

DOI: <https://doi.org/10.32839/2304-5809/2020-9-85-24>

УДК 004.67:[37:502]

Корнієнко Ю.М., Шерман М.І.
Херсонський державний університет

ІНФОРМАЦІЙНО-ДОВІДКОВА СИСТЕМА НАВЧАЛЬНОГО ПРИЗНАЧЕННЯ З БІОЛОГІЇ

Анотація. В статті йде мова про використання інформаційних технологій в освіті, обумовлюються причини використання та їх переваги. Дається уточнене визначення поняття парсингу, його призначення та основні етапи. Також стаття розкриває основні аспекти створення інформаційно-довідкової системи навчального призначення для студентів біологічних факультетів, орієнтованої на вирішення завдань комп'ютерно-інформаційної підготовки майбутніх біологів в умовах сучасної вищої освіти. Розглянуто основні етапи процесу проектування та розроблення інформаційно-довідкової системи, обґрунтування відповідно до функціональних вимог до системи алгоритмів та програмних засобів їх реалізації. Реалізацію проекту було здійснено за допомогою серверної мови програмування PHP та системи керування вмістом Wordpress. Розроблена інформаційно-довідкова система виступає засобом формування професійної компетентності студентів закладів вищої освіти, що надає користувачам можливість здійснювати пошук в системі на тему, яка їх цікавить. Результатом пошуку є перелік статей, які відповідають пошуковому запити. Цей результат є засобом динамічного оновлення навчального матеріалу, передбаченого з біологічних та біолого-технологічних дисциплін, що є складовими професійної підготовки студентів спеціальності «Водні біоресурси та аквакультура», що, в свою чергу, сприяє одержанню суб'єктивно нових знань студентами у процесі навчання.

Ключові слова: інформаційно-довідкова система, ІДС, інформаційні технології в освіті, парсинг, пошук інформації, автоматичний збір даних, PHP, cURL, Simple HTML DOM, WordPress.

Korniienko Yurii, Sherman Mychailo
Kherson State University

THE INFORMATION AND REFERENCE SYSTEM FOR EDUCATIONAL PURPOSES IN BIOLOGY

Summary. The article deals with the usage of information technology in education, stipulates the reasons for its usage and its benefits. Gathering information on the Internet is a time-consuming and routine job that takes a lot of time, but which must be performed to constantly replenish the base of educational materials of the educational institution and maintain the current level of training. The development of an automated information collection system will significantly speed up this process. That is why parsers were developed. A parser is a program or a script that collects and saves the necessary data from web resources and displays it in a structured form. The concept of parsing, its purpose and main stages are defined. The article also describes the main aspects of creating an information and reference system for educational purposes for students of biological faculties, focused on the solution of problems of computer and information training for future biologists in modern higher education. The main stages of the information and reference system development process are considered. Examples of articles that testify to the relevance of research in the field of parsing are given. A feature of the created system is the automatic replenishment of the database of educational materials using a parser program. The developed information and reference system serves as a means of forming the professional competence of students of higher education institutions, which gives users the opportunity to search the system on a topic that interests them. The search result is a list of articles that match the search query. The project was implemented by means of the PHP programming language, cURL libraries and Simple HTML DOM and the Wordpress content management system. Thanks to this resource, users will be able to receive relevant information on a topic that interest them. The system also provides for the possibility of students learning materials by sorting training materials in the right order. The structure of the MySQL database for the informational system and the description of the fields of its main tables were given. The chart of precedents concerning the roles characteristic of the system (the administrator, the editor, the member, the guest), the chart of the sequences in the system, the chart of the sequences of the parser are presented.

Keywords: information and reference system, information technology in education, parsing, information retrieval, automatic data collection, PHP, cURL, Simple HTML DOM, WordPress.

Постановка проблеми. Враховуючи сучасні тенденції розвитку суспільства та технологій, інновації в управлінні навчальним закладом на базі інформаційних технологій є ключовим механізмом, який дозволяє перемагати в конкурентній боротьбі на ринку освітніх послуг та забезпечувати підготовку кваліфікованих фахівців на сучасному рівні [1].

Інформатизація освіти спрямована на підвищення ефективності, якості та доступності освіти. Її основне завдання – впливати на діяльність освітніх закладів так, щоб зменшити затрати ресурсів (часових, матеріальних, фінансових тощо) [2; 6].

Повсюдне використання інформаційних технологій в освітній сфері обумовлюється необхідністю підтримання сучасного високого рівня підготовки спеціалістів, як результат вдосконалення методики навчання.

Використання інформаційних технологій в освіті визначається наступними причинами:

- широкими можливостями інформаційних технологій по індивідуалізації освіти;
- підвищенням мотивації учнів та студентів при використанні інформаційних технологій та посилення емоційного фону освіти;

- надання широкого простору для активної самостійної діяльності учнів та студентів;
- високою наочністю представлення навчального матеріалу;
- все зростаючі інтерактивні можливості інформаційних технологій;
- доступність інформаційних технологій в будь-який зручний час [3].

