# АНАЛІЗ ПІДХОДІВ ДО АВТОМАТИЗОВАНОЇ ГЕНЕРАЦІЇ ОПИСУ ТОВАРУ ЗА ВІДГУКАМИ ПРО НЬОГО

**Рогинський О.В., Бабкова Н.В., Кочуєва З.А.**
***Національний технічний університет «Харківський політехнічний інститут», м. Харків***

Проблеми обробки текстів виникли практично відразу за появою обчислювальної техніки. Але незважаючи на піввікову історію досліджень в області штучного інтелекту, величезний стрибок у розвитку ІТ та суміжних дисциплінах, задовільного вирішення більшості практичних завдань обробки тексту поки не існує.

Комп’ютерна лінгвістика – розділ науки, який вивчає застосування математичних моделей для опису лінгвістичних закономірностей. Її можна розділити на дві великі частини. Одна з них вивчає способи застосування обчислювальної техніки в лінгвістичних дослідженнях – застосування відомих математичних методів (наприклад, статистична обробка) для виявлення закономірностей. Виявлені закономірності використовуються іншою частиною, що вивчає питання осмислення текстів, написаних природною мовою, – створення математичних моделей для розв’язання лінгвістичних задач та розробка програм, які функціонують на основі цих моделей. Ця частина комп’ютерної лінгвістики тісно пов’язана з розділом штучного інтелекту, який займається розробкою систем обробки текстів природної мови.

Найбільш важливе завдання комп’ютерної лінгвістики – вилучення інформації з текстів та представлення її у вигляді формальної системи знань (наприклад, у вигляді семантичної мережі). Виконано ряд експериментальних розробок у даному напрямку, які орієнтовані на конкретні предметні області, проте повністю працездатних програмних продуктів немає.

Витяг інформації з текстів – основа для «розкопки» тексту, а також для створення систем завантаження текстів в сховища даних. Подібні системи існують та призначені для інтеграції й очищення даних, які розміщені в сховищах, але вони не надають ніяких засобів введення даних, що містяться в текстовому вигляді.

Поряд з отриманням інформації існує й зворотна задача генерації правильно побудованих текстів. Вихідними даними для таких систем є чітко формалізовані знання. На перший погляд, ця задача може здатися дивною, адже в більшості випадків формалізовані знання можна представляти у вигляді бланків, що мають чітку, заздалегідь визначену систему полів. Але це не завжди так. Якщо предметна область має складну та розгалужену структуру, то більшість полів бланка часто виявляються порожніми, що сильно ускладнює сприйняття інформації; для кінцевого користувача було б набагато простіше та зручніше мати справу не з такими бланками, а з неформалізованими (але коректна побудованим) текстовим описом тих же самих даних.

Таким чином, задачі вилучення факів з текстів та генерація генерації зв’язних текстів природною мовою досі являються актуальними проблемами, які потребують удосконалення існуючих методів вирішення.