

АНОТАЦІЯ

Дана робота присвячена рішення для управління потоком інформації з інтернет-ЗМІ шляхом збору та агрегування новин за їхнім змістом. Ресурс такого типу дозволить за лічені хвилини ознайомитись із всіма останніми актуальними подіями у світі.

Для досягнення даної мети було проведено аналіз існуючих підходів до побудови систем збору та агрегації, та алгоритми, що базуються на даних підходах, виявлено їхні недоліки.

Робота містить 4 розділи:

Розділ 1 присвячений аналізу актуальності даної роботи, огляду існуючих підходів та рішень до вирішення задачі. В ньому описані найпопулярніші алгоритми збору даних та оцінювання подібності тексту. Також наведено предметну область та постановку задачі.

У другому розділі наведено математичні основи алгоритмів, вибраних для реалізації програмного продукту, який розв'язуватиме дану проблему. Наведена інформація про задачу оцінювання подібності текстів в загальному випадку і в контексті вирішення задачі агрегації новин. Приводяться існуючі підходи до розв'язання даної задачі, розглянуті основні методи та дослідження. Описуються алгоритми, які будуть застосовані в даному дослідженні, та критерії якості роботи системи.

Розділ 3 детально описує архітектуру програмного продукту. У даному розділі зазначено обґрунтування вибору платформи, наведено аналіз вимог користувача до програмного продукту. Також даний розділ містить керівництво користувача для використання розробленої системи.

У розділі 4 наведений функціонально-вартісний аналіз програмного продукту.