Министерство науки и высшего образования Российской Федерации

федеральное государственное бюджетное образовательное учреждение высшего образования

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

МОРДОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

им. Н.П. ОГАРЁВА»

(ФГБОУ ВО «МГУ им. Н.П. Огарёва»)

Факультет математики и информационных технологий

Кафедра фундаментальной информатики

ОТЧЁТ

по производственной практике

преддипломная практика

студента 4 курса

Направление подготовки 02.03.02 Фундаментальная информатика и информационные технологии

Место прохождения практики: г. Саранск, ФГБОУ ВО «НИ МГУ им. Н.П. Огарева», факультет математики и информационных технологий, кафедра фундаментальной информатики, научно-исследовательская лаборатория №439

Срок прохождения практики: 15.05.2024 – 28.05.2024

Автор отчёта А. Е. Конышев

подпись, дата

Обозначение отчёта: ОП–02069964–П–02.03.02–08–24

Руководитель практики

от университета

к.ф.-м.н., доц. Н. М. Куляшова

подпись, дата

Отчёт защищён Оценка

дата

Саранск 2024

ЗАДАНИЕ

на производственную практику

преддипломная практика

Студенту Конышеву Артему Евгеньевичу

4 курса 402 группы

Направление подготовки 02.03.02 Фундаментальная информатика и информационные технологии

Профиль Информатика и компьютерные науки

Место прохождения практики: г. Саранск, ФГБОУ ВО «НИ МГУ им. Н.П. Огарева», факультет математики и информационных технологий, кафедра фундаментальной информатики, научно-исследовательская лаборатория №439

Срок прохождения практики: 15.05.2024 – 28.05.2024

Срок представления отчёта студента и отзыва руководителя практики от профильной организации на защиту

дата

1 Цели и задачи практики: Разработка алгоритма преобразования XML формата в стилизованный текст и его реализация

2 Компетенции студента, формируемые в результате прохождения практики: ПК-5, ПК-6, ПК-7, ПК-8, ПК-9, ПК-10

3 Задание на практику: разработать модель хранения данных, модель настроек отображения данных, разработать XML парсер для создания документов.

Руководитель практики

от университета

к.ф.-м.н., доц. Н. М. Куляшова

подпись, дата

Задание к исполнению принял А. Е. Конышев

подпись, дата

ДНЕВНИК

по производственной практике

преддипломная практика

Студента Конышева Артёма Евгеньевича

4 курса 402 группы

Направление подготовки 02.03.02 Фундаментальная информатика и информационные технологии

Профиль Информатика и компьютерные науки

ЗАПИСИ

о работах, выполненных в период практики

|  |  |  |
| --- | --- | --- |
| Дата | Содержание/Результаты работы | Замечания и подпись руководителя |
| 15.05.24 -18.05.24 | Первичное описание модели хранения данных |  |
| 18.05.24 -20.05.24 | Модификация модели хранения данных |  |
| 20.05.24 -22.05.24 | Формализация правил представления данных |  |
| 22.05.24 -27.05.24 | Разработка XML парсера |  |
| 27.05.24 | Создание отчета |  |

Руководитель практики

от университета

к.ф.-м.н., доц. Н. М. Куляшова

подпись, дата

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 5](#_Toc167830509)

[1 Описание модели. 7](#_Toc167830510)

[1.1 Правила написания текста 7](#_Toc167830511)

[1.1.1 Простой текст. Абзацы 8](#_Toc167830512)

[1.1.2 Заголовки 8](#_Toc167830513)

[1.1.3 Перечисления 9](#_Toc167830514)

[1.1.4 Код на языке программирования 9](#_Toc167830515)

[1.1.5 Формулы и различные математические объекты 10](#_Toc167830516)

[1.2 Вставка изображений 11](#_Toc167830517)

[1.3 Добавление таблиц 11](#_Toc167830518)

[1.4 Формализация требований 13](#_Toc167830519)

