“Львівська політехніка”. – 2014. – № 805. – С. 298–305. 4. Кустова Г. И. Словарь как лексическая база данных / Г. И. Кустова, Е. В. Падучева // Вопросы языкознания, 1994. – № 4. – С. 96–106. 5. Перебийніс В. І., Сорокін В. М. Традиційна та комп'ютерна лексикографія: навч. посібник. – К.: Вид. центр КНЛУ, 2009. — 218 с. 6. Рабулець О. Г. Дієслово в лексикографічній системі / О. Г. Рабулець, Н. М. Сухарина, В. А. Широков, К. М. Якименко; НАН України; Укр. мов.-інформ. фонд. – К.: Довіра, 2004. – 259 с. 7. Сухарина Н. М. Граматична та лексична семантика українського дієслова в лексикографічній системі: автореф. дис. ... канд. філол. наук: 10.02.01 / Н. М. Сухарина ; НАН України. Ін-т мовознав. ім. О. О. Потебні. – К., 2003. – 19 с. 8. Чепик Е. Ю. Политическое слово в структуре электронного словаря // Культура народов Причерноморья. 2005. – № 69. – С. 205–209. 9. Широков В. А. Інформаційна теорія лексикографічних систем / В. А. Широков. – К.: Довіра, 1998. – 331 с. 10. Широков В. А. Технологічні основи сучасної тлумачної лексикографії / В. А. Широков, О. Г. Рабулець, О. М. Костишин, І. В. Шевченко, К. М. Якименко // Мовознавство, 2002. – № 6. – С.49–86.

УДК 004.9

В. В. Литвин, В. А. Висоцька, Д. Г. Досин, М. Г. Гірняк
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

РОЗРОБЛЕННЯ МЕТОДІВ ТА ЗАСОБІВ ПОБУДОВИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ОПРАЦЮВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ З ВИКОРИСТАННЯМ ОНТОЛОГІЧНОГО ПІДХОДУ

© Литвин В. В., Висоцька В. А., Досин Д. Г., Гірняк М. Г., 2015

Розглянуто питання розроблення методів та програмних засобів опрацювання інформаційних ресурсів. Розроблено формальний опис процесу опрацювання текстового контенту. Сформульовано новий підхід до застосування та впровадження бізнес-процесів.

Ключові слова: контент, онтологія, інформаційний ресурс, бізнес-процес, система управління контентом, життєвий цикл контенту.

The article discusses the development of unified methods and software tools for processing information resources. The models of information resource processing are proposed. A new approach to application and implementation of business processes is formulated.

Key words: content, ontology, information resource, business-process, content management system, content lifecycle.

Вступ. Загальна постановка проблеми

Зміст дослідження полягає у вирішенні важливої науково-прикладної проблеми побудови інтелектуальних систем опрацювання інформаційних ресурсів (ядром баз знань яких є онтології) для аналізу текстових масивів даних при формуванні спільних тематик досліджень у міжвузівських консорціумах. Використання онтологій у складі баз знань інтелектуальних систем опрацювання інформаційних ресурсів допомагає вирішити низку методологічних та технологічних проблем, які виникають під час розроблення таких систем. Зокрема, для України це відсутність концептуальної цілісності й узгодженості окремих прийомів та методів інженерії знань; нестача кваліфікованих фахівців у цій галузі; жорсткість розроблених програмних засобів та їх низька адаптивна здатність;

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

складність впровадження інтелектуальних систем опрацювання інформаційних ресурсів, що зумовлено психологічними аспектами [1–7]. Все це свідчить про актуальність проблематики досліджень використання онтологій у процесі побудови інтелектуальних систем опрацювання інформаційних ресурсів.

Основною компонентою інтелектуальних систем опрацювання інформаційних ресурсів є база знань, що формується відповідно до предметної області, на яку зорієнтоване функціонування цієї системи. Традиційні методи інженерії знань (отримання знань від експерта, машинне навчання тощо) не ґрунтуються на системі вивірених та загальноприйнятих стандартів, тому побудовані на їхній основі бази знань із часом втрачають свою функціональність через низьку ефективність функціонування. Як стандарт інженерії знань використовують онтологічний інжиніринг, у результаті застосування якого отримують онтологію бази знань. Онтологія – це детальна формалізація деякої області знань, подана за допомогою концептуальної схеми. Така схема складається з ієрархічної структури термінів, зв'язків між ними, теорем та обмежень, які прийнято у певній предметній області. Пропонується зважувати елементи онтології з метою прийняття оптимальних рішень у певній предметній області на основі опрацювання інформаційних ресурсів та для оптимізації структури та змісту онтології.

Основою верифікованих, повних, несуперечливих, цілісних інформаційних моделей предметних областей, поданих у формі баз знань, повинні бути об'єктивні знання (збірники наукових статей, енциклопедії, підручники, монографії тощо) у вигляді онтології. Такими предметними областями, в яких знання чітко формалізуються, є наукові дослідження, військова, правова сфера тощо. Це зумовлює проблему розроблення та запровадження уніфікованих методів побудови інтелектуальних систем опрацювання інформаційних ресурсів з використанням онтологічного підходу для підвищення ефективності як баз знань, так і процесів функціонування таких систем. У статті пропонується вирішити цю проблему у вигляді теоретично обґрунтованих моделей функціонування та методів побудови інтелектуальних систем на основі адаптивних онтологій, суть яких полягає в адаптації баз знань цих систем до специфіки задач відповідної предметної області за рахунок ваг важливості термінів та відношень онтологій. Також пропонується розробити методи автоматизованої розбудови та оптимізації онтологій, критеріями яких слугує стандарт якості інформаційних систем ISO/IEC 12207:2008.

Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями

Аналіз основних підходів, методів та засобів побудови інтелектуальних систем опрацювання інформаційних ресурсів і напрямів досліджень використання онтологій показує, що у складі таких систем використовуються не всі можливості онтологій. Переваги використання онтологій порівняно з іншими методами побудови баз знань очевидні, оскільки саме онтології відображають об'єктивні знання та слугують стандартом інженерії знань. Зокрема, не розв'язаними є такі задачі: моделювання процесів опрацювання інформаційних ресурсів та виведення нових знань на основі онтологій; критерії наповнення та оптимізації онтологій; оцінка новизни знань онтологій. Розв'язання цих задач необхідне для побудови ефективних прикладних інформаційних систем опрацювання інформаційних ресурсів для аналізу текстових масивів даних при формуванні спільних тематик досліджень у міжвузівських консорціумах.

Об'єктом дослідження в межах статті є процес побудови інтелектуальних систем опрацювання інформаційних ресурсів для аналізу текстових масивів даних при формуванні спільних тематик досліджень у міжвузівських консорціумах.

Предметом дослідження є методи та засоби побудови інтелектуальних систем опрацювання інформаційних ресурсів (сховищ та баз даних, текстових документів) на основі онтологічного підходу для аналізу текстових масивів даних при формуванні спільних тематик досліджень у міжвузівських консорціумах (наприклад, для програм ERASMUS, TEMPUS – автоматичне визначення спільних тематик досліджень між різними командами науковців різних навчальних закладів в різних країнах). ERASMUS – це програма обмінів студентів, викладачів та науковців країн-членів Євросоюзу, а також країн-учасників програми ERASMUS. Програма надає можливість

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

навчатися, проходити стажування чи викладати в іншій країні, що бере участь в програмі. У Єврокомісії ERASMUS назвали найуспішнішою освітньою програмою ЄС і важливим інструментом боротьби з молодіжним безробіттям [10]. Терміни навчання і стажування можуть становити від 3 місяців до 1 року.

Структура досліджень: 1) розроблення методу побудови базової інтелектуальної системи опрацювання текстової інформації. Основні компоненти такої системи: онтологія, лінгвістичний модуль, модуль розпізнавання предикатів, модуль навчання онтології, модуль логічного виведення, модуль автоматичного планування та прийняття оптимальних рішень; 2) розроблення методу машинного навчання онтології; 3) побудова методу розрахунку цінності інформації в базах та сховищах даних, природномовних текстових документах на основі онтології предметної області.

Очікуваними результатами є:

1. Розроблення методів та засобів використання адаптивних онтологій у складі інтелектуальних систем опрацювання інформаційних ресурсів. Для цього пропонується модифікувати структуру традиційних онтологій шляхом введення в їх структуру ваг важливості термінів та відношень. Це дасть змогу, завдяки налаштуванню цих ваг, адаптувати онтологію до специфіки задач предметної області та до потреб користувача системи. Така модель онтології задає не лише експліцитні, а й імпліцитні знання, що дає змогу під час процесу прийняття рішень використовувати набутий інтелектуальною системою досвід та здійснювати процес навчання таких систем.

2. Розроблення математичного забезпечення функціонування інтелектуальних систем опрацювання інформаційних ресурсів на основі адаптивної онтології, що дасть змогу формалізувати процес функціонування таких систем. Буде побудована семантична метрика на основі адаптивної онтології, яка на відміну від інших існуючих семантичних метрик, враховуватиме причинно-наслідкові залежності між термінами, а не лише їх таксономію. Цю метрику буде використано для оцінювання важливості інформаційних ресурсів.

3. Розроблення методу навчання адаптивної онтології, який полягає у налаштуванні ваг важливості її термінів та відношень. Суть методу полягає у визначенні ваг важливості базових понять онтології та їх поширення на всю онтологію для семантичних та ознакових задач, що уможливлює автоматизоване налаштування онтології до специфіки предметної області та задач, які розв’язують в її межах.

4. Розроблення методу оптимізації адаптивної онтології на основі критеріїв стандарту якості інформаційних систем ISO/IEC 12207:2008, який полягає у періодичному її доповненні новими поняттями та зв'язками з вилученням тих елементів, семантичне значення яких для системи найменше. В методі також враховано необхідність виявлення і усунення суперечності та надлишковості під час наповнення онтології, що відповідає критерію її цілісності. Завдяки цьому в складі інтелектуальних систем опрацювання інформаційних ресурсів можна використовувати онтологію, адаптовану до предметної області.

