

В.В. ЗОСИМОВ, канд. техн. наук, доцент кафедри комп'ютерних наук та прикладної математики, Миколаївський національний університет імені В.А. Сухомлинського, вул. Нікольська, 24, 54030, м. Миколаїв, Україна, zosimovvv@gmail.com

ПОБУДОВА ЗАГАЛЬНОЇ СТРУКТУРИ КОРПОРАТИВНИХ ВЕБ-РЕСУРСІВ

На основі аналізу структури та інформаційного вмісту корпоративних веб-ресурсів визначено їхню загальну структуру. Результати дослідження є невід'ємним блоком розв'язання задачі розробки спеціалізованої онтології корпоративних веб-ресурсів, враховуючи особливості вітчизняного ринку розробки Інтернет-ресурсів.

Ключевые слова: корпоративні веб-ресурси, видобування інформації, структура веб-ресурсів, веб-навігація, usability.

Вступ

Стрімкий розвиток інформаційних технологій в останні десятиріччя ставить перед суспільством цілий клас задач ефективної обробки великих масивів слабкоструктурованої інформації, представленої у вигляді веб-сторінок. Серед них — створення стандартів розробки веб-ресурсів визначених типів, а також пошук, вилучення, обробка, аналіз, збереження та відображення інформації.

Інтеграція бізнес-процесів, що відбуваються на підприємствах України, в середовище закордонних партнерів диктує необхідність розвитку сфери електронної комерції як необхідної умови існування сучасних компаній.

Значні результати досягнуто в галузі розробки моделей семантичної розмітки Інтернет-магазинів як основного інструменту ведення електронної комерції.

Як стандарт мікророзмітки продуктів електронної торгівлі з 2008 р. використовується інструментарій *Good Relations* [1, 2], який надає можливість вказати спеціальні властивості для:

- компанії — контактні дані, місцезнаходження, логотип;

- магазину — адреса, час роботи, телефон;
- конкретного товару — категорія товару, стислий опис, код, способи оплати тощо.

В той же час досить мало уваги приділяється електронному ринку послуг, а саме структурним та семантичним стандартам розробки корпоративних веб-ресурсів. В українському сегменті всесвітньої павутини лише незначний відсоток веб-ресурсів розроблений з використанням стандартів семантичної розмітки.

Враховуючи зазначене, виникає необхідність розробки комплексної технології оперування веб-даними, яка включатиме методи, моделі та програмні засоби створення веб-ресурсів з інтегрованою семантичною розміткою та пошук інформації з використанням пошукових агентів [3]. Для забезпечення ефективної роботи такої технології необхідно дослідити вміст корпоративних веб-ресурсів, розробити семантичну структуру та словник семантичних тегів для опису побудованої структури.

Тому мета даної статті — на основі аналізу структури та інформаційного вмісту корпоративних веб-ресурсів, побудувати їх загальну структуру, що є першим кроком до побудови:

- спеціалізованого семантичного профілю корпоративних веб-ресурсів у вигляді семантичних класів та їх властивостей для опису всього інформаційного вмісту;
- системи керування вмістом з вбудованою можливістю інтеграції семантичної розмітки в *HTML*-код веб-сторінок.

Структурні елементи веб-ресурсу

Структура веб-ресурсу це набір веб-сторінок, зв'язаних між собою гіперпосиланнями (елементами навігації) згідно заданої ієрархії.

Структурними елементами веб-ресурсу є сторінки, зв'язок між якими забезпечують гіперпосилання, об'єднані в групи, що формують навігаційні меню. За типом їх поділяють на основне та додаткові [4].

Основне навігаційне меню є обов'язковим елементом корпоративного веб-ресурсу. Воно забезпечує навігацію між основними розділами, такими як «Про компанію», «Послуги», «Відгуки», «Контакти» тощо.

Додаткові меню не є обов'язковими елементами. Вони використовуються для:

- забезпечення швидкого доступу до ключових сторінок, наприклад, конкретних послуг. Такі навігаційні панелі розташовуються в бокових колонках веб-ресурсу або безпосередньо під основним навігаційним меню у верхній частині сайту і дублюють ті пункти основного меню, на яких власник ресурсу хоче зробити особливий акцент;
- забезпечення більш ефективного індексування веб-ресурсу пошуковими роботами. Додаткові навігаційні панелі забезпечують більш швидку індексацію веб-ресурсу в *глибину*. Такі навігаційні меню розташовуються в бокових колонках як список посилань на інформаційні статті або в нижній частині сайту – як дублікат основного меню.

Структура ресурсу може бути побудована у вигляді дерева на основі головного навігаційного меню [5].

