

УДК 004.318

*Барташевська Ю.М.***ВИКОРИСТАННЯ СЕМАНТИЧНИХ ТЕХНОЛОГІЙ У ДІЯЛЬНОСТІ ЗВО:
ПРИКЛАДНИЙ АСПЕКТ****ВНЗ «Університет імені Альфреда Нобеля», м. Дніпро, Україна**

В статті розглянуто проблему використання великих даних у сучасній економіці та суспільному житті. Обсяги та складність інформації стрімко зростають, але сучасні технології не можуть забезпечити ефективне їх використання. Існує відставання в технологіях, алгоритмах та практиці використання великих даних. Таку диспропорцію здатні змінити семантичні технології, які характеризуються іншим підходом до обробки та використання даних. Даний підхід заснований на використанні знань. Досліджено, що незважаючи на доволі тривалий час існування семантичних технологій і семантичних мереж існує багато перепон для їх ефективного впровадження. Це проблеми доступності семантичного контенту, доступності онтологій, їх еволюції, масштабованості та багатомовності. А оскільки далеко не всі дані, подані у мережі, створені у термінах семантичної розмітки і навряд чи будуть приведені до неї у перспективі, то проблема доступності семантичного контенту є однією з основних. В статті показано відмінність семантичної мережі і семантичного Веба, а також вказані технології розвитку останнього. В якості предмета дослідження обрано модуль курсів ВНЗ «Університет імені Альфреда Нобеля». Детально розглянуто склад окремого модуля або окремого курсу: дані щодо закладу вищої освіти (ЗВО), викладача-лектора, дані щодо забезпечення курсу та мови його викладання, набутих навичок, вмінь, результатів тощо. Побудовано граф модуля курсів на прикладі Університету імені Альфреда Нобеля в термінах онтології, розглянуті його окремі, найбільш значущі класи – складові частини. Розглянуті основні класи, підкласи та їх наповнення, вказано типи даних (дата, текст, URL). Переведено онтологічну схему у RDF-формат, як такого, що необхідний для моделювання даних для семантичної мережі та подальшого дослідження. Окреслено перспективи подальших досліджень щодо застосування обраної моделі представлення знань, використання мови запитів, отримання та інтерпретації даних інших ЗВО тощо.

Ключові слова: семантичні технології, семантичні мережі, онтології, SmartTools, граф модуля курсів.

DOI: 10.32434/2521-6406-2020-1-7-3-9

Постановка проблеми

Сьогодні семантичні технології стають все більш популярними та затребуваними. Насамперед це пов'язано зі зростанням кількості та складності інформації. За підрахунками міжнародних експертів [1] вже сьогодні існує велике відставання між обсягами даних і технологіями і системами, що забезпечують їх ефективне використання. Причому це стосується не тільки алгоритмів і технологій обробки, а й їх зберігання. За оцінюваннями спеціалістів, реальний рівень використання наявних у компаній сис-

тем зберігання в середньому становить 25–35% [2].

Про зростання кількості даних свідчать і результати дослідження компанії IDC. За їх підрахунками, наданими у аналітичному звіті [3], глобальний обсяг даних зростає з 33 Зетабайт у 2018 р. до 175 Зетабайт у 2025 р.

Всі ці дані свідчать про те, що кількість даних у сучасному світі зростає постійно, а великі масиви даних зросли більш ніж на порядок. Користувачі і технології наразі відстають від цього зростання і відрив стрімко зростає.

Актуальним для компаній стає питання зберігання великих даних, які зібрані у різний час різними підрозділами, з контролем їх якості та узгодженості. Наслідок цього — зростання витрат на інфраструктуру та переробку даних. Компанії накопичують дані, не приділяючи особливої уваги їх інтеграції, і не завжди точно знають, яка інформація є у їх розпорядженні. Сам факт зберігання створює помилкове уявлення переробки даних, у разі потреби, на щось цінне та корисне, спрямоване на благо компанії. Але це уявлення хибне, а компанії часто недооцінюють витрати на пошук, обробку та аналіз даних. За даними аналітиків [4], тільки 3% зібраних наборів даних компаній, які відображають її ключові показники роботи, не містили помилок чи критичних помилок. Також стає необхідною стандартизація термінології однієї організації, оскільки вміння зіставляти інформацію з різних джерел і швидко реагувати на зовнішні зміни цілком може стати визначальним фактором її фінансового успіху.