5. Розроблення програмного забезпечення функціонування інтелектуальних систем опрацювання інформаційних ресурсів, яке ґрунтується на побудованих моделях, методах та алгоритмах, що дасть можливість реалізувати окремі компоненти та функціональні модулі інтелектуальних систем опрацювання інформаційних ресурсів, ядром баз знань яких є онтологія.

Результати 1–3 можна науково обґрунтувати та довести на основі використання теорії концептуальних графів, методів штучного інтелекту (машинне навчання, логічне виведення) та теорії оптимального управління. Результати 4–5, отримані на основі практичного досвіду, будуть корисними методичними і технічними напрацюваннями. Пропонована у проекті методика ґрунтується на оцінці новизни (цінності) знань, які пропонується додавати в онтологію за рахунок введення ваг важливості термінів та відношень онтології (адаптивна онтологія). Такий підхід дасть змогу отримати релевантну онтологію для певної предметної області, оптимізувати її структуру та зміст. Отримана таким чином онтологія дасть змогу адекватно аналізувати інформаційні ресурси предметної області для прийняття оптимальних рішень.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

Аналіз останніх досліджень та публікацій

У [4, 5] введено поняття адаптивної онтології, що дає змогу будувати математичні моделі функціонування інтелектуальних систем на основі онтологій. Аналізуючи роботи [4–8, 21–24] загалом, можна зробити висновок, що наукові дослідження в галузі розроблення та використання онтологій під час побудови прикладних інформаційних систем активно розвиваються. Однак сьогодні відсутні фундаментальні дослідження з використання онтологій для прийняття оптимальних рішень на основі опрацювання інформаційних ресурсів у певній предметній області. Таке використання ґрунтується на методах навчання онтологій (ontology learning).

У [1] розв'язано задачу розроблення науково обґрунтованих методів опрацювання інформаційних ресурсів інтелектуальних Web-систем та побудови на їх основі технологічних програмних засобів для створення, поширення і сталого розвитку систем е-бізнесу. У роботі досліджено закономірності, особливості та залежності під час опрацювання інформаційних ресурсів у таких системах. Аналіз наведених чинників дає змогу зробити висновок про існування певного протиріччя між активним розвитком і поширенням ІТ та прикладних інформаційних систем опрацювання інформаційних ресурсів, з одного боку, та порівняно незначним обсягом наукових досліджень з цієї тематики та їх локальністю, з іншого [20, 28–33]. Це протиріччя породжує проблему стримування інноваційного розвитку цього сектору через створення і запровадження відповідних новітніх прогресивних ІТ, що негативно впливає на темпи зростання цієї частини прикладних досліджень.

Наукові дослідження за напрямом використання онтологій під час розроблення та функціонування інформаційних систем, зокрема інтелектуальних систем опрацювання інформації, розпочалися наприкінці минулого століття і зараз динамічно та інтенсивно розвиваються. Основні теоретичні засади формальних математичних моделей онтологій розроблено у роботах Т. Грубера, Дж. Солтона, А. Гомес-Переса, які запропонували розглядати онтологію як тривимірний кортеж; у працях Н. Гуаріно, П. Фолтса, М. Шамбарда наведено методики побудови онтологій та їх можливі шляхи розвитку; Дж. Сова ввів поняття концептуального графу, а М. Монтес-Гомес використав його для подання онтологій; використання онтологій під час функціонування прикладних інформаційних систем описано в роботах Р. Кнаппе, К. Джонса, Е. Кауфмана, Е. Мена, М. Бориса, А. Каллі, І. П. Норенкова, М. Ю. Уварова, Ю. В. Рогушина; проблему побудови інтелектуальних систем на основі онтологій розглянуто у роботах Т. Андреасена, Т. Бернерса-Лі, Д. Хендлера, О. Лазсіла, О. В. Палагіна, А. В. Анісімова, А. Я. Гладуна, зокрема опрацювання української природної мови та інформаційних ресурсів. Аналізуючи наукові дослідження закордонних вчених у галузі опрацювання інформаційних ресурсів на основі онтологій, доходимо висновку, що навчання онтологій охоплює такі аспекти, як оцінювання якості онтології, видобування знань з різнорідних джерел для розбудови онтології, розроблення методики інтеграції різних онтологій у межах певної предметної області. Сьогодні можна виділити такі напрями досліджень, пов'язані з навчанням онтологій та їх використанням в інтелектуальних системах опрацювання інформаційних ресурсів:

1. Навчання онтологій на основі аналізу природномовних текстів [2, 11–13].
2. Використання онтологій під час побудови інтелектуальних систем підтримки прийняття рішень [14].
3. Розвиток програмних засобів розроблення онтологій як вручну, так і в автоматизованому режимі (OntoEdit, Ontosaurus, OpenCyc Knowledge Server, Protégé) [15–17].
4. Використання онтологій для розв'язування прикладних задач на основі запитів до бази знань [2, 18].
5. Розвиток мов опису онтологій (XML, OWL, RDF, DAML+OIL) [19].

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

Формування мети

Erasmus Mundus – освітня програма ЄС, спрямована на активізацію міжнародного співробітництва та підвищення мобільності серед студентів, викладачів, науковців європейських університетів та вищих навчальних закладів третіх країнах на всіх континентах [10]. Намагаючись перетворити ЄС на світового лідера в освіті, а європейські університети – на осередки знань і центри інновацій, програма Erasmus Mundus також має на меті сприяння взаєморозумінню між людьми, активізацію міжкультурного діалогу.

Програму Erasmus Mundus започаткував Європейський Союз в 2004 році для країн, які не входять до ЄС. Студенти старших курсів і науковці з різних країн мають змогу отримувати стипендії від ЄС для продовження навчання або проведення наукових досліджень у країнах ЄС.

Erasmus Mundus має три напрями:

1. Спільні курси і програми Erasmus Mundus: спільні магістерські та докторські курси та програми, а також стипендії для студентів цих програм;

2. Партнерство Erasmus Mundus: утворення партнерств між університетами ЄС та третьої країни з метою обміну студентами всіх академічних рівнів навчання та обміну науково-педагогічними кадрами. За цим компонентом програми з конкретною метою утворюється партнерство між європейськими університетами та вищими навчальними закладами третьої країни або групи країн. Передбачено двосторонній обмін, коли в рамках утвореного партнерства в українських університетах викладають зарубіжні викладачі та вчаться студенти-іноземці.

3. Проекти з пропагування освіти взагалі та міжнародного співробітництва в освіті зокрема. Проекти в рамках зазначеного компоненту спрямовані на підвищення привабливості вищої освіти в ЄС, її популярізації в третіх країнах, посилення міжкультурного співробітництва, поширення та поглиблення міжкультурного діалогу.

Участь у цьому компоненті програми можуть брати вищі навчальні заклади та науково-дослідні інституції з країн-членів ЄС і третіх країн, а також інші організації освітньої галузі, орієнтовані на підтримку та сприяння розвитку сфери вищої освіти. Мінімальний склад партнерства – мінімум 3 інституції країн ЄС та 2 з інших країн, які можуть брати участь у програмі Erasmus Mundus [10]. Участь у першому напрямі можуть брати:

– українські студенти або випускники, які вже здобули вищу освіту (рівні бакалавра, магістра або їх еквіваленти), аспіранти, докторанти, молоді науковці. Охочим навчатися за програмою Erasmus Mundus потрібно обрати програму із списку та подати заявку, виконуючи інструкції та рекомендації розробників курсу.

– українські науковці та викладачі на запрошення від своїх колег з європейських університетів теж можуть долучитися до розроблення курсів і програм Erasmus Mundus (за чинними умовами, треті країни, включаючи Україну, не можуть виступати ініціаторами та заявниками на створення консорціуму та розроблення курсів і програм Erasmus Mundus).

Українці в рамках програми можуть навчатися за понад 140 магістерськими та 40 докторськими програмами за кордоном. Окрім доступу до нових країн і знань, учасники ERASMUS MUNDUS отримують стипендії: 1000 євро на місяць для магістрантів, 2800 євро – для докторантів [10]. У конкурсі на 2013–2014 навчальний рік 87 українських студентів вибороли право на проходження навчання за програмою Erasmus Mundus. Загалом у 2004–2013 роках гранти на навчання за магістерськими та докторськими програмами Erasmus Mundus отримали 329 українських студентів. У 2012 році переможцями чергового конкурсу за компонентом "Партнерство Erasmus Mundus" стали дев'ять партнерств за участю сімнадцяти українських вищих навчальних закладів. На наступні чотири роки учасниками програм академічної мобільності із загальним обсягом фінансування понад 35 мільйонів євро мають стати близько 400 студентів, викладачів, адміністративний персонал університетів Азербайджану, Білорусі, Вірменії, Грузії, Молдови та України; їх прийматимуть університети Великої Британії, Ірландії, Іспанії, Італії, Нідерландів, Німеччини, Польщі, Португалії, Румунії, Словаччини, Фінляндії, Франції та Швеції.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

З 2007 року реалізовано чотири проекти за участі Міжнародного освітньо-консультативного центру, Одеської національної академії харчових технологій, Національного університету харчових технологій, Національного технічного університету "Київський політехнічний інститут" [10].

Кількість студентів, що було обрано для навчання на магістерських програмах [10]

2004–2013	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
329	4	23	18	24	34	29	30	28	52	87

Кількість обраних для навчання на докторських програмах [10]

2010–2013	2010	2011	2012	2013
14	1	4	2	7

У 2014–2020 роках у Євросоюзі діятиме нова програма ERASMUS+. Вона поєднає сім чинних європейських програм у сфері освіти, науки та спорту, зокрема й ERASMUS та Erasmus Mundus [10]. У межах ERASMUS+ надаватимуть більше стипендій. Програма надаватиме фінансування для більшої мобільності, міжнародних партнерств та спільних науково-дослідних проектів, а також для розширення можливостей та розвитку кадрів у країнах-партнерах по всьому світу [10].

