



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»

**Лабораторная работа № 4  
по дисциплине «Анализ Алгоритмов»**

**Тема** Параллельные вычисления на основе нативных потоков

**Студент** Шахнович Дмитрий Сергеевич

**Группа** ИУ7-52Б

**Преподаватели** Волкова Л.Л., Строганов Д.В.

# СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ . . . . .</b>	<b>3</b>
<b>1 Входные и выходные данные . . . . .</b>	<b>3</b>
<b>2 Преобразование входных данных в выходные . . . . .</b>	<b>3</b>
<b>3 Примеры работы программы . . . . .</b>	<b>5</b>
<b>4 Тестирование . . . . .</b>	<b>5</b>
<b>5 Описание исследования . . . . .</b>	<b>6</b>
<b>ЗАКЛЮЧЕНИЕ . . . . .</b>	<b>8</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ . . . . .</b>	<b>9</b>

# ВВЕДЕНИЕ

**Параллелизм** описывает последовательности, которые происходят одновременно [1]. Таким образом параллельные вычисления таковы, что они выполняются одновременно. Распараллеливание вычислений может привести к росту временной эффективности программы при использовании на многопроцессорных и однопроцессорных, если программа часто блокируется ожиданиями ввода/вывода, системах [2].

В современных системах различают параллельность реализованную на потоках и процессах. При этом стоит учитывать, что часто программы выполняются не строго параллельно, а конкурентно, то есть выполняются последовательно, постоянно переключаясь между друг другом [2].

Целью данной работы является разработка ПО, выполняющего скачивание страниц и парсинг рецептов с сайта menunedeli.ru.

Задачами работы являются:

- рассмотрение структуры сайта;
- разработка ПО, выполняющего скачивание и парсинг в одном потоке;
- разработка многопоточного ПО, выполняющего скачивание и парсинг;
- исследование временных характеристик разработанных программ.

## 1 Входные и выходные данные

Входными данными для программы являются ссылки на страницы сайта menunedeli.ru. с рецептами. Каждая ссылка содержит один рецепт в веб-ресурсе. Выходными данными являются файлы – загруженные рецепты по ссылкам из входных данных, при этом в файлах содержится часть сайта с рецептом.

## 2 Преобразование входных данных в выходные

Для получения ссылок с рецептами с сайта в виде файла разработан скрипт parse.py на языке python3:

Листинг 2.1 — parse.py – скрипт для получения ссылок с рецептами,

```
import requests as re
import bs4
import argparse
import sys

parser = argparse.ArgumentParser(
    prog="parse" ,
```

```

description="Скачивает ссылки на статьи с сайта menunedeli.ru"
)
parser.add_argument("-c", "--count", type=int, default=1000, help="Количество ссылок для скачивания")
parser.add_argument("-s", "--save", type=str, default="links.txt", help="Файл, куда сохранять ссылки")

args = parser.parse_args()

saveTo = args.save
UpToLinks = args.count
catalogFormat = "https://menunedeli.ru/novye-stati/page/{}"
with open(saveTo, "w") as f:
    links = 0
    page = 1
    while True:
        catalogPage = re.get(catalogFormat.format(page))

        bs = bs4.BeautifulSoup(catalogPage.content, "lxml")

        for a in bs.find_all('article'):
            for link in a.find_all('meta', attrs={'itemprop': 'url'}):
                print(link['content'], file=f)
                links += 1
                if links >= UpToLinks:
                    exit(0)
    page += 1

```

Полученный в результате работы программы 2.1 файл выступает входными данными для разработанного ПО.

Веб-страница по каждой из ссылок скачивается и обрабатывается одной из программ либо в однопоточном режиме, либо многопоточном, при этом из всей полученной страницы остаётся только часть с рецептом.

В результате работы программ для каждой из ссылок создаётся html файл, в котором хранится рецепт.

Программы, выполняющие скачивание и обработку страниц был использован язык программирования С [4], при этом для реализации многопоточности были использованы posix threads из библиотеки pthreads [5].

## 3 Примеры работы программы

Один из примеров обработки веб-страниц [6] представлен на рисунках 3.1-3.2.

