

**Кравчук Є.С.**

Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»

**Сергєєв-Горчинський О.О**

Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»

## АНАЛІЗ ВЗАЄМОДІЇ ЗМІСТУ НАВЧАЛЬНИХ ДИСЦИПЛІН ЗА ЇХ СЕМАНТИЧНИМ ОПИСОМ

*Метою роботи є дослідження підходів до опису та побудови онтологій, зокрема, у домені навчальних дисциплін та методів автоматизації формування доменних знань із корпусу документів. Досліджено можливості застосування онтологій для представлення доменної області навчальних дисциплін і курсів. Описано поняття онтології в контексті комп'ютерних наук та семантичного вебу, розглянуто наявні онтології навчальних дисциплін, що використовуються у світі, та виявлено слабкі сторони ручного створення та підтримки онтологій. Проведено аналіз методів і підходів, що автоматизують процес побудови онтологій. Розглянуто обмеження, які постають під час вирішення таких задач як із теоретичної сторони, так і в контексті практичних аспектів побудови програмних додатків, що реалізують такі моделі. Досліджено сучасні інноваційні методи автоматичної побудови графа знань у межах доменної області.*

**Ключові слова:** онтологія, навчальна дисципліна, граф знань, семантичний веб, семантичний пошук.

**Постановка проблеми.** У наш час із ростом кількості інформації дуже гостро постає проблема структуризації накопичених даних для отримання корисної інформації з них, зокрема у сфері навчання. Стрімкий ріст ринку масових онлайн-курсів та поява нових дисциплін і напрямів навчання робить таку проблему актуальною і в даній сфері.

Наявні засоби опису онтологій навчальних дисциплін і курсів вже не можуть повною мірою відповідати вимогам до швидкості створення, підтримки й обслуговування. Отже, тема набуває особливої актуальності в контексті опису навчальних дисциплін зокрема та представлення доменних знань загалом.

З огляду на стрімку автоматизацію навчальних процесів, з'являється потреба в зборі, агрегації та пов'язуванні навчальних дисциплін та матеріалів в одну систему. Розроблення методики, що охоплює ці процеси, дозволить створювати розподілені системи дистанційного навчання і використовувати різні навчальні матеріали.

Сьогодні для дистанційного навчання використовуються курси, які містять інформаційний документ із переліком основних понять та пред-

метів і посилань на літературу. Основною проблемою розвитку дистанційного навчання є «механічність» розроблення освітніх курсів, оскільки створення освітнього курсу зазвичай супроводжується переносом матеріалу лекції вручну в програмний додаток. У даний час проблемою є ефективний опис та організація наявного контенту для полегшення обміну, повторного використання та модифікації.

**Аналіз основних досліджень і публікацій.** Семантичні мережі вирішують цю проблему. Семантична мережа – область досліджень штучного інтелекту, що має на меті створення метаданих. Основний виклик семантичного вебу – представити сутність контенту в зрозумілій для всіх користувачів формі.

«Онтологія» – термін, запозичений із філософії, який означає науку, що описує форми буття і те, як вони співвідносяться між собою. Інакше кажучи, онтологія – явний опис концептуалізації [1]. Веб-онтологія може включати опис класів, їхніх властивостей, а також індивідів класів. Онтології грають критично важливу роль в організації оброблення знань на базі веб, їх спільному використанні й обміні ними між додатками.

Онтології в загальному вигляді визначаються як спільно використовувані формальні концептуалізації конкретних предметних областей у вигляді тверджень, якими можуть обмінюватися люди і програмні додатки.

Компонентами онтології є:

– множина понять предметної області та їхніх атрибутів;

– множина відношень або асоціацій між виділеними поняттями;

– множина аксіом і правил виведення, заданих на виділених поняттях.

Консорціумом World Wide Web Consortium (далі – W3C) розроблені два основні стандарти для представлення онтологій, а саме: Resource Description Framework Schema (далі – RDFS) і мова Web Ontology Language (далі – OWL). RDFS і OWL є розширеннями моделі загального призначення Resource Description Framework (RDF) для представлення даних в Інтернеті.