Програма "ТЕМПУС" (TEMPUS) – Схема співпраці між країнами ЄС і країнами-партнерами в галузі вищої освіти. "ТЕМПУС" – транс'європейська програма взаємообмінів між університетами (Trans-European Mobility Programme for University Studies, TEMPUS), розпочалася 1990 року на підтримку реформи вищої освіти в країнах Центральної та Східної Європи; відтоді вона тричі оновлювалась ("ТЕМПУС II", "ТЕМПУС Iibis" і "ТЕМПУС III" – на 2000 — 2006 рр.). Нині до країн-партнерів ЄС у цій програмі належать західнобалканські країни, країни Східної Європи та Центральної Азії (так звані "країни "ТАСІС") і країни Середземномор'я.

Програма фінансує міжуніверситетську співпрацю в сфері:

- розроблення навчальних програм;
- управління університетами;
- взаємодії науковців та громадянського суспільства;
- партнерства освіти і бізнесу;
- структурних реформ у сфері вищої освіти.

З 2007 року розпочато новий етап програми – Tempus IV (2007-2013).

Національний Темпус-офіс в Україні координує взаємодію між Представництвом Європейського Союзу в Україні, Міністерством освіти та науки України, Європейським виконавчим агентством з питань освіти, аудіовізуальних засобів та культури, яке відповідає за впровадження Програми Темпус, Генеральним директоратом європейської комісії з питань освіти та культури, вищими навчальними закладами, іншими зацікавленими сторонами.

Виділення проблеми

Існує мовний бар'єр в обміні інформацією. Більшість наукових досліджень опубліковано рідною мовою науковців, тому є певні перепони для ознайомлення з результатами наукових досліджень (якщо вони опубліковані не англійською або рідною мовою дослідника) для учасників сучасного наукового простору як в ЄС, так і поза його межами. З побудовою онтологій класифікацій сучасних наукових досліджень певної предметної області (наприклад, комп'ютерних наук та інформаційних технологій) та використанням онтологій перекладачів професійної тематик (наприклад, технічний переклад) значно спрощується процес пошуку необхідної інформацій та визначення множини актуальних авторських робіт з певної шуканої предметної області. Це ж допомагає визначити спільні тематики наукових досліджень учасників програми ERASMUS для обміну досвідом та пошуку партнерів для наукових досліджень та участі в проектах.

Аналіз отриманих наукових результатів

Використовуючи чітку ієрархію визначення належності наукових публікацій (авторство, тематика, навчальний заклад, де провадилось дослідження, країна, ключові слова роботи тощо), можна спростити процес ідентифікації спільних тематичних напрямів наукових досліджень (рис. 1).

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

Рис. 1. Ієрархія для загальної класифікації наукових робіт

При побудові *онтології* = {*база знань*, *правила виведення*} для сучасних напрямів наукових досліджень спрощується процес контентного пошуку спільних характеристик у різних роботах різних авторів, написаних різними мовами, тобто $\exists w_i \in \{R\}$, де w_i – опублікована наукова робота, яка може належати множині спільних рубрик R, але одній рубриці вона все-таки більше належить, що і формує основний напрям дослідження автора цієї роботи.

Модель процесу аналізу текстових масивів даних: при формуванні спільних тематик досліджень у міжвузівських консорціумах введемо деякі позначення

$$M =< C, U, P, R, T, A, W, K, a, b, g >,$$

де C – країни; U – навчальні заклади; P – напрями наукових досліджень; R – рубрики напрямів досліджень; T – теми наукових робіт; A – автори наукових робіт; W – опубліковані наукові роботи (статті, тези конференцій, монографії, посібники, підручники); K – ключові слова; a – процес пошуку спільних робіт за ключовими словами; b – процес формування множини спільних рубрик згідно знайдених наукових робіт; g – процес визначення спільних напрямів наукових досліджень між учасниками міжнародних освітянських програм для формування міжвузівських консорціумів обміну досвідом та пошуку партнерів для наукових досліджень та участі в проектах.

Насамперед учасниками наукових консорціумів є навчальні заклади та університети U, де $U_i^C = \{u_{i1}^C, u_{i2}^C, ..., u_{in_U}^C\}$ – множина навчальних закладів c_i країни при $U_i^C \subset U_i$, $U_i^C \cup U_i^{\overline{C}} = U_i$, $\bigcup_{i=1}^{n_S} U_i = U^S$, де U^S – навчальні заклади як учасники програми ERASMUS при $U^S \subset U$, $U^S \cup U^{\overline{S}} = U$, $C = \{c_1, c_2, ..., c_{n_C}\}$ та $\exists c_i \forall u_{ij}^C Belongs(c_i, u_{ij}^C)$, де *Belong* – оператор належності c_i країні u_{ij}^C навчального закладу.

Якщо напрями досліджень $P = \{p_1, p_2, ..., p_{n_P}\}$, рубрики досліджень $R^P = \{r_1^P, r_2^P, ..., r_{n_R}^P\}$, автори наукових робіт $A^U = \{a_1^U, a_2^U, ..., a_{n_A}^U\}$ з певних навчальних закладів $U_i^C = \{u_{i1}^C, u_{i2}^C, ..., u_{in_U}^C\}$, теми робіт $T^R = \{t_1^R, t_2^R, ..., t_{n_T}^R\}$, наукові роботи $W = \{w_1, w_2, ..., w_{n_W}\}$ при $W^A = \{w_1^A, w_2^A, ..., w_{n_{WA}}^A\}$, $W^U = \{w_1^U, w_2^U, ..., w_{n_{WU}}^U\}$, $W^R = \{w_1^R, w_2^R, ..., w_{n_{WR}}^R\}$, $W^P = \{w_1^P, w_2^P, ..., w_{n_{WP}}^P\}$, $W^T = \{w_1^T, w_2^T, ..., w_{n_{WT}}^T\}$, де $W_i^A \subset W_i$, $W_i^U \subset W_i$, $W_i^R \subset W_i$, $W_i^P \subset W_i$, $W_i^T \subset W_i$, $\bigcup_{i=1}^{n_W} W_i = W$, тоді для

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

- формування бази даних наукових робіт $W_i = W_i^A \cup W_i^U \cup W_i^T \cup W_i^R \cup W_i^P$,
- знаходження спільних напрямів досліджень $W_i = (W_i^A \cup W_i^U \cup W_i^T) \cap (W_i^R \cup W_i^P)$
- знаходження спільних тематик $W_i = (W_i^A \cup W_i^U \cup W_i^R \cup W_i^P) \cap W_i^T$

Якщо ключові слова $K = \{k_1, k_2, ..., k_{n_K}\}$, тоді $K = K^W \cup K^R$, де $K^W = \{k_1^W, k_2^W, ..., k_{n_{KW}}^W\}$ ключові слова робіт, а $K^R = \{k_1^R, k_2^R, ..., k_{n_{KR}}^R\}$ ключові слова рубрик.

При наступних вхідних даних

- $\exists r_i \forall k_{ij}^R Belongs(r_i, k_{ij}^R)$;
- $\exists w_i \forall k_{ij}^W Belongs(w_i, k_{ij}^W)$;
- $\exists w_i \exists t_j^R Belongs(w_i, t_j^R)$;
- $\exists t_i^R \exists u_{ij}^C Belongs(t_i^R, u_{ij}^C)$;

існують такі властивості результату аналізу текстових масивів даних зі спільними ключовими словами, де $p_i \in P$, $p_l \in P$, $r_{ij}^P \in R^P$, $r_{lm}^P \in R^P$ для $i = \overline{1, n_P}$, $l = \overline{1, n_P}$, $j = \overline{1, n_R}$, $m = \overline{1, n_R}$

1) $\exists p_i \exists r_{ij}^P \exists p_l \exists r_{lk}^P (Belongs(p_i, r_{ij}^P) \vee Belongs(p_l, r_{lk}^P))$;
2) $\exists w_i \exists k_{ij}^W \exists r_l \exists k_{lj}^R ((Belongs(w_i, k_{ij}^W) \wedge Belongs(r_l, k_{lm}^R)) \rightarrow Belongs(w_i, r_l))$ при $(K^W \cap K^R) \subset K^R$;
3) $\exists w_i \exists t_j^R \exists w_l \exists t_m^R (((Belongs(w_i, t_j^R) \wedge Belongs(w_l, t_m^R)) \rightarrow (Belongs(w_i, t_j^R) \approx Belongs(w_l, t_m^R))) \rightarrow$
$\rightarrow (t_m^R \approx t_j^R))$ при $(w_i \cap w_j) \subset K^R$;
4) $\exists t_i^R \exists u_{ij}^C \exists t_l^R \exists u_{lm}^C (((Belongs(t_i^R, u_{ij}^C) \wedge Belongs(t_l^R, u_{lm}^C)) \rightarrow (Belongs(t_i^R, r_z^P) \wedge Belongs(t_l^R, r_z^P))) \rightarrow$
$\rightarrow (Belongs(t_i^R, p_z) \wedge Belongs(t_l^R, p_z))$.

