreach ($parseResult as $phrase){
            if (!in_array($phrase['text'],$signoredConcepts)){
                if (isset($linkTitles[$phrase['text']])){
                    $linkTitles[$phrase['text']]['count']++;
                    $linkTitles[$phrase['text']]['urls'][] = $phrase['link'];
                } else {
                    $linkTitles[$phrase['text']]['count'] = 1;
                    $linkTitles[$phrase['text']]['urls'][] = $phrase['link'];
                }
            }
        }
    }
    $count = round(count($linkTitles)/100,3);
}

```

```

arsort($linkTitles);
$result = array();
foreach ($linkTitles as $title => $link){
    $fr = round(($link['count']/$count),3);
    $result[] = array(
        'concept'=> $title,
        'num'    => $link['count'],
        'fr'     => $fr
    );
}
return json_encode($result);
}

public function getLink($queryid,$pageid){
    $output = "";
    $links = $this->db->getParsedResult($queryid, $pageid);
    if (is_array($links[0])){
        $output.='
        Link: '.$links[0]['Link'].'<br>
        Parsed urls:<br>';
        $parsed_content = unserialize($links[0]['parsed_content']);
        foreach ($parsed_content as $value) {
            $output.= $value['text'] . ' ' . $value['link'] . '<br>';
        }
    }
    return $output;
}

public function getSearchString($id){
    $result = $this->calculateFinal($id);
    $result = json_decode($result,1);
    $output = "";
    foreach ($result as $value) {
        $output[] = $value['concept'];
    }
    $output = implode(',', $output);
    return $output;
}
}

```

**Додаток В. Лістинг модуля управління діалогом підсистем
структурування контенту.**

```

var counter = 0;
Ext.example = function(){
    var msgCt;
    function createBox(t, s){
        return ['<div class="msg">',
'<div      class="x-box-tl"><div      class="x-box-tr"><div      class="x-box-
tc"></div></div></div>',
'<div class="x-box-ml"><div class="x-box-mr"><div class="x-box-mc"><h3>', t,
'</h3>', s, '</div></div></div>',
'<div      class="x-box-bl"><div      class="x-box-br"><div      class="x-box-
bc"></div></div></div>',
'</div>'].join(""); }
    return { msg : function(title, format){
        if(!msgCt){ msgCt = Ext.DomHelper.insertFirst(document.body,
        {id:'msg-div'}, true); }
        var      s      =      Ext.String.format.apply(String,
        Array.prototype.slice.call(arguments, 1));
        var m = Ext.DomHelper.append(msgCt, createBox(title, s), true);
        m.hide();
        m.slideIn('t').ghost("t", { delay: 2000, remove: true}); } };
    }());
$(document).ready(
function(){
$("#start").click(function(){
$(".bar").width(0);
$(".barHidden").width(0);
doStart();
});
});
function doStart(){
var keyword = $("#keyword");
var query = "/natalia/main.php?action=get_search_result&q="+keyword.val();
$.ajax({ url: query, dataType: "json", success: function(data){
    getPages(data); } });
}
function getPages(data){
var progressbar = $("#progressbar");
var result = $("#result");
var resultCount = data['count'];
var keyword = data['query'];
var id = data['id'];

```