У зв'язку з цим актуальним стає перехід від обробки даних до обробки знань, тобто зміна характеру роботи з інформацією. Одним з шляхів реалізації обробки знань є використання семантичних мереж, як необхідній інструмент моделювання та структурування онтологій; застосування підходу формального представлення знань в базі знань або в графах знань (Knowledge Graph). І хоча бази знань, як і бази даних з'явилися давно, тільки зараз компанії починають замислюватися над використанням онтологій, логічних висновків тощо. Це свідчить про поступове зміщення акценту від ведення бізнесу на основі даних до ведення бізнесу на основі знань, а отже якісного використання бізнес-інформації.

Мета статті

Метою статті є дослідження використання семантичних технологій, моделі графу знань (онтологічної моделі) у практичній діяльності ЗВО.

Аналіз останніх досліджень і публікацій

Ідея семантичної мережі була розроблена консорціумом W3 в середині 90-х років. Продовження та розвитку вони набули у 2001 р. у журналі Scientific American у статті Тіма Бернерса-Лі, Джеймса Хендлера і Орі Лассила під назвою «The Semantic WEB» [5].

Сьогодні проблематикою семантичних мереж, онтологій займаються такі провідні вітчизняні та зарубіжні вчені, як: Грінков В.О. [2], Хлизов Є. [4], Хорошевський В.Ф. [6], Лянган Ю. [7], Петренко М.Г., Зеленцов Д.Г. [8] та ін.

Отже, ідеї автоматизованої обробки знань висувуються вже давно, але повноцінно використовуватися почали лише зараз. На сьогодні створені всі необхідні компоненти методик і технологій, потрібних для роботи з онтологічними моделями, які є предметом обробки за допомогою семантичних технологій. Однак, на шляху впровадження цих технологій та методик є декілька проблем науково-технічного та людського характеру, а саме: доступність семантичного контенту та онтологій, еволюція онтологій, їх масштабованість та багатомовність. У зв'язку з цим особливого значення набувають проблеми онтологічного інжинірингу та роботи з існуючими онтологіями, обробки та зберігання великих обсягів знань, їх візуалізація.

Викладення основного матеріалу дослідження

Загальновідомо, що семантичні мережі є технологічним втіленням семантичних технологій.

Сьогодні існує багато визначень семантичної мережі. Скористаємось визначенням, наданим в [9]: «Семантична мережа — це модель предметної області, надана у вигляді графа, де у вершинах стоять поняття, а ребра — відносини між ними». Тобто семантична мережа відображає семантику предметної області у вигляді понять і відношень, кількість типів яких визначається її автором. Ідеї семантичної мережі використовуються і в мережі Інтернет — семантичний Веб (Semantic Web) або семантичне павутиння. Головна мета Semantic Web — формування глобальної семантичної мережі, в якій ресурси Інтернет стандартизовані і придатні для машинної обробки.

Технологіями розвитку семантичного Веб є [7]:

- XML (eXtensible Markup Language) — розширювана мова розмітки;
- RDF (Resource Description Framework) — мова моделювання даних для семантичної мережі. Вся інформація семантичної павутини зберігається і представляється в RDF;
- SPARQL (протокол SPARQL і мова запитів RDF) — мова запитів семантичної мережі. Він спеціально призначений для запиту даних в різних системах;
- OWL (мова вебонтологій) — мова схеми або мова подання знань семантичної мережі. OWL дозволяє визначати поняття спільно, щоб ці поняття можна було використовувати як можна частіше.

