

С. І. Гайко,
В. В. Приходнюк

ПІДХІД ДО АВТОМАТИЗОВАНОЇ СТРУКТУРИЗАЦІЇ ОСВІТНІХ РЕСУРСІВ НА ОСНОВІ МЕТОДУ РЕКУРСИВНОЇ РЕДУКЦІЇ

Анотація. Статтю присвячено питанням автоматизованої структуризації освітніх інформаційних ресурсів, представлених природномовними текстами, для відображення їх користувачам у зручній (зокрема, інтерактивній) формі. Розглянуто проблеми обробки, представлення, використання та інтеграції великих обсягів міждисциплінарних знань. Проведено аналіз найбільш поширених програмних засобів, що призначені для обробки природномовних текстів і можуть бути використані для вирішення поставленої задачі. Запропоновано спосіб оцінки таких програмних засобів на основі багатокритеріальної оптимізації. Визначено і обґрунтовано критерії такої оцінки, а також наведено її результати. На основі отриманих результатів показано, що найбільш ефективним засобом для вирішення задачі є система «Рекурсивний редуктор», що дає можливість аналізувати широкий спектр документів і представляти результати в різній формі, включаючи XML-файли онтологій, які можуть бути використані в роботі когнітивної ІТ-технології (КІТ) «Поліедр». У статті надано короткий опис математичної моделі, на основі якої працює «Рекурсивний редуктор». Зазначена модель базується на використанні λ-виразів. λ-вирази застосовуються для формування спеціалізованих рекурсивних функцій, що виконують ідентифікацію лексем у вхідному тексті і формують на їх основі об'єкти і їх атрибути. Додатково описано застосування онтологічного класифікатора документів для підвищення ефективності роботи рекурсивного редуктора. Також наведено приклад його роботи — автоматизовану структуризацію освітнього інформаційного ресурсу, представленого навчальною програмою з хімії (поглиблене навчання), відображення її у вигляді онтології, а в подальшому — створення на її основі трансдисциплінарного інтерактивного документа засобами КІТ «Поліедр».

Ключові слова: освітні ресурси, обробка природномовних текстів, онтологічне представлення текстів, рекурсивний редуктор, інтерактивний документ.

Постановка проблеми. Однією з базових умов прогресивного розвитку суспільства є науково-освітній розвиток. Ця сфера людської діяльності характеризується високим рівнем інформатизації, а саме накопиченням великих обсягів даних, виникненням нових типів цифрових відносин між науковою спільнотою і суспільством. При цьому, як свідчить дослідження компанії IDC «Big Data, Bigger Digital Shadows,

and Biggest Growth in the Far East», у світі використовується лише мізерна частка інформації при продовженні зростання її обсягів [1]. Отже, значна частина цифрового нарративу досі лишається пасивною, що викликає постійний інтерес фахівців до розвитку інформаційних технологій, спрямованих на структуризацію та інтеграцію інформаційних ресурсів.

Бажаючи надати учасникам освітнього процесу найактуальнішу інформацію, заклади освіти постійно змінюють і вдосконалюють навчальні

програми, і попри це підготовка майбутніх фахівців та випускників шкіл засвідчує, що ці методики застаріли на кілька років. Тому процес навчання дедалі більше перетворюється на активний обмін знаннями у вигляді семінарів, дискусій, конференцій.

Для нинішнього століття характерні нові засоби та сфери спілкування і взаємодії людей. На якісно інший рівень виходять наші інтелектуальні та практичні можливості. Це пов'язано із застосуванням нових механізмів одержання будь-яких даних, інформації, знань, із мобільністю людини, відкритістю кордонів тощо [2].

У цих умовах на перший план виходять вивчення і вдосконалення процесів створення засобів та методик формування інтелектуальних інформаційно-освітніх середовищ; забезпечення доступу учнів до науково-інформаційних баз провідних наукових центрів і університетів; підтримка колективної взаємодії учні — викладачі — науковці; проектування та побудова представлення предметних знань на основі обробки природномовної (ПМ) текстової інформації.

Зазначимо, що на сьогодні спостерігається розрив між достатньо розвинутими засобами граматичної обробки природномовної інформації для розв'язання вузькоспеціалізованих задач і недостатністю таких засобів для розв'язання комплексних задач, пов'язаних із функціональним аналізом і розумінням ПМ-інформації, її формально-логічним представленням, добуванням предметних знань з їх подальшим використанням під час проведення експериментальних досліджень [2].

Саме тому найбільш перспективними напрямками створення та використання систем обробки та представлення інформації вважаються ті, що здатні опрацьовувати розподілені, політематичні, великі масиви слабо- або неструктурованих даних (зокрема, природномовних документів) і тим самим надавати технологічну підтримку науково-освітній галузі в ефективному та конструктивному застосуванні накопичених людством знань [1; 3; 4].

Аналіз останніх досліджень і публікацій. Дослідження з питань обробки природної мови проводяться вже близько 70 років. Розпочавшись наприкінці 40-х років ХХ ст., вони пройшли чотири стадії розвитку, що характеризуються, відповідно, акцентуванням уваги на створенні систем машинного перекладу, домінуванням

теорій штучного інтелекту, розробкою обчислювальних граматик і логічного програмування і застосуванням статистичного підходу [5].