RDF базується на XML, що дозволяє представляти твердження про ресурси у вигляді триплетів «суб'єкт – предикат – об'єкт», які визначають відношення між довільними суб'єктом і об'єктом за допомогою предиката (дієслова чи дієслівного звороту). Триплети придатні для машинного оброблення й інтелектуального аналізу даних (Data Mining) [2].

Мова OWL призначена для запису семантики предметних областей у вигляді онтологій. OWL стандартизує подання знань у веб, шляхом представлення дійсності в моделі даних «об'єкт – властивість» [2].

OWL придатна для опису не тільки веб-сторінок, але й будь-яких об'єктів дійсності. Кожному елементу опису в даній мові (зокрема, властивості, що зв'язує об'єкти) відповідає уніфікований ідентифікатор ресурсу (Uniform Resource Identifier, URI). OWL дозволяє описувати значення термінів у словниках і відношення між ними, наслідування класів, метахарактеристики, потужності зв'язків та еквівалентності.

**Постановка завдання.** Пошук та огляд наявних стандартів і специфікацій, присвячених формалізації процесів створення, зберігання й опису онтологій у домені навчальних дисциплін та пов'язаних із ними сферах, дозволили виокремити специфікації та стандарти онтологій, які широко використовуються у світі для опису доменних моделей навчальних курсів.

1. Learning Object Metadata Ontology (далі – LOM) – онтологія та словник, що застосовуються для представлення стандарту метаданих для

освітнього контенту IEEE LOM. Виконує функцію моста для зв'язування навчальних метаданих із Linked Open Data.

2. Academic Institution Internal Structure Ontology (далі – AIISO) – онтологія та словник, що надають класи та властивості для опису курсів, модулів, практичних і теоретичних навчальних матеріалів академічної установи [3].

3. Бібліографічна онтологія (Bibliographic Ontology (далі – BIBO)) описує бібліографічну інформацію в семантичній павутині у форматі RDF. Дана онтологія може використовуватися як антологія цитування, як онтологія класифікації документів чи як спосіб опису будь-якого документа у RDF. Ця специфікація служить документом простору імен «Бібліографічна онтологія». По суті, вона описує онтологію (класи та властивості) і терміни, з яких вона складається, щоб семантичні веб-додатки могли використовувати ці терміни в різних форматах, сумісних із RDF.

4. Linked Science Core Vocabulary Specification (далі – LSC) – науковий словник призначений для опису наукових ресурсів, серед яких елементи дослідження, їхній контекст та взаємозв'язок. LSC є зразком «будівельних блоків» для опису зв'язків між дисциплінами, зрозумілим для комп'ютера чином. LSC фокусується на властивостях, які можна використати для опису змісту дослідницької роботи, тобто для зіставлення досліджень, гіпотез, експериментів, даних та публікацій. LSC визначає тільки основні терміни для науки. Більш конкретні терміни, необхідні для різних наукових спільнот, можна записати як розширення LSC.

5. Learning Resource Metadata Initiative (далі – LRMI) – ініціатива, що має на меті зробити легшим пошук освітніх матеріалів через пошукові та спеціалізовані сервіси викладачів і учнів. Підходом, що лежить в основі LRMI, є розширення онтології schema.org таким чином, щоб можна було виражати важливі освітні характеристики та відношення [4].

6. Teaching Core Vocabulary Specification (далі – TEACH) – легкий словник основ викладання, який містить терміни, що дозволяють викладачам пов'язувати свої курси разом. Основний навчальний словник заснований на практичних вимогах, пред'явлених до проведення семінарів та опису курсів як зв'язаних даних [5].

7. Semantic units (далі – SemUnit) – проект, ініційований французькими вищими навчальними закладами. Мета проекту – акумулювати переваги семантичного веб та пов'язаних даних для покращення електронного навчання для широкого

кола французьких вищих навчальних закладів. У межах проекту розроблено онтологію OWL з урахуванням семантики елементів LOM [6].