Основою даного дослідження є побудова графу знань модуля курсів ЗВО, відповідно до

стандартів розробки програм Університету імені Альфреда Нобеля. Створення такого графу та його використання у складі інформаційної системи дозволить точно та швидко отримувати необхідну інформацію, а отже більш ефективно управляти інформацією ЗВО.

Для більш зручного та наочного надання даних графу знань використовуються різні програми та фреймворки. Одним з найбільш популярних є фреймворк *Protégé*, який дозволяє розмішувати та використовувати онтологію як локально, так і спільно у форматах owl та xml і з повною підтримкою стандарту W3C. Інструментарій фреймворку розповсюджується з відкритим кодом [8].

Однак, з нашої точки зору, для візуалізації та передачі знань більш доречно використовувати концепт-карти. Вони дозволяють побачити суть проблеми, виявити нові знання у даних. Концепт-карти доволі широко використовуються як у освіті, так і в бізнесі з метою: короткого викладення проблеми; виявлення знань; виявлення концепцій, взаємодій ієрархій з аналізу джерел; генерування нових знань; трансформації прихованих знань в явні структуровані; збереження даних в асоціативних зв'язках; моделювання спільних групових знань; обмін думками з прогнозами, тенденціями в робочих групах; вибудовування ланцюжків аргументації тощо [10].

В якості такої концепт-карти було використано *SmartTools*, яка є безкоштовною, дозволяє стимулювати співпрацю і обмін думками при створенні моделей знань, дозволяє здійснювати розробку колективних моделей знань.

Приклад графу знань модуля курсів Університету імені Альфреда Нобеля наведено на рис. 1.

Схема типового курсу, надана на рис. 1, створена за допомогою семантичної розмітки *Shema.org* у програмі *SmartTools*. Основна схема курсу складається з супер (глобальних) класів: курс або модуль курсів (*Course*); дисципліна з курсу (їх може бути декілька; *CreativeWork*); викладач (*Person*) та ін. А також локальних класів: дані щодо курсу (*Property Value Specification*); елементи курсу (*ItemList*); література з курсу (*Book, Article, Digital Document*); інформація щодо результатів навчання та освітнього ступеня (*Educational Occupational Credential*); дані щодо ЗВО (*Organization*) та ін.

Отже, клас курсу Університету має такі об'єкти та їх властивості:

– *name* (назва), яка потім може бути ініційована у класі *CreativeWork/name*. Напри-

клад, курс «Вища математика» містить кілька блоків, які визначені на рівні робочої програми курсу (лінійна алгебра, дискретна математика, теорія ймовірностей тощо);

– *inLanguage* (мова). Більшість курсів (*programs*) викладаються українською мовою, але для програм, затверджених Університетом Трійці Сент-Давид (Уельс) або за деякими окремими курсами, мова викладання – англійська;

– *datePublished* (дата публікації, впровадження) програми. Може співпадати з *copyrightYear* (*CreativeWork*).

Цитування курсу (література, *citation*) надано трьома видами: цифровий документ (*DigitalDocument*), стаття (*Article*) і книга (*Book*). Для книги характеристиками (властивостями) є: автор (*author*), видавець (*bookEdition*) та кількість сторінок (*numberOfPages*). Для статті: автор (*author*), назва статті (*headline*), діапазон сторінок (*pagination*). Для цифрового документа: автор (*author*) та ім'я вмісту (*name*), розташування мережі (*URL*).

Крім того, при описі об'єкта *accountable Person* було використано клас *Person* без зазначення *Lecturer*, оскільки лектор відповідає за щорічну програму навчання та її зміст. Об'єкт *job Title* додано до класу *Person*, оскільки Університет вказує не тільки науковий ранг (*honorific Prefix*) та ступінь (*honorific Suffix*), а й посаду.

Також до класу *Organization* додано об'єкт кафедри (*department*), оскільки програма вказує не тільки організацію (ЗВО), а й випускаючу кафедру. Введено клас *Property Value Specification*, який є підкласом *Intangible*, який описує додаткові властивості курсу: форми навчання, бали з курсу тощо.