Використання програмного та технічного забезпечення надає великі можливості в традиційній освіті, одним із яких є електронні освітні ресурси. Саме освітні ресурси являють собою базу для забезпечення високого рівня якості проходження учнями навчального процесу.

Аналіз останніх досліджень і публікацій. Дослідження в області парсингу сайтів є актуальними й донині. Стаття В.А. Ільїна, Д.А. Скоसेлева [4] присвячена методу парсингу вебсторінок, який визначає посилання на певні HTML елементи. В статті Комарова А.В., Меншиков А.А. [5] розглядає метод автоматизованого вилучення адрес з неструктурованих текстів. В роботі Р.М. Яхшісарова [6] описуються методи й інструменти для парсингу даних з вебсайту. Проветов В.Л., Конєва Н.Є. [7] розглядають найбільш поширені методи і засоби агрегації даних, одним з яких є парсинг.

Виділення не вирішених раніше частин загальної проблеми. Збір інформації в мережі Інтернет – це трудомістка та рутинна робота, яка займає багато часу, але яку потрібно виконувати для постійного поповнення бази навчальних матеріалів освітнього закладу та підтримання сучасного рівня підготовки спеціалістів. Розробка системи автоматизованого збору інформації дозволить значно прискорити цей процес [2]. Саме тому були розроблені парсери.

Парсинг – це синтаксичний аналіз сайтів, який виконує спеціальна програма або скрипт. Після цього зібрана інформація відображається в певному вигляді, по певним правилам, або алгоритмам та виконується на одній із мов програмування. Головна особливість парсингу в тому, що аналіз і збір інформації відбувається не з книг, а з вебресурсів [6; 7].

Етапи парсингу даних:

1. збір контенту;
2. вилучення інформації;
3. збереження результатів [7].

Мета статті. Мета дослідження полягає в створенні інформаційно-довідкової системи навчального призначення для студентів біологічних факультетів з автоматичним поповненням бази навчальних матеріалів за допомогою парсингу.

Викладення основного матеріалу дослідження. Реалізація проекту по розробці інформаційно-довідкової системи навчального призначення була здійснена за допомогою серверної мови програмування PHP, бібліотек Simple HTML DOM та cURL, а також системи керування вмістом WordPress. В якості сховища інформації була обрана база даних MySQL.

Процес розробки інформаційно-довідкової системи навчального призначення включає такі етапи:

1. Визначення теми, мети та цілей проекту.
2. Вибір макету інформаційно-довідкової системи з бібліотеки тем системи керування вмістом WordPress.
3. Пошук інформаційного ресурсу, що містить наукові матеріали з біології, який буде використовуватись для автоматизованого поповнення бази навчальних матеріалів інформаційно-довідкової системи.
4. Створення HTML і CSS-файлів.
5. Використання мови програмування PHP, бібліотек Simple HTML DOM та cURL для розроблення парсеру статей з біології.
6. Первинне інформаційне наповнення вебресурсу.
7. Розробка адаптивності сторінки для різного типу пристроїв [8].

В якості макету інформаційно-довідкового макету була обрана тема University Hub, автором якої є WEN Themes.

Даний макет був переоформлений відповідно до потреб системи. Були вказані дані про університет, його ректора, завідувача кафедри, контактні дані, логотип тощо.

Структура інформаційно-довідкової системи складається з п'яти вкладинок:



Рис. 1. Макет інформаційно-довідкової системи

- головна сторінка;
- сторінка з навчальними дисциплінами;
- сторінка з переліком додаткових матеріалів, отриманих за допомогою програми-парсера та посилання на них;
- галерея;
- сторінка з контактними даними, картою та формою зворотного зв'язку.

Для того, щоб користуватися розробленою системою не потрібно мати ніяких спеціальних навичок, достатньо лише володіти базовими навичками роботи з персональним комп'ютером.

В зазначеній інформаційно-довідковій системі присутні чотири ролі користувачів:

1. Адміністратор – має можливість видалення або створення користувачів, редагування консолі, файлів, сторінок, записів, приватних сторінок, опублікованих сторінок і записів, редагування тем і так далі.

2. Редактор – має можливість створення та редагування контенту, редагування категорій та посилань, модерування коментарів. Також має можливість редагування чужих записів, сторінок, приватних записів і опублікованих сторінок.

3. Учасник – зареєстрований користувач системи. Має можливість створення нового контенту. Але не може публікувати його напряму, оскільки публікація повинна бути виконана редактором після того, як він перевірить роботу учасника. Також має змогу коментувати записи.

4. Гість – незареєстрований користувач системи. Його можливості дуже обмежені. Може лише переглядати публічні записи та сторінки ресурсу.

