**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ**

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук

Образовательная программа бакалавриата «Программная инженерия»

|  |  |  |
| --- | --- | --- |
| СОГЛАСОВАНО  Научный руководитель,  Заместитель декана по учебно-методической работе факультета компьютерных наук  доцент департамента [больших данных и информационного поиска](https://cs.hse.ru/big-data/), канд. социологических. наук  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ И.Ю. Самоненко  «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 2022 г. |  | УТВЕРЖДАЮ  Академический руководитель образовательной программы «Программная инженерия»  профессор департамента программной инженерии, канд. техн. наук  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.В. Шилов  «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 2022 г. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | ***Подп. и дата*** |  | | ***Инв. № дубл.*** |  | | ***Взам. инв. №*** |  | | ***Подп. и дата*** |  | | ***Инв. № подл*** |  | | **Приложение для визуализации алгоритма Фараха**  **Техническое задание**  **ЛИСТ УТВЕРЖДЕНИЯ**  **RU.17701729.10.03-01 ТЗ 01-1-ЛУ** | | |
|  |  | |
| Исполнитель:  студент группы БПИ206  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ / Вавилов Г.В /  «\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2022 г. | |
|  | | |
|  | |  |

**Москва 2022**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| УТВЕРЖДЕН  RU.17701729.10.03-01 ТЗ 01-1-ЛУ |  | |  | |
| |  |  | | --- | --- | | ***Подп. и дата*** |  | | ***Инв. № дубл.*** |  | | ***Взам. инв. №*** |  | | ***Подп. и дата*** |  | | ***Инв. № подл*** |  | | **Приложение для визуализации алгоритма Фараха**  **Техническое задание**  **RU.17701729.10.03-01 ТЗ 01-1**  **Листов \_\_** | | | | |
|  | |  | | |
|  | | |
|  | | | | |
|  | | | |  |

**Москва 2022**

**Содержание**

[Аннотация 4](#_Toc58433075)

[1. Введение 5](#_Toc58433076)

[1.1. Наименование 5](#_Toc58433077)

[1.2. Краткая характеристика и область назначения: 5](#_Toc58433078)

[2. Основания для разработки 6](#_Toc58433079)

[2.1. Документ, на основании которого ведется разработка 6](#_Toc58433080)

[2.2. Наименование темы разработки 6](#_Toc58433081)

[3. Назначение разработки 7](#_Toc58433082)

[3.1. Функциональное назначение 7](#_Toc58433083)

[3.2. Эксплуатационное назначение 7](#_Toc58433084)

[4. Требования к программе 8](#_Toc58433085)

[4.1. Требования к функциональным характеристикам 8](#_Toc58433086)

[4.2. Требования к надежности. 8](#_Toc58433087)

[4.3. Условия эксплуатации. 8](#_Toc58433088)

[4.4. Требования к составу и параметрам технических средств. 8](#_Toc58433089)

[5. Требования к программной документации 10](#_Toc58433090)

[5.1. Предварительный состав программной документации 10](#_Toc58433091)

[6. Технико-экономические показатели 11](#_Toc58433092)

[6.1. Предполагаемая потребность 11](#_Toc58433093)

[6.2. Ориентировочная экономическая эффективность 11](#_Toc58433094)

[6.3. Экономические преимущества разработки по сравнению с отечественными и зарубежными аналогами. 11](#_Toc58433095)

[7. Стадии и этапы разработки 12](#_Toc58433096)

[7.1. Техническое задание 12](#_Toc58433097)

[7.1.1. Разработка технического задания. 12](#_Toc58433098)

[7.1.2. Согласование технического задания. 12](#_Toc58433099)

[7.1.3. Утверждение технического задания. 12](#_Toc58433100)

[7.2. Эскизный проект 12](#_Toc58433101)

[7.3. Технический проект 12](#_Toc58433102)

[7.4. Рабочий проект 12](#_Toc58433103)

[7.5. Внедрение 12](#_Toc58433104)

[8. Порядок контроля и приемки 13](#_Toc58433105)

[8.1. Виды испытаний. 13](#_Toc58433106)

[8.2. Общие требования к приемке работы. 13](#_Toc58433107)

[Терминология 14](#_Toc58433108)

[1. Web-форум 14](#_Toc58433109)

[2. Парсер 14](#_Toc58433110)

[3. Javascript 14](#_Toc58433111)

[4. HTML 14](#_Toc58433112)

[5. Electron.js 14](#_Toc58433113)

[6. JSON 14](#_Toc58433114)

[Список используемой литературы 15](#_Toc58433115)

[Лист регистрации изменений 16](#_Toc58433116)

# Аннотация

Данное Техническое задание по разработке приложения для сбора информации из сайтов содержит следующие разделы: «Введение», «Основание для разработки», «Назначение разработки», «Требования к программе», «Требования к программным документам», «Технико-экономические показатели», «Стадии и этапы разработки», «Порядок контроля и приемки» и приложения.

