горитмів підсистеми виробки рішень, які забезпечать необхідний рівень оперативності.

УДК 004.912

ТАНАСЕВСКИЙ В.В, КУРСИН А.И.

АВТОМАТИЧЕСКОЕ ВЫЯВЛЕНИЕ ПЛАГИАТА В ТЕКСТАХ

В настоящее время в связи с широким использованием интернета увеличилась вероятность появления плагиата в текстах. В связи с тем, что задача выявления плагиата не может быть эффективно выполнена вручную, актуальной является её автоматизация. В сети Интернет доступны сервисы проверки на плагиат (примеры - http://www.antiplagiat.ru/, http://www.copyscape.com/, http://istio.com/). Документ, проверяемый на наличие плагиата, загружается на веб-сервер, где он сверяется с базой данных документов ресурса или с документами, доступными в интернете. Недостатки таких систем – отсутствуют гарантии конфиденциальности документа, проверяемого на плагиат, и нет возможности использовать свою базу данных документов. Поэтому целесообразна разработка программы проверки текста на плагиат, работающей на локальном компьютере. Потенциальные пользователи данной программы – издательства, СМИ (проверка публикуемых материалов), учебные заведения (проверка студенческих работ). В данной работе была создана программа позволяющая создать базу данных, хранящую уже известные документы и проводить в ней поиск фрагментов текста из документов, проверяемых на плагиат. Предложения проверяемого текста ищутся в базе данных программы. Выявляются фрагменты, состоящие из нескольких предложений подряд, идущие в таком порядке в каком-либо из текстов базы данных программы. Для эффективного поиска плагиата необходимо хранить и обрабатывать большие объемы документов. Отсюда понятно, что основной проблемой при определении плагиата является задача поиска образца в больших массивах информации. Для ее решения в данной работе использованы B-деревья, так как скорость поиска в этих структурах с ростом количества данных замедляется незначительно.