

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІНЖЕНЕРНИЙ НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ім. Ю.М. ПОТЕБНИ

Кафедра інформаційної економіки, підприємництва та фінансів
(повна назва кафедри)

Кваліфікаційна робота
другий (магістерський)
(рівень вищої освіти)

на тему Управління оцінюванням контенту web-ресурсів університету

Виконав: студент 2 курсу, групи 8.0511-іе-д
спеціальності 051 Економіка
(код і назва спеціальності)
освітньої програми Інформаційна економіка
(код і назва освітньої програми)
спеціалізації _____
(код і назва спеціалізації)

Ломейко Сергій Дмитрович
(ініціали та прізвище)

Керівник професор кафедри інформаційної економіки,
підприємництва та фінансів, д.е.н. Глущевський В.В.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)

Рецензент д.е.н., доц. Клопов І.О.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)

Запоріжжя
2022

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІНЖЕНЕРНИЙ НАВЧАЛЬНО-НАУКОВИЙ
ІНСТИТУТ ім. Ю.М. ПОТЕБНІ

Кафедра Інформаційної економіки, підприємництва та фінансів
Рівень вищої освіти другий (магістерський)
Спеціальність 051 Економіка
(код та назва)
Освітня програма Інформаційна економіка
(код та назва)
Спеціалізація _____
(код та назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри д.е.н. доц.

Шапуров О.О.

«_____» _____ 20__ року

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТОВІ

Ломейко Сергій Дмитрович

(прізвище, ім'я, по батькові)

1. Тема роботи (проекту) Управління оцінюванням контенту web-ресурсів університету

керівник роботи Глушчевський В'ячеслав Валентинович
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом ЗНУ від від «25» жовтня 2022 року № 1445-с

2. Строк подання студентом роботи 9 грудня 2022 року

3. Вихідні дані до роботи аналітичні показники роботи web-ресурсів університетів та структурних підрозділів

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) 1) дослідити загальні особливості організації інформаційних web-ресурсів; 2) проаналізувати методи та моделі порівняльного аналізу в задачах раціональної організації контенту web-ресурсів університетів; 3) обґрунтувати підхід до моделювання базових показників оцінки ефективності функціонування інформаційних web-ресурсів; 4) розробити алгоритм збору інформації за допомогою парсингу та збереження інформації web-ресурсів ЗВО; 5) проаналізувати контент web-ресурсу кафедри ІЕПФ ЗНУ з аналогами інших ЗВО; 6) провести моделювання процесу оцінки ефективності функціонування контенту web-ресурсів кафедри ІЕПФ ЗНУ.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) Напрямки та етапність організації контенту, Вимоги щодо організації контенту з

позицій класифікації web-ресурсів ЗВО, Алгоритм порівняльного аналізу web-сайтів при організації його контенту, Алгоритм збору інформації за допомогою парсингу та збереження інформації web-ресурсів, Алгоритм процедури виявлення некоректної та застарілої інформації на web-ресурсі кафедри ІЕПФ ЗНУ.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	професор, д.е.н. кафедри інформаційної економіки, підприємництва та фінансів Глущевський В.В.	17.09	19.10
2	професор, д.е.н. кафедри інформаційної економіки, підприємництва та фінансів Глущевський В.В.	19.10	29.10
3	професор, д.е.н. кафедри інформаційної економіки, підприємництва та фінансів Глущевський В.В.	29.10	16.11

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Призначення наукових керівників. Затвердження тем дипломних робіт	30.06.2022	
2	Напрацювання теоретичного матеріалу: дослідження сутності об'єкту та предмету дослідження, критичний аналіз існуючих методологічних засад, вибір та обґрунтування напрямку проведення дослідження	01.09.2022	
3	Апробація результатів на Міжнародних та Всеукраїнських конференціях	Протягом навчального року	
4	Розробка інформаційної моделі забезпечення основних елементів концептуального підходу.	01.10.2022	
5	Збір та систематизація статистичного та нормативного матеріалу дослідження.	15.10.2022	
6	Узагальнення отриманих результатів. Оформлення роботи	30.10.2022	
7	Надання роботи та автореферату до рецензії. Нормоконтроль	25.11.2022	
8	Прилюдний захист дипломної роботи на засіданні ЕК	15.12.2022	

Студент _____
(підпис)

Ломейко С.Д.
(ініціали та прізвище)

Керівник роботи (проекту) _____
(підпис)

Глущевський В.В.
(ініціали та прізвище)

Нормоконтроль пройдено

Нормоконтролер _____
(підпис)

Н.Г. Метеленко
(ініціали та прізвище)

АНОТАЦІЯ

Ломейко С. Д. Управління оцінюванням контенту web-ресурсів університету.

Кваліфікаційна випускна робота для здобуття ступеня вищої освіти магістра за спеціальністю 051 – Економіка, науковий керівник В.В. Глущевський. Запорізький національний університет Інженерний навчально-науковий інститут ім. Ю.М. Потебні. Кафедра інформаційної економіки, підприємництва та фінансів, 2022.

В роботі вирішено актуальну проблему ефективної організації та оцінювання контенту web-ресурсів структурних підрозділів університетів. На прикладі кафедри інформаційної економіки, підприємництва та фінансів ІННІ ім. Ю.М. Потебні ЗНУ було удосконалено методологічний підхід до процедури виявлення некоректної та застарілої інформації на web-ресурсах ЗВО.

Ключові слова: WEB-РЕСУРС, КОНТЕНТ, ЗВО, МОДЕЛЮВАННЯ.

ABSTRACT

Lomeiko S. Managing an Evaluation of University Web Content.

Qualifying final work for obtaining a master's degree in higher education by specialty 051 - Economics, supervisor V. Glushchevsky Zaporizhzhya National University Engineering Educational and Scientific Institute named after Y.M. Potebni. Department of Information Economics, Entrepreneurship and Finance, 2022.

The work solves the actual problem about effective organization and evaluation of the web-resources content in universities structural divisions. On the example of the Department of Information Economics, Entrepreneurship and Finance ZNU. At the request of the Higher Education Institution, the methodological approach to the procedure for identifying incorrect and outdated information on the web-resources of the Higher Education Institution was improved.

Key words: WEB-RESOURCE, CONTENT, HIGHER EDUCATION INSTITUTION, MODELING.

ЗМІСТ

ВСТУП	6
РОЗДІЛ 1. ТЕОРЕТИЧНИЙ ОГЛЯД ТА ПОРІВНЯЛЬНИЙ АНАЛІЗ ПІДХОДІВ ДО ОРГАНІЗАЦІЇ КОНТЕНТУ НА СТОРІНКАХ WEB-РЕСУРСІВ	9
1.1. Загальні поняття та показники організації контенту інформаційних web-ресурсів	9
1.2. Аналіз специфіки сайтів ЗВО в межах класифікації web-ресурсів	13
1.3. Дослідження систем контент-аналізу web-ресурсів	15
Висновки до розділу 1	24
РОЗДІЛ 2. МЕТОДИ ТА МОДЕЛІ ОЦІНЮВАННЯ КОНТЕНТУ WEB -РЕСУРСІВ	25
2.1. Математичні методи порівняльного аналізу web-ресурсів при оцінюванні контенту	25
2.2. Математична модель ефективності функціонування інформаційних web-ресурсів в умовах різної їх структурованості	33
2.3. Розробка алгоритму виявлення некоректної та неактуальної інформації web-ресурсів	37
Висновки до розділу 2	42
РОЗДІЛ 3. АНАЛІЗ ТА ОЦІНКА КОНТЕНТУ WEB-РЕСУРСУ КАФЕДРИ ІНФОРМАЦІЙНОЇ ЕКОНОМІКИ, ПІДПРИЄМНИЦТВА ТА ФІНАНСІВ ЗНУ	43
3.1. Порівняльний аналіз контенту web-ресурсів структурних підрозділів університетів	43
3.2. Аналіз переваг та недоліків в організації контенту web-ресурсу кафедри ІЕПФ ЗНУ	52
3.3. Реалізація моделі ефективності функціонування контенту web-ресурсів	57
Висновки до розділу 3	62
ВИСНОВКИ	64
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	67

ВСТУП

Розвиток інформаційних технологій забезпечує розвиток суспільства, переоцінювання його пріоритетів, особливо модернізацію існуючих галузей економіки та створення і стійкий розвиток нових. Цифровізація усіх сфер діяльності людини сьогодні є життєвою необхідністю. Одним із засобів розв'язування потреб суспільства є інформаційні web-ресурси.

Зокрема, web-ресурси навчальних закладів відіграють все більшу роль в створенні їх іміджу та залученні абітурієнтів в умовах конкуренції, епідемії коронавірусу та війни в Україні. Відповідно зростає значимість раціональної організації контенту на сторінках web-ресурсів з метою ефективного сприйняття відвідувачами наявної інформації. Аналіз публікацій вітчизняних та зарубіжних вчених показав, що ефективність функціонування більшої кількості web-ресурсів є невисока через слабку структурованість контенту а також наявність в контенті застарілої та неактуальної інформації. Тому її виявлення та оновлення є актуальною задачею, розв'язування якої лежить у сфері розробки засобів оцінювання, структурування та розпізнавання контенту.

Саме така ситуація склалась зокрема на кафедрі інформаційної економіки, підприємництва та фінансів ІННІ ім. Ю.М. Потебні ЗНУ, що одержала додаткову актуальність у зв'язку із необхідністю оцінки web-ресурсу <https://znuierpf.com.ua>.

Об'єктом дослідження є процеси контент-аналізу web-ресурсів університету.

Предметом дослідження є інструментальні засоби та методи забезпечення підтримки функціонування інформаційних web-ресурсів за умов різної їх структурованості.

Метою дослідження є підвищення ефективності управління контентом інформаційних web-ресурсів в умовах різної структурованості контенту на

прикладі сайту кафедри Інформаційної економіки, підприємництва та фінансів Інженерного навчально-наукового інституту ім. Ю.М. Потебні ЗНУ.

Для досягнення цієї мети поставлено та вирішено такі завдання:

- 1) дослідити загальні особливості організації інформаційних web-ресурсів;
- 2) проаналізувати методи та моделі порівняльного аналізу в задачах раціональної організації контенту web-ресурсів університетів;
- 3) обґрунтувати підхід до моделювання базових показників оцінки ефективності функціонування інформаційних web-ресурсів;
- 4) розробити алгоритм збору інформації за допомогою парсингу та збереження інформації web-ресурсів ЗВО;
- 5) проаналізувати контент web-ресурсу кафедри ІЕПФ ЗНУ з аналогами інших ЗВО;
- б) провести моделювання процесу оцінки ефективності функціонування контенту web-ресурсів кафедри ІЕПФ ЗНУ.

Методи дослідження. Для розробки методу моделювання базового показника оцінки ефективності функціонування інформаційних web-ресурсів залежно від структурованості контенту використано стохастичний підхід та методи параметричної ідентифікації моделей нелінійних систем. Для розробки методів структурування, розпізнавання та порівняння текстового контенту web-ресурсів використовувалися методи семантичного опису та аналізу контенту із елементами машинного навчання.

Наукова новизна одержаних результатів. У кваліфікаційній роботі вирішено актуальну наукову проблему ефективної організації та оцінювання контенту web-ресурсів структурних підрозділів університетів. Найбільш суттєві наукові результати полягають у такому:

дістала подальшого розвитку:

алгоритм порівняльного аналізу web-сайтів при організації його контенту, що має практичне значення для кафедр та факультетів ЗВО та

дозволяє обґрунтувати заходи з підвищення ефективності інформаційних ресурсів;

удосконалено:

методологічний підхід до процедури виявлення некоректної та застарілої інформації на web-ресурсах ЗВО, який, на відміну від існуючих, враховує внутрішні та зовнішні джерела інформації через бази ЕДЕБО та власні розробки інформаційних систем ЗВО, що дозволяє синхронізувати контент web-ресурсів із еталонною та актуальною інформацією з бази даних ЗВО.

Практичне та теоретичне значення дослідження полягає в тому, що запропонований підхід до моделювання базових показників оцінки ефективності функціонування інформаційних web-ресурсів дозволяє здійснювати управління контентом web-ресурсів ЗВО на якісно новому рівні.

Структура та обсяг кваліфікаційної роботи. Магістерська робота складається зі вступу, трьох розділів, висновків, списку використаних літературних джерел. Загальний обсяг роботи - 71 сторінка, 10 таблиць, 31 рисунок, 43 джерела літератури.

РОЗДІЛ 1

ТЕОРЕТИЧНИЙ ОГЛЯД ТА ПОРІВНЯЛЬНИЙ АНАЛІЗ ПІДХОДІВ ДО ОРГАНІЗАЦІЇ КОНТЕНТУ НА СТОРІНКАХ WEB-РЕСУРСІВ

1.1. Загальні поняття та показники організації контенту інформаційних web-ресурсів

Розглянемо особливості організації інформаційних web-ресурсів. Структура сайту є найважливішим інструментом із точки зору SEO (пошукова оптимізація). Неправильна структура сайту ускладнює його відвідуваність. Тому при розробці архітектури web-ресурсу необхідно проаналізувати розміщення кожного розділу та підрозділу, щоб зробити його таким, який би задовільнив потреби користувача і реагував на вимоги пошукових роботів [10].

Немає чіткого визначення щодо того, якою є правильна структура. Це залежить від типу веб-сайту, семантичного ядра та цільової аудиторії, тому він завжди індивідуальний. Однак існують еталонні типи конструкцій, а також основні правила його розвитку. Структура сайту - логічна побудова всіх сторінок ресурсу, схема, за якою розподіляється шлях до папок, категорій, підкатегорій (рис.1.1).

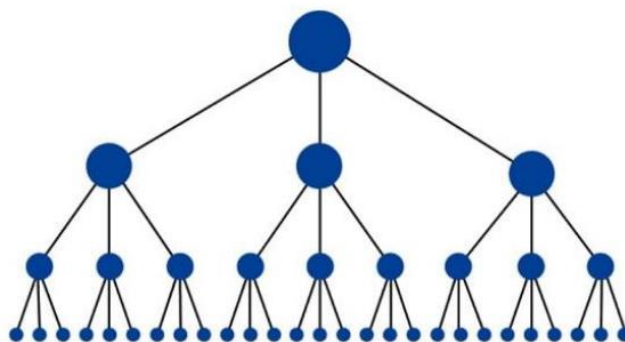


Рис. 1.1. Структура сторінок web-ресурсу [5]

Пошукові роботи переходять за всіма посиланнями сайту. Якщо структура веб-сайту правильна, пошуковий робот витратить менше часу на аналіз усього ресурсу, що пришвидшить індексацію сайту для ефективного

SEO просування. Правильно та логічно розміщені документи на сайті можуть потрапити в індекс пошукової системи вже наступного дня. Це невід'ємна частина для SEO-просування [23, 37].

Раціональна організація контенту web-ресурсів має різні важливі аспекти (наприклад, існують окремі напрямки оптимізації з позиції просування сайту або з позицій сприйняття контенту користувачем). Також дослідження можуть виконуватись із різними цілями та на різних етапах практичного створення web-рішень. Існує багато різних категорій web-сайтів, відносно яких може виконуватись подібне дослідження. Уточнимо напрямки та рамки, в яких буде виконуватись ця кваліфікаційна робота магістра, спираючись на завдання - рис.1.2.



Рис. 1.2 Напрямки та етапність організації контенту

* Розроблено автором

Перш за все, виділимо напрямки організації контенту за принципово різними критеріями. Це зокрема напрямок оптимізації просування сайтів (search engine optimization, SEO) із використанням пошукових систем [39], а також напрямки забезпечення юзабіліті та комфортного сприйняття змісту для користувачів. Спираючись на формулювання завдання на дипломну роботу, обмежимося виключно дослідженням раціонального розміщення контенту з позицій його сприйняття користувачем.

Дослідження щодо раціональної організації контенту web-ресурсів в прикладному плані може передувати розробці, як інструмент аналізу сайтів-аналогів. Надалі власно розробка включає моделювання розміщення контенту, яке передує втіленню сайту. І уже на цьому етапі методика дослідження може використовуватись для попереднього оцінювання одержаних результатів та виявлення напрямків їх удосконалення.

Також, на рис. 1.2, зазначено категорію сайтів, для яких буде виконуватись дослідження. Виходячи із умови обґрунтування рекомендацій щодо оцінки ефективності кафедри ІЕПФ, логічним є обрати для аналізу веб-сайти споріднених кафедр та університетів. Саме з цього будемо виходити в при проведенні дослідження.

Аналіз літературних джерел, свідчить про те, що існує декілька ключових показників, які визначають загальну релевантність вмісту [32, 39,40]:

1. Для визначення актуальності чи недостовірності контенту пошукові системи звертають увагу на дату створення сторінок. На перших позиціях розміщені ті сайти, інформація яких була опублікована трохи пізніше.

2. Важливим є оновлення контенту на сайті, що також впливає на пошукові системи при визначенні його актуальності. Щоб постійно мати важливий і корисний контент на ресурсі, недостатньо просто змінити кілька слів у статтях або навіть видалити одне речення, оскільки численні

конкуренти, які змінюють зміст на пару абзаців, обов'язково обійдуть ресурс у процесі пошуку.

3. Контент потрібно оновлювати часто. Частота вимірюється в днях, а не місяцях, упродовж багатьох років.

4. Пошукові системи приділяють увагу свіжому контенту, численним новоствореним сторінкам. Зважаючи на це, необхідно забезпечити збільшення загальної кількості сторінок щороку приблизно на 30%.

5. Загальна поведінка користувача відіграє певну роль, яка також показує, наскільки релевантними є вміст. Якщо оновлення контенту ресурсу викликає серйозне поживлення користувачів до нього, то вони набагато довше залишаються на сторінках, відповідно, можна судити про актуальність інформації.

6. Якщо ресурс згадується все частіше, якщо він з'являється у все більшій загальній кількості посилань на ресурс із інших сайтів, репостів, лайків у соціальних мережах, пошукові системи роблять висновок, що сайти містять надзвичайно актуальний контент.

Актуальність та вагомість контенту описуються такими параметрами:

1. Зміна документу, домену або веб-сторінки. Сайти, які додають нові та створюють абсолютно новий контент, привертатимуть увагу цільової аудиторії.

2. Внесення змін до важливих областей документів, які показують важливість сторінки. Менш важливим вмістом є реклама та навігація. Найважливіша інформація, як правило, знаходиться на головній сторінці ресурсу.

3. Збільшення кількості посилань та частоти їх додавання. На новизну сайтів впливають посилання із найсвіжіших сторінок ресурсу. Новинними є також ті ресурси, на які всі постійно посилаються. Якщо сайт кардинально зміниться сам і своє розташування, новий якір також буде змінений. Пошукова система визначає, що ресурс настільки змінився, що старий якірний текст більше не актуальний. Це повністю знецінює всі старі зв'язки.

Таким чином, ефективність функціонування інформаційних web-ресурсів визначається рядом чинників, серед яких варто виділити: продуктивність апаратного і програмного забезпечення, способи організації джерела, зокрема структурованість контенту та складність запиту.

Разом з тим ефективність функціонування більшої кількості web-ресурсів є невисока через слабку структурованість контенту, а також наявність у ньому застарілої та неактуальної інформації.