[2 Алгоритм преобразования XML документа 14](#_Toc167830520)

[ЗАКЛЮЧЕНИЕ 17](#_Toc167830521)

[СПИСОК ИСПОЛЬЗОВАНЫХ ИСТОЧНИКОВ 18](#_Toc167830522)

[ПРИЛОЖЕНИЕ А 19](#_Toc167830523)

ВВЕДЕНИЕ

Информация – неотъемлемая часть нашей жизни. Мы потребляем и обрабатываем ее из различных источников в различных видах и различными способами. Один из видов информации – текст. Он может быть получен из документов, книг, статей, журналов, интернет-сайтов. В цифровую эпоху важным элементом работы с текстом является не только его хранение и передача, но и представление.

В разных структурах и организациях есть свои нормативы для составления документов – стандарты. Они затрагивают любой элемент документа: от заголовков и содержания до таблиц и рисунков. Стандарты позволяют создавать документы в единообразном стиле, избегать коллизий при оформлении того или иного участка (если такое предусмотрено стандартом), а также позволяют описывать структуру документа.

Однако, следовать всем требованиям, накладываемыми нормативами, для людей, создающих документы впервые – задача трудоемкая. Можно запомнить основные правила составления заголовков, абзацев, рисунков, нумерации – этого будет достаточно для создания простого документа. Но могут возникнуть сложности при составлении таблиц, списков (особенно вложенных), математических формул, колонтитулов, списка источников. Также не простым будет процесс редактирования документа для соответствия другому стандарту. В связи с этим становится актуальным вопрос упрощения создания электронных документов, на которые наложены требования оформления, использовав определенный формат представления данных, а также формализация этих требований для обработки программами.

Целью выпускной квалификационной работы является разработка парсера, автоматизирующего процесс представления различных текстовых данных, которые могут быть статьей, отчетом, документом, литературным произведением, в требуемый вид, в зависимости от объекта, где этот текст будет размещен. Поставленная цель требует решения проблемы изменения стиля содержимого, вследствие непреднамеренного упущения какого-либо правила при составлении документа. Для достижения поставленной цели целесообразно использовать язык разметки XML, необходимый для хранения текстовых данных, и файл стилей по аналогии с файлом CSS для HTML.

Новизна данной работы обусловлена отсутствием на сегодняшний день простых способов создания и последующего изменения электронных документов.

1 Описание модели.

Для создания документов нам необходимо придерживаться правила разделения данных от представления. Контент будет содержаться в XML файле. Правила его отображения – в другом файле, либо в специальной секции того же файла.

Документ может содержать различные типы данных: текст, рисунки, таблицы, формулы. Необходимо описать, как эти данные должны быть отображены в «сыром» виде на языке разметки XML. Далее потребуется задокументировать правила, которые могут быть наложены на контент, например шрифт или выравнивание. Данный раздел по своей сути является документацией к оформлению содержимого и его представления.

Преимуществом размещения данных, используя XML, является простота встраивания в HTML разметку. Т. о. можно отобразить документ в сети Интернет без использования сторонних инструментов преобразования, хотя правила отображения все же потребуется зафиксировать. Эти настройки можно описать 2 способами: отдельно в css файле, или внутри тега <style>.

1.1 Правила написания текста

Любой текст можно представить как набор заголовков абзацев, и других структур специального вида. К таким структурам относятся:

* списки или перечисления,
* текст на языке программирования,
* формулы.

К каждому из этих видов применяются различные стили оформления, поэтому целесообразно размещать эти структуры в соответствующих им тегах. Правила их написания на языке разметки XML будут отражены в соответствующих пунктах.

1.1.1 Простой текст. Абзацы

Абзацем назовем такой текст, который состоит только из предложений, не содержащий остальных, перечисленных ранее структур. Всё содержимое абзаца размещается внутри тега <p>.

