

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

Группа ИУ7-51Б

Тема работы **Организация параллельных вычислений по конвейерному** принципу

Студент

Баранов Николай Алексеевич

Преподаватель

Волкова Лилия Леонидовна

СОДЕРЖАНИЕ

В	ВЕДЕНИЕ	3
1	Входные и выходные данные	3
2	Преобразование входных данных в выходные	3
3	Примеры работы программы	4
4	Тестирование	5
5	Описание исследования	5
3.	АКЛЮЧЕНИЕ	9
\mathbf{C}	ПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	10

ВВЕДЕНИЕ

В данной лабораторной работе рассматривается организация параллельных вычислений по конвейерному принципу.

Цель работы – получение навыка организации параллельных вычислений по конвейерному принципу.

Задачи работы:

- анализ предметной области;
- разработка алгоритма обработки данных с выгруженных страниц сайта eda.ru [1];
- создание ПО, реализующего разработанный алгоритм;
- исследование характеристик созданного ПО.

1 Входные и выходные данные

Входными данными является папка с HTML-файлами. Каждый файл содержит одну страницу рецепта, выгруженную с сайта eda.ru [1]. Выходными данными является база данных, содержащая извлечённую из страниц информацию, а именно: URL страницы рецепта, ингредиенты, шаги рецепта и URL основного изображения рецепта. Каждая запись в базе данных дополнительно содержит уникальный идентификатор задачи, а также номер задачи из Redmine (9134).

2 Преобразование входных данных в выходные

Программа находит все файлы в папке. Из каждого файла считываются данные, после чего из них извлекаются URL страницы рецепта, ингредиенты, шаги рецепта и URL основного изображения рецепта. Полученные данные сохраняются в базе данных. Процесс завершается после обработки всех файлов в папке.

3 Примеры работы программы

Для реализации данной лабораторной работы был выбран язык Java, так как он содержит все необходимые средства для реализации алгоритмов. Нативные потоки создавались при помощи класса Thread [2] через явный вызов конструктора. В качестве СУБД была выбрана SQLite, так как она соответствует требованиям из условия лабораторной работы [3].

На рисунке 3.1 представлен пример запуска программы из терминала. На рисунке 3.2 представлена таблица в базе данных с выгруженными данными.

```
кпіfe@knife-Swift-SF314-5106:~/labs/aa/lab5/lab5$ java -jar build/libs/lab5.jar
Обработка рецептов завершена
Среднее время существования задачи: 1.803053703
Среднее время ожидания в очереди 1: 0.795731311
Среднее время ожидания в очереди 2: 0.714282865
Среднее время ожидания в очереди 3: 0.000174197
Среднее время ожидания в очереди 4: 0.000179046
Среднее время выполнения этапа read: 0.098508496
Среднее время выполнения этапа parse: 0.173286863
Среднее время выполнения этапа save: 0.020890926
```