```

var outText = 'Запит : ' + data['query'].toString() + '<br>Отримано
результатів : ' + resultCount.toString();
outText += ' <button id="details" class="btn btn-mini"
type="button">деталі</button>';
result.html(outText);
result.animate({'opacity':'show'});
$("#details").click(function() {
    $("#processed").toggle();
});
progressbar.animate({'opacity':'show'});
for(var i = 0; i < resultCount; i++){ getSinglePage(id,i,resultCount); }
}
function getSinglePage(id,i,resultCount){
var query = "/natalia/main.php?action=process_page&q="+id+"&num="+i;
$.ajax({ url: query, dataType: "json", success: function(data){
    getProcessedPage(data,resultCount,id); counter++;
    if (counter == resultCount){ getFinal(id); } }
});
}
function getProcessedPage(data,resultCount,id){
var processed = $("#processed");
var resultText = processed.html();
var bar = $(".bar");
var barHidden = $(".barHidden");
var width = barHidden.width();
var step = 600/resultCount;
var newWidth = width+step;
resultText += "Опрацьовано сторінку : <a href=\"http://"+data['url']+\"\"
target=\"_blank\">"+data['url']+\"</a>"; resultText += " <a
href=\"main.php?action=get_link&q="+id+"&num="+data['pageid']+\"\"
target=\"_blank\">посилання</a><br>"; // resultText += 'sfsdf';
if (newWidth > 600){newWidth = 600; }
barHidden.width(newWidth);
bar.width();
bar.css("width",newWidth);
processed.html(resultText);
// processed.show();
}
function getFinal(id){
var finalResult = $("#finalResult");
Ext.onReady(function() {
    Ext.define('myModel',{ extend:'Ext.data.Model', fields:['concept','num','fr'] });
var concept = Ext.create('Ext.data.Store', {

```



```

    ]]
  }
],
//height: 300,
width: 575,
renderTo: 'finalResult'
});
});
finalResult.animate({'opacity':'show'});

var ignoredResult = $("#ignoredResult");
Ext.onReady(function() {

  Ext.define('myModel', {
    extend: 'Ext.data.Model',
    fields: ['concept','fr']
  });

  var ignored = Ext.create('Ext.data.Store', {
    model: 'myModel',
    pageSize: 50,
    proxy: {
      type: 'ajax',
      url: "/natalia/main.php?action=get_ignored&q="+id,
      reader: {
        type: 'json'
      }
    },
    autoLoad: true
  });
  Ext.create('Ext.grid.Panel', {
    store: ignored,
    viewConfig: {
      enableTextSelection: true
    },
    tbar: [
      {
        xtype: 'button',
        text: 'Оновити',
        scale : 'small',
        handler: function(event, toolEl, panelHeader) {
          ignored.reload();
        },
        border: 1,

```

```

style: {
  borderColor: 'grey',
  borderStyle: 'solid'
}
},
{
  xtype: 'button',
  text: 'Видалити всі',
  scale : 'small',
  handler: function(event, toolEl, panelHeader) {
    Ext.Ajax.request({
      url: '/natalia/main.php?action=clear_ignored&q='+id,
      success: function(response){
        ignored.reload()
      }
    });
  },
  border: 1, style: { borderColor: 'grey', borderStyle: 'solid' }
}],
columns: [
  { text : 'Концепт',
    dataIndex : 'concept',
    width : 475 },
  { text : 'Залучити',
    menuDisabled: true,
    sortable : false,  xtype : 'actioncolumn',  align : 'center',  width : 100,
    items: [{ icon : 'img/bullet_green.gif', // Use a URL in the icon config
      handler: function(grid, rowIndex, colIndex) {
        var rec = ignored.getAt(rowIndex);
        Ext.MessageBox.msgButtons['yes'].text = "Save";
        Ext.MessageBox.msgButtons['no'].text = "Dont save";
        Ext.MessageBox.msgButtons['cancel'].text = "Cancel";
        Ext.MessageBox.confirm({
          title:'Враховувати',
          msg:'Залучити концепт "' + rec.get('concept') +'" з розрахунків',
          buttonText: {
            yes: "Так",
            no: "Ні"
          },
          fn:function(button) {
            if (button === 'yes') {
              Ext.Ajax.request({
                url:
'/natalia/main.php?action=unset_ignore&q='+id+'&ignore='+rec.get('concept'),

```



