

**СИСТЕМА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ВІДПОВІДНОСТІ ЛЕКЦІЙНИХ МАТЕРІАЛІВ  
НАВЧАЛЬНИХ ДИСЦИПЛІН СТАНДАРТАМ ОСВІТИ**

*У статті досліджено розвиток дистанційної освіти як сучасного способу надання якісних освітніх послуг. Доведено необхідність контролю вмісту лекційних матеріалів навчальних дисциплін. Запропоновано інформаційні технології побудови повних онтологій опису та лекційних матеріалів. Викладено основні засади розробки системи інтелектуального аналізу відповідності лекційних матеріалів навчальних дисциплін стандартам освіти.*

*This article investigates the development of modern distance education as a way of providing quality education, proves the necessity of controlling the lectures contents of the academic disciplines, proposes the information technologies to create the comprehensive ontologies and lecture materials, describes the basic principles of the development of data mining systems that check an accordance of the lectures contents with the education standards.*

Стан освіти в сучасному світі складний і суперечливий. З одного боку, освіта в 20-му столітті стала однією з найважливіших сфер людської діяльності; величезні досягнення в цій галузі лягли в основу грандіозних соціальних і науково-технологічних перетворень, характерних для минаючого століття [1]. З іншого боку, розширення сфери освіти та зміну її статусу супроводжуються загостреннями проблем у цій сфері, які свідчать про кризу освіти. І, нарешті, в останні десятиліття в процесі пошуку шляхів подолання кризи освіти відбуваються радикальні зміни у цій сфері та формування нової освітньої системи.

Сфера освіти істотно перетинається в суспільстві з інформаційно-технічною сферою життя суспільства, а освітня діяльність стає найважливішим компонентом його розвитку. Не треба також забувати, що інформація та теоретичне знання є стратегічними ресурсами країни і, разом з рівнем розвитку освіти, багато в чому визначають її суверенітет і національну безпеку. Можна говорити про становлення освіти як найважливішого чинника подолання відсталості в розвитку більшої частини людства. Від вирішення проблем освіти, які завжди гостро стояли в країнах, що розвиваються і які ще більше посилюються в останні десятиліття в зв'язку з бурхливим розвитком інформаційних технологій, залежать зараз перспективи соціально-економічного розвитку цих країн, вирішення глобальної проблеми подолання відсталості у світі [2].

На сьогоднішній день, дистанційна освіта (distance learning) є найбільш ефективним засобом застосування інформаційних технологій для надання освітніх послуг та контролю отриманих знань [3]. Такий спосіб освіти має ряд переваг – окрім незалежності від географічної віддаленості, можна виділити ще й незалежність від часових рамок. Адже навчання відбувається у вільний час, у комфортних для студента умовах – вдома чи на роботі. Студенти дистанційних програм одержують навчальні матеріали через супутникове або кабельне телебачення, можуть взяти матеріали через Інтернет, беруть участь у відео-конференціях і консультаціях з викладачами, перебуваючи на будь-якій відстані від нього. На відміну від заочної форми навчання, де контроль за вивченням матеріалу припадає на певний час (зокрема на період сесії), дистанційна освіта дозволяє контролювати знання після кожного пройденого курсу – має більш гнучкий графік складання іспитів [3]. На сучасному етапі велика кількість навчальних закладів пропонують послуги цієї форми навчання – як в світі, так і на Україні [4, 5].

Для організації навчання використовується система дистанційної освіти (СДН). Система дистанційного навчання – це клієнт-серверне програмне забезпечення. Для її роботи функціонує головний комп'ютер (сервер), на якому встановлена сама програма, а користувачі можуть підключатися до серверу через мережу Інтернет або через локальну комп'ютерну мережу університету.

Для забезпечення повноцінного навчання студентів-дистанційників та контролю їх знань викладачі наповнюють свої навчальні курси необхідними навчальними матеріалами, створюють тести. Хоча існує багато можливих засобів надання знань та контролю за їх засвоєнням (лекційний матеріал, відео- та аудіо матеріали, віртуальні лабораторні роботи, тестові завдання, контрольні роботи, онлайн-семінари та ін.), проте загальноприйнятим є підхід [6] по застосуванню в якості інструменту навчання лекційних матеріалів визначеної структури, і тестів в якості інструменту контролю рівня отриманих знань [7].

