Карпов І.А., [karpovilya5@gmail.com](mailto:karpovilya5@gmail.com), ДНУ імені Олеся Гончара, Антоненко С.В.

**ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ТЕКСТУ. ПРОБЛЕМИ ТА МЕТОДИ**

Все більшої актуальності в різних областях людської діяльності набуває глибинний аналіз різноманітних даних. На сучасних підприємствах, в дослідницьких проектах або в Інтернеті утворюються великі обсяги даних.

Глибинний (інтелектуальний) аналіз даних (Data mining) здійснюється автоматично шляхом застосування методів математичної статистики, штучних нейронних мереж, теорії нечітких множин або генетичних алгоритмів. Метою аналізу є виявлення правил та закономірностей, наприклад, статистичних подій. Так, наприклад, можливо виявити зміни у поведінці клієнтів або груп клієнтів для покращення стратегії підприємства.

Один із напрямів інтелектуального аналізу даних та штучного інтелекту є інтелектуальний аналіз тексту. Мета інтелектуального аналізу тексту є отримання інформації з колекцій текстових документів, ґрунтуючись на застосуванні ефективних, у практичному плані, методів машинного навчання та обробки природної мови. Інтелектуальний аналіз тексту використовує всі ті ж підходи до переробки інформації, що й глибинний аналіз даних, однак різниця між цими напрямками проявляється лише в кінцевих методах, а також у тому що, що глибинний аналіз даних має справу з сховищами та базами даних, а не електронними бібліотеками та корпусами текстів.

Ключовими завданнями інтелектуального аналізу тексту є наступні: категоризація текстів, пошук інформації, обробка змін в колекціях текстів, а також розробка засобів представлення інформації для користувача. Одним із найпоширеніших напрямків використання інтелектуального аналізу даних є аналіз тексту та перевірка його на коректність за граматичної чи смислової точки зору.

Найбільша проблема систем інтелектуального аналізу тексту – це навчити систему не тільки розуміти запит, а ще й давати зрозумілу людині відповідь. Цим займається підрозділ інтелектуального аналізу тексту – question answering (відповідь за запитання). Приклад алгоритму роботи цього підрозділу:

1. Пошук інформації в документі для відбору тих частин, які потенційно можуть містити відповідь.
2. Фільтрація фраз, схожих на очікувану відповідь. На цьому етапі усе залежить від питання: наприклад, якщо він починається зі слова «де», фільтр виділить, зокрема, фрази, які містять географічні назви.
3. Пошук правильної відповіді. За допомогою допоміжних методів виводиться відповідь на запит.

Інтелектуальний аналіз тексту, здебільшого, використовує наступні методи:

* тематичне індексування – присвоєння документам/запитам тематичних індексів, які відображають деякі атрибути їх класифікації (за принципом бібліотечних каталогів);
* пошук за ключовими словами – використовує результати тематичного індексування для пошуку документів, які відповідають вказаним вимогам, зокрема, що містять зазначені користувачем ключові слова.

Використовуючи розглянуті методи та алгоритм, планується спроектувати та створити прикладну програмну систему інтелектуального аналізу тексту, дослідити існуючі методи інтелектуального аналізу тексту, розробити нові або модифікувати та покращити існуючі алгоритми для отримання більш точних результатів та більш кращого аналізу тексту.