1.2. Аналіз специфіки сайтів ЗВО в межах класифікації web-ресурсів

В ході дослідження та оцінки контенту web-ресурсів ЗВО, доцільно уточнити позиціонування сайтів університетів та кафедр в класифікації. На перший погляд, сайти ЗВО або їх підрозділів (таких, як окремі кафедри або факультети) найбільше підходять класу «візиток», отже їх головна функція — створення відповідного іміджу закладу освіти. Однак, аналіз існуючих сайтів такої категорії, дозволяє виділити декілька їх суттєвих особливостей:

- На відміну від «класичних» візиток, як їх зокрема характеризує [8], web-ресурси ЗВО та їх підрозділів мають значно більший обсяг (зазвичай йдеться про 10-100 сторінок);
- Оскільки в останні роки суттєво загострилась конкуренція ЗВО за абітурієнтів, на їх сайтах відповідно підсилюється рекламний аспект, тобто вони все більше набувають рис сайтів послуг, активно просуваючи послуги освіти в конкурентному середовищі;
- В силу великого обсягу інформації, яку необхідно донести потенційним одержувачам послуг освіти, щоб ті мали можливість всебічно оцінити їх переваги, сайти ЗВО також набувають і риси «статейників».

Таким чином, вимоги до розміщення контенту на сайтах ЗВО з позицій класифікації сайтів повинні враховувати їх «гібридне» позиціонування, яке включає риси сайтів-візиток, сайтів послуг та статейників (рис. 1.3).

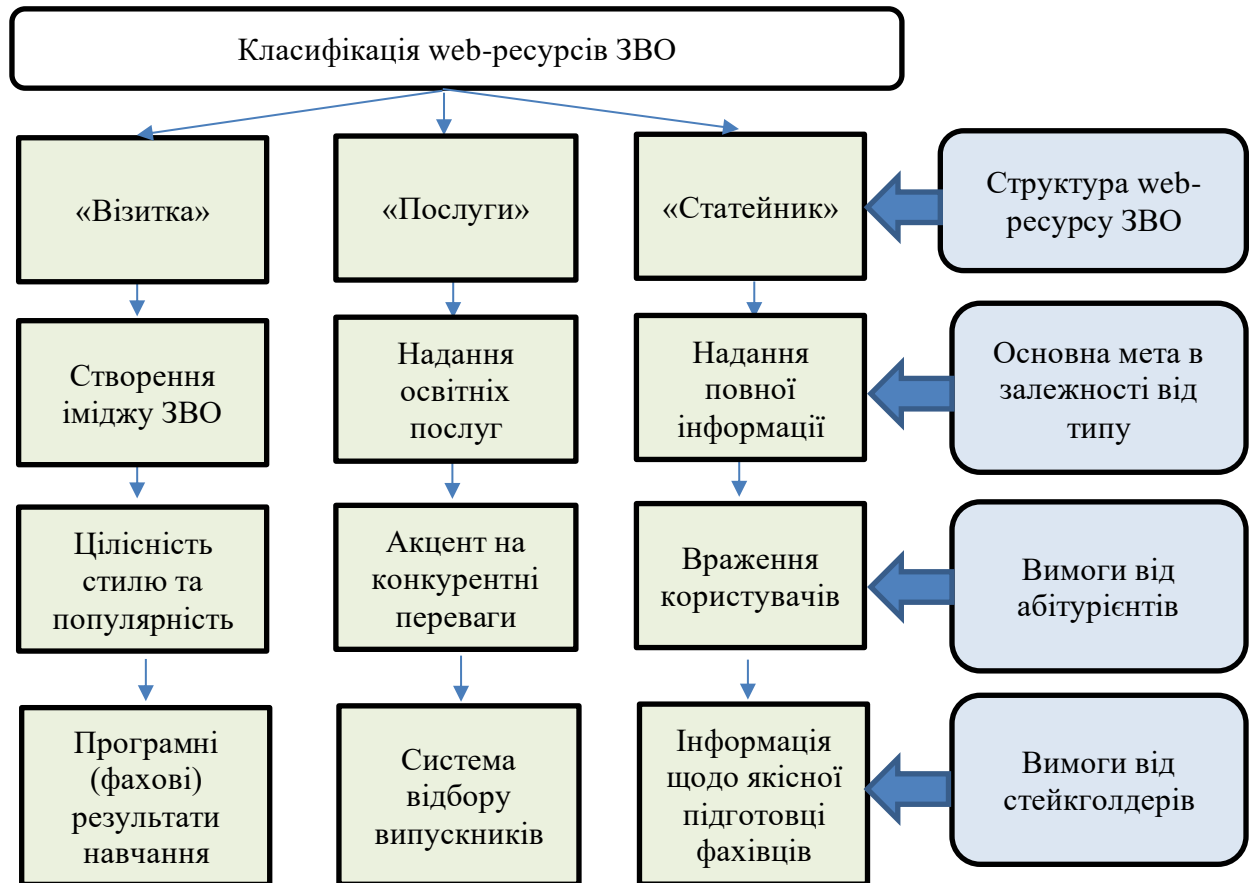


Рис. 1.3 Вимоги щодо організації контенту з позицій класифікації web-ресурсів ЗВО

* Вдосконалено автором за [8]

– Для сайтів-візиток основною метою є створення іміджу ЗВО. Відповідно вимогою до організації контенту тут має бути впізнаваність та популярність, яка досягається зокрема за рахунок UI/UX компонентів. Останнє стосується як візуальної складової (якість дизайну, фірмовий графічний стиль), так і текстового наповнення (стиль подачі, структурованість тексту та його зручність для сприйняття). Основною вимогою стейкхолдерів на таких web-ресурсах є наявність інформації щодо освітніх програм та програмних результатів навчання для формування рівня знань та навичок потенційних співробітників серед випускників.

– Основна функція сайтів послуг це активне просування саме власних послуг в конкурентному середовищі. Відповідно, вимоги до розміщення контенту включають акцентування корисності послуг для потенційного клієнта (наприклад, виважений та змістовний опис цінності спеціальності «Інформаційна економіка» в сучасному конкурентному середовищі), а також одержання продукту саме від цього постачальника (виділення переваг навчання саме в цьому ЗВО). Важливим аспектом такого сайту є форми зворотнього зв'язку, анкети на подання за програмою 2D диплом, відеоплатформа записів лекцій та презентацій курсів тощо;

– Сайт «статейник» має за мету створення такого інформаційного середовища, яке максимально сприяє детальному розгляду предмета і зокрема його сприйняття через враження користувачів (що має створити довіру і емоційне підключення). Такий підхід передбачає певний формат подачі — розгорнутий, деталізований з різних позицій для абітурієнтів, студентів, викладачів та інших зацікавлених сторін. Актуальність та змістовність інформаційного ресурсу є ключовим такого типу сайтів.

Таким чином, класифікація web-ресурсів ЗВО щодо організації контенту та його погодженням є складною задачею і потребує подальшого дослідження.

1.3. Дослідження систем контент-аналізу web-ресурсів

Одним із відомих методів оцінки текстової інформації web-ресурсів є контент-аналіз – стандартна методика дослідження, предметом якої є аналіз змісту текстових масивів і продуктів комунікативної кореспонденції (наприклад, коментарі, форуми, електронне листування, статті тощо) [25, 41].

Поняття контент-аналізу не має однозначного визначення, що породжує проблему: системи, побудовані на основі різних підходів до контент-аналізу, в загальному випадку несумісні [33].

Основними складовими дослідження контент-аналізу [6, 9, 17]:

1. Виокремлення контенту.
2. Структурування - гіпотеза структурування, сегментація, поділ текстів (повторення в усіх/ряді текстів) у всьому досліджуваному масиві.
3. Формалізація – використання однотипних сегментів та застосування високого ступеня формалізації, визначених правил та формальних алгоритмів у аналітичних процедурах.
4. Реферування – формальний поділ текстів та визначення окремих елементів для наступного пошуку із застосуванням аналітико-синтетичної процедури.
5. Аналіз – застосування методів математичної статистики та методів теорії ймовірності для аналізу тексту.

Аналіз основних методів контент-аналізу web-ресурсів представлений в таблиці 1.1.

Таблиця 1.1.

Аналіз методів контент аналізу

Назва методу	Показники		
	Складність реалізації	Часові параметри	Можливість програмної інтеграції
NLP - Методи об'єктно - якісного та систематичного дослідження змісту засобів комунікації (Д. Джері, Дж. Джері)	Висока	Великі часові витрати на формалізацію	Часткова інтеграція в існуюче ПЗ
NLP - Систематичне кількісне опрацювання, оцінювання та інтерпретація форми і змісту інформаційного джерела (Д. Мангейм, Р. Рич)	Середня	Великі часові витрати на формалізацію	Інтеграція на рівні окремого ПЗ
NLP - Якісно-кількісний метод дослідження документів (характеризується об'єктивністю висновків і строгістю процедури) та квантифікаційного опрацювання тексту з подальшою інтерпретацією результатів; предмет дослідження – проблеми соціальної дійсності, які висловлюються і приховуються у документах, та внутрішні закономірності самого об'єкту дослідження (В. Іванов)	Висока	Великі часові витрати на формалізацію	Інтеграція на рівні окремого ПЗ
Text mining - Пошук у тексті визначених змістовних понять (одиниць аналізу), виявлення частоти їх появи та співвідношення із змістом всього документа (Б. Краснов)	Середня	Середні часові витрати на формалізацію	Часткова інтеграція в існуюче ПЗ
Word2vec - метод векторного представлення, коли вкладання слів будують множини векторних представлень слів на основі вхідних текстів (Г. Міколов)	Середня	Великі часові витрати на формалізацію	Інтеграція на рівні окремого ПЗ
Skip-gram (словосполучення з пропусками) визначає навколишній контекст	Середня	Середні часові витрати на	Інтеграція на рівні окремого ПЗ

на основі заданого слова із врахуванням порядку слів; СВОВ (Continuous Bag Of Words, неперервний мішок слів) - обернений до попереднього метод		формалізацію	
---	--	--------------	--

* Вдосконалено автором на основі [1, 12]

Проведемо дослідження відомих систем контент-аналізу. Системи контент-аналізу поділяються за типами [19].

Перший тип – повністю автоматизовані пакети для контент-аналізу. Включають в себе розроблені програмно-аналітичні словники, в які, зазвичай, неможливо або досить важко внести зміни. Такі програми застосовуються для перевірок лінгвістичних гіпотез, закладених в основу програмних словників та аналітичних схем. Адже сучасний рівень розвитку "штучного інтелекту" не дозволяє аналітику передавати програмі повний контроль над процесом інтерпретації масиву документів. Жодна сучасна програма не може розуміти текст в людському значенні цього поняття [11, 28 34].

Другий тип програм – надзвичайно прості та універсальні пакети підрахунку частот різних слів в текстах. До недоліків таких програм слід віднести відсутність категоризації підрахованих слів, що призводить до викривлення результатів в бік окремо беззмистовних слів, що належать до службових частин мови. Крім того, різні форми змістовних слів рахуються як різні слова, що не дозволяє поставити знак рівності між одиницями підрахунку та одиницями значень. Останній із згаданих недоліків був частково розв'язаний розробниками через процедуру лематизації (виділення незмінних частин слів). Однак подібних програм з лематичним вдосконаленням для слов'янських мов, які мають надзвичайно гнучку морфологію, яка передбачає досить відмінні форми для різних родів, чисел, відмінків тощо, не існує [30, 31, 37, 38].

Третій тип програм – напівавтоматичні пакети для ручного кодування текстових даних. Вони спрощують деякі технічні аспекти кодування: користування ними аналогічне альтернативі "писати олівцем чи писати на комп'ютері". Дані комплекси позбавляють кодувальника постійно звертатись

до кодувальної таблиці, упорядковують результати кодування багатьох кодувальників, зберігають дані в зручній для пошуку і маркування формі, візуалізують знайдені зв'язки. Дозволяють заощадити певний час та підвищують ефективність, однак у випадку масивів інформації, що вимірюється сотнями і тисячами сторінок, дане програмне забезпечення відчутно не допоможе. Тому такі програми найчастіше використовуються для упорядкування якісного контент-аналізу та для контент-аналітичних кейс-стаді [22, 24, 26, 27].

Під час проведення контент-аналізу web-ресурсу кожен із сторінок сайту можна представити в аналітичній моделі у вигляді системи масового обслуговування, що моделює певну затримку (час перебування користувача на сторінці).

Використовуються такі показники:

1. Кількість сторінок web-ресурсу.
2. Інтенсивність вхідного потоку заявок на сторінку ззовні.
3. Частина відвідувачів сайту, для якої обчислюється максимальний час перебування на web-ресурсі.
4. Середній час перебування користувача на web-ресурсі.

Виділено такі критерії оцінки якості веб-ресурсів:

- зовнішній вигляд web-документу, дизайн web-ресурсу є головним і найчастіше єдиним критерієм оцінки його якості;

- структура та навігація - керування вмістом web-ресурсів, а також різні сервіси. Наприклад, для web-ресурсів ЗВО це «Форма зворотнього зв'язку», «Пошук», «Форма реєстрації за програмою 2D диплом», «Чат-бот для відповідей на загальні питання абітурієнтів». Функціональність web-ресурсу - простота та зручність, оскільки працюватимуть із ним користувачі, які можуть нічого не знати про web-дизайн та HTML;

- загальна зручність використання web-ресурсу. За допомогою дружнього UI/UX дизайну можна завоювати довіру, а неякісний сайт тільки відштовхне відвідувачів. Сайти, створені відповідно до правил оцінки якості

та зручності використання web-сторінок, найпопулярніші та мають більш високу статистику відвідувань [5, 36].

Таким чином, навігація веб-сайту повинна бути простою та зручною, web-сторінки мають завантажуватися швидко.

Розглянуто системи для оцінки ефективності web-ресурсів.

PageSpeed Insights - це інструмент, що вимірює швидкість завантаження сторінки, аналізує параметри і дає рекомендації щодо покращення. Результат перевірки - оцінка за стобальною шкалою. Багато оптимізаторів і експертів не радять повністю орієнтуватися на цю оцінку, оскільки немає точних даних, що вона безпосередньо впливає на ранжування сайту.

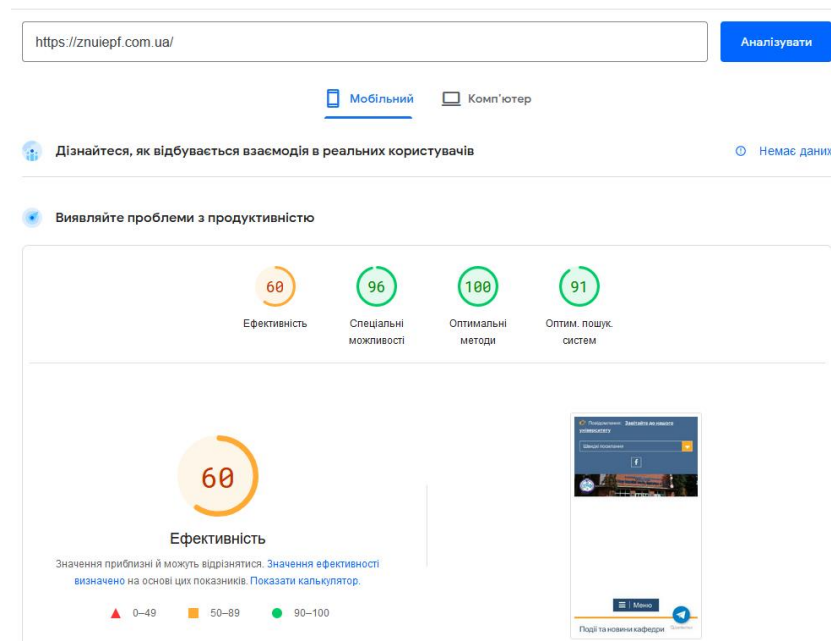


Рис. 1.4. Сторінка сайту кафедри ІЕПФ в системі PageSpeed Insights
Інструмент показує результати для мобільних і десктопних пристроїв
й аналізує сторінку за такими параметрами:

1. Перша візуалізація вмісту.
2. Затримка в завантаженні даних.
3. Індекс швидкості.
4. Період до повного завантаження.
5. Перший простій центрального процесора.

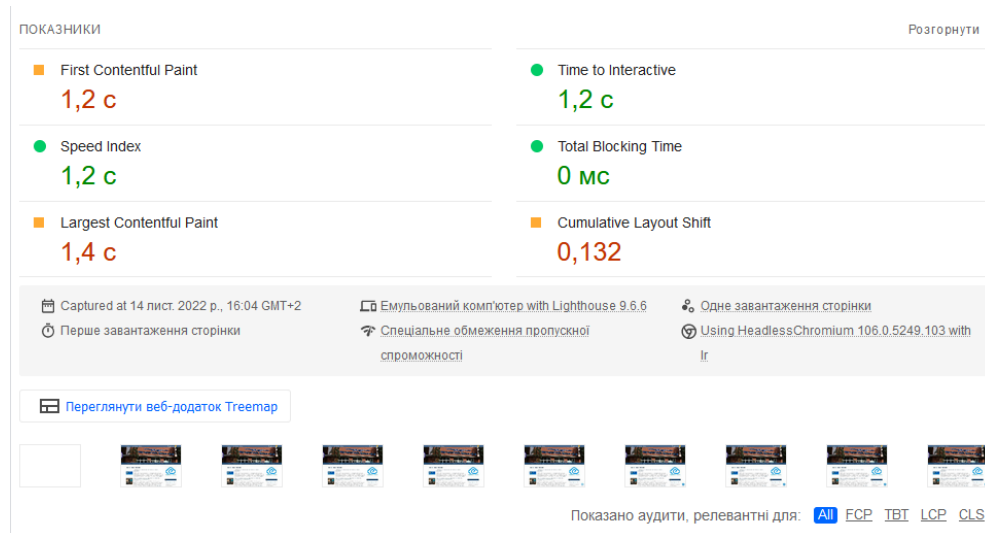


Рис. 1.5. Сторінка системи PageSpeed Insights визначення ефективності web-ресурсу

Цей інструмент користується великою популярністю у плані зменшення розмірів скриптів, зображень та інших елементів, що корисно для підвищення швидкості завантаження сайту. Використання кешу браузера, рекомендований сервісом у деяких випадках, теж дуже важлива частина прискорення ресурсу [6, 35].

Tools.pingdom.com – інструмент для оцінки швидкості роботи web-ресурсів. Відображає кожний пошуковий запит, що формується при зверненні відвідувачів до ресурсу. Інколи на сайтах є безліч зайвих запитів, яких візуально не видно, а фізично вони присутні.

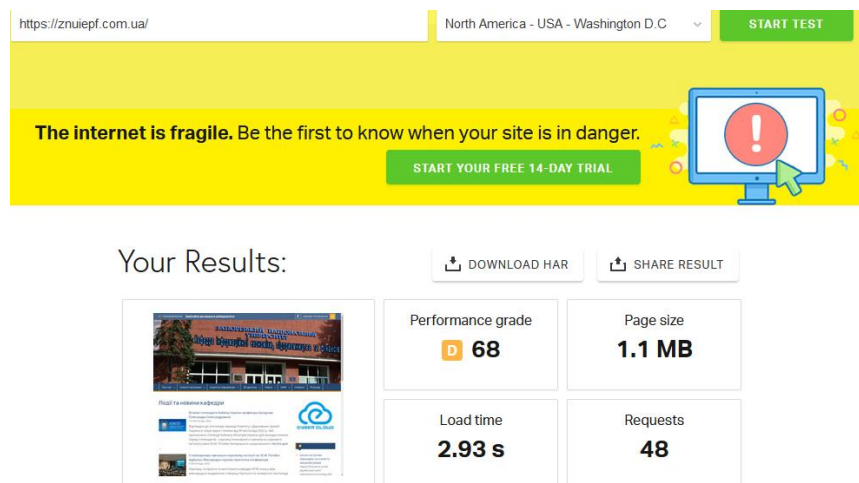


Рис. 1.6. Сторінка системи Tools.pingdom.com

Загальне значення показника web-ресурсу представлено на рис. 1.7.

