Университет ИТМО Кафедра вычислительной техники

Отчет по прохождению практики Параллельные вычисления

Выполнил студент группы Р3310 Шаймарданов Р. Р.

Руководитель Соснин В. В.

Содержание

Вве	едение		
Ин	струме	ентарий	
2.1	Систе	ема компьютерной верстки T _E X(L ^A T _E X)	
	2.1.1	Описание	
	2.1.2	Сравнение с другими программными средствами	
	2.1.3	Выбор инструмента редактирования	
2.2	Систе	ема контроля версий Git	
	2.2.1	Описание	
	2.2.2	Сравнению с другими системами контроля версий	
	2.2.3	Основные команды	
	2.2.4	GitHub	
Пеј	речень	ь обучающих источников	
3.1	Крите	ерий сравнения	
3.2	Иност	гранные материалы	
	3.2.1	Книги	
	3.2.2	Бесплатные книги	
	3.2.3	Журналы	
	3.2.4	Онлайн-курсы	
3.3	Росси	йские материалы	
	3.3.1	Книги	
	3.3.2	Весплатные книги	
	3.3.3	Онлайн-курсы	
Ист	топьзу	емые сайты	6

1 Введение

Цель: составить составить список из наиболее авторитетных литературных источников по "Параллельным вычислениям".

Задачи

- 1. Описать инструментарий, необходимый для выполнения практического задания.
- 2. Составить перечень современных (новее 2010 года) англоязычных источников, посвящённых параллельным вычислениям в системах с общей памятью.
- 3. Разбить найденные источники на четыре группы:
 - платные книги
 - журналы
 - бесплатные книги
 - онлайн-курсы
- 4. Сформулировать критерий уровня авторитетности для найденных материалов, выполнить ранжирование источников внутри групп.
- 5. Найти, провести разбиение и ранжировать русскоязычные источники, аналогично с иностранными.

2 Инструментарий

2.1 Система компьютерной верстки ТЕХ(ЕТЕХ)

2.1.1 Описание

Тех—это низкоуровневый язык разметки и программирования, созданный Дональдом Кнутом для единообразной вёрстки документов. Кнут начал разрабатывать систему набора текста Техв 1977 году для исследования потенциальных возможностей оборудования цифровой печати, которое в то время начинало проникать в издательское дело. Главным образом он надеялся улучшить качество печатной продукции, которое расстраивало в его собственных книгах и статьях. После выпуска в 1989 году поддержки восьмибитных символов разработка Техприостановилась, только иногда выходили версии с исправленными ошибками.

IATEX— основанный на ТЕХпакет макросов, созданный Лесли Лампортом. Основная цель — упростить вёрстку текста, особенно в документах с математическими формулами. Значительно позднее авторы разработали для IATEX расширения, которые называются пакетами или стилями. Некоторые из них распространяются вместе с большинством дистрибутивов ТЕХ/IATEX.

Так как IATEX содержит часть команд TEX, то создание документа в IATEX- тоже программирование: создаётся текстовый файл в IATEX разметке, макросы LaTeX обрабатывают его и производят конечный документ.

2.1.2 Сравнение с другими программными средствами

Подход IATEX к созданию документа называется WYSIWYM¹: во время набора текста Вы не видите окончательный вариант документа, толь-

¹What You See Is What You Mean (То, что ты видишь, есть то, что ты имеешь в виду)

ко логическую структуру этого документа. Оформлением занимается сам \LaTeX Такой подход имеет как достоинства, так и недостатки по сравнению с \LaTeX программами, такими как Openoffice.org Writer или Microsoft Word.

Достоинства:

- Файлы с исходными текстами можно просмотреть в любом текстовом редакторе, они понятнее в отличие от сложных бинарных файлов и форматов XML, используемых WYSIWYG программами.
- Вы полностью сосредотачиваетесь на структуре и содержании документа и забываете о том, как будет выглядеть печатный вариант.
- Не нужно вручную настраивать шрифты, размер текста, высоту строк или читаемость текста.
- Легко скопировать структуру документа в другой документ, в WYSIWYG программах не всегда ясно, какое именно было использовано форматирование.
- Разметка, шрифты, таблицы и т.д. согласованы во всём документе.
- Легко набирать математические формулы.
- Легко создаются алфавитные указатели, сноски, ссылки и библиографические списки.
- Так как исходный документ содержит просто текст, с помощью программных средств на любом языке программирования можно создать таблицы, рисунки, формулы и т.д.