В разделе «Введение» указано наименование и краткая характеристика области применения программы.

В разделе «Основания для разработки» указан документ, на основании которого ведется разработка и наименование темы разработки.

В разделе «Назначение разработки» указано функциональное и эксплуатационное назначение программного продукта.

Раздел «Требования к программе» содержит основные требования к функциональным характеристикам, к надежности, к условиям эксплуатации, к составу и параметрам технических средств, к информационной и программной совместимости, к маркировке и упаковке, к транспортировке и хранению.

Раздел «Требования к программным документам» содержит предварительный состав программной документации и специальные требования к ней.

В разделе «Ожидаемые технико-экономические показатели» указана предполагаемая потребность и экономические преимущества разрабoтки по сравнению с отечественными и зарубежными образцами или аналогами.

Раздел «Стадии и этапы разработки» содержит стадии разработки, этапы и содержание работ.

В разделе «Порядок контроля и приемки» указаны общие требования к приемке работы.

Настоящий документ разработан в соответствии с требованиями:

ГОСТ 19.101-77 Виды программ и программных документов [2];

ГОСТ 19.102-77 Стадии разработки [1];

ГОСТ 19.103-77 Обозначения программ и программных документов [3];

ГОСТ 19.104-78 Основные надписи [4];

ГОСТ 19.105-78 Общие требования к программным документам [5];

ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом [6];

ГОСТ 19.404-79 Пояснительная записка. Требования к содержанию и оформлению [7].

# Введение

* 1. Наименование**:** Приложение для визуализации алгоритма Фараха.
  2. Характеристика и область назначения:

Алгоритм Фараха используется для построения сжатого суффиксного дерева. Алгоритм интересен тем, что имеет линейную временную сложность и может работать с бесконечными алфавитами.

Алгоритм очень объемный и сложный для понимания. Приложение для визуализации должно помочь пользователю с формированием представления о работе алгоритма и используемых в нем структурах данных.

Образование – основная область применения.

# Основания для разработки

## Документ, на основании которого ведется разработка

Программа выполнена в рамках курсового проекта в соответствии с учебным планом подготовки бакалавров (НИУ ВШЭ, факультет компьютерных наук) по направлению «Программная инженерия». Приказ декана факультета компьютерных наук Национального исследовательского университета «Высшая школа экономики» № \_\_\_\_\_\_\_\_\_\_\_\_ от \_\_\_\_\_\_\_ "Об утверждении тем, руководителей курсовых работ студентов образовательной программы Программная инженерия факультета компьютерных наук".

## Наименование темы разработки

* + 1. **Название темы разработки на русском языке:** Приложение для визуализации алгоритма Фараха.
    2. **Название темы разработки на английском языке:** Farach algorithm visualization application.

# Назначение разработки

## Функциональное назначение

Приложение наглядно показывает, как должно строиться сжатое суффиксное дерево для введенной пользователем строки символов при помощи алгоритма Фараха. Приложение визуализирует каждый этап алгоритма и объясняет выполненные операции.

## Эксплуатационное назначение

Приложение позволяет детально и наглядно рассмотреть каждый этап работы алгоритма с целью его понимания. Также визуализация должна помочь при реализации алгоритма.

# Требования к программе

## **Требования к функциональным характеристикам**

Приложение состоит только из клиентской части, которая должна быть реализована в виде веб-приложения, запускаемого в браузере.

