



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«Московский государственный технический университет имени  
Н. Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н. Э. Баумана)

---

ФАКУЛЬТЕТ «Информатика и системы управления»  
КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

## Отчет по лабораторной работе №5 по дисциплине «Анализ Алгоритмов»

Тема Организация параллельных вычислений по конвейерному принципу

Студент Смирнов И.В.

Группа ИУ7-52Б

Преподаватель Волкова Л. Л., Строганов Д.В.

2024 г.

# СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Входные и выходные данные	4
2 Преобразование данных	4
3 Пример работы программы	5
4 Тестирование	7
5 Описание исследования	8
ЗАКЛЮЧЕНИЕ	11
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	12

## ВВЕДЕНИЕ

В современном программировании эффективная обработка данных играет ключевую роль в повышении производительности приложений. Одним из подходов к оптимизации выполнения задач является конвейерная обработка (*pipeline processing*), которая позволяет разделить процесс обработки данных на последовательные стадии. Каждая стадия выполняется независимо, что обеспечивает параллельную обработку и более эффективное использование системных ресурсов.

**Цель лабораторной работы** — получение навыка организации параллельных вычислений по конвейерному принципу. Для достижения поставленной цели необходимо выполнить следующие задачи:

- описать входные, выходные данные, а также преобразования входных данных в выходные;
- реализовать алгоритм для парсинга рецептов с веб-сайта с использованием конвейерной обработки и многопоточности на основе нативных потоков;
- протестировать разработанный алгоритм по методологии черного ящика;
- описать пример работы программы на конкретном случае;
- сделать выводы, основываясь на информации в полученных лог-файлах.

# 1 Входные и выходные данные

Входные данные: базовый *URL*—адрес веб-сайта, с которого будут загружаться страницы с рецептами; количество страниц для обработки рецептов.

Выходные данные: база данных *SQLite*, содержащая информацию о каждом обработанном рецепте (*ID*, *IssueID*, *URL*, название, ингредиенты, шаги, *URL* изображения); лог-файлы, фиксирующие процесс выполнения программы и возникающие ошибки.

## 2 Преобразование данных

По нажатии кнопки «Начать парсинг» программа считывает базовый *URL*—адрес, а также количество страниц из полей ввода. Далее программа выделяет из всех страниц рецепты и записывает их в файл `inputfile.txt`, который подается алгоритму на вход. Алгоритм организован в виде конвейера из трёх стадий, каждая из которых выполняется в отдельном потоке:

### 1. Стадия 1: Загрузка *HTML*

- Чтение ссылок на рецепты из файла `inputfile.txt`;
- Загрузка *HTML*-контента каждой страницы с использованием библиотеки `HtmlAgilityPack`.

### 2. Стадия 2: Парсинг *HTML*

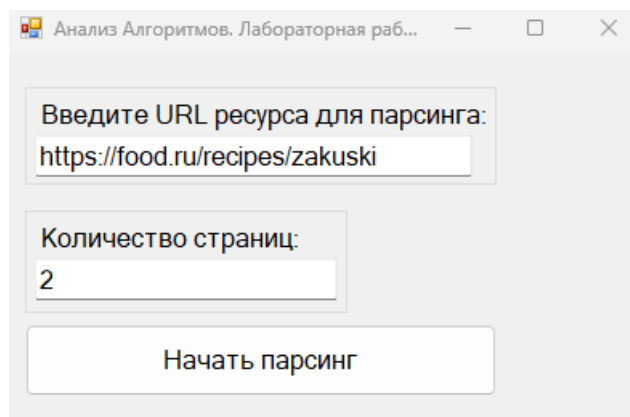
- Извлечение заголовка рецепта;
- Извлечение *URL* изображения рецепта на основе заголовка;
- Извлечение ингредиентов и шагов приготовления.

### 3. Стадия 3: Запись в базу данных *SQLite*

- Запись обработанных данных в базу данных `recipes.db`.