The screenshot shows a web page from the menunedeli.ru website. At the top, there is a navigation bar with links for 'О сайте' (About), 'Бесплатные рецепты' (Free recipes), 'Бесплатные мероприятия и подарки' (Free events and gifts), 'Лучшие курсы' (Best courses), 'Бесплатный урок' (Free lesson), a search bar with placeholder 'По рецептам' (Search by recipe), and a search icon. Below the navigation bar, there is a sidebar titled 'Школа хозяйки' (Cooking School) with links for 'Кулинарная школа' (Culinary school), 'Меню на все случаи' (Menu for all occasions), 'Как готовить быстро?' (How to cook quickly?), 'Как готовить экономно?' (How to cook economically?), 'Организация пространства на кухне: что нужно на кухне' (Organizing kitchen space: what is needed in the kitchen), 'Что приготовить?' (What to cook?), 'Что приготовить быстро и вкусно для завтрака' (What to cook quickly and deliciously for breakfast), 'Что приготовить на ужин для всей семьи вкусно и быстро' (What to cook for dinner for the whole family quickly and deliciously), 'Что приготовить на ужин недорого: рецепты на каждый день' (What to cook for dinner inexpensively: recipes for every day), 'Рецепты' (Recipes), 'Средневенные' (Midday), and 'Салаты' (Salads). The main content area features a title 'Блины с творогом: рецепт начинки' (Cheese blini: filling recipe) with a photo of the dish. Below the photo, there is a note: 'Можно встретить как сладкую, так и не сладкую версию. Но, однозначно, лидирует первая.' (You can encounter both sweet and non-sweet versions. However, it is clear that the first one leads.) There is also a note: 'В него можно добавлять изюм, курагу, шоколад или мак. Тут уже полет фантазии и ваши предпочтения.' (You can add raisins, currants, chocolate or macadamia. Here your imagination and preferences come into play.) A purple button labeled 'СОХРАНИТЬ СЕБЕ ЭТУ РЕЦЕПТ' (Save this recipe) is visible. The footer of the page includes a logo for 'menunedeli' and a copyright notice: '© 2018 menunedeli.ru'.

Рисунок 3.1 — Веб-страница по ссылке [6]

### Блины с творогом: рецепт начинки



Одна из самых популярных начинок для тонких блинчиков - это творог. Можно встретить как сладкую, так и не сладкую версию. Но, однозначно, лидирует первая.

Для начинки лучше выбирать мелкозернистый творог или кремовый. В него можно добавлять изюм, курагу, пукаты, шоколад или мак. Тут уже полет фантазии и ваши предпочтения. Ну, а я чаще всего готовлю две начинки из творога: вариант с сырным яйцом и второй со сметаной.

[Сохранить себе этот рецепт](#)

• [Телеграм](#)

- Общ.\актив. время готовки: 10 минут \ 10 минут
- Рейтинг: 5 / 5 2 отзыва [Подробнее о рейтинге](#)

X  
★★★★★  
2  
★★★★★  
0  
★★★  
~

Рисунок 3.2 — Результат обработки программой

## 4 Тестирование

Тестирование программы проводилось загрузкой в качестве входных данных 1000 ссылок на различные рецепты веб-ресурса menunedeli.ru. При этом отслеживалось количество удачных

скачиваний и обработок страницы, а также выборочная ручная проверка 10 случайных выходных файлов.

Тестирование проводилось как для однопоточной реализации, так и для многопоточной, при этом выходные файлы обоих реализаций при тестировании оказались идентичными.

Тестирование было успешно пройдено.

## 5 Описание исследования

В ходе исследования были собраны данные о временных характеристиках однопоточной и многопоточной реализаций, при этом для многопоточной реализации были собраны данные при количестве потоков равному 1, 2, 4, 8, 16, 32, 64, 128.

Для каждого из вариантов программы выполнение проводилось 10 раз и бралось среднее арифметическое времён, на вход подавался файл с 1000 ссылок.

Технические характеристики устройства, на котором проводились замеры:

- процессор: AMD Ryzen 7 5800H (16) @ 4.46 ГГц;
- оперативная память: 16 ГБ;
- операционная система: Arch Linux x86\_64.

При проведении замеров ноутбук был включён в сеть и были запущены только системные приложения.

Временные характеристики реализаций представлены в таблице 5.1 и на графике 5.1.