Недостатки:

• Во время редактирования документа нельзя (обычно) увидеть его окончательный вариант.

²What You See Is What You Get (Что видишь, то и получишь)

- Необходимо знать нужные команды разметки РТДХ.
- Иногда сложно получить требуемый вид документа.

Документ LATEX— обычный текстовый файл, в котором указано содержание документа вместе с дополнительной разметкой. При обработке исходного файла макросами LATEX можно получить документ в разных форматах. Изначально LATEX поддерживает форматы DVI и PDF, но при использовании другого ПО можно легко получить PostScript, PNG, JPG и т.д.

2.1.3 Выбор инструмента редактирования

В ходе изучения всех возможных вариантов работа с LATEX для создания данного отчета, была выбрана программа Textmaker

Выбор Textmaker'a обусловлен следующими его особенностями:

- Автоматическая подсветка синтаксиса
- Функция автодополнения команд ИТЕХ
- Сокрытие блоков кода
- Быстрая навигация по структуре документа
- Указание на строку с ошибкой, для быстрой отладки
- Интегрированный просмотр PDF

2.2 Система контроля версий Git

2.2.1 Описание

Система управления версиями — программное обеспечение для облегчения работы с изменяющейся информацией. Система управления версиями позволяет хранить несколько версий одного и того же документа, при необходимости возвращаться к более ранним версиям, определять, кто и когда сделал то или иное изменение, и многое другое.

Git — это гибкая, распределённая система управления версиями. Проект был создан Линусом Торвальдсом для управления разработкой ядра Linux, первая версия выпущена 7 апреля 2005 года. На сегодняшний день его поддерживает Джунио Хамано. Программа является свободной и выпущена под лицензией GNU GPL версии 2.

У каждого разработчика, использующего Git, есть свой локальный репозиторий, позволяющий локально управлять версиями. Затем, сохраненными в локальный репозиторий данными, можно обмениваться с другими пользователями. Часто при работе с Git создают центральный репозиторий, с которым остальные разработчики синхронизируются. В этом случае все участники проекта ведут свои локальны разработки и беспрепятственно скачивают обновления из центрального репозитория. Когда необходимые работы отдельными участниками проекта выполнены и отлажены, они, после удостоверения владельцем центрального репозитория в корректности и актуальности проделанной работы, загружают свои изменения в центральный репозиторий. Работа над версиями проекта в Git может вестись в нескольких ветках, которые затем могут с легкостью полностью или частично объединяться, уничтожаться, откатываться и разрастаться во все новые и новые ветки проекта.

2.2.2 Сравнению с другими системами контроля версий

Достоинства:

- Надежная система сравнения ревизий и проверки корректности данных, основанные на алгоритме хеширования Secure Hash Algorithm 1.
- Гибкая система ветвления проектов и слияния веток между собой.
- Наличие локального репозитория, содержащего полную информацию обо всех изменениях, позволяет вести полноценный локальный контроль версий и заливать в главный репозиторий только полностью прошедшие проверку изменения.
- Высокая производительность и скорость работы.
- Удобный и интуитивно понятный набор команд.
- Множество графических оболочек, позволяющих быстро и качественно вести работы с Git'ом.
- Возможность делать контрольные точки, в которых данные сохраняются без дельта компрессии, а полностью. Это позволяет уменьшить скорость восстановления данных, так как за основу берется ближайшая контрольная точка, и восстановление идет от нее. Если бы контрольные точки отсутствовали, то восстановление больших проектов могло бы занимать часы.
- Широкая распространенность, легкая доступность и качественная документация.
- Гибкость системы позволяет удобно ее настраивать и даже создавать специализированные контроля системы или пользовательские интерфейсы на базе git.