Однако внутри абзацев могут содержаться слова, различающиеся написанием от других слов. Например, слова могут быть выделены курсивом, иметь полужирное начертание и т.д. Для этого их нужно заключить в соответствующие теги:

* <b> для выделения полужирным,
* <i> для написания курсивом,
* <u> для подчёркивания,
* <s> для зачёркивания,
* <sup> для надстрочного написания,
* <sub> надстрочного написания.

1.1.2 Заголовки

Заголовки необходимо включать в тег <h>. Они могут различаются уровнем, поэтому тег <h> содержит необязательный атрибут lvl, принимающий целые значения, начиная с 1.

Пример, содержащий заголовки разных уровней:

<h lvl="1">

Заголовок уровня 1

</h>

<h lvl="2">

заголовок уровня 2

</h>

1.1.3 Перечисления

Разделяют два типа перечислений: нумерованные (упорядоченные) и ненумерованные. В некоторых случаях используется комбинация этих типов, когда элементом одного списка является другой список.

Нумерованный список следует размещать в теге <ol> (ordered list), ненумерованный размещать внутри тега <ul> (unordered list). Тег <ol> содержит атрибут type для обозначения вида перечисления: значение abc для буквенного или значение num для числового; а также атрибут sep отвечающий за разделитель между цифрой или буквой и элементом списка. Разделителем может быть точка, скобка или он может отсутствовать вовсе. Сами же объекты списка следует размещать внутри тега <li> (list item).

Пример нумерованного списка на языке разметки XML:

<ol type="num" sep="">

<li>[первый объект списка]</li>

<li>[второй объект списка]</li>

[…]

</ol>

Пример ненумерованного списка на языке разметки XML:

<ul>

<li>яблоки</li>

<li>бананы</li>

<li>груши</li>

</ul>

1.1.4 Код на языке программирования

Текст, написанный на каком-либо языке программирования или на языке текстовой разметки, можно разместить в документах в единообразном с обычным текстом виде. Но в этом случае теряется одна отличительная черта такого рода текста – его читаемость. Она предназначена для человека, читающего код, а не для компьютеров, которые этот код исполняют. Стиль оформления кода отличается от обычного текста следующими правилами:

* используются моноширинные шрифты, например Consolas или Courier New,
* текст выровнен по левому краю, но рекомендуется соблюдение табуляции для различных структур кода,
* междустрочный интервал обычно равен одной строке,
* если код содержит много символов, разумно использовать меньший размер шрифта.

Поэтому рациональным решением является размещение кода в отдельный тег – <code>. Этот тег так же целесообразно наделить атрибутом lang, который будет нести дополнительную информацию о языке, на котором написан текст, будь то язык программирования или язык разметки. Она может пригодиться для подсветки синтаксиса языка в различных редакторах.

Пример кода:

<code lang="js">

console.log(“Hello world”)

</code>

1.1.5 Формулы и различные математические объекты

Формулы и объекты из сферы математики так же характеризуются отличительным написанием от остального текста. Такого рода объекты размещаются внутри тега <math> в формате LaTeX. Поэтому для их размещения необходимо предварительно перевести формулу с помощью соответствующих программ или веб-сервисов, например latexeditor.lagrida.com.

Возьмём формулу нормального распределения, изображённую на рисунке 4, и запишем в формате LaTeX.

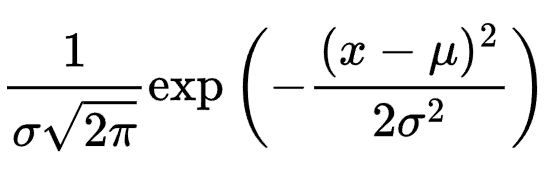


Рисунок 4 – Формула нормального распределения

В теге <math> будет содержать следующее выражение:

<math>

\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)

</math>

1.2 Вставка изображений

Так как XML файл — это текстовый файл и в него нельзя вставить изображения, то для их отображения в конечном документе необходимо использовать ссылки. Эти ссылки обозначают расположение изображений на устройстве.

