

«

»

“ ”

“ ”

**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Технологии и решения для анализа больших данных**

: 09.04.01

:
: 2, : 3

		3
1	()	3
2		108
3	, .	46
4	, .	18
5	, .	18
6	, .	0
7	, .	24
8	, .	2
9	, .	8
10	, .	62
11	(, ,)	
12		

(): 09.04.01

1420 30.10.2014 ., : 25.11.2014 .

: 1, ,

(): 09.04.01

, 7 20.06.2017

, 6 21.06.2017

:

,

:

,

:

.

1.

1.1

Компетенция ФГОС: ПК.9 способность проектировать системы с параллельной обработкой данных и высокопроизводительные системы и их компоненты; в части следующих результатов обучения:	
2.	
3.	
1.	
Компетенция НГТУ: ПК.20.В способность управлять средой функционирования объектов профессиональной деятельности; в части следующих результатов обучения:	
1.	

2.

2.1

	(
,	,)

.9. 2	
1.понятие матакомпьютинга и способы организации матакомпьютеров	; ;
.9. 3	
2.методы управления ресурсами и системное ПО организации вычислений на кластерах и матакомпьютерах	; ;
.9. 1	
3.разрабатывать прикладные программы для вычислений на кластерах и матакомпьютерах с использованием приемы, методов и языков параллельного программирования	; ;
.20. . 1	
4.использовать специализированные программные средства при решении профессиональных задач	; ;

3.

3.1

	,	.		
: 3				
:				
1.	0	2	1, 2	
2. Big Data .	0	2	1, 2	
:				
3. Hive	0	2	1, 2	
4. SQL vs NoSQL. NoSQL	2	2	1, 2	

:				
5.	Hadoop	2	2	1, 2
6.	MapReduce. MapReduce	0	4	1, 2
7.	HDFS	0	2	1, 2
8.	R: ,	2	2	3, 4

3.2

:				
: 3				
:				
1.	SQL vs NoSQL. NoSQL	4	4	4
:				
2.	R: ,	6	6	1, 2, 3, 4
3.	HDFS	4	4	4
4.	MapReduce. MapReduce	4	4	3, 4

4.

:				
: 3				
1		1, 2, 3	36	2
: . . Data Mining : / . . . - ., 2006. - 382 .: .				
2		1	18	5
: . . Data Mining : / . . . - ., 2006. - 382 .: .				
3		1, 4	8	1
: . . Data Mining : / . . . - ., 2006. - 382 .: .				

5.

- , (. 5.1).

5.1

-	
	e-mail
	e-mail
	e-mail
	;

6.

(), - 15- ECTS.

. 6.1.

2

6.1

: 3	
<i>Практические занятия:</i>	30
<i>РГЗ:</i>	30
<i>Экзамен:</i>	40

6.2

6.2

.9	2.	+	+
	3.	+	+
	1.	+	+
	.20. 1.	+	+

1

7.

1. ЭБС Znanium.com [Электронный ресурс] : электронно-библиотечная система. - [Россия], 2011. - Режим доступа: <http://znanium.com>. - Загл. с экрана.

2. ЭБС Znanium.com [Электронный ресурс] : электронно-библиотечная система. - [Россия], 2011. - Режим доступа: <http://znanium.com>. - Загл. с экрана.

1. Корытникова Н. В. Online Big Data как источник аналитической информации в online-исследованиях / Н. В. Корытникова // Социологические исследования. - 2015. - № 8. - С. 14-24.

2. Просто о больших данных : пер. с англ. / Джудит Гурвиц [и др.]. - Москва, 2015. - 393, [2] с. : ил. - Парал. тит. л. англ..

1. ЭБС НГТУ : <http://elibrary.nstu.ru/>
2. ЭБС «Издательство Лань» : <https://e.lanbook.com/>
3. ЭБС IPRbooks : <http://www.iprbookshop.ru/>
4. ЭБС "Znaniium.com" : <http://znaniium.com/>
5. :

8.

8.1

1. Чубукова И. А. Data Mining : учебное пособие / И. А. Чубукова. - М., 2006. - 382 с. : ил.

8.2

- 1 Microsoft Windows
- 2 Microsoft Office

9. -

1	(- , ,)	

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»

Кафедра автоматизированных систем управления

“УТВЕРЖДАЮ”
ДЕКАН АВТФ
к.т.н., доцент И.Л. Рева
“ ___ ” _____ г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

УЧЕБНОЙ ДИСЦИПЛИНЫ

Технологии и решения для анализа больших данных

Образовательная программа: 09.04.01 Информатика и вычислительная техника, магистерская
программа: Компьютерное моделирование систем

1. Обобщенная структура фонда оценочных средств учебной дисциплины

Обобщенная структура фонда оценочных средств по дисциплине Технологии и решения для анализа больших данных приведена в Таблице.