Структура та зміст лекційного матеріалу регламентуються освітньо-професійними програмами підготовки за відповідними освітньо-кваліфікаційними рівнями (галузевими стандартами вищої освіти) [8]. В умовах, коли якість дистанційної освіти забезпечується відповідними лекційними матеріалами та набором тестових завдань, контроль за їх якістю набуває особливого значення. Відповідно, для контролю якості послуг у системі дистанційної освіти першочерговим є завдання визначення відповідності лекційних матеріалів навчальних дисциплін стандартам освіти.

Не завжди якість лекційного матеріалу правильно може оцінити людина. У більшості випадків варто скористатись інструментами, що дають інформаційні технології. Це допоможе уникнути

суб'єктивізму, знизить вірогідність помилки, охопить повний обсяг вихідної інформації, дозволить оцінити якість матеріалу швидко, без втручання людини.

Метою роботи є розробка інформаційної технології інтелектуального аналізу відповідності лекційних матеріалів навчальних дисциплін стандартам освіти.

Метод співставлення лекційних матеріалів навчальних дисциплін та галузевих стандартів вищої освіти визначається особливостями опису навчальних дисциплін освітньо-професійними програмами підготовки фахівців. На рисунку 1 показано приклад опису навчальної дисципліни «Організація баз даних та знань» згідно освітньо-професійною програмою підготовки бакалаврів напряму «Комп'ютерні науки» [9]. Очевидно, що ієрархія змістовних блоків визначає верхній рівень вертикальної онтології відповідної навчальної дисципліни. Тому порівняння онтології лекційних матеріалів навчальної дисципліни та онтології опису навчальної дисципліни згідно стандартів освіти є закономірним засобом визначення їх відповідності.

Шифр навчальної дисципліни	Назва навчальної дисципліни	Назва блока змістового модуля	Шифр блока змістових модулів	Назва теми (змістового модуля)	Шифр теми (змістового модуля)
3.04	Організація баз даних та знань	Моделювання даних	3.04.01	Системи баз даних. Основні поняття й архітектура	3.04.01.01
				Моделі даних	3.04.01.02
				Реляційна модель даних	3.04.01.03
				Теорія нормалізації реляційної моделі даних	3.04.01.04
		Мови запитів	3.04.02	Мова SQL	3.04.02.01
				Мова QBE	3.04.02.02
		Проектування та захист баз даних	3.04.03	Проектування баз даних	3.04.03.01
				Цілісність даних	3.04.03.02
				Захист баз даних	3.04.03.03
				Навігаційна обробка даних	3.04.03.04
		Класифікація баз даних	3.04.04	Розподілені бази даних	3.04.04.01
				Паралельні бази даних	3.04.04.02
				Делуксовані бази даних	3.04.04.03
				Об'єктно-орієнтовані бази даних	3.04.04.04
				Бази даних в Інтернеті	3.04.04.05
				Бази знань	3.04.04.06

Рисунок 1 – Приклад опису навчальної дисципліни

Зважаючи на наявні рівні онтології опису навчальної дисципліни (Назва, Модуль, Тема) та природно необхідність структуризації тем (підтеми, розділи чи параграфи), було отримано онтологію опису навчальної дисципліни, показану на рисунку 2. Елементи структуризації тем не обов'язково потребують верифікації, але включаються до онтології опису навчальної дисципліни, оскільки цей рівень є присутнім у онтології лекційних матеріалів.

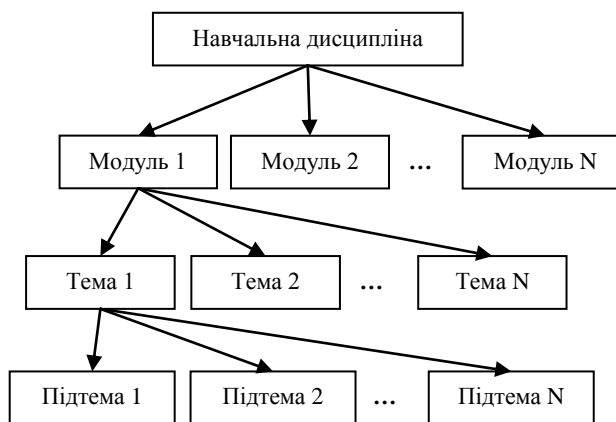


Рисунок 2 – Онтологія опису навчальної дисципліни

Онтологія опису навчальної дисципліни в показаному вигляді регламентує структуру лекційних матеріалів, яка є основою для побудови відповідної онтології. Структура лекційних матеріалів як електронних документів регламентується мовами розмітки документів (наприклад, WordprocessingML для XML) [10] й реалізується через систему заголовків. По відношенню до онтології лекційних матеріалів навчальної дисципліни, заголовки співставлені їй відповідним рівням, як це показано у таблиці 1.