В приложении должно быть реализовано следующее:

1. Поле ввода для строки, по которой будет построено суффиксное дерево.
2. Поэтапная визуализация алгоритма Фараха
   1. Разбиение строки на пары символов и дальнейшее сжатие.
      1. Манипуляции с символами и парами символов показываются при помощи перемещения текстовых элементов.
   2. Построение четного суффиксного дерева.
      1. Дерево состоит из вершин и ребер. Вершина показывается в виде круга. Ребро показывается в виде отрезка, соединяющего вершины. Ребро подписывается текстом, который оно обозначает.
   3. Построение нечетного суффиксного дерева.
      1. Требования к дереву аналогичны требованиям в пункте “ Построение четного суффиксного дерева”.
      2. Суффиксный массив представляется в виде таблицы, содержащей суффиксы и их индексы в строке.
      3. Построение суффиксного дерева по суффиксному массиву (Алгоритм Карккайнена-Сандерса) визуализировать не требуется.
   4. Слияние деревьев.
      1. Требования к дереву аналогичны требованиям в пункте “ Построение четного суффиксного дерева”.
      2. Ребра четного дерева имеют цвет, отличный от цвета ребер нечетного дерева.
   5. Удаление двойных ребер.
3. Возможность остановить визуализацию.
4. Возможность перейти к следующему кадру визуализации.
5. Возможность перейти к предыдущему кадру визуализации.

## Требования к надежности.

* + 1. Должна быть предусмотрена корректная обработка ошибок, связанных с некорректными действиями пользователя. В случае некорректных входных данных программа не должна завершаться аварийно или уходить в бесконечный цикл. Пользователь должен быть уведомлен о некорректных данных.
    2. Во время загрузки программы должно быть обеспечено стабильное соединение с сетью интернет.
    3. После загрузки программа должна иметь возможность корректной работы без соединения с сетью интернет.

## Условия эксплуатации.

* + 1. **Климатические условия эксплуатации.**

Требования к климатическим условиям эксплуатации не предъявляются.

* + 1. **Требования к видам обслуживания**

Обслуживание не требуется.

* + 1. **Требования к численности и квалификации персонала.**

До работы с программой допускается один оператор и любое количество наблюдателей. Оператор должен владеть базовыми навыками работы с компьютером и иметь опыт использования веб-браузера.

## Требования к составу и параметрам технических средств.

Для работы программы необходимы следующие компоненты:

* + - Процессор: Intel Pentium 4 / Athlon 64 или более поздней версии с поддержкой SSE2
    - Операционная система: Windows 7 / Windows 8 / Windows 10 / Windows 11 / MacOS X 11 и выше / Linux
    - Свободное место на жёстком диске: 256 МБ.
    - Оперативная память: 2ГБ.
    - Доступ в интернет.
    - Клавиатура.
    - Мышь или заменяющее устройство ввода.

Требования обусловлены необходимостью использования веб-браузера для работы программы.

* 1. **Требования к информационной и программной совместимости.**
     1. **Требования к информационным структурам на входе.**

Входные данные – строка символов, по которой требуется построить суффиксное дерево.

* + 1. **Требования к информационным структурам на выходе.**

В процессе работы программы должен быть визуализирован каждый этап алгоритма. Результат работы алгоритма – построенное сжатое суффиксное дерево. Данные и структуры данных, использованные во время работы алгоритма, должны быть представлены пользователю в удобном для восприятия виде.

Узлы дерева могут быть представлены в виде окружностей, ребра дерева в виде отрезков, соединяющих вершины.

* + 1. **Требования к методам решения.**

Среда разработки: Visual Studio Code.

Используемые библиотеки и технологии: платформа Node.js, TypeScript, Babel, Jest.

* + 1. **Требования к исходным кодам и языкам программирования.**

Исходный код программы должен быть написан на языках программирования HTML, CSS и TypeScript с возможностью транспиляции в JavaScript. Допускается использование библиотек и фреймворков.

* + 1. **Требования к программным средствам, используемым программой.**

Для работы приложения необходима установленная операционная система из п. 4.4 “Требования к составу и параметрам технических средств”.

Также необходим веб-браузер Firefox > 96.0.0 / Chrome > 97.0.0 или их аналоги на движках Gecko/Blink.

* 1. **Требования к маркировке и упаковке.**

Программа состоит из трех частей:

1. .html файл с разметкой страницы.
2. .css файл со стилями страницы.
3. .js файл – результат транспиляции и сборки исходного кода.

Программа должна храниться на выделенном сервере или на устройстве пользователя.