• Универсальный сетевой доступ с использованием протоколов http, ftp, rsync, ssh и др.

Недостатки:

- Unix ориентированность. На данный момент отсутствует зрелая реализация Git, совместимая с другими операционными системами.
- Возможные (но чрезвычайно низкие) совпадения хеш кода отличных по содержанию ревизий.
- Не отслеживается изменение отдельных файлов, а только всего проекта целиком, что может быть неудобно при работе с большими проектами, содержащими множество несвязных файлов.
- При начальном (первом) создании репозитория и синхронизации его с другими разработчиками, потребуется достаточно длительное время для скачивания данных, особенно, если проект большой, так как требуется скопировать на локальный компьютер весь репозиторий.

2.2.3 Основные команды

add: Добавляет содержимое рабочей директории в индекс для последующего коммита.

status: Показывает состояния файлов в рабочей директории и индексе: какие файлы изменены, но не добавлены в индекс; какие ожидают коммита в индексе. Вдобавок к этому выводятся подсказки о том, как изменить состояние файлов.

diff: Используется для вычисления разницы между любыми двумя Git деревьями.

difftool: Запускает внешнюю утилиту сравнения для показа различий в двух деревьях, на случай если вы хотите использовать что-либо отличное от встроенного просмотрщика git diff.

commit: Берёт все данные, добавленные в индекс с помощью git add, и сохраняет их слепок во внутренней базе данных, а затем сдвигает указатель текущей ветки на этот слепок.

reset: Используется в основном для отмены изменений. Она изменяет указатель HEAD и, опционально, состояние индекса.

rm: Используется в Git для удаления файлов из индекса и рабочей директории. Она похожа на git add с тем лишь исключением, что она удаляет, а не добавляет файлы для следующего коммита.

mv: Удобный способ переместить файл, а затем выполнить git add для нового файла и git rm для старого.

clean:Удаление мусора из рабочей директории. Это могут быть результаты сборки проекта или файлы конфликтов слияний.

2.2.4 GitHub

GitHub — крупнейший веб-сервис для хостинга IT-проектов и их совместной разработки. Основан на системе контроля версий Git и разработан на Ruby on Rails и Erlang компанией GitHub, Inc (ранее Logical Awesome).

Для выполнения практической работы создан репозиторий в аккаунте RandomRuslan на GitHub'e.

3 Перечень обучающих источников

В ходе данной работы необходимо найти современные (новее 2010 года) материалы, оказывающие поддержку в изучении параллельных вычислений в системах с общей памятью. Все источники должны быть разбиты на иностранные и русскоязычные, а внутри этих групп на платные и бесплатные книги, журналы и онлайн-курсы. Выполнить внутри каждой группы ранжирование источников по авторитетности, а также сформулировать данный критерий на основе различных показателей материалов.

3.1 Критерий сравнения

Так как все исследуемые материалы имеют различные особенности, то и критерии для сравнения и сортировки будут различными. Каждый критерий формируется на основе различных численных значений, играющих важную роль в оценке авторитетности данного материала. Среди них:

- *Цитирование* число ссылок на данную работу. Данное число вычисляется с использованием бесплатных общедоступных баз данных в Интернете, с помощью Google Scholar.
- h-Индекс (Индекс Хирша) наукометрический показатель, представляющий собой суммарное число ссылок на работы учёного. Критерий основан на учёте числа публикаций исследователя и числа цитирований этих публикаций, является количественной характеристикой продуктивности учёного за весь период научной деятельности. Учёный имеет индекс h, если h из его N статей цитируются как минимум h раз каждая, в то время как оставшиеся (N h) статей цитируются менее, чем h раз каждая. Данное число также вычисляется с использованием Google Scholar. Если у источника несколько авторов, берется наибольший из h-Индексов.
- ullet g-Uн $de\kappa c$ показатель, схожий с Индексом Хирша, рассчитываемый на

основе распределения цитирований, полученных публикациями ученого. Для множества статей, отсортированного в порядке убывания количества цитирований, которые получили эти статьи, g-индекс это наибольшее число, такое, что g самых цитируемых статей получили суммарно не менее g^2 цитирований. Данное число также вычисляется с использованием Google Scholar. Если у источника несколько авторов, берется наибольший из g-Индексов.

