

## ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АВТОМАТИЗОВАНОГО ВИЗНАЧЕННЯ ТЕРМІНІВ У НАВЧАЛЬНИХ МАТЕРІАЛАХ

Рассмотрен вопрос автоматизированного поиска ключевых терминов в контенте учебных материалов как составляющая задачи автоматизации процесса разработки тестовых заданий. Предложена информационная технология, основанная на дисперсионной оценке важности слов, которая позволяет с достаточно высокой эффективностью определять семантические термины в контенте учебных материалов. Рассмотрены факторы, затрудняющие эффективное определение семантических терминов в учебных материалах.

Ключевые слова: ключевые термины, анализ текста, дисперсионная оценка, тесты, учебные материалы.

The question automated search of key terms in the content of educational materials as part of automation of process development tests task have been considered. Proposed the information technology, based on the dispersion assessment of the importance of words, which allows a high enough efficiency to determine the semantic content in terms of educational materials. The factors that hinder effective analysis of educational materials have been considered.

Keywords: key terms, text analysis, disperse evaluation, tests, educational materials.

Тести на сучасному етапі є загальноприйнятим методом контролю рівня отриманих знань як у дистанційній освіті, так і в інших формах навчання. Автоматизація процесу розробки тестових завдань дозволяє суттєво розширити спектр застосування й якість тестового матеріалу [1]. Для досягнення цієї мети основним завданням є автоматизований пошук ключових термінів у контенті матеріалу, що вивчається, оскільки саме терміни є семантичними об'єктами вивчення та перевірки.

Знайдені терміни є базовими вузлами для автоматизованої побудови семантичних конструкцій, що використовуються для формування контенту тестових завдань. Метод генерації тестових завдань (параметризовані тести, семантичні мережі, понятійно-тезова модель) визначає спосіб формування контенту тестових завдань на основі одержаного переліку термінів навчального матеріалу.

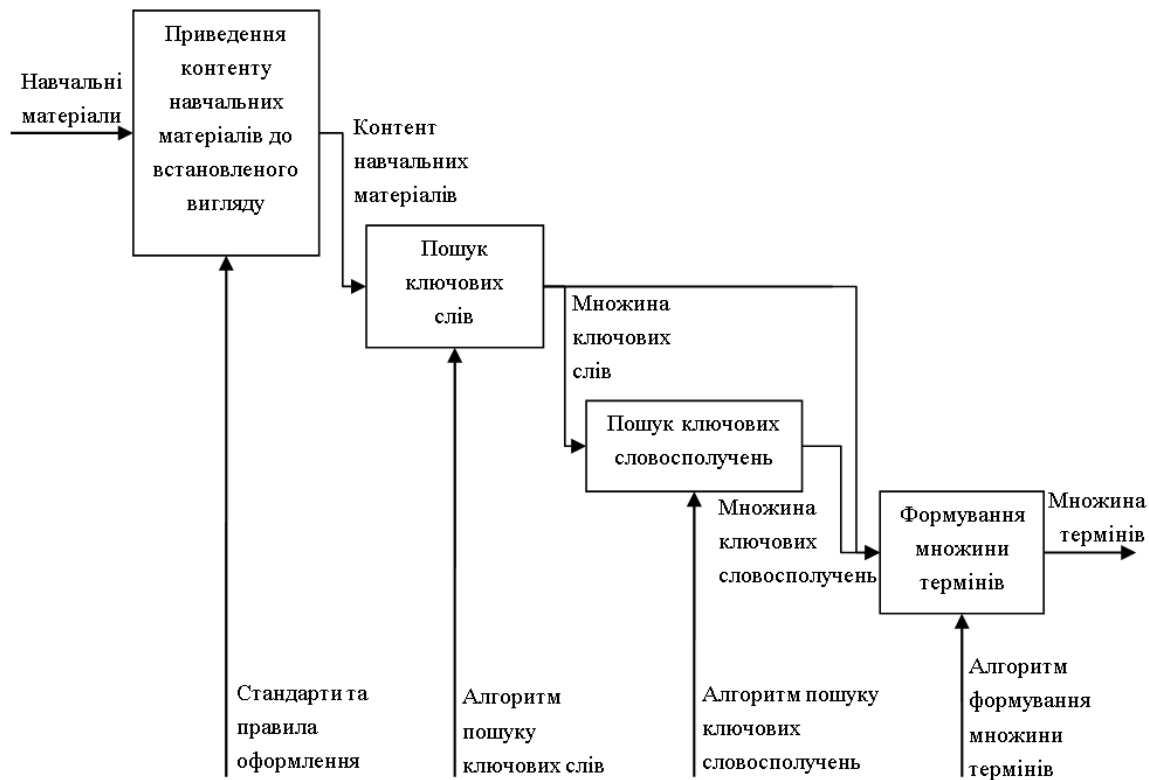
Термінами можуть виступати як слова, так і словосполучення й аббревіатури. Якщо аббревіатури з технічної точки зору розглядаються як слова, то словосполучення вимагають окремого алгоритму ідентифікації.