Для узагальнення структури корпоративних веб-ресурсів, необхідно провести експерименти:

- побудувати загальну структуру верхнього рівня; експеримент полягає у видобуванні

текстових значень зі структурних елементів основного навігаційного меню для великої кількості веб-ресурсів, складанні таблиці кількості входжень кожного елемента, об'єднанні та узагальненні результатів;

- побудувати внутрішню структуру для кожного елемента верхнього рівня. Експеримент полягає в дослідженні інформаційного наповнення та вкладених елементів кожного типового структурного елемента верхнього рівня для великої кількості веб-ресурсів, складанні таблиці кількості входжень кожного елемента, об'єднанні та узагальненні результатів.

Алгоритм розробки навігаційних меню

Для створення ефективного алгоритму розробки навігаційних меню необхідно визначити їх характерні особливості.

Кожен елемент навігації в *HTML*-коді визначається тегом `<a>` з обов'язковим атрибутом `href = "URL сторінки"`.

Однозначною ознакою навігаційного меню є тег `<nav>`, який доданий в специфікації мови *HTML5* для семантичної розмітки елементів навігації. Стандарт *HTML5* опубліковано наприкінці 2014 р., тому існує великий відсоток веб-ресурсів, розроблених на базі ранніх специфікацій мови *HTML*, відповідно без використання тегу `<nav>`.

Характерними особливостями навігаційного меню є наявність більше двох послідовно розміщених тегів `<a>` в межах одного блоку *HTML*-сторінки, не розділених текстовими даними.

На основі зазначених вище характерних особливостей можна побудувати алгоритм видобування текстових значень з елементів навігаційного меню (рис. 1).

Визначення загальної структури

Важливо відзначити, що запропонований алгоритм видобування елементів навігації не розрізняє основне та додаткові меню. Але для даного експерименту це не критично. Елементи додаткових меню, як правило, вказують на