Рисунок 3.1 – Запуск программы

	✓ SELECT *	from recipes;						
<u> </u>	Soutput ⊞ main.recipes × Tx, ■ ⊞							
_								
Ш	<u>₩</u> K < 1	5 rows 🗸 🗦 🤣 (<u>9, ■ + − 5 @ 6</u>	Tx: Auto V DD	L ¥ Q LLJ			
	∏qid 7 ÷	□ issue_id 7 ÷	□ image_url 7 ÷	□ steps 7 ÷	□ ingredients ▽ ÷	□ title 7	÷ □url 7 ÷	
	1	9134	https://eda.ru/ima	["Картофель х	[{"пате":"Молодой ка…	Картофель «Айдахо».	https://eda.ru/recepty	
	2	9134	https://eda.ru/ima	["Нарежьте бо…	[{"пате":"Куриная гр	Курица «Пикассо».	https://eda.ru/recepty	
	3	9134	https://eda.ru/ima	["Мясо промыт	[{"пате":"Говядина",…	Азу по-татарски.	https://eda.ru/recepty	
	4	9134	https://eda.ru/ima	["Подготовить	[{"пате":"Картофель"	Картофель хассельбак.	https://eda.ru/recepty	
	5	9134	https://eda.ru/ima	["В сотейник …	[{"пате":"Мясной фар…	Лазанья классическая с мясом.	https://eda.ru/recepty	
	6	9134	https://eda.ru/ima	["Вскипятите …	[{"name":"Спагетти",…	Спагетти карбонара со сливками.	https://eda.ru/recepty	
	7	9134	https://eda.ru/ima	["Положите ве…	[{"name":"Творог","u	Сырники из творога.	https://eda.ru/recepty	
	8	9134	https://eda.ru/ima	["Мелко нарез…	[{"пате":"Рис арбори…	Ризотто с белыми грибами.	https://eda.ru/recepty	
	9	9134	https://eda.ru/ima	["Просеять му…	[{"пате":"Пшеничная …	Американский тыквенный пирог с кориц	e https://eda.ru/recepty	
	10	9134	https://eda.ru/ima	["Взбить яйца…	[{"пате":"Пшеничная …	Тонкие блины на молоке.	https://eda.ru/recepty	
	11	9134	https://eda.ru/ima	["Подготовить	[{"name":"Молоко","u	Легендарные пирожки с грибами из «Ц.	Д https://eda.ru/recepty	
	12	9134	https://eda.ru/ima	["Шоколад раз…	[{"пате":"Темный шок…	Брауни (brownie).	https://eda.ru/recepty	
	13	9134	https://eda.ru/ima	["Разогреть д	[{"name":"Caxap","un	Классическая шарлотка.	https://eda.ru/recepty	
	14	9134	https://eda.ru/ima	["Взять 200 г	[{"name":"Творог","u	Творожный кекс с яйцами.	https://eda.ru/recepty	
	15	9134	https://eda.ru/ima	["Для начала	[{"пате":"Желатин","	Десерт «Яблоко».	https://eda.ru/recepty	

Рисунок 3.2 – Полученная папка с рецептами

4 Тестирование

Выполнено тестирование программы по методологии чёрного ящика. В таблице 4.1 представлены функциональные тесты. Все тесты пройдены успешно.

Таблица 4.1 – Результаты выполнения функциональных тестов

	Исходное	Ожидаемое	Полученное
$N_{\overline{0}}$	количество	количество	количество
	файлов	записей	записей
1	10	10	10
2	15	15	15
3	25	25	25
4	100	100	100

5 Описание исследования

В ходе исследования требуется сформировать лог обработки задач, а также получить среднее время существования задачи, среднее время ожидания задачи в каждой из очередей, а также среднее время обработки задачи на каждой из стадий. Для формирования лога программа получала на вход папку с 10 выгруженными рецептами. Замеры времени проводились с помощью метода System.nanoTime() [4]. Все замеры проводились на ноутбуке AcerSwift3x, 11thGenIntel(R)Core(TM)i7 - 1165G7.

Лог обработки задач представлен в таблице 5.1. Обозначения событий:

- created создание задачи;
- $start_read$ начало чтения файла;
- $stop_read$ окончание чтения файла;
- $start_parse$ начало извлечения данных;
- *stop_parse* окончание извлечения данных;
- start_save начало сохранения данных;

- *stop_save* окончание сохранения данных;
- $-\ destroyed$ уничтожение задачи;

Таблица 5.1 — Лог обработки задач

№	Метка	Номер	Событие
] \ ~	времени	задачи	Сооытие
1	5914.566054132	1	created
2	5914.567273217	2	created
3	5914.567277993	1	start_read
4	5914.567498916	3	created
5	5914.567775497	4	created
6	5914.567929022	5	created
7	5914.568164955	6	created
8	5914.568311197	7	created
9	5914.568466558	8	created
10	5914.568661928	9	created
11	5914.568818656	10	created
12	5914.780051346	1	end_read
13	5914.780431769	2	start_read
14	5914.780494712	1	start_parse
15	5914.832295667	2	end_read
16	5914.832559352	3	start_read
17	5914.908876301	3	end_read
18	5914.909236768	4	start_read
19	5915.018138436	4	end_read
20	5915.018316692	5	start_read
21	5915.058474246	5	end_read
22	5915.058698517	6	start_read
23	5915.064677421	1	end_parse
24	5915.064829694	2	start_parse
25	5915.064893024	1	start_save
26	5915.092590470	6	end_read