Improve page performance

GRADE	SUGGESTION
F 0	Compress components with gzip
F 0	Use cookie-free domains
F 0	Add Expires headers
F 20	Make fewer HTTP requests
B 90	Make favicon small and cacheable
B 90	Avoid URL redirects
A 95	Avoid HTTP 404 (Not Found) error

Рис. 1.7. Сторінка системи Tools.pingdom.com із загальними значення показника web-ресурсу

Розмір web-ресурсу кафедри Інформаційної економіки, підприємництва та фінансів, яка тестувалася 1.1 МВ, що означає дуже добре. В даному випадку web-ресурс швидко працює, але через велику кількості картинок на ньому, якщо розмір сторінки буде більше 10 МВ інформації, тоді є можливість втратити відвідуваність.

На рис. 1.8. представлено деталізовану інформацію по об'єму завантаження функціональних частин web-ресурсу кафедри ІЕПФ. Як видно контент (image, html) складають приблизно 25% всі інше програмний код для відображення web-ресурсу у браузері.

Content size by content type






	PERCENT	
 Script	29.35%	317.4 KB
 CSS	24.71%	267.1 KB
 Image	19.85%	214.7 KB
 Font	19.55%	211.4 KB
 HTML	6.47%	69.9 KB

Рис. 1.8.

GTmetrix – це безкоштовна програма, що перевіряє швидкість web-сторінки та подає детальний звіт. Вона генерує оцінки продуктивності web-сторінок та надає рекомендації щодо їх покращення.

Особливості:

1. Тестувати web-ресурсів у різних країнах, а також у браузерах.
2. Надає короткий зміст ключових показників ефективності.

3. Можливість відстежувати ефективність web-ресурсів за допомогою моніторингу та бачити його за допомогою графіків.

4. Здатність перевірити web-ресурс на різних модельованих пристроях.

5. Дозволяє відтворювати завантаження web-сторінок із відео.

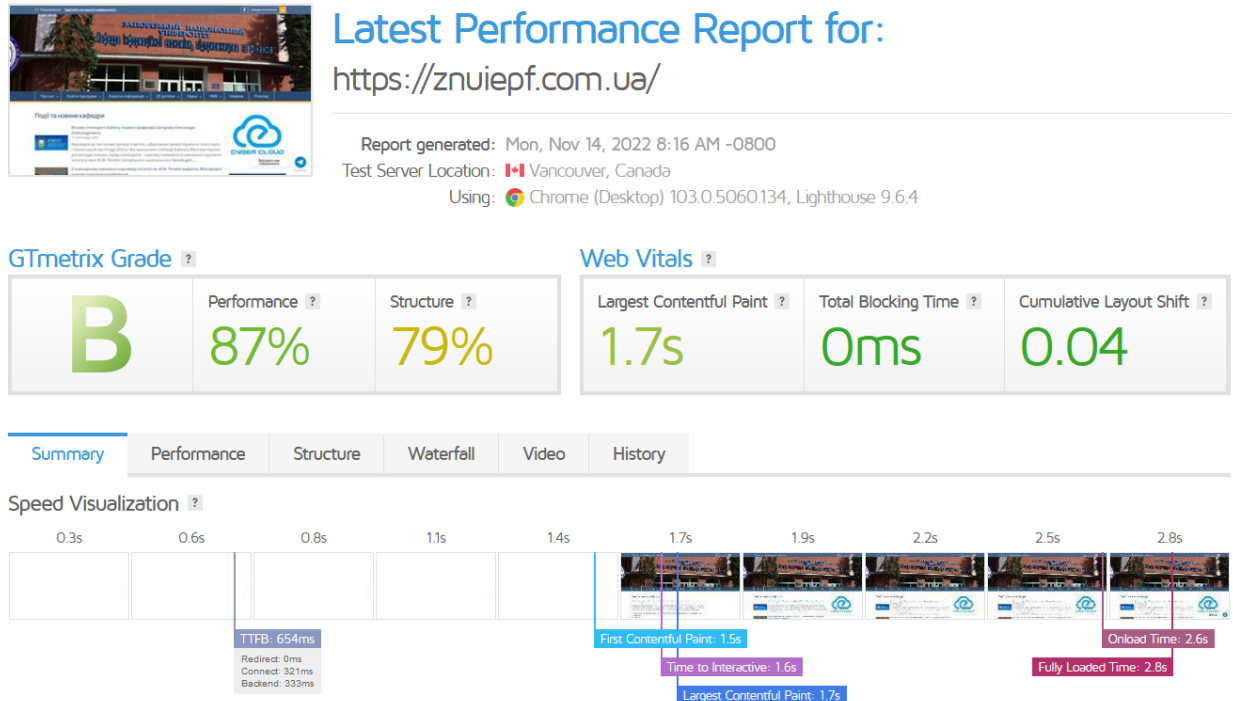


Рис. 1.9. Сторінка системи GTmetrix

Performance Metrics – показник ефективності Lighthouse, зафіксований GTmetrix, з спеціальними аудитами, параметрами аналізу, специфікаціями браузера та обладнання.

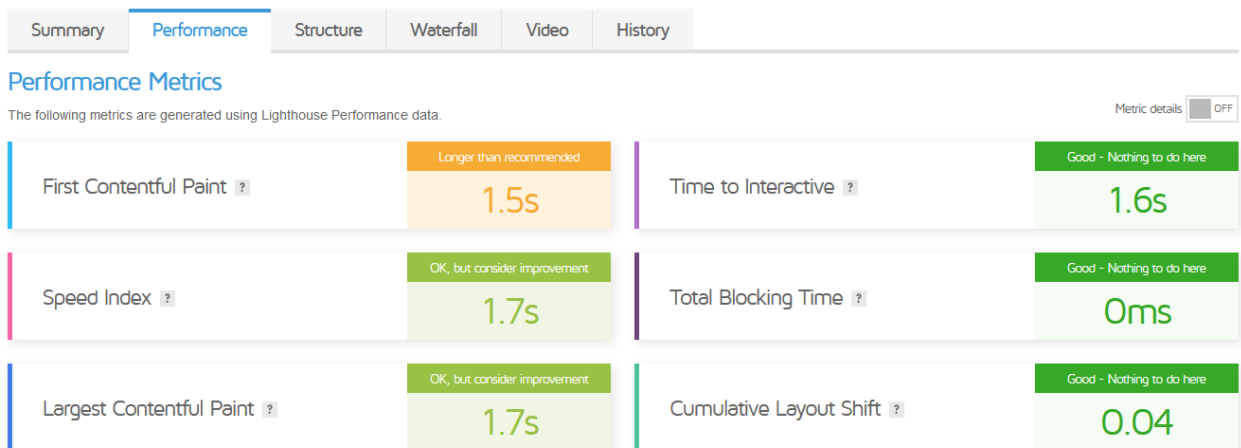


Рис. 1.10. Сторінка системи GTmetrix - показник ефективності

Structure – відображає, наскільки добре побудована ваша сторінка для оптимальної роботи.

IMPACT	AUDIT	Show Audits Relevant to
Med-High	Eliminate render-blocking resources <small>FCP LCP</small>	Potential savings of 561ms
Med	Enable text compression <small>FCP LCP</small>	Potential savings of 403KB
Med	Serve static assets with an efficient cache policy	Potential savings of 844KB
Med-Low	Use a Content Delivery Network (CDN)	35 resources found
Low	Reduce unused CSS <small>FCP LCP</small>	Potential savings of 225KB
Low	Use passive listeners to improve scrolling performance	1 event listener not passive
Low	Ensure text remains visible during webfont load <small>FCP LCP</small>	4 fonts found
Low	Avoid chaining critical requests <small>FCP LCP</small>	23 chains found
Low	Avoid an excessive DOM size <small>TBT</small>	470 elements
Low	Avoid enormous network payloads <small>LCP</small>	Total size was 1.02MB
Low	Properly size images	Potential savings of 10.1KB
Low	Reduce JavaScript execution time <small>TBT</small>	45ms spent executing JavaScript

Рис. 1.11. Сторінка системи GTmetrix структурованість сторінки кафедри ІЕПФ

Таким чином, у даному пункті обґрунтовується показники оцінки контенту web-ресурсів, які стосуються витрат часу на обробку запитів. Як впливає із вищенаведеного аналізу, існує спектр систем для оцінки ефективності контенту web ресурсів, проте усі вони використовуються для констатації факту ефективності того чи іншого web-ресурсу. Разом з тим, жоден із них не дає можливості прогнозувати та моделювати показник ефективності. Вони лише придатні для збору поточної інформації про інформаційний web-ресурс. Тому для використання показника ефективності необхідно розробити його математичне представлення і провести комп'ютерне моделювання та верифікацію запропонованого математичного опису з метою дослідження впливу структурованості контенту на ефективність web-ресурсу.

Висновки по розділу 1.

В першому розділі було розглянуто особливості організації інформаційних web-ресурсів та спираючись на поставлені завдання, уточнено напрямки, в яких буде виконуватись ця кваліфікаційна робота магістра. Зазначено, що ефективність функціонування інформаційних web-ресурсів визначається рядом чинників, серед яких варто виділити: продуктивність апаратного і програмного забезпечення, способи організації джерела web-ресурсів, зокрема структурованість контенту та складність запиту.

Визначено, що ефективність функціонування більшої кількості web-ресурсів є незначною, через слабку структурованість контенту, а також наявність у ньому застарілої та неактуальної інформації.

В ході дослідження та оцінки контенту web-ресурсів ЗВО, уточнено позиціонування сайтів університетів та кафедр в класифікації сайтів, та вдосконалено вимоги щодо організації контенту web-ресурсів ЗВО для задоволення потреб всіх зацікавлених сторін.

Обґрунтовано показники оцінки контенту web-ресурсів, які стосуються витрат часу на обробку запитів. Зазначено, що жоден з наявних показників не дає можливості прогнозувати та моделювати ефективність контенту web-ресурсу. Вони лише придатні для збору поточної інформації про інформаційний web-ресурс. Тому для використання показника ефективності необхідно розробити його математичне представлення і провести комп'ютерне моделювання та верифікацію запропонованого математичного опису з метою дослідження впливу структурованості контенту на ефективність web-ресурсу.

Такий підхід забезпечить не тільки встановлення цього впливу для конкретного web-ресурсу, але і забезпечить моделювання й оцінку структурованості контенту для розв'язування задач щодо його оптимізації, аналізу та перевірки на цілісність, достовірність й актуальність.

РОЗДІЛ 2

МЕТОДИ ТА МОДЕЛІ ОЦІНЮВАННЯ КОНТЕНТУ WEB - РЕСУРСІВ

2.1 Математичні методи порівняльного аналізу web-ресурсів при оцінюванні контенту

Існує різноманіття математичних методів та моделей при проведенні порівняльного аналізу. Враховуючи вибір задачі порівняльного аналізу організації контенту на сайтах-аналогах доцільно зосередитись саме на відповідних математичних методах та моделях. Методи та моделі порівняльного аналізу, які підходять для проведення дослідження web-ресурсів, описані зокрема у наукового керівника Глущевського В.В. [4]. Розглянемо основні підходи, що запропоновані в роботі.

Першим пропонуємо розглянути порівняння за багатьма критеріями – задача оцінки якості відноситься до багатокритеріальних задач оптимізації. Існує велика кількість підходів до рішення такого типу задач [3]:

- Теорії корисності для багатокритеріального вибору альтернатив з множини в умовах ризику і невизначеності;
- Зведення багатокритеріальної задачі до скалярної оптимізації;
- Розробка людино-машинних процедур рішення багатокритеріальних задач оптимізації в інтерактивному режимі;
- інші.

При оцінюванні об'єктів за багатьма критеріями можна виділити наступні недоліки:

- Суперечливість критеріїв;
- Неможливість аналітичного зв'язку між критеріями;
- Різний тип оцінок за критеріями, наприклад чисельні та змістовні;
- Чисельні оцінки відрізняються за розмірністю;
- Відмінність критеріїв за важливістю;

Ці недоліки можливо зняти за допомогою виявлення та врахування суб'єктивних суджень експертів. Зазвичай від людини потрібна наступна інформація:

- Оцінка об'єктів за критеріями;
- Важливість критеріїв;
- Перелік порівнювальних об'єктів;
- Перелік порівняльних критеріїв;

За принципом приведення оцінок об'єктів до єдиної оцінки можна виділити наступні методи [4]:

- Метод на основі вибору головного критерію;
- Метод на основі компенсації критеріїв;
- Метод на основі розрахування загальних оцінок;
- Метод на основі парного порівняння;

Розглянемо основні методи порівняння об'єктів наведені у роботі [3, 4, 7]:

Метод аналізу ієрархій – при застосуванні даного методу використовується така послідовність етапів: спочатку виконується структуризація задачі, далі будується ієрархічне представлення задачі, виявляються експерти оцінки та виконується обробка експертних оцінок.

Метод комплексної оцінки – засновано на розрахунках загальної оцінки з урахуванням оцінок усіх критеріїв.

Метод порівняння з використанням функції корисності, під функцією корисності розуміються функції $P = F(X)$, яка описує залежність корисності альтернатив P від оцінок цих альтернатив X .

Таким чином, існує значна кількість математичних методів, які можна розглядати в плані їх використання для порівняльного аналізу організації контенту на різних web-ресурсах.

На нашу думку доцільно обрати метод, який пристосований для використання оцінок експертів, які неодмінно мають враховуватись при оцінюванні організації контенту web-ресурсу.

Зокрема зручною для використання в нашому випадку є модель прийняття рішень в нечітких умовах Беллмана-Заде [13], яка поєднує у собі методи парного порівняння та експертних оцінок за допомогою критеріїв важливості об'єктів.

Нехай $X=\{x\}$ – множина альтернатив. Нечітку мету \tilde{G} будемо ототожнювати нечіткою множиною \tilde{G} в X . Наприклад, якщо альтернативами є дійсні числа $X = R$, і нечітка мета сформульована як « x має бути близько 10», то її можна представити нечіткою множиною з наступною функцією приналежності:

$$\mu_G(x) = \frac{1}{1+(x-10)^2}, x \in X \quad (2.1)$$

Аналогічним чином нечітке обмеження \tilde{C} визначається як деяка нечітка множина на універсальній множині X .

Нечітке рішення \tilde{D} (найкращий варіант) функція приналежності якої показує, наскільки подібне рішення задовольняє нечітким цілям і нечітким обмеженням. Логічній операції « \cap », яка пов'язує цілі з обмеженнями, відповідає перетин нечітких множин. Отже, рішення - це перетин нечіткої мети \tilde{G} з нечітким обмеженням \tilde{C} (2.2):

$$\tilde{D} = \tilde{G} \cap \tilde{C} \quad (2.2)$$

Взаємозв'язок між нечіткими метою, обмеженням та рішенням показана на рис. 2.1:

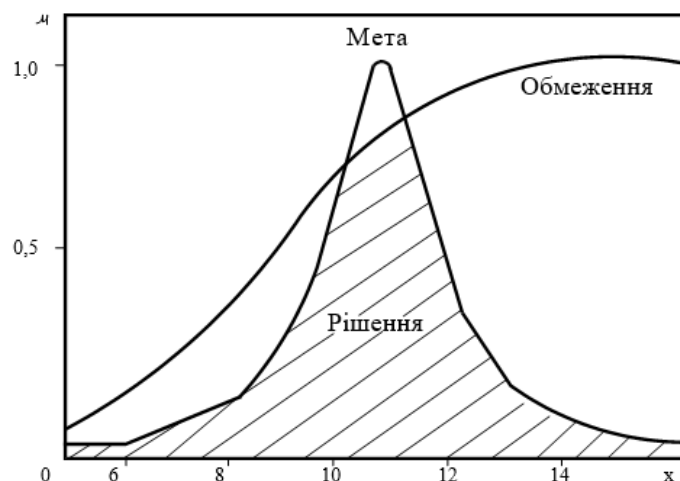


Рис. 2.1 Прийняття рішень за принципом Беллмана-Заде [13]

Мета і обмеження конфліктують між собою, тому в нечіткій множині \tilde{D} немає жодного елемента зі ступенем приналежності, що дорівнює одиниці. Отже, не існує альтернативи, яка повністю задовольняє і цілі, і обмеженню. В якості чіткого рішення в таких випадках зазвичай вибирають альтернативу з максимальним ступенем приналежності нечіткій множині \tilde{D} .

При прийнятті рішень за схемою Беллмана-Заде не створюється будь-яка різниця між метою та обмеженнями. В даному випадку робиться допущення, що поділ на ціль і обмеження є умовним. У традиційній теорії прийняття рішень подібні заміни функції переваги на обмеження не допустимі. Однак і тут простежується деяка приховане схожість між цілями та обмеженнями. Вона стає явним при використанні методу невизначених множників Лагранжа та штрафних функцій, коли мета і обмеження об'єднуються в одну функцію.

У загальному випадку, коли є n цілей і m обмежень, результуюче рішення за схемою Беллмана-Заде визначається перетином всіх цілей і обмежень (2.3) [13]:

$$\tilde{D} = \tilde{G}_1 \cap \tilde{G}_2 \cap \dots \cap \tilde{G}_n \cap \tilde{C}_1 \cap \tilde{C}_2 \cap \dots \cap \tilde{C}_m \quad (2.3)$$

і відповідно (2.4):

$$\mu_D = \mu_{G_1} \cap \mu_{G_2} \cap \dots \cap \mu_{G_n} \cap \mu_{C_1} \cap \mu_{C_2} \cap \dots \cap \mu_{C_m} \quad (2.4)$$

Всі цілі та обмеження, що входять до \tilde{D} , мають однакову важливість. Більш звична ситуація, в якій задоволення одним цілям і (або) обмеженням важливіше, ніж іншим. Позначимо через $\alpha_i \in (0,1)$ - коефіцієнт відносної важливості i -й мети, а через $\beta_j \in (0,1)$ - коефіцієнт відносної важливості j -го обмеження:

$$\sum_{i=1,n} \alpha_i + \sum_{j=1,m} \beta_j = 1 \quad (2.5)$$

Тоді функція приналежності рішення визначається так (2.6):

$$\mu_D = (\mu_{G_1})^{\alpha_1} \cap (\mu_{G_2})^{\alpha_2} \cap \dots \cap (\mu_{G_n})^{\alpha_n} \cap (\mu_{C_1})^{\beta_1} \cap (\mu_{C_2})^{\beta_2} \cap \dots \cap (\mu_{C_m})^{\beta_m} \quad (2.6)$$

Чим менше коефіцієнт відносної важливості, тим відповідна нечітка множина мети або обмеження стає більш розмазаним і, отже, його роль в ухваленні рішення знижується.

Будемо вважати відомими:

$P = \{P_1, P_2, \dots, P_k\}$ - множина варіантів, які підлягають багатокритеріальному аналізу (web-ресурси);

$G = \{G_1, G_2, \dots, G_n\}$ - множина критеріїв, за яким оцінюють варіанти (критерії оцінки контенту).

Завдання багатокритеріального аналізу полягає в упорядкуванні елементів множини P за критеріями з множини G .