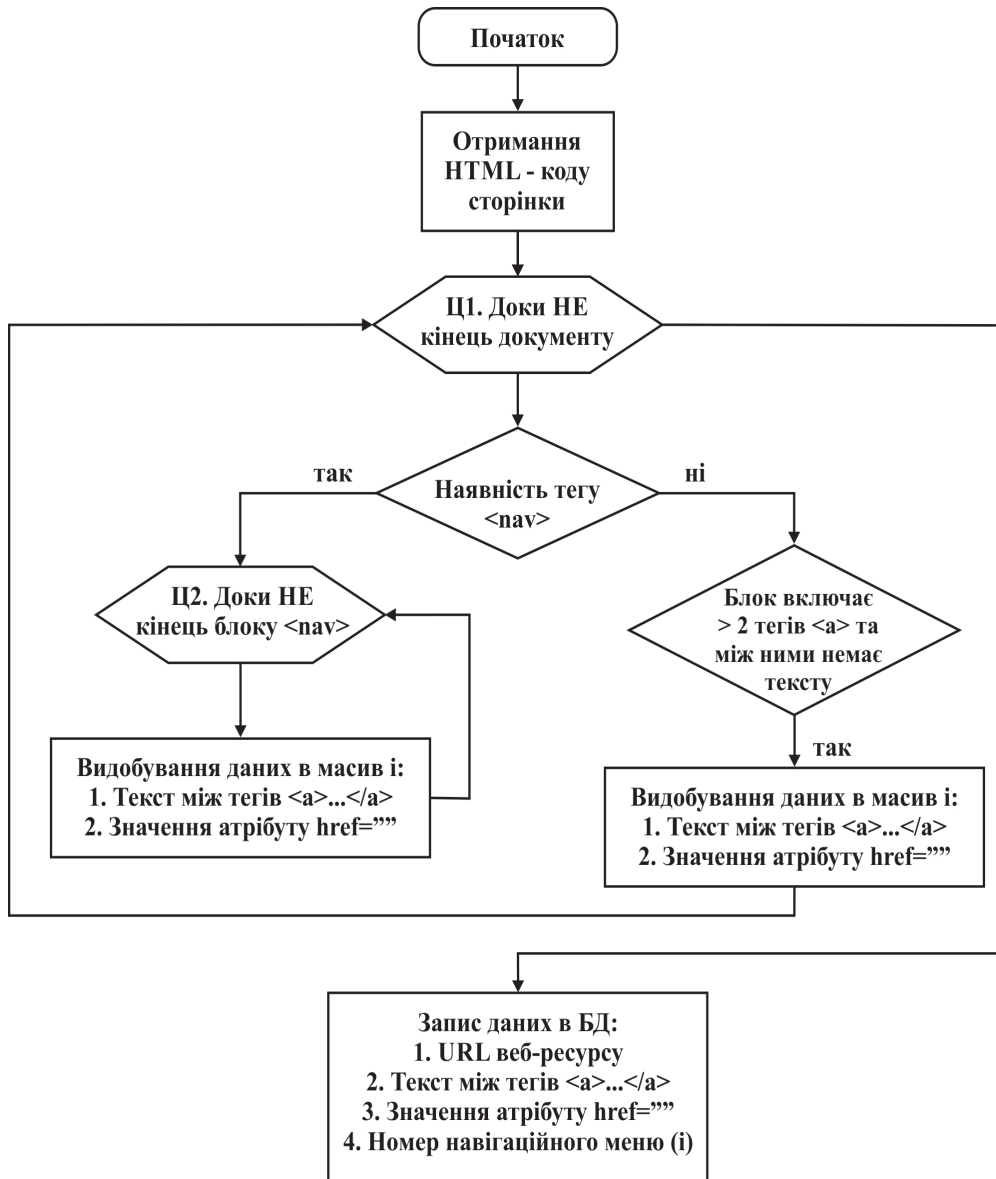


Рис. 1. Схема алгоритму видобування текстових значень з елементів навігаційного меню

специфічні для конкретного веб-ресурсу сторінки і тому будуть відсіянні за загальною кількістю входжень після складання зведеної таблиці.

В ході експерименту досліджено структуру 1000 корпоративних веб-ресурсів. Аналіз веб-ресурсу та видобування даних здійснювалось за допомогою автоматичного парсеру, реалізованого засобами мови оперування веб-даними, розробленої автором [3].

Веб-ресурси для проведення аналізу автоматично відібрані пошуковою системою *Google* на запит «*Наша компанія*». Таке формулювання пошукового запиту, очевидно, забезпечую високу імовірність наявності в результатах пошуку саме корпоративних сайтів з різних галузей діяльності, а також з каталогу веб-ресурсів України *ukraina.net.ua*. Аналізувалась тільки головна сторінка веб-ресурсів.

Загалом оброблено 1284 веб-ресурсів, з яких відсіяно повторні посилання на веб-ресурси, дошки об'яв та сайти-агрегатори послуг.

З кожного веб-ресурсу видобувались елементи навігаційного меню у вигляді *Назва елемента* -> *Посилання* і зберігались в базі даних. В табл. 1 представлено перелік структурних елементів впорядкованих за зменшенням частоти їх входжень до веб-ресурсів. Всього відібрано 5371 унікальних елементів.

Елементи з числом входжень менше 50 були проігноровані, як неінформативні. Це специфічні для деяких компаній елементи з посиланням:

- на конкретні послуги, які в загальному вигляді є елементами другого рівня для кореневого «Послуги»;

- на конкретні статті, присвячені діяльності компанії, які є елементами другого рівня для корневих «Статті» або «Про компанію»;

Відсіяно як неінформативні — 5326 елементів.

Список інформативних структурних елементів представлено в табл. 1 в порядку зменшення загальної кількості входжень.

Побудова внутрішньої структури для кожного елемента верхнього рівня

В ручному режимі переглянуто 100 обраних довільним чином веб-ресурсів з попередньої вибірки для дослідження інформаційного вмісту базових структурних елементів.

Таблиця 1. Перелік структурних елементів корпоративних веб-ресурсів

№ з/п	Назва елемента	Кількість входжень	№ з/п	Назва елемента	Кількість входжень
1	Контакти	891	24	Нагороди	84
2	Про компанію	637	25	Партнери	81
3	Послуги	611	26	Клієнти про нас	78
4	Новини	363	27	FAQ	75
5	Продукція	232	28	Проекти	74
6	Вакансії	217	29	Дилерам	72
7	Про нас	193	30	Документи	72
8	Співпраця	185	31	Наші об'єкти	71
9	Сертифікати	177	32	Акції	68
10	Інформація	169	33	Умови співпраці	65
11	Сервіс	156	34	Для дилерів	63
12	Відгуки	146	35	Наші контакти	62
13	Головна	142	36	Події	60
14	Кар'єра	138	37	Відзнаки	59
15	Статті	125	38	Для партнерів	57
16	Пресс-центр	124	39	Клієнти	56
17	Ліцензії	102	40	Наші проекти	56
18	Наші послуги	98	41	Наша компанія	54
19	Наші роботи	97	42	Наші партнери	53
20	Наша продукція	93	43	Виробництво	52
21	Об'єкти	93	44	Галерея робіт	51
22	Наші клієнти	87	45	Наші новини	50
23	Магазин	86			