На наступному рисунку подано діаграму прецедентів інформаційно-довідкової системи.

База даних, яка використовувалась при створенні інформаційно-довідкової системи, містить в собі більш ніж дванадцять таблиць. Далі наведено список та характеристику основних таблиць бази даних розроблюваної системи:

– wp_comments – ця таблиця містить всі коментарі та їх дані, залишені на сайті.

– wp_options – в цій таблиці зберігаються всі налаштування системи. Деякі плагіни додають свої налаштування в цю таблицю.

– wp_posts – в цій таблиці зберігаються всі параметри записів вебсайту. В ній знаходяться автор, дату створення, дата публікації, заголовок, посилання, статус коментування тощо.

– wp_terms – ця таблиця містить назви, слаги (slug) та кількість записів для категорій і тегів.

– wp_users – ця таблиця містить дані облікових записів користувачів.

На наступному рисунку подано діаграму послідовностей розроблюваної інформаційно-довідкової системи [9].

На головній сторінці створеної системи відображається календар, а також список останніх новин.

На сторінці додаткових матеріалів відображається список статей, коротка інформація по кожній із них, а також посилання на повну версію. Відображення статей відбувається за допомогою запиту до бази даних. Після цього формується вибірка із 10 останніх статей та відображається на сторінці. Внизу сторінки присутнє меню для навігації по всім статтям, які зберігаються в базі даних.