На сучасному етапі розвитку науки і техніки в напрямі комп'ютерної обробки природномовних текстів виокремлюють два основні підходи: лінгвоаналітичний і статистичний [6]. Найбільш перспективними й ефективними з них визнано, відповідно, експліцитні методи семантичного аналізу текстової інформації (алгоритми онтологічного семантичного аналізу) та методи латентно-семантичного аналізу. Саме ці методи дають змогу визначити та побудувати смислову структуру природномовного тексту у формалізованому вигляді (наприклад, у вигляді фрейму, семантичної мережі, онтологічного представлення тексту тощо). Цим питанням присвячені роботи Глушкова В. М., Гладуна В. П., Палагіна О. В., Широкова В. А., Хорошевського В. Ф., Поспелова Д. А., Овдій О. М., Гаврилової Т. А., Загорюлька Ю. А., Малишевського А. В., Андона П. І., Валькмана Ю. Р., Соловйової К. О., Стрижка О. Є., Яловця А. Л., Gruber T., Guarino N., Noy N., Corcho O., Gomez-Perez A., Gruninger M., Fernandez Lopez M., Happel H., McGuinness D., Lee J.

Мета статті. Визначити найбільш ефективний спосіб аналізу освітніх ресурсів шляхом аналізу наявних у відкритому доступі програмних засобів і побудови їх рейтингу.

Виклад основного матеріалу.

І. Аналіз систем обробки природної мови. Найбільш ефективним способом розв'язання задачі структуризації освітніх ресурсів є застосування сучасних програмних засобів. Однак при цьому виникає додаткове завдання — вибору програмного засобу, що буде застосовуватись. Для цього пропонується використання багатокритеріальної оптимізації [7].

Для вибору оптимальної системи (або систем), що можуть використовуватись у межах реалізації поставленої задачі, необхідно:

- 1) відібрати системи для аналізу;
- 2) сформулювати критерії якості систем;
- 3) провести оцінку обраних систем за критеріями;
- 4) визначити важливість кожного із критеріїв для реалізації поставленої задачі;
- 5) виконати ранжування систем і вибрати систему з найвищою оцінкою.

Для аналізу обрано 26 систем [8], серед яких представлено системи для обробки природномовних текстів українською, російською,

англійською та іншими мовами; охоплено системи різних рівнів лінгвістичного аналізу; доступні безкоштовно і комерційні, а саме: «Рекурсивний редуктор»; General Architecture for Text Engineering; Affx Grammars Over a Finite Lattice (AGFL); RapidMiner; LexicoSyntactic Pattern Language (LSPL); «Томіта-парсер»; «Конспект (TextTermin)»; «ИСИДА-Т»; Link Grammar Parser; Link Grammar Parser for Russian; RCO (Russian Context Optimizer); libmorphukr; Mystem; Cíbola/Oleada; pymorphy2; Russian Morphological Dictionary; Stemka; MonoConc; TextAnalyst 2.0; AskNet; Vivissimo; Quintura Searchcrystal; «МедіаЛінгва»; ABBYY Retrieval & Morphology (ARM) Engine; Apache Lucene/Solr; Convera RetrievalWare.

Критеріями оцінки систем вибрано наявність чи відсутність у них певних функцій, важливих при обробці освітніх ресурсів. Зокрема, такими функціями є:

- обробка текстів, написаних певною мовою, — в межах процесу обробки освітніх ресурсів найважливішою є підтримка української мови, другорядне значення має підтримка російської та англійської;
- обробка файлів заданого формату — найважливішою є підтримка doc/docx і pdf, другорядне значення має підтримка xls, xlsx, txt;
- формування вихідних файлів у заданому форматі — найважливіше значення мають форми HTML (для відображення користувачу), XML (для інтеграції з КІТ «Полюдр»), другорядне значення мають PDF, JSON, TXT;
- доступні види аналізу (семантичний, синтаксичний, морфологічний, аналіз розмітки, аналіз стилю, аналіз метаданих) визначають

спектр інформації, що може бути виокремлена з інформаційних ресурсів.

Також системи можуть оцінюватись за більш спеціалізованими критеріями, пов'язаними з особливостями представлення результату. Такими критеріями є:

- можливість виокремлення з текстів термінів, зв'язків, атрибутів, контекстів, зображень — визначає, наскільки повно буде представлена наявна в освітньому ресурсі інформація в результатах структуризації;
- спосіб представлення опрацьованого тексту (у вигляді тексту, вебдокумента, онтології, інтерактивного документа) — визначає, чи потрібно буде застосовувати допоміжне програмне забезпечення для відображення результатів структуризації освітніх ресурсів.

Значення параметрів пропонується оцінювати за шкалою від нуля до десяти, де нуль — найгірше значення, десять — найкраще.

На рис. 1 наведено фрагмент експертної оцінки обраних систем.

Важливість критеріїв пропонується оцінювати за лінгвістичною шкалою [7]. Вагомість критеріїв представлена в табл. 1.

Результати ранжування представлено на рис. 2.