### 3 Пример работы программы

На рисунках 3.1 — 3.6 представлен пример работы программы. В данном случае программа обрабатывает несколько страниц сайта, извлекает ссылки на рецепты, парсит информацию о каждом рецепте и сохраняет её в базу данных *SQLite*. Сообщения о ходе выполнения и возможных ошибках фиксируются в соответствующих лог-файлах.



The screenshot shows a window titled "Анализ Алгоритмов. Лабораторная раб...". Inside, there is a label "Введите URL ресурса для парсинга:" followed by a text input field containing "https://food.ru/recipes/zakuski". Below this is another label "Количество страниц:" followed by a text input field containing "2". At the bottom, there is a large button labeled "Начать парсинг".

Рисунок 3.1 – Ввод данных в интерфейс приложения

```
2024-10-31 15:41:37.642 - Найдена ссылка на рецепт: https://food.ru/recipes/226447-bento-boks-s-sendvichem-i-salatom
2024-10-31 15:41:37.642 - Найдена ссылка на рецепт: https://food.ru/recipes/228002-tvorozhnye-percytvorozhnye-percy
2024-10-31 15:41:37.643 - Найдена ссылка на рецепт: https://food.ru/recipes/227850-opjata-po-koreiski
2024-10-31 15:41:37.643 - Найдена ссылка на рецепт: https://food.ru/recipes/227933-marinovannaja-kapusta-s-lukom
2024-10-31 15:41:37.643 - Найдена ссылка на рецепт: https://food.ru/recipes/227366-sendvich-s-arbuzom
2024-10-31 15:41:37.644 - Найдена ссылка на рецепт: https://food.ru/recipes/227364-briusselskaja-kapusta-na-zimu
2024-10-31 15:41:37.644 - Найдена ссылка на рецепт: https://food.ru/recipes/227351-pomidory-konfi
2024-10-31 15:41:37.644 - Найдена ссылка на рецепт: https://food.ru/recipes/227347-maslo-s-petrushkoi-i-chesnokom
2024-10-31 15:41:37.644 - Всего собрано 20 уникальных ссылок на рецепты.
2024-10-31 15:41:37.645 - Страница 2 обработана, собрано ссылок: 20 из 20
2024-10-31 15:41:37.645 - Страница 2 обработана, собрано ссылок: 20
2024-10-31 15:41:37.646 - Файл inputfile.txt с ссылками успешно создан.
2024-10-31 15:41:37.646 - Поток стадии 1 запущен.
2024-10-31 15:41:37.649 - Поток стадии 2 запущен.
2024-10-31 15:41:37.652 - Поток стадии 3 запущен.
2024-10-31 15:41:37.654 - Поток накопителя запущен.
2024-10-31 15:41:37.657 - Поток накопителя запущен и готов к логированию статистики.
2024-10-31 15:41:37.692 - Генерация задач завершена.
2024-10-31 15:42:37.700 - Среднее время существования задачи: 27861,073542 мс
2024-10-31 15:42:37.701 - Среднее время ожидания в очереди стадии 1: 24715,885195 мс
2024-10-31 15:42:37.701 - Среднее время ожидания в очереди стадии 2: 0,052732 мс
2024-10-31 15:42:37.702 - Среднее время ожидания в очереди стадии 3: 0,000000 мс
2024-10-31 15:42:37.702 - Среднее время обработки на стадии 1: 24715,885195 мс
2024-10-31 15:42:37.703 - Среднее время обработки на стадии 2: 2,308495 мс
2024-10-31 15:42:37.703 - Среднее время обработки на стадии 3: 12,577458 мс
```