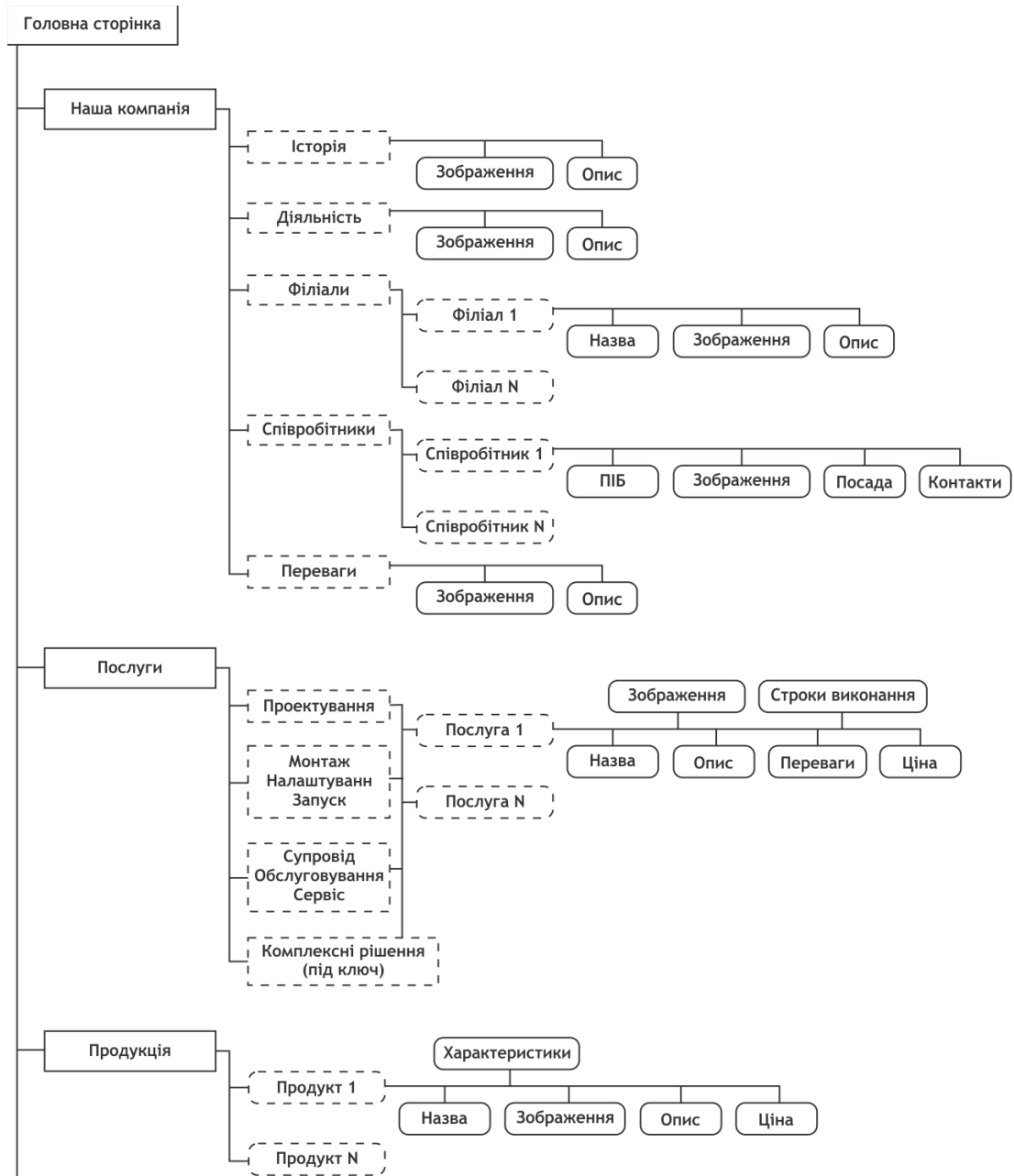


Рис. 2. Структура елемента "Головна сторінка" корпоративних веб-ресурсів

Наша компанія. Містить інформацію про компанію, види її діяльності, історію виникнення тощо. Частково дублюється 72 відсотка інформації головної сторінки. Тому доцільно об'єднати ці розділи.

Проекти. Містить перелік реалізованих замовлень у вигляді портфолію з назвою,

зображенням і характеристиками. Доцільно об'єднати з розділами *Наші клієнти*, *Наші об'єкти*, *Галерея робіт* як близькі за змістом.

Документи. Містить перелік таких документів, що засвідчують легальність компанії, зразки документів для заповнення, різні договори.

Сертифікати, Ліцензії, Відзнаки. Містить перелік документів з їх описом.