Як свідчать результати оцінювання, система «Рекурсивний редуктор» має значно кращі результати, ніж інші розглянуті програмні засоби. Це пов'язано з тим, що вона об'єднує підтримку релевантних мов, широкі можливості аналізу текстів (включно з морфологічним і синтаксичним аналізом) і засоби для представлення результатів аналізу в різних формах.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1			Підтримка мов			Вхідні формати файлів			Вихідні формати файлів					Доступні види аналізу					
2	nodeproperties	Клас системи	Укр.	Рос.	Англ.	doc, docx	xls, xlsx	txt	pdf	XML	HTML	PDF	JSON	TXT	семан-тичний	синтаксичний	морфологічний	аналіз розмітки	аналіз стилю
3	Рекурсивний редуктор	Лінгвістичний фреймворк	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
4	General Architecture for Text Engineering	Лінгвістичний фреймворк			10			10						10	10	10	10	10	10
5	Affx Grammars Over a Finite Lattice (AGFL)	Лінгвістичний фреймворк			10			10						10	10	10	10	10	10
6	RapidMiner	Лінгвістичний фреймворк			10			10						10	10	10	10	10	10
7	LexicoSyntactic Pattern Language (LSPL)	Лінгвістичний фреймворк			10			10						10	10	10	10	10	10
8	Томіта-парсер	Лінгвістичний фреймворк		10	10			10						10	10	10	10	10	10
9	Конспект (TextTermin)	Синтаксичний аналізатор		10	10			10						10	2	10	10	10	10
10	ИСИДА-Т	Синтаксичний аналізатор			10			10						10		10	10	10	10
11	Link Grammar Parser	Синтаксичний аналізатор			10			10						10		10	10	10	10
12	Link Grammar Parser for Russian	Синтаксичний аналізатор			10			10						10		10	10	10	10
13	RCO (Russian Context Optimizer)	Синтаксичний аналізатор			10			10						10		10	10	10	10
14	libmorphukr	Морфологічний аналізатор		10				10						10		10	10	10	10
15	Mystem	Морфологічний аналізатор		10	10			10						10		10	10	10	10
16	Cíbola/Oleada	Морфологічний аналізатор			10			10						10		10	10	10	10
17	pymorphy2	Морфологічний аналізатор			10			10						10		10	10	10	10
18	Russian Morphological Dictionary	Морфологічний аналізатор			10			10						10		10	10	10	10
19	Stemka	Морфологічний аналізатор			10			10						10		10	10	10	10
20	MonoConc	Пошукова машина			10			10						10		10	10	10	10
21	TextAnalyst 2.0	Пошукова машина			10			10						10		10	10	10	10

Рис. 1. Оцінка систем обробки текстів за обраними критеріями

Таблиця 1

Важливість	Критерії
Мала важливість	Клас системи
Слабка важливість	Вхідні формати файлів xls, xlsx, txt Вихідні формати файлів PDF, JSON, TXT
Середня важливість	Підтримка російської мови Підтримка англійської мови Доступні види аналізу
Значна важливість	Вхідні формати файлів doc/docx і pdf Вихідні формати файлів HTML, XML Виокремлення з текстів термінів, зв'язків, атрибутів, контекстів, зображень
Абсолютна важливість	Підтримка української мови Представлення опрацьованого тексту у вигляді тексту, вебдокумента, онтології, інтерактивного документа

II. Математична модель методу рекурсивної редукції. Робота рекурсивного редуктора базується на використанні методу рекурсивної редукції, призначеного для формування онтологій виду [9–13].

$$O = \langle X, R, F \rangle, \quad (1)$$

де X — множина концептів предметної області,
 R — множина відношень між концептами,
 F — множина функцій інтерпретації X та R .

Структуризацію певного природномовного тексту T^T можна представити як певне перетворення — перетворення структуризації.

$$F_{str} : T^T \rightarrow O \quad (2)$$

Структуризація методом рекурсивної редукції полягає в рекурсивному застосуванні оператора редукції. Оператор редукції записується в термінах λ -виразів [11–14] і, в свою чергу, є комбінацією трьох інших операторів, що застосовуються послідовно.

$$F_{rd} = F_x \circ F_{smr} \circ F_{ct}, \quad (3)$$

де F_x — оператор ідентифікації об'єктів X ,

F_{smr} — оператор ідентифікації зв'язків R ,

F_{ct} — оператор ідентифікації контекстів, через які визначаються функції інтерпретації в кінцевій онтології.

1 Рекурсивний редуктор	0,922	14 Convera RetrievalWare	0,345
2 Apache Lucene/Solr	0,415	15 Link Grammar Parser	0,323
3 AskNet	0,415	16 Link Grammar Parser for Russian	0,323
4 Конспект (TextTermin)	0,385	17 ИСИДА-Т	0,323
5 Томіта-парсер	0,376	18 RCO (Russian Context Optimizer)	0,297
6 Affix Grammars Over a Finite Lattice (AGFL)	0,376	19 MonoConc	0,259
7 General Architecture for Text Engineering	0,376	20 TextAnalyst 2.0	0,259
8 LexicoSyntactic Pattern Language (LSPL)	0,376	21 Mystem	0,234
9 RapidMiner	0,376	22 Cibola/Oleada	0,209
10 Quintura Searchcrystal	0,361	23 libmorphukr	0,209
11 Vivisimo	0,361	24 pymorphy2	0,209
12 ABBYY Retrieval & Morphology (ARM) Engine	0,361	25 Russian Morphological Dictionary	0,209
13 Медіалінгва	0,361	26 Stemka	0,209

Рис. 2. Результати ранжування обраних систем

Кожен з елементів оператора редукції, у свою чергу, формується правилами виду.

$$g = \langle f_{ap}^g, f_{tr}^g \rangle, \quad (4)$$

де f_{ap}^g — функція застосовності, що визначає, чи може правило бути застосоване до певного набору вхідної інформації, f_{tr}^g — функція перетворення, що задає перетворення вхідної інформації.

Задане правилом g перетворення має вигляд.

$$F_g(x) = \begin{cases} f_{tr}^g(x), f_{ap}^g(x) \\ x, \neg f_{ap}^g(x) \end{cases} \quad (5)$$

Для використання методу рекурсивної редукції необхідно створювати функції застосовності й перетворення, що відповідають вхідному масиву документів і потрібному результату.

Для підвищення ефективності роботи «Рекурсивного редуктора» при обробці текстів різних стилів (у яких значною мірою різняться способи представлення інформації) у процесі обробки можуть використовуватися додаткові правила, що відображають особливості відповідного стилю. Зазначені правила формуються на основі об'єктів онтологічного класифікатора стилів документів. Об'єкти, у свою чергу, мають атрибути, які й описують особливості оформлення тексту конкретних документів, таких як програма, інструкція, стандарт тощо.