Зауважимо, що клас *Course* є підкласом більш потужного класу *Creative Work*, пов'язаного з *Thing*.

Також додано підклас *Educational Occupational Credential*, який відповідає за результати навчання, теоретичні та практичні навички:

– *competency Required* – знання, вміння, здібності, які повинен продемонструвати учень;

– *educational Level* – курс;

– *credential Category* – категорія або «ступінь» бакалавра або магістра.

Отриману схему (рис. 1) можна надати у RDF-форматі, форматі необхідному для моделювання даних для семантичної мережі. Фрагмент коду наведено на рис. 2.

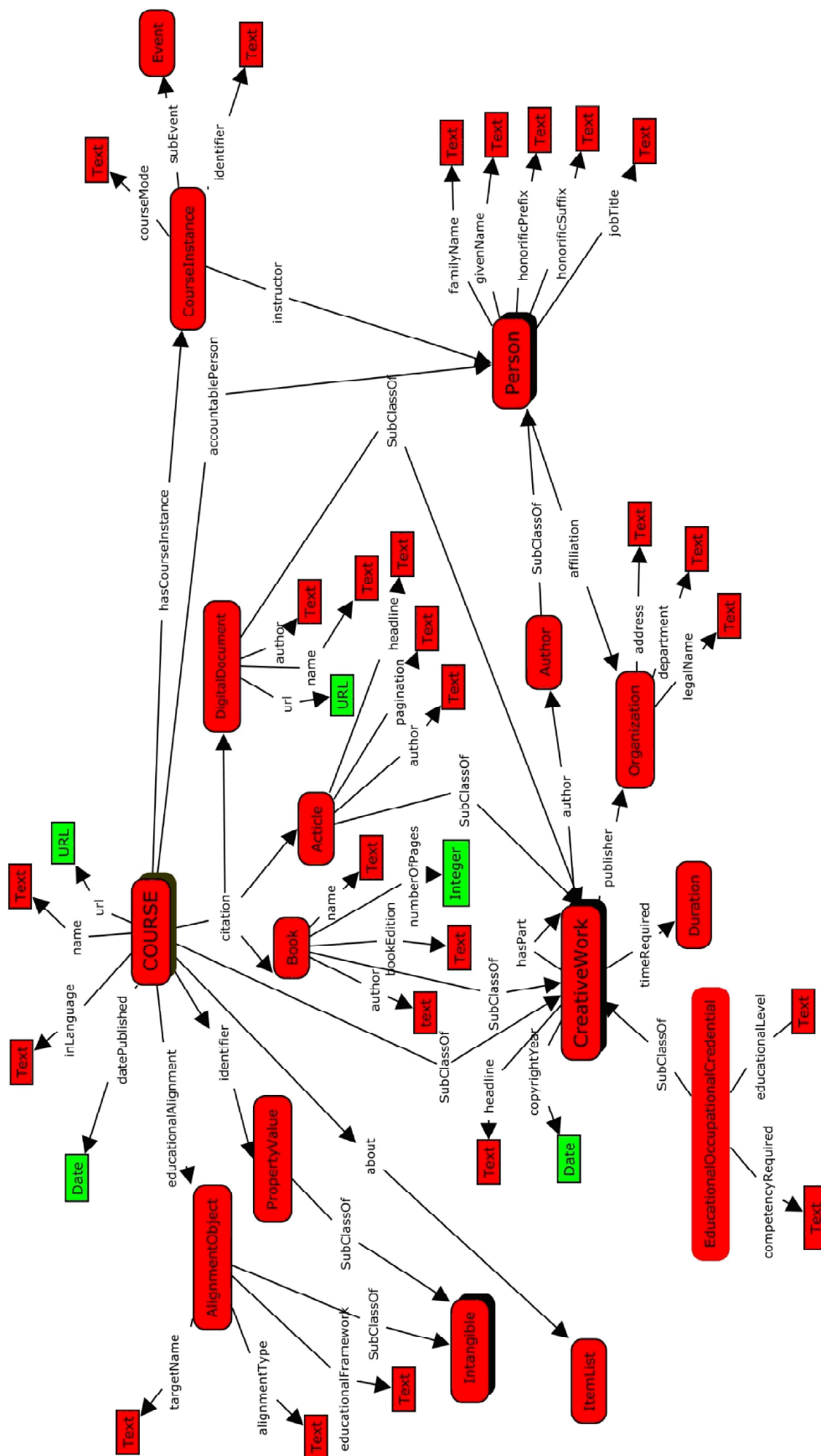


Рис. 1. Граф знань модуля курсів Університету

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#"
3   xml:base="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10"
4   xmlns:owl="http://www.w3.org/2002/07/owl#"
5   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
6   xmlns:xml="http://www.w3.org/XML/1998/namespace"
7   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
8   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
9   xmlns:schema="public-schemaorg@w3.org">
10 <owl:Ontology rdf:about="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10">
11   <owl:versionIRI rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10/1.0"/>
12 </owl:Ontology>
13
14
15 <owl:ObjectProperty rdf:about="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#accountablePerson">
16   <rdfs:domain rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#Course"/>
17   <rdfs:range rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#Lecturer"/>
18 </owl:ObjectProperty>
19
20 <owl:ObjectProperty rdf:about="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#citation">
21   <rdfs:domain rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#Course"/>
22   <rdfs:range rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#Article"/>
23   <rdfs:range rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#Book"/>
24   <rdfs:range rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#DigitalDocument"/>
25 </owl:ObjectProperty>
26
27
28 <owl:ObjectProperty rdf:about="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#educationalAlignment">
29   <rdfs:domain rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#Course"/>
30   <rdfs:range rdf:resource="http://www.semanticweb.org/home/ontologies/2019/9/UAN_10.10#AlignmentObject"/>
31 </owl:ObjectProperty>

```

Рис. 2. Фрагмент RDF-коду семантичної моделі курсу

Висновки

В статті досліджено необхідність поступового переходу до використання семантичних технологій, що оперують знаннями, а не даними. Така необхідність викликана збільшенням обсягів та складністю даних, що нас оточують та бажанням більш ефективного їх використання. Показано застосування семантичних технологій в діяльності ЗВО для підвищення ефективності роботи з інформацією. Побудовано граф знань модуля курсів ЗВО, визначено глобальні та локальні класи, їх наповнення. Розроблено RDF-модель. Крім того, створена модель випробувана для збору інформації щодо окремих курсів ЗВО, що буде розвинуто та описано у подальших дослідженнях. Також існує можливість створення повноцінної інформаційної системи для збору та обробки подібної інформації у ЗВО.

