Інформаційна технологія тематичного сортування текстової інформації

Федорко Андрій Петрович

Шепетівське територіальне відділення Малої академії наук України

Шепетівський НВК №1, 10 клас, м.Шепетівка

Колісецький Вілен Іванович, вчитель інформатики Шепетівського НВК №1

Мазурець Олександр Вікторович, старший викладач кафедри КНІТ Хмельницького національного університету

Метою роботи є розробка інформаційної технології тематичного сортування текстової інформації та програмного забезпечення для перевірки його ефективності при автоматизованому сортуванні новин по рубриках. При цьому використовуються розроблений підхід до визначення множин ключових слів для рубрик новин на основі методу оцінки TFIDF та розроблені математико-алгоритмічні моделі для визначення приналежності тестової новини до актуальних рубрик новин.

Автоматизація сортування текстової інформації є ефективним інструментом, що заощаджує час користувача та підвищує якість роботи новинних агрегаторів, тому даний напрям досліджень є актуальним.

На основі розробленої інформаційної технології тематичного сортування текстової інформації в середовищі .NET Framework з використанням мови програмування C# було створено два програмних продукти: систему визначення множин ключових слів для рубрик новин та систему тематичного сортування новин.

Одержані результати дослідження ефективності інформаційної технології показали, що в переважній більшості випадків програмна система, виконана відповідно до запропонованої інформаційної технології тематичного сортування текстової інформації, успішно виконала сортування новин за рубриками, й середня успішність сортування за рубриками склала 94,4%.