Розділи *Документи, Сертифікати, Ліцензії, Відзнаки* доцільно об'єднати як близькі.

Відгуки. Містить відгуки відвідувачів. Важливо зазначити, що в 17 відсотках веб-ресурсів з вже опублікованими відгуками користувачів, відсутня можливість розмістити відгук. Це викликає сумніви в реальності відгуків. При аналізі веб-ресурсів та подальшому прийнятті рішень враховувати можна лише ті відгуки, власники яких підтвердили свою особистість шляхом авторизації на веб-ресурсі або спеціалізованих сторонніх сервісах.

Вакансії. Містить перелік вакансій, опис можливостей кар'єрного зростання, загальну інформацію про працевлаштування.

Співпраця. Містить інформацію про умови співпраці.

Наші партнери. Містить перелік партнерів, в більшості випадків у вигляді списку логотипів.

Доцільно об'єднати розділи *Співпраця та Наші партнери, Дилерам*, як близькі за змістом.

Таблиця 2. Групи структурних елементів веб-ресурсів

Назва структурного елемента	Частота входження
Статті, Інформація, FAQ, Новини, Наші новини, Пресс-центр, Акції, Події	1041
Головна, Про нас, Про компанію, Наша компанія	1026
Контакти, Наші контакти	953
Послуги, Наші послуги, Сервіс	865
Наші клієнти, Клієнти, Об'єкти, Проекти, Наші роботи, Наші проекти, Наші об'єкти, Галерея робіт	585
Партнери, Дилерам, Співпраця, Умови співпраці, Для партнерів, Для дилерів, Наші партнери	577
Документи, Ліцензії, Сертифікати, Нагороди, Відзнаки	494
Продукція, Наша продукція, Виробництво, магазин, товари	463
Вакансії, Кар'єра	355
Відгуки, Клієнти про нас	220

Контакти. Містить необхідну користувачу контактну інформацію, карту проїзду. Також в цей розділ доцільно перемістити всі типи контактних форм, такі як *Зворотний зв'язок, Виклик замірника, Зворотний дзвінок, Замовлення столиків* тощо. Такі форми значно перевантажують сторінки веб-ресурсу і часто бувають дуже нав'язливими.

Послуги. Містить перелік послуг, що надає компанія, у разі необхідності, може бути розділений на типи: проектування, монтаж, сервісне обслуговування тощо.

Продукція. Містить інформацію про представлену продукцію.

Новини. Цей розділ використовується «за призначенням», а саме для висвітлення новин щодо діяльності компанії лише для крупного бізнесу. Це зумовлено тим, що в них є великий штат працівників, велике виробництво, регіональні представництва, а великі масштаби підприємства генерують велику кількість подій, які можна висвітлювати. На веб-ресурсах малих та приватних компаній сторінка *Новини* в більшості випадків заповнена загальними новинами, які авто-матично завантажуються з новинних ресурсів або взагалі залишаються порожніми.

Статті. Спираючись на статистику сайту американської асоціації веб-майстрів (AWA), власники веб-ресурсів наповнюють цей розділ корисними інформаційними статтями стосовно діяльності компанії, послуг, продукції, правил використання лише в 12 відсотках випадків. Ще 46 відсотків замовляють послуги копірайтерів для наповнення розділу *SEO-статтями* для підвищення штучного рейтингу за результатами пошукових систем. 23 відсотка містять дещо перероблені копії вже існуючих статей з інших ресурсів. Останні 19 відсотків залишаються порожніми.

В 31 відсотках переглянутих веб-ресурсів цей розділ містить не більше двох статей, розміщених незадовго після створення веб-ресурсу або взагалі порожніх. У 52 відсотках були присутні *SEO-статті*. Їх характеризують:

- велика кількість ключових слів, виділених жирним шрифтом;