СПИСОК ЛІТЕРАТУРИ

1. Глобальные тенденции 2030: альтернативные миры. Отчет Национального Совета по разведке. – 2012. – Вып.5. – 169 с. Режим доступа: http://www.nkibrics.ru/system/asset_publications/data/53c7/b3a1/676c/7631/400a/0000/original/Global-Trends-2030-RUS.pdf?1408971903.

2. Гриньков В.О. Метод кількісної оцінки стійкості системи зберігання даних // Збірник наукових праць ВІПІ. – 2019. – № 3. – С.16-22.

3. Reinsel David, Gantz John, Rydning John. The Digitization of the World From Edge to Core. // IDC: Analyze the future, 2018. – 28 p. Режим доступа: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-data-age-whitepaper.pdf>

4. Хлызов Е. Управление данными на основе графов знаний // Открытые системы. СУБД. – 2018. – № 3. – С.30-31. Режим доступа: <https://www.osp.ru/os/2018/03/13054405/>

5. Ефимов А.А., Рожина И.В. Актуальность и методические аспекты обучения семантическим сетям в педагогическом ВУЗЕ // Инновационные технологии в педагогике высшей школы: материалы V Международной научной конференции. – Екатеринбург: УрГПУ, 2008. – С.215-220.

6. Хорошевский В.Ф. Семантические технологии: ожидания и тренды // OpenSemantic Technologies for Intelligent Systems (OSTIS-2012): материалы II Международной научно-технической конференции. – Минск: БГУ-ИР, 2012. – С.143-158.

7. Yu Liyang. Introduction to the Semantic Web and Semantic Web Services. – Taylor & Francis Group, LLC, 2007. – 367 p. <https://doi.org/10.1201/9781584889342>

8. Петренко М.Г., Зеленцов Д.Г. Про практичне використання онтологічних моделей предметних областей // Комп'ютерне моделювання: аналіз, управління, оптимізація. – 2019. – № 2(6). – С.58-73.

<https://doi.org/10.32434/2521-6406-2019-6-2-58-73>

9. Козлов Ф.А. Методы агрегирования и анализа данных в системах электронного обучения с использованием семантических технологий: Дис... канд.техн. наук: 05.13.06.СПб.: Санкт-Петербургский Национальный исследовательский университет информационных технологий, механики и оптики, 2015. — 121 с.

10. Прохоров А. Обзор программ класса conceptmapping. // Компьютерпресс. — 2007. — № 3. — С.176-187.

Надійшла до редакції 08.05.2020

ИСПОЛЬЗОВАНИЕ СЕМАНТИЧЕСКИХ ТЕХНОЛОГИЙ В ДЕЯТЕЛЬНОСТИ ВУЗА: ПРИКЛАДНЫЕ АСПЕКТЫ

Барташевская Ю.Н.

В статье рассмотрена проблема использования больших данных в современной экономике и общественной жизни. Объемы и сложность информации стремительно растут, но современные технологии не могут обеспечить эффективное их использование. Существует отставание в технологиях, методах и практике использования больших данных. Такую диспропорцию способны изменить семантические технологии, характеризующиеся другим подходом к обработке и использованию данных. Данный подход основан на использовании знаний. Доказано, что несмотря на довольно длительное время существования семантических технологий и семантических сетей существует много препятствий для их эффективного применения. Это проблемы доступности семантического контента, доступности онтологий, их