Дослідження базових понять онтологій у контексті семантичного веб і відповідної термінології дозволяють зробити такий висновок: тема семантичного вебу та використання онтологій як основного апарату представлення доменних знань є інноваційною й актуальною, що активно розвивається та має право на впровадження у вітчизняну освіту.

**Виклад основного матеріалу дослідження.** В епоху аналітики великих даних, пошукові та рекомендаційні системи стали основними механізмами, за допомогою яких користувачі знаходять і виявляють корисну інформацію. Тому важливо, щоб ці системи оброблення даних могли надавати цільові, релевантні результати, які цілком відповідають намірам користувача.

Пошукові та рекомендаційні двигуни рідко можуть конкурувати, якщо вони не використовують моделі, що містять детальну інформацію про категорії заданих питань і, що більш важливо, категорії відповідей. Одним із найбільш типових шляхів представлення доменної області для наочної демонстрації цих ідей є використання онтологій – комбінацій таксономій, що містять попередньо відомі сутності, їхні властивості та взаємозв'язки. Такі онтології можна інтегрувати в пошуковий сервіс для того, щоб покращити його спроможність надавати кінцевому користувачу саме ті дані, які він очікує отримати. Наприклад, якщо хтось шукає термін «сервіс» у домені «комп'ютерних наук», то цей термін має дуже різні поняття, наприклад, в апаратному забезпеченні – «довічний сервіс» (lifelong service) чи протилежне значення в інформаційних технологіях – «веб-сервіс» (web service), коли хтось шукає вакансію, то запит може мати різноманітні значення залежно від контексту [7].

Онтології зазвичай будуються вручну експертами, що робить їх створення, підтримку й оновлення витратним рішенням. У науковому просторі розробляються підходи до пошуку зв'язків у доменних областях для часткової автоматизації процесу побудови онтологій.

Розглянемо методи й алгоритми, які використовують потужність великих аналітичних даних і розподілених обчислень для автоматичного створення мовно-незалежних семантичних баз знань. Такі семантичні бази знань дозволяють значно покращити узгодженість між запитом та документами, а отже, пропонувати більш релевантні результати для будь-якого пошукового або рекомендаційного запиту.

1. Пошукові двигуни є одним із найпоширеніших способів взаємодії людей із цифровою інформацією, вони можуть отримати більше користі від інтеграції із семантичними базами знань, які підвищують загальну спроможність пошукової системи точно інтерпретувати запити та відповідати на них. Можливість шукати будь-яку комбінацію ключових слів у трильйонах документів і оцінювати релевантність всіх результатів запиту за часом відповіді в кілька секунд потребує наявності вузькоспеціалізованих структур даних і підходів до їх моделювання, що працюють паралельно в розподіленій системі. Головним із них є інвертована індексація, відокремлення та реплікація даних, денормалізована модель даних, а також розподілена модель агрегації й оцінки.

2. Інвертований індекс – механізм швидкого пошуку за ключовими словами. Для створення інвертованого індексу потрібно спочатку створити прямий індекс, який зіставить документи та списки термінів, які містяться в них. В інвертованому індексі кожне ключове слово відповідає набору документів, які містять його. За інвертованим індексом можна швидко, без циклічного проходження за списком слів, визначити ті документи, які відповідають певному набору ключових слів у пошуковому запиті (див. рис. 1).

Єдине, що пропущено в спрощеному варіанті індексу, це те, що додаються ще і позиції терміну в документі, ці позиції (часто разом із метаданими) зберігаються разом з ідентифікатором документа. Фразові запити (наприклад, «дерев'яний бампер для телефону») можна знайти, використовуючи позиції в списку, щоб відфільтрувати документи, де всі терміни («дерево», «бампер», «телефон») з'являються послідовно.

