

Інформаційні технології

УДК 004

Степаніщева Валерія Сергіївна

студент

*Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»*

Степанищева Валерия Сергеевна

студент

*Национального технического университета Украины
«Киевский политехнический институт имени Игоря Сикорского»*

Stepanishcheva Valeriia

Student of the

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

**ПОБУДОВА БАЗИ ЗНАНЬ ДЛЯ СИСТЕМИ ДИСТАНЦІЙНОГО
НАВЧАННЯ**

**ПОСТРОЕНИЕ БАЗЫ ЗНАНИЙ ДЛЯ СИСТЕМЫ
ДИСТАНЦИОННОГО ОБУЧЕНИЯ**

**BUILDING A KNOWLEDGE BASE FOR A DISTANCE LEARNING
SYSTEM**

***Анотація.** У дослідженні проаналізовано теорію створення бази знань для дистанційного навчання на основі онтології, реалізовано опис сутностей та зв'язків між ними, описано та створено методи заповнення бази.*

***Ключові слова:** система дистанційного навчання, база знань, представлення знань, онтологія.*

Аннотация. В исследовании проанализированы методы создания и реализации рекомендательных систем. Приведено сравнение результатов работы различных методов. Предложены способы внедрения рекомендательных алгоритмов в систему дистанционного обучения.

Ключевые слова: система дистанционного обучения, база знаний, представление знаний, онтология.

Summary. The research carried out an analysis of the methods of creating and implementing recommender systems. The comparison of the results of various methods is given. The methods of implementation of advisory algorithms in the system of distance learning are offered.

Key words: distance learning system, knowledge base, knowledge representation, ontology.

Постановка проблеми. Однією з рушійних сил людства була саме освіта. Проте для тих, хто хоче опанувати нові знання є дуже багато рішень – книжки, відеоуроки, сайти з курсами. Але все ж таки тут не все ідеально. Що робити людині, якщо вона не знає напевно, що їй вчити, де знайти те, що зараз актуально? Саме на ці питання потрібно звернути увагу, створюючи систему дистанційного навчання.

Виклад основного матеріалу. Загальна мета створення бази знань заключається в тому, щоб представити знання предметної області. Це має бути структура, що буде втілювати в собі зручний інструмент для зберігання даних, їхнього накопичення та проведення маніпуляцій над ними.

Існує багато моделей представлення знань, найпопулярніші з них це: продукційні моделі, семантичні мережі, фрейми та онтології [1].

Продукційні моделі та фрейми мають занадто просту структуру. Семантичні мережі навпаки важко структуровані. Онтології ж поєднують в

собі переваги усіх методів: чітке представлення об'єктів та можливість реалізовувати складні зв'язки [2].

Онтологія визначається як $O = \langle X, R, F \rangle$, де X – кінцева множина понять предметної області, R – кінцева множина відносин між поняттями, F – кінцева множина функцій інтерпретації [3].

Онтологія буде реалізована за допомогою створення кількох типів сутностей та зв'язків між ними.

Тож, оглянемо функціонал, який надає База знань.

Сутність «Категорія».

База знань надаватиме доступ до інформації про категорії, за якими відбувається навчання. Існує дев'ять категорій: Веб розробка, Мобільні додатки, Мови програмування, Розробка ігор, Бази даних, Тестування програмного забезпечення, Розробка програмного забезпечення, Інструменти розробки, Електронна комерція.

Сутність Категорія має декілька властивостей (рис 3.1): назва категорії та ключові слова.

Ключові слова Категорії мають критерії, що виражають ступінь того, наскільки дане ключове слово характеризує категорію. Це: загальна частота зустрічання слова в метаданих, кількість вакансій з цим словом та частота згадування в курсах. На рис. 1 показаний зв'язок Категорії та її ключових слів.