Таблиця 1 – Відповідність елементів структури цифрових документів рівням онтології лекційних матеріалів

Порядок в ієрархії	Рівень онтології лекційних матеріалів	Елемент структури цифрового документу
1	Навчальна дисципліна	Заголовок 1
2	Модуль	Заголовок 2
3	Тема	Заголовок 3
4	Підтема	Заголовок 4

Автоматизоване порівняння онтології опису навчальної дисципліни та онтології лекційних матеріалів дозволяє визначити показники їх відповідності, а відповідно – якість структури лекційних матеріалів навчальної дисципліни.

На відміну від структури, інформаційний вміст лекційних матеріалів вимагає оперування ключовими термінами (поняття, визначення) [11]. З семантичної точки зору, ключові терміни є найнижчим рівнем онтології лекційних матеріалів навчальної дисципліни. Їх особливістю є наявність властивості рівня прив'язки; адже поряд із термінами, що використовуються суто в рамках певних підтем, інші терміни можуть бути актуальними в масштабах тем, модулів, і навіть всієї навчальної дисципліни.

Структура лекційного матеріалу визначає місцезнаходження визначених термінів, але не включає їх безпосередньо. Наприклад, назва підтеми «Основні поняття» ніяким чином не визначає власне перелік понять. Тому формування цього найнижчого рівня онтології опису навчальної дисципліни покладається на експерта як необхідний етап деталізації опису навчальної дисципліни. Створення переліку ключових термінів навчальної дисципліни та їх онтоприв'язка дозволяють створити повну онтологію опису навчальної дисципліни (рис. 3).

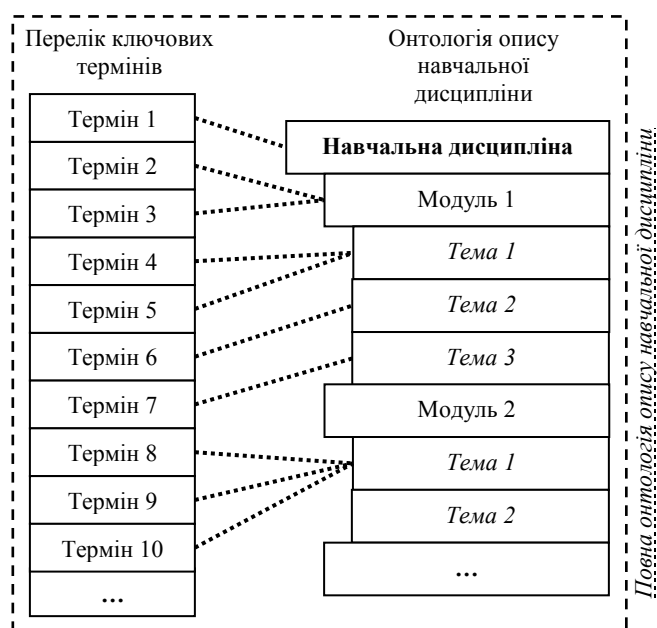


Рисунок 3 – Приклад повної онтології опису навчальної дисципліни

Для побудови повної онтології лекційних матеріалів навчальної дисципліни, окрім аналізу їх структури як електронних документів, також потрібна побудова переліку ключових термінів. Для цього є доцільним використання алгоритмів аналізу інформаційного вмісту текстових матеріалів [12]. Всі ці підходи дозволяють створювати переліки слів на підставі текстів, в яких окремим словам або словосполученнями деяким спеціальним чином співставлені у відповідність числові вагові значення. Як функцію, що ставить у відповідність слову число, можна розглядати, наприклад, порядковий номер унікального слова в тексті, довжину слова, «вагу» слів у текстах, оцінку TFIDF, а також інші вагові оцінки. Серед всіх алгоритмів аналізу інформаційного змісту текстів найбільш відомими є частотний аналіз, TFIDF та дисперсійна оцінка [13].

Частотний аналіз передбачає визначення кількості зустрічань кожного із слів у тексті. В результаті застосування цього методу, слова у тексті визнаються тим більш значущими, чим більша кількість їх повторів має місце у тексті, що аналізується.