Отже, функція застосовності, сформована на основі додаткового правила, буде задаватися множиною предикатів ідентифікації, сформованих на основі атрибутів певного об'єкта (8)

$$p_i(l, x) = \begin{cases} 1, a_i \in P_l \\ 0, a_i \notin P_l \end{cases}, \quad (8)$$

де $a_i \in A_x$ — певний атрибут об'єкта x , P_l — множина ознак лексеми l .

Отже, можна задати особливості текстової розмітки типових документів, при цьому основною вимогою є сумісність множин A_x і P_l , що даватиме можливість ефективно використовувати операції порівняння.

III. Використання рекурсивного редуктора для обробки освітніх ресурсів. Розглянемо роботу методу рекурсивної редукції на прикладі освітнього ресурсу — програми з хімії (10–11 класи, поглиблене навчання) [13].

Програма містить:

- 1) титульну сторінку з назвою;
- 2) пояснювальну записку, що включає фрагмент тексту й таблицю;
- 3) набір підрозділів, що стосуються різних курсів.

Автоматизоване виділення вершин було реалізовано на основі порівняння типового зразка навчальної програми, формальний опис якого міститься в класифікаторі стилів, і документа, що оброблявся. Так, класифікатор містить об'єкт «Навчальна програма» (рис. 3), якому належать такі атрибути зі значеннями:

- заголовок документа: шрифт — Times New Roman, кегль — 18, регістр — верхній, вирівнювання — по центру;
- заголовок першого рівня: шрифт — Times New Roman, кегль — 12, регістр — верхній, вирівнювання — по центру; накреслення — напівжирне;
- заголовок другого рівня: шрифт — Times New Roman, кегль — 12, регістр — нижній, вирівнювання — по центру; накреслення — звичайне і т. д.

Елементи тексту (у вхідному файлі), параметри розмітки яких збігаються із шаблонними, було автоматично віднесено до визначеної ієрархії. У процесі роботи методу рекурсивної редукції була сформована онтологія у форматі KIT «Поліедр».

KIT «Поліедр» призначена для підтримки процесів лінгвістично-семантичного аналізу великих обсягів просторово розподіленої неструктурованої інформації (Big Data), їх структуризації, визначення контекстних зв'язків між документами, що обробляються, прогнозування та підтримки процесів раціонального вибору



Рис. 3. Фрагмент онтологічного класифікатора стилів

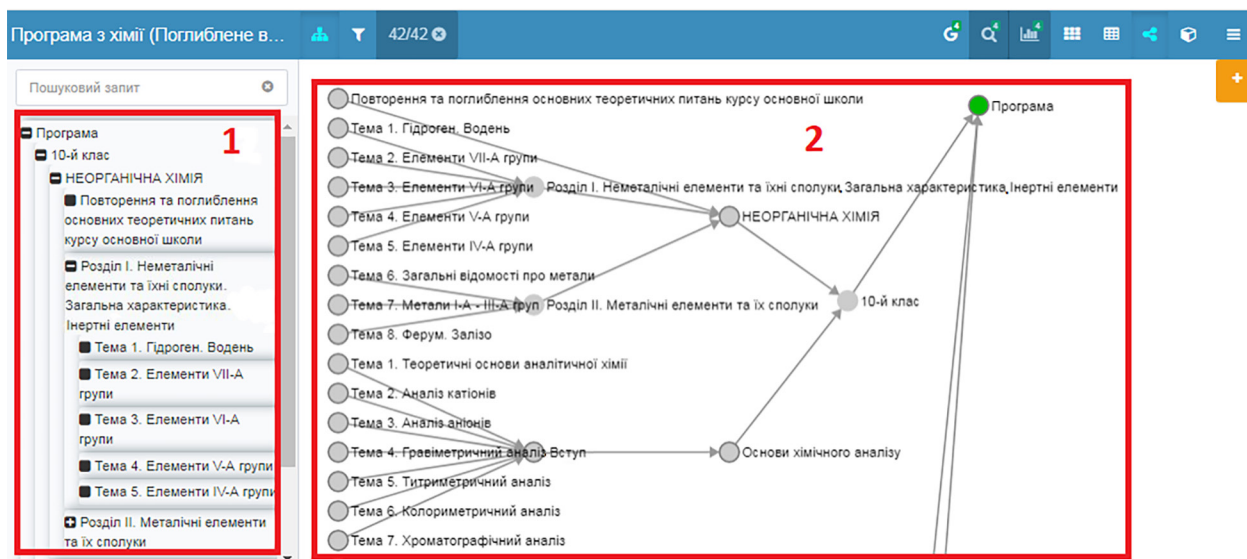


Рис. 4. Програма з хімії у вигляді таксономії (1), графа (2)

з наступним формуванням інформаційно-аналітичних WEB-орієнтованих рішень.

Онтологія програми з хімії візуалізується в середовищі КІТ «Поліедр» за допомогою таксономії або орієнтованого графа (рис. 4).

Представлення онтології у вигляді графа дає можливість одночасно відображати велику кількість об'єктів і зв'язків між ними. Це забезпечує значне підвищення ефективності роботи експерта з інформаційним ресурсом.

Порівняння структурованого (інтерактивно-го) і неструктурованого представлень інформаційного ресурсу продемонстровано на рис. 5. На цьому рисунку показано пояснювальну записку, що складається з текстової частини і таблиці критеріїв.

Так само кожний розділ представлений таблицею, що містить назви тем у межах курсу, кількість годин, зміст навчального матеріалу, вимоги до рівня підготовки учня (рис. 6).

