

АНОТАЦІЯ

Дипломну роботу виконано на 50 аркушах, вона містить 2 додатки та перелік посилань на використані джерела з 20 найменувань. У роботі наведено 8 рисунків та 1 таблицю.

Роботу присвячено розробці математичного та програмного забезпечення системи збору і ранжування даних із мережі Інтернет.

У роботі проведено порівняльний аналіз існуючих рішень для збору даних з мережі Інтернет та методи кластеризації даних. Було виділено їх переваги та недоліки.

На базі фреймворку Scrapy створено систему пошуку та збереження інформації з Web-сайтів новин. Зібрану інформацію проаналізовано за допомогою реалізації на мові програмування Python методу нечіткої кластеризації, що дозволило відфільтрувати та згрупувати новини відповідно до спільних характеристик.

Ключові слова: система збору даних, нечітка кластеризація, алгоритм *c*-середніх, Scrapy, Python.