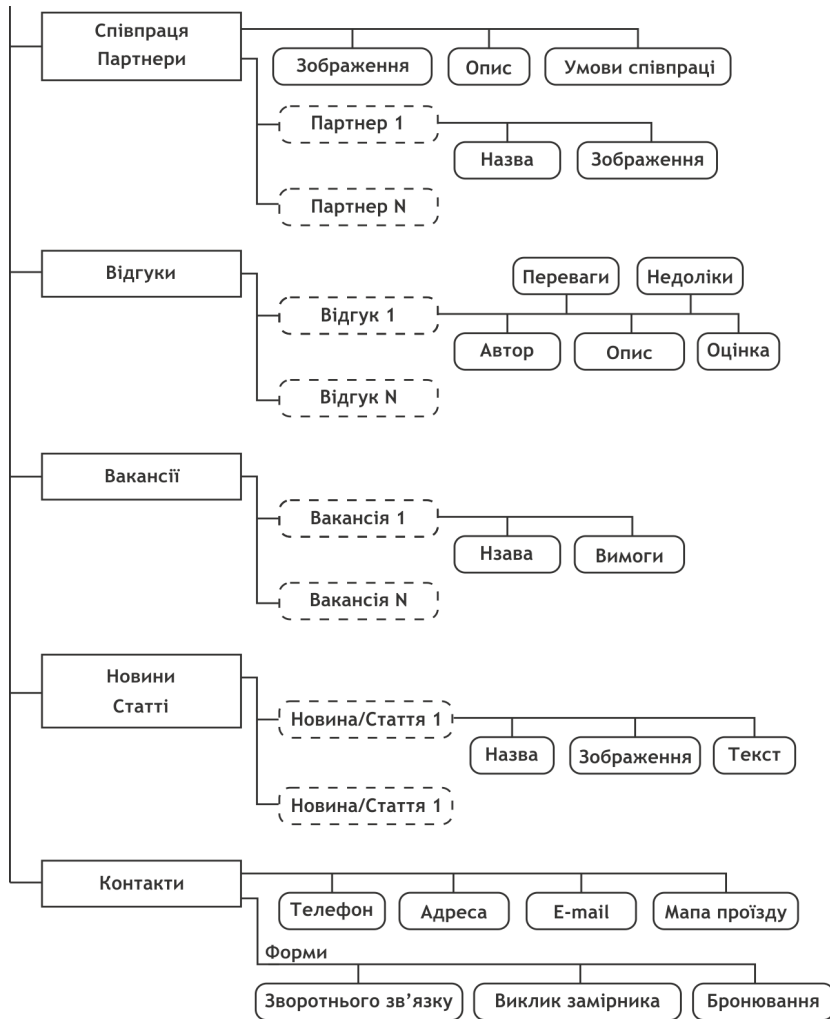


Рис. 3. Групи структурних елементів веб-ресурсів

- наявність загальної інформації за темою, але без подробиць, оскільки написані не спеціалістами з предметної галузі;

- невеликий обсяг, близько 1000–1500 символів (200–250 слів), що є достатнім для пошукової системи.

У 19 відсотках наявні великі докладні статті, що глибоко розкривають тему.

Часто короткі анонси новин та статей виносяться в ліву або праву бокову частину веб-ресурсу з посиланнями на повний текст статей. Це забезпечую кращу індексацію розділу статей пошуковими системами, а також створює ілюзію того, що інформаційне наповнення сторінок часто оновлюється, що

є одним з позитивних факторів підвищення рейтингу веб-ресурсу за результатами пошуку, але заважає користувачу сприймати основну інформацію на сторінці, перевантажуючи її зайвою інформацією.

Враховуючи досить низький відсоток використання розділів *Новини* та *Статті* для розміщення корисної для користувачів інформації, доцільно об'єднати ці елементи в один, а також, приєднати до них розділи *Акції*, *Події*, *Інформація* як близькі за змістом. При перевантаженні основного меню, можна винести його підпунктом в розділ *Наша компанія*.

В результаті дослідження сформовано структуру елементів верхнього рівня. Схожі за сен-

сом структурні елементи об'єднані в тематичні групи.

Загальна структура корпоративних веб-ресурсів побудована за результатами експерименту, а також спираючись на результати досліджень провідних фахівців в галузі веб-дизайну та зручності використання (*usability*) веб-ресурсів [6, 7]. Згідно з рекомендаціями, кількість кореневих елементів основної навігації повинна бути не більше восьми.

Побудована структура включає 10 елементів, але слід зазначити, що на жодному з досліджених веб-ресурсів не були присутні всі ці елементи одночасно. Максимальна їх кількість налічувала дев'ять. Загальна схема включає всі можливі варіанти. При розробці кожного конкретного проекту з них обираються тільки відповідні до вимог компанії.

В табл. 2. представлено групи структурних елементів та загальна частота входження всіх елементів групи. Кількість входжень, яка перевищує загальну кількість проаналізованих веб-ресурсів, зумовлена тим, що надіяких з них присутні два чи більше пунктів меню, які потім були об'єднані в одну групу.

Дані табл. 2 використано як структурні групи, до яких додавались елементи, близькі за змістом, для забезпечення зручності навігації за веб-ресурсом.

Наприклад, в групу *Про компанію* додавались *Наш бізнес, Команда, Співробітники, Представництва, Філіали, Джерела натхнення, Історія компанії, Розвиток компанії* тощо.