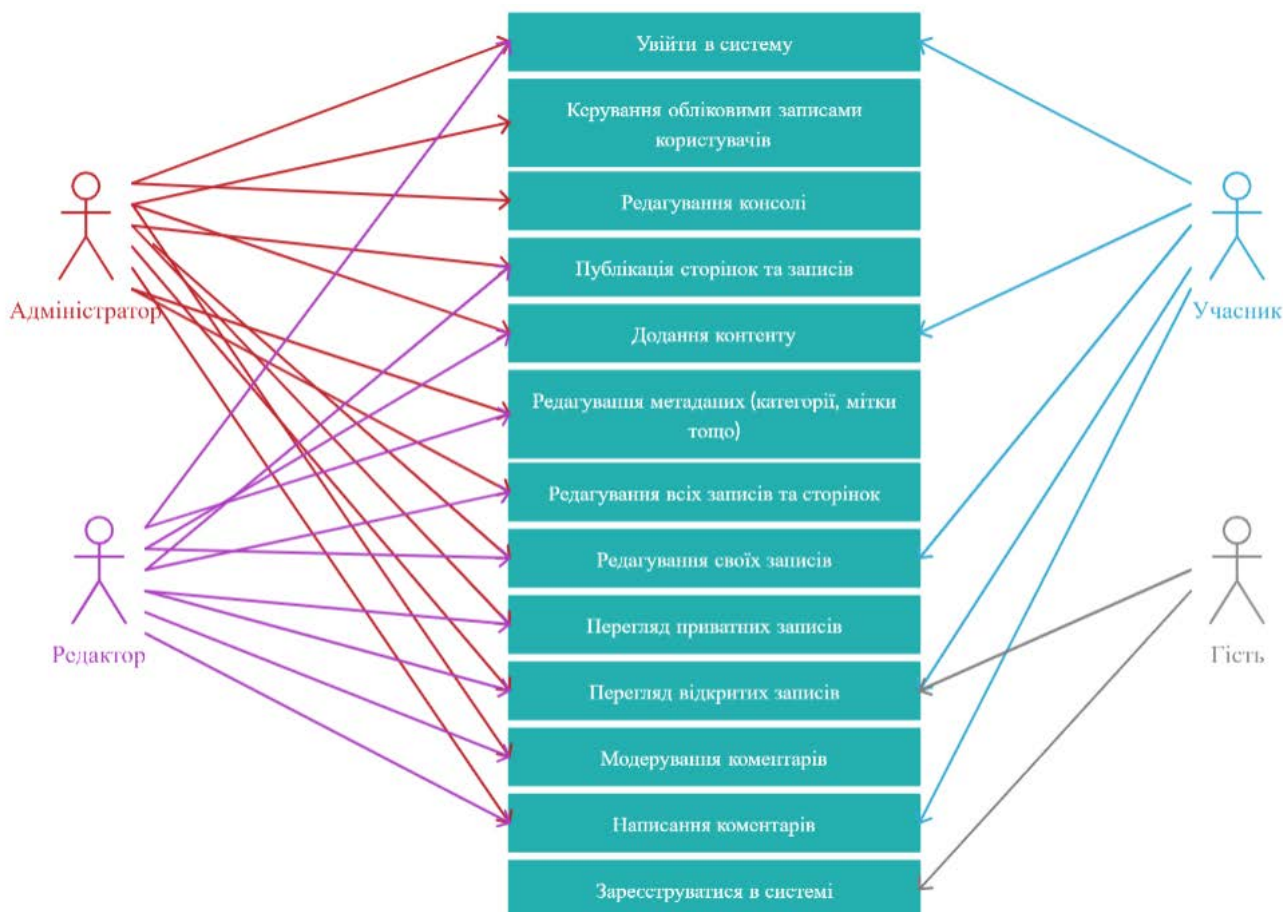


Рис. 2. Діаграма прецедентів системи

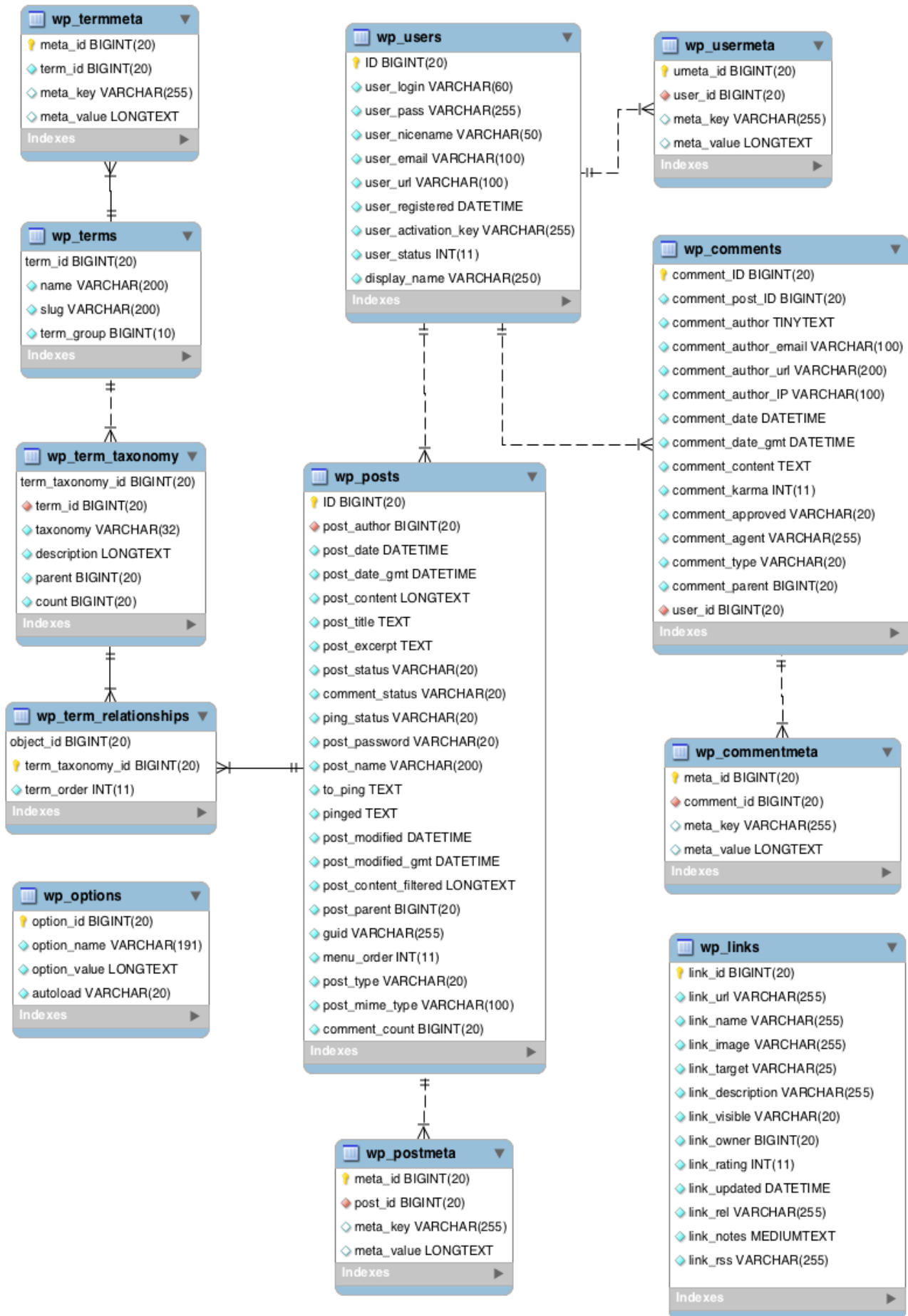


Рис. 3. Структура бази даних системи

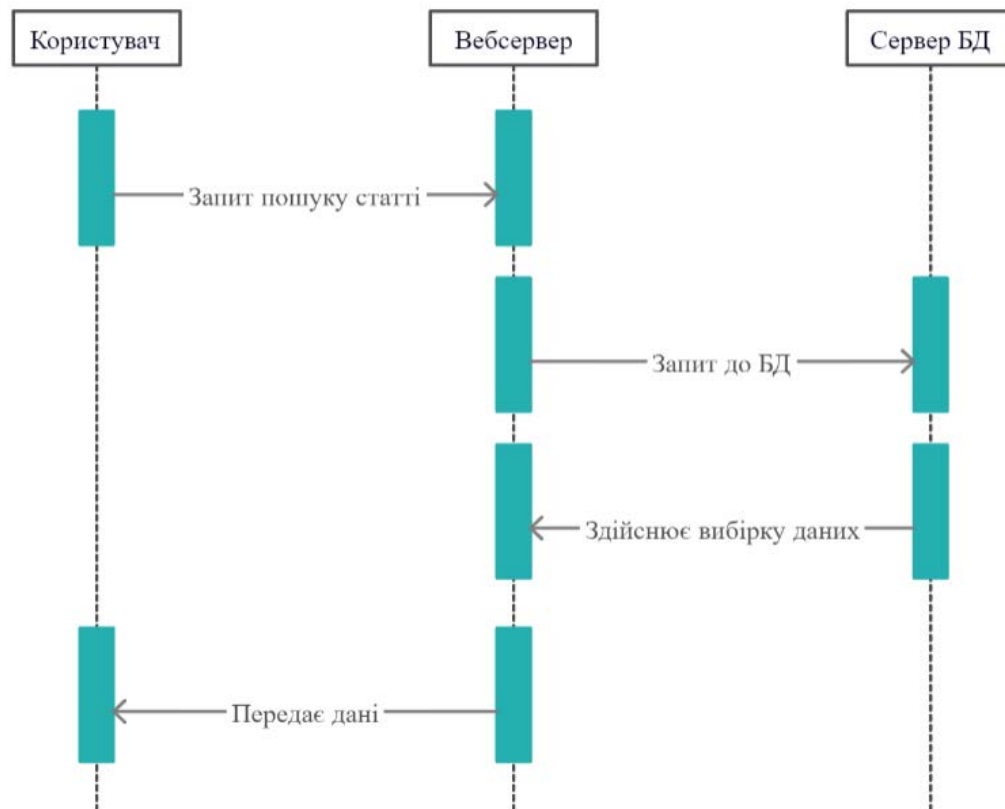


Рис. 4. Діаграма послідовностей системи

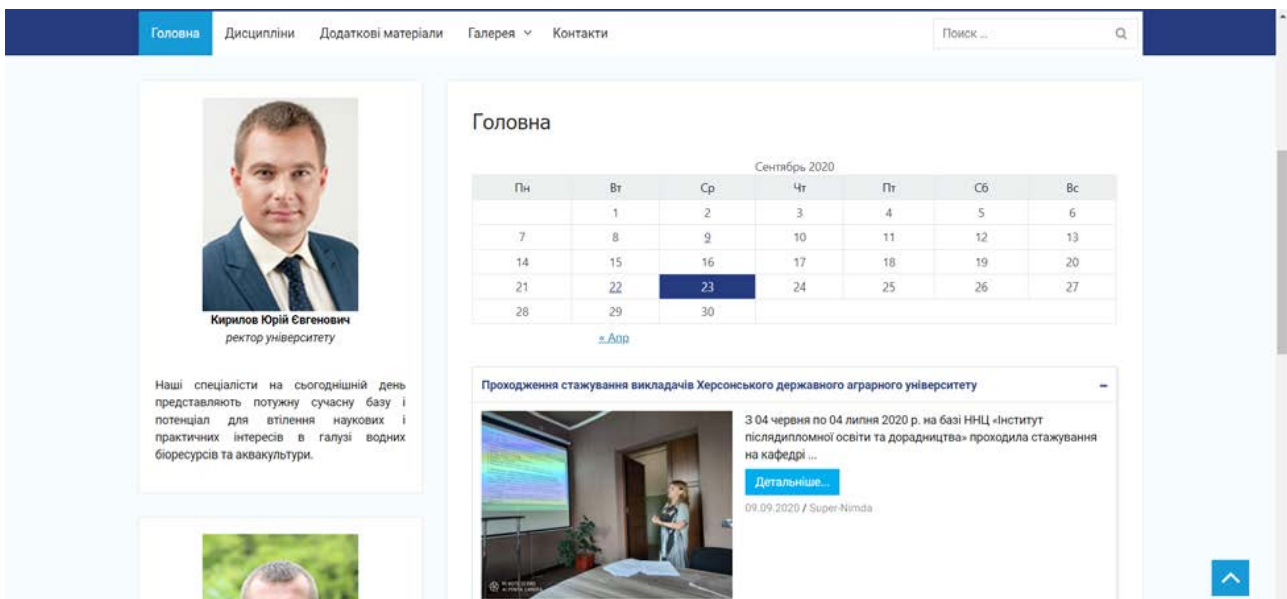


Рис. 5. Головна сторінка вебсистеми

Поповнення бази даних інформаційно-довідкової системи новими статтями відбувається за допомогою розробленої програми-парсера. Ця програма спочатку робить запит до вебресурсу, отримує від нього HTML сторінку, потім за допомогою бібліотеки Simple HTML DOM перебирає кожну статтю, виділяє важливу інформацію та структурує її по заданому алгоритму. Після цього отримана інформація додається до бази даних шляхом створення окремого запису для кожної статті за допомогою вбудованих функцій системи керування вмістом WordPress.

