

УДК 004.4

Голуб Т.В.¹, Зеленьова І.Я.²

¹асист. НУ «Запорізька політехніка»

²канд. техн. наук, доц. НУ «Запорізька політехніка»

ПІДХІД ДО АПАРАТНОГО ПРИСКОРЕННЯ ПРОЦЕСУ КЛАСИФІКАЦІЇ ТЕКСТОВИХ ДОКУМЕНТІВ

Класифікація текстових документів в одну із заздалегідь визначених категорій є складним процесом. І, як будь-який складний процес, вимагає значних часових витрат на його реалізацію.

В загальному сенсі класифікація складається з двох значних етапів: визначення ознак категорій і безпосередньої класифікації певного тексту в одну з категорій.

Визначення ознак категорій виконується на основі опорної вибірки текстів, категорії приналежності яких визначені заздалегідь експертами. Даний процес супроводжується великим обсягом розрахунків і порівнянь, в результаті чого вимагає значних часових витрат. При цьому він виконується в більшості випадків один раз. Тому вдосконалення даного процесу за критерієм часу не є доцільним.

Безпосередня класифікація певного тексту в одну із заздалегідь визначених категорій виконується на основі порівняння ознак цього тексту з ознаками кожної категорії. Таке порівняння також вимагає значних часових витрат, пов'язаних з великою кількістю ознак. При цьому даний процес повторюється багаторазово для кожного документу, що аналізується. Тому його оптимізація за часом є актуальною.

На даний момент класифікатори текстових документів, на які покладено виконання другого етапу, реалізовано в більшості випадків в програмному варіанті. Це супроводжується складністю прискорення даного процесу, зумовленого підвищенням вимог до розрахункових потужностей обчислювальних приладів, що виконують відповідні функції.

Побудова ж класифікатора на основі програмно-апаратної реалізації дозволить використати внутрішні ресурси, які надають можливість паралельного виконання процесів без необхідності збільшення розрахункових потужностей приладу.

Програмна частина програмно-апаратної реалізації виконує етап формування простору ознак категорій, а на апаратну частину покладено етап класифікації тексту. При чому останній включає крок попередньої підготовки тексту, який полягає у виділенні ознак документу, що аналізується, і крок безпосередньої класифікації тексту.

Таким чином, перенесення обох кроків етапу класифікації на апаратну частину дозволяє створити автономний прилад для класифікації текстових документів, перелік ознак категорій якого формується програмною частиною.