эволюции, масштабируемости и многоязычия. А поскольку далеко не все данные, представленные в сети, созданы в терминах семантической разметки и вряд ли будут приведены к ней в перспективе, то проблема доступности семантического контента является одной из основных. В статье показано отличие семантической сети и семантического Веба, а также указаны технологии развития последнего. В качестве предмета исследования выбран модуль курсов ВУЗа «Университет имени Альфреда Нобеля». Подробно рассмотрен состав отдельного модуля или отдельного курса: данные относительно ВУЗа, преподавателя-лектора, данные по обеспечению курса и языка его преподавания, приобретенных навыков, умений, результатов и тому подобное. Построен граф модуля курсов на примере ВУЗ «Университет имени Альфреда Нобеля» в терминах онтологии, рассмотрены его отдельные, наиболее значимые классы — составные части. Рассмотрены основные классы, подклассы и их наполнения, указаны типы данных (дата, текст, URL). Онтологическая схема переведена в RDF-формат, как такой, который необходим для моделирования данных в семантической сети и дальнейшего исследования. Определены перспективы дальнейших исследований по применению выбранной модели представления знаний, использования языка запросов, получение и интерпретация данных других ВУЗов и т.д.

Ключевые слова: семантические технологии, семантические сети, онтологии, CmapTools, граф модуля курсов.

THE USE OF SEMANTIC TECHNOLOGIES IN THE ACTIVITIES OF THE UNIVERSITY: APPLIED ASPECTS

Bartashevskaya Yu.N.

Alfred Nobel University, Dnipro, Ukraine

The article considers the problem of using Big Data in a modern economics and public life. The volumes and complexity of information are growing rapidly, but modern technologies cannot ensure their effective use. There is a lag in technologies, methods, and practices for using Big Data. The imbalance can be changed by semantic technologies, characterized by a different approach to the processing and use of data. This approach is based on the use of knowledge. Proved that despite the rather long time of the existence of semantic technologies and semantic networks, there are many obstacles to their effective application. These are the problems of accessibility of semantic content, accessibility of ontologies, their evolution, scalability and multilingualism. And since far from all the data presented on the network is created in terms of semantic markup and is unlikely to be brought to it in the future, the problem of accessibility of semantic content is one of the main ones. The article shows the difference between the semantic network and the semantic Web, and also indicates the development technologies of the latter. As the subject of study, the module of the courses of the Alfred Nobel University was selected. The composition of a separate module or a separate course is examined in detail: data on the university, lecturer, data on the provision of the course and language of its teaching, acquired skills, abilities, results and the like. A graph of the module of courses has been built on the example of the Alfred