На наступному рисунку подано діаграму послідовностей для програми-парсера.

Далі наведено детальний алгоритм роботи програми-персера:

1. Зробити запит на сторонній вебсервер по протоколу https за допомогою бібліотеки cURL та передати всі необхідні заголовки.

2. Отримати від стороннього вебсерверу HTML сторінку та за допомогою бібліотеки Simple HTML DOM створити об'єкт зі всіма елементами DOM дерева, до яких можна швидко отримати доступ за допомогою вбудованих функцій.

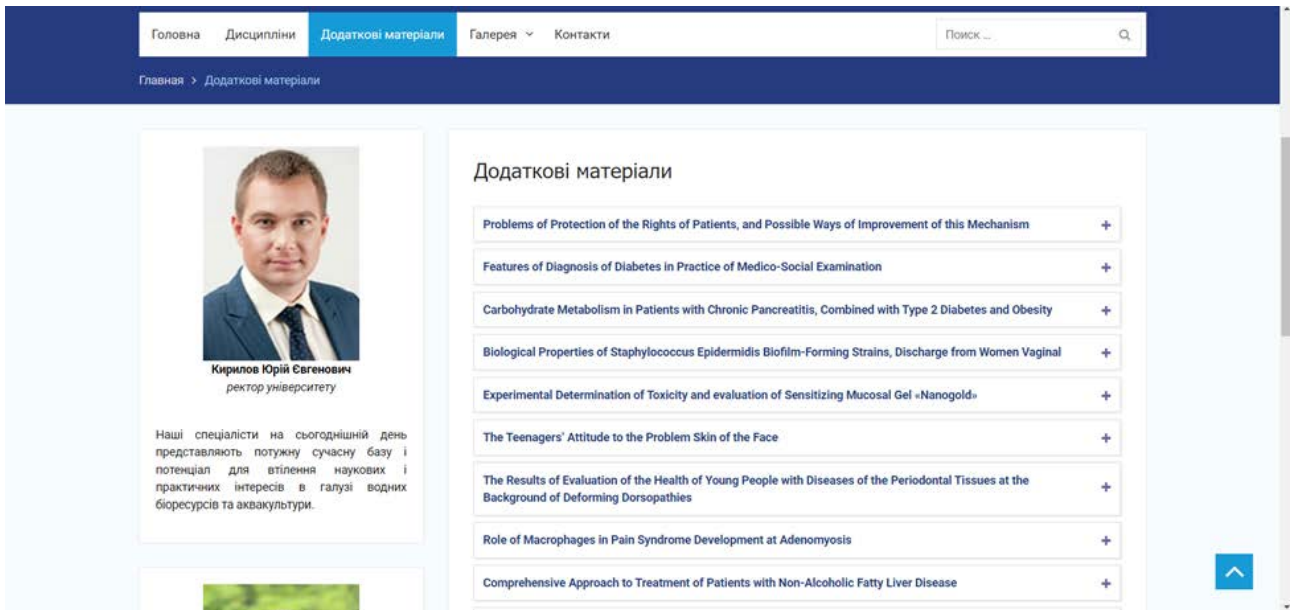


Рис. 6. Сторінка з додатковими матеріалами вебсистеми

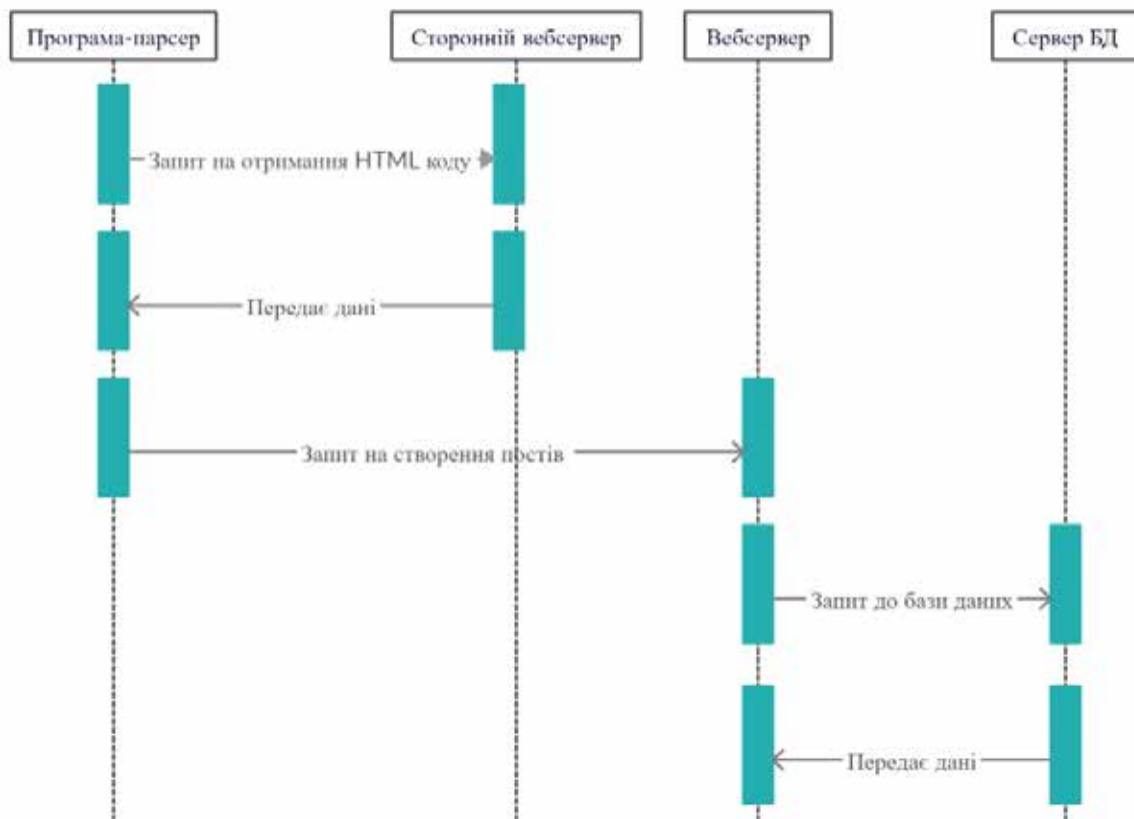


Рис. 7. Діаграма послідовностей програми-парсеру

3. В створеному об'єкті знайти всі посилання на статті та зберегти їх.

4. Обійти всі збережені посилання, вилучити необхідну інформацію про статті та зберегти цю інформацію в масив.

5. Перебрати масив статей та за допомогою вбудованих функцій WordPress створити для кожної статті окремий запис в системі: за допомогою функції `wp_insert_post()` створити новий запис, а за допомогою функції `wp_set_object_terms()` присвоїти кожному запису від-

повідну категорію для зручної навігації по статтям.

6. Попередньо в системі було створено новий шаблон запису для статей, в якому використовується спеціальний плагін `Advanced Custom Fields`, за допомогою якого було створено власні поля для запису, які не передбачені стандартними можливостями WordPress. Для заповнення таких полів використовувалася функція `update_field()`.

Після того, як парсер завершить свою роботу, в спеціальній таблиці бази даних будуть створені

нові записи зі статтями, які, в свою чергу, будуть відображені на сайті в посторінковому вигляді, як це було показано на рис. 6.