3. Шардінг даних – корисна властивість інвертованого індексу, яка полягає в зростанні ймовірності використання термінів, вже наявних в індексі, у разі збільшення кількості індексованих документів. Це означає, що для великих наборів даних кількість термінів в індексі стабілізується за умови збільшення кількості документів, оскільки кількість термінів залежить від розподілу термінів у мові документа. Це дає можливість легко розділити інвертований індекс на декілька підіндексів і потім розподіляти запити до кожного з них паралельно й агрегувати отримані документи. Таке розбиття часто називається шардінгом (відокремленням) індексу.

4. Рекомендаційні системи – системи, що автоматизують процес виявлення інтересів користувача шляхом застосування технології інтелектуального

аналізу даних (Data Mining) для виявлення об'єктів, цікавих для окремих користувачів [8]. Протягом багатьох років архітектури рекомендаційних систем розвивалися як в академічному середовищі, так і в інформаційній індустрії (електронна комерція/електронний шопінг, електронні бібліотеки, електронне навчання, електронний туризм тощо) через експоненційне збільшення обсягу даних.

Рекомендаційні системи можна поділити на три основні категорії: контентні, зі спільною фільтрацією та гібридні технології [9]. Контентні є найбільш чутливими до розуміння текстового змісту, оскільки ці системи засновані на зіставленні елементів/користувачів на основі подібності між їхнім текстовим описом [10]. Наявність семантичної бази знань має вирішальне значення для підвищення ефективності контентних рекомендаційних систем [11].

5. Семантичне виявлення – процес побудови семантичних баз знань, що традиційно зосереджений на використанні онтологій/таксономій, які створюються та підтримуються їхнім розробником. Ручна побудова онтологій/таксономій не масштабується, важко підтримується і є дуже витратною роботою. Для виявлення семантичних

зв'язків між термінами даного корпусу використовують кластеризацію та зменшення розмірності, що супроводжується появою незрозумілих зв'язків. Окрім названих підходів для виявлення семантичних зв'язків між фразами можна спиратися на журнали пошуку користувачів.

Щоб краще зрозуміти проблему, подумаємо про різні значення слова «архітектор» у контексті «архітектури будівлі» та «архітектора програмного забезпечення». Якщо хтось вводить слово «архітектор» у формі пошуку, пошукова система на базі ключових слів поверне змішаний набір документів, серед яких деякі будуть стосуватися архітектури програмного забезпечення, проте інші – архітектури будівель. Типовий пошуковий двигун трансформує запит «архітектор будівель» у булевий вигляд «будівель» і «архітектор», в якому слова розглядаються як незалежні терміни, що призводить до вибору документів, в яких записано про когось, хто є «архітектором програмного забезпечення», який «будує додатки».

Маючи доступ до історії пошуку мільйонів користувачів, за допомогою аналізу великих даних можна виявити взаємозв'язки між пошуковими фразами та найпоширенішими значеннями кож-

| Корпус текстів   |               |   | Прямий індекс     |               |                   | Інвертований індекс |              |              |
|--|---------------|---|-------------------|---------------|-------------------|---------------------|--------------|--------------|
| <p><b>Текст: 1</b><br/> <b>Посада: зареєстрована медсестра</b><br/> <b>Опис: зареєстрована медсестра у лікарню</b><br/> <b>Навички: онкологія, флеботомія</b></p> <p><b>Текст: 2</b><br/> <b>Посада: Інженер-програміст</b><br/> <b>Опис: програмний інженер у компанію</b><br/> <b>Навички: .net, java</b></p> <p><b>Текст: 3</b><br/> <b>Посада: Java-інженер</b><br/> <b>Опис: інженер Java у компанію</b><br/> <b>навички: Java, Swift</b></p> |               |   | <b>Атрибут</b>    | <b>Текст</b>  | <b>Слово</b>      | <b>Атрибут</b>      | <b>Слово</b> | <b>Текст</b> |
|  | <b>Опис</b>   | 1 | зареєстро<br>вана | <b>Опис</b>   | зареєстро<br>вана | 1                   |              |              |
|  |               |   | медсестра         |               | медсестра         | 1                   |              |              |
|  |               |   | у                 |               | у                 | 1, 2, 3             |              |              |
|  |               |   | лікарню           |               | лікарню           | 1                   |              |              |
|  |               | 2 | програмни<br>й    |               | програмни<br>й    | 2                   |              |              |
|  |               |   | інженер           |               | інженер           | 2, 3                |              |              |
|  |               |   | у                 |               | компанію          | 2, 3                |              |              |
|  |               |   | компанію          |               | Java              | 3                   |              |              |
|  |               | 3 | інженер           | <b>Посада</b> | зареєстро<br>вана | 1                   |              |              |
|  |               |   | Java              |               | медсестра         | 1                   |              |              |
|  |               |   | у                 |               | інженер           | 2, 3                |              |              |
|  |               |   | компанію          |               | програміс<br>т    | 2                   |              |              |
|  | <b>Посада</b> | 1 | зареєстро<br>вана |               | Java              | 3                   |              |              |
|  |               |   | медсестра         |               |                   |                     |              |              |
|  |               | 2 | інженер           |               |                   |                     |              |              |
|  |               |   | програміст        |               |                   |                     |              |              |
|  |               | 3 | Java              |               |                   |                     |              |              |
|  |               |   | інженер           |               |                   |                     |              |              |