Метод TFIDF для визначення дискримінантної сили слів використовує інверсну частоту появи слова в окремих документах масиву. Запропонований метод зважування слів має позначення TF IDF, де TF указує на частоту появи слів у документі, а IDF – на величину, зворотну кількості документів у масиві, що містять дане слово [14, 15]. За цим методом, текст розбивається на фрагменти, що містять задану кількість слів (наприклад, 500). Потім для кожного слова, що входить у текст, підраховується кількість фрагментів, в які це слово входить, а також загальна кількість входжень даного слова в текст. Після цього розраховується середнє значення TFIDF вагової оцінки кожного слова.

Деякі автори пропонують визначити дескримінантну силу слів у тексті на основі дисперсійної оцінки [16]. Припускається, що оцінка нерівномірності входження слів, яка визначається за даним методом, вказує на рівень смислової важливості даного слова у тексті.

Таким чином, дисперсійна оцінка дозволяє відокремити слова, що зустрічаються в тексті відносно рівномірно (для рівномірно розподілених слів ця оцінка дорівнює нулю), від слів, розподілених нерівномірно. Тобто це є оцінка відмінної, дискримінантної сили слів, зокрема, для інформаційного пошуку.

Ідея дисперсійної оцінки дуже близька до TFIDF, при цьому менш поширена, однак більш коректно застосовна до повних одиничних текстів на кшталт лекційних матеріалів навчальних дисциплін [17], а не до масивів текстів, як TFIDF.

Дисперсійна оцінка реалізується наступним чином [12]: нехай текст складається з  $N$  слів ( $n = 1, \dots, N$ ),  $N$  – порядковий номер слова в тексті, позиція слова. Деяке слово, наприклад  $A$ , позначається як  $A_k^n$ , де індекс  $k = 1, 2, \dots, K$  – номер появи даного слова в тесті, а  $n$  – позиція даного слова в тексті. Наприклад,  $A_2^{50}$  означає, що на 50-ій позиції тексту знаходиться слово  $A$ , яке зустрілось третій раз.

Інтервалом між послідовними появами слова при таких позначеннях буде величина  $\Delta A_k = A_{k+1}^m - A_k^n = m - n$ , де на  $m$ -ій і  $n$ -ій позиціях в тесті знаходиться слово  $A$ , яке зустрілось  $k+1$ -ий і  $k$ -ий рази.

Тоді дисперсійна оцінка розраховується як

$$\sigma_A = \frac{\sqrt{(\Delta A^2)} - (\Delta A)}{(\Delta A)},$$

де  $(\Delta A)$  – середнє значення послідовності  $\Delta A_1, \Delta A_2, \Delta A_k, (\Delta A^2)$  – послідовності  $A_1^2, A_2^2, A_k^2, K$  – кількість появи слова  $A$  в тексті.

Деякі автори пропонують у якості ефективної надбудови над алгоритмом визначення дисперсійної оцінки побудову компактифікованого графа горизонтальної видимості для встановлення інформаційно-значущих елементів тексту з оглядом на його структуру [12]. Мережа слів із використанням алгоритму горизонтальної видимості будується в три етапи. На першому етапі на горизонтальній осі відзначається ряд вузлів, кожен з яких відповідає словам у порядку появи в тексті, а по вертикальній осі відкладаються вагові чисельні оцінки. На другому етапі будується традиційний граф горизонтальної видимості, у якому між вузлами існує зв'язок, якщо вони знаходяться в межах «прямої видимості», тобто якщо їх можна з'єднати горизонтальною лінією, що не перетинає жодну іншу вертикальну лінію. На третьому, заключному етапі, отримана на попередньому етапі мережа компактифікується – всі вузли з певним словом об'єднуються в один вузол, всі зв'язки таких вузлів також об'єднуються, в результаті ступінь (число зв'язків) кожного слова визначається кількістю компактифікованих зв'язків із іншими словами у межах «прямої видимості».

З метою вибору найбільш ефективного методу аналізу вмісту текстових матеріалів для формування переліків ключових слів у лекційних матеріалах навчальних дисциплін, авторами було розроблено тестовий програмний продукт. Наприклад, даний програмний продукт на основі контенту лекції «Реляційна модель даних» (рис. 4) навчальної дисципліни «Організація баз даних та знань» буде переліки значущих слів (рис. 5) згідно описаних вище методів (починаючи з максимальних вагових значень, по спаданню). У таблиці 2 подано результати у вигляді списків з 30 найбільш значимих слів у даній лекції згідно методів частотного аналізу, TFIDF та дисперсійної оцінки.