Рисунок 3.2 – Сообщения о ходе выполнения парсинга

2024-10-31 15:41:38.638 - Задача 1 успешно обработана на стадии 1.  
 2024-10-31 15:41:38.639 - Задача 1 поставлена в очередь стадии 2.  
 2024-10-31 15:41:38.639 - Задача 2 начала обработку на стадии 1: <https://food.ru/recipes/229515-kimchi-iz-daikona>  
 2024-10-31 15:41:39.600 - Задача 2 успешно обработана на стадии 1.  
 2024-10-31 15:41:39.600 - Задача 2 поставлена в очередь стадии 2.  
 2024-10-31 15:41:39.601 - Задача 3 начала обработку на стадии 1: <https://food.ru/recipes/229153-shampinony-pod-shuboi>  
 2024-10-31 15:41:42.711 - Задача 3 успешно обработана на стадии 1.  
 2024-10-31 15:41:42.712 - Задача 3 поставлена в очередь стадии 2.  
 2024-10-31 15:41:42.712 - Задача 4 начала обработку на стадии 1: <https://food.ru/recipes/229142-kvashenaja-kapusta-s-morkoviu-i-jablokami>  
 2024-10-31 15:41:45.078 - Задача 4 успешно обработана на стадии 1.  
 2024-10-31 15:41:45.079 - Задача 4 поставлена в очередь стадии 2.  
 2024-10-31 15:41:45.080 - Задача 5 начала обработку на стадии 1: <https://food.ru/recipes/229361-spadini-s-sibasom-i-cukini>  
 2024-10-31 15:41:47.008 - Задача 5 успешно обработана на стадии 1.

Рисунок 3.3 – Сообщения о ходе выполнения стадии 1

2024-10-31 15:41:38.639 - Задача 1 начала обработку на стадии 2.  
 2024-10-31 15:41:38.644 - Задача 1 успешно обработана на стадии 2.  
 2024-10-31 15:41:38.644 - Задача 1 поставлена в очередь стадии 3.  
 2024-10-31 15:41:39.600 - Задача 2 начала обработку на стадии 2.  
 2024-10-31 15:41:39.601 - Задача 2 успешно обработана на стадии 2.  
 2024-10-31 15:41:39.602 - Задача 2 поставлена в очередь стадии 3.  
 2024-10-31 15:41:42.712 - Задача 3 начала обработку на стадии 2.  
 2024-10-31 15:41:42.715 - Задача 3 успешно обработана на стадии 2.  
 2024-10-31 15:41:42.715 - Задача 3 поставлена в очередь стадии 3.  
 2024-10-31 15:41:45.079 - Задача 4 начала обработку на стадии 2.

Рисунок 3.4 – Сообщения о ходе выполнения стадии 2

2024-10-31 15:41:38.644 - Задача 1 начала обработку на стадии 3.  
 2024-10-31 15:41:38.731 - Задача 1 успешно записана в базу данных.  
 2024-10-31 15:41:38.731 - Задача 1 завершена и записана в базу данных.  
 2024-10-31 15:41:39.602 - Задача 2 начала обработку на стадии 3.  
 2024-10-31 15:41:39.608 - Задача 2 успешно записана в базу данных.  
 2024-10-31 15:41:39.608 - Задача 2 завершена и записана в базу данных.  
 2024-10-31 15:41:42.715 - Задача 3 начала обработку на стадии 3.  
 2024-10-31 15:41:42.722 - Задача 3 успешно записана в базу данных.  
 2024-10-31 15:41:42.722 - Задача 3 завершена и записана в базу данных.  
 2024-10-31 15:41:45.081 - Задача 4 начала обработку на стадии 3.