Nobel University in terms of ontology, its individual, most significant classes — components are considered. The main classes, subclasses and their contents are considered, data types (date, text, URL) are indicated. The ontological scheme has been converted to the RDF format, such as is necessary for modelling data in the semantic network and further research. The prospects for further research on the application of the selected model for representing knowledge, using the query language, obtaining and interpreting data from other universities, etc. are determined.

Keywords: semantic technologies, semantic networks, ontologies, CmapTools, course module graph.

REFERENCES

1. *Global'nye tendencii 2030: al'ternativnye miry. Otchet Natsionalnogo Soveta po razvedke* [Global Trends 2030: Alternative Worlds. National Intelligence Council Report]. 2012, vol. 5, 169 p. URL: <http://www.kmu.gov.ua/control/uk/photogallery/gallery?galleryId=15725757&>. (in Russian).
2. Hrinkov V.O. *Metod kilkisnoi otsinky stikosti systemy zberihannia danykh* [Method of quantitative assessment of storage system stability]. Collection of VITI scientific works, 2019, vol. 3, pp.16-22. (in Ukrainian).
3. Reinsel David, Gantz John, Rydning John. *The Digitization of the World From Edge to Core*. IDC: Analyze the future, 2018, 28 p. URL: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>.
4. Hlyzov E. *Upravlenie dannymi na osnove grafov znanij* [Data management based on knowledge graphs]. Open systems. SCDB, 2018, vol. 3, pp.30-31. URL: <https://www.osp.ru/os/2018/03/13054405/> (in Russian).

5. Efimov A.A., Rozhina I.V. *Aktual'nost' i metodicheskie aspekty obuchenija semanticheskim setjam v pedagogicheskom VUZE* [The relevance and methodological aspects of teaching semantic networks in a pedagogical university]. Innovative Technologies in Higher Education Pedagogy: Materials of the V International Scientific Conference, 2008, pp.215-220. (in Russian).

6. Horoshevskij V.F. *Semanticheskie tehnologii: ozhidanijai trendy* [Semantic technologies: expectations and trends]. Open-SemanticTechnologiesforIntelligentSystems (OSTIS-2012). 2012, pp.143-158. (in Russian).

7. Yu Liyang Introduction to the Semantic Web and Semantic Web Services. Taylor & Francis Group, LLC, 2007, 367 p. <https://doi.org/10.1201/9781584889342>

8. Petrenko N.G., Zelentsov D.G. *Pro praktichne vikoristannja ontologichnih modelej predmetnih oblastej* [On the practical use of ontological models of subject areas]. Computer Modeling: Analysis, Control, Optimization, 2019, no. 2, pp.58-73. (in Russian).
<https://doi.org/10.32434/2521-6406-2019-6-2-58-73>

9. Kozlov F.A. *Metody agregirovanija i analiza dannyh v sistemah elektronnoho obuchenija s ispol'zovaniem semanticheskikh tehnologij* [Methods of aggregating and analyzing data in e-learning systems using semantic technologies]: Dis ... cand. tech. of sciences: 05.13.06. SPb, 2015, 121 p. (in Russian).

10. Prohorov A. *Obzor program klassa concept mapping* [Review of concept mapping class programs]. Computerpress, 2007, vol. 3, pp.176-187. (in Russian).