Отримані результати свідчать, що методам частотного аналізу та TFIDF заважають попадання до переліку значущих слів не тільки змістовно значущих слів, а й значущих слів для зв'язування семантичних конструкцій (сполучники, займенники та ін.), у той час як метод дисперсійної оцінки дозволяє одразу побудувати перелік значущих слів саме з змістовної точки зору. Даний метод дає достатньо якісний результат навіть без подальшого застосування горизонтального графу видимості та лематизації. Оскільки метод дисперсійної оцінки є більш ефективним для пошуку значущих слів у текстах, то є доцільним його використання при аналізі лекційних матеріалів навчальної дисципліни для побудови їх повної онтології.

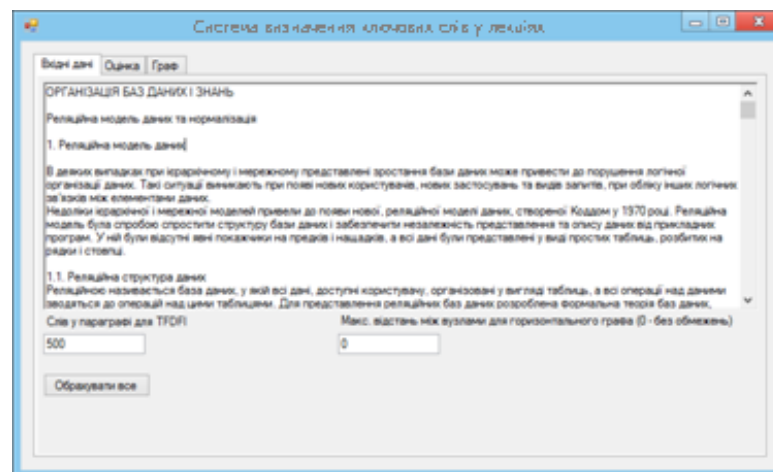


Рисунок 4 – Завантаження лекційного матеріалу до системи пошуку ключових слів

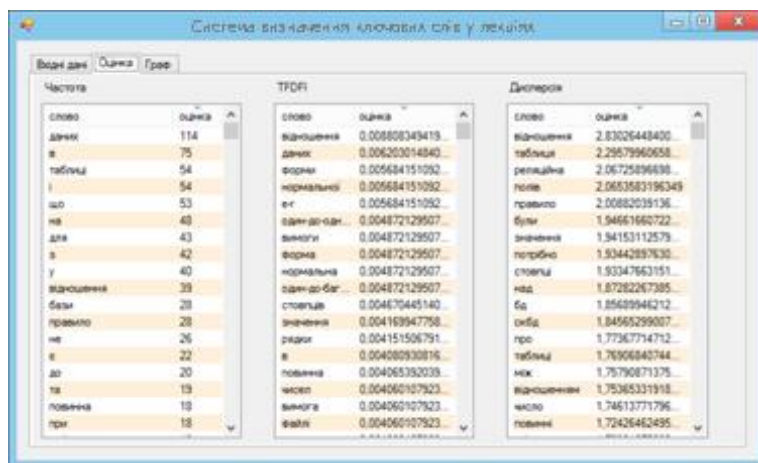


Рисунок 5 – Результат аналізу лекційного матеріалу системою пошуку ключових слів

Таблиця 2 – Приклад результату при пошуку ключових слів

<i>Частотний аналіз</i>	<i>TFIDF</i>	<i>Дисперсійна оцінка</i>
даних	даних	відношення
в	і	таблиця
таблиці	бути	значення
і	це	полів
що	в	число
на	у	реляційна
для	бази	правило
з	що	стовпці
у	з	повинні
відношення	на	скбд
правило	для	відношенням
бази	таблиці	цілісності
не	відношення	бази
є	нормальної	таблиці
до	форми	таблицями
та	нормальна	порядок
бути	форма	правила
повинна	вимоги	операції
при	стовпців	зв'язків
таблиць	рядки	ключа
які	файли	первинний
ключа	вимога	накладаються
щоб	чисел	реляційний
таблиця	правило	реалізації
баз	таблиця	номер
від	роботи	стовпець
повинні	між	ключів
між	відновлення	рядки
ключ	є	нормалізації
це	вимагає	таблиця