Рисунок 3.5 – Сообщения о ходе выполнения стадии 3

	id	issue_id	url	title	ingredients	steps
1	1	9162	<a href="https://food.ru/recipes/229700-buterbrod-gribnoi">https://food.ru/recipes/229700-buterbrod-gribnoi</a>	Грибные бутерброды	Тостовый хлеб: 80 г; ...	Пропустите чеснок через пресс.   ...
2	2	9162	<a href="https://food.ru/recipes/229515-kimchi-iz-daikona">https://food.ru/recipes/229515-kimchi-iz-daikona</a>	Кимчи из дайкона	Дайкон: 550 г; Соли: 12.5 г; ...	Нарежьте дайкон небольшими кубиками ...
3	3	9162	<a href="https://food.ru/recipes/229153-shampinony-pod-shuboi">https://food.ru/recipes/229153-shampinony-pod-shuboi</a>	Шампиньоны под ...	Шампиньоны: 384 г; ...	Разрежьте картофель на половинки или ...
4	4	9162	<a href="https://food.ru/recipes/229142-kvashenaja-kapusta-s-morkoviu-i-jablokami">https://food.ru/recipes/229142-kvashenaja-kapusta-s-morkoviu-i-jablokami</a>	Квашеная капуста с ...	Белокочанная капуста: 500 ...	Нашинуйте капусту тонкой соломкой. ...
5	5	9162	<a href="https://food.ru/recipes/229361-spadini-s-sibasom-i-cukini">https://food.ru/recipes/229361-spadini-s-sibasom-i-cukini</a>	Спадини с сибасом ...	Сибас: 700 г; Лайм: 60 г; ...	Снимите цедру с лайма при помощи ...
6	6	9162	<a href="https://food.ru/recipes/229355-brusketta-s-baklazhannoi-ikroi">https://food.ru/recipes/229355-brusketta-s-baklazhannoi-ikroi</a>	Брускетта с ...	Баклажан: 630 г; Красный ...	Выложите баклажаны целиком на угли и ...
7	7	9162	<a href="https://food.ru/recipes/228923-patissony-farshirovannye">https://food.ru/recipes/228923-patissony-farshirovannye</a>	Фаршированные ...	Патиссон: 500 г; Фарш из ...	Срежьте с патиссона верхушку. ...
8	8	9162	<a href="https://food.ru/recipes/229351-nektariny-s-syrom-halumi-na-grile">https://food.ru/recipes/229351-nektariny-s-syrom-halumi-na-grile</a>	Нектарины с сыром ...	Сыр халуми: 200 г; Нектарины ...	Прогрейте гриль до 20-220 °C. Тщательно ...
9	9	9162	<a href="https://food.ru/recipes/228421-tykva-farshirovannaja-mjasom-v-duhovke">https://food.ru/recipes/228421-tykva-farshirovannaja-mjasom-v-duhovke</a>	Фаршированная ...	Тыква: 1800 г; Свинина: 600 ...	Срежьте с тыквы верхушку. Аккуратно ...
10	10	9162	<a href="https://food.ru/recipes/228162-liubov-morkov">https://food.ru/recipes/228162-liubov-morkov</a>	Закуска «Любовь»	Слоеное тесто: 500 г; ...	Разрежьте половину лука на две равные ...
11	11	9162	<a href="https://food.ru/recipes/227774-vietnamskii-sendvich-s-farshem-i-ovoshchami">https://food.ru/recipes/227774-vietnamskii-sendvich-s-farshem-i-ovoshchami</a>	Вьетнамский сэндвич	Багет: 300 г; Свиной фарш: ...	Нарежьте очищенную морковь тонкой ...
12	12	9162	<a href="https://food.ru/recipes/226448-bento-boks-s-onigiri-i-japonskim-omletom">https://food.ru/recipes/226448-bento-boks-s-onigiri-i-japonskim-omletom</a>	Бенто бокс с онигири	Рис для суши: 180 г; Норик: ...	Сварите рис по инструкции на упаковке. В...
13	13	9162	<a href="https://food.ru/recipes/226447-bento-boks-s-sendvichem-i-salatom">https://food.ru/recipes/226447-bento-boks-s-sendvichem-i-salatom</a>	Бенто бокс с ...	Белый хлеб: 50 г; Тунец: ...	Нарежьте мелким кубиком сельдерея и ...
14	14	9162	<a href="https://food.ru/recipes/228002-tvorozhnye-percy-tvorozhnye-percy">https://food.ru/recipes/228002-tvorozhnye-percy-tvorozhnye-percy</a>	Фаршированный ...	Творог 5: 270 г; Болгарский ...	Разрежьте перец вдоль пополам. Аккурат...
15	15	9162	<a href="https://food.ru/recipes/227850-opjata-po-koreiski">https://food.ru/recipes/227850-opjata-po-koreiski</a>	Опята по-корейски	Опята: 500 г; Репчатый лук: ...	Налейте в кастрюлю 2 литра воды. Всып...
16	16	9162	<a href="https://food.ru/recipes/227933-marinovannaja-kapusta-s-lukom">https://food.ru/recipes/227933-marinovannaja-kapusta-s-lukom</a>	Маринованная ...	Белокочанная капуста: 445 ...	Нашинуйте капусту полосками шириной...
17	17	9162	<a href="https://food.ru/recipes/227366-sendvich-s-arbutom">https://food.ru/recipes/227366-sendvich-s-arbutom</a>	Сэндвич с арбузом	Арбуз: 300 г; Моцарелла: 60 ...	Вырежьте из арбуза 2 квадратных ломтик...
18	18	9162	<a href="https://food.ru/recipes/227364-brusselskaja-kapusta-na-zimu">https://food.ru/recipes/227364-brusselskaja-kapusta-na-zimu</a>	Брюссельская ...	Брюссельская капуста: 500 ...	Налейте в кастрюлю 1.5-2 л воды. ...