Для вставки ссылки изображения используется конструкция вида:

<img src="" [desc=""] />

Атрибут src содержит ссылку на изображение, а атрибут desc, не являющийся обязательным, – описание изображения.

1.3 Добавление таблиц

Таблицы, написанные на языке разметки XML, отличаются тем, что данные организованы в виде дерева, а не в виде классических строк и столбцов. В XML каждая запись таблицы, которая в табличной форме соответствует строке, является отдельным тегом, содержащим вложенные элементы для каждой ячейки данных. В отличие от традиционных таблиц, где столбцы имеют фиксированное положение и идентификацию, в XML данные структурированы иерархически, что позволяет гибко добавлять и изменять элементы. Т. о., используя подход «любая таблица – набор строк», она отобразится в XML в следующем виде:

<table [desc=””]>

[ <th>

<td>[data]<td>

</th>]

<tr>

<td>[data]<td>

</tr>

…

</table>

Тег <table> содержит необязательный атрибут desc, служащий для описания таблицы. Внутри тега может сначала располагаться тег <th> (table header) необходимый для отображения заголовков таблицы. Но он может отсутствовать, в таком случае таблица целиком будет состоять из наборов тегов <tr> (table row). Внутри тегов <th> или <tr> располагаются теги <td>, внутри которых располагается информация.

Покажем, как таблица 1, содержащая некоторую информацию о студентах должна быть описана в виде XML дерева.

Таблица 1 – Информация о студентах

|  |  |  |  |
| --- | --- | --- | --- |
| Студент | Возраст | Группа | Оценка |
| Иванов И. | 20 | М01 | 4 |
| Петров П. | 19 | М02 | 5 |

Описание таблицы 1 в виде XML дерева:

<table desc=” Информация о студентах”>

<th>

<td>Студент</td>

<td>Возраст</td>

<td>Группа</td>

<td>Оценка</td>

</th>

<tr>

<td>Иванов И.</td>

<td>20</td>

<td>М01</td>

<td>4</td>

</tr>

<tr>

<td>Петров П.</td>

<td>19</td>

<td>М02</td>

<td>5</td>

</tr>

</table>

1.4 Формализация требований

Формализации подлежат правила, относящиеся к внешнему виду документа, чем его структурное содержание. Это объясняется тем фактом, внутреннее содержание документа, определяет его автор, тогда как отображение содержимого от автора не зависит.

Из описанной в разделах 1.1, 1.2, 1.3 модели требуется сформировать правила отображения структурных элементов. Данные правила будут описаны в виде xml дерева. Каждый элемент характеризуется своими свойствами. Например, текст имеет кегль (параметр), который может быть задан в разных единицах измерения: pt, px, em, cm и т.п. В общем случае правила описаны следующим образом:

<тег (элемент)>

<свойство1 параметр1="" [параметр1="" ]\* />

…

</тег>

XML файл, соответствующий нашей модели описан в приложении А.

2 Алгоритм преобразования XML документа

Для создания документа требуемого вида необходимо сначала проанализировать 2 файла: содержимого и представления. Прочитав программой XML файл содержимого, необходимо далее соотнести настройки его отображения, хранящиеся в файле представления. Сформированные таким образом данные позволят создавать документы любого удобного формата соответствующими инструментами.

Покажем на примере использования средств .Net Framework и Windows Forms программу, реализующую данный алгоритм.

Требуется сначала указать пути расположения XML документов. После программа проанализирует и сопоставит их содержимое. Первый файл имеет следующее содержимое:

<?xml version="1.0" encoding="UTF-8"?>

<document>

<h lvl="1">ВВЕДЕНИЕ</h>

<p>Информация – неотъемлемая часть нашей <b>жизни</b>...</p>

<ol type="num" sep=" ">

<li>[первый объект списка]</li>

<li>[второй объект списка]</li>

</ol>

<ul>

<li>яблоки</li>

<li>бананы</li>

<li>груши</li>

</ul>

<code lang="js">

console.log(“Hello world”)

</code>

<math>

\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)

</math>

<img src="cat.jpg" desc="котик"/>

<table desc="">

<th>

<td>студент</td>

<td>номер</td>

</th>

<tr>

<td>иванов</td>

<td>4</td>

</tr>

</table>

</document>

Файл настроек является модификацией файла из приложения А, добавлением в него параметров и их значений.