* 1. **Требования к транспортированию и хранению.**

Программа запрашивается пользователем через веб-браузер при помощи http-запросов на удаленный сервер. После загрузки программы она хранится на жестком диске пользователя. Транспортирование и хранение программы выполняет веб-браузер, поэтому пользователю не требуется

* 1. **Специальные требования**
     1. Специальные требования к программе не предъявляются

# Требования к программной документации

## Предварительный состав программной документации

* + 1. Техническое задание «Методы сбора информации из Web-форумов» (ГОСТ 19.201-78) [8]
    2. Программа и методика испытаний «Методы сбора информации из Web-форумов» (ГОСТ 19.301-78) [12]
    3. Пояснительная записка «Методы сбора информации из Web-форумов» (ГОСТ 19.404-79) [7]
    4. Руководство оператора «Методы сбора информации из Web-форумов» (ГОСТ 19.505-79) [13]
    5. Руководство программиста «Методы сбора информации из Web-форумов» (ГОСТ 19.504-79) [14]
    6. Текст программы «Методы сбора информации из Web-форумов» (ГОСТ 19.401-78) [15]

# Технико-экономические показатели

## Предполагаемая потребность

Программа будет использоваться специалистами сферы анализа данных и обычными пользователями. Также можно использовать программу в качестве инструмента для поиска уязвимостей на сайтах.

## Ориентировочная экономическая эффективность

Программа позволяет автоматизировать рутинные работы, на которые требуется большое количество времени. Это позволяет сократить затраченные на работу деньги и время.

## Экономические преимущества разработки по сравнению с отечественными и зарубежными аналогами.

Существует огромное количество аналогов данного программного продукта.

Среди них:

*Import.io*

*Webhose.io*

*Dexi.io*

*Scrapinghub*

*ParseHub*

*VisualScraper*

*Spinn3r*

*80legs*

*Scraper*

*OutWit Hub*

Преимущества программы:

* + 1. Бесплатная
    2. С открытым исходным кодом
    3. Модульная

# Стадии и этапы разработки

## Техническое задание

## Разработка технического задания.

## Согласование технического задания.

## Утверждение технического задания.

## Эскизный проект

* + 1. Разработка алгоритмa рабoты программы.
    2. Разработка эскиза структуры программы.
    3. Разработка макета для графического интерфейса.

## Технический проект

* + 1. Разработка алгоритмов для решения конкретных задач.
    2. Разработка структуры программы.
    3. Утверждение технического проекта.

## Рабочий проект

* + 1. Программирование.
    2. Отладка.
    3. Разработка программной документации в соответствии с требованиями ГОСТ 19.101-77.
    4. Испытание программы.
    5. Корректировка программы по результатам испытания.

## Внедрение

* + 1. Подготовка и передача программы заказчику.

# Порядок контроля и приемки

## Виды испытаний.

Тестируется правильность выполнения программой выполняемого ей функционала.

Тестирование осуществляется в соответствии с документом «Программа и методика испытаний» (ГОСТ 19.301-79) [11].

## Общие требования к приемке работы.

Работа принимается в случае полной работоспособности программы и успешного прохождения тестирования.

**Приложение 1**

# Терминология

## Web-форум

Сайт в сети интернет, где пользователи могут задавать тематические вопросы и отвечать на них

## Парсер

Программа, собирающая определенные данные со страницы сайта.

## Javascript

Интерпретируемый язык программирования.

## HTML

Язык гипертекстовой разметки для создания графических интерфейсов.

## Electron.js

Фреймворк для создания кроссплатформенных приложений для персональных компьютеров на базе языка программирования JavaScript.

## JSON

Текстовый формат обмена данными, основанный на JavaScript.

**Приложение 2**

# Список используемой литературы

1. ГОСТ 19.102-77 Стадии разработки. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.101-77 Виды программ и программных документов. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.103-77 Обозначения программ и программных документов. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.104-78 Основные надписи. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.105-78 Общие требования к программным документам. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.404-79 Пояснительная записка. Требования к содержанию и оформлению. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.201-78 Техническое задание. Требования к содержанию и оформлению. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.603-78 Общие правила внесения изменений. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. ГОСТ 19.604-78 Правила внесения изменений в программные документы, выполненные печатным способом. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
11. ГОСТ 19.602-78 Правила дублирования, учета и хранения программных документов, выполненных печатным способом. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
12. ГОСТ 19.301-79 Программа и методика испытаний. Требования к содержанию и оформлению. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
13. ГОСТ 19.505-79 Руководство оператора. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
14. ГОСТ 19.504-79 Руководство программиста». Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
15. ГОСТ 19.401-78 Текст программы «Методы сбора информации из Web-форумов». Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
16. **ГОСТ Р 59212-2020 «Сыры для пиццы термизированные. Технические условия»**. Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Лист регистрации изменений | | | | | | | | | |
| Номера листов (страниц) | | | | | Всего листов (страниц в докум.) | № документа | Входящий № сопроводительного докум. и дата | Подп. | Дата |
| Изм. | Измененных | Замененных | Новых | Аннулированх |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

# Лист регистрации изменений