Загалом, автоматизована побудова повних онтологій опису та лекційних матеріалів навчальної дисципліни в рамках відповідної експертної системи дозволяє досягти наступних основних результатів (рис. 6):

- допомога в побудові структури та вмісту лекційних матеріалів шляхом автоматичної побудови структури документа та переліків термінів;
- оцінка відповідності лекційних матеріалів вимогам стандарту як результат порівняльного аналізу повних онтологій опису та лекційних матеріалів;
- рекомендації по вдосконаленню лекційних матеріалів у вигляді згенерованого експертного висновку за показниками відповідності лекційних матеріалів вимогам стандарту та повною онтологією опису навчальної дисципліни.



Рисунок 6 – Узагальнена схема роботи системи інтелектуального аналізу відповідності лекційних матеріалів стандартам освіти

Таким чином, у статті досліджено розвиток дистанційної освіти як сучасного способу надання якісних освітніх послуг. Доведена необхідність контролю вмісту лекційних матеріалів навчальних дисциплін. Запропоновано інформаційні технології побудови повних онтологій опису та лекційних матеріалів. Викладено основні засади розробки системи інтелектуального аналізу відповідності лекційних матеріалів навчальних дисциплін стандартам освіти.

### Література

1. Нові інформаційні технології в освіті [Електронний ресурс]. – Режим доступу: <http://it-tehnolog.com/statti/novi-informatsiyuni-tehnologiyi-navchannya/>
2. Концепція якості освіти [Електронний ресурс]. – Режим доступу: <http://osvita.ua/>
3. Мазур М.П. Розвиток дистанційного навчання в Україні як складової інформатизації сучасного суспільства. - Інформатика та інформаційні технології в навчальних закладах. - №1, 2007. – с. 71-75.
4. University of the People [Електронний ресурс]. – Режим доступу: <http://www.uopeople.org/>
5. Факультет заочно-дистанційного навчання, післядипломної освіти та довузівської підготовки ХНУ [Електронний ресурс]. – Режим доступу: <http://dn.tup.km.ua/>
6. Мазур М.П., Яновський М.Л. Нова модель цифрової дистрибуції на прикладі навчального процесу // Вимірювальна та обчислювальна техніка в технологічних процесах. – Хмельницький, 2009. - №1. – с.159-166.
7. Павлыш В.Н., Зайцева М.Н. Теоретические основы контроля в учебном процессе высшей школы // Научные труды ДНТУ. – 2011. – № 9. – С. 130-136.
8. Гончаренко С.У. Державні стандарти професійної освіти: теорія і методика: Моногр. / АПН України. Ін-т педагогіки і психології проф. освіти, Технол. ун-т Поділля. – Хмельницький: ТУП, 2002. – 334с.
9. Галузевий стандарт вищої освіти України з напрямку підготовки 6.050101 «Комп'ютерні науки»: Збірник нормативних документів вищої освіти. – К.: Видавнича група BHV, 2011, – 85с.
10. Сайт Microsoft для розробників OpenXMLDeveloper [Електронний ресурс]. – Режим доступу: <http://openxmldeveloper.org/>
11. Палагин А.В., Петренко Н.Г. Системно-онтологический анализ предметной области // Управляющие системы и машины. - 2009. – № 4. - С. 3-14.
12. Д.В. Ландэ, А.А. Снарский, Е.В. Ягунова Использование графов горизонтальной видимости для выявления слов, определяющих информационную структуру текста // Труды 15-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции — RCDL-2013» – Ярославль: 2013. – с. 158-164.
13. Giora R. Segmentation and Segment Cohesion: On the Thematic Organization of the Text // Text. An Interdisciplinary Journal for the Study of Discourse Amsterdam. 1983, № 2. – P. 155-181.
14. João Ventura and Joaquim Ferreira da Silva Ranking and Extraction of Relevant Single Words in Text // Frontiers in Brain, Vision and AI
15. Salton G., McGill M. J. Introduction to Modern Information Retrieval. – New York: McGraw-Hill, 1983. – 448 p.
16. Ortuño M., Carpena P., Bernalola P., Muñoz E., Somoza A.M. Keyword detection in natural languages and DNA // Europhys. Lett., 2002, 57. – P. 759-764.

17. Снитюк В.Е., Юрченко К.Н. Интеллектуальное управление оценением знаний. – Черкассы, 2013. – 262с.