Внешний вид программы показан на рисунке 5.

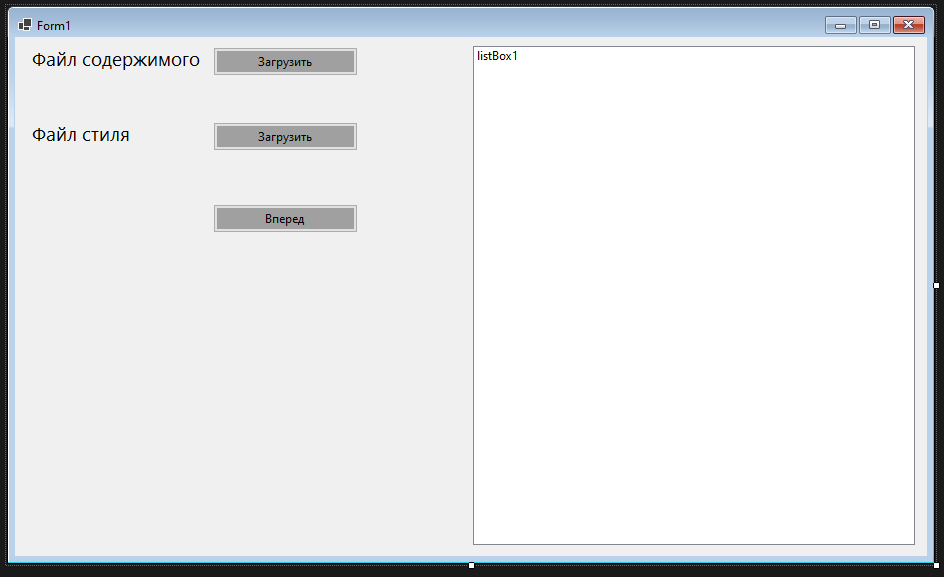


Рисунок 5 – Стартовое окно программы

После указания необходимых путей файлов результат работы программы будет отображено в правой части её окна. Продемонстрируем этот результат на рисунке 6.

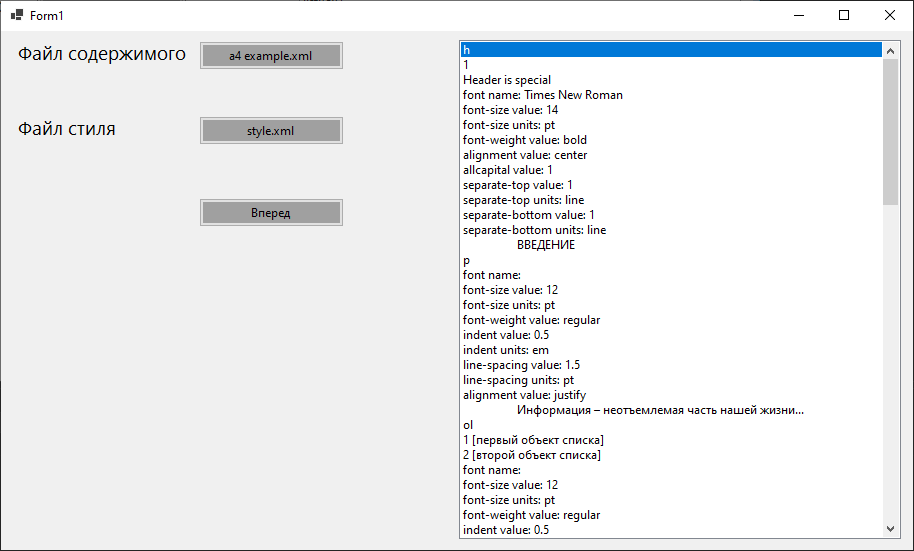


Рисунок 6 – Результат работы программы

Можем заметить, что сначала указывается тег, который соответствует типу объекта, далее значение атрибута если оно указано, следом перечисляются параметры и их значения. В конце указывается содержимое. Этот процесс повторяется для каждого тега. Используя полученный результат, мы можем создать документы, используя, например, средства C#, или созданием HTML файла с настройкой стиля.