Формируемые компетенции	Показатели сформированности компетенций (знания, умения, навыки)	Темы	Этапы оценки компетенций	
			Мероприятия текущего контроля (курсовой проект, РГЗ(Р) и др.)	Промежуточная аттестация (экзамен, зачет)
ПК.20.В способность управлять средой функционирования объектов профессиональной деятельности	у1. использовать специализированные программные средства при решении профессиональных задач	Язык R: основы, особенности, применение	РГЗ	Экзамен, вопросы теста 1-15
ПК.9/П способность проектировать системы с параллельной обработкой данных и высокопроизводительные системы и их компоненты	з2. понятие матакомпьютинга и способы организации матакомпьютеров	SQL vs NoSQL. Модели данных NoSQL Архитектура Hadoop Архитектура HDFS Введение в проблематику управления большими данными Ведение в MapReduce. Компоненты MapReduce Обзор Hive и базовых принципов работы с БД Определение Big Data . Источники и применение Язык R: основы, особенности, применение	РГЗ	Экзамен, вопросы теста 1-15
ПК.9/П	з3. методы управления ресурсами и системное ПО организации вычислений на кластерах и матакомпьютерах	SQL vs NoSQL. Модели данных NoSQL Архитектура Hadoop Архитектура HDFS Введение в проблематику управления большими данными Ведение в MapReduce. Компоненты MapReduce Обзор Hive и базовых принципов работы с БД Определение Big Data . Источники и применение	РГЗ	Экзамен, вопросы теста 1-15
ПК.9/П	у1. разрабатывать прикладные программы для вычислений на кластерах и матакомпьютерах с использованием приемы, методов и языков параллельного программирования	Язык R: основы, особенности, применение	РГЗ	Экзамен, вопросы теста 1-15

2. Методика оценки этапов формирования компетенций в рамках дисциплины.

Промежуточная аттестация по дисциплине проводится в 3 семестре - в форме экзамена, который направлен на оценку сформированности компетенций ПК.20.В, ПК.9/П.

Экзамен проводится в форме письменного тестирования, варианты теста составляются из вопросов, приведенных в паспорте экзамена, позволяющих оценить показатели сформированности соответствующих компетенций

Кроме того, сформированность компетенций проверяется при проведении мероприятий текущего контроля, указанных в таблице раздела 1.

В 3 семестре обязательным этапом текущей аттестации является расчетно-графическое задание (работа) (РГЗ(Р)). Требования к выполнению РГЗ(Р), состав и правила оценки сформулированы в паспорте РГЗ(Р).

Общие правила выставления оценки по дисциплине определяются балльно-рейтинговой системой, приведенной в рабочей программе учебной дисциплины.

На основании приведенных далее критериев можно сделать общий вывод о сформированности компетенций ПК.20.В, ПК.9/П, за которые отвечает дисциплина, на разных уровнях.

Общая характеристика уровней освоения компетенций.

Ниже порогового. Уровень выполнения работ не отвечает большинству основных требований, теоретическое содержание курса освоено частично, пробелы могут носить существенный характер, необходимые практические навыки работы с освоенным материалом сформированы не достаточно, большинство предусмотренных программой обучения учебных заданий не выполнены или выполнены с существенными ошибками.

Пороговый. Уровень выполнения работ отвечает большинству основных требований, теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые виды заданий выполнены с ошибками.

Базовый. Уровень выполнения работ отвечает всем основным требованиям, теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые из выполненных заданий, возможно, содержат ошибки.

Продвинутый. Уровень выполнения работ отвечает всем требованиям, теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.

Паспорт экзамена

по дисциплине «Технологии и решения для анализа больших данных», 3 семестр

1. Методика оценки

Экзамен проводится в письменной форме, по тестам. В ходе экзамена преподаватель вправе задавать студенту дополнительные вопросы из общего перечня (п. 4).

Пример теста для экзамена

Вопрос № 1

Какие СУБД полностью полагаются на оперативную память при хранении информации:

- а) Oracle Exalytics
- б) SAP HANA
- в) BigTable
- г) HBase

Вопрос № 2

Для чего аналитику необходима «песочница»?