Рис. 1. Зіставлення текстів у прямому й інвертованому індексах

ного терміна. Мірою схожості змісту двох термінів є семантична схожість [12]. Два основних підходи, використовувані для обчислення семантичної схожості, – семантичні мережі (підхід, заснований на знаннях), а також обчислення зв'язності термінів у великому корпусі текстів (структурний підхід) [13]. Основні методи – точкова взаємна інформація (Point-wise Mutual Information (далі – PMI)) та латентний аналіз семантики (Latent Semantic Analysis (далі – LSA)). Дослідження показують, що PMI перевершує LSA за знаходженням синонімів в Інтернеті [14].

Методи формування семантичних знань. Неоднозначний термін – термін, який відноситься до більш ніж одного значення, залежно від контексту. Наприклад, термін “java” може посилатися на мову програмування Java або на тип кави під назвою “java”, або на острів Індонезії Java.

Щоб виявити неоднозначні терміни, використовують метод побудови ймовірнісної моделі ієрархічних даних (Probabilistic Graphical Model for Massive Hierarchical Data Problems (далі – PGMHD)), який заснований на зборі пошукових термінів, введених користувачами для розрахунку оцінки смислової схожості термінів та класів користувачів. Метод розміщує класи користувачів у верхньому рівні, а пошукові терміни – у нижньому рівні семантичної моделі. Далі два рівні з'єднуються ребрами, які вказують на те, скільки користувачів із даного класу у верхньому рівні шукали певний термін у нижньому рівні. Якщо кількість з'єднань вище, ніж певне граничне значення для більш ніж одного базового класу, термін вважається неоднозначним.

Після виявлення неоднозначності термінів наступною задачею є вирішення цієї неоднозначності. Зменшення неоднозначності означає визначення ймовірних семантично пов'язаних значень неоднозначного терміна. Кожна група семантично пов'язаних термінів являє собою можливе значення первинного терміна з урахуванням контексту, в якому терміни були використані.

Наприклад, неоднозначний термін “driver” має семантично пов'язані терміни “transportation”, “truck driver”, “software”, “embedded system”, “CDL”. Класифікуючи ці терміни за допомогою класів користувачів, які згадували їх в журналах пошуку, у кінцевому підсумку їх можна розподілити на дві групи: 1) “transportation”, “truck driver”, “CDL” та 2) “software”, “embedded system”.

Оскільки PGMHD потрібен для виявлення неоднозначних термінів, ідентичну модель можна застосувати, щоб знайти для будь-якого заданого терміна семантично пов'язані терміни, що належать до одного класу. Для цього потрібно розрахувати оцінку подібності, засновану на ймовірнісних показниках, між даним терміном X та терміном Y [15].

Група семантично пов'язаних термінів, які класифікуються під тим самим батьківським класом, формує ймовірний зміст неоднозначного терміна. Даний підхід не має обмежень за кількістю синонімів: для деяких термінів є лише два можливих значення, деякі мають три і більше значень.