- *Оценка на Amazon.com* среднее значение выставленных пользователями оценок по пятибальной шкале для данного товара на сайте.
- SNIP (Source Normalized Impact per Paper), разработанный Центром CWTS, отражает влияние контекстной цитируемости журнала, что позволяет непосредственно сравнивать журналы различной тематики, принимая во внимание частоту, с которой авторы цитируют другие источники, скорость развития влияния цитаты и степень охвата литературы данного направления базой данных. Данное число вычисляется с использованием бесплатных общедоступных баз данных в Интернете, с помощью www.journalmetrics.com.
- SJR (SCImago Journal Rank) представляет собой рейтинг журналов, разработанный исследовательской группой SCImago. Он дает возможность оценить научный престиж работ ученых, исходя из количества весомых цитат на каждый документ. Журнал наделяет собственным «престижем» или статусом другие журналы, цитируя опубликованные в них материалы. Фактически это означает, что цитата из источника с относительно высоким показателем SJR имеет большую ценность, чем цитата из источника с более низким показателем SJR. Данное число вычисляется с использованием бесплатных общедоступных баз данных в Интернете, с помощью www.journalmetrics.com.

3.2 Иностранные материалы

3.2.1 Книги

Материалы данного раздела разделены на две категории:

- Посвящённы параллельным вычислениям, разбирая системы с общей памятью Таблица 1
- посвящённых параллельным вычислениям в целом (возможно, для систем с общей памятью) Таблица 2

Критерий оценивания формируется по формуле:

$$K = 10 * C * A + g/10^n$$
, где

С — количество цитирования книги

A — оценка книги на Amazon.com

g — g-Индекс автора

n — количество разрядов в наибольшем g-index

В итоге, критерий представляется десятичной дробью, где:

• Целая часть (10 * C * A)

Оценка на Amazon.com выступает в роли коэффициента количества цитирования. Если оценка отсутствует, коэффициент принимается равным 3-м, так как это среднее возможное значение оценки (т. е. оценка выше увеличит значение кол-ва цитирования). Умножение на 10 необходимо, чтобы произведение оставалось целым числом.

• Дробная часть $(g/10^n)$

Деление на степень десяти необходимо, чтобы сместить g-Индекс в дробную часть критерия. Является дополнительным, учитывается только в случае, если целые части равны.

Таблица 1

Название	Автор	Цитиро-	Амазон	g-index	Крите-
		вание			рий
The Art of Multiprocessor	Maurice	951	4.1	50	38991.5
Programming, Revised	Herlihy,				
Reprint 1st Edition	Nir Shavit				
An Introduction to Parallel	Peter	196	4.0	35	7840.35
Programming 1st Edition	Pacheco				
Multicore Application	Darryl	33	4.5	13	1485.13
Programming: for Windows,	Gove				
Linux, and Oracle Solaris					
(Developer's Library) 1st					
Edition					
Multicore and GPU	Gerassimos	4	3.7	26	148.26
Programming: An	Barlas				
Integrated Approach 1st					
Edition					
Programming Models	Pavan	-	-	39	0.39
for Parallel Computing	Balaji				
(Scientific and Engineering					
Computation)					
Shared Memory Application	Victor	0	-	18	0.18
Programming: Concepts	Alessandrini				
and Strategies in Multicore					
Application Programming					
1st Edition					

Multicore	Software	Robert	-	-	13	0.13
Development	Techniques:	Oshana				
Applications,	Tips, and					
Tricks (News	nes Pocket					
Books) 1st Edit	tion					