Висновки і пропозиції. Під час дослідження було створено інформаційно-довідкову систему навчального призначення, яка орієнтована на вирішення завдань по комп'ютерно-інформаційній підготовці майбутніх біологів.

Особливістю створеної системи є автоматичне поповнення бази навчальних матеріалів за допомогою програми-парсеру. Парсер – це програма або скрипт, яка виконує збір та збереження необхідних даних з вебресурсів та відображає їх в структурованому вигляді. Програма-парсер була реалізована за допомогою серверної мови програмування PHP, бібліотек cURL та Simple HTML DOM.

Було визначено поняття парсингу, його призначення та основні етапи. Приведено приклади статей, що свідчать про актуальність дослідження в області парсингу.

Розроблена інформаційно-довідкова система виступає засобом формування професійної компетентності студентів закладів вищої освіти, яка надає користувачам можливість здійснювати пошук в системі на тему, яка їх цікавить. Результатом пошуку є перелік статей, які відповідають пошуковому запиту.

Також під час дослідження було наведено структуру бази даних розробленої системи, описані та охарактеризовані її основні таблиці. Представлено макет системи, діаграму прецедентів в системі, діаграму послідовностей системи, діаграму послідовностей програми-парсеру, наведені зображення сторінок розробленої системи.

Подальшого дослідження потребують питання розробки предметно-орієнтованого програмного забезпечення та навчально-методичних ресурсів, спрямованих на підвищення якості професійної підготовки студентів біологічних спеціальностей в умовах сучасної освіти.