- а) Для высокопроизводительной аналитики за счёт использования оперативной памяти и inDB операций.
- б) Для хранения всех полученных от заказчика данных.
- в) Для построения отчётов о результатах анализа
- г) Для снижения затрат, связанных с репликацией данных

2. Критерии оценки

- Ответ на тест считается **неудовлетворительным**, если студент при ответе на вопросы не дает определений основных понятий, не способен показать причинно-следственные связи явлений, при решении задачи допускает принципиальные ошибки, оценка составляет *0 баллов*.
- Ответ на тест засчитывается на **пороговом** уровне, если студент при ответе на вопросы дает определение основных понятий, может показать причинно-следственные связи явлений, при решении задачи допускает непринципиальные ошибки, например, вычислительные, оценка составляет *10 баллов*.
- Ответ на тест засчитывается на **базовом** уровне, если студент при ответе на вопросы формулирует основные понятия, законы, дает характеристику процессов, явлений, проводит анализ причин, условий, может представить качественные характеристики процессов, не допускает ошибок при решении задачи, оценка составляет *20 баллов*.
- Ответ на тест засчитывается на **продвинутом** уровне, если студент при ответе на вопросы проводит сравнительный анализ подходов, проводит комплексный анализ, выявляет проблемы, предлагает механизмы решения, способен представить количественные характеристики определенных процессов, приводит конкретные примеры из практики, не допускает ошибок и способен обосновать выбор метода решения задачи, оценка составляет *40 баллов*.

3. Шкала оценки

Экзамен считается сданным, если сумма баллов по всем заданиям билета оставляет не менее 10 баллов (из 40 возможных).

В общей оценке по дисциплине экзаменационные баллы учитываются в соответствии с правилами балльно-рейтинговой системы, приведенными в рабочей программе дисциплины, следующим образом: общая оценка по дисциплине складывается из баллов, полученных на экзамене (максимум 40), и баллов за выполнение практической части (максимум 30) и РГЗ (максимум 30).

Таблица соответствия баллов, традиционной оценки и буквенной оценки ECTS приведена ниже.

Диапазон баллов рейтинга	Оценка ECTS	Традиционная шкала оценки	
96 -100	A	отлично	зачтено
93 - 96	A		
90 - 93	A-		
86 - 89	B		
83 - 86	B	хорошо	
80 - 83	B-		
76 - 79	C		
73 - 76	C		
70 - 73	C-		
66 - 69	D	удовлетворительно	зачтено
63 - 66	D		
60 - 63	D-		
50 - 59	E		
25 - 49	FX	неудовлетворительно	не зачтено
0 - 24	F		

4. Вопросы теста к экзамену по дисциплине «Технологии и решения для анализа больших данных»

Вариант 1

1 Hadoop это:

- а) Набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах.
- б) Распределённая СУБД, позволяющая обрабатывать большие данные.
- в) Язык выполнения заданий в парадигме MapReduce.
- г) Распределённая файловая система, предназначенная для хранения файлов большого объёма

2 Начиная с каких размеров данных обоснованно применение кластера Hadoop для хранения данных?

- а) 100Гб
- б) 1Тб
- в) 100Тб
- г) 1Пб

3 Пример благоразумного использования Hadoop

- а) Анализ 10 Гб данных.
- б) Ежедневное сохранение данных температуры, поступающих со всех городов России (по одному показанию на город, всего городов 1100 шт).
- в) Посекундное сохранение данных температуры, поступающих со всех городов России (по одному показанию на город, всего городов 1100 шт).
- г) Построение графика пульса пациента в реальном времени.

4 Выберите верное утверждение:

- а) Data Warehouse создаются для проверки гипотез при анализе больших данных.
- б) «Песочница» используется для снижения нагрузки на основной Data Warehouse.
- в) Каждый Data Warehouse должен содержать «песочницу».
- г) «Песочница» необходима для любого процесса аналитики.

5 Какие из следующих средств разумно использовать для анализа данных, представленного единственным csv - файлом размера более 100Гб:

- а) Hadoop
- б) Data Warehouse
- в) «Песочница»
- г) Python

6 Для чего аналитику необходима «песочница»?

- а) Для высокопроизводительной аналитики за счёт использования оперативной памяти и inDB операций.
- б) Для хранения всех полученных от заказчика данных.
- в) Для построения отчётов о результатах анализа
- г) Для снижения затрат, связанных с репликацией данных

7 В чём преимущество колоночно-ориентированных СУБД?

- а) Они позволяют выполнять более сложные SQL-запросы по сравнению с реляционными СУБД
- б) Они позволяют динамически дополнять содержание записей новыми полями
- в) Они имеют более гибкие возможности аналитики.
- г) Они позволяют эффективно делать межколоночные сравнения.