Критерії оцінювання навчальних досягнень учнів з хімії

При оцінюванні рівня навчальних досягнень з хімії враховується:

- рівень засвоєння теоретичних знань;
- оволодіння хімічною мовою як засобом відображення знань про речовини і хімічні явища;
- сформованість експериментальних умінь, необхідних для виконання хімічних дослідів, передбачених навчальною програмою;
- здатність учнів застосовувати набуті знання на практиці;
- уміння розв'язувати розрахункові задачі.

За відмінностями між обсягом і глибиною досягнутих результатів, ступенем самостійності у виконанні завдань, здатністю використовувати знання у нових ситуаціях впрокредено рівні навчальних досягнень учнів, що оцінюються за 12-бальною шкалою.

Кожний наступний рівень вбирає в себе вимоги до попереднього, а також додає нові характеристики.

Визначальними в оцінюванні рівня навчальних досягнень учнів є особистісні результати пізнавальної діяльності, в яких відбиваються загальнопредметні компетентності, набуті учнями в процесі навчання хімії.

Рівні навчальних досягнень учнів	Бали	Критерії оцінювання навчальних досягнень учнів
Початковий	1	Учень (учениця) розпізнає деякі хімічні об'єкти (хімічні символи, формули, явища, посуд тощо) і називає їх (на побутовому рівні); знає правила безпеки під час проведення практичних робіт
	2	Учень (учениця) описує деякі хімічні об'єкти за певними ознаками; знає призначення лабораторного обладнання

Програма

Пояснювальна записка

Критерії оцінювання навчальних досягнень учнів

Рівні навчальних досягнень учнів	Бали	Критерії оцінювання навчальних досягнень учнів
Початковий	1	Учень (учениця) розпізнає деякі хімічні об'єкти (хімічні символи, формули, явища, посуд тощо) і називає їх (на побутовому рівні); знає правила безпеки під час проведення практичних робіт
	2	Учень (учениця) описує деякі хімічні об'єкти за певними ознаками; знає призначення лабораторного обладнання
	3	Учень (учениця) має фрагментарні уявлення з предмета вивчення і може відтворити окремі його частини; під керівництвом учителя виконує найпростіші хімічні дослідів
Середній	4	Учень (учениця) знає окремі факти, що стосуються хімічних сполук і явищ; складає прості прилади для проведення дослідів і виконує їх під керівництвом учителя; складає з допомогою вчителя скорочену умову задачі
	5	Учень (учениця) з допомогою вчителя відтворює окремі частини навчального матеріалу, дає визначення основних понять;

Рис. 5. Пояснювальна записка в текстовому та інтерактивному вигляді

<p>Розділ I. Неметалічні елементи та їхні сполуки. Загальна характеристика. Інертні елементи. Тема 1. Гідроген. Водень (7 год) Гідроген. Місце у періодичній системі. Будова атома. Ізотопи. Поширення Гідрогену в природі. Водень. Склад молекули і будова речовини. Добування. Фізичні та хімічні властивості: взаємодія з неметалами, металами, оксидами металів. Застосування водню. Об'ємні відношення газів у хімічних реакціях. Закон об'ємних відношень газів. Розрахунки за законом об'ємних відношень газів.</p>	<p>Учень: <i>називає</i> інертні елементи, ізотопи та сполуки Гідрогену; <i>формулює</i> закон об'ємних відношень газів; <i>складає</i> електронну та графічно-електронну формули атома Гідрогену та інертних елементів; рівняння відповідних хімічних реакцій; схеми електронного балансу окисно-відновних реакцій; <i>характеризує</i> інертні елементи, Гідроген за його місцем у періодичній системі та будовою атома; поширення в природі; фізичні та хімічні властивості водню; способи добування; <i>обґрунтовує</i> місце Гідрогену в періодичній системі; валентність та</p>
--	--

Рис. 6. Фрагмент навчальної програми у вигляді таблиці

У картці відповідного об'єкта всі складові такої таблиці будуть відображені у вигляді атрибутів «Годин, всього», «Зміст навчального матеріалу», «Державні вимоги до рівня загальноосвітньої підготовки учня» із числовим або текстовим значенням (рис. 7).

Додатковою функцією КІТ «Поліедр» є об'єднання інформації, що міститься в різних інформаційних ресурсах. Це здійснюється за допомогою процедури пошуку (рис. 8). Будь-який фрагмент контексту будь-якого об'єкта (наприклад, слово «водень» з картки об'єкта, показаного на рис. 7) може бути використаний як пошуковий запит, після виконання якого об'єкта з різних інформаційних ресурсів (як структурованих — онтологій, так і не структурованих) видаються користувачеві у вигляді онтології (пошукової онтології).

Таким чином реалізується технологія онтологічної інтеграції розподілених інформаційних ресурсів шляхом побудови онтології, об'єктами якої є поняття і процеси (концепти), що належать різним предметним областям.

Зазначимо, що на сьогодні інтеграція інформаційних ресурсів є актуальним завданням для широкого кола спеціалістів, а створення єдиного інформаційного простору — важливою складовою відкритої та дистанційної освіти. Завдяки тому, що до середовища КІТ «Поліедр» може бути залучено потенційно необмежену кількість як навчальних ресурсів, так і учасників навчального процесу, ця технологія може бути базисом для створення єдиного інформаційного простору.