```

        success: function(response){
            ignored.reload();
        }
    });
    Ext.example.msg('Концепт ', rec.get('concept')+' залучено до
    розрахунків')
        //store.reload()
        }
    }
    });
    }
    ]]
    }
    ],
    //height: 300,
    width: 575,
    renderTo: 'ignoredResult'
    });
    });
    ignoredResult.animate({'opacity':'show'});
    }

```

**Додаток Г. Лістинг модуля агрегування та згладжування підсистеми
моделювання відвідуваності.**

```

clc;
clear all;
close all;
%13-01-2013-проіндексовано тайтли в GOOGLE
%11-03-2013 - виставлено тематику на сайт
drive='i';
% drive='i';
% drive='d';
folder='\Stattia6\PageVisit_Api2Cart\';
file_all='api_AllPgs_20120305_20130728.txt';
argRead=[drive,folder,file_all]
y_all=textread([drive,folder,file_all]);
folder='\Stattia6\PageVisit_Api2Cart\Sources\';
file_File2Cart='api2cart_File2Cart_20120305_20130728.txt';
file_Magn='api2cart_Magn_20120305_20130728.txt';
file_ShopCart='api2cart_ShopCart_20120305_20130728.txt';
file_Organic='api2cart_organic_20120305_20130728.txt';
y_File2Cart=textread([drive,folder,file_File2Cart]);
y_Magn=textread([drive,folder,file_Magn]);
y_ShopCart=textread([drive,folder,file_ShopCart]);
y_Organic=textread([drive,folder,file_Organic]);
uwr0=zeros(54,1);
sz=size(y_all)
nt=sz(2)
tt=1:nt;
nf=0;

nw=round(nt/7)
for i=1:nw
sMag=0;
sal=0;
sOrg=0;
is=0;
for j=1:7
jj=j+(i-1)*7;
if (jj<=nt)
j7=j+(i-1)*7;
sal=sal+y_all(jj+(i-1)*7);
sMag=sMag+y_File2Cart(j7)+y_Magn(j7)+y_ShopCart(j7);
sOrg=sOrg+y_Organic(j7);

```

```

    is=is+1;
    end
end
y7al(i)=sal/is;
y7mag(i)=sMag/is;
y7org(i)=sOrg/is;
end
t7=1:nw;
nsm=1;
nsm1=3;
nf=nf+1;
figure(nf)
smOrg1=smooth(y7org);
smMag1=smooth(y7mag);
smy0=smooth(y7al);
uwr1=smooth(uwrT);
for jj=2:nsm1
    smOrg1=smooth(smOrg1);
    smMag1=smooth(smMag1);
    smy0=smooth(smy0);
    uwr1=smooth(uwr1);
end
end
szT7=size(t7)
plot (t7,y7al,'-k',t7,5*y7mag,':k',t7,2*y7org,'--k',t7,2*uwrT,'-.k','LineWidth',2)
title(['Загальна відвідуваність ','FontName','Arial CYR'])
h = legend('allPages','5*PagesMagnetic','2*PagesOrg','2*email',2);
xlabel('тижні')
ylabel('перегляди сторінок в день')
grid on
smMag1(45)=smMag1(45)+1.5;
smMag1(46)=smMag1(46)+2.2;
smMag1(47)=smMag1(47)+1.9;
smMag1(48)=smMag1(48)+0.5;

nf=nf+1;
My7=max(y7al);
sy7=size(y7al)
sy0=size(smy0)
ey=abs(y7al'-smy0)/My7;
Me7=max(ey);
Aey=mean(ey);
figure(nf)
plot (t7',y7al','--k',t7',smy0,'-r','LineWidth',2)

```

```

title(['Згладжування відвідуваності, середн. відхил.=',
num2str(Aey*100),'%',...
', макс. відхил.=', num2str(Mey*100),'%'],'FontName','Arial CYR')
xlabel('тижні')
ylabel('перегляди сторінок в день')
grid on

nf=nf+1;
figure(nf)
plot (t7',smy0','-k',t7,5*smMag1','k',t7,2*smOrg1,'--k',t7,2*uw1,'-
.k','LineWidth',2)
title(['Total Attendance Trend'],'FontName','Arial CYR')
h = legend('Attendance','5*Announcing','2*Publications','2*Delivering',2);
xlabel('Weeks')
ylabel ('Page Views')
grid on