Список літератури:

1. Вдовін І.В. Актуальні питання автоматичного вилучення даних з веб-сторінок. *Перспективи розвитку інформаційних технологій*. Новосибірськ, 2015. С. 11–16.
2. Абрамова Т.А. Розробка парсинг-системи для отримання прихованих посилань зі сторінок соціальних мереж. *Вісник ПензГУ*. 2016. № 3. С. 22–31.
3. Вдовін І.В. Застосування технології Web Mining до вилучення наукових даних в мережі інтернет. *ОмГТУ*. 2015. С. 10–14.
4. Ильин В.А., Скоселев Д.А. Парсинг веб-сайтов с использованием шаблонов. *Мир современной науки*. 2018. № 2. С. 8–12.
5. Комарова А.В., Менщиков А.А. Метод автоматизованого вилучення адрес з неструктурованих текстів. *International Journal of Open Information Technologies*. 2017. № 11. С. 43–45.
6. Яхшисарова Р.М. Программная реализация парсинга данных раздела поиска недвижимости сайта avito.ru. *Математическое моделирование процессов и систем*: материалы VII Международной молодежной научно-практической конференции (Стерлитамак, 30 октября – 1 ноября 2019 г.). Стерлитамак, 2019. С. 437–441.
7. Просветов В.Л., Конева Н.Е. Анализ методов и средств автоматизации процессов обработки данных веб-сайтов. *Евразийское Научное Объединение*. 2019. № 1. С. 89–94.
8. Шерман М.І., Самчинська Я.Б. Інформаційно-довідкова система рідкісних та зникаючих видів тварин як засіб формування цифрової компетентності майбутніх екологів. *Інформаційні технології і засоби навчання*. 2019. № 4. С. 121–135.
9. Шерман М.І., Степаненко Н.В. Інформаційно-комунікаційні технології у професійній підготовці майбутніх екологів в умовах аграрного університету. *Нові технології навчання: Наук.-метод. зб.* 2008. № 54. С. 7–10.

References:

1. Vdovin, I.V. (2015). Aktual'ni pytannya avtomatychnoho vyluchennya danykh z veb-storinok [Current issues of automatic data extraction from web pages]. *Prospects for the development of information technology*, pp. 11–16.
2. Abramova, T.A. (2016). Rozrobka parsynh-systemy dlya otrymannya prykhovanykh posylan' zi storinok sotsial'nykh merezh [Development of a parsing system to obtain hidden links from social media pages]. *The Penza State University Herald*, vol. 3, pp. 22–31.
3. Vdovin, I.V. (2015). Zastosuvannya tekhnolohiyi Web Mining do vyluchennya naukovykh danykh v merezhi internet [Application of Web Mining technology to the extraction of scientific data on the Internet]. *The Omsk State Technical University Herald*, pp. 10–14.
4. Yl'yn, V.A., & Skoselev, D.A. (2018). Parsynh veb-saytov s yspol'zovanyem shablonov [Parsing websites using templates]. *The world of modern science*, vol. 2, pp. 8–12.
5. Komarova, A.V., & Menshchikov, A.A. (2017). Metod avtomatyzovanoho vyluchennya adres z nestrukturnovanykh tekstiv [Method of automated extraction of addresses from unstructured texts]. *International Journal of Open Information Technologies*, vol. 11, p. 43–45.
6. Yakhshysarova, R.M. (2019). Prohrammnaya realizatsyya parsynha danykh razdela poyska nedvyzhymosti sayta avito.ru [Software implementation of data parsing of the real estate search section of the avito.ru website]. *Proceedings of the Mathematical modeling of processes and systems: materials of the VII International youth scientific-practical conference (Russia, Sterlitamak, October 30 – November 1, 2019)*. Sterlitamak, pp. 437–441.
7. Prosvetov, V.L., & Koneva, N.E. (2019). Analiz metodov y sredstv avtomatyzatsyy protsessov obrabotky danykh veb-saytov [Analysis of methods and tools for automating the processing of website data]. *Eurasian Scientific Association*, vol. 1, pp. 89–94.
8. Sherman, M.I., & Samchyns'ka, Ya.B. (2019). Informatsiyno-dovidkova sistema ridkisnykh ta znykayuchykh vydiv tvaryn yak zasib formuvannya tsyfrovoyi kompetentnosti maybutnikh ekolohiv [The information and reference system on rare and endangered species of animals as a computer tool for the formation of future ecologists' digital competency]. *Information technologies and teaching aids*, vol. 4, pp. 121–135.
9. Sherman, M.I., & Stepanenko, N.V. (2008). Informatsiyno-komunikatsiyni tekhnolohiyi u profesiyinyi pidhotovtsi maybutnikh ekolohiv v umovakh ahrarnoho universytetu [Information and communication technologies in the professional training of future ecologists in the conditions of agrarian university]. *New learning technologies: Scientific and methodical collection*, vol. 54, pp. 7–10.