

УДК 629.735.072.8.08:004(043.2)

Бабинець К. – ст. гр. СП-41, Пастух О.

Тернопільський національний технічний університет імені Івана Пулюя

ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ КЛАСИФІКАЦІЇ ТЕКСТОВОЇ ІНФОРМАЦІЇ

Науковий керівник: д.т.н., професор Пастух О.А.

Babynets Kh., Pastukh O.

Ternopil Ivan Puluuj National Technical University

SOFTWARE FOR CLASSIFICATION OF TEXT INFORMATION

Supervisor: Dr.Sc., Professor Pastukh O.A.

У сьогоднішній день очевидним є факт потоків великих даних різної природи інформації, серед яких важливе місце займає текстова інформація. Говорять навіть про те, що ми живемо не лише в еру інформаційних технологій, але і в еру технологій даних. З плином часу кількість даних збільшується по експоненційному закону, тому виникає гостра потреба автоматичної обробки текстової інформації на основі комп'ютерів, які у свою чергу потребують використання сучасного програмного забезпечення для класифікації текстової інформації. На рис.1 наведено програмний конвеєр для класифікації текстової інформації.



Рис.1 – Програмний конвеєр для класифікації текстової інформації

Програмний конвеєр для класифікації текстової інформації містить базові кроки, що зображені на рис.1. Перший крок – це попередня обробка текстової інформації: узгодження реєстерних форм даних, виправлення граматичних помилок, тощо. Другий крок – це класифікація на основі алгоритму машинного навчання. Третій крок – це процес тестування. Четвертий крок – це візуалізація даних і зберігання окремих програмних компонент для їх використання сторонніми програмами.