Нехай $\mu_G(P_j)$ - число в діапазоні $[0,1]$, якими оцінюється варіант $P_j \in P$ за критерієм $G_j \in G$ чим більше число $\mu_G(P_j)$, тим краще варіант P_j за критерієм $G_i, i=1, n, j=1, k$. Тоді критерій можна представити нечіткою множиною \tilde{G}_i на універсальній множині варіантів P (2.7):

$$G_i = \left\{ \frac{\mu_{G_i}(P_1)}{P_1}, \frac{\mu_{G_i}(P_2)}{P_2}, \dots, \frac{\mu_{G_i}(P_k)}{P_k} \right\} \quad (2.7)$$

де $\mu_G(P_j)$ - ступінь приналежності P_j нечіткій множині \tilde{G} .

Знаходити ступені приналежності нечіткої множини (2.7) зручно методом побудови функцій приналежності на основі парних порівнянь. При використанні цього методу необхідно сформулювати матриці парних порівнянь варіантів за кожним критерієм. Загальна кількість таких матриць дорівнює кількості критеріїв.

Найкращим варіантом буде той, який одночасно кращий за всіма критеріями. Нечітке рішення \tilde{D} знаходиться як перетин приватних критеріїв (2.8):

$$\tilde{D} = \tilde{G}_1 \cap \tilde{G}_2 \cap \dots \cap \tilde{G}_n = \left\{ \frac{\min_{i=1, n} \mu_{G_i}(P_1)}{P_1}, \frac{\min_{i=1, n} \mu_{G_i}(P_2)}{P_2}, \dots, \frac{\min_{i=1, n} \mu_{G_i}(P_k)}{P_k} \right\} \quad (2.8)$$

Згідно з отриманою нечіткою множиною \tilde{D} найкращим варіантом слід вважати той, у якого найбільша ступінь приналежності (2.9):

$$D = \arg \max(\mu_D(P_1), \mu_D(P_2), \dots, \mu_D(P_k)) \quad (2.9)$$

При нерівноважних умовах ступеня приналежності нечіткої множини \tilde{D} знаходять так:

$$\mu_D(P_j) = \min_{i=1,n}(\mu_{G_i}(P_j))^{\alpha_i}, \quad j = 1, k \quad (2.10)$$

де α_i - коефіцієнт відносної важливості критерію G_i , $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$.

Коефіцієнти відносної важливості критерію G_i можуть бути визначені різними методами, наприклад, за допомогою парних порівнянь за шкалою Сааті.

Для оцінки скористаємося критеріями: G_1, G_2, \dots, G_n .

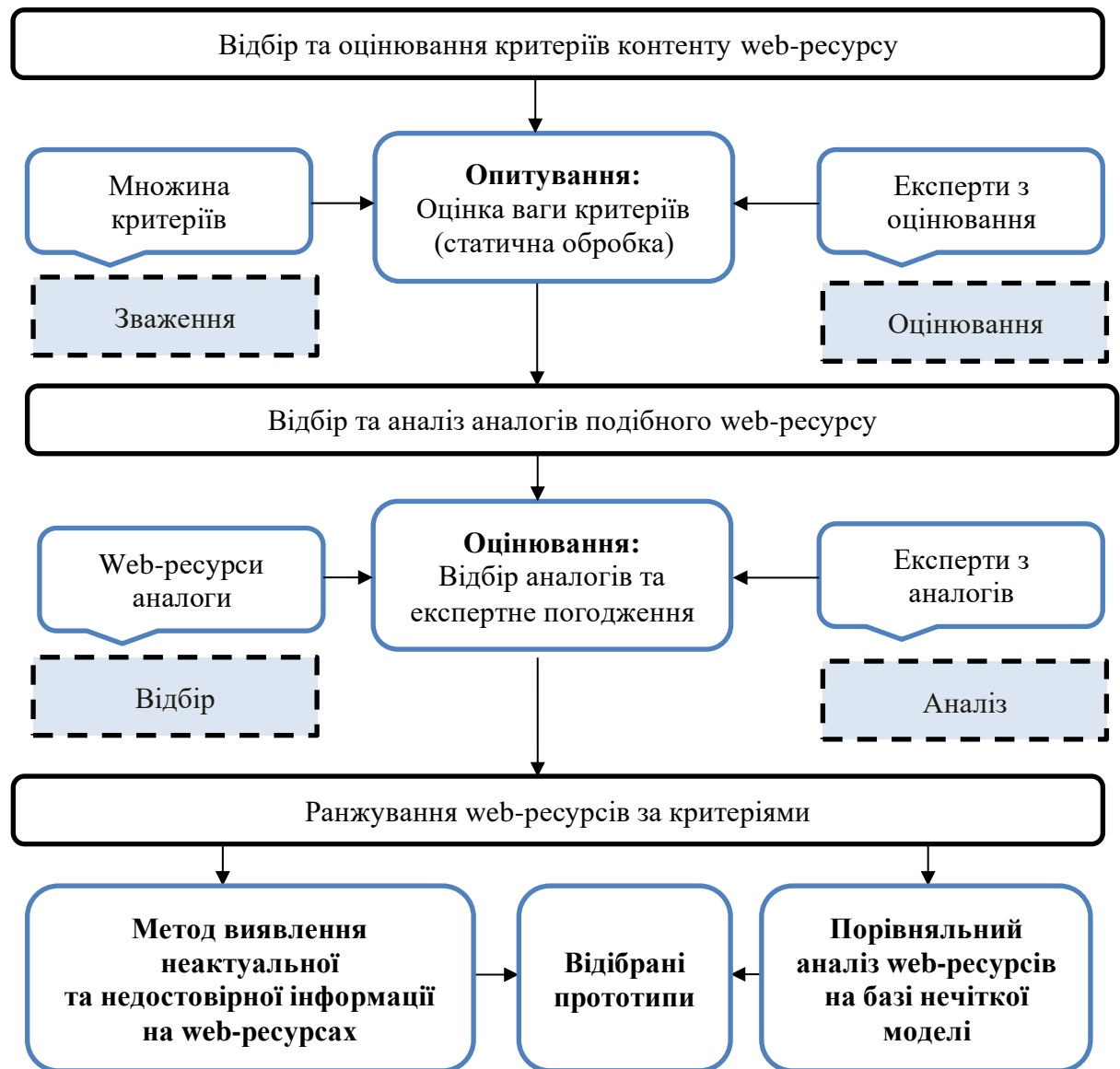
Розглядаються експертні парні порівняння всіх пар проектів за критеріями G_1 / G_n . При цьому використовуємо таблицю парних порівнянь за шкалою Сааті (табл. 2.1).

Таблиця 2.1.

Парні порівняння за шкалою Сааті

Коеф. шкали	Значення
1	якщо відсутня перевага елемента G_i над елементом G_j ;
3	якщо є слабка перевага G_i над G_j ;
5	якщо є істотна перевага G_i над G_j ;
7	якщо є явна перевага G_i над G_j ;
9	якщо є абсолютна перевага G_i над G_j ;
2,4,6,8	проміжні порівняльні оцінки.

Для проведення порівняльної оцінки значимості критеріїв оцінювання web-ресурсів доцільно розробити алгоритм порівняльного аналізу web-сайтів при організації його контенту (рис. 2.2).



* Модифіковано автором

Рис. 2.2. Алгоритм порівняльного аналізу web-сайтів при організації його контенту

Практичним втіленням експертного оцінювання може стати опитування експертів, яким пропонується визначити значимість всіх критеріїв згідно заданої шкали (наприклад, 100-бальної). Кількість експертів за цим напрямком може бути досить значною, щоб відобразити спектр оцінок різних фахівців та користувачів. В сучасних умовах зручним і поширеним інструментом такого опитування є Googl-форми.

Для визначення підсумкових оцінок значимості критеріїв можна застосувати статистичну обробку, зокрема оцінювання середньозважених

значень із урахуванням вагових коефіцієнтів експертів, які відображають їх кваліфікацію.

Ранжування експертів само по собі може бути непростою задачею. Зокрема тут можливий підхід, коли вся множина експертів розділяється на групи (наприклад, ті, хто має практичний досвід в цій області чи його не має). Ситуація може спроститись, якщо виявиться, що оцінки представників окремих груп різняться несуттєво. В цьому разі всі експерти одержують однакову вагу при врахуванні їх оцінок.

На другому етапі має бути виконаний відбір сайтів аналогів та їх попередній аналіз. Тут також маємо спиратись на оцінки експертів щодо відповідності конкретних аналогів кожному із критеріїв. В ході аналізу змісту web-ресурсів аналогів встановлюється їх типова структура та функціонал, зокрема тематика основних розділів та супутні функції щодо подання контенту.

Для відібраних аналогів виконується їх детальне оцінювання за всіма критеріями (зокрема за 100-бальною шкалою). Таке оцінювання є попереднім, оскільки воно не враховує вагу окремих критеріїв. Кінцеве оцінювання і відбір аналогів повинні виконуватись з використанням нечітких моделей на наступному кінцевому етапі.

Третім етапом для відібраної групи аналогів web-ресурсів проводиться ретельний порівняльний аналіз з урахуванням критеріїв оцінки контенту. Даний порівняльний аналіз буде виконаний за допомогою метода нечітких множин Беллмана-Заде. На цьому етапі алгоритму також застосуємо метод виявлення некоректного та недостовірного контенту web-ресурсів із використанням засобів машинного навчання, який розглянемо детально у наступному підрозділі. В кінцевому етапі роботи алгоритму експерти з аналогів виконують остаточний відбір прототипів, враховуючи як результати кількісного аналізу, так і супутні додаткові умови ефективності функціонування web-ресурсів.

2.2. Математична модель ефективності функціонування інформаційних web-ресурсів в умовах різної їх структурованості

Виходячи із припущень щодо оцінки ефективності виконання запитів інформаційними web-ресурсами, замість кількісної оцінки ефективності пропонується ввести деякі умовні значення оцінки ефективності, визначивши категорії, за якими вони можуть бути згруповані. Ці категорії формуються із часових витрат на виконання запитів щодо певного виду web-ресурсу.

Згрупуємо усі запити до web-ресурсів за категоріями [2]:

- σ_f - запити, що виконуються швидко;
- σ_m – запити, що виконуються із середньою швидкістю;
- σ_s – запити, що виконуються повільно.

Для кожної категорії модель спостережень можна подати у вигляді рівняння:

$$m_t = d_\sigma + e_\sigma \varphi_t \quad (2.11)$$

де m_t – середній час виконання запиту web-ресурсом за останній інтервал спостереження;

d_σ – середній очікуваний час виконання запиту web-ресурсом для поточної категорії σ запитів;

e_σ – значення, що визначає відхилення спостережуваних значень від середніх величин для поточної категорії запитів;

φ_t – випадкове збурення, що впливає на спостереження.

Для web-ресурсу S відомий показник r_t , що характеризує поточний режим його функціонування у певний момент часу t . Область значень R^1 показника r_t розіб'ємо на інтервали точками:

$$d: -\infty = d_0 < d_1 < \dots < d_{n-1} < d_n = +\infty,$$

де n – число режимів ефективності, підставивши у відповідність кожному режиму один інтервал. Режим, у якому знаходиться web-ресурс S , визначимо як приналежність r_t до цього інтервалу.

Вважаємо, що у будь-якому режимі середній час виконання запиту web-ресурсом за останній інтервал спостереження можна описати рівнянням (2.11), тобто, при зміні режиму m_t модель не змінюється, змінюються лише параметри. Останні можна замінити функціями $D(r_t)$ та $E(r_t)$, які повертають значення параметрів рівняння (2.11) залежно від поточного режиму, що визначається значенням показника r_t .

Виходячи з припущення, що залежність вимірюваних характеристик (час виконання запиту) для web-ресурсу S від його перебування в одному з режимів ефективності є лінійною, то можна запропонувати наступне рівняння для вимірювання у момент часу t характеристики:

$$m_t = D(r_t) + E(r_t)\varphi_t \quad (2.12)$$

Для $D(r_t)$ та $E(r_t)$ напишемо наступні вирази для їх зображення:

$$D(r_t) = \left\{ \begin{array}{l} D_1, -\infty < x < d_1 \\ D_2, d_1 \leq x < d_2 \\ \dots \\ D_n, d_{n-1} \leq x < +\infty \end{array} \right\} \quad (2.13)$$

$$E(r_t) = \left\{ \begin{array}{l} E_1, -\infty < x < d_1 \\ E_2, d_1 \leq x < d_2 \\ \dots \\ E_n, d_{n-1} \leq x < +\infty \end{array} \right\}$$

Значення D_1, D_2, \dots, D_n та E_1, E_2, \dots, E_n є відомими, виходячи із наведених вище припущень.

Для знаходження параметрів роботи web-ресурсу, що описуються відношенням (2.12) розбиваємо інтервал спостережень $[t_0; +\infty]$ на деякі рівні відрізки $t_0 < t_1 < t_2 \dots t_{k-1} < t_k < \dots$. На поточному інтервалі $[t_{k-1}; t_k]$ для кожного інформаційного ресурсу вимірюємо та зберігаємо час виконання відповідних запитів. У момент часу t_n формуємо із отриманих даних необхідну досліджену вибірку, що доповнюється середнім часом виконання запитів для кожного інформаційного ресурсу. Довжина часового інтервалу $[t_{k-1}; t_k]$ повинна визначатися достатнім часом, щоб затрати обчислювальних

ресурсів, які використовуються для оптимізації web-ресурсу, були меншими, ніж затрати ресурсів, що витрачаються на роботу всього web-ресурсу.

Для показника r_t повинна бути сформована модель, що описує закон, за яким ресурс S на поточному інтервалі спостереження $[t_{k-1}; t_k]$ має бути віднесений до тієї чи іншої категорії ефективності σ .

Зазначену модель, як і модель оцінки ефективності ресурсу, практично не можливо сконструювати виходячи з фізичного змісту процесів. Тому єдиним способом побудови таких моделей для стороннього спостерігача є застосування підходу на основі «чорної скриньки» із залученням методів параметричної ідентифікації [18].

Загальна рекомендація теорії ідентифікації моделей систем полягає в тому, щоб починати ідентифікацію моделі на основі її елементарних структур [5]. Тому спочатку для побудови моделі обираємо просту параметричну модель. Припустимо, що для ресурсу S показник r_t є випадковим процесом, який описується рівнянням авторегресії першого порядку [36]:

$$r_t = ar_{t-1} + s + b\psi_t, \quad (2.14)$$

де a , s , b - відомі невідомі числа (задані, або такі, що підпадають під подальшу ідентифікацію), ψ_t - послідовність незалежних, однаково розподілених випадкових величин, r_0 - випадкова величина, що не залежить від ψ_t .

Для знаходження невідомих коефіцієнтів a , s , b авторегресії (2.14), на основі відомих експериментальних значень тривалості виконання запиту ресурсу, у тому числі спостереження, використаємо середньоквадратичне відхилення між обчисленим значенням тривалості запиту та тим, що спостерігається, тобто:

$$\psi = \sum_{i=1}^{25} (ar_{t-1} + s + b\varphi_t - m_t)^2 \quad (2.15)$$

Тепер диференціюємо отриманий функціонал ψ (2.15) відносно невідомих коефіцієнтів a , s , b авторегресії і результати диференціювання квадратичної функції прирівняємо до нуля:

$$\begin{cases} \frac{\partial \psi}{\partial a} = 0 \\ \frac{\partial \psi}{\partial s} = 0 \\ \frac{\partial \psi}{\partial b} = 0 \end{cases} \quad (2.16)$$

Отримана система (2.16) є системою лінійних алгебричних рівнянь. Її розв'язками є значення коефіцієнтів a , s , b авторегресії.

В якості об'єктивної числової характеристики T_t ефективності джерела S на момент часу t можна розглядати абсолютне значення часу виконання запитів ресурсом на поточному інтервалі спостереження $[t_{k-1}; t_k]$. Використання абсолютного значення T_t означає, що на його величину можуть впливати тільки ті чинники невизначеності, які безпосередньо пов'язані із цим ресурсом (складність виконаного запиту, обсяг даних ресурсу, технічні характеристики апаратного забезпечення тощо).

Для унеможливлення впливу різномірності реальних джерел доцільно взяти відношення показника T_t до довжини інтервалу спостереження $[t_{k-1}; t_k]$, тобто перейти від абсолютного значення до відносного:

$$E_q = \frac{T_{t_k}}{t_k - t_{k-1}}, \quad (2.17)$$

де E_q – показник відносної ефективності.

Досліджуючи динаміку цього показника, можна вважати, що

$$r_{t_k} = r_{t_{k-1}} + \frac{T_{t_k}}{t_k - t_{k-1}} \quad (2.18)$$

Використавши для опису r_t модель, що представлена відношенням (2.12), залишається вибрати параметри, що забезпечують адекватність моделі спостереження до реально отриманих даних.

Наявність процесу r_t пов'язано з необхідністю моделювати поточний режим ефективності інформаційного web-ресурсу. Даний процес необхідний, щоб забезпечити послідовність зміни режимів, що відповідає спостережуваним даним. Зміна режиму відбувається при виході значень

процесу за межі інтервалу, визначеного для обраного режиму. У даному контексті величина та розмірність показника r_t не має значення.

В якості моделі, яка описує динаміку показника ефективності web-ресурсу, пропонується стохастична динамічна система спостережень Ковбасістого А.В. [1]:

$$\begin{cases} r_t = a_t r_{t-1} + s_t + b_t \psi_t, & t = 1, 2, \dots, \\ m_t = D(r_t) + E(r_t) \varphi_t \end{cases}, \quad (2.19)$$

де φ_t не залежить від ψ_t, r_0 .

Таким чином, є можливість обґрунтованого вибору параметрів розглянутої моделі за рахунок нескладного статистичного аналізу середовища функціонування вибраного web-ресурсу.

2.3. Розробка алгоритму виявлення некоректної та неактуальної інформації web-ресурсів

Задачі аналізу контенту і виявлення некоректної та застарілої інформації є достатньо складними. Для їх розв'язання потрібно створювати інтелектуальні системи із моделюванням контенту на основі використання онтологічного підходу. Проте на початкових етапах необхідно встановити певні особливості як існуючих алгоритмів аналізу контенту, так і майбутніх – націлених не тільки на звичайний контент, а й на виявлення, наприклад, неактуальної чи недостовірної інформації.

Система аналізу контенту повинна складатися з двох частин. Перша – встановлює семантичну структуру сайту, спираючись на конкретний код, наприклад на HTML, проводити семантичний аналіз (парсинг) і запис результатів у базу даних. Друга частина повинна із аналізу доступних web-ресурсів з подібним контентом проводити семантичний аналіз (парсинг) і сформувати аналогічну базу даних для зіставлення із даними бази, сформованої першою частиною. Функції першої системи опрацьовано в ряді

праць [5, 6, 19]. Зокрема, побудовано моделі представлення семантичної структури веб-сайтів за умови доступу до програмного коду сторінки.

Більшість web-сторінок створюють за допомогою мови HTML (англ. HyperText Markup Language – мова розмітки гіпертекстових документів) або XHTML. Браузер опрацьовує документ HTML та відтворює на екрані. Отже, усі web-сайти мають однакову HTML-розмітку, що дає можливість репродуктувати HTML-документ за допомогою браузера у звичному для користувача вигляді. Власне, контент розміщують у блоці, але також у цьому блоці теги можуть містити різні імена та послідовність. Останнє означає певні частини контенту, що суттєвим чином ускладнює його аналіз.

Наступною функцією системи є парсинг сайтів – послідовний синтаксичний аналіз інформації, розміщеної на інтернет-сторінках [17]. Текст інтернет-сторінок – ієрархічний набір даних, структурований за допомогою природних і комп'ютерних мов. Web-парсинг використовуємо для автоматизованого збору контенту або даних із будь-якого сайту чи сервісу. Результат парсингу записуємо у базу даних попередньо встановленої структури. Спираючись на аналіз існуючих систем контент-аналізу web-сайтів, можемо сформулювати основні вимоги до алгоритмів автоматизованого виявлення застарілої та недостовірної інформації.

