

## МЕТОДИ СЕМАНТИЧНОГО АНАЛІЗУ РОЗГОРНУТИХ ВІДПОВІДЕЙ У СИСТЕМАХ АДАПТИВНОГО НАВЧАННЯ

При розробці автоматизованих систем адаптивного навчання постає проблема аналізу розгорнутих відповідей студентів. Одним із рішень поставленої проблеми може бути використання методів семантичного аналізу тексту.

Серед існуючих методів аналізу текстових даних чинне місце посідає метод з використанням векторної моделі. Основна ідея якого полягає в алгебраїчному представленні колекції текстових масивів векторами одного спільного для всієї колекції векторного простору. Векторна модель є основою для вирішення багатьох завдань, таких як: пошук документа за запитом, класифікація документів та кластеризація документів.

Для знаходження близьких за сенсом речень можна знаходити косинус кута між векторами речень. Слова у реченнях можуть повторюватись, але бути у різній формі, саме тому використовують стемінг – це процес скорочення слова до основи шляхом відкидання допоміжних частин, таких як закінчення чи суфікс. Результати іноді дуже схожі на визначення кореня слова, але його алгоритми базуються на інших принципах. Стемінг застосовується в лінгвістичній морфології та в інформаційному пошуку. Використовуючи стемінг ми отримуємо вектори, які найбільш вдало демонструють речення. Саме такі вектори використовують для знаходження косинусу кута між ними.

Іншим ефективним і актуальним засобом вилучення семантики із тексту та її представлення є метод латентно-семантичного аналізу (ЛСА) [1].

Д. В. Ланде розглянув метод ЛСА в рамках технологій глибинного аналізу текстової інформації Text Mining, що розроблені на основі статистичного та лінгвістичного аналізу, і дозволяють не лише здійснювати відбір релевантних документів, а й виділяти їх семантику, яка досить часто буває прихованою [2].

ЛСА є теорією і методом екстракції і представлення контекстно-залежного змісту слів шляхом статистичної обробки великого корпусу текстів. Головною ідеєю методу є те, що сукупність усіх контекстів, в яких певне слово вживається або, навпаки, не вживається, обумовлює набір обмежень, які визначають подібність значень слів або множини слів. Отже, простежується думка, що між словами і контекстом, в якому вони вживаються існують приховані (латентні) зв'язки. Метод ЛСА дозволяє визначити асоціативну і семантичну близькість та вирахувати кореляції між двома термами, двома документами, або між термом і документом.

Ефективність застосування методу ЛСА в сфері знань людини підтверджена різноманітними прикладами його роботи. Зокрема, вперше зазначений метод був застосований з метою автоматичного індексування текстів та виявлення їх асоціативно-семантичної структури. Використання методу ЛСА знайшло своє відображення у системах вилучення, представлення семантичної інформації із тексту. Метод ЛСА пройшов випробування та підтвердив свою ефективність у таких напрямках обробки природної мови як моделювання концептуальних знань людини; інформаційний пошук, при реалізації якого ЛСА показує набагато кращі результати порівняно із звичайними векторними методами.

Для детальнішої обробки розгорнутих відповідей студента в системах адаптивного тестування якості знань можна використовувати синонімічні ряди слів при латентно-семантичному аналізі. Це дозволить ефективніше вирішувати задачу визначення унікальності розгорнутої відповіді студента, як серед відповідей інших студентів курсу, так і серед текстів популярних інтернет-ресурсів.

Використання синонімічних сполучень в алгоритмі вилучення семантики із тексту дозволить досягнути меншої чутливості до окремих формулювань думок студентів.

### Список літератури

1. Заболеева-Зотова А. В. Латентный семантический анализ: новые решения в Internet / А. В. Заболеева-Зотова., 2001. – (Информационные технологии).
2. Ландэ Д. В. Поиск знаний в Internet. Профессиональная работа / Д. В. Ландэ., 2005. – 272 с.