Перелічені структурні елементи виділені як елементи другого рівня для кореневого

Про компанію. Така організація навігаційної панелі зменшує перенавантаження веб-сторінки елементами навігації першого рівня і буде інтуїтивно зрозумілу для користувача систему навігації.

Базуючись на результатах дослідження та визначеній загальній структурі елементів верхнього рівня, побудовано дерево структурних елементів корпоративного веб-ресурсу, представлено на рис. 2, 3.

Висновки

Для визначення структури корпоративних веб-ресурсів проведено експерименти, за допомогою яких покроково досліджено:

- загальну структуру;
- елементи навігації;
- інформаційні розділи.

В результаті дослідження сформована структура елементів верхнього рівня. Схожі за сенсом структурні елементи об'єднані в тематичні групи для відображення.

Базуючись на результатах дослідження та побудованій загальній структурі елементів верхнього рівня, побудовано дерево структурних елементів корпоративного веб-ресурсу.

Наявність загальної структури для визначеного типу веб-ресурсів, відкриває широкі можливості підвищення ефективності інтелектуальної обробки та представлення результатів видачі пошукових систем.

Побудована структурна схема корпоративних веб-ресурсів є базою для розробки спеціалізованої онтології корпоративних веб-ресурсів, враховуючи особливості вітчизняного ринку розробки Інтернет-ресурсів.

СПИСОК ЛІТЕРАТУРИ

1. <https://www.w3.org/DesignIssues/Semantic.html> — основні положення семантичного веб.
2. <http://www.heppnetz.de/projects/goodrelations/> — Офіційний сайт семантичного словника Good Relations.
3. Зосімов В.В. Комплексний підхід до підвищення ефективності обробки веб-даних на основі семантичної розмітки. УСиМ, 2018, №4, с. 32–45.
4. Якоб Нильсен, Хоа Лоранжер Web-дизайн: удобство использования Web-сайтов - Prioritizing Web Usability. М.: Вильямс, 2007. 368 с. ISBN 0-321-35031-6.
5. Речинский А.В., Сергеев С.Ф. Разработка пользовательских интерфейсов. Юзабилити-тестирование интерфейсов информационных систем. СПб.: Изд-во Политехн. ун-та, 2012, 145 с. (Учебное пособие). ISBN 978-5-7422-3717-4.

6. Круг С. Как сделать сайт удобным. Целесообразность по методу Стива Круга = Rocket Surgery Made Easy: The Do-It-Yourself Guide to Finding and Fixing Usability Problems. СПб.: Питер, 2010. С. 208. ISBN 978-5-49807-515.
7. Гарретт Д. Веб-дизайн: книга Джесса Гарретта. Элементы опыта взаимодействия = The Elements of User Experience: User-Centered Design for the Web. Символ-Плюс, 2008. 192 с. ISBN 5-93286-108-8.

Надійшла 06.02.2019

REFERENCES

1. <https://www.w3.org/DesignIssues/Semantic.html> — The basic provisions of the semantic web. (In Ukrainian).
2. <http://www.heppnetz.de/projects/goodrelations/> — Official site of the semantic dictionary Good Relations. (In Ukrainian).
3. Zosimov, V.V., 2018. "An Integrated Approach to Improving the Efficiency of Web Data Processing Based on Semantic Markup". *Upravl iе sistemy i ma iny*, 4, pp. 32–45. (In Ukrainian).
4. Nyl sen, Ya., Loran zher, K., 2007. Web design: Web site usability. Prioritizing Web Usability. M.: Vilyams, 368 p. ISBN 0-321-35031-6. (In Russian).
5. Rechinskiy, A.V., Serge yev, S.F., 2012. Razrabotka pol'zovatel'skikh interfe ysov. Yuzabiliti-testirovaniye interfe ysov informatsionnykh sistem. SPb.: Izd-vo Politekhn. un-ta, 145 p. ISBN 978-5-7422-3717-4. (In Russian).
6. Krug, S., 2010. Kak sdelat' sayt udobnym. Tselesoobraznost' po metodu Stiva Kруга. Rocket Surgery Made Easy: The Do-It-Yourself Guide to Finding and Fixing Usability Problems. SPb.: Piter, 208 p. ISBN 978-5-49807-515. (In Russian).
7. Garrett, D., 2008. Veb-dizayn: kniga Dzhesa Garretta. Elementy opyta vzaimodeystviya. The Elements of User Experience: User-Centered Design for the Web. Simvol-Plyus, 192 p. ISBN 5-93286-108-8. (In Russian).

