



## ПРОБЛЕМЫ РАСПАРАЛЛЕЛИВАНИЯ ПРОЦЕССОВ ПРИ ОПРЕДЕЛЕНИИ АВТОРСТВА ТЕКСТОВ

**Груздо И.В., Россоха С.В., Шостак И.В.**

*Национальный аэрокосмический университет*

*им. Н.Е. Жуковского «ХАИ»*

*г. Харьков, ул. Чкалова 17,*

*e-mail: tigratwovna@rambler.ru, sergey.rossokha@gmail.com,*

*Iv\_shostak@rambler.ru*

Современный этап развития сферы образования в Украине характеризуется, с одной стороны, стремительным ростом объемов письменных работ, а с другой – расширением их номенклатуры. Проверка этих работ нуждается в тщательном анализе со стороны преподавателей с целью установления оригинальности контента, а также в одновременном использовании при оценивании работ четырех разнотипных систем. Указанные обстоятельства порождают проблему, суть которой состоит в недостаточной эффективности существующих, «ручных» способов обработки преподавателями текстов письменных учебных работ (ПУР), установлении фактов наличия в них текстовых заимствований, и формировании на основании этих фактов объективных оценок работ в четырех разнотипных системах оценивания.

В докладе рассмотрена методика поиска заимствований в ПУР. Отличием описанной методики анализа текстовых документов является учет особенностей, присущих письменным учебным работам, что, в конечном итоге, приводит к повышению эффективности организации учебного процесса в заведениях различного уровня аккредитации, а так же выбор соответствующего метода анализа текстов в зависимости от типа ПУР.

На основе данной методики была разработана специальная ИТ анализа и оценивания письменных учебных работ с учетом наличия в них текстовых заимствований, а так же был рассмотрен процесс анализа ПУР в техническом университете, а именно, Национальном аэрокосмическом университете им. Н.Е.Жуковского «ХАИ». При этом автоматизация анализа ПУР была осуществлена с использованием ПС «Plagiarizm», который представляет собой исследовательский прототип, функционирующий на основе специально разработанной технологии.

Использование ПС «Plagiarizm» в учебном процессе дает следующий эффект: накопление опыта преподавателей и нормоконтролеров, а также для поддержки принятия решений ими по оцениванию письменных учебных работ. Кроме того, следует отметить, что применение ПС «Plagiarizm» на практике предоставляет возможность оценивать эффективность результатов, достигнутых в ходе исследования.

Для того, чтобы использовать преимущества ПС «Plagiarizm» необходимо выполнить преобразование последовательно выполняемых процессов программы в несколько параллельно взаимодействующих процессов. При этом необхо-



димо, что бы машина автоматически выполняла однозначное определение отдельных файлов для каждого обнаруженного позаимствованного объекта, так же связывала полученную информацию по каждой проанализированной работе, с соответствующим файлом источником, и производила анализ полученных данных. Для этого была выполнена формулировка задачи распараллеливания процессов при определении авторства текстов и рассмотрены существующие типы распараллеливания.

В работе были проанализированы существующие архитектуры многопоточных приложений:

- многопроцессные приложения с автономными процессами;
- многопроцессные приложения, взаимодействующие через трубы, сокеты и очереди System V IPC;
- многопроцессные приложения, взаимодействующие через разделяемую память;
- собственно многопоточные приложения;
- событийно-ориентированные приложения;
- гибридные архитектуры.

В ходе анализа был сделан вывод о том, что по своей направленности ПС «Plagiarizm» относится к событийно-ориентированным приложениям, а так же имеет сложно организованную структуру, когда в составе программы имеются и параллельные участки и циклические. При преобразовании ПС «Plagiarizm» в параллельную программу необходимо гарантировать, что все обработчики событий будут завершаться достаточно быстро.

Было отмечено, что важное место в преобразовании ПС «Plagiarizm» в параллельную программу занимают теоретические вопросы, включающие выбор модели распараллеливания, определение эквивалентности программ, определение степени параллелизма и синхронизация параллельных процессов, анализ потенциального параллелизма алгоритмов и программ, определение максимального параллелизма и др.

В докладе было приведено описание параллельного алгоритма поиска заимствований и определения авторства текстов, а также разработанная группа алгоритмов, позволяющих за приемлемое время получать распределения параллельных ветвей анализа текстов ПУР по процессорным ядрам. Первая группа алгоритмов относится к классу вероятностных алгоритмов локальной оптимизации, вторая основана на алгоритмах обхода графов и их разбиении.