ЗАКЛЮЧЕНИЕ

В ходе проделанной работы была описана модель, которая позволила отделить данные от их представления. Сформированы требования составления файла, содержащего данные, такие как: текст и различные структуры из него, таблицы, изображения. Также были формализованы правила отображения данных в документе с использованием языка разметки XML. Была написана программа - XML парсер, способная считать два файла и сформировать общую структуру содержимого.

СПИСОК ИСПОЛЬЗОВАНЫХ ИСТОЧНИКОВ

1. EPUB 3.3: сайт. – URL: https://www.w3.org/publishing/epub3/ (дата обращения 23.04.2024)
2. EPUB vulnerabilities: Electronic reading systems riddled with browser-like flaws: сайт. – URL: https://portswigger.net/daily-swig/epub-vulnerabilities-electronic-reading-systems-riddled-with-browser-like-flaws (дата обращения 23.04.2024)
3. ECMA-376 ECMA International: сайт. – URL: https://ecma-international.org/publications-and-standards/standards/ecma-376/ (дата обращения 23.04.2024)
4. Gribuser/FB3: сайт. – URL: https://github.com/gribuser/FB3 (дата обращения 23.04.2024)
5. Litres/FB3Reader: сайт. – URL: https://github.com/Litres/FB3Reader (дата обращения 23.04.2024)
6. Litres/FB3Editor: сайт. – URL: https://github.com/Litres/FB3Editor (дата обращения 23.04.2024)
7. Электронные книги и их форматы: DjVu – его история, плюсы, минусы и особенности: сайт. – URL: https://habr.com/ru/companies/maccentre/articles/411545/ (дата обращения 23.04.2024)