Задачу автоматизації визначення семантичних термінів у контенті навчальних матеріалів пропонується вирішувати шляхом використання розробленої інформаційної технології (рис. 1), яка включає наступні етапи перетворення інформації:

- попередня технічна обробка тексту, що полягає в усуненні неоднозначного іменування термінів та обробці розділових знаків;
- пошук ключових слів у контенті навчального матеріалу методом на базі дисперсійного оцінювання [2] й формування вихідної множини ключових слів;
- пошук неперервних скупчень важливих слів та їх аналіз з метою формування вихідної множини ключових словосполучень;
- заміщення словосполученнями слів, які є елементами словосполучень й формування вихідної множини термінів.

Для формування об'єднаної множини термінів на основі множин ключових слів та ключових словосполучень формується узагальнена множина термінів, до якої входять словосполучення й ті слова, значення дисперсійної оцінки яких суттєво перевищує значення дисперсійної оцінки зв'язаних із цим словом колокацій. Об'єднана множина термінів сортується за значеннями їх дисперсійної оцінки.

Таким чином, вхідними даними для інформаційної технології є контент навчальних матеріалів або його визначена частина; вихідними даними є ранжована множина семантичних термінів навчальних матеріалів.



**Рис. 1. Діаграма етапів пошуку семантичних термінів у контенті навчальних матеріалів**

З метою перевірки ефективності розробленої інформаційної технології було проведено порівняння результату автоматизованого визначення ключових семантичних термінів із переліком, сформованим експертом (автором відповідних матеріалів).

Аналіз отриманих результатів виявив, що відсутність програмно визначених термінів у множині автора не завжди характеризує недолік запропонованої технології. Деякі семантично важливі терміни автори суб'єктивно ігнорують, в той час як іншу категорію складають поняття, на яких автори акцентують надмірну увагу попри їх другорядність в рамках матеріалу, що викладається. Тому для різносторонньої оцінки результатів дослідження слід розглядати не тільки терміни із множини експерта, що не були знайдені програмно, а й автоматично знайдені терміни, які не увійшли до множини експерта.

В рамках дослідження розробленим тестовим програмним забезпеченням було оброблено вибірку з 30 лекцій із різних навчальних курсів. Практичне тестування розробленої інформаційної технології [3] виявило її середню ефективність 87,3%, показавши при цьому мінімальну ефективність 65,1% та максимальну – 100%.

Висока ефективність запропонованої технології надає підставу до її практичного застосування у вирішенні ряду актуальних задач, таких як оцінка відповідності навчальних матеріалів змістовим вимогам, оцінка відповідності наборів тестових завдань навчальним матеріалам, автоматизована генерація переліків ключових термінів та анотацій, а також автоматизованого формування тестових завдань.

#### Перелік посилань

1. Снитюк В.Е., Юрченко К.Н. Интеллектуальное управление оцениванием знаний. – Черкассы, 2013. – 262с.
2. Ventura, J. & Silva, J. (2007). New Techniques for Relevant Word Ranking and Extraction. In Proceedings of 13th Portuguese Conference on Artificial Intelligence, Springer-Verlag, pp. 691-702.
3. Бармак О.В., Мазурець О.В. Інформаційна технологія автоматизованого визначення термінів у навчальних матеріалах // Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах» – Хмельницький, 2015. – №2. – С.94–102.