Правила аналізу наукових робіт за ключовими словами для пошуку спільних тем:

1) $\exists K_i \exists K_j \exists W_i^K \exists W_j^K ((K_i \cap K_j) \rightarrow (W_i^K \cup W_j^K))$;
2) $\exists K_i \exists K_j \exists W_i \exists W_j \exists T_i \exists T_j (((K_i \cap K_j) \cap (W_i^K \cup W_j^K)) \rightarrow (T_i \approx T_j))$;

при таких властивостях множин

- $K_i \subseteq K_i^W$, $K_j \subseteq K_j^W$, $(K_i \cup K_j) \subseteq K^W$
- $W_i^K \subseteq W_i^T$, $W_j^K \subseteq W_j^T$, $(W_i^K \cup W_j^K) \subseteq W^K \subseteq (W_i^T \cup W_j^T) \subseteq W^T$, тоді $(W_i^K \cup W_j^K) \subseteq W^T$
- $T_i \subset T^R$, $T_j \subset T^R$, $(T_i \cup T_j) \subset T^R$

Побудуємо основні відношення на множинах K_i, W_i^K, T_i^R, W_i^T, W_i^A, W_i^R, R_i^P, W_i^P, P_i^U, W_i^U, U_i^C, C_i^E, U_i^S та C_i:

1) $R_{i1} = \{(K_i, W_i^K) \mid K_i \in W_i^K\}$
2) $R_{i2} = \{(W_i^K, T_i^R) \mid W_i^K \in T_i^R\}$
3) $R_{i3} = \{(W_i^T, T_i^R) \mid W_i^T \in T_i^R\}$
4) $R_{i4} = \{(W_i^A, T_i^R) \mid W_i^A \in T_i^R\}$
5) $R_{i5} = \{(W_i^R, R_i^P) \mid W_i^R \in R_i^P\}$
6) $R_{i6} = \{(W_i^P, P_i^U) \mid W_i^P \in P_i^U\}$
7) $R_{i7} = \{(W_i^U, U_i^C) \mid W_i^U \in U_i^C\}$
8) $R_{i8} = \{(T_i^R, R_i^P) \mid T_i^R \in R_i^P\}$

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

9) $R_{i9} = \{(R_i^P, P_i^U) \mid R_i^P \in P_i^U\}$

10) $R_{i10} = \{(P_i^U, U_i^C) \mid P_i^U \in U_i^C\}$

11) $R_{i11} = \{(U_i^C, C_i^E) \mid U_i^C \in C_i^E\}$, C_i^E – країни, які підтримують програми типу Erasmus Mundus при $U^C \subset U$

12) $R_{i12} = \{(U_i^S, C_i) \mid U_i^S \in C_i\}$, U^S – навчальні заклади як учасники програми ERASMUS при $U^S \subset U$ та $(U^C \cup U^S) \subseteq U$, $C_i^E \subseteq C_i$

Алгоритм аналізу текстових масивів даних при формуванні спільних тематик досліджень у міжвузівських консорціумах

Етап 1. Знаходження ключових слів у наукових роботах.

Етап 2. Рубрикація наукових робіт згідно із знайденими ключовими словами з порівнянням тематичних словників рубрик (задача класифікації текстових документів).

Етап 3. Визначення множини перетину тематик наукових робіт з різних навчальних закладів зі спільними рубриками.

Етап 4. Визначення списку авторів з перетину множини робіт.

Етап 5. Визначити список навчальних закладів з перетину множин авторів аналізованих робіт.

Етап 6. Формування рейтингу навчальних закладів за кількістю спільних тематик та активності в наукових дослідженнях та спільних тематиках авторів наукових робіт з цих навчальних закладів.

Етап 7. Формування множини актуальних спільних тематик наукових робіт авторів з різних навчальних закладів.

Етап 8. Формування множини авторів наукових робіт зі спільними науковими тематиками.

Етап 9. Формування множини навчальних закладів, які можуть провадити спільні наукові дослідження.

Задача класифікації тестових документів (ТД) визначається так. Існує множина текстів наукових робіт $W = \{w_1, w_2, ..., w_{n_W}\}$; множина $R^P = \{r_1^P, r_2^P, ..., r_{n_R}^P\}$ рубрик, які розглядатимемо як класи Class=$\{\text{Class}_1, \text{Class}_2, \ldots, \text{Class}_N\}$. Кожна рубрика подається деяким описом (онтологією $\hat{O}$), яка має певну внутрішню структуру. Класифікація f текстів наукових робіт $w_i \in W$ полягає у виконанні певних процедур, на основі яких робиться висновок про відповідність w_i одній із структур Class_j, що означає зарахування w_i до класу Class_j, або висновок про неможливість класифікації w_i. Тобто у множину класів необхідно додати пустий клас Class_0, якому ставитимемо у відповідність тексти наукових робіт, які не були прокласифіковані. У нашому випадку елементами множини T є електронні версії сканованих текстових наукових робіт. Отже, загальну модель класифікації ТД подамо у вигляді

$$f : W \xrightarrow{\hat{O}} \text{Class}. \qquad (1)$$

Крім того, існує чітка відповідність між класом та його онтологією, який виражається як структурою онтології, так й важливістю термінів (ключових слів), які належать відповідній рубриці. Тобто класу відповідає деяка адаптивна онтологія (АО) [3]:

$$\text{Class} \leftrightarrow \hat{O}_i = \left\langle \hat{B}_i, \hat{Q}_i, F \right\rangle.$$

де $\hat{B}_i =< B_i, \Delta_i >$ – підмножина понять онтології B_i з їх вагами Δ_i; $\hat{Q}_i =< Q_i, n_i >$ – підмножина відношень онтології Q_i з їх вагами n_i; F – ітерпретація елементів онтології. Ітерпретація F та важливість відношень Q не змінюються, а залишаються однаковими для всіх класів. Відображення f не має ніяких обмежень, так що можливі ситуації, коли деякий текст наукової роботи можна зарахувати до декількох рубрик одночасно. Крім сформульованої задачі класифікації визначається

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

завдання навчання g рубрикатора, під якою мають на увазі редагування адаптивної онтології, яка відповідає рубрикатору:

$$g : \hat{O} \to \hat{O}. \qquad (2)$$

Основні види класифікаторів. Для розв'язування задачі (2) використовують різні види класифікаторів.

1. *Статистичні класифікатори на основі ймовірнісних методів*. Найвідомішими серед них є сімейство байєсівських класифікаторів. Загальною ознакою для таких методів є процедура f, в основу якої покладено формулу Байєса для умовної ймовірності. Аналізований текст наукової роботи w_i подаємо у вигляді послідовності термінів $(B_{i_1}, B_{i_2}, ..., B_{i_{N_i}})$. Кожна рубрика Pr_j характеризується безумовною ймовірністю її вибору $p(\mathrm{Pr}_j)$ під час класифікації деякого тексту наукової роботи w_i (сукупність таких подій для всіх рубрик утворюють систему гіпотез, де $\sum_{j=1}^{N} p(\mathrm{Pr}_j) = 1$), та умовною ймовірністю $p(B \mid \mathrm{Pr}_j)$ – зустріти термін B у тексті наукової роботи w_i за умови вибору рубрики Pr_j. Ці величини утворюють елементи V_j множини $V = \{V_1, V_2, ..., V_N\}$ описів рубрик і використовуються під час розрахунку ймовірностей $p_{ij} = p(w_i \mid \mathrm{Pr}_j)$ того, що текст наукової роботи w_i буде класифікований за умови вибору рубрики Pr_j. Під час розрахунку p_{ij} враховується подання тексту наукової роботи w_i у вигляді послідовності термінів $(B_{i_1}, B_{i_2}, ..., B_{i_{N_i}})$. Підставивши ці величини у формулу Байєса, отримуємо ймовірність

$$\tilde{p}_{ji} = p(\mathrm{Pr}_j \mid w_i) = \frac{p(\mathrm{Pr}_j) \cdot p(w_i \mid \mathrm{Pr}_j)}{\sum_{k=1}^{N} p(\mathrm{Pr}_k) \cdot p(w_i \mid \mathrm{Pr}_k)}$$

того, що буде обрана рубрика Pr_j, за умови, що текст наукової роботи w_i пройде успішну класифікацію. Процедура f зводиться до підрахунку $\tilde{p}_{ji}$ для всіх рубрик Pr_j і вибору тієї, для якої ця величина максимальна. Навчання рубрикатора зводиться до складання словника $(B_1, B_2, ..., B_{N_K})$ та визначення для кожної рубрики величин $p(\mathrm{Pr}_j)$ і $p(B \mid \mathrm{Pr}_j)$, де $B \in (B_1, B_2, ..., B_{N_K})$.

2. *Класифікатори, що використовують методи на основі штучних нейронних мереж.* Цей вид класифікаторів добре зарекомендував себе в задачах розпізнавання зображень. Опис класів Pr, як правило, являє собою багатовимірні вектори дійсних чисел, закладені в синаптичних важливостях штучних нейронів, а процедура класифікації f характеризується способом перетворення аналізованого тексту наукової роботи w_i до аналогічного вектору, видом функції активації нейронів, а також топології мережі. Навчання класифікатора в цьому разі збігається з процедурою навчання нейронної мережі та залежить від обраної топології [9].

3. *Класифікатори, засновані на функціях подібності.* Характерною ознакою цього методу є універсальність описів V, які, з одного боку, використовуються для подання змісту рубрик, а з іншого, – змісту аналізованих текстів. Процедура класифікації f використовує міру подібності вигляду $d : V \times V \to [0,1]$, що дає змогу кількісно оцінювати тематичну близькість описів $V_W \in V$ і $V_i \in V$, де опис V_W подає зміст аналізованого тексту наукової роботи, а V_i – зміст i-ї рубрики. Дії процедури класифікації f зводяться до перетворення аналізованого тексту W на подання $V_W \in V$, оцінки подібності опису V_W з описами рубрик V_i. Тобто обчислюється відстань між описами $d(V_W, V_i)$. На основі цієї відстані здійснюється висновок про належність тексту наукової роботи одній або декільком рубрикам. Найхарактернішим для таких класифікаторів є використання лексичних векторів моделі терм-документ в описах V, які так само застосовуються і в нейронних

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

класифікаторах [3]. По суті пропонований підхід є частинним випадком цих класифікаторів. Описами V пропонуємо використовувати адаптивну онтологію рубрик $\hat{O}$, як відстані використовувати метрику Дайеса [26]. Коефіцієнт Серенсена–Дайса є бінарною мірою подібності стрічок і має в загальному вигляді такий запис:

$$d(V_W, V_i) = \frac{2 \times n(V_W \cap V_i)}{n(V_W) + n(V_i)}.$$

де $d(V_W, V_i)$ – коефіцієнт подібності описів або $d(V_W, V_i)$, $0 \le p \le 1$; V_W і V_i – описи аналізованих текстів; $n(V)$ – функція, що визначає ключові слова у описі V.