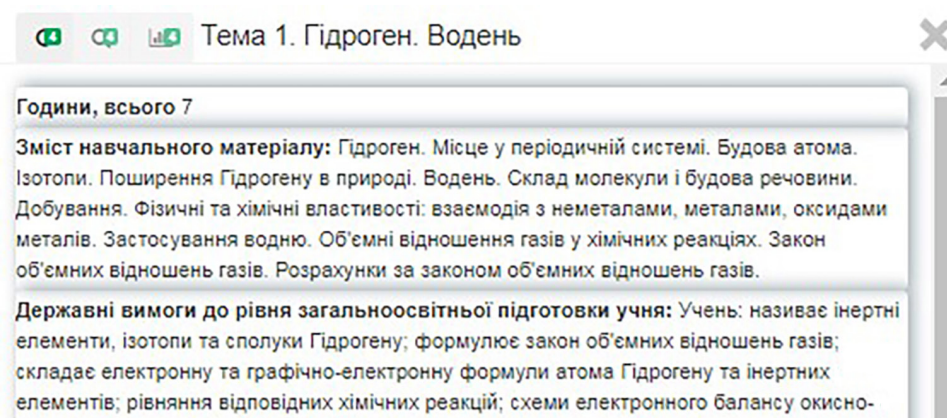


Рис. 7. Картка об'єкта «Тема 1. Гідроген. Водень»

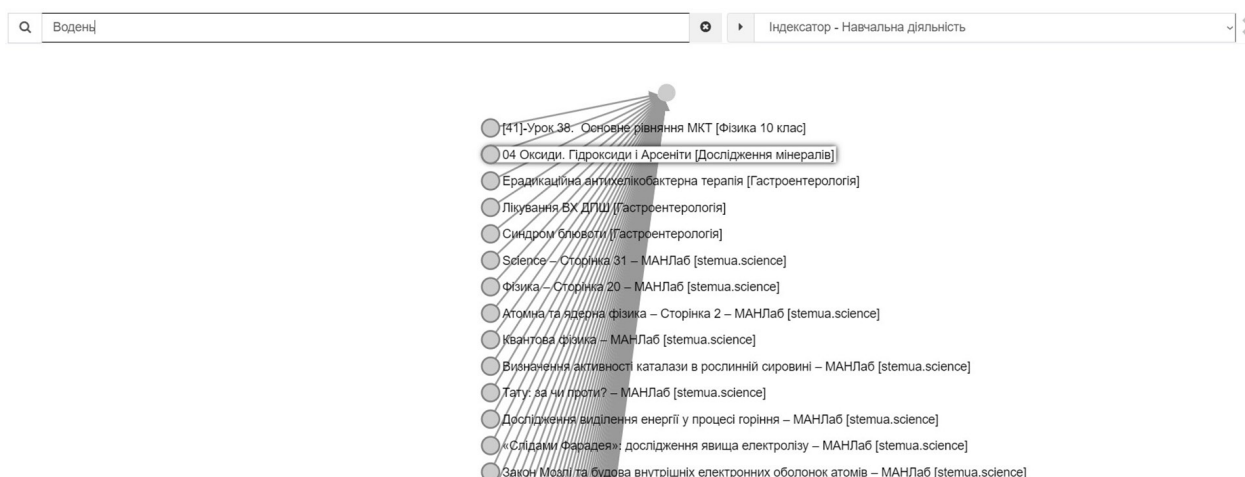


Рис. 8. Об'єднання інформації з різних інформаційних ресурсів за допомогою процедури пошуку

Висновки.

В умовах швидкого розвитку глобальної мережі і неперпинного зростання обсягів інформації найперспективнішим стає розвиток інформаційних технологій, призначених для створення формалізованих систем мультидисциплінарних знань.

Огляд і аналіз наявних систем, призначених для обробки масивів природномовних текстів, виявив, що найбільш ефективною системою для рішення багатоаспектних задач є «Рекурсивний редуктор».

Представлення освітніх ресурсів в онтологічній формі (зокрема, автоматизоване з допомогою «Рекурсивного редуктора») дає можливість збільшити ефективність роботи користувачів з такими ресурсами.

Представлення масивів освітніх ресурсів в онтологічній формі дає змогу створити єдиний інформаційний простір у межах КІТ «Поліедр».

Список використаних джерел

1. Стрижак О. Є. Управління знаннями — головна парадигма сучасної освіти. *Комп'ютер у школі та сім'ї*. 2016. № 5. С. 9–11.
2. Комп'ютерні онтології та їх використання у навчальному процесі. Теорія і практика : монографія / С. О. Довгий та ін. Київ : Інститут обдарованої дитини, 2013. 310 с.
3. Величко В. Ю., Попова М. А., Приходнюк В. В., Стрижак О. Є. ТОДОС — ІТ-платформа формування трансдисциплінарних інформаційних середовищ. *Системи озброєння і військова техніка*. 2017. № 1. С. 10–19.
4. Гладун В. П. Процессы формирования новых знаний. София : СД «Педагог 6», 1994. 189 с.
5. Natural language processing: a historical review. *International encyclopedia of linguistics / Jones K., Bright W. (Ed.). New York : Oxford University Press, 1992. Vol. 3. P. 53–59.*
6. Комарницька О. Методи автоматизованого семантичного аналізу природномовної інформації. *Філологічний дискурс*. 2018. № 7. С. 92–100.
7. Горбурюков В. В., Стрижак О. Є., Франчук О. В. Використання онтологій у системах підтримки прийняття рішень. *Математичне моделювання в економіці : зб. наук. пр. / редкол. : С. О. Довгий (голов. ред.) та ін. Київ : НАН України ; Інститут телекомунікацій і глобального інформаційного простору ; Інститут економіки та прогнозування ; Інститут кібернетики ім. В. М. Глушкова, 2013. С. 33–39.*
8. Гайко С. І., Приходнюк В. В. Засоби трансдисциплінарного представлення освітніх ресурсів. *Актуальні аспекти розвитку STEM-освіти у навчанні природничо-наукових дисциплін*. Кропивницький, 2020. С. 1–64.
9. Палагин А. В., Кривый С. Л., Величко В. Ю., Петренко Н. Г. К анализу естественно-языковых объектов. *«Intelligent Processing» International Book Series «INFORMATION SCIENCE & COMPUTING»: Supplement to the International Journal «INFORMATION TECHNOLOGIES & KNOWLEDGE»*. Sofia, Bulgaria : ITHEA, 2009. Вып. 3. С. 36–43.
10. Приходнюк В. В. Технологічні засоби трансдисциплінарного представлення геопросторової інформації : дис. ... канд. техн. наук : 05.13.06. Київ, 2017. 267 с.
11. Палагин А. В., Петренко Н. Г. Системно-онтологический анализ предметной области. *Управляющие системы и машины*. 2009. Вып. 4. С. 3–14.
12. Барендрегт Х. Лямбда-исчисление. Его синтаксис и семантика. Москва : Мир, 1985. 606 с.

