

ДОСЛІДЖЕННЯ МЕТОДІВ СЕМАНТИЧНОГО АНАЛІЗУ ТА РОЗРОБКА ПРОГРАМНОГО ЗАСОБУ ДЛЯ АВТОМАТИЗОВАНОЇ ОБРОБКИ ТЕКСТОВОЇ ІНФОРМАЦІЇ

А.Ю. Запорожець¹, О.А. Татарінова²

¹ магістрант кафедри Комп'ютерного моделювання процесів і систем, НТУ «ХПІ», Харків, Україна

² доцент кафедри Комп'ютерного моделювання процесів і систем, НТУ «ХПІ», Харків, Україна

Доповідь, представлена на увагу спеціалістів та науковців, розглядає одне з найбільш прогресивних галузей комп'ютерних наук – автоматизацію обробки текстової інформації, яка виступає в ролі ключової компоненти в системах штучного інтелекту. Важливість цього напрямку зумовлена зростаючою потребою в розумінні природної мови, що є основним вектором спілкування між людьми, а також основою для формування і передачі знань. Непересічний інтерес до цієї теми викликаний потребою удосконалення взаємодії між людиною та машинами, де інтелектуальні системи здатні ідентифікувати, обробляти та інтерпретувати мовні конструкції з людською точністю та витонченістю.

Для забезпечення ефективності та точності процесів обробки природної мови, критично важливо розробляти алгоритми, що враховують лінгвістичні особливості мов, такі як синтаксичний порядок слів, морфологічні характеристики та семантичні відтінки. Реалізація обробки текстової інформації безпосередньо залежить від детального аналізу цих аспектів, оскільки вони визначають контекстуальне і функціональне навантаження слова в тексті.

В контексті класифікації емоційного забарвлення текстів, проблема розподілення текстової інформації за емоційним спектром виступає в якості складної задачі семантичного аналізу. Задля вирішення цього завдання пропонується реалізація комбінованого підходу, який передбачає використання як нейронних мереж, так і наївного байєсівського класифікатора. Нейронні мережі забезпечують високу точність за рахунок своєї здатності виявляти складні нелінійні взаємозв'язки в даних, тоді як наївний байєсівський класифікатор пропонується як лінійний метод, який може ефективно працювати з розділеними і незалежними ознаками текстів. Ця дуплексна архітектура дозволяє забезпечити більш високу точність визначення емоційного забарвлення тексту, підвищуючи ефективність системи класифікації.

У процесі розробки алгоритму класифікації, виконано його імплементацію у вигляді програмного коду, що включає в себе інтерфейс програмного застосування для ефективного зчитування та обробки необхідних даних. Програмування алгоритму виконувалося на мові Python та набору бібліотек, таких як NumPy, Pandas, scikit-learn, що надають значні переваги для роботи з текстовими даними. Інтеграція цих інструментів дозволяє створити потужну та адаптивну систему, яка може точно класифікувати текстову інформацію за емоційними характеристиками, забезпечуючи широкі можливості для подальшого застосування в різних областях діяльності, від соціальних медіа до клієнтського обслуговування.