8 Какие СУБД полностью полагаются на оперативную память при хранении информации:

- а) Oracle Exalytics
- б) SAP HANA
- в) BigTable
- г) HBase

9 Какие из следующих технологий СУБД не используют принцип MapReduce

- а) Hadoop
- б) Cassandra
- в) HDInsight
- г) Redis

10 Выберите одно неверное высказывание про MapReduce:

- а) Интерфейс для массово-параллельной обработки данных, где вычисления производятся на узлах, где информация изначально была сохранена
- б) MapReduce –это две операции: распределения и сборки данных
- в) MapReduce был придуман разработчиками Hadoop
- г) MapReduce был анонсирован разработчиками Google

11 Принцип MapReduce состоит в том, чтобы

- а) Производить вычисления на узлах, где информация изначально была сохранена
- б) Использовать вычислительные мощности систем хранения
- в) Использовать функциональное программирование для решения задач массивно-параллельной обработки
- г) Использовать две операции - две операции: распределение и сборка данных

12 Отметьте неверное понимание Variety в контексте характеристик Big Data

- а) Высокая скорость генерирования данных.
- б) Разные типы данных в колонках таблиц реляционных СУБД.
- в) Разнообразие отраслей, являющихся источниками данных.
- г) Разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные.

13 Выберите неверное высказывание

- а) Большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных.
- б) Увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации.
- в) Удешевление систем хранения на единицу информации привело к росту рынка больших данных.
- г) Большое разнообразие источников данных

14 Перечислите четыре основных характеристики Big Data:

- а) Virtualization, Volume, Variability, Vehicle
- б) Variety, Velocity, Volume, Value
- в) Verification, Volume, Velocity, Visualization
- г) Video, Value, Variety, Volume

15 Отметьте те из вариантов, в которых данные структурированы:

- а) Данные о продажах компании, представленные в виде помесечных отчётов в формате MS Word.
- б) Таблица с ежедневными показаниями температуры помещения за год в файле формата csv.
- в) Текст книги А.С. Макаренко Педагогическая поэма, представленный в формате PDF.
- г) Библиотека фильмов, представленных в формате mpeg

Вариант 2

1 Выберите неверный ответ:

- а) Большие данные – это данные объёма свыше 1 Тб
- б) Проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна.
- в) Большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров.
- г) Большие данные как правило не структурированы.

2 Выберите верный ответ

- а) Большие данные – это обработка или хранение более 1 Тб информации.

- б) Проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна.
- в) Большие данные – это огромная PR-акция крупных вендоров и не более того.
- г) Большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект.

3 Отметьте значимые события, повлиявшие на формирование тренда больших данных:

- а) Разработка Hadoop
- б) Изобретение принципа MapReduce
- в) Разработка языка Python
- г) Победа Deepblue в матче с Г.Каспаровым.

4 Укажите фактор, способствовавший появлению тренда больших данных

- а) Маркетинговые кампании крупных корпораций
- б) Снижение издержек на хранение данных
- в) Появление новых технологий обработки потоковых данных
- г) Выпуск баз данных с обработкой данных в памяти

5 На каком из этапов процесса CRISP-DM происходит проверка гипотез?

- а) Понимание бизнеса (Business understanding)
- б) Понимание данных (Data Understanding)
- в) Моделирование (Modeling)
- г) Оценка (Evaluation)

6 Выберите верное утверждение:

- а) Data Warehouse создаются для проверки гипотез при анализе больших данных.
- б) «Песочница» используется для снижения нагрузки на основной Data Warehouse.
- в) Каждый Data Warehouse должен содержать «песочницу».
- г) «Песочница» необходима для любого процесса аналитики.

7 Отметьте неверное понимание Varietув контексте характеристик Big Data

- а) Высокая скорость генерирования данных.
- б) Разные типы данных в колонках таблиц реляционных СУБД.
- в) Разнообразие отраслей, являющихся источниками данных.
- г) Разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные.

8 Какие из следующих средств разумно использовать для анализа данных, представленных единственным csv - файлом размера более 100Гб:

- а) Hadoop
- б) Data Warehouse
- в) «Песочница»
- г) Python

9 Пример благоразумного использования Hadoop

- а) Анализ 10 Гб данных.
- б) Ежедневное сохранение данных температуры, поступающих со всех городов России (по одному показанию на город, всего городов 1100 шт).
- в) Посекундное сохранение данных температуры, поступающих со всех городов России (по одному показанию на город, всего городов 1100 шт).
- г) Построение графика пульса пациента в реальном времени.

10 Какие СУБД полностью полагаются на оперативную память при хранении информации:

- а) Oracle Exalytics
- б) SAP HANA
- в) BigTable
- г) HBase

11 Для чего аналитику необходима «песочница»?