27	5915.092770066	7	start_read
28	5915.126669438	7	end_read
29	5915.126906159	8	start_read
30	5915.175440624	8	end_read
31	5915.175696939	9	start_read
32	5915.229258635	9	end_read
33	5915.229924620	10	start_read
34	5915.260029949	1	end_save
35	5915.260302422	1	destroyed
36	5915.263341521	10	end_read
37	5915.292567461	2	end_parse
38	5915.292742219	3	start_parse
39	5915.292771230	2	start_save
40	5915.304765031	2	end_save
41	5915.305110055	2	destroyed
42	5915.535943401	3	end_parse
43	5915.536102866	4	start_parse
44	5915.536132931	3	start_save
45	5915.546228427	3	end_save
46	5915.546405107	3	destroyed
47	5915.732401342	4	end_parse
48	5915.732542625	5	start_parse
49	5915.732543652	4	$start_save$
50	5915.742133994	4	end_save
51	5915.742264356	4	destroyed
52	5915.866717891	5	end_parse
53	5915.866923255	5	$start_save$
54	5915.866924900	6	start_parse
55	5915.876488213	5	end_save
56	5915.876626804	5	destroyed
57	5915.971145058	6	end_parse
58	5915.971294991	6	start_save
59	5915.971296282	7	start_parse

60	5915.976689610	6	end_save
61	5915.976842919	6	destroyed
62	5916.075069522	7	end parse
63	5916.075299060	8	start_parse
64	5916.075302356	7	start_save
65	5916.085196926	7	end_save
66	5916.085309083	7	$\frac{-}{destroyed}$
67	5916.186589994	8	end_parse
68	5916.186721748	8	start_save
69	5916.186722932	9	start_parse
70	5916.197654520	8	end_save
71	5916.197762992	8	destroyed
72	5916.307698707	9	end_parse
73	5916.307823192	10	start_parse
74	5916.307825498	9	start_save
75	5916.317253554	9	end_save
76	5916.317369350	9	destroyed
77	5916.414392967	10	end_parse
78	5916.414509037	10	start_save
79	5916.420278978	10	end_save
80	5916.420386453	10	destroyed

В таблице 5.2 представлены результаты замеры среднего времени работы. Очередь 1 передаёт задачи из потока создания задач в поток чтения файлов, очередь 2 — из потока чтения файлов в поток извлечения данных, очередь 3 — из потока извлечения данных в поток сохранения данных, очередь 4 — из потока сохранения данных в поток уничтожения задачи.

Таблица 5.2 – Лог обработки задач

$N_{\overline{0}}$	Характеристика	Значение
1	Среднее время существования задачи	1.304942546
2	Среднее время ожидания в очереди 1	0.411286480
3	Среднее время ожидания в очереди 2	0.632964180

4	Среднее время ожидания в очереди 3	0.000171396
5	Среднее время ожидания в очереди 4	0.000166034
6	Среднее время выполнения этапа чтения файлов	0.069331781
7	Среднее время выполнения этапа извлечения данных	0.163242528
8	Среднее время выполнения этапа сохранения данных	0.027780148

Дольше всего выполнялся этап извлечения данных, из-за чего в очереди 2 среднее время ожидания дольше, чем в остальных, а в очередях 3 и 4 извлечение задачи происходило почти сразу же после постановки.

ЗАКЛЮЧЕНИЕ

Цель работы достигнута. Решены все поставленные задачи:

- анализ предметной области;
- разработка алгоритма обработки данных с выгруженных страниц сайта eda.ru [1];
- создание ПО, реализующего разработанный алгоритм;
- исследование характеристик созданного ПО.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1. «Еда» [Электронный ресурс]. Режим доступа: https://eda.ru (дата обращения: 4.11.2024).
- 2. Документация языка программирования Java класс Thread [Электронный ресурс]. Режим доступа: https://docs.oracle.com/en/java/javase/21/docs/api/java.base/java/lang/Thread.html (дата обращения: 4.11.2024).
- 3. Документация базы данных SQLite [Электронный ресурс]. Режим доступа: https://www.sqlite.org/docs.html (дата обращения: 15.11.2024).
- 4. Документация языка программирования Java класс System [Электронный ресурс]. Режим доступа: https://docs.oracle.com/en/java/javase/21/docs/api/java.base/java/lang/System.html#nanoTime() (дата обращения: 4.11.2024).