4. *Кластерний аналіз*. Метод, пов'язаний з поданням даних в n-вимірному просторі, є кластерним аналізом і належить до групі статистичних методів автоматичної класифікації. Він дає змогу кількісно формалізувати лінгвістичні знання. Це допомагає наочно зображати структуру текстів з різноманіттям їх лексичних, синтаксичних, семантичних зв'язків і може слугувати базою для виконання багатьох складних завдань автоматизованого опрацювання тексту (інформаційного пошуку, автоматизованого перекладу, атрибуції авторства тощо). Основна мета кластерного аналізу – виділити у вихідних даних такі однорідні групи, об'єкти яких були схожі (подібні) в певному сенсі один на одного, а об'єкти з різних груп – не схожі (різнорідні). Процедури кластер-аналізу, які проводять кількісну систематизацію вихідних даних, виділяють структуру даних на основі попарного порівняння елементів вихідного масиву. Самі об'єкти, що підлягають структурній класифікації (кластеризації), розташовуються в просторі, вимірами якого є ознаки. Якщо ознак n, то простір n-вимірний, але у зв'язку з можливою кореляцією ознак розмірність простору може бути зменшена. Причому елементи одного кластера не обов'язково формуються у компактній ділянці простору ознак, а можуть бути *розсипані* у ньому. До того ж приналежність до класу визначається: 1) загальними значеннями всіх або декількох ознак; 2) найбільшою кількістю загальних значень ознак. Перший тип класифікації називається монотетичний і корисний у побудові класів для дуже вузької групи завдань (наприклад, створення таксономічних ключів). Другий тип є політетичною класифікацією. Принцип її побудови є основою для породження природної систематизації об'єктів. Більшість методів кластерного аналізу спрямовані на отримання політетичних класів.

Незважаючи на широке застосування кластерного аналізу, загальноприйнятого визначення кластера не існує. Є лише інтуїтивне уявлення про те, що елементи одного кластера ближчі один до одного, ніж до інших елементів, які розташовані поза цим кластером.

Також не існує точної постановки задачі кластерного аналізу. Однак на основі загального поняття про кластери формулюється така задача. Множину текстів $W = \{w_1, w_2, ..., w_{n_w}\}$ зарахувати до деякого елемента множини Pr, використовуючи деякий критерій подібності елементів. Подібність елементів множини текстів наукових робіт w_i зумовлюється наявністю множини ознак $K^W = \{k_1^W, k_2^W, ..., k_{n_{KW}}^W\}$, які характеризують елементи множини V. Виміряні значення ознак для елемента w_i утворюють вектор $E_i = (e_{i1}, e_{i2}, ..., e_{in_{KW}})$. Ввівши для множини векторів $E = (E_1, E_2, ..., E_M)$ $U = \{U_1, U_2, ..., U_M\}$ деяку міру їх подібності, можна виконати різні процедури групування (кластеризації) текстів наукових робіт.

Основні підходи до розроблення інтелектуальних систем класифікації текстових наукових робіт. Інформаційні ресурси зображено у вигляді певного формату даних. Таким форматом може бути текстовий документ, web-сторінка тощо. Головним джерелом інформаційних ресурсів є текст. Текст – це інформація, подана за допомогою сукупності символів.

Для роботи з текстом необхідно визначити дві його характеристики:

- кодування – метод зображення символів певної мови або групи мов.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

- мова тексту – сукупність довільно відтворюваних загальноприйнятих у межах конкретного суспільства звукових знаків для об'єктивно існуючих явищ і понять, а також загальноприйнятих правил їх комбінування під час вираження думок.

Текст складається з речень. Речення – граматична конструкція, побудована з одного чи кількох слів певної мови, яка містить окрему, порівняно незалежну думку. Речення, своєю чергою, складаються зі слів і можуть містити розділові знаки. Слово – низка морфем, об'єднаних згідно з граматичними правилами певної мови і співвідносних з певним елементом позамовної реальності. Розділові знаки – сукупність графічних знаків, що не належать до алфавіту конкретної мови, а слугують засобом відображення тих ознак (сторін) писемної мови, які не можна передати літерами та іншими писемними позначеннями: цифрами, знаками рівності, подібності тощо. До розділових знаків в українській мові належать: крапка, кома, крапка з комою, двокрапка, тире, дефіс, три крапки, знак оклику, знак запитання, лапки, дужки, абзац. Для аналізу структури слів використовується морфологічний аналіз, про що вже зазначалось у третьому розділі.

Розроблена інтелектуальна система класифікації текстових ресурсів складається із чотирьох підсистем (див. рис. 2).

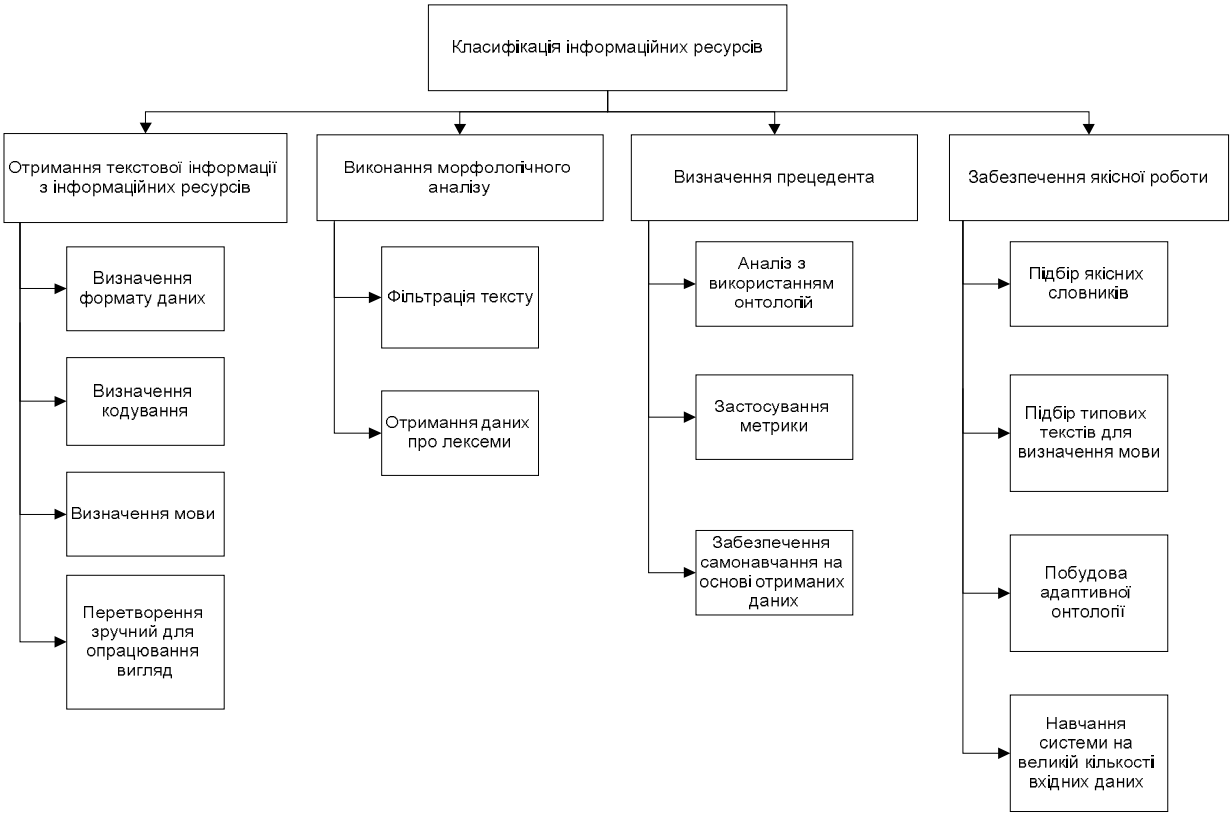

Рис. 2. Декомпозиція модулів інтелектуальної системи класифікації текстових ресурсів

Основне завдання системи – класифікація інформаційних ресурсів. Для виконання поставленої задачі необхідно виконати підзавдання. Першим таким підзавданням є отримання текстової складової з інформаційних ресурсів, тобто необхідно з деякого початкового фізичного зображення інформаційного ресурсу отримати текст та базову інформацію про нього. Для цього потрібно визначити формат даних та чи підтримується він системою (перелік форматів, які підтримуються розробленою системою, наведено у табл. 1). Якщо система підтримує цей формат, необхідно видобути з нього текстову складову. Часто текст, отриманий у результаті попередньої дії, подано у певному кодуванні. Подальша робота з різними кодуваннями може призвести до низки проблем, тому необхідно деяким способом визначити його кодування. Для того, щоб працювати з текстовою інформацією, необхідно знати, якою мовою або мовами він написаний, тобто необхідно

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

визначити мову тексту. Отримання текстової складової завершується зображенням усієї отриманої інформації у зручному для подальшого опрацювання вигляді.

Таблиця 1

Формати даних, які підтримуються інтелектуальною системою

Назва формату	Розширення файла	Тип mime	Коментар
Microsoft Word Document	doc	application/msword	Найпопулярніший формат для текстів наукових робіт
Office Open XML	docx, docm	application/vnd. openxmlformats-officedocument	Стандарт ECMA-376, ISO/IEC 29500
Open Document Format for Office Applications	odt	application/vnd. oasis.opendocument text	Стандарт ISO/IEC 26300:2006
Portable Document Format	pdf	application/pdf	Стандарт ISO 320-1:2008
Rich Text Format	rtf	text/rtf	
Plain text	txt	text/plain	
Hyper Text Markup Language	html, htm	text/html	Стандарт W3C HTML 4.01

Наступне підзавдання – морфологічний аналіз. Для цього насамперед необхідно виконати фільтрацію тексту від будь-яких засобів розмітки, розділових знаків тощо та подати його як набір слів. Після цього необхідно проаналізувати слова та на основі цього отримати дані про лексеми, які можна використати для класифікації.

Визначення класу (рубрики) полягає в опрацюванні набору отриманих лексем на основі адаптивної онтології. Для цього використовують формули (1)–(2), на основі яких приймається рішення про приналежність вихідного тексту до певної рубрики. Ваги важливості понять класів онтології обчислені на основі статистичного аналізу текстових документів, для яких відомо, до якої рубрики вони належать. Тобто при кожному входженні деякого поняття B_i у текст, який належить до класу Class_j, важливість цього поняття збільшувалась на одиницю $\Delta_{ji} = \Delta_{ji} + 1$, тобто використано частотний аналіз. Очевидно, що на початку всі важливості $\Delta_{ji} = 0$.