13. Stryzhak O., Prychodniuk V., Podlipaiev V. Model of Transdisciplinary Representation of GEOspatial Information. In: Ilchenko M., Uryvsky L., Globa L. (eds.) *Advances in Information and Communication Technologies*. UKRMICO 2018. *Lecture Notes in Electrical Engineering*. 2019. Vol. 560. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-16770-7_3.
14. Гайко С. І., Приходнюк В. В. Представлення освітніх ресурсів у вигляді інтерактивних документів. *Інтернет — Освіта — Наука*. Вінниця : ВНТУ, 2020. С. 249–251.

References

1. Stryzhak, O. Ye. (2016). Upravlinnia znanniamy — holovna paradyhma suchasnoi osvity [Knowledge management is the main paradigm of modern education]. *Kompiuter u shkoli ta simi — Computer at school and family*, 5, 9–11 [in Ukrainian].
2. Dovhyi, S. O., Velychko, V. Yu., Hloba, L. S., Stryzhak, O. Ye., Andrushchenko, T. I. et al. (2013). *Kompiuterni ontolohii ta yikh vykorystannia u navchalnomu protsesi. Teoriia i praktyka [Computer ontologies and their use in the learning process. Theory and Practice]*. Kyiv : Instytut obdarovanoi dytyny [in Ukrainian].
3. Velychko, V. Yu., Popova, M. A., Prykhodniuk, V. V., & Stryzhak, O. Ye. (2017). TODOS — IT-platforma formuvannia transdystyplinarnykh informatsiynykh seredovyshch [TODOS — IT-platform for the formation of transdisciplinary information environments]. *Systemy ozbroiennia i viiskova tekhnika — Weapons systems and military equipment*, 1, 10–19 [in Ukrainian].
4. Hladun, V. P. (1994). *Protsessy formirovaniya novykh znaniy [Processes of formation of new knowledge]*. Sofiya : SD «Pedagog 6» [in Bulgarian].
5. Jones, K., Bright, W. (Ed.) (1992). *Natural language processing: a historical review. International encyclopedia of linguistics*. New York : Oxford University Press. (Vol. 3), (pp. 53–59).
6. Komarnytska, O. (2018). Metody avtomatyzovanoho semantichnoho analizu pryrodnomovnoi informatsii [Methods of automated semantic analysis of natural-language information]. *Filolohichni dyskurs — Philological discourse*, 7, 92–100 [in Ukrainian].
7. Horborukov, V. V., Stryzhak, O. Ye., & Franchuk, O. V. (2013). Vykorystannia ontolohii u systemakh pidtrymky pryiniattia rishen [Using Ontologies in Decision Support Systems]. S. O. Dovhyi et al (Eds.), *Matematychni modeliuvannia v ekonomitsi — Mathematical Modeling in Economics* : Scientific works collection. (Issue 3), (pp. 33–39). Kyiv : NAN Ukrainy, Instytut telekomunikatsii i hlobalnoho informatsiinoho prostoru, Instytut ekonomiky ta prohnozuvannia, Instytut kibernetiky im. V. M. Hlushkova [in Ukrainian].
8. Haiko, S. I., & Prykhodniuk, V. V. (2020). Zasoby transdystyplinarnoho predstavlenia osvitnikh resursiv [Means of transdisciplinary representation of educational resources]. *Aktualni aspekty rozvytku STEM-osvity u navchanni pryrodnycho-naukovykh dystsyplin — Current aspects of the development of STEM-education in the teaching of natural sciences* (pp. 61–64). Kropyvnytskyi [in Ukrainian].
9. Palagin, A. V., Kryvyy, S. L., Velichko, V. Yu., & Petrenko, N. G. (2009). K analizu estestvenno-jazykovykh obektov [To the analysis of natural language objects]. *“Intelligent Processing” International Book Series “INFORMATION SCIENCE & COMPUTING”: Supplement to the International Journal “INFORMATION TECHNOLOGIES & KNOWLEDGE”*, 3, 36–43 [in Bulgarian].
10. Prykhodniuk, V. V. (2017). Tekhnolohichni zasoby transdystyplinarnoho predstavlenia heoprostorovoi informatsii [Technological means of transdisciplinary presentation of geospatial information]. *Candidate’s thesis*. Kyiv [in Russian].
11. Palagin, A. V., & Petrenko, N. G. (2009). Sistemno-ontologicheskii analiz predmetnoy oblasti [System-ontological analysis of the subject area]. *Upravlyayushchie sistemy i mashiny — Control systems and machines*, 4, 3–14 [in Russian].
12. Barendregt, X. (1985). *Lambda-ischislenie. Ego sintaksis i semantika [Lambda calculus. Its syntax and semantics]*. Moscow : Mir [in Russian].
13. Stryzhak O., Prychodniuk V., Podlipaiev V. (2019) Model of Transdisciplinary Representation of GEOspatial Information. In: Ilchenko M., Uryvsky L., Globa L. (Eds.) *Advances in Information and Communication Technologies*. UKRMICO 2018. *Lecture Notes in Electrical Engineering*, (vol. 560). Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-16770-7_3.
14. Haiko, S. I., & Prykhodniuk, V. V. (2020). Predstavlenia osvitnikh resursiv u vyhliadi interaktyvnykh dokumentiv [Presentation of educational resources in the form of interactive documents]. *Internet — Osvita — Nauka. — Internet — Education — Science*, 249–251 [in Ukrainian].