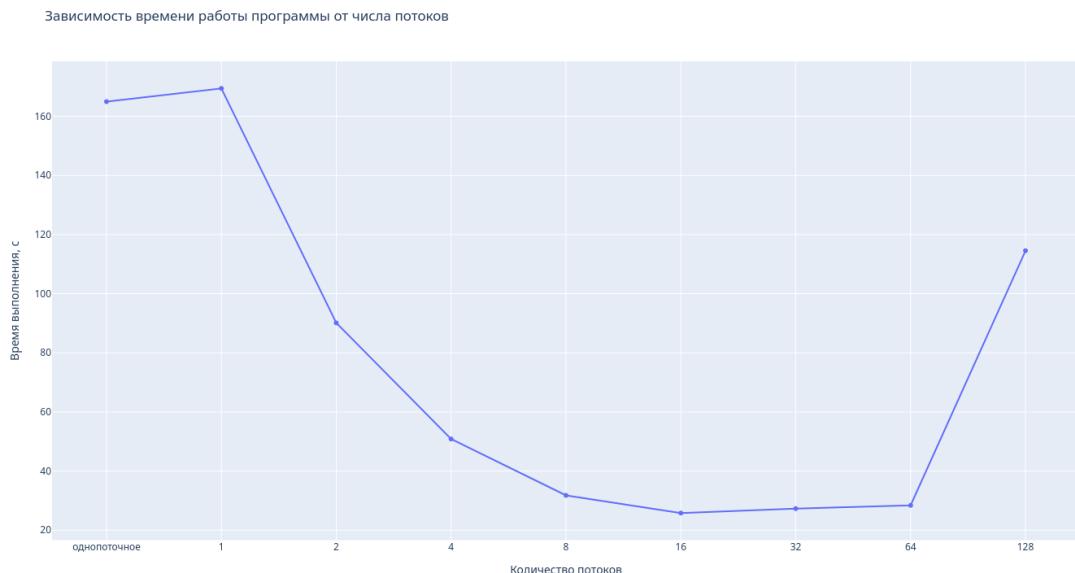


Рисунок 5.1 — График временных характеристик разработанных программ

Таблица 5.1 — Временные характеристики разработанных программ (начало)

Количество потоков	Время работы, с
Однопоточная реализация	
1	164.98
Многопоточная реализация	
1	169.45
2	90.11
4	50.85
8	31.79
16	25.78
32	27.32
64	28.42
128	114.53

В результате исследования сделан вывод, что использование параллельных вычислений на нативных потоках может ускорить программу, однако использование потоков, больше числа потоков процессора (в случае устройства, на котором проводились замеры >16 потоков), не даст увеличения временной производительности.

# **ЗАКЛЮЧЕНИЕ**

Целью – разработка ПО, выполняющего скачивание страниц и парсинг рецептов с сайта menunedeli.ru – была выполнена.

В ходе работы были решены следующие задачи:

- рассмотрена структура сайта;
- разработано ПО, выполняющее скачивание и парсинг в одном потоке;
- разработано многопоточное ПО, выполняющее скачивание и парсинг;
- исследованы временные характеристики разработанных программ.

В ходе исследования было выявлено, что использование параллельных вычислений на нативных потоках ускоряет программу, однако использование большего числа потоков, чем может выполнять процессор не даёт прироста производительности, а замедляет программу.

## **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. David R. Butenhof Programming with POSIX threads [Текст] / David R. Butenhof — . — : Library of Congress Cataloging-in-Publication Data, 1997 — 202 с.
2. Таненбаум Э., Бос Х. Современные операционные системы. [Текст] / Таненбаум Э., Бос Х. — 4-е изд.. — СПб.: Питер, 2015 — 1120 с.
3. Python / [Электронный ресурс] // Python : [сайт]. — URL: <https://www.python.org/> (дата обращения: 20.09.2024).
4. Керниган, Б. У., Ритчи, Д. М. Язык программирования С [Текст] / Б. У. Керниган, Д. М. Ритчи — 2-е издание. — М.: Издательский дом "Вильямс 2013 — 304 с.
5. Michael Kerrisk. pthreads. / Michael Kerrisk. [Электронный ресурс] // Linux man-pages : [сайт]. — URL: <https://man7.org/linux/man-pages/man7/pthreads.7.html> (дата обращения: 20.10.2024).
6. Ольга Котельникова. Блины с творогом: рецепт начинки. / Ольга Котельникова. [Электронный ресурс] // Меню недели : [сайт]. — URL: <https://menunedeli.ru/recipe/bliny-s-tvorogom-recept-nachinki/> (дата обращения: 20.10.2024).