Серед переліку вимог: налаштування фільтрів URL, щоб не парсити зайві сторінки, налаштування глибини парсингу, якісне скачування контенту, обробка у декілька потоків та збереження контенту у різних форматах (Рис.2.3).

Як бачимо з рис. 2.3, після здійснення парсингу згенеровані результати необхідно опрацьовувати, щоб надати інформації вигляд, придатний для подальшого використання. Конкретний формат залежить від того, як у подальшому будуть оброблятися зібрані дані. Доволі часто з отриманого контенту за допомогою XML формується RSS-потік, що зручний для використання даних без процедури ререйтингу. Іноді результат парсингу розміщують у CSV-файл, оскільки цей текстовий формат дуже простий у

подальшій обробці, легко конвертується у SQL-запити і без проблем відкривається в Excel. В особливих випадках потрібно, щоб кінцеві дані були представлені у вигляді електронних таблиць XLS. Складнішим завданням, є реалізація функцій другої частини системи виявлення некоректної та неактуальної інформації на основі контент-аналізу web-ресурсів. У нашій кваліфікаційній роботі для реалізації системи запропоновано використати порівняння результатів парсингу із відомою базою організації – власника інформаційного web-ресурсу.

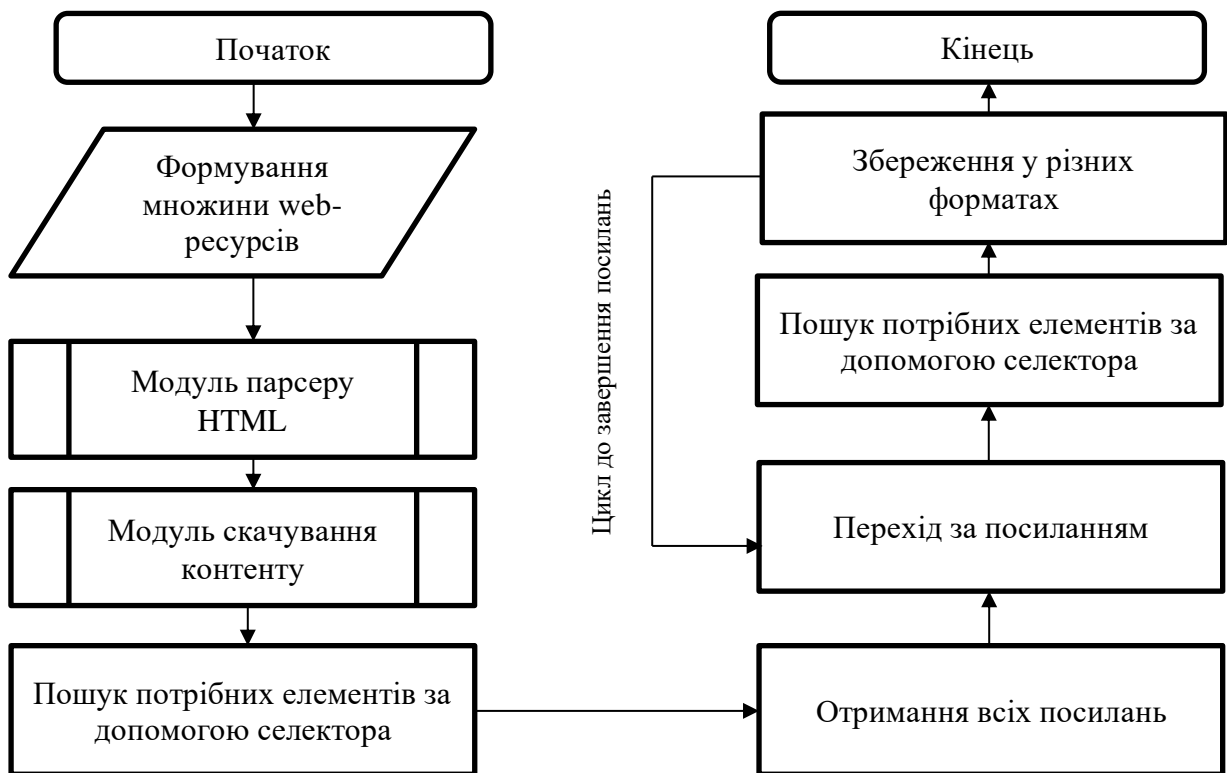
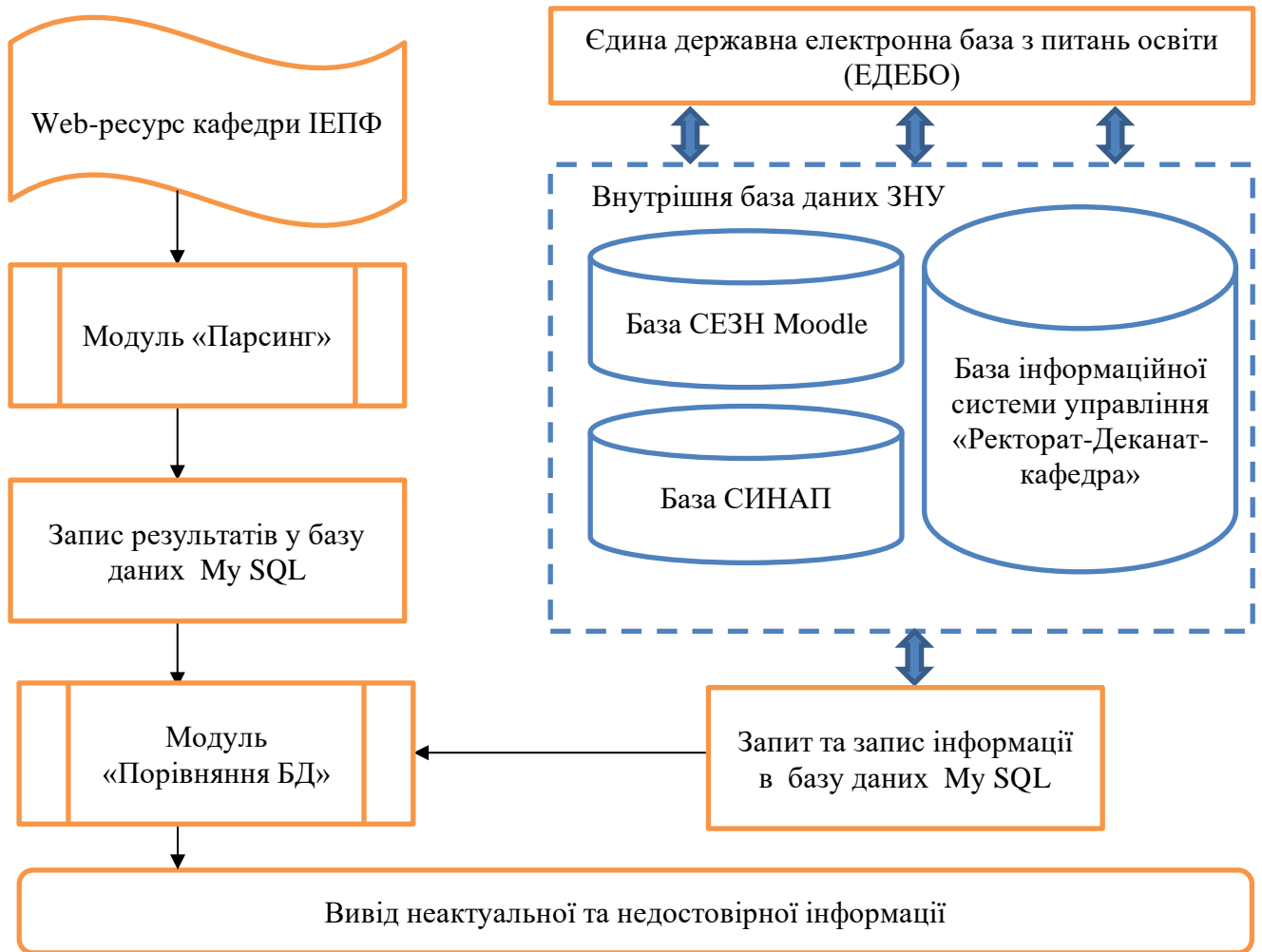


Рис. 2.3. Алгоритм збору інформації за допомогою парсингу та збереження інформації web-ресурсів

* модифіковано автором за [17]

Загальний алгоритм реалізації процедури виявлення некоректної та застарілої інформації на web-ресурсах на прикладі кафедри інформаційної економіки, підприємництва та фінансів і порівняння із існуючою базою даних Запорізького національного університету наведена на рис. 2.4.



* Розроблено автором

Рис. 2.4. Алгоритм процедури виявлення некоректної та застарілої інформації на web-ресурсі кафедри ІЕПФ ЗНУ

Як бачимо із алгоритму на рис. 2.4., процес підготовки контенту до процедури виявлення неактуальної та недостовірної інформації на web-ресурсах кафедри передбачає виконання таких дій:

1. Збір контенту, найчастіше шляхом скачування web-сторінок.
2. Отримані дані HTML-формату трансформуються у заданий формат.
3. Перетворення отриманих результатів у зручний для користувача та доступний файл, наприклад, формування бази даних MySQL.

Після отримання результатів та запису у базу даних потрібно їх синхронізувати із еталонною та актуальною інформацією з бази Запорізького національного університету.

Для початку синхронізації даних у БД варто отримати перелік усіх елементів, які співставляються. Для цього необхідно створити запит, що порівнює стовпці (імена) у таблицях між двома базами даних (схемами) MySQL.

```
set @database_kafedra = 'real_object';
```

```
set @database_universitet = 'rezult';
```

Основним застосуванням команди SELECT мови SQL є перегляд рядків з таблиці. Нижче наведено листинг 2.1 команди SELECT, що відобразить усі рядки з таблиці:

Лістинг 2.1 – Фрагмент SQL коду.

```
select *
from (select COALESCE(c1.table_ rezult, c2.table_real_object) as table_name,
COALESCE(c1.column_name, c2.column_name) as table_column, c1.column_name as
database_kafedra , c2.column_name as database_universitet
from (select table_ rezult, column_name from information_schema.columns c where
c.table_schema = @database_kafedra) c1
right join (select table_real_object, column_name from information_schema.columns c
where c.table_schema = @database_universitet) c2 on c1.table_name = c2.table_name and
c1.column_name = c2.column_name
union
```

Функція COALESCE приймає ряд аргументів і повертає перший аргумент, що не має значення NULL. Якщо всі аргументи NULL, функція COALESCE повертає NULL.

У результаті застосування подібних запитів при порівнянні двох БД можливо встановити подібність структурованої інформації на різних web-ресурсах. Це означає, що якщо на 2 ресурсах було встановлено подібний контент, який стосується деякого об'єкту, і основна БД університету надала інформацію, яка співпадає, тоді приймаємо рішення, що контент коректний та актуальний.

Висновки по розділу 2.

У другому розділі кваліфікаційної роботи виконаний огляд математичних методів порівняльного аналізу для використання в задачі раціональної організації контенту web-ресурсів. Підтверджено доцільність застосування підходу із використанням нечітких множин, надано розгорнуту характеристику метода нечітких множин Беллмана-Заде з прийняття рішень для аналізу аналогів web-ресурсів. На підставі виконаного аналізу запропоновано методику, яка може бути задіяна на стадії порівняння web-сайтів аналогів.

Сформульовано основні вимоги до засобів автоматичного аналізу контенту, основним завданням яких є виявлення застарілої та некоректної інформації. Встановлено, що засоби повинні бути адаптовані на типові структури web-сторінок, а також на відомі структури опису web-сторінки мовою розмітки HTML.

Запропоновано та обґрунтовано підхід до моделювання базових показників оцінки ефективності функціонування інформаційних web-ресурсів. Він дозволяє врахувати характерні для нелінійних систем особливості, такі як: залежність збурень від поточних значень показника, циклічність процесів, стрибкоподібні зміни характеристик.

Запропоновано та обґрунтовано алгоритм збору інформації за допомогою парсингу та збереження інформації web-ресурсів. Основна ідея побудови алгоритму полягає у перетворенні інформації із інформаційних web-ресурсів у формат, придатний для порівняння з базами даних, що існують в організаціях або виявлені на інших web-ресурсах. Розроблено алгоритм процедури виявлення некоректної та застарілої інформації на web-ресурсі кафедри ІЕПФ ЗНУ.

РОЗДІЛ 3
АНАЛІЗ ТА ОЦІНКА КОНТЕНТУ WEB-РЕСУРСУ КАФЕДРИ
ІНФОРМАЦІЙНОЇ ЕКОНОМІКИ, ПІДПРИЄМНИЦТВА ТА ФІНАНСІВ
ЗНУ

3.1. Порівняльний аналіз контенту web-ресурсів структурних підрозділів університетів

Кваліфікаційна робота виконувалася на кафедрі інформаційної економіки, підприємництва та фінансів ЗНУ, де існує потреба в оцінці контенту web-сайту кафедри, який розташований за посиланням <https://znuiepf.com.ua/>. Надалі результати аналізу будуть використовуватись для редизайну та оптимізації контенту web-ресурсу кафедри згідно із завдання на кваліфікаційну роботу.

На першому етапі аналізу виконаний відбір web-ресурсів конкуруючих кафедр інших університетів, які випускають студентів за освітньою програмою 051 Економіка, та їх попереднє експертне оцінювання автором. За результатами такого оцінювання обмежена множина аналогів буде відібрана для подальшого детального порівняння з використанням методу нечіткого оцінювання. Усі розглянуті web-ресурси наведені у таблиці 3.1.

Таблиця 3.1.

Аналіз web-ресурсів конкурентних кафедр

№	Назва закладу	Кафедра	Посилання	Змістовна наповненість	Візуальне відображення	Підлягання більш ретельному аналізу
1	2	3	4	5	6	7
1	Запорізький національний університет	Кафедра економічної кібернетики	https://econom.znu.edu.ua/кафедра-економічної-кібернетики/	Мала	Погане	Ні
2	Запорізький інститут економіки та інформаційних технологій	Кафедра інформаційних технологій	https://www.zieit.edu.ua/?page_id=2226	Мала	Середнє	Ні

Продовження табл. 3.1.

1	2	3	4	5	6	7
3	Класичний приватний університет	Кафедра економіки	http://virtuni.education.zp.ua/info_cpu/node/580	Мала	Середнє	Ні
4	Університет імені Альфреда Нобеля	Кафедра інформаційних технологій	https://duan.edu.ua/university-ukr/kafedry/15-pages/325-kafedra-informatsiinykh-tekhnologii.html	Середня	Високе	Так
5	НТУУ "Київський політехнічний інститут імені Ігоря Сікорського"	Кафедра економічної кібернетики	https://ecocyber.fmm.kpi.ua/	Висока	Середнє	Так
6	Львівський національний університет імені Івана Франка	Кафедра економічної кібернетики	https://cyber.lviv.ua/	Висока	Середнє	Так
7	ХНЕУ імені Семена Кузнеця	Кафедра економічної кібернетики	https://ek.hneu.edu.ua/	Висока	Середнє	Так
8	Національний ТУ «Дніпровська політехніка»	Кафедра "Економіка та економічна кібернетика"	https://ekit.nmu.org.ua/ua/	Середня	Середнє	Ні
9	Національний університет "Полтавська політехніка імені Юрія Кондратюка"	Кафедра економічної теорії та економічної кібернетики	https://nupp.edu.ua/page/kafedra-yekonomichnoi-teorii-ta-yekonomichnoi-kibernetiki.html	Середня	Середнє	Ні
10	Харківський національний університет радіоелектроніки	Кафедра економічної кібернетики та управління економічною безпекою	https://nure.ua/department/kafedra-ekonomichnoyi-kibernetiki-ta-upravlinnya-ekonomichnoyu-bezpekoyu-ek	Мала	Середнє	Ні
11	Західноукраїнський національний університет	Кафедра економічної кібернетики та інформатики	https://eki.wunu.edu.ua/	Висока	Високо	Так
12	Черкаський Державний Технологічний Університет	Кафедра економічної кібернетики та маркетингу	https://ekm.chdtu.edu.ua/	Висока	Високе	Так
13	Національний університет біоресурсів і природокористування України	Кафедра економічної кібернетики	https://nubip.edu.ua/node/2971	Середнє	Середнє	Ні
14	Харківський національний університет ім.В.Н.Каразіна	Кафедра економічної кібернетики та прикладної економіки	http://www.cyber.kharkov.ua/	Середнє	Високе	Так
15	КНЕУ	Інститут інформаційних технологій в економіці	https://fisit.kneu.edu.ua/	Середнє	Середнє	Ні

Спираючись на данні таблиці 3.1 для подальшого аналізу методом нечітких множин Беллмана-Заде було обрано 7 найбільш вдалих аналогів web-ресурсів наведених у таблиці 3.2.

Таблиця 3.2

Відібрані web-ресурси структурних підрозділів університетів

№ Аналогу	Назва закладу	Кафедра	Посилання
1	Університет імені Альфреда Нобеля	Кафедра інформаційних технологій	https://duan.edu.ua/university-ukr/kafedry/15-pages/325-kafedra-informatsiinykh-tekhnologii.html
2	НТУУ "Київський політехнічний інститут імені Ігоря Сікорського"	Кафедра економічної кібернетики	https://ecocyber.fmm.kpi.ua/
3	Львівський національний університет імені Івана Франка	Кафедра економічної кібернетики	https://cyber.lviv.ua/
4	ХНЕУ імені Семена Кузнеця	Кафедра економічної кібернетики	https://ek.hneu.edu.ua/
5	Західноукраїнський національний університет	Кафедра економічної кібернетики та інформатики	https://eki.wunu.edu.ua/
6	Черкаський Державний Технологічний Університет	Кафедра економічної кібернетики та маркетингу	https://ekm.chdtu.edu.ua/
7	Харківський національний університет ім.В.Н.Каразіна	Кафедра економічної кібернетики та прикладної економіки	http://www.cyber.kharkov.ua/

Далі отримані аналоги проаналізуємо методом нечітких множин Беллмана-Заде з урахуванням критеріїв оцінювання web-ресурсів. На основі проведеного аналізу [21, 29, 42] нами було відібрано 9 критеріїв (К1 — К9), які представлені в таблиці 3.3.

Таблиця 3.3

Критерії для аналізу web-ресурсів за методом Беллмана-Заде

Критерій	Назва критерію
К1	Ієрархія змісту
К2	Візуальна ієрархія
К3	Мапа сайту
К4	Адаптивність та кросбраузерність
К5	Гармонійна кольорова гамма
К6	Швидкість завантаження контенту
К7	Відповідність дизайну та змісту
К8	Креативність подачі
К9	Врахування тенденцій моди

Виходячи із складу відібраних web-ресурсів та критеріїв виконаємо послідовність дій з порівняльного аналізу за схемою Беллмана-Заде.

1) Формуємо 9 матриць парних порівнянь аналогів за критеріями (7x7) та 1 матрицю парних порівнянь критеріїв (матриця α - 9x9).

2) Для матриць 7x7 знаходяться ціни альтернатив – середні геометричні рядків матриці. Елементи строки перемножуються та з їх добутків знаходиться корінь 7-ого ступеня.

3) Знаходиться сума цін web-ресурсів та їх відносна вага. Загальна сума повинна дорівнювати 1.

4) За одним критерієм, найбільш переважною, на думку експерта, є альтернатива, що має максимальну вагу.

