

СКОРОЧЕННЯ ЧАСОВИХ ВИТРАТ НА ОБРОБКУ ІНФОРМАЦІЇ

Із збільшенням кількості інформації, що потребує обробки та аналізу, підвищуються вимоги до розрахункових можливостей апаратних ресурсів, на які покладено вирішення цієї задачі. При цьому очевидно, що збільшення обсягу вхідної інформації при сталих розрахункових ресурсах супроводжується зростанням часових витрат на даний процес.

При наявності обмежень збільшення можливостей апаратних ресурсів логічно виникають умови для необхідності вдосконалення певних етапів обробки інформації. При чому вдосконалення має проводитися з акцентом на скорочення часових витрат як безпосередньо, так і опосередковано шляхом зменшення кількості розрахунків. Одним із шляхів вдосконалення є також реалізація програмного та/або апаратного розпаралелювання даного процесу. Програмне розпаралелювання, реалізоване на базі розрахункової машини, вимагає перерозподілу з виділенням додаткових ресурсів на вирішення даної задачі. Апаратне розпаралелювання, при своєму незначному збільшенні початкових витрат на реалізацію цього рішення, дозволяє суттєво прискорити процес обробки вхідної інформації.

Прикладом обробки інформації з метою її подальшого аналізу є класифікація текстів, що представляє собою складний процес (рис.1). Одночасно з тим, що на кожному його етапі можна досягти певного скорочення часових витрат, перенесення частини розрахункового навантаження на апаратну платформу дозволяє досягти значних результатів. В якості апаратної платформи можна використати FPGA (Field-Programmable Gate Array) [1].



Рисунок 1 – UML-діаграма дій процесу класифікації текстів

Зокрема FPGA надає можливість паралельного виконання однотипних дій над різними вхідними даними. Апаратне розпаралелювання може бути використане як на етапі ідентифікації категорії, так і на кроках видалення

«стоп-слів» та стемінга слів. Це доцільно у випадку, коли зазначені етапи також реалізуються апаратно на тому ж кристалі FPGA.

UML-діаграма розгортання ресурсів класифікатору, реалізованого на зазначених принципах, наведена на рис. 2.

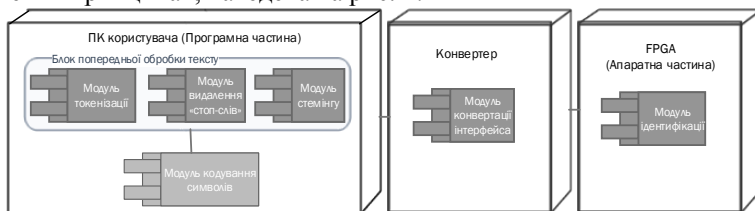


Рисунок 2 - UML-діаграма розгортання програмно-апаратного комплексу

Для даного класифікатору було розроблено різні засоби вдосконалення процесу на кожному із основних етапів. Використання окремо кожного із розроблених засобів на відповідному етапі класифікації текстів призводить прямо або опосередковано до скорочення часових витрат, а комплексне їх застосування дозволяє досягти ще більшого ефекту.

Результати застосування розроблених засобів наступні:

а) в рамках програмної частини:

1) в результаті модифікації алгоритму стемінга досягнуто зменшення часу на стемінг на 22%;

2) при фільтрації простору ознак категорій досягнуто зменшення зазначеного простору в середньому на 24%;

б) в рамках апаратної частини: паралельна організація схеми блоку ідентифікації текстів на основі апаратної реалізації забезпечила можливість прискорення процесу класифікації в n разів.

Комплексне використання засобів скорочення часових витрат на основі програмної та апаратної частин дозволяє досягти зменшення часу класифікації текстів на 32%.

Таким чином, на прикладі виконання класифікації текстів було показано, що використання комплексного підходу до обробки вхідної інформації із аналізом і вдосконаленням кожного з кроків, що потребують виконання, дозволяє досягти значного скорочення часових витрат певного складного процесу в цілому.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Програмно-апаратний спосіб прискорення процесу класифікації текстових документів / Т.В. Голуб, І.Я. Зеленьова, С.С. Грушко, М.А. Павлішин, А.О. Котенко // міжнар. наук.-практ. конф.: «Technical sciences: history, the present time, the future, EU experience», 27-28 september, 2019: тези доп., - Wlowlawek, Republic of Poland, 2019. – P. 90-93.