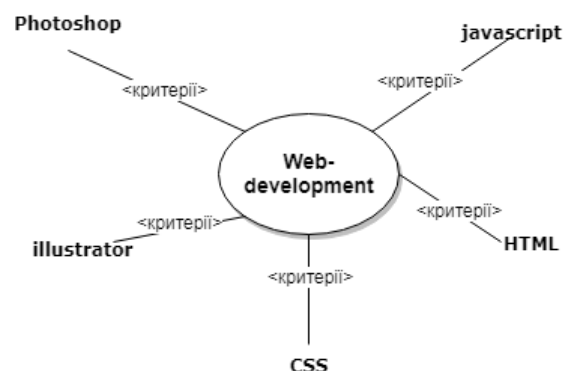


Рис. 1. Відношення між Категорією та її ключовими словами

Сутність «Курс».

Наступна сутність, яку необхідно розглянути, - це Курс. Курс є об'єктом, що представлений на сайтах з курсами дистанційного навчання.

Курс має такі властивості: назва, посилання, ключові слова, рівень, мова, мова субтитрів, ціна, тривалість, рейтинг та особливості.

Кожен Курс має список, пов'язаних з ним ключових слів (рис. 2). Дані ключові слова не мають критеріїв схожості, адже вони лише спрощують пошук курсів та представляють собою опис курсу.

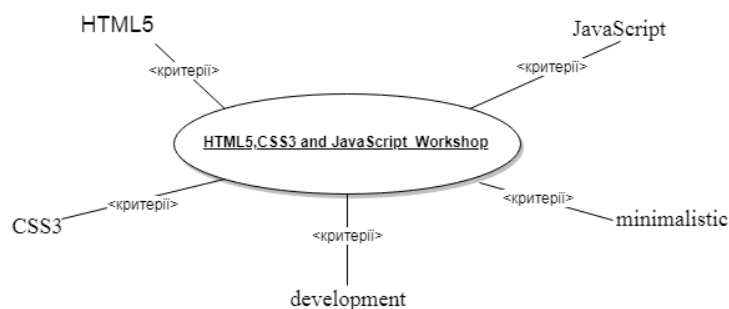


Рис. 2. Зв'язок Курсу та його ключових слів

Сутність «Користувач».

Користувач має такі властивості: юзернейм, стать, дата народження, рід діяльності, улюблені категорії та рівень користувача.

Перейдемо до зв'язків між сутностями.

Зв'язки між Категоріями.

Категорії, як і у реальному житті, пов'язані між собою (рис. 3), та мають критерії схожості.

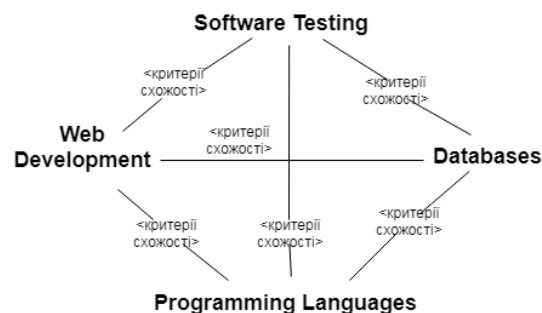


Рис. 3. Схема зв'язку Категорій

Значення критеріїв схожості обраховуються як кількість спільних ключових слів з вибірок розміром N.

Зв'язки Категорії та Курсів.

Категорії та Курси також мають зв'язки, адже будь-який курс відноситься до певної категорії або ж до декількох (рис 4).

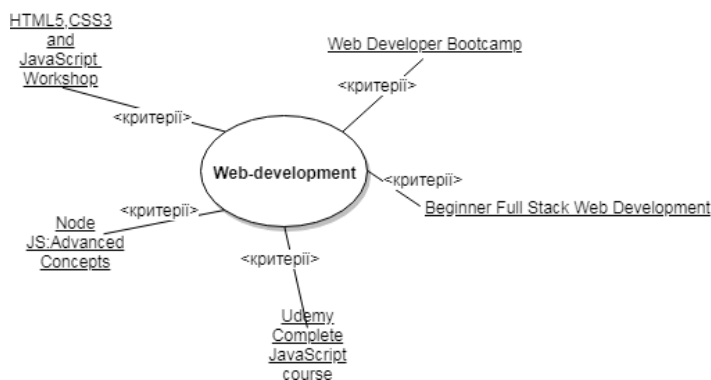


Рис. 4. Схема зв'язків Категорії та Курсів, що відносяться до неї

Зв'язки мають критерії пов'язаності, які показують, наскільки той чи інший курс відноситься до цієї категорії.