5) Перевірка експертних оцінок на несуперечливість:

- Знаходяться суми стовбців матриці парних порівнянь.
- Розраховується допоміжна величина L шляхом додавання добутків сум стовбців матриці на ваги альтернатив.

- Знаходиться величина, яка називається індексом узгодженості (I_y):

$$I_y = (L-X)/(X-1) \quad (3.1)$$

де X – розрядність матриці.

- В залежності від розмірності матриці парних порівнянь знаходиться величина випадкового узгодження. Значення випадкового узгодження наведені в таблиці 3.4.

Таблиці 3.4

Значення випадкового узгодження

Розмірність матриці	3	4	5	6	7	8	9	10
Випадкове узгодження	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49

На підставі значення випадкового узгодження для матриць парних порівнянь web-ресурсів (7x7) дорівнює 1,32, а для матриці парних порівнянь критеріїв (9x9) дорівнює 1,45.

- Знаходиться відношення узгодженості (B_y):

$$B_y = I_y / \text{Випадкове узгодження} \quad (3.2)$$

Якщо значення B_U перевищує 0,2, то потрібне уточнення матриці парних порівнянь.

Для оцінювання відносної важливості критеріїв організації контенту було виконано опитування множини експертів за допомогою розробленої форми на сайті кафедри ІЕПФ ІННУ ім. Ю.М. Потебні ЗНУ за посиланням https://znuiepf.com.ua/?page_id=5512 (рис.3.1).

Критерії оцінки контенту web-ресурсів

Оберіть web-ресурс для оцінювання

Університет імені Альфреда Нобеля

Ієрархія змісту

Критерій	1	2	3	4	5	6	7	8	9	Оцінка
Низько	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Візуальна ієрархія	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Мапа сайту	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Адаптивність та кросбраузерність	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Гармонійна кольорова гамма	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Швидкість завантаження контенту	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Відповідність дизайну та змісту	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Креативність подачі	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Врахування тенденцій моди	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо
Низько	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Дуже високо

Відправити

Листопад 2022

Пн	Вт	Ср	Чт	Пт	Сб	Нд
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

« Жов

Рис. 3.1. Форма експертного опитування з оцінки контенту web-ресурсів

Відповіді експертів автоматично зберігаються у Формам сайту WordPress та електронних таблицях Google.

За результатами опитування виконаємо розрахунки та формування матриць.

В якості прикладу наведемо формування матриці парних порівнянь web-ресурсів (7x7) й розрахунок коефіцієнтів відносної важливості – ієрархії змісту (K1) – рис.3.2.

	A	B	C	D	E	F	G	H	I
1		w-r1	w-r2	w-r3	w-r4	w-r5	w-r6	w-r7	Ціна web-ресурсу
2	w-r1	1	0,5	0,333333	0,5	0,333333	0,166667	1	0,463987849
3	w-r2	2	1	3,333333	5	3,333333	2,857143	2,5	2,5955702
4	w-r3	3	0,3	1	1,666667	2,222222	1	2	1,311297547
5	w-r4	2	0,2	0,6	1	1,25	2	4	1,133224334
6	w-r5	3	0,3	0,45	0,8	1	3,333333	1,666667	1,087595747
7	w-r6	6	0,35	1	0,5	0,3	1	2,5	0,966448356
8	w-r7	1	0,4	0,5	0,25	0,6	0,4	1	0,531615116
9	Сумма	18	3,05	7,216667	9,716667	9,038889	10,75714	14,66667	8,089739148
10									
11									
12									Відносна вага web-ресурсу
13	X=	7							w-r1 0,06
14	L=	8,00601							w-r2 0,32
15									w-r3 0,16
16	ly	0,167668							w-r4 0,14
17									w-r5 0,13
18	By	0,127021							w-r6 0,12
19									w-r7 0,07
20									Сума 1
21									
22									
23									

Рис. 3.2. Формування та розрахунки матриці парних порівнянь за критерієм оцінки контенту – ієрархія змісту

Оскільки показник $B_y=0,16$ і не перевищує $0,2$, то відношення вважається узгодженим також на рисунку 3.3 наведена діаграма, на якій зображено порівняння web-ресурсів за наведеним критерієм (ієрархія змісту).

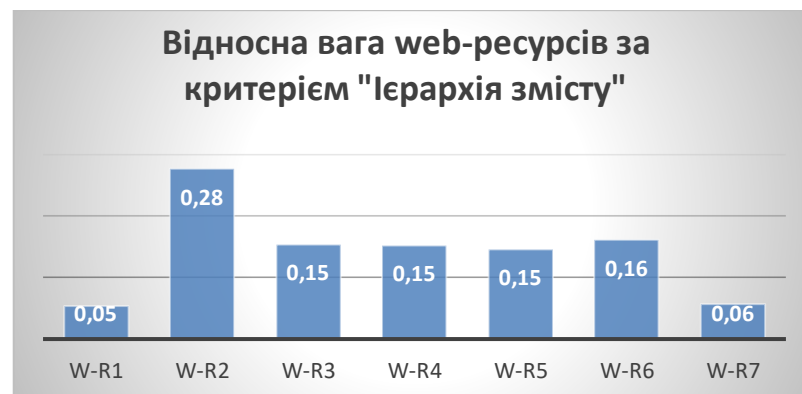


Рис. 3.3. Відносна вага альтернативних web-ресурсів ієрархії змісту

З наведеної на рисунку 3.3 діаграми ми бачимо, що за критерієм ієрархії змісту лідируючу позицію займає другий web-ресурс, а саме сайт кафедри Економічної кібернетики НТУУ "Київський політехнічний інститут імені Ігоря Сікорського".

Так як при розрахунках у всіх критеріях значення B_y не перевищувало $0,2$, надалі обмежимо аналіз підсумкових діаграм для інших критеріїв.

На рисунку 3.4 наведені діаграми, на яких зображено порівняння web-ресурсів за іншими критеріями оцінювання контенту.

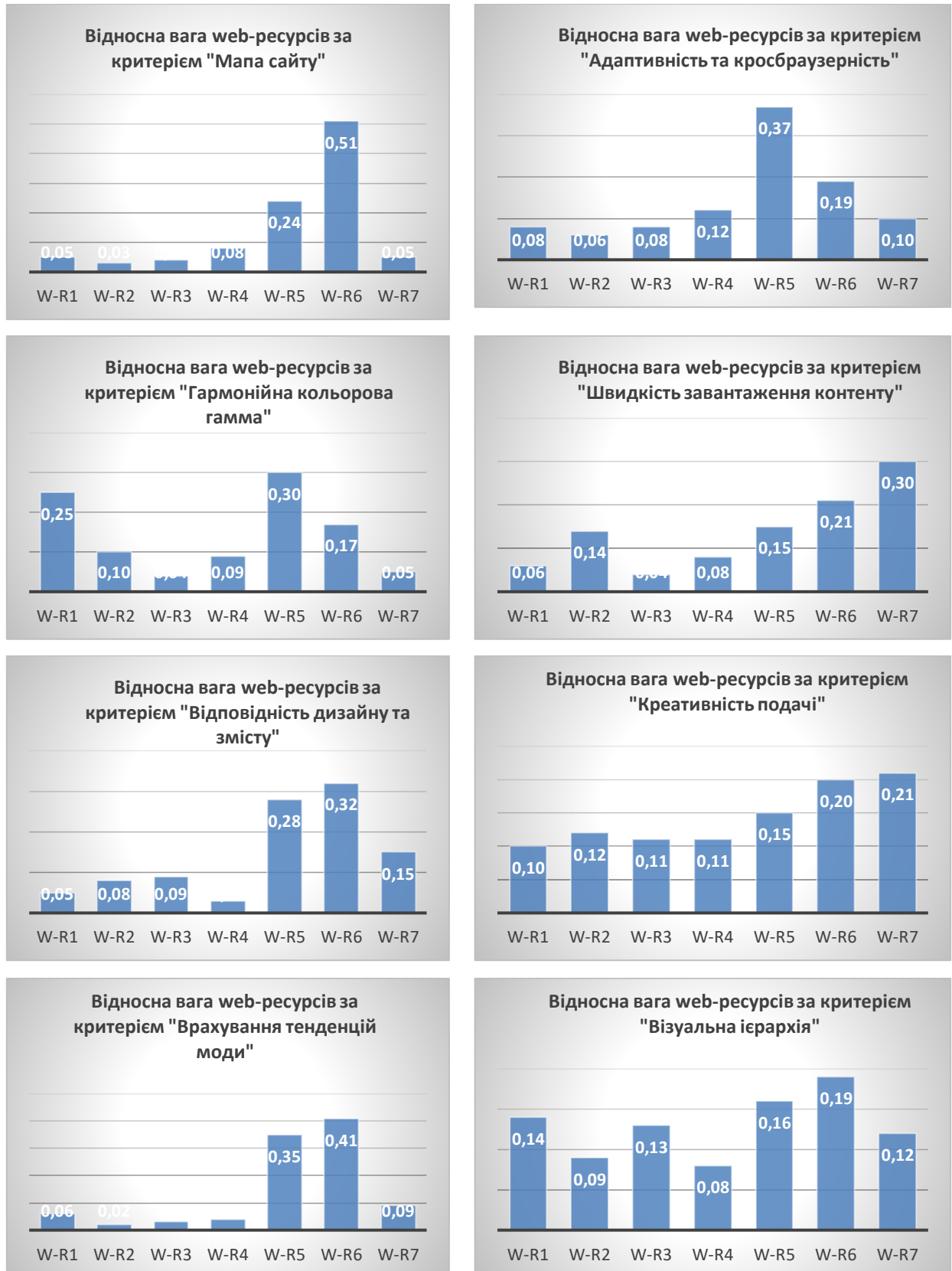


Рис.3.4 Відносна вага альтернативних web-ресурсів за критеріями оцінювання контенту

Таким чином на базі аналізу 9 матриць парних порівнянь аналогів були визначені web-ресурси, які є кращими за відповідними критеріями.

Далі як зображено на рисунку 3.5 застосуємо 4 етап до матриці парних порівнянь критеріїв розміщення контенту на web-ресурсах (9x9) та знайдемо вагу коефіцієнтів відносної важливості ($\alpha_1 - \alpha_9$) для K_1, K_2, \dots, K_9 .

	A	B	C	D	E	F	G	H	I	J	K	L	M
1											Ціна критерію		
2	K1	1	0,5	0,333333	0,5	0,333333	0,166667	1	0,5	2	0,550321208		
3	K2	2	1	3,333333	0,166667	3,333333	0,5	2,5	2,5	1,1111	1,328156663		
4	K3	3	0,3	1	0,142857	2,222222	0,333333	2	2,5	0,5	0,852609326		
5	K4	2	6	7	1	1,25	0,5	4	3,3333	1,4286	2,15443469		
6	K5	3	0,3	0,45	0,8	1	0,2	1,666667	0,3333	1,25	0,708527883		
7	K6	6	2	3	2	5	1	2,5	3,3333	2	2,629024198		
8	K7	1	0,4	0,5	0,25	0,6	0,4	1	0,5	3,3333	0,647478803		
9	K8	2	0,4	0,4	0,3	3	0,3	2	1	5	0,983888697		
10	K9	0,5	0,9	2	0,7	0,8	0,5	0,3	0,2	1	0,627665094		
11	Сумма	20,5	11,8	18,01667	5,859524	17,53889	3,9	16,96667	14,2	17,623	10,48210656		
12													
13		X=	7										
14		L=	8,45293										
15													
16		ly	0,242155										
17													
18		By	0,183451										
19													
20													
21													
22													
23													
24													
25													
26													

а	Відносна вага критерію
K1	0,05
K2	0,13
K3	0,08
K4	0,21
K5	0,07
K6	0,25
K7	0,06
K8	0,09
K9	0,06
Сума	1,00

Рис. 3.5. Формування та розрахунки матриці парних порівнянь α для критеріїв розміщення контенту на web-ресурсах університетів

Заповнення матриці виконувалося на підставі відредагованих даних отриманих при проведенні анкетування, на основі цих даних було побудовано діаграму важливості критеріїв зображену на рисунку 3.6

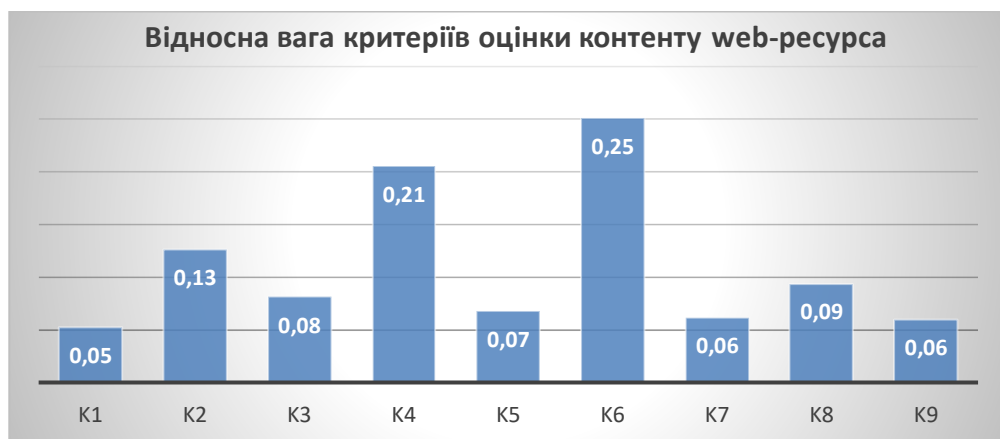


Рис. 3.6. Розрахунок ваг критеріїв оцінки контенту web-ресурсів

Як бачимо з отриманої діаграми коефіцієнтів відносної важливості найбільшу важливість при прийнятті рішень має критерій – Швидкість завантаження контенту (K6), на другому місці – Адаптивність та кросбраузерність (K4), третє місце – Візуальна ієрархія (K2), четверте місце – Креативність подачі (K8), п'яте місце – Мапа сайту (K3), шосте місце – Гармонійна кольорова гамма (K5), сьоме місце та восьме місце поділяють – Відповідність дизайну та змісту (K7) та Врахування тенденцій моди (K9), та дев'яте місце – Ієрархія змісту (K1).

Для отримання результату ранжирування аналогів web-ресурсів було проведено розрахунок експертних порівнянь усіх аналогів, який наведено на рис. 3.7.

	A	B	C	D	E	F	G	H	I	J	K	L
1	a	Відносна вага критерію			w-r1	w-r2	w-r3	w-r4	w-r5	w-r6	w-r7	
2	K1	0,05		K1	0,06	0,32	0,16	0,14	0,13	0,12	0,07	
3	K2	0,13		K2	0,14	0,09	0,13	0,08	0,16	0,19	0,12	
4	K3	0,08		K3	0,05	0,03	0,04	0,08	0,24	0,51	0,05	
5	K4	0,21		K4	0,08	0,06	0,08	0,12	0,37	0,19	0,10	
6	K5	0,07		K5	0,25	0,10	0,04	0,09	0,30	0,17	0,05	
7	K6	0,25		K6	0,06	0,14	0,04	0,08	0,15	0,21	0,30	
8	K7	0,06		K7	0,05	0,08	0,09	0,03	0,28	0,32	0,15	
9	K8	0,09		K8	0,10	0,12	0,11	0,11	0,15	0,20	0,21	
10	K9	0,06		K9	0,06	0,02	0,03	0,04	0,35	0,41	0,09	
11	Сума	1										
12												
13					w-r1	w-r2	w-r3	w-r4	w-r5	w-r6	w-r7	
14				K1	0,86	0,94	0,91	0,90	0,90	0,89	0,87	
15				K2	0,78	0,74	0,77	0,73	0,79	0,81	0,76	
16				K3	0,78	0,75	0,77	0,81	0,89	0,95	0,78	
17				K4	0,60	0,56	0,60	0,65	0,82	0,71	0,62	
18				K5	0,91	0,86	0,80	0,85	0,92	0,89	0,82	
19				K6	0,49	0,61	0,45	0,53	0,62	0,68	0,74	
20				K7	0,83	0,86	0,86	0,81	0,92	0,93	0,89	
21				K8	0,81	0,82	0,81	0,81	0,84	0,86	0,86	
22				K9	0,84	0,79	0,81	0,82	0,94	0,95	0,87	
23				min=	0,49	0,56	0,45	0,53	0,62	0,68	0,62	
24												
25												

Рис. 3.7. Розрахунок показника ступені приналежності web-ресурсів

На підставі отриманих розрахунків було побудовано діаграму порівняння web-ресурсів з урахуванням важливості критеріїв розміщення контенту на сторінках web-сайтів, яка зображена на рисунку 3.8.

З отриманої діаграми можемо зробити такі висновки:

Найкращими аналогами для подальшого розгляду будуть аналоги 6 та 5 так як вони отримали однакову та лідируючу відмітку серед усіх інших

аналогів. Шостий аналог – веб-сайт кафедри «Комп’ютерних систем та мереж» Чернівецького національного університету буде прийнятий для ретельного аналізу.

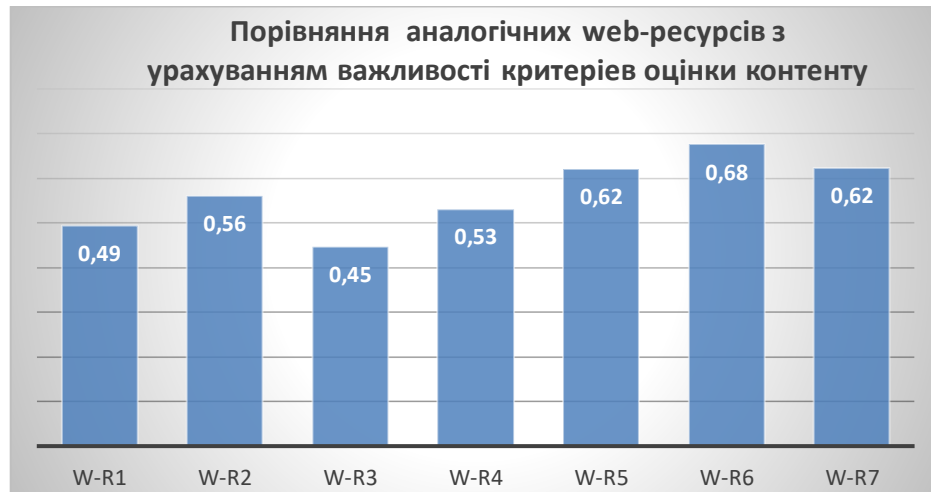


Рис. 3.8. Результати оцінки контенту web-ресурсів університетів

Таким чином для подальшого аналізу та оцінки контенту конкуруючих web-ресурсів університетів рекомендовано розглядати перш за все Черкаський державний технологічний університет, який отримав оцінку 0,68. Також потребують ретельного аналізу web-ресурси Харківського національного університету ім. В.Н. Каразіна та Західноукраїнський національний університет (мають однакову оцінку 0,62).

3.2. Аналіз переваг та недоліків в організації контенту web-ресурсу кафедри ІЕПФ ЗНУ

На підставі більш ретельного аналізу методом нечітких множин з сімох розглянутих аналогів було обрано web-ресурс кафедри економічної кібернетики та маркетингу Черкаського державного технологічного університету та проведено детальний аналіз у порівнянні з web-ресурсом кафедри інформаційної економіки, підприємництва та фінансів ЗНУ з вказівками їх переваг та недоліків.

Розглянемо організацію контенту web-ресурсу кафедри економічної кібернетики та маркетингу Черкаського державного технологічного університету – <https://ekm.chdtu.edu.ua/>.