Окрім аналізу журналів запитів для автоматичного створення семантичних баз знань, також можна використовувати взаємозв'язки між словами і фразами, закодовані як в тексті, так і в

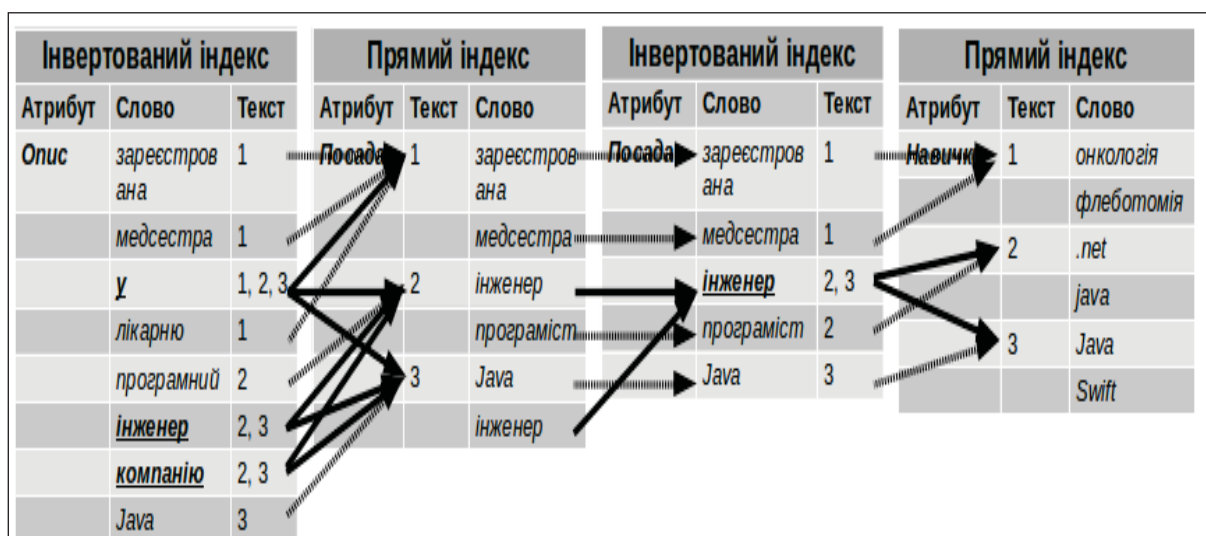


Рис. 2. Матеріалізація ребер із використанням документів

структурованому змісті множини документів. Шляхом оброблення вмісту даних, що представляють доменну область, можна автоматично згенерувати представлення області знань у вигляді графа.

Після того, як граф знань побудовано, треба пройти по всім вузлам, щоб отримати взаємозв'язки між ключовими словами, фразами, сутностями й іншими лінгвістичними варіаціями, представленими у вхідних даних. Ця модель називається «графом семантичних знань» (Semantic Knowledge Graph (далі – SKG)).

Інші системи побудови онтологій зазвичай намагаються визначити певні сутності в колекції і заповнити заздалегідь сформований граф взаємозв'язків між сутностями. Це, на жаль, призводить до значної втрати інформації тоді, коли значення терміна або фрази змінюється залежно від його лінгвістичного контексту. Одна із цілей підходу SKG полягає в тому, щоб цілком передбачити всі смислові взаємозв'язки, що містяться в наборі текстових документів.

Зміст слів може варіюватися залежно від контексту, в якому вони знайдені. Наприклад, слова «архітектор» та «інженер» добре відомі, але, коли вони трапляються всередині фраз, як-от «архітектор програмного забезпечення» або «інженер-електрик», вони мають більш обмежену інтерпретацію.