%виділення факторів впливу
% побудова похідної відвідуваності
DMу=0.04; %максимальний рівень частки похідної у відвідуваності
dsmy=size(nw-1,1);
t7d=1:nw-1;
Msmy0=max(smy0);
for i=1:nw-1
    dsmy(i)=smy0(i+1)-smy0(i);
end
dsmy=dsmy/Msmy0;
d2smy=size(nw-1,1);
t7d2=1:nw-1;
d2smy(1)=0;
for i=1:nw-2
    d2smy(i+1)=dsmy(i+1)-dsmy(i);
end

nf=nf+1;
figure(nf)
plot (t7d',dsmy','-k','LineWidth',2)
title(['Похідна згладженої відвідуваності'],'FontName','Arial CYR')
h = legend('похідна 1 порядку','похідна 2 порядку',2);
grid on

nf=nf+1;
figure(nf)
plot (t7',smy0','-k',t7,3*smMag1','k',t7,smOrg1,'--k',t7,uwr1,'-.k','LineWidth',2)

```

hold on

```

hglev=60
axis([2 72 0 hglev])
iAct=0;
vert=1:hglev;
point=ones(1,hglev);
for i=1:nw-1
if (dsmy(i)>DMy)&&(iAct==0)
% активізація відвідуваності
iAct=1
% друк вертикалі
plot (i*point,vert,'--b','LineWidth',2)
elseif (dsmy(i)<DMy)&&(iAct==1)
% деактивізація відвідуваності
iAct=0
end
end
title(['Total Attendance Trend'],'FontName','Arial CYR')
h = legend('Attendance','3*Announcing','Publications','Delivering',2);
xlabel('Weeks')
ylabel('Page Views')
grid on
ntst=44+15;
ntfn1=73;
ntfn=73;
zt7=t7(ntst:ntfn)-ntst+1;
szT7=size(t7)
OrgApr=smOrg1(ntst:ntfn1)
yApr=smy0(ntst:ntfn1)
zApr=t7(ntst:ntfn1);
ntst2=63;
ntfn2=73;
yOrg=smy0(ntst2:ntfn2);
tOrg=[ntst2:ntfn2]-ntst2;
Org2=smOrg1(ntst2:ntfn2);
yOrgM(1)=yOrg(1);
nT=ntfn2-ntst2+1;
a1= 0.2644 ;
a2= 1.7490;
for jj=2:nT
yOrgM(jj)=yOrgM(1)+a1*(Org2(jj)-Org2(1))^a2;
end

```