Одним із пріоритетів в розробленні системи є забезпечення якісної роботи. Передусім необхідно підібрати словники, що містять якнайбільшу кількість слів та мають якнайбільше можливостей для виконання морфологічного аналізу. Необхідно навчити систему якісно розпізнавати мову методом навчання й основне – побудувати якісну адаптивну онтологію, яка точно визначає рубрику. Форматом словників було обрано формат, який використовується для словників Hunspell. Для кожної мови використано декілька файлів, а саме словник, який містить слова, файл афіксів, який визначає значення спеціальних позначок у словнику, файл стоп-слів, які фільтруються у разі визначення термінів, та файл біграм, який використовується для визначення мови. Цей файл використовується під час використання *N*-грамних моделей для визначення мови [25]. Для автоматизованого визначення кодування використано метод розподілу символів [27]. Основні модулі системи написано мовою програмування Python, онтологію розроблено в редакторі Protégé-OWL.

Процес рубрикації текстового контенту

Аналіз лексико-граматичної та семантико-прагматичної побудови тексту використовують для автоматичної рубрикації контенту та формування дайджестів, що призводить до формування тематично підібраних масивів контенту (табл. 2).

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

Таблиця 2

Основні етапи рубрикації текстового контенту

Назва	Призначення етапу
Підготовка	Визначення тематики/мети/об`єкта аналізу, хронологічні та географічні рамки, принципи відбору.
Збір даних	Формування класифікатора відбору ключових цитат та інструкції для кодувальника.
Формальний аналіз	Перетворення фрагментів тексту без аналізу його змісту. Морфологічні дані забезпечують доступ до змісту, опосередкованого через співвідношення одиниць змісту з одиницями виразу.
Змістовий аналіз	Аналіз елементів і логіко-семантичних відношень між ними для подання семантики контенту.
Синтаксичний аналіз	Автоматично за наявності лексико-граматичних та граматичних даних до кожного слова синтаксично прив'язують словоформи у реченні.
Морфемний аналіз	Сегментування тексту, де виділення префіксів можливе без знання частин мови, а суфіксів – ні: потрібні різні їх набори та процедури відсікання суфіксів для кожної частини мови окремо.
Класифікація	Автоматичне опрацювання фрагментів текстового контенту для розпізнавання змісту.
Кодування	Кодування фрагментів текстового контенту.
Архівація	Збереження фрагментів текстового контенту в базі даних.

За змістовний аналіз контенту відповідає процес витягування граматичних даних зі слова через графемний аналіз та корегування результатів морфологічного аналізу через аналіз граматичного контексту лінгвістичних одиниць (алг. 1).

Алгоритм 1. Рубрикація текстового контенту

Етап 1. Поділ текстового контенту на блоки.

Крок 1. Подання на вхід блоку побудови дерева блоки текстового контенту.

Крок 2. Створення нового блоку в таблиці блоків.

Крок 3. Накопичення символів до символу нового рядка.

Крок 4. Перевірка на наявність крапки перед символом нового рядка. Якщо є, то перехід до кроку 5, якщо ні, то збереження послідовності у таблиці, розбір нового блоку та перехід до крок 3.

Крок 5. Перевірка наявності кінця тексту. Якщо кінець тексту, то перехід до кроку 6, якщо ні, то зберігається накопичена послідовність у таблицю, розбір нового блоку та перехід до кроку 2.

Крок 6. Отримання на виході дерево блоків у вигляді таблиці.

Етап 2. Поділ блоку на речення зі збереженням структури.

Крок 1. На вхід подається таблиця блоків. Створення таблиці речень із зв'язком за полем Код_розділу типу *n-to-*1 із таблицею блоків.

Крок 2. Створення нового речення в таблиці речень.

Крок 3. Накопичення символів до крапки, крапки з комою або символу нового рядка.

Крок 4. Перевірка на наявність скорочення. Якщо скорочення, то перехід до кроку 5, якщо ні, то збереження послідовності у таблиці, розбір нового речення та перехід до кроку 2.

Крок 5. Перевірка наявності кінця тексту блоку. Якщо кінець тексту, то перехід до кроку 6, якщо ні, то збереження послідовності у таблиці, розбір нового речення та перехід до кроку 2.

Крок 6. Отримують на виході дерево речень у вигляді таблиці.

Крок 7. Перевірка наявності кінця тексту. Якщо кінець тексту, то перехід до кроку 8, якщо ні, то розбір нового блоку та перехід до кроку 1.

Крок 8. Отримання на виході дерева речень у вигляді таблиць.

Етап 3. Поділ речень на лексеми із вказанням належності до речень.

Крок 1. Формування на основі таблиці речень таблиці лексем із полями Код_лексеми (унікальний ідентифікатор), Код_речення (число, рівне коду речення з лексемою), Номер_лексеми (число, що дорівнює номеру лексеми в реченні), Текст (текст лексеми).

Крок 2. Подання на вхід для розбору на лексеми речення з таблиці речень.

Крок 3. Створення нової лексеми в таблиці лексем.

Крок 4. Накопичення символів до крапки, пропусків або кінця речення та збереження в таблиці лексем.

Крок 5. Перевірка кінця речення. Якщо так, то перехід до кроку 6, якщо ні, то збереження накопиченої послідовності у таблиці, розбір нової лексеми та перехід до кроку 3.

Крок 6. Проведення синтаксичного аналізу на основі даних, одержаних на виході.

Крок 7. Проведення морфологічного аналізу на основі даних, одержаних на виході.

Етап 4. Визначення тематики текстового контенту.

Крок 1. Побудова ієрархічної структури властивостей кожної лексичної одиниці тексту, що містить граматичну та семантичну інформацію.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

Крок 2. Формування лексикону з ієрархічною організацією типів властивостей, коли кожен тип-нащадок успадковує і перевизначає властивості предка.

Крок 3. Уніфікація – базовий механізм побудови синтаксичної структури.

Крок 4. Визначення ключових слів *KeyWords* текстового контенту.

Крок 5. Визначення *TKeyWords* – тематичні ключові слова в множині *KeyWords* для *Topic* – тема контенту та *Category* – категорія контенту.

Крок 6. Визначення *FKeyWords* – частота вживання ключових слів та *QuantitativeryTKey* – частота вживання тематичних ключових слів в тексті текстового контенту.

Крок 7. Визначення *Comparison* – порівняння вживання ключових слів різних тематик Розрахунок *CofKeyWords* – коефіцієнт тематичних ключових слів контенту, *Static* – коефіцієнт статистичної важливості термів, *Addterm* – коефіцієнт наявності додаткових термів.. Порівняння множини ключових слів контенту з ключовими поняттями тем. Якщо є збіг, то перехід до кроку 9, якщо ні, то перехід до кроку 8.

Крок 8. Формування нової рубрики з набором ключових понять аналізованого контенту.

Крок 9. Присвоєння визначеній рубриці аналізованого текстового контенту.

Крок 10. Розрахунок *Location* – коефіцієнт розташування контенту в тематичній рубриці.

Етап 4. Заповнення бази пошукових образів для атрибутів *Topic* – тема контенту, *Category* – категорія контенту, *Location* – коефіцієнт розташування контенту в тематичній рубриці, *CofKeyWords* – коефіцієнт тематичних ключових слів в контенті, *Static* – коефіцієнт статистичної важливості термів, *Addterm* – коефіцієнт наявності додаткових термів, *TKeyWords* – тематичні ключові слова, *FKeyWords* – частота вживання ключових слів, *Comparison* – порівняння вживання ключових слів різних тематик, *QuantitativeryTKey* – частота вживання тематичних ключових слів в тексті текстового контенту.

Побудова тексту визначається темою, вираженою інформацією, умовами спілкування, завданням повідомлення та стилем викладення. Із семантичною, граматичною та композиційною структурою контенту пов'язані його стильові/стилістичні характеристики, залежні від індивідуальності автора та підпорядковані тематичній/стильовій домінанті тексту.

Процес рубрикації контенту у вигляді діаграми варіантів подано на рис. 3.

Рис. 3. Діаграма варіантів використання для процесу рубрикації текстового контенту

Основні етапи визначення морфологічних ознак одиниць тексту: визначення граматичних класів слів – частин мови і принципів їх класифікаційного виділення; виокремлення частини семантики слова як морфологічної; обґрунтування набору морфологічних категорій та їх природи; опис сукупності формальних засобів, закріплених за частинами мови та їх морфологічними категоріями. Процес рубрикації через автоматичне індексування складових текстового контенту розбита на послідовні блоки: морфологічний аналіз, синтаксичний аналіз, семантико-синтаксичний аналіз лінгвістичних конструкцій та варіювання змістовного запису текстового контенту (рис. 4).

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

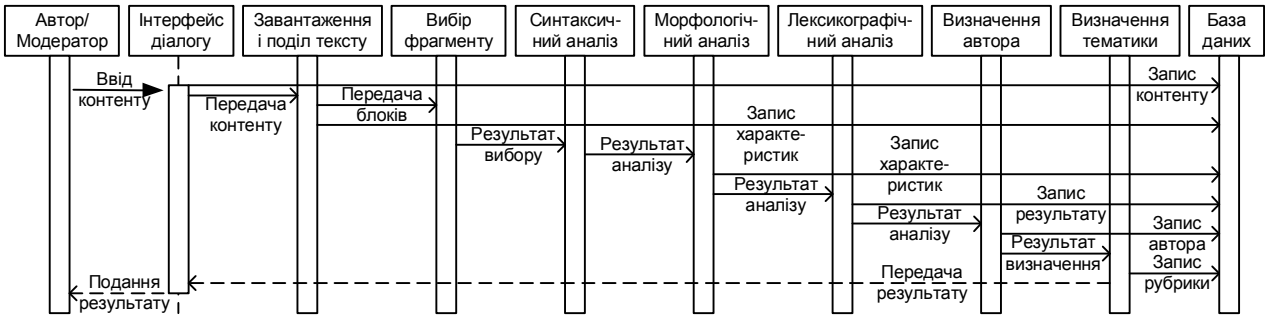

Рис. 4. Діаграма послідовності для процесу рубрикації текстового контенту

Використано такі способи виразу граматичного значення: синтетичний, аналітичний, аналітико-синтетичний та суплетивний. Граматичні значення узагальнені через однотипні характеристики та підлягають поділу на часткові значення. Для позначення класів однотипних граматичних значень використано поняття граматичної категорії. До морфологічних значень належать категорії роду, числа, відмінка, особи, часу, способу, стану, виду, об'єднувані у парадигми для класифікації частин тексту. Об'єктом морфологічного аналізу є структура слова, форми словозміни, способи виразу граматичних значень. Морфологічні ознаки одиниць тексту – це інструменти дослідження зв'язку між лексикою, граматикою, використанням їх у мовленні, парадигматикою (відмінкові форми відмінюваних слів) і синтагматикою (лінійні зв'язки слів, сполучення).