Критеріїв у кожного зв'язку три: рейтинг, кількість користувачів, які проходять курс та частота рекомендування цього курсу користувачам.

Зв'язки Користувача та Категорій.

Користувач має зв'язки з Категоріями (рис. 5).

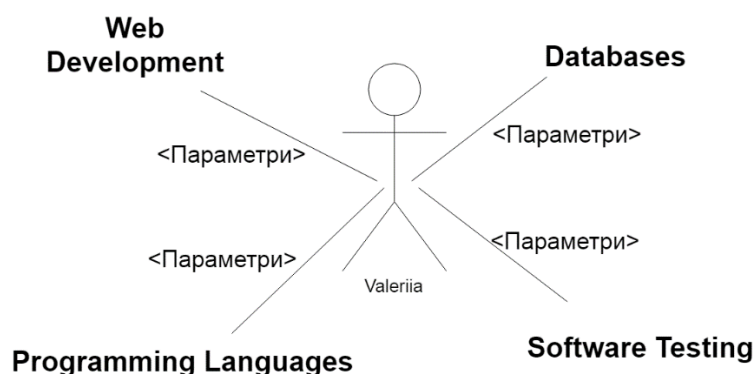


Рис. 5. Структура зв'язків між Користувачем та Категоріями

На відміну від попередніх зв'язків цей не має чітких критеріїв пов'язаності, проте тут присутні деякі параметри: кількість пройдених

Курсів з цієї Категорії, оцінка Користувача цієї Категорії та середня оцінка за усі Курси Користувача з цієї Категорії.

Зв'язки Користувача та Курсів.

Зважаючи на те, що Користувачі проходять Курси, мають бути і зв'язки між ними (рис. 6).

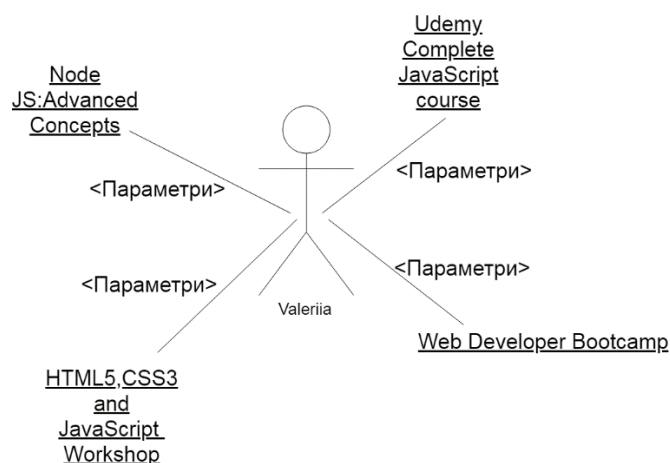


Рис. 6. Зв'язки Користувача та Курсів

За аналогією зі зв'язками Користувач – Категорія, даний тип зв'язків також не має чітких критеріїв пов'язаності – лише певні параметри, які визначають так звані організаційні моменти.

Кожен зв'язок має такі параметри: Оцінка, Дата початку проходження даного Курсу та Статус кінця.

Перейдемо до заповнення бази знань.

Заповнення бази знань відбувається кількома шляхами. Один з них – через графічний інтерфейс користувача. Відповідно, інформацію цим способом надає Користувач. На сторінці реєстрації користувач вводить основні свої дані, що є параметрами сутності Користувач. В особистому кабінеті Користувач може змінювати особисту інформацію та виставляти оцінки категоріям. Ці оцінки характеризуватимуть рівень цікавості цієї чи іншої категорії для користувача. На сторінці курсів «Мої курси» користувач може виставляти рейтинги курсам, які він пройшов.

Пошук ключових слів.