S. I. Haiko,
V. V. Prykhodniuk

APPROACH TO AUTOMATED STRUCTURING OF EDUCATIONAL RESOURCES BASED ON THE METHOD OF RECURSIVE REDUCTION

Abstract. The article is devoted to the issues of automated structuring of educational information resources, presented in natural language texts, with the purpose of creating interactive documents based on them. The problems of formation, processing, presentation, use and integration of large volumes of interdisciplinary knowledge are considered. Prospects for the development of information technologies designed to create a formalized system of knowledge in specific subject areas are substantiated. The article analyzes the most common analytical systems that are designed to process natural language texts and can be used to solve this problem. The article proposes a way to evaluate software based on multicriteria optimization. The criteria for such an assessment are identified and justified and its results are presented. Based on the obtained results, it is shown that the most effective tool for solving the problem is the system "Recursive reducer", which allows to analyze a wide range of documents and present the results in various forms, including XML ontologies that can be used in the work of cognitive IT technology (CIT) "Polyhedron". The article gives a brief description of the mathematical model on which the "Recursive reducer" based on the use of λ -expressions works. λ -expressions are used to form specialized recursive functions that perform the tasks of identifying tokens in the input text and forming objects. The use of an ontological document classifier to improve the efficiency of a recursive reducer is further described. Article also gives an example of its usage — automated structuring of educational information resource, presented by the curriculum in chemistry (advanced training), its presentation in the form of an ontology and creation on its basis of an interactive document by means of CIT "Polyhedron".

Keywords: educational resources, processing of natural language texts, ontological representation of texts, recursive reducer, interactive document.

С. И. Гайко,
В. В. Приходнюк

ПОДХОД К АВТОМАТИЗИРОВАННОЙ СТРУКТУРИЗАЦИИ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ НА ОСНОВЕ МЕТОДА РЕКУРСИВНОЙ РЕДУКЦИИ

Аннотация. Статья посвящена вопросам автоматизированной структуризации образовательных информационных ресурсов, представленных естественно-языковыми текстами, с целью представления их пользователям в удобной (в частности, интерактивной) форме. Рассмотрены проблемы обработки, представления, использования и интеграции больших объемов междисциплинарных знаний. Проведен анализ наиболее распространенных программных средств, которые предназначены для обработки естественно-языковых текстов и могут быть использованы для решения поставленной задачи. Предложен способ оценки таких программных средств на основе многокритериальной оптимизации. Определены и обоснованы критерии такой оценки, а также приведены ее результаты. На основе полученных результатов показано, что наиболее эффективным способом решения задачи является система «Рекурсивный редуктор», позволяющая анализировать широкий спектр документов и представлять результаты в разной форме, включая XML-файлы онтологий, которые могут быть использованы в работе когнитивной ИТ-технологии (КИТ) «Полиэдр». В статье дано краткое описание математической модели, на основе которой работает «Рекурсивный редуктор». Данная модель основывается на использовании λ -выражений. λ -выражения применяются для формирования специализированных рекурсивных функций, выполняющих идентификацию лексем во входном тексте и формирующих на их основе объекты и их атрибуты. Дополнительно описано применение онтологического классификатора документов для повышения эффективности работы рекурсивного редуктора. Также приведен пример его работы – автоматизированное структурирование образовательного информационного ресурса, представленного учебной программой по химии (углубленное обучение), представление ее в виде онтологии, а в дальнейшем — создание на ее основе трансдисциплинарного интерактивного документа посредством КИТ «Полиэдр».

Ключевые слова: образовательные ресурсы, обработка естественно-языковых текстов, онтологическое представление текстов, рекурсивный редуктор, интерактивный документ.

ІНФОРМАЦІЯ ПРО АВТОРІВ

Гайко Світлана Іванівна — аспірантка, Інститут телекомунікацій і глобального інформаційного простору НАН України, м. Київ, Україна, svitgai@i.ua; ORCID ID: <https://orcid.org/0000-0002-3564-475X>

Приходнюк Віталій Валерійович — канд. техн. наук, старший науковий співробітник, НЦ «Мала академія наук України», м. Київ, Україна, tangens91@gmail.com; ORCID ID: <https://orcid.org/0000-0002-2108-7091>

INFORMATION ABOUT THE AUTHORS

Haiko S. I. — PhD student, Institute of Telecommunications and Global Information Space NAS of Ukraine, Kyiv, Ukraine, svitgai@i.ua; ORCID ID: <https://orcid.org/0000-0002-3564-475X>

Prykhodniuk V. V. — PhD in Engineering, Senior Researcher, the NC “Junior Academy of Sciences of Ukraine”, Kyiv, Ukraine, tangens91@gmail.com; ORCID ID: <https://orcid.org/0000-0002-2108-7091>

ИНФОРМАЦИЯ ОБ АВТОРАХ

Гайко С. И. — аспирант, Институт телекоммуникаций и глобального информационного пространства НАН Украины, г. Киев, Украина, svitgai@i.ua; ORCID ID: <https://orcid.org/0000-0002-3564-475X>

Приходнюк В. В. — канд. техн. наук, старший научный сотрудник, НЦ «Малая академия наук Украины», г. Киев, Украина, tangens91@gmail.com; ORCID ID: <https://orcid.org/0000-0002-2108-7091>

Стаття надійшла до редакції / Received 29.05.2021