Основними перевагами в організації контенту зазначеного web-ресурсу є:

1) Головна сторінка (рис. 3.9) виконана за схемою візуалізації Z, що візуально підходить для неї, бо вона не містить великої кількості тексту.

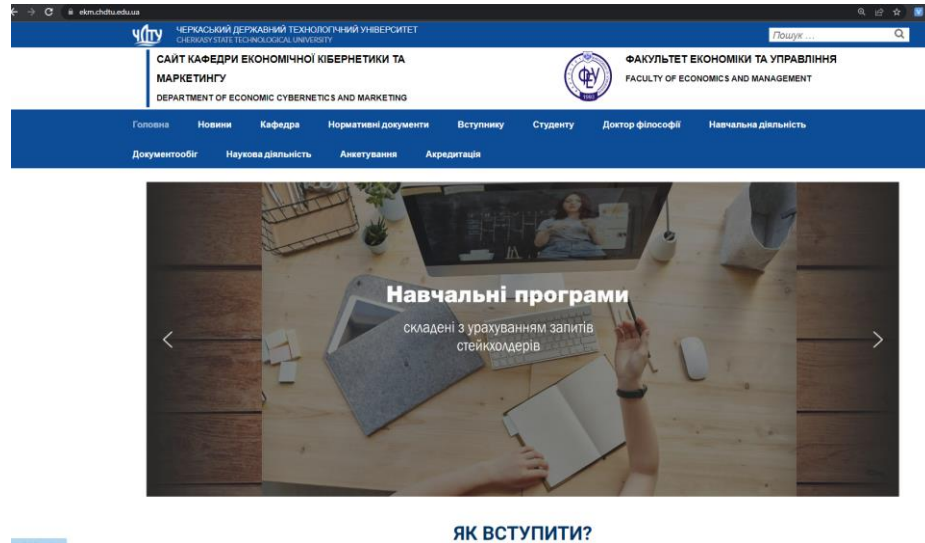


Рис. 3.9. Головна сторінка web-сайту кафедри економічної кібернетики та маркетингу ЧДТУ

2) Сторінки які містять текстову наповненість (рис. 3.10) виконані за схемою візуалізації F, для покращення та прискорення пошуку потрібної інформації використовується сайдбар, з більш детальною навігацією.

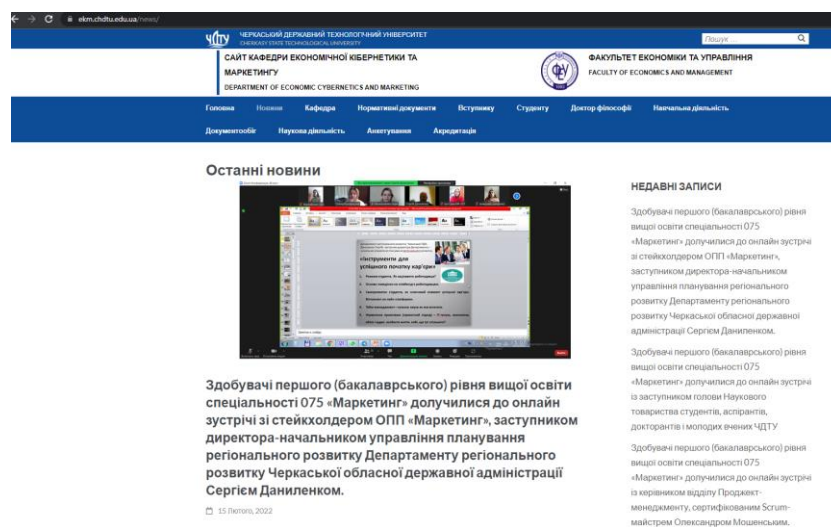
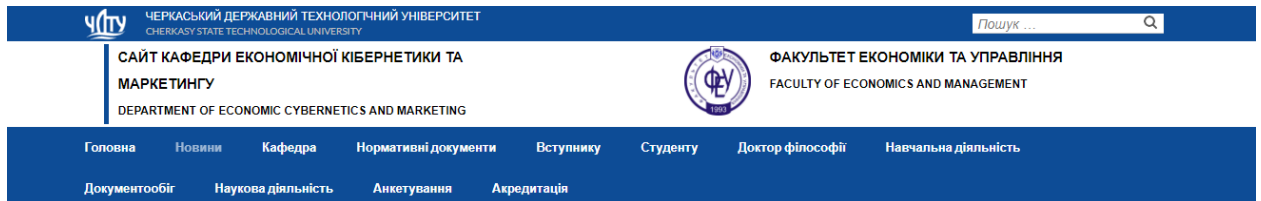


Рис. 3.10. Структура контенту web-сайту кафедри економічної кібернетики та маркетингу ЧДТУ

3) Присутня чітка візуальна ієрархія з розбиттям контенту на окремі блоки, що несуть різну за змістом та значенням інформацію.

- 4) Дуже вдало продумана та побудована структурна ієрархія.
- 5) Оформлення посилань та тексту відповідає загально прийнятим стандартам.
- 6) Змістовна інформація супроводжується візуальною.
- 7) Хедер та футер оформлені гармонійно та без перенавантажень (рис.3.11).

Хедер



Футер

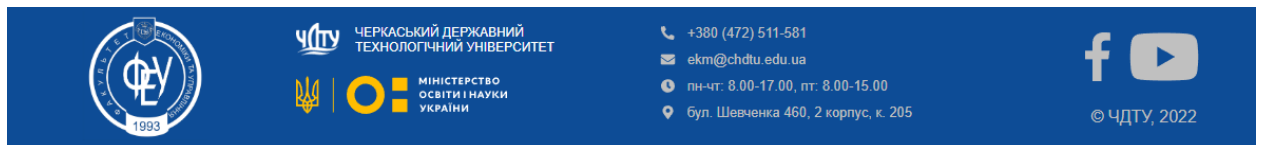


Рис.3.11. Оформлення хедеру та футеру web-сайту

- 8) Використовується однорідна структура.
- 9) На сторінці контактів присутня велика кількість корисної інформації.
- 10) Гармонійно підібрана кольорова гамма.
- 11) Web-ресурс має адаптивний дизайн, сторінка змінює відображення в залежності від розширення екрану та має мобільну версію.

Серед недоліків слід відмітити наступне:

- 1) Присутнє перенавантаження сторінки, а саме перенавантаження сайдбару, він занадто великий (довгий), як рішення цієї проблеми можна зробити заголовки з випадаючим меню, як це реалізовано у хедері.
- 2) Меню web-ресурсу розташоване не на одному рядку.
- 3) Оформлення більшої частини сторінок виконане у застарілій манері, не використовуючи нинішні тенденції моди.
- 4) На сторінці «Контакти» присутня форма, яка не несе ніякої функціональної та змістовної наповненості.

5) Використовується не зовсім гармонійний шрифт для основного тексту.

6) Немає можливості змінювати мову web-сайту.

7) Присутні ознаки застарілої та неактуальної інформації.

Далі розглянемо переваги та недоліки web-сайту кафедри інформаційної економіки, підприємництва та фінансів ЗНУ.

Переваги:

1) Гармонійне використання білого синього кольорів в оформленні web-сторінок.

2) Присутня чітка візуальна ієрархія з розбиттям контенту на окремі блоки, що несуть різну за змістом та значенням інформацію.

3) Хедер (зображений на рис. 3.12) оформлений гармонійно та без перенавантажень.

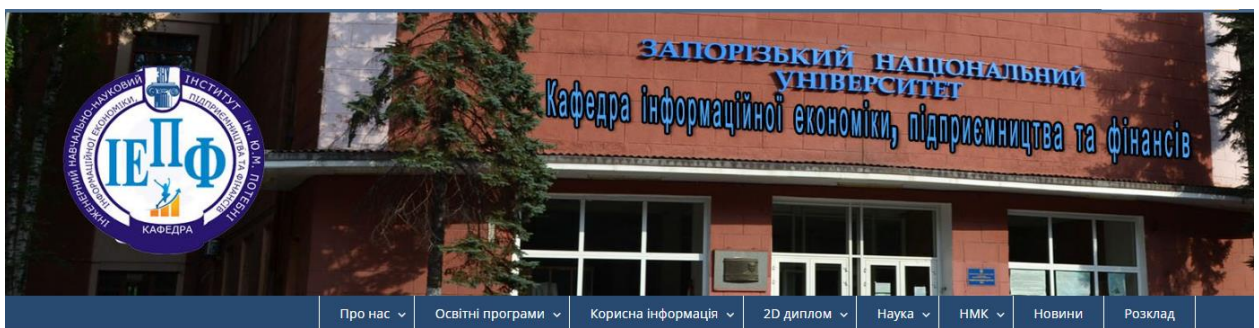


Рис. 3.12. Хедер web-сайту кафедри

4) Використовується однорідна структура.

5) Використовується невелика кількість шрифтів та присутнє їх гармонійне поєднання.

6) Web-сайт має адаптованість під різні розширення екрану та підтримку мобільної версії.

7) Акцентування уваги до контенту проводиться завдяки виділення елементів розміром.

8) Оформлення виконане у гармонійній кольоровій гаммі.

9) Web-сайт виконано у сучасній манері з використанням візуальної ієрархії, це гарно відображено на головній сторінці рисунок 3.13.



Рис. 3.13. Головна сторінка web-сайту кафедри ІЕПФ ЗНУ

Недоліки:

- 1) Web-сайт має занадто великий розмір шрифту у місцях з текстовим наповненням.
- 2) Не вся розділи сайту мають структуровану наповненість та навіть є пусті (сторінка «Кейс-клуб»).
- 3) Присутня погана побудова змістовної ієрархії у навігації web-сайту.
- 4) Сторінка «Заявка на 2D диплом» не містить відповідної форми.
- 5) Як зображено на рисунку 3.14 у футері відсутня стандартна контактна інформація та посилання на соціальні мережі.

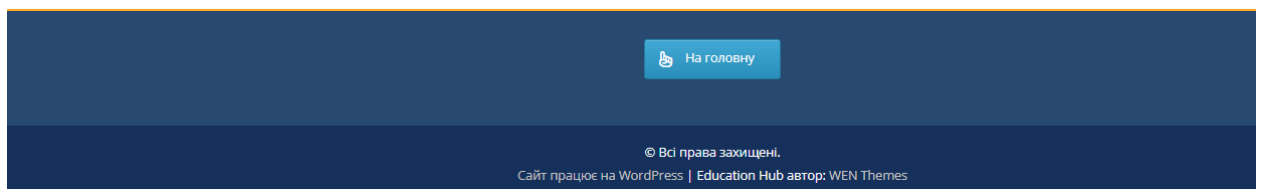


Рис. 3.14. Наповненість футеру web-сайту ІЕПФ ЗНУ

- 6) На сторінці з контактами дуже мало інформації, що робить її марною.
- 7) Немає можливості змінювати мову web-сайту.
- 8) Є признаки неактуальної та застарілої інформації.

Таким чином аналіз та порівняння з найкращім аналогом серед web-ресурсів структурних підрозділів університетів вказав на недоліки в організації контенту web-сайту кафедри ІЕПФ та необхідність в управлінні оцінкою контентом задля виявлення неактуальної та застарілої інформації, редизайну певних структурних елементів web-сайту.

3.3. Реалізація моделі ефективності функціонування контенту web-ресурсів

Проведемо реалізацію розглянутої системи математичних моделей оцінки ефективності web-ресурсів в підрозділі 2.2 на сайті кафедри інформаційної економіки, підприємництва та фінансів ЗНУ. Нами було проведено збір статистичних даних часу виконання запитів трьома інформаційними джерелами: інтегрованою системою інформаційних новин (IIS), персональними сторінками викладачів кафедри (PPT) і системою зберігання неструктурованого контенту web-ресурсу (UCWR). Для кожного джерела необхідно виміряти середній час відповіді в секундах із часовим інтервалом 2 хвилини.

На рис. 3.15. представлено фрагменти вибірки даних для кожного інформаційного джерела (IIS), (PPT), (UCWR).

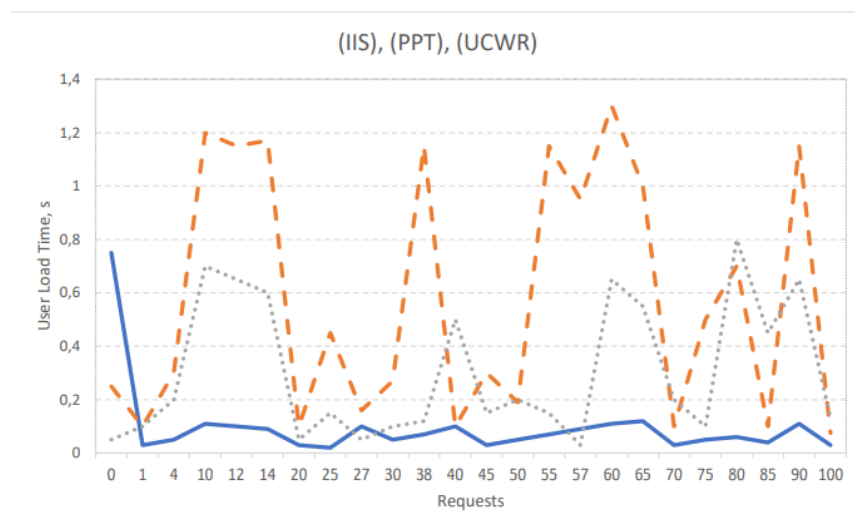


Рис.3.15. Експериментальні дані для кожного інформаційного джерела (IIS), (PPT), (UCWR) ресурсу <https://znuiepf.com.ua/>

На рисунку 3.15 видно, що тривалість виконання запитів розподілено нерівномірно, існують деякі значення, біля яких знаходиться велика частина спостережень. Особливо великі часові затримки для спостережень, що характеризують зберігання неструктурованого контенту досліджуваного web-ресурсу. Результати вимірювання тривалості виконання запиту зводили до середнього для серій з декількох вимірювань (чотирьох). Результати усереднених вимірювань тривалості виконання запитів у розрізі категорій, наведено у таблиці 3.5.

Таблиця 3.5.

**Результати вимірювання ефективності джерел із різним рівнем
структурованості контенту (IIS) (PPT), (UCWR) ресурсу**

<https://znuiepf.com.ua/>

№	IIS (інтегрована система інформаційних новин)	PPT (персональні сторінки викладачів кафедри)	UCWR (система зберігання неструктурованого контенту web-ресурсу)
Серії експерименту	r_{ft}	r_{m_i}	r_{st}
1	0.5	0.2	0.3
2	0.1	0.2	0.3
3	0.2	0.2	0.3
4	0.1	0.2	0.8
5	0.1	0.4	0.9
6	0.1	0.5	1.1
7	0.1	0.5	1.1
8	0.2	0.7	1.3
9	0.1	0.9	0.5
10	0.1	0.5	0.2
11	0.1	0.5	0.2
12	0.1	0.2	1
13	0.1	0.2	1
14	0.1	0.2	0.2
15	0.1	0.1	1

Середньоквадратичні значення є результатом обчислень на основі даних, наведених у таблиці 3.5 для кожної категорії інформаційного web-ресурсу. Тобто, середні та середньоквадратичні відхилення тривалості

запитів для кожної категорії є результатом, отриманим на основі 15 спостережень експерименту для кожної категорії.

Усі запити до контенту розділені за трьома категоріями: D_f , D_m , D_s - середній очікуваний час виконання запиту web-ресурсом для поточної категорії σ запитів. E_f , E_m , E_s - значення, що визначає відхилення спостережуваних значень від середніх величин для поточної категорії запитів.

Параметри D_f , D_m , D_s , E_f , E_m , E_s визначалися як середній час виконання запиту відповідним web-ресурсом, який відносився до певної категорії, та середньоквадратичним відхиленням, що характерний для ресурсу відповідної категорії, відповідно.

Результати оцінювання параметрів D_f , D_m , D_s , E_f , E_m , E_s наведено у таблиці 3.6.

Таблиця 3.6

Результати обчислення середніх значень та середньоквадратичних відхилень тривалості запитів для кожної категорії з різним рівнем структурованості контенту (IIS) (PPT), (UCWR) ресурсу <https://znuiepf.com.ua/>

Середнє значення тривалості запитів для IIS	Середнє значення тривалості запитів для PPT	Середнє значення тривалості запитів для UCWR
D_f	D_m	D_s
0,14	0,366	0,666
Середньоквадратичне відхилення тривалості запитів для IIS	Середньоквадратичне відхилення тривалості запитів для PPT	Середньоквадратичне відхилення тривалості запитів для UCWR
E_f	E_m	E_s
0,105	0,228	0,382

Значення із таблиці 3.5 були використані для визначення відповідних категорій ефективності. До категорії σ_f віднесено спостереження зі значенням близько 0.1 секунди, до категорії σ_s - спостереження зі значенням близько 1 секунди. До категорії σ_m віднесені спостереження в інтервалі між

0.1 та 1 секундою. Далі інтервали були скориговано так, щоб врахувати збурення значень. Ці інтервали також зроблено такими, що не перетинаються, щоб виключити частину значень, які знаходяться поблизу меж інтервалів і тому не можуть бути однозначно віднесені до тієї чи іншої категорії.

Таким чином, для отриманих даних визначено три інтервали $\Delta_f^r = (0; 0.15]$, $\Delta_m^r = (0.2; 0.38]$, $\Delta_s^r = (0.66; +\infty)$. Із вибірки для кожного типу контенту сформовано три множини значень, відповідно до обраних інтервалів. Далі оцінено ймовірності віднесення кожного з них до відповідної категорії. Ймовірності визначалися частотами попадання спостережуваних даних у відповідні інтервали.

Ймовірності визначалися частотами попадання спостережуваних даних у відповідні інтервали

Таблиця 3.7.

Оцінка ймовірності приналежності контенту до відповідної категорії ефективності

Тип контенту	$P(r_t \in \Delta_f)$	$P(r_t \in \Delta_m)$	$P(r_t \in \Delta_s)$
IS	0,9	0,07	0,03
IS	0,71	0,2	0,09
UCWR	0,24	0,39	0,37

Безпосередньо зі спостережень не можна зробити висновок про визначення параметра a (для малої початкової кількості вимірювань), тому проведено кілька експериментів із моделювання ефективності ресурсів із різними значеннями параметра a . Із отриманих результатів вибрано значення параметра a , що презентує найбільш близьку поведінку моделі порівняно з фактичними значеннями. На пізніших стадіях, коли кількість вимірювань тривалості запиту була більшою від кількості коефіцієнтів авторегресії, була використано методику, наведену у попередньому підрозділі (мінімізації середньоквадратичного відхилення вимірюваного та прогнозованого значень).

Оцінки значення параметра a й обчислених значень параметрів s, b рівняння (2.14) наведено у таблиці 3.8.

Таблиця 3.8.

Оцінки параметрів a, s, b для відповідної категорії контенту

Тип контенту	a	s	b
PS	0,1	-79	60,3
PPT	0,2	-21	21,7
UCWR	0,3	0,78	27,9

На основі знайдених показників a, s, b – заздалегідь відомі не випадкові числа (задані, або такі, що передбачають подальшу ідентифікацію у процесі вибірки), що впливають на показник r_t , який є випадковим процесом, що описується рівнянням авторегресії першого порядку для досліджуваного web-ресурсу S . Дані показники впливають на функції $D(r_t)$ та $E(r_t)$, які повертають значення параметрів рівняння (2.11) залежно від поточного режиму, що визначається значенням показника r_t .