Рисунок 3.6 – Содержимое выходной базы данных

## 4 Тестирование

Выполнено тестирование реализованной программы по методологии черного ящика. В таблице 4.1 представлено описание тестовых случаев. Все тесты пройдены успешно.

Таблица 4.1 – Описание тестовых случаев

№	Входные данные	Ожидаемый результат	Результат теста
1	Пустой базовый URL	Вывод сообщения об ошибке, запрос корректного URL	Пройден
2	Некорректный базовый URL (без протокола)	Вывод сообщения об ошибке, запрос корректного абсолютного URL	Пройден
3	Отключенное интернет-соединение	Вывод сообщений об ошибках при загрузке страниц	Пройден
4	Корректный базовый URL, 1 страница	Успешная загрузка рецептов со страницы, сохранение в <code>recipes.db</code>	Пройден
5	Корректный базовый URL, 5 страниц	Успешная загрузка рецептов из 5 страниц, сохранение в <code>recipes.db</code>	Пройден
6	Корректный URL, 0 страниц	Вывод сообщения об ошибке, запрос корректного числа страниц	Пройден
7	Корректный URL, отрицательное число страниц	Вывод сообщения об ошибке, запрос корректного числа страниц	Пройден

## 5 Описание исследования

В ходе исследования требуется сформировать лог обработки задач. В таблицах 5.1 — 5.4 приведены фрагменты лога обработки.

Таблица 5.1 – Общий лог

Время	Статус
2024-10-31 17:22:05.967	Поток стадии 1 запущен.
2024-10-31 17:22:05.970	Поток стадии 2 запущен.
2024-10-31 17:22:05.972	Поток стадии 3 запущен.
2024-10-31 17:22:05.975	Поток накопителя запущен.
2024-10-31 17:22:05.978	Поток накопителя готов к логированию статистики.
2024-10-31 17:22:06.001	Генерация задач завершена.
2024-10-31 17:23:05.999	Среднее $t$ существования задачи: 32269,128355 мс
2024-10-31 17:23:05.999	Среднее $t$ ожидания в очереди стадии 1: 26951,583527 мс
2024-10-31 17:23:06.000	Среднее $t$ ожидания в очереди стадии 2: 0,000000 мс
2024-10-31 17:23:06.000	Среднее $t$ ожидания в очереди стадии 3: 0,090936 мс
2024-10-31 17:23:06.000	Среднее $t$ обработки на стадии 1: 26951,583527 мс
2024-10-31 17:23:06.001	Среднее $t$ обработки на стадии 2: 3,180491 мс
2024-10-31 17:23:06.001	Среднее $t$ обработки на стадии 3: 7,464700 мс