```

nid1=2;
AA=zeros(nid1,2);
bb=zeros(nid1,1);
for i=2:nid1+1
    AA(i,1)=1;
    AA(i,2)=log(smOrg1(i)-smOrg1(1));
    bb(i)=log(smy0(i)-smy0(1));
end
AAA=AA
bbb=bb
a=(AA'*AA)\(AA'*bb)
a(1)=exp(a(1)) % побудова значення a1
a1= a(1) ;
a2= a(2);
for jj=2:nT
    yOrgM2(jj)=yOrgM(1)+a1*(Org2(jj)-Org2(1))^a2;
end
Vidv1=smy0(27:39)
Vidv2=smy0(39:50)
Vidv3=smy0(54:64)
Vidv4=smy0(65:73)
Fact1Print=smOrg1(27:39)
Fact2Print=smMag1(41:52)
Fact3Print=uwr1(53:63)
Fact4Print=smOrg1(65:73)

```

**Додаток Д. Акти про впровадження результатів дисертаційної роботи
Д.1 Впровадження результатів роботи в навчальний процес ТНЕУ**

„ЗАТВЕРДЖУЮ”

Перший проректор

Тернопільського національного
економічного університету

Шинкарик М.І.

2014 р.



АКТ

про впровадження в навчальний процес Тернопільського національного економічного університету результатів дисертаційної роботи
Пасічник Наталії Романівни
„ Математичні моделі відвідуваності Веб-сайтів та методи їх ідентифікації”

Даний акт складений про те, що результати дисертаційної роботи аспіранта кафедри комп'ютерних наук Пасічник Наталії Романівни на тему „Математичні моделі відвідуваності Веб-сайтів та методи їх ідентифікації” використані в навчальному процесі факультету комп'ютерних інформаційних технологій Тернопільського національного економічного університету для студентів напряму підготовки 6.050103 – „Програмна інженерія”. Зокрема, при викладанні дисциплін „Web-дизайн” та „Управління продажами в ІТ”:

- надається інформація щодо побудови та застосування математичних моделей в процесі розробки та підтримки Веб-сайтів;
- надається інформація щодо побудови та застосування математичних моделей для оцінки ефективності заходів підвищення відвідуваності Веб-сайтів;
- використовується у процесі виконання лабораторних робіт розроблене у дисертаційній роботі програмне забезпечення для генерації типових структур тематичних Веб-сайтів.

Декан факультету комп'ютерних
інформаційних технологій,
зав. кафедри комп'ютерних наук,
д.т.н., проф.

М. П. Дивак

Доцент кафедри комп'ютерних наук,
к.т.н.

М. Я. Шпінталь

Додаток Д. Акти про впровадження результатів дисертаційної роботи
Д.2. Впровадження результатів роботи на ПП «МагнетікВан»



Юридична адреса: вул. Симоненка 4/164, м. Тернопіль, Україна, 46016

Поштова адреса: вул. Бродівська 5Б, м. Тернопіль, Україна, 46002

тел. +38 (0352) 523-831

ЗАТВЕРДЖУЮ

Виконавчий директор

ПП «МагнетікВан»

Череватий О. П.

“ 20 ” березня 2014 р.

АКТ

про впровадження результатів дисертаційної роботи

Пасічник Наталії Романівни

на тему "«Математичні моделі відвідуваності Веб-сайтів та методи їх ідентифікації»

у виробничий процес підприємства

Цей акт свідчить про те, що при формуванні контенту Веб-сайту проекту API2Cart та моніторингу його відвідуваності впроваджено такі результати, розроблені Пасічник Н.Р.

- неавтономну модель відвідуваності Веб-сайтів, яка включає модель динаміки фактора впливу на основі модифікованої системи звичайних диференціальних рівнянь Моно;
- методи ідентифікації моделі процесу короткотермінового та довготермінового росту відвідуваності Веб-сайтів;
- метод побудови структур тематичних Веб-сайтів на основі частотного аналізу та ранжування, який дозволяє будувати типову структуру тематичного Веб-сайту, узгоджену із структурою відповідної предметної області та будувати шляхи оновлення контенту
- розроблене програмне забезпечення в середовищі «Matlab» для моделювання відвідуваності Веб-сайтів та з використанням технологій PHP і JavaScript для формування контенту Веб-сайтів на основі онтологічного підходу.

Виконавчий директор



Череватий О. П.

Додаток Е. Участь у виконанні прикладних розробок

Додаток Е.1. Довідка про участь у виконанні держбюджетних прикладних розробок



Тернопільський національний економічний університет
Ternopil National Economic University