- а) Для высокопроизводительной аналитики за счёт использования оперативной памяти и inDB операций.
- б) Для хранения всех полученных от заказчика данных.
- в) Для построения отчётов о результатах анализа
- г) Для снижения затрат, связанных с репликацией данных

12 Укажите фактор, способствовавший появлению тренда больших данных

- а) Маркетинговые кампании крупных корпораций
- б) Снижение издержек на хранение данных
- в) Появление новых технологий обработки потоковых данных
- г) Выпуск баз данных с обработкой данных в памяти

13 Hadoop это:

- а) Набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах.
- б) Распределённая СУБД, позволяющая обрабатывать большие данные.
- в) Язык выполнения заданий в парадигме MapReduce.
- г) Распределённая файловая система, предназначенная для хранения файлов большого объёма

14 Выберите неверное высказывание

- а) Большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных.
- б) Увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации.
- в) Удешевление систем хранения на единицу информации привело к росту рынка больших данных.
- г) Большое разнообразие источников данных

15 Начиная с каких размеров данных обоснованно применение кластера Hadoop для хранения данных?

- а) 100Гб
- б) 1Тб
- в) 100Тб
- г) 1Пб

**ТАБЛИЦА ПРАВИЛЬНЫХ ОТВЕТОВ
НА ТЕСТОВЫЕ ЗАДАНИЯ
ДЛЯ ПРОВЕРКИ ЗНАНИЙ СТУДЕНТОВ**
Вариант 1

Номер задания	Вариант ответа				Номер задания	Вариант ответа			
	А	Б	В	Г		А	Б	В	Г
1	+				13	+			
2			+	+	14		+		
3			+		15		+		
4		+			16				
5	+			+	17				
6	+			+	18				
7		+			19				
8	+	+			20				
9				+	21				
10	+		+		22				
11	+	+		+	23				
12	+	+	+		24				

**ТАБЛИЦА ПРАВИЛЬНЫХ ОТВЕТОВ
НА ТЕСТОВЫЕ ЗАДАНИЯ
ДЛЯ ПРОВЕРКИ ЗНАНИЙ СТУДЕНТОВ**

Вариант 2

Номер зада- ния	Вариант ответа				Номер зада- ния	Вариант ответа			
	А	Б	В	Г		А	Б	В	Г
1	+				13	+			
2		+			14	+			
3	+	+			15			+	+
4	+	+			16				
5			+		17				
6		+			18				
7	+	+	+		19				
8	+		+	+	20				
9			+		21				
10	+	+			22				
11	+				23				
12					24				

Паспорт расчетно-графического задания (работы)

по дисциплине «Технологии и решения для анализа больших данных», 3 семестр

1. Методика оценки

В рамках расчетно-графического задания (работы) по дисциплине студенты должны подготовить реферат на одну из предложенных тем (см. пункт 4). Целью реферата является ознакомление с материалом, расширяющим и углубляющим знания студента по выбранной теме.

Реферат должен быть объемом 10 – 20 страниц формата А4 и включать:

- титульный лист установленного образца (см. пункт 1.4.1);
- оглавление;
- основной материал;
- заключение (выводы);
- список использованных источников по ГОСТ 1.7-2003, 7.05-2008;
- необязательные приложения.

Основной материал реферата должен представлять обзор источников по выбранной теме.

1.1 Составление библиографии по теме

Список использованных источников характеризует уровень анализа выбранной темы и поэтому в него надо включать не только рекомендуемую литературу (учебную, учебно-методическую), но и статьи из научных журналов, сборников научных трудов, и профессиональных сайтов. Список литературы должен содержать адреса всех сайтов, материал которых использован в реферате (нежелательно использовать информацию из Википедии, поскольку существуют профессионально-ориентированные сайты). Желательно использование специализированной литературы на иностранном языке.

При использовании печатных источников должны приводиться полные библиографические данные источника. Для периодических изданий к реферату прикладываются ксерокопии первых страниц статей, а для книг – ксерокопии страниц с выходными данными

1.2 Сроки сдачи РГЗ

Текст реферата, оформленный в соответствии с требованиями, должен быть представлен в печатном виде на рецензию руководителю работы не позже, чем за 5 дней до начала экзаменационной сессии.

В случае необходимости работа возвращается для исправления выявленных

недочётов.

1.3 Оценка руководителем результата выполнения РГЗ даётся следующим образом:

- реферат удовлетворяет всем требованиям и сдан в срок - **20 баллов**;
- реферат удовлетворяет всем требованиям, но сдан после контрольного срока без уважительной причины – **10 