Реалізація автоматичного кодування слів тексту, тобто приписування їм кодів граматичних класів, пов'язане з граматичною класифікацією. Морфологічний аналіз містить такі етапи: виділення основи у словоформі; пошук основи у словнику основ; порівняння структури словоформи з даними у словниках основ, коренів, префіксів, суфіксів, флексій. У процесі аналізу ідентифікують значення слів та синтагматичних відношень між словами контенту. Інструментами аналізу є словники основ/флексій/омонімів та статистичних/синтаксичних словосполучень, зняття лексичної омонімії, семантичний аналіз іменних безприйменникових конструкцій, таблиці семантико-синтаксичного сполучення іменників/прикметників та компонентів прийменникових конструкцій, алгоритми аналізу для визначення послідовностей перевірок і звертань до словника і таблиць; система поділу слів тексту на флексію й основу; тезаурус еквівалентностей для заміни еквівалентних слів одним/кількома номерами понять, які слугують ідентифікаторами змісту замість основ слів; тезаурус у вигляді ієрархії понять для пошуку для цього поняття загального/асоційованого з ним поняття; система обслуговування словників. Процес індексування залежить від дескрипторного словника або інформаційно-пошукового тезауруса (рис. 5).

Дескрипторний словник має структуру таблиці з трьома колонками: основи слів; набори дескрипторів, приписані кожній основі; граматичні ознаки дескрипторів. Індексування складається з виділення інформативних словосполучень з тексту; розшифрування абревіатури; заміна слів з основами-дескрипторами на код дескриптора; зняття омонімії.

Метою роботи є розвиток методів та засобів побудови інтелектуальних систем опрацювання інформаційних ресурсів, ядром баз знань яких є онтології, та підвищення ефективності функціонування таких систем завдяки адаптації онтологій до специфіки задач предметної області.

Запропонований підхід спрямований на вирішення таких завдань:

– аналіз специфіки та методів побудови інтелектуальних систем опрацювання інформаційних ресурсів шляхом їх класифікації за принципом та простором функціонування, що дасть змогу виділити типи таких систем в залежності від математичного забезпечення їх функціонування;

– розроблення математичних моделей та методів функціонування різних класів інтелектуальних систем опрацювання інформаційних ресурсів, ядром баз знань яких є онтології, для дослідження ефективності таких систем;

– розроблення методів та алгоритмів адаптування онтології до специфіки задач предметної області з метою підвищення ефективності функціонування інтелектуальних систем опрацювання інформаційних ресурсів в межах цієї предметної області;

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

– визначення характеристик якості інтелектуальних систем опрацювання інформаційних ресурсів та критеріїв оптимізації онтологій у складі їх баз знань для підвищення показників ефективності функціонування таких систем.

Обґрунтування актуальності та/або доцільності виконання завдань, виходячи із:

– стану досліджень проблематики і тематики;

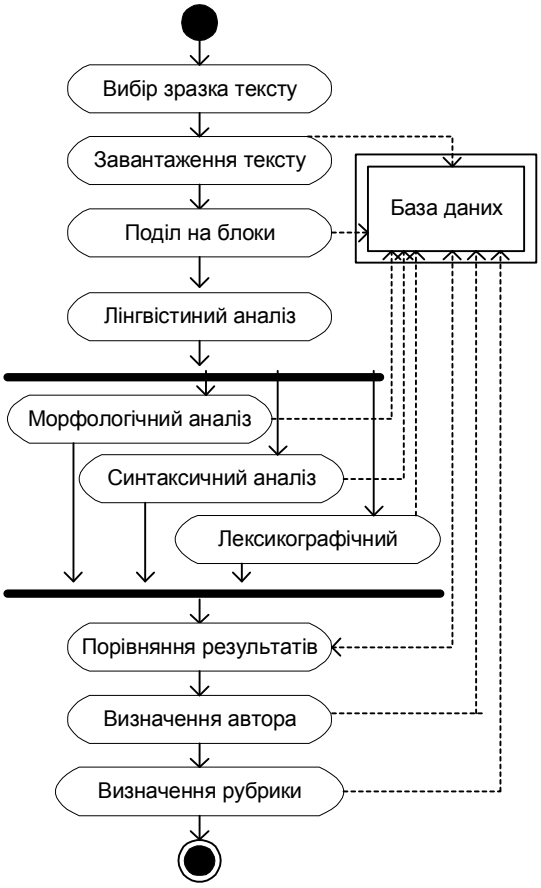

Рис. 5. Діаграма діяльності підсистеми рубрикації текстового контенту

У роботі пропонується агентний підхід до побудови інтелектуальних систем, що передбачає визначення чотирьох головних компонент: критеріїв ефективності (perfomance), середовища функціонування (environment), засобів впливу на середовище (actuators) та засобів моніторингу стану середовища (sensors) (разом PEAS). Такий підхід ґрунтується на побудові та навчанні онтології у складі бази знань інтелектуальної системи, яка містить понятійний апарат заданої предметної області, а також сформульовану в термінах онтології ієрархічну мережу задач (цілей), що стоять перед інтелектуальною системою. Для цього буде використано недавно розроблену і затверджену консорціумом W3C мову подання знань OWL2, що порівняно з попередньою версією OWL1 має суттєво більші виразні засоби описової логіки (DL). Це дасть змогу реалізувати методи і засоби автоматичного планування діяльності агентів для конкретних предметних областей, а також приймати оптимальні рішення інтелектуальним агентом, який функціонує в межах цієї предметної області.

Запропонований підхід забезпечує можливість розроблення принципово нових методів опрацювання інформаційних ресурсів у тих предметних областях, для яких інтелектуальна система має належним чином побудовану і навчену онтологію. Зокрема, на відміну від існуючих методів, які основані на статистичному аналізі частоти використання термінів і словосполучень у текстах, пропонований підхід ґрунтується на розпізнаванні змісту речень текстового документа побудовою в процесі їх опрацювання предикатів мовою описової логіки та зіставлення розпізнаних таким чином предикатів з контекстною інформацією, що міститься в онтології, яка перед тим пройшла

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

попереднє навчання. Для опрацювання сховищ та баз даних пропонується зважувати їх елементи оцінками онтологічних знань.

Розроблені методи та засоби буде використано для побудови сучасних прикладних інтелектуальних систем опрацювання інформаційних ресурсів, ядром баз знань яких є онтології, в таких предметних областях, як управління та супровід інформаційних ресурсів у мережі Інтернет, під час побудови систем on-line та off-line реалізації контенту, cloud storage, cloud computing, у військовій та правовій сферах. Розширення галузей застосування спроектованих та розроблених інтелектуальних систем опрацювання інформаційних ресурсів, що ґрунтуються на онтологіях, сприятиме подоланню соціальних, організаційних, методичних та психологічних бар'єрів, що виникають як реакція на введення інноваційних підходів до ведення діяльності у відповідних галузях господарства України та нормативних сферах функціонування.

Висновки і перспективи подальших наукових розвідок

Розроблено математичне забезпечення опрацювання сховищ та баз даних, природомовних текстових документів в тих предметних областях, які чітко формалізуються за допомогою онтологічних моделей. На основі математичного забезпечення функціонування інтелектуальних систем опрацювання інформаційних ресурсів розроблено методи та засоби побудови інтелектуальних систем опрацювання інформаційних ресурсів для прийняття оптимальних рішень у межах певної предметної області. Такі системи доволі поширені, зокрема, для формування, управління та супроводу зростаючих обсягів контенту в Інтернеті під час побудови систем on-line та off-line реалізації контенту, cloud storage, cloud computing.

Розроблено математичне забезпечення та методи опрацювання інформаційних ресурсів на основі онтологічних знань. Отримані наукові результати є цінними з погляду побудови сучасних інтелектуальних систем прийняття оптимальних рішень у різних предметних областях. Зокрема цінними є методи навчання та оптимізації онтологій у таких предметних областях. Фахівці, що працюватимуть над цим фундаментальним дослідженням, набудуть додаткових професійних навичок розроблення математичних моделей, методів та алгоритмів навчання, оптимізації та адаптації онтології до специфіки предметних областей. Пропонована тематика є сучасною й активно розвивається у всьому науковому світі.

