

## ИСПОЛЬЗОВАНИЕ ТЕОРИИ КОНЕЧНЫХ АВТОМАТОВ ДЛЯ ФИЛЬТРАЦИИ КОНТЕНТА

В современных информационных системах все чаще возникает необходимость фильтрации контента. Данный процесс подразумевает широкий диапазон действия и может включать: установление родительского контроля, поиск и замену символов и словосочетаний в тексте, редактирование изобразительной информации (например, снижение количества оттенков для повышения эффективности алгоритмов сжатия), цензуру и другое.

Для обеспечения эффективной работы в сети *Internet* основной целью фильтрации контента является удаление ненормативной лексики из передаваемой информации. Для ее достижения поставлены следующие задачи:

- выбор абстрактной модели, используемой для предоставления и управления потоком выполнения команд в информационной системе;
- адаптация способов представления текстовой информации в информационных системах в соответствие выбранной модели;
- реализация соответствующего фильтра.

В качестве абстракции выбран конечный автомат, который является моделью вычислений, основанный на гипотетической машине состояний. При этом в один момент времени только одно из состояний может быть активно и для выполнения каких-либо действий машина изменяет свое состояние.

В данной работе в качестве способа представления текстовой информации выбрана последовательность байтов, которая соответствует стандарту *Unicode* и кодированию символов в соответствии с правилом *UTF-8*. Интерпретация последовательности байтов позволяет операционной системе рендерить глифы из шрифта, содержащего определенное множество символов на экране, т. е. отображать графическое представление символов.

Реализация фильтра осуществляется с помощью любого языка программирования, а сам фильтр должен быть компонентом программного обеспечения, которое или получает информацию по сети (например, браузер) или входит в состав сетевого экрана.

Достоинством работы с потоком байтов является его быстродействие, поскольку операция удаления определенной последовательности, по сути, является операцией незаписи или непередачи байтов.