Ключові слова характеризують Курси та Категорії. Їх треба якось знаходити та додавати в базу знань. Це відбуватиметься за допомогою парсингу сайту з вакансіями. Адже там є великий об'єм інформації, що містить актуальні описи вакансій – те, що потрібно на трудовому ринку зараз. Система робить парсинг сайту вакансій в тих випадках, коли не були знайдені ключові слова за запитом або їх знайдено замало.

Парсинг здійснюється по сайту robo.ua.

Спочатку система зберігає посилання на вакансії, потім отримує їхній опис та видаляє слова на кшталт «and», «with», «the» тощо.

В результаті ми отримуємо список ключових слів для даного запиту. За схожими або вже існуючими словами в сутності ключових слів, отримані слова записуються по категоріях, обраховуються критерії.

Отримання Курсів.

Було вирішено брати курси з сайту Udemy.com. Для цього був надісланий запит адміністрації сайту з проханням надати доступ до API клієнта сайту. Тож, Курси будуть отримуватись не шляхом парсингу, а шляхом надсилання GET запитів.

Для того, щоб отримати список курсів потрібен запит вигляду: GET /api-2.0/courses/ПАРАМЕТРИ ЗАПИТУ

Параметри запиту можуть бути такими: номер сторінки результатів пошуку, кількість курсів на одній сторінці, пошуковий запит, назва категорії, назва підкатегорії, ціна, мова, рівень курсів, дані про те, чи є вправи на написання коду, значення того, як впорядковувати курси, мінімальний рейтинг курсів в результаті та тривалість курсів.

Результат отримується у вигляді JSON рядку, який з легкістю розпізнається та дані записуються в сутність Курс.

Висновки. Метою створення бази знань є інтелектуалізація Системи дистанційного навчання.

Проаналізувавши усі моделі представлення знань, було зроблено висновок, що для бази знань в СДН найкращою моделлю представлення знань є онтологія. Завдяки чітко вираженим сутностям та гнучкій системі зв'язків можна з легкістю представити будь-яку предметну область у вигляді онтології.

Визначено, що онтологія для СДН має складатись з кількох типів сутностей, які матимуть велику кількість екземплярів, що пов'язані між собою за допомогою зв'язків. До того ж, зв'язки мають певні ваги пов'язаності однієї сутності з іншою. визначено такі основні сутності: Категорія курсів, Курс, Користувач. Відповідно, між собою усі сутності пов'язані різними типами зв'язків, ваги пов'язаності яких були описані.

Наведені способи заповнення бази знань системи ДН. Описано три методи, якими наповнюється база знань: за допомогою інтерфейсу, за допомогою автоматичного витягнення даних з інших джерел, а саме парсинг сайту з вакансіями та аналіз описів вакансій для отримання ключових слів та витягання даних про курси з сайту Udemu.com.

Наприкінці роботи було проведено тестування модулів бази знань, а саме: пошуку курсів за запитом користувача з використанням певних фільтрів, пошуку ключових слів за допомогою парсингу та аналізу тексту вакансій, та отримання курсів з сайту Udemu.com.

Література

1. Моделі представлення знань [Електронний ресурс] // Вікіпедія. Вільна енциклопедія – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%9C%D0%BE%D0%B4%D0%B5%D0%BB%D1%96_%D0%BF%D1%80%D0%B5%D0%B4%D1%81%D1%8

2%D0%B0%D0%B2%D0%BB%D0%B5%D0%BD%D0%BD%D1%8F_
%D0%B7%D0%BD%D0%B0%D0%BD%D1%8C. Дата доступу:
19.05.2018.

2. Семантична мережа [Електронний ресурс] // Енциклопедія Вікіпедія – Режим доступу: Семантична мережа [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%BD%D0%B0_%D0%BC%D0%B5%D1%80%D0%B5%D0%B6%D0%B0. – Дата доступу: 19.05.2018.
3. Онтологія [Електронний ресурс] // Енциклопедія Вікіпедія – Режим доступу: [http://ru.science.wikia.com / wiki / Онтологія](http://ru.science.wikia.com/wiki/Онтологія). – Дата доступу: 19.05.2018.