Lvivska Str. 11, Ternopil, 46020, Ukraine
 Tel./Fax +380 (352) 47 50 51
 E-mail: academ@tneu.edu.ua
 http://www.tneu.edu.ua

вул. Львівська, 11, Тернопіль, 46020, Україна
 Тел./факс +380 (352) 47 50 51
 E-mail: academ@tneu.edu.ua
 http://www.tneu.edu.ua

№ 126-21/1365

“28” 05 2014р.

На № _____ від _____

**Національний університет «Львівська
 політехніка»**

**Голові спеціалізованої вченої ради Д 35.052.05
 д.т.н, професору Мандзію Б. А.**

ДОВІДКА

Видана **Пасічник Наталії Романівні** про те, що вона брала участь у виконанні держбюджетної прикладної розробки з теми «Інформаційна технологія для ідентифікації і візуалізації зворотнього гортанного нерва в процесі хірургічної операції на щитовидній залозі» (номер державної реєстрації 0112U000078).

На даний час бере участь у виконанні держбюджетної прикладної розробки з теми «Теорія побудови та методи реалізації в реальному часі міждисциплінарних математичних моделей зміни стану складних об'єктів» (номер державної реєстрації 0114U000569) та кафедральних науково-дослідних робіт з таких тем: «Макромодельовання складних систем та процесів в умовах структурної невизначеності на основі неточних даних» (державний реєстраційний номер 0111U010356); «Математичні моделі процесів навантаження та релаксації в складних системах» (державний реєстраційний номер 0113U000848).

Проректор з наукової роботи
д.е.н., професор



З.-М. В. Задорожний

Вик. Письменний В.І.
 (0352)47-50-59

Додаток Е. Участь у виконанні прикладних розробок
Додаток Е.2. Акт про використання результатів дисертаційної роботи

«ЗАТВЕРДЖУЮ»

Проректор з наукової роботи



Тернопільського національного економічного
університету

д.е.н., проф. З.-М.В. Задорожний

05 _____ 2014 р.

про використання результатів кандидатської дисертаційної роботи

Пасічник Наталії Романівни

**«Математичні моделі процесів відвідуваності Веб-сайтів
та методи їх ідентифікації»**

Комісія у складі голови - завідувача кафедри комп'ютерних наук, керівника науково-дослідної роботи, д.т.н., проф. Дивака М.П. та членів: директора Науково-дослідного інституту інноваційного розвитку та державотворення Монастирського Г. Л., начальника відділу прогнозування і маркетингу Кушніра О.Р. склали цей акт про те, що дослідження та результати дисертаційної роботи Пасічник Н.Р. використані під час виконання науково-дослідних робіт на кафедрі комп'ютерних наук з безпосередньою участю автора, а саме:

- науково-дослідної роботи «Інформаційна технологія для ідентифікації та візуалізації зворотного гортанного нерва в процесі хірургічної операції на щитовидній залозі» (номер державної реєстрації 0112Ш00078, 2012-2013рр.), у якій автором розроблено критерії ефективності Веб-сайтів та метод побудови структур тематичних Веб-сайтів на основі онтологічного підходу;
- науково-дослідної роботи на тему: «Теорія побудови та методи реалізації в реальному часі міждисциплінарних математичних моделей зміни стану складних об'єктів» (номер державної реєстрації 0114Ш00569), де автором розроблено дискретну динамічну модель відвідуваності Веб-сайтів та метод її ідентифікації;
- науково-дослідної роботи на тему: „Макромодельовання складних систем та процесів в умовах структурної невизначеності на основі неточних даних" (номер державної реєстрації 011113010356), де автором розроблено метод ідентифікації моделі короткотермінової динаміки відвідуваності Веб-сайту;
- науково-дослідної роботи на тему „Математичні моделі процесів навантаження та релаксації в складних системах" (номер державної реєстрації 011311000848), де автором розроблено метод ідентифікації моделі довготермінової динаміки відвідуваності Веб-сайту.

Голова комісії

завідувач кафедри комп'ютерних наук,
керівник НДР, д.т.н., проф.

Дивак М.П.

Члени комісії:

Директор НДІ ІРД

Монастирський Г. Л.

Начальник відділу
прогнозування і маркетингу

Кушнір О.Р.