Модель SKG відрізняється від більшості традиційних графових структур, бо дозволяє неорієнтованість ребра, яке з'єднує будь-які два вузла. Щоб перейти від вихідного вузла  $v_i$  у вузол  $v_j$  по ребру  $e_{ij}$ , треба мати інвертований індекс пошуку, що встановлює відповідність вузла  $v_i$  до базового набору документів, а також прямий індекс пошуку, який ставить ці документи у відповідність до будь-якого іншого вузла  $v_j$ , з яким ці доку-

менти також пов'язані. Ця комбінація інвертованого і прямого індексів дозволяє відображати всі терміни або комбінації термінів як вузли в графі, виконувати обхід між будь-якими двома вузлами через набір спільно використовуваних документів, як показано на рис. 2.

Наявність ребер між документами зумовлена наявністю в них спільних термінів. Ваги ребер розраховуються за допомогою функції статистичного розподілу документів у вузлах. Граф семантичних знань забезпечує як без втрат, так і з високим ступенем стиснення, представлення всіх потенційних ребер, які пов'язують можливі матеріалізовані вузли.

**Висновки.** Розглянуто методи й інструменти формування семантичних знань, серед яких аналіз великих обсягів запитів за допомогою ймовірнісної моделі ієрархічних даних PGMHD, що дозволяє знайти семантично пов'язані терміни та фрази, які можна застосувати для побудови онтологій. Також описано метод виявлення та зменшення неоднозначності змісту термінів і фраз, які знаходяться в журналах запитів.

Розглянуто модель Semantic Knowledge Graph, яка застосовує співвідношення між словами та фразами в межах документів, щоб автоматично генерувати граф співвідношень між фразами. Із графа знань можна виявити й оцінити зв'язки між будь-якими суб'єктами, що містяться в ньому, суто за змістом індексованих документів.

Застосування семантичних методів є необхідною умовою для побудови пошукових і рекомендаційних систем, здатних індексувати документи, формувати семантичні зв'язки та гнучко інтерпретувати вхідні запити. Прикладом вдалої інтеграції графа знань є підвищення на 59–76% релевантності результатів пошуку в одній із найбільших у світі систем для пошуку вакансій [16].

#### Список літератури:

1. Грибова В., Клещев А. Онтологическая парадигма программирования. Open Semantic Technologies for Intelligent Systems, 2012. С. 213–220. URL: [http://conf.ostis.net/images/a/ab/%D0%93%D1%80%D0%B8%D0%B1%D0%BE%D0%B2%D0%B0\\_%D0%92.%D0%92..2012%D1%81%D1%82-%D0%9E%D0%BD%D1%82%D0%BE%D0%BB%D0%9F%D0%9F.pdf](http://conf.ostis.net/images/a/ab/%D0%93%D1%80%D0%B8%D0%B1%D0%BE%D0%B2%D0%B0_%D0%92.%D0%92..2012%D1%81%D1%82-%D0%9E%D0%BD%D1%82%D0%BE%D0%BB%D0%9F%D0%9F.pdf) (дата звернення: 15.04.2018).
2. Resource Description Framework, Web Ontology Language. URL: [https://www.w3.org/\(/RDF чи OWL/\)](https://www.w3.org/(/RDF%20чи%20OWL/)) (дата звернення: 12.02.2018).
3. Academic Institution Internal Structure Ontology (AIISO). URL: <http://vocab.org/aiiso/schema#> (дата звернення: 21.02.2018).
4. Learning Resource Metadata Initiative (LRMI). URL: <https://wiki.creativecommons.org/wiki/LRMI> (дата звернення: 23.02.2018).
5. Teaching Core Vocabulary Specification. URL: <http://linkedscience.org/teach/ns/#sec-reference> (дата звернення: 26.02.2018).
6. Ontologies informatiques at SemUnit. URL: [http://semunt.supelec.fr/pubby/page/ressource/unit/\\_2166\\_Ontologies\\_informatiques](http://semunt.supelec.fr/pubby/page/ressource/unit/_2166_Ontologies_informatiques) (дата звернення: 27.02.2018).