V.V. Zosimov, Ph.D in Techn. Sciences, Associate Professor of the Computer Science and Applied Mathematics Department, V.O. Sukhomlynsky Mykolaiv National University, Nikolska str., 24, Mykolaiv, 54030, Ukraine, zosimovvv@gmail.com

CONSTRUCTION OF THE CORPORATE WEB RESOURCE GENERAL STRUCTURE

Introduction. Today, enough attention is paid to the electronic market of services, namely structural and semantic standards for the corporate web resources development. In the Ukrainian segment of the World Wide Web, only a small percentage of web resources is developed using the semantic markup standards.

Therefore, there is a need to develop a comprehensive web service technology, which will include methods, models and software for creating web resources with integrated semantic markup and information search using search agents. To ensure the effective operation of the technology, it is necessary to examine the content of corporate web resources, to develop a structure and dictionary of semantic tags to describe the structure constructed.

Purpose. Based on the analysis of the corporate web resources structure and its content, it is necessary to determine the general structure of corporate web resources.

Methods. The structure of the upper level elements is formed. Structural elements that are similar in meaning are merged into thematic groups for displaying. Based on the research results and the general structure of the upper-level elements built, the tree of the structural elements of the corporate web resource is constructed.

Results. During the experiment, the structure of 1000 corporate web resources is investigated. Web resource analysis and data retrieval are performed using an automated parser implemented by means of a web-based operating language.

The navigation menu elements are extracted from each web resource (5371 unique items were seized). Elements with an entry number less than 50 are ignored as non-informative. Formed list of informative structural elements.

The next step in manual mode is to revise 100 randomly selected web resources to study the content of the basic structural elements.

As a result of the research, the structure of the upper level elements is formed. Structural elements that were similar in meaning are merged into thematic groups for displaying.

Conclusion. Based on the structure and content analysis of corporate web resources, the overall structure of corporate web resources is determined. The study results are an integral part of the task of developing a specialized ontology of corporate web resources, taking into account the peculiarities of the domestic market for the development of Internet resources.

Keywords: corporate web-resources, information extraction, web-resource structure, web-navigation, usability.

В.В. Зосімов, канд. техн. наук,
доцент кафедры компьютерных наук и прикладной математики,
Николаевский национальный университет им. В.А. Сухомлинского,
ул. Никольская 24, 54030, г. Николаев, Украина,
zosimovvv@gmail.com

ПОСТРОЕНИЕ ОБЩЕЙ СТРУКТУРЫ КОРПОРАТИВНЫХ ВЕБ-РЕСУРСОВ

Введение. Сегодня достаточно мало внимания уделяется электронному рынку услуг, а именно структурным и семантическим стандартам разработки корпоративных веб-ресурсов. В украинском сегменте Всемирной паутины лишь небольшой процент веб-ресурсов разрабатывается с использованием стандартов семантической разметки.

Следовательно, существует необходимость в разработке комплексной технологии веб-службы, которая будет включать методы, модели и программное обеспечение для создания веб-ресурсов с интегрированной семантической разметкой и получения информации с использованием поисковых агентов.

Чтобы обеспечить эффективную работу технологии, необходимо изучить содержание корпоративных веб-ресурсов, разработать структуру и словарь семантических тегов для описания построенной структуры.

Цель. На основе анализа структуры и содержания корпоративных веб-ресурсов определить общую структуру корпоративных веб-ресурсов.

Методы. Сформирована и описана структура элементов верхнего уровня корпоративных веб-ресурсов. Структурные элементы, похожие по смыслу, объединены в тематические группы для отображения. На основе анализа результатов исследования и общей структуры построенных элементов верхнего уровня, получено дерево структурных элементов корпоративного веб-ресурса.

Результаты. В ходе эксперимента исследована структура 1 тыс. корпоративных веб-ресурсов. Анализ веб-ресурса и извлечения данных осуществлялись с помощью автоматического парсера, реализованного средствами разработанного языка оперирования веб-данными.

С каждого веб-ресурса извлекались значения элементов навигационного меню (всего получено 5371 уникальных элементов). Элементы с числом вхождений менее 50 проигнорированы, как неинформативные. Сформирован список информативных структурных элементов.

В ручном режиме просмотрено 100 выбранных произвольным образом веб-ресурсов из предварительной выборки для исследования информационного содержания базовых структурных элементов.

В результате исследования сформирована структура элементов верхнего уровня.

Выводы. На основе анализа структуры и содержания корпоративных веб-ресурсов определена общая структура корпоративных веб-ресурсов. Результаты исследования — неотъемлемая часть задачи создания специализированной онтологии корпоративных веб-ресурсов с учетом особенностей отечественного рынка разработки Интернет-ресурсов.

Ключевые слова: корпоративные веб-ресурсы, извлечение информации, структура веб-ресурсов, веб-навигация, usability.