

УДК 004.415

РЕАЛИЗАЦИЯ И ОПТИМИЗАЦИЯ АЛГОРИТМА ОБХОДА УЗЛОВ ДОКУМЕНТА ODT

Ларионова Г. С. (Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»)

Марцинкевич В. И. (Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»)

Терещенко В. В. (Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»)

Научный руководитель – ктн, доцент, Горлушкина Н.Н. (Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»)

Аннотация: В докладе приводится процесс реализации и оптимизации алгоритмов обхода узлов документа ODT.

Введение.

В процессе написания научных работ студенты и ученые допускают ошибки при оформлении. В результате содержание документов может не соответствовать ГОСТам или другим стандартам и нормативным документам. Для решения этих проблем разрабатывается система автоматизированного нормоконтроля документов, ключевой частью которого является парсер документов формата ODT [1]. Цель работы - реализация и оптимизация алгоритма обхода узлов документа.

Основная часть.

Алгоритм обхода узлов документа формата ODT требуется для получения характеристик объектов, которые могли бы указать на ошибки в оформлении. Большинство их хранится в стилях. Сложность состоит в том, что все стили состоят в иерархической структуре, в которой нужный атрибут может находиться на любом уровне. Для решения этой проблемы уже были написаны и представлены алгоритмы. Однако в ходе апробации алгоритмов было выявлено, что наследование есть не только у стилей, но и у структурных элементов, поэтому необходимая характеристика может оказаться в другом иерархическом дереве стилей [2,3].

В ходе решения этой проблемы реализовано два алгоритма, в основе которых лежит поиск в глубину. Для определения лучшего варианта был проведен сравнительный анализ, который не только показал сильные и слабые стороны представленного кода, но и обозначил общую проблему – время работы слишком велико. Поэтому применен ряд способов оптимизации алгоритмов.

Выводы.

Реализованный алгоритм обхода узлов документа формата ODT – необходимый элемент системы парсинга документов. После оптимизации он позволяет быстро и качественно определить структурные элементы и получить необходимые атрибуты для дальнейшего анализа системой нормоконтроля документов.

Список использованных источников:

1. Марцинкевич В. И., Ларионова Г. С., Терещенко В. В., Ситникова К. А. Анализ возможностей парсинга электронных текстовых документов для автоматизации

нормоконтроля // Экономика. Право. Инновации. - 2022. - Т. 1, №3. - С. 39-49.

2. Open Document Format for Office Applications (OpenDocument) Version 1.3. Part 3: OpenDocument Schema. — URL: <https://docs.oasis-open.org/office/OpenDocument/v1.3/os/part3-schema/OpenDocument-v1.3-os-part3-schema.html>.

3. Open Document Format for Office Applications (OpenDocument) Version 1.2. Part 1: OpenDocument Schema. 2010-06-10. — URL: <http://docs.oasis-open.org/office/v1.2/cd05/OpenDocument-v1.2-cd05-part1.html>.

Ларионова Г. С. (автор) Подпись

Марцинкевич В. И. (автор) Подпись

Терещенко В. В. (автор) Подпись

Горлушкина Н. Н. (научный руководитель) Подпись