Таблица 2

Название	Автор	Цитиро-	Амазон	g-index	Крите-
		вание			рий
Parallel Programming:	Thomas	244	5	34	12200.34
for Multicore and Cluster	Rauber,				
Systems 2nd ed.	Gudula				
	Rünger				
Structured Parallel	Michael	162	4.7	50	7614.5
Programming: Patterns	McCool,				
for Efficient Computation	James				
1st Edition	Reinders,				
	Arch				
	Robison				
Masterkurs Parallele	Günther	27	_	47	810.47
und Verteilte Systeme:	Bengel,				
Grundlagen und	Christian				
Programmierung von	Baun,				
Multicore-Prozessoren,	Marcel				
Multiprozessoren, Cluster,	Kunze,				
Grid und Cloud (German	Karl-Uwe				
Edition) 2nd Edition	Stucky				

Parallel Con	nputer	Michel	0	4.8	50	0.5
Organization and Desi	ign 1st	Dubois,				
Edition		Murali				
		Annavaram,				
		Per				
		Stenström				
Parallel Computing fo	r Data	Norman	0	3	31	0.31
Science: With Exa	amples	Matloff				
\mid in R, C++ and	CUDA					
(Chapman & Hall/CR	RC The					
R Series)						

3.2.2 Бесплатные книги

Критерий оценивания формируется по той же формуле, что для платных иностранных книг.

Таблица 3

Название	Автор	Цитиро-	Амазон	g-index	Крите-
		вание			рий
Introduction to Parallel	Blaise	201	-	21	6030.21
Computing	Barney				
Introduction to High-	Victor	27	3.5	50	945.5
Performance Scientific	Eijkhout				
Computing					
The Practice of Parallel	Sergey	3	4.5	2	135.02
Programming	Babkin				
Programming on Parallel	Norm	0	-	7	0.07
Machines: GPU, Multicore,	Matloff				
Clusters and More					

3.2.3 Журналы

Критерий для данного раздела формируется на основе двух популярных показателях оценки журналов — SNIP и SJR. Так как они равноправны и близки по значению (не отличаются порядком), критерий формируется как простая сумма данных показателей.

Таблица 4

Название	Редактор	SNIP	SJR	Критерий
Transactions on Parallel and	David A.	3.892	2.017	5.909
Distributed Systems	Bader			
Parallel Computing	Jeffrey	2.141	1.232	3.373
	Hollingsworth			
Journal of Parallel and	Viktor	1.991	1.093	3.084
Distributed Computing	Prasanna			
Parallel Processing Letters	Selim G. Akl	1.453	0.499	1.952

3.2.4 Онлайн-курсы

Формирование данного критерия авторитетности осложнено тем, что разные ресурсы предоставляют разную информацию о курсах. Из общих показателей возможно выделить только g-Индекс и h-Индекс авторов и лекторов курсов, которые и лягут в основу критерия авторитетности. Так как они равноправны и близки по значению (не отличаются порядком), критерий формируется как простая сумма данных показателей.

Таблица 5

Название	Автор/Лектор	g-index	h-index	Крите-
				рий
Heterogeneous Parallel	Wen-mei W. Hwu	50	50	100
Programming				
Parallel Computing	Alan Edelman	50	38	88
Parallel Programming for	Kathy Yelick	8	36	44
Multicore				
Multicore Programming	Tom Van Cutsem	14	26	40
Parallel Programming for	Constantinos	10	27	37
Multicore Machines Using	Evangelinos			
OpenMP and MPI				
High Performance	Randy LeVeque	2	3	5
Scientific Computing				

3.3 Российские материалы

3.3.1 Книги

Из-за недостатка показателей для русскоязычных материалов, критерий оценивания формируется по формуле:

$$K = C + g/10^n$$
, где

С — количество цитирования книги

g — g-Индекс автора

 ${\rm n}-{\rm количество}$ разрядов в наибольшем g-index

ко в случае, если целые части равны.

В итоге, критерий представляется десятичной дробью, где:

• *Целая часть* (C) Число ссылок на данную работу, показывающее ее значимость.