Таблица 5.2 – Лог стадии 1

Время	Статус
2024-10-31 17:22:13.079	Задача 2 начала обработку на стадии 1: (ссылка)
2024-10-31 17:22:17.636	Задача 2 успешно обработана на стадии 1.
2024-10-31 17:22:17.636	Задача 2 поставлена в очередь стадии 2.
2024-10-31 17:22:17.637	Задача 3 начала обработку на стадии 1: (ссылка)
2024-10-31 17:22:23.061	Задача 3 успешно обработана на стадии 1.
2024-10-31 17:22:23.062	Задача 3 поставлена в очередь стадии 2.
2024-10-31 17:22:23.062	Задача 4 начала обработку на стадии 1: (ссылка)

Таблица 5.3 – Лог стадии 2

Время	Статус
2024-10-31 17:22:13.082	Задача 1 поставлена в очередь стадии 3.
2024-10-31 17:22:17.636	Задача 2 начала обработку на стадии 2.
2024-10-31 17:22:17.638	Задача 2 успешно обработана на стадии 2.
2024-10-31 17:22:17.638	Задача 2 поставлена в очередь стадии 3.
2024-10-31 17:22:23.062	Задача 3 начала обработку на стадии 2.
2024-10-31 17:22:23.064	Задача 3 успешно обработана на стадии 2.
2024-10-31 17:22:23.064	Задача 3 поставлена в очередь стадии 3.
2024-10-31 17:22:27.787	Задача 4 начала обработку на стадии 2.

Таблица 5.4 – Лог стадии 3

Время	Статус
2024-10-31 17:22:13.095	Задача 1 завершена и записана в базу данных.
2024-10-31 17:22:17.638	Задача 2 начала обработку на стадии 3.
2024-10-31 17:22:17.645	Задача 2 успешно записана в базу данных.
2024-10-31 17:22:17.645	Задача 2 завершена и записана в базу данных.
2024-10-31 17:22:23.064	Задача 3 начала обработку на стадии 3.
2024-10-31 17:22:23.070	Задача 3 успешно записана в базу данных.
2024-10-31 17:22:23.070	Задача 3 завершена и записана в базу данных.
2024-10-31 17:22:27.789	Задача 4 начала обработку на стадии 3.

В результате проведённого исследования было подтверждено, что конвейерная обработка выполняет различные этапы параллельно, обеспечивая более высокую скорость обработки по сравнению с простой последовательной обработкой. Однако, поскольку стадия 1 занимает наибольшее время выполнения, вторые и третьи потоки часто простаивают в ожидании следующих задач. Это свидетельствует о возможности дальнейшего улучшения системы для более эффективного использования ресурсов и повышения общей производительности.

## ЗАКЛЮЧЕНИЕ

Цель работы достигнута: получен навык организации параллельных вычислений по конвейерному принципу.

В ходе выполнения данной лабораторной работы были решены следующие задачи:

- описаны входные, выходные данные, а также преобразования входных данных в выходные;
- реализован алгоритм для парсинга рецептов с веб-сайта с использованием конвейерной обработки и многопоточности на основе нативных потоков;
- протестирован разработанный алгоритм по методологии черного ящика;
- описан пример работы программы в конкретном случае;
- сделаны выводы, основанные на информации в полученных лог-файлах.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Threading Model Overview [Электронный ресурс]. URL: <http://justin.harmonize.fm/index.php/2008/09/threading-model-overview/> (дата обращения: 20.10.2024).
- [2] Pipeline Principle [Электронный ресурс]. URL: [https://www.lkouniv.ac.in/site/writereaddata/siteContent/202004221613338445rohit\\_engg\\_pipelining\\_and\\_hazzard.pdf](https://www.lkouniv.ac.in/site/writereaddata/siteContent/202004221613338445rohit_engg_pipelining_and_hazzard.pdf) (дата обращения: 31.10.2024).