

Веб-сервис для определения плагиата в LMS CATS

Попова Ю. Б.

Белорусский национальный технический университет

Для проверки документов на плагиат в сети Интернет существует несколько программных продуктов. Однако не все обучающиеся размещают свои работы во всемирной сети и часто передают их младшим курсам. Поэтому в учебных заведениях существует проблема плагиата работ с прежних лет. Одним из решений указанной проблемы является использование автоматизированных систем управления обучением (англ., Learning Management System, LMS). Обучающиеся закладывают свои лабораторные, контрольные и курсовые работы в такие системы, образуя архивы прошлых лет, и преподавателю не составляет труда проверить работы на плагиат. Ключевым моментом здесь является возможность LMS делать такие проверки. Автором предлагается веб-сервис, который может быть интегрирован в любую LMS, для поиска заимствованных работ и составления отчетов с указанием процентов схожести с аналогами. Он реализует модифицированный алгоритм векторной модели текстового документа, заключающийся в формировании отдельного N -списка для каждого анализируемого документа. Вследствие чего будет происходить попарное сравнение документов и формирование образа одного документа относительно N -списка другого. Таким образом, в i -й строку матрицы схожести будут записываться коэффициенты схожести рассматриваемых документов относительно i -го документа. Предлагаемая модификация позволяет ускорить процесс вычислений, поскольку нет необходимости искать общие термины для всех документов. Следует отметить, что в ходе учебного процесса преподаватель сталкивается с большим количеством работ студентов, которые необходимо проверять одновременно, например, контрольные или курсовые работы целого потока обучающихся. Поэтому выполнять такие проверки для каждой работы отдельно – это очень затратный процесс как по времени выполнения, так и по вычислительным ресурсам. Для решения данной проблемы предлагается использовать кластерный анализ текстовой информации, основная цель которого – разбить множество объектов на группы таким образом, чтобы объекты внутри одной группы были максимально похожи друг на друга, но в то же время максимально отличались от объектов другой группы. Алгоритм кластерного анализа текстовой информации реализован и интегрирован в LMS CATS (Care About The Students), разработанную и используемую на кафедре программного обеспечения вычислительной техники БНТУ.