• Дробная часть $(g/10^n)$ Деление на степень десяти необходимо, чтобы сместить g-Индекс в дробную часть критерия. Является дополнительным, учитывается толь-

Таблица 6

Название	Автор	Цитиро-	g-index	Критерий
		вание		
Технологии параллельного	Александр Анто-	18	25	18.25
программирования MPI и	нов			
OpenMP				
Современные языки и	Виктор Гергель	10	10	10.1
технологии параллельного				
программирования				

Инструменты параллель-	Кирилл Корня-	8	12	8.12
ного программирования в	ков, Валенти-			
системах с общей памятью	на Кустикова,			
	Иосиф Мееров,			
	Алексей Сиднев,			
	Александр Сы-			
	соев, Александр			
	Шишков			
Введение в параллельные	Михаил Якобов-	7	4	7.04
методы решения задач.	ский			
Учебное пособие				
Практикум по методам па-	Александр Стар-	3	23	3.23
раллельных вычислений	ченко, Евгений			
	Данилкин, Вален-			
	тина Николаева,			
	Сергей Проханов			
Основы параллельно-	Кирилл Богачев	2	9	2.9
го программирования.				
Учебное пособие				
Вычисления на многопро-	Алексей Генна-	0	3	0.3
цессорных компьютерах.	дьевич Абрамов			
Параллельные вычисле-				
ния на основе технологии				
OpenMP: учебное пособие				

3.3.2 Бесплатные книги

Критерий оценивания формируется по той же формуле, что для платных русскоязычных книг.

Таблица 7

Название		Автор	Цитиро-	g-index	Критерий
			вание		
Модели	параллельного	Федотов Илья	2	2	2.2
программирования					

3.3.3 Онлайн-курсы

Формирование данного критерия авторитетности осложнено тем, что разные ресурсы предоставляют разную информацию о курсах. Из общих показателей возможно выделить только g-Индекс авторов и лекторов курсов. В итоге, критерий формируется по формуле:

$$K = g + S/10^n$$
, где

g — g-Индекс автора

S—среднее арифметическое просмотров каждой лекции (округленное до целого) n—количество разрядов в наибольшем среднем просмотров В итоге, критерий представляется десятичной дробью, где:

- Целая часть (g)
 g-Индекс автора
- Дробная часть (S/10ⁿ)

 Деление на степень десяти необходимо, чтобы сместить среднем просмотров в дробную часть критерия. Является дополнительным, учитывается только в случае, если целые части равны.

Таблица 7

Название	Автор/Лектор	g-index	Среднее	Крите-
			прос-	рий
			мотров	
Высокопроизводительные	Гергель Виктор Пав-	15	_	15
вычисления для мно-	лович, Сысоев Алек-			
гопроцессорных много-	сандр Владимирович,			
ядерных систем	Козинов Евгений			
	Александрович, Лабу-			
	тина Анна Андреевна			

Инструменты парал-	Сысоев Александр	12	_	12
лельного программи-	Владимирович, Мее-			
рования в системах с	ров Иосиф Борисович,			
общей памятью	Сиднев Алексей Алек-			
	сандрович, Шишков			
	Александр Валерье-			
	вич, Корняков Кирилл			
	Владимирович, Ку-			
	стикова Валентина			
	Дмитриевна			
Параллельное програм-	Евгений Калишенко,	2	2416	2.2416
мирование	Алексей Злобин			
Параллельное програм-	Роман Елизаров	0	3174	0.3174
мирование				
Эффективные парал-	Александр Тискин	0	1573	0.1573
лельные алгоритмы:				
методика BSP				

4 Используемые сайты

- 1. ru.wikibooks.org
- 2. git-scm.com
- 3. Amazon.com
- 4. ozon.ru
- 5. Google Scholar
- 6. arxiv.org
- 7. journalmetrics.com
- 8. computing.llnl.gov
- 9. tacc.utexas.edu
- $10.\ free computer books.com$
- 11. freetechbooks.com
- 12. elsevier.com/journals
- 13. sciencedirect.com
- 14. computer.org
- 15. ocw.mit.edu
- 16. coursera.org
- 17. cs.berkeley.edu
- 18. coursetalk.com
- $19.\ soft.vub.ac.be/soft$

- 20. books.google.ru
- 21. elib.spbstu.ru
- 22. scanlibs.com
- 23. pp-book.narod.ru
- 24. lektorium.tv
- 25. hpcc.unn.ru