Результат моделювання для системи зберігання неструктурованого контенту аналізованого web-ресурсу представлено на рисунку 3.16. Отримані моделі забезпечують значення статистичних параметрів (середнє значення і середньоквадратичне відхилення) кожної категорії ефективності web-ресурсу і частот віднесення контенту до категорії, що близькі до параметрів спостережень.

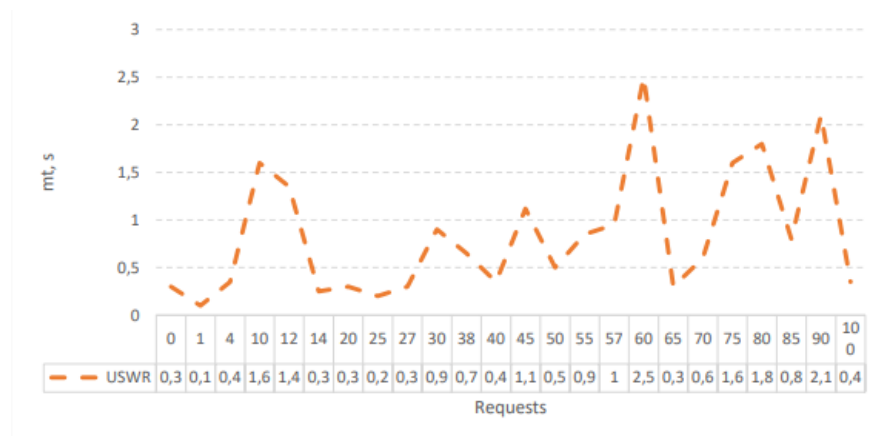


Рис. 3.16. Результати моделювання для неструктурованого контенту web-ресурсу <https://znuiepf.com.ua/>

Слід зауважити, що таблиця 3.7 відображає результати, що засвідчують найчастіше звертання користувача, а саме до добре структурованого контенту. Зауважимо, що у цьому випадку імовірності належності до діапазону тривалості виконання запитів добре структурованого контенту для різних категорій контенту розподілилися таким чином: $P(r_t \in \Delta_f) = \{0.9; 0.71; 0.24\}$. В інших випадках це б показувало службі підтримки потребу у структуруванні контенту, бо проміжок тривалості виконання запитів був би віднесений до інших категорій: другої $r_t \in \Delta_m$ та третьої $r_t \in \Delta_s$. Тобто це свідчило б про низьку ефективність інформаційного web-ресурсу.

Висновки до розділу 3.

У третьому розділі було проведено порівняльний аналіз web-ресурсів аналогів сайту кафедри ІЕПФ ЗНУ та їх попереднє оцінювання. Зокрема розглянуті 15 web-сайтів структурних підрозділів університетів, які експертно оцінювались автором за показниками змістовності та якості інформаційних ресурсів. Для подальшого детального аналізу із застосуванням методу нечітких множин було відібрано 7 web-сайтів аналогів.

Для ефективної організації процесу оцінювання контенту web-ресурсів, було розроблено форму опитування на web-сайті кафедри ІЕПФ ЗНУ, результати якої збирались у базу даних та застосовувались для подальшого аналізу даних.

Виконаний детальний порівняльний аналіз відібраних сайтів-аналогів за дев'ятьма критеріями, що відібрані з урахуванням обмежень даного сегменту. За результатами аналізу згідно схеми Беллмана-Заде відібрані три web-сайти, які характеризуються кращими показниками з урахуванням відносної ваги всіх задіяних критеріїв. Зокрема сюди потрапили web-ресурси Черкаського державного технологічного університету, Харківського

національного університету ім. В.Н. Каразіна та Західноукраїнського національного університету.

Зазначені переваги та недоліки web-ресурсу кафедри ІЕПФ ЗНУ у порівнянні з аналогами та визначено необхідність в управлінні оцінкою контентом задля виявлення неактуальної та застарілої інформації, редизайну певних структурних елементів web-сайту.

Наведено приклад застосування процедур оцінювання параметрів моделі на основі експериментальних даних, зібраних у процесі функціонування сайту кафедри ІЕПФ ЗНУ <https://znuiepf.com.ua>.

Отримані на основі запропонованого підходу математичні моделі можуть бути використані при вирішенні цілого ряду практичних завдань, що виникають при розробці і впровадженні web-систем: налаштування параметрів продуктивності, аналіз впливу структурованості контенту, оцінка ефективності функціонування під навантаженням.

ВИСНОВКИ

У результаті проведеного дослідження вирішено актуальну проблему ефективної організації та оцінювання контенту web-ресурсів структурних підрозділів університетів. На прикладі кафедри інформаційної економіки, підприємництва та фінансів ІННІ ім. Ю.М. Потебні ЗНУ було удосконалено методологічний підхід до процедури виявлення некоректної та застарілої інформації на web-ресурсах ЗВО, який, на відміну від існуючих, враховує внутрішні та зовнішні джерела інформації через бази ЕДЕБО та власні розробки інформаційних систем ЗВО, що дозволяє синхронізувати контент web-ресурсів із еталонною та актуальною інформацією з бази даних ЗВО.

Основними результатами кваліфікаційної роботи є:

Розглянуто особливості організації інформаційних web-ресурсів та спираючись на поставлені завдання, уточнено напрямки, в яких буде виконуватись ця кваліфікаційна робота магістра. Зазначено, що ефективність функціонування інформаційних web-ресурсів визначається рядом чинників, серед яких варто виділити: продуктивність апаратного і програмного забезпечення, способи організації джерела web-ресурсів, зокрема структурованість контенту та складність запиту.

Визначено, що ефективність функціонування більшої кількості web-ресурсів є незначною, через слабку структурованість контенту, а також наявність у ньому застарілої та неактуальної інформації.

В ході дослідження та оцінки контенту web-ресурсів ЗВО, уточнено позиціонування сайтів університетів та кафедр в класифікації сайтів, та вдосконалено вимоги щодо організації контенту web-ресурсів ЗВО для задоволення потреб всіх зацікавлених сторін.

Обґрунтовано показники оцінки контенту web-ресурсів, які стосуються витрат часу на обробку запитів. Зазначено, що жоден з наявних показників не дає можливості прогнозувати та моделювати ефективність контенту web-ресурсу. Вони лише придатні для збору поточної інформації

про інформаційний web-ресурс. Тому для використання показника ефективності необхідно розробити його математичне представлення і провести комп'ютерне моделювання та верифікацію запропонованого математичного опису з метою дослідження впливу структурованості контенту на ефективність web-ресурсу.

Такий підхід забезпечить не тільки встановлення цього впливу для конкретного web-ресурсу, але і забезпечить моделювання й оцінку структурованості контенту для розв'язування задач щодо його оптимізації, аналізу та перевірки на цілісність, достовірність й актуальність.

Виконаний огляд математичних методів порівняльного аналізу для використання в задачі раціональної організації контенту web-ресурсів. Підтверджено доцільність застосування підходу із використанням нечітких множин, надано розгорнуту характеристику метода нечітких множин Беллмана-Заде з прийняття рішень для аналізу аналогів web-ресурсів. На підставі виконаного аналізу запропоновано методичку, яка може бути задіяна на стадії порівняння web-сайтів аналогів.

Сформульовано основні вимоги до засобів автоматичного аналізу контенту, основним завданням яких є виявлення застарілої та некоректної інформації. Встановлено, що засоби повинні бути адаптовані на типові структури web-сторінок, а також на відомі структури опису web-сторінки мовою розмітки HTML.

Запропоновано та обґрунтовано підхід до моделювання базових показників оцінки ефективності функціонування інформаційних web-ресурсів. Він дозволяє врахувати характерні для нелінійних систем особливості, такі як: залежність збурень від поточних значень показника, циклічність процесів, стрибкоподібні зміни характеристик.

Обґрунтовано алгоритм збору інформації за допомогою парсингу та збереження інформації web-ресурсів. Основна ідея побудови алгоритму полягає у перетворенні інформації із інформаційних web-ресурсів у формат, придатний для порівняння з базами даних, що існують в організаціях або

виявлені на інших web-ресурсах. Розроблено алгоритм процедури виявлення некоректної та застарілої інформації на web-ресурсі кафедри ІЕПФ ЗНУ.

Проведено порівняльний аналіз web-ресурсів аналогів сайту кафедри ІЕПФ ЗНУ та їх попереднє оцінювання. Зокрема розглянуті 15 web-сайтів структурних підрозділів університетів, які експертно оцінювались автором за показниками змістовності та якості інформаційних ресурсів. Для подальшого детального аналізу із застосуванням методу нечітких множин було відібрано 7 web-сайтів аналогів.

Для ефективної організації процесу оцінювання контенту web-ресурсів, було розроблено форму опитування на web-сайті кафедри ІЕПФ ЗНУ, результати якої збирались у базу даних та застосовувалися для подальшого аналізу даних.

Виконаний детальний порівняльний аналіз відібраних сайтів-аналогів за дев'ятьма критеріями, що відібрані з урахуванням обмежень даного сегменту. За результатами аналізу згідно схеми Беллмана-Заде відібрані три web-сайти, які характеризуються кращими показниками з урахуванням відносної ваги всіх задіяних критеріїв. Зокрема сюди потрапили web-ресурси Черкаського державного технологічного університету, Харківського національного університету ім. В.Н. Каразіна та Західноукраїнського національного університету.

Зазначені переваги та недоліки web-ресурсу кафедри ІЕПФ ЗНУ у порівнянні з аналогами та визначено необхідність в управлінні оцінкою контентом задля виявлення неактуальної та застарілої інформації, редизайну певних структурних елементів web-сайту.

Отримані на основі запропонованого підходу математичні моделі можуть бути використані при вирішенні цілого ряду практичних завдань, що виникають при розробці і впровадженні web-систем: налаштування параметрів продуктивності, аналіз впливу структурованості контенту, оцінка ефективності функціонування під навантаженням.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Дивак М. П., Ковбасистий А.В. Особливості побудови методу виявлення некоректної та застарілої інформації. Сучасні комп'ютерні інформаційні технології: Матеріали VI Всеукраїнської школи-семінару молодих вчених і студентів АСІТ'2016. Тернопіль: ТНЕУ 2016. С. 119–120.
2. Дивак М.П., Мельник А. М., Ковбасистий А. В., Папа О. А. Підхід до математичного моделювання ефективності WEB-ресурсів // Оптико-електронні інформаційно-енергетичні технології, Міжнародний науково-технічний журнал. - №2 (38). - 2019. - С. 29-38.
3. Использование метода Бокса в задачах многокритериальной оптимизации / [А. Г. Данилкович, С.В. Брановицкая, С.Г. Бондаренко, О.В.Сангинова] // Восточноевропейский журнал передовых технологий. – 2013. – № 3/4 (63). – С. 4-8.
4. Математичні моделі та методи ринкової економіки. Навчально-методичний посібник для студентів ЗДІА спеціальності „Економічна кібернетика” всіх форм навчання / Укл. В.В. Глушцевський. ЗДІА, 2007. 210 с.
5. Пасічник Н. Метод та алгоритм побудови структур тематичних Веб-сайтів на основі онтологічного підходу. Наукові праці Донецького національного технічного університету: серія Інформатика, кібернетика та обчислювальна техніка, 2012. Вип 15 (203). С. 184–189.
6. Пасічник Н. Метод формування онтологічного контенту на основі аналізу інформації спеціалізованих Веб-сайтів. Вісник Хмельницького національного університету: серія Технічні науки. 2012. Вип. 5. С. 241–244.
7. Теслюк В.М. Моделі та інформаційні технології синтезу мікроелектромеханічних систем: Монографія. – Львів: Видавництво ПП ”Вежа і Ко”, 2008 – 192 с.
8. Типи Сайтів. Landing Page [Електронний ресурс] / Режим доступу: http://znet.ru/raskrutka/vidyi-saytov-i-ih-klassifikatsiya-s-primerami/#Landing_Page.

9. Юзабіліті веб-сайту. Основні принципи юзабіліті: від особливостей дизайну до контенту, від інтерфейсу форми замовлення до пошуку. Можливості тестування. [Електронний ресурс] / Вікторія Закірова – Режим доступу: <https://www.ashmanov.com/education/articles/yuzabiliti/>

10. Analysis of site structure. URL: <http://www.web-patrol.net/audit-sitestructur.html>.

11. Asana I.M.D.P., Wiguna I.K.A.G., Atmaja K.J., Sanjaya I.P.A. FPGrowth Implementation in Frequent Itemset Mining for Consumer Shopping Pattern Analysis Application. Mantik J. 2020, 4, P. 2063–2070.

12. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. Zhang Y., Wallace B., 2015. [Електронний ресурс] – Режим доступу до ресурсу: <https://arxiv.org/abs/1510.03820>

13. Bellman R.E., Zadeh L.A. Decision-making in a fuzzy environment. Management Science. 1970. Vol. 17, N 4 P. B-141–B-164.

14. Duncan Temple Lang. XML: Tools for parsing and generating XML within R and S-Plus. URL: <http://CRAN.R-project.org/package=XML>

15. Duncan Temple Lang. RCurl: General network (HTTP/FTP/...) client interface for R. URL: <http://CRAN.R-project.org/package=RCurl>

16. Duriez B., Chilaka S., Bercher J.-F., Hercul E. and Prioleau M.-N. Replication dynamics of individual loci in single living cells reveal changes in the degree of replication stochasticity through S phase. Nucleic Acids Research. 2019. vol. 47. P. 5155-5169.

17. Dyvak M. P., Melnyk A. M., Kovbasisty A.V., Turchyn L. Y., Martsenyuk Y. O. System for web resources content structuring and recognizing with the machine learning elements // Radio Electronics, Computer Science, Control. - Zaporizhzhia, No 3(46). - 2018. - P. 128-134. (Web of Science).

18. Dyvak M., Melnyk A., Kovbasisty A., Shcherbiak I., Huhul O. Recognition of relevance of web resource content based on analysis of semantic

components. *Advanced Computer Information Technologies (ACIT'2019): IX International conference (Czech Republic)*. Ceske Budejovice, 2019. P. 297–302 (IEEE Xplore Digital Library, DOI: 10.1109/ACITT.2019.8779897).

19. Dyvak M., Kovbasisty A., Stakhiv P., Lipiński P. Generalized structure of the algorithm for automated detection of non relevant and wrong information on web resources // *Journal of Applied Computer Science*. - 2017. - P. 23-37.

20. Dyvak M., Pasichnyk N. Formalism in setting the task of creating a quality website. *Scientific papers of Donetsk National Technical University: ser. Informatics, Cybernetics and Computer tehnika*. 2011. Vol. 14 (188). P. 325–329.

21. Club Runner «10 Golden Rules of Website Design & Functionality» [Електронний ресурс] - Режим доступу: <https://site.clubrunner.ca/page/10-golden-rules-of-website-design-functionality>

22. Guha R., Brickley D., Macbeth S. Schema.org: evolution of structured files on the on-line. *Commun. ACM*. 2016. Vol. 59, 2. P. 44–51. [URL:https://doi.org/10.1145/2844544](https://doi.org/10.1145/2844544)

23. Gutiérrez S., Torres N., Domínguez N. Análisis con «big data» de las respuestas de los hoteles en TripAdvisor. *Esic Mark*. 2018, P. 160, 339–378.

24. Haller A., Janowicz K., Cox S., Phuoc D., Taylor K., Lefrancois M. (Eds.). *Semantic Sensor Network Ontology*. W3C Recommendation 19 October, 2017. URL:<http://www.w3.org/TR/vocabssn/>.

25. Harrington P. *Machine Learning in Action*. New York: Manning Publications Co., Shelter Island, 2012 (ISBN 9781617290183).

26. Hammar K. et al. Easy research questions relating ontology uncover patterns. *Ontology Engineering with Ontology Salvage Patterns—Foundations and Functions*. Hitzler P., Gangemi A., Janowicz K., Krisnadhi A., Presutti V. (Eds.). *Studies on the Semantic Web 25*. IOS Press, 2016. P. 189–198.

27. Hitzler P., Bianchi F., Ebrahimi M., Sarker M. Neural-symbolic integration and the Semantic Web. *Semantic Web*. 2020. Vol. 11, 1. P. 3–11.

28. Kraus M., Feuerriegel S., Oztekin A. Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*. 2020. Vol. 281(3). P. 628–641. URL: <https://dx.doi.org/10.1016/j.ejor.2019.09.018>
29. Jamie Juviler «9 Guidelines for Exceptional Web Design, Usability, and User Experience» [Електронний ресурс] – Режим доступу: <https://blog.hubspot.com/blog/tabid/6307/bid/30557/6-guidelines-for-exceptional-website-design-and-usability.aspx>.
30. Matsyi O. Approaches to solving basic problems of closed routes. *IEEE Proceedings of 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET-2020)*. LvivSlavske, 2020. February 25–29. P. 69–72. (Scopus).3
31. Saurkar A., Pathare K., Gode S. An Overview On Web Scraping Techniques And Tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*. 2018. Vol. 4(4). P. 363–367.
32. Singh N., Baack D.W. Web Site Adaptation: A Cross-cultural Comparison of U.S. and Mexican Web sites. *Journal of Computer-mediated Communication*. 2004. Vol. 9(4).
33. Smola A., S.V.N. Vishwanathan: *Introduction to Machine Learning*, Cambridge: Cambridge University Press, 2008; eBook (October 1, 2010). P. 234.
34. Spinelli A.S., Compagnoni C.M., Lacaita A.L. Reliability of NAND Flash Memories: Planar Cells and Emerging Issues in 3D Devices. *Computers*. 2017. Vol. 6. P. 16. [CrossRef]. Ielmini D. Brain-inspired computing with resistive switching memory (RRAM): Devices, synapses and neural networks. *Microelectron. Eng.* 2018. Vol. 190. P. 44–53. [CrossRef]
35. Pasichnyk N., Dyvak M. Formalism in the quality site creating problem. *Naukovi pratsi DonNTY: ser. Informatyka, kibernetyka ta obchysliuvalna tekhnika*. 2011, Vol. 14(188). P. 325–329 (in Ukrainian).
36. Pasichnyk N. Mathematical modeling of the Website quality characteristics in dynamics / N.Pasichnyk, M.Dyvak, R.Pasichnyk. *Journal of*

Applied Computer Science. Lodz: Technical University Press, 2014. Vol. 22 № 1. P. 171–183.

37. Thomas DM, Mathur S. Data analysis by web scraping using python. 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA). 2019. P. 450–454. URL: <https://ieeexplore.ieee.org/abstract/document/8822022>.

38. Vania C., Kementchedjhieva Y., Søgaard A., Lopez A. A systematic comparison of methods for low-resource dependency parsing on genuinely lowresource languages. Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, November, 2019. P. 1105–1116.

39. vTools for parsing in the work of an SEO specialist. URL: <https://netpeak.net/ru/blog/instrumenty-dlya-parsinga-v-rabote-seo-spetsialista/>.

40. Weare C., Lin W.Y., Content Analysis of the World Wide Web: Opportunities and Challenges. Social Science Computer Review. 2002. Vol. 18(272).

41. Witten I., Frank E., Hall M. Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition. San Mateo: Morgan Kaufmann, CA, 2011.

42. WordStream «A Beginner's Guide to Content Strategy for the Web: 10 Things You Need to Know» [Электронный ресурс] - Режим доступа: <https://www.wordstream.com/blog/ws/2012/11/28/content-strategy>

43. Xu Z., Wei X., Liu Y., Mei L., Hu C., Choo K. K. R., Sugumaran V. Building the search pattern of web users using conceptual semantic space model. International Journal of Web and Grid Services. 2016. Vol. 12 (3). P. 328–347.