ПРИЛОЖЕНИЕ А

**Файл настроек представления содержимого**

<?xml version="1.0" encoding="UTF-8"?>

<style>

<p>

<!-- Используемый шрифт-->

<font name="" />

<!-- Размер шрифта -->

<font-size value="12" units="pt" />

<!-- Ширина шрифта-->

<font-weight value="1" units="pt" />

<!-- Абзацный отступ -->

<indent value="0.5" units="em" />

<!-- Расстояние между строками -->

<line-spacing value="1.5" units="pt" />

<!-- Выравнивание текста на странице-->

<alignment value="justify" />

</p>

<h>

<font name="" />

<font-size value="" units="" />

<indent value="" units="" />

<line-spacing value="" units="" />

<alignment value="" />

<!-- Отделение пустой строкой сверху-->

<separate-top value="1" units="line" />

<!-- Отделение пустой строкой снизу-->

<separate-bottom value="1" units="line" />

<!-- Специальные настройки -->

<special>

<!-- Список отдельных разделов-->

<names></names>

<font name="" />

<font-size value="" units="" />

<font-weight value="" units="" />

<alignment value="" />

</special>

</h>

<ol>

<indent value="" units="" />

<alignment value="" />

</ol>

<ul>

<indent value="" units="" />

<alignment value="" />

**Продолжение приложения А**

</ul>

<li>

<font name="" />

<font-size value="" units="" />

</li>

<code>

<font name="" />

<font-size value="" units="" />

<indent value="" units="" />

<line-spacing value="" units="" />

<alignment value="" />

<separate-top value="" units="" />

<separate-bottom value="" units="" />

</code>

<math>

<font name="" />

<font-size value="" units="" />

<indent value="" units="" />

<line-spacing value="" units="" />

<alignment value="" />

<separate-top value="" units="" />

<separate-bottom value="" units="" />

</math>

<img>

<alignment value="" />

<width value="" units="" />

<height value="" units="" />

<separate-top value="" units="" />

<separate-bottom value="" units="" />

</img>

<table>

<indent value="" units="" />

<separate-top value="" units="" />

<separate-bottom value="" units="" />

<alignment value="" />

</table>

<th>

<font name="" />

<font-size value="" units="" />

<alignment value="" />

</th>

<td>

<font name="" />

**Окончание приложения А**

<font-size value="" units="" />

<alignment value="" />

</td>

</style>

АНКЕТА

производственная практика

преддипломная практика

1 Удовлетворены ли Вы местом прохождения практики? (один ответ)

☑ да ⬜ нет

2 Удовлетворены ли Вы качеством разработки методических указаний, содержащихся в программе практики? (один ответ)

* в методических указаниях недостаточно информации для составления отчета
* методические указания не подходят к профильной организации
* не ознакомлен с программой практики

☑ методические указания мне понятны

3 Считаете ли Вы достаточными для выполнения работ, предусмотренных программой практики, те теоретические знания, которые Вы получили в Университете? (один ответ)

☑ да

* скорее да, чем нет
* скорее нет, чем да
* нет

4 Дала ли производственная практика возможность применить и развить навыки, необходимые для дальнейшей профессиональной деятельности (в соответствии с Вашим направлением подготовки/Вашей специальностью)? (один ответ)

☑ да

* скорее да, чем нет
* скорее нет, чем да
* нет

5 Хотели бы Вы в дальнейшем продолжить свою трудовую деятельность в организации, в которой проходили практику? (один ответ)

☑ да

* нет
* со мной заключили договор о дальнейшем трудоустройстве

6 Как Вы оцениваете итоги практики с точки зрения ее результативности (один ответ)

☑ на практике я еще больше убедился в правильности выбора профессии

* практика разочаровала меня в выбранной профессии
* практика обнаружила пробелы в моей специальной подготовке
* практика носила формальный характер

7 Оцените степень удовлетворенности местом прохождения практики (материально-техническая оснащенность, кадровый состав) по пятибалльной шкале (1 – очень плохо, 5 – отлично) 5 – отлично

8 Оцените степень удовлетворенности взаимоотношениями с руководителем практики от Университета по пятибалльной шкале (1 – очень плохо, 5 – отлично) 5 – отлично

9 Оцените удовлетворенность в целом условиями прохождения практики (один ответ)

☑ полностью удовлетворен

* удовлетворен частично
* полностью не удовлетворен

10 Ваши предложения по организации практики

Предложений нет

Министерство науки и высшего образования Российской Федерации

федеральное государственное бюджетное образовательное учреждение высшего образования

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

МОРДОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

им. Н.П. ОГАРЁВА»

(ФГБОУ ВО «МГУ им. Н.П. Огарёва»)

ОТЧЁТ

по производственной практике

преддипломная практика

Срок прохождения практики: 15.05.2024 – 28.05.2024

|  |  |  |
| --- | --- | --- |
| Фамилия Имя Отчество студента | Населенный пункт/профильная организация/структурное подразделение Университета | Оценка |
| Конышев Артем Евгеньевич | г. Саранск, ФГБОУ ВО «НИ МГУ им. Н.П. Огарева», факультет математики и информационных технологий, кафедра фундаментальной информатики, научно-исследовательская лаборатория №439 |  |

1 Цели и задачи практики: Разработка алгоритма преобразования XML формата в стилизованный текст

2 Компетенции студента, формируемые в результате прохождения практики: ПК-5, ПК-6, ПК-7, ПК-8, ПК-9, ПК-10

3 Анализ анкеты студентов–практиканта показал, что: студент полностью удовлетворен условиями прохождения практики. Предложений по совершенствованию процессов организации и прохождения практики от студента не поступило.

Руководитель практики

от университета

к.ф.-м.н., доц. Н. М. Куляшова

подпись, дата