7. Navigli R., Velardi P. Learning domain ontologies from document warehouses and dedicated web sites. *Computational Linguistics*. 2004. Т. 30. №. 2. P. 151–179.
8. Bobadilla J., Ortega F., Hernando A., Gutierrez A., Recommender systems survey. *Knowledge-based systems*. 2013. Vol. 46. P. 109–132.
9. Aggarwal C. Content-based recommender systems. *Recommender Systems*. Springer; Cham, 2016. P. 139–166.
10. Pazzani M., Billsus D. Content-based recommendation systems. *The adaptive web*. Springer; Berlin; Heidelberg, 2007. P. 325–341.
11. Semantics-aware content based recommender systems / M. de Gemmis, P. Lops, C. Musto, F. Narducci, G. Semeraro. *Recommender Systems Handbook*. Springer; Boston; MA, 2015. P. 119–159.
12. Semantic measures for the comparison of units of language, concepts or instances from text and knowledge base analysis / S. Harispe, S. Ranwez, S. Janaqi, J. Montmain. *arXiv preprint arXiv: 1310.1285*. 2013.
13. Mihalcea R., Corley C., Strapparava C. Corpus-based and knowledge-based measures of text semantic similarity. *AAAI*. 2006. Vol. 6. P. 775–780.
14. Turney P. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. *Proceedings of the 12th European Conference on Machine Learning, EMCL 01, (London, UK)*. Springer-Verlag, 2001. P. 491–502.
15. PGMHD: A scalable probabilistic graphical model for massive hierarchical data problems / K. AlJadda, M. Korayem, C. Ortiz, T. Grainger, J. Miller, and W. York. *Big Data (Big Data)*, 2014. *IEEE International Conference on, IEEE*, 2014. P. 55–60.
16. Grainger T., AlJadda K., Korayem M., Smith A. The Semantic Knowledge Graph: A compact, auto-generated model for real-time traversal and ranking of any relationship within a domain. *Data Science and Advanced Analytics (DSAA)*, 2016. *IEEE International Conference on. IEEE*, 2016. С. 420–429.

#### **АНАЛИЗ ВЗАИМОДЕЙСТВИЯ СОДЕРЖАНИЯ УЧЕБНЫХ ДИСЦИПЛИН ПО ИХ СЕМАНТИЧЕСКОМУ ОПИСАНИЮ**

*Целью данной работы является исследование существующих подходов к описанию и построению онтологий, в частности, в домене учебных дисциплин, и методов автоматизации формирования доменных знаний из корпуса документов. Исследованы возможности применения онтологий для представления доменной области учебных дисциплин и курсов. Описаны понятия онтологии в контексте компьютерных наук и семантического веба, рассмотрены существующие онтологии учебных дисциплин, используемых в мире, и выявлены слабые стороны ручного создания и поддержания онтологий. Проведен анализ методов и подходов, автоматизирующих процесс построения онтологий. Рассмотрены ограничения, возникающие при решении подобных задач как с теоретической стороны, так и в контексте практических аспектов построения приложений, реализующих такие модели. Исследованы современные инновационные методы автоматического построения графа знаний в рамках доменной области.*

**Ключевые слова:** онтология, учебная дисциплина, граф знаний, семантический веб, семантический поиск.

#### **ANALYSIS OF THE INTERACTION OF THE EDUCATIONAL DISCIPLINES CONTENT BY THEIR SEMANTIC DESCRIPTION**

*The purpose of this work is to study existing approaches to the description and construction of ontologies, especially in the domain of academic disciplines and methods for automating the presentation of domain knowledge from the body of documents. Possibilities of application of ontologies for representation of the domain area of educational disciplines and courses are studied. The concept of ontology in the context of computer science and semantic web is described, various existing ontologies of educational disciplines used in the world are considered and the weaknesses of manual creation and support of ontology are revealed. The analysis of methods and approaches that automate the process of construction of ontologies is carried out. The limitations, which arise when solving similar problems from the theoretical side and in the context of practical aspects of constructing software applications implementing such models are considered. Innovative modern methods of automatic building of a knowledge graph within the domain area are investigated.*

**Key words:** ontology, educational discipline, knowledge graph, semantic web, semantic search.