1. Висоцька В. А. Методи і засоби опрацювання інформаційних ресурсів в системах електронної контент-комерції / В. А. Висоцька: дис. … канд. техн. наук за спеціальністю 05.13.06 – інформаційні технології. На правах рукопису. – Національний університет "Львівська політехніка" Міністерства освіти і науки України, Львів, 2014. 2. Ландэ Д. В. Подход к созданию терминологических онтологий / Д. В. Ландэ, А. А. Снарский // Онтология проектирования. – 2014. – N 2(12). – С. 83–91. 3. Леонтьева Н. Н. К теории автоматического понимания естественных текстов. Ч. 2: Семантические словари: состав, структура, методика создания / Н. Н. Леонтьева. – М.: Изд-во МГУ, 2001. 4. Литвин В. В. Методи та засоби побудови інтелектуальних систем підтримки прийняття рішень на основі адаптивних онтологій / В. В. Литвин: автореф. дис. … д-ра техн. наук за спеціальністю 01.05.03 – математичне і програмне забезпечення обчислювальних машин і систем. На правах рукопису. – Національний університет "Львівська політехніка", Львів, 2012. 5. Литвин В.В. Бази знань інтелектуальних систем підтримки прийняття рішень: монографія / В. В. Литвин. – Львів: Видавництво Львівської політехніки, 2011. – 240 с. 6. Литвин В. В. Бази знань інтелектуальних систем підтримки прийняття рішень: монографія / В. В. Литвин. – Львів : Видавництво Львівської політехніки, 2011. – 240 с. 7. Литвин В. В. Процес підтримки прийняття управлінських рішень у системі раннього попередження / В.В. Литвин, О. І. Цмоць // Актуальні проблеми економіки. – № 11(149). – 2013. – SNIP 0,085.– С. 222–229. – (SCOPUS). – Режим доступу: http://irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/ape_2013_11_32.pdf. 8. Литвин В. В. Підхід до побудови інтелектуального агента визначення групи банківського ризику на основі онтології / В. В. Литвин //

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

Актуальні проблеми економіки : наук. економіч. журн. – Київ, 2011. – №7(121).– (SCOPUS).– С. 314–320. – http://web.a.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authtype=craw ler&jrnl=19936788&AN=64499457&h=EF90VVTkovPqw%2fqVW01ABrqOE%2bXguBTMOkxR6JTowN2POrmrZNHUR9zCN3zrSlruxic8O8YdYTEIuMvZdegzvw%3d%3d&crl=f&resultNs=AdminWebAuth&resultLocal=ErrCrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authtype%3dcrawler%26jrnl%3d19936788 %26AN%3d64499457. 9. Круглов В. В. Искусственные нейронные сети. Теория и практика / В. В. Круглов, В. В. Борисов. – М.: Горячая линия – Телеком, 2001. 10. Рашкевич Ю. М. "Болонський процес та нова парадигма вищої освіти". – Львів: Видавництво Львівської політехніки, 2014 – 166 с. 11. Benz Dominik. Capturing emergent semantics from social tagging systems / Dominik Benz, Andreas Hotho // In Jens Lehmann and Johanna Völker, editors, Perspectives on Ontology Learning, Studies on the Semantic Web. – AKA Heidelberg. – IOS Press. – 2014. 12. Blomqvist Eva. Ontology design patterns in ontology learning / Eva Blomqvist, Aldo Gangemi, Francesco Draicchio // In Jens Lehmann and Johanna Völker, editors, Perspectives on Ontology Learning, Studies on the Semantic Web. – AKA Heidelberg. – IOS Press. – 2014. 13. Bühmann Lorenz. Universal OWL axiom enrichment for large knowledge bases / Lorenz Bühmann, Jens Lehmann // In Proceedings of EKAW 2012. – Springer, 2012. – P. 57–71. 14. Elena Simperl. Ontocom: A reliable cost estimation method for ontology development projects / Simperl Elena, Buerger Tobias, Hangl Simon, Woelger Stephan, Popov Igor // Web Semantics: Science, Services and Agents on the World Wide Web. – 16(0):1. – 16, 2012. 15. Kordjamshidi P. Learning to interpret spatial natural language in terms of qualitative spatial relations / P. Kordjamshidi, J. Hois, M. van Otterlo, M. F. Moens // In: T. Tenbrink, J. Wiener, C. Claramunt (Eds.), Representing Space in Cognition: Interrelations of Behavior, Language, and Formal Models. Series Explorations in Language and Space. – Oxford University Press, 2013, – P. 115–146. 16. Wong W. Ontology learning from text: a look back and into the future / W. Wong, W. Liu, M. Bennamoun // ACM Computing Surveys. – 44 (4). – 2012. – 20:1–20:36. 17. Mani I. Interpreting Motion: Grounded Representations for Spatial Language / I. Mani, J. Pustejovsky // Explorations in language and space. – Oxford University Press. – 2012. 18. Saleh M. E. Semantic-Based Query in Relational Database Using Ontology / M. E. Saleh // Canadian Journal on Data, Information and Knowledge Engineering. – 2011. – Vol. 2. – № 1. – P.1–16. 19. Lehmann Jens. Class expression learning for ontology engineering / Jens Lehmann, Sören Auer, Lorenz Bühmann, and Sebastian Tramp // Journal of Web Semantics, 9:71 – 81, 2011. 20. Chyrun L. Electronic content commerce system development / Lyubomyr Chyrun, Victoria Vysotska, Vasyl Andrunyk // MEST Journal. – Vol.4 No.2. – 2015. – PP. 120-138 [Online]. – ISSN 2334-7058 (Online). – http://www.mest.meste.org/MEST_Najava_clanaka.html http://mest.meste.org/MEST_Najava/VI_Chyrun.pdf. 21. Lytvyn V. V. The similarity metric of scientific papers summaries on the basis of adaptive ontologies / V. V. Lytvyn // Proceedings of VIIth International Conference on Perspective Technologies and Methods in MEMS Design, Polyana, Ukraine, 11-14 May 2011. – Lviv : IEEE; LPNU, 2011. – P. 162. – (SCOPUS). http://ieeexplore.ieee.org/ xpl/login.jsp?tp=&arnumber=5960319&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5 %2F5954677 %2F5960241 %2F05960319.pdf%3Farnumber%3D5960319. 22. Lytvyn V. Design of intelligent decision support systems using ontological approach / V. Lytvyn // An international quarterly journal on economics in technology, new technologies and modelling processes. – ISSN 2084-5715. – Lublin-Lviv. – 2013. – Vol. II. – No 1. – 31-38. – http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.baztech-d7349648-4893-4610-806b-023b87367d5f. 23. Lytvyn V. Definition of the semantic metrics on the basis of thesaurus of subject area / V. Lytvyn, O. Semotuyk, O. Moroz // An international quarterly journal on economics in technology, new technologies and modelling processes. – ISSN 2084-5715.– Lublin-Lviv. – 2013. – Vol. II. – No 4. – P. 47–51. – http://ena.lp.edu.ua:8080/bitstream/ntb/23678/1/7-47-51.pdf. 24. Lytvyn V. An ontology based intelligent diagnostic systems of steel corrosion protection / V. Lytvyn, D. Dosyn, A. Smolarz. – Elektronika. – ISSN 0033-2089.– Poland. – N 8. – 2013. – P. 22–24. – http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-78298beb-92d6-4e2f-84ad-c4b2fb4db8bc. 25. Shanjian L. A composite approach to language [Електронний ресурс] / L. Shanjian, M. Katsuhiko // encoding detection. – Режим доступу: http://www. mozilla.org/projects/intl/ UniversalCharset Detection.html. 26. Sörensen T. A method of establishing groups of equal amplitude in plant sociology

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

based on similarity of species content // Kongelige Danske Videnskabernes Selskab. Biol. krifter. Bd V. № 4. 1948. P. 1-34. 27. Yang Y. An Evaluation of Statistical Approaches to Text Categorization / Y. Yang // Journal of Information Retrieval. – 1999. – № 1. 28. Vysotska V. Linguistic analysis and modelling semantics of textual content for digest formation / Victoria Vysotska, Lyubomyr Chyrun // MEST Journal (Management Education Science & Society Technologie). – Vol.3 No.1. – PP. 127-148 [Online]. – ISSN 2334-7171, ISSN 2334-7058 (Online), DOI 10.12709/issn.2334-7058. This issue: DOI 10.12709/mest.02.02.02.0. – Режим доступу: http://mest.meste.org/MEST_1_2015/Sadrzaj_eng.html http://mest.meste.org/MEST_1_2015/5_15.pdf. 29. Vysotska V. Features of the content-analysis method for text categorization of commercial content in processing online newspaper articles / Victoria Vysotska, Lyubomyr Chyrun // Applied Computer Science. ACS journal. – Volume 11, Number 1. – Poland, 2015. – ISSN 2353-6977 (Online), ISSN 1895-3735 (Print) – PP. 5-19 [Online]. –www.acs.pollub.pl, http://www.acs.pollub.pl/index.php/-current-issue/applied-computer-science-volume-11-number-1-2015.html, http://www.acs.pollub.pl/pdf/v11n1/2.pdf. 30. Vysotska V. The Means Structure of Information Resources Processing in Electronic Content Commerce Systems / Victoria Vysotska, Lyubomyr Chyrun // Journal of Information Sciences and Computing Technologies (JISCT). – Vol 3, No 3 (2015). – Punjab, India, 2015. – ISSN: 2394-9066. – P. 241–248. – http://scitecresearch.com/journals/index.php/jisct/issue/view/24, http://scitecresearch.com/ journals/index.php/jisct/article/view/134, http://scitecresearch.com/journals/ index.php/ jisct/ article/view/ 134/119. 31. Vysotska V. Designing features of architecture for e-commerce systems / Victoria Vysotska, Lyubomyr Chyrun // MEST Journal (Management Education Science & Society Technologie). – Vol. 2 No.1. – PP. 57-70 [Online]. – ISSN 2334-7171, ISSN 2334-7058 (Online), DOI 10.12709/issn.2334-7058. This issue: DOI 10.12709/mest.02.02.02.0. – Режим доступу: http://mest.meste.org/R3.html, http://www.meste.org/mest/Archive/MEST_II_2_1.pdf, http://mest.meste.org/ MEST_1_2014/_06.pdf, http://mest.meste.org/MEST_1_2014/K1_eng.html. 32. Vysotska V. Comprehensive method of commercial content support in the electronic business systems / Victoria Vysotska, Lyubomyr Chyrun, Liliya Chyrun // Комп'ютерні системи проектування. Теорія і практика, Вісник Нац. ун-ту "Львівська політехніка". – ISSN 0321-0499. – № 777. – Львів, 2013. – С. 21–30. – http://ena.lp.edu.ua:8080/bitstream/ntb/25752/1/6-21-30.pdf. 33. Vysotska Victoria. Web Content Processing Method for Electronic Business Systems / Victoria Vysotska, Lyubomyr Chyrun // International Journal of Computers & Technology. – Vol 12, No 2. – December 2013. – P. 3211–3220. – ISSN 2277-3061. – [Online] http://cirworld.org/journals/index.php/ijct/article/view/3299. Impacr factor 1,532. (Index Copernicus, NASA ADS, DOAJ, Google Scholar, Eyesource, EBSCO, CiteSeer, UlrichWeb, ScientificCommons, ProQuest CSA Technology Research Database).

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua