

УДК 004

**ПОВЫШЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ КЛИЕНТ-СЕРВЕРНОГО  
ВЗАИМОДЕЙСТВИЯ ДЛЯ СЕРВИСА АВТОМАТИЗАЦИИ  
НОРМОКОНТРОЛЯ ДОКУМЕНТОВ**

**Филиппов Р.К. Университет ИТМО, Санкт-Петербург**

**Бауэр В.С. Университет ИТМО, Санкт-Петербург**

**Тартыньских П.С. Университет ИТМО, Санкт-Петербург**

**Научный руководитель – к.т.н. Горлушкина Н.Н. Университет ИТМО,  
Санкт-Петербург**

В рамках разработки сервиса автоматизированного нормоконтроля документов и обучения оформлению документации ведутся работы по созданию архитектуры клиент-серверного взаимодействия. Клиентом является плагин для MS Word, функциями которого передача на серверное приложение актуальной версии документа и получения инструкций для отображения вспомогательной информации пользователю. В упрощенном варианте архитектура представляет собой одну конечную точку на сервере, принимающую полностью весь документ и после анализа возвращающую JSON объект со списком всех комментариев, необходимых для отображения.

Во время работы над документом пользователь, увидев комментарии сервиса, предположительно начнет исправление допущенных ошибок. Неактуальные комментарии необходимо убирать, но для этого также желательно убедиться, что исправления не породили новых ошибок. Что в свою очередь требует повторного запроса на сервер с актуальной версией документа. Но так как исправления происходят в реальном времени, то количество запросов отрицательно влияет на производительность сервера и требует передачи большого объема информации средствами сети интернет. Особенно данная ситуация негативна при обработке больших документов, а также документов, содержащих изображения. Поэтому целью работы является модификация упрощенной архитектуры для повышения производительности клиент-серверного взаимодействия.

Во-первых, большой объем памяти в документе занимают изображения. Но в рамках разработки сервиса не предусмотрен анализ их содержания, из чего следует, что можно заметно сократить объем передаваемой информации заменив изображения картинкой подобного размера, но занимающей меньший объем памяти или же вовсе удалить их из состава документа, оставив код разметки в файле document.xml.

Во-вторых, нет необходимости при изменении нескольких строк в документе передавать его полностью. Для этого нужно обеспечить хранения документов пользователя и механизм отправки и замены на сервер только измененных параграфов и значений документа.

При создании такого механизма со стороны сервера была развернута база данных, содержащая документы пользователя и комментарии, получившиеся при последнем анализе. Когда сервер получает измененные данные или параграфы, он обновляет содержащийся в базе данных документ и анализирует результат. После анализа происходит сравнение последних отправленных на клиент комментариев и формируется разница между новым и старым набором. Вдобавок к новым комментариям, для исчезнувших комментариев на клиент будет отправлен их идентификатор с пометкой «удалить». И напоследок соответственно обновляется список актуальных комментариев в базе данных.

Со стороны клиента было необходимо хранить хеш каждого параграфа или любого другого элемента документа для обнаружения изменившихся данных. При получении сигнала о редактировании с использованием механизма debounce (с англ.: затухание) начинается проверка хешей, формирование списка изменившихся элементов и отправка их на сервер в формате JSON. В данном случае обновление возможно только строковых данных формата

xml. В случае изменения данных, не предусмотренных системой, будет отправляться конкретный файл, но опять же не весь архив docx.

При получении плагином инструкций комментариев, он удаляет помеченные в документе и добавляет новые. В случаях незначительного изменения комментария он все равно полностью удаляется и добавляется как новый.

Для проверки некоторых правил необходимо точное расположение элементов в документе и на его страницах. Не всегда эту информацию можно найти из xml документа. Поэтому приходится прибегать к нестандартным методам. Например, из плагина можно получить pdf файл и отправить на сервер. Но опять же при каждом изменении такая отправка будет нежелательной. В данном случае можно переводить pdf формат в изображения и отправлять только измененные страницы. А найдя подходящий алгоритм сжатия можно ограничить цветовую палитру до двух цветов и сократить размеры изображения в десятки раз. Для определения местоположения объектов документа нет необходимости в информации о цвете и прозрачности.

Таким образом удалось сократить количество передаваемой информации между клиентом и сервером в сервисе автоматизированного нормоконтроля документов путем оставления только необходимых для анализа данных и хранения копии документа на сервере, что позволило без издержек поддерживать актуальность выводимых пользователю комментариев.

Филиппов Р.К. (автор)

Бауэр В.С. (автор)

Тартынских П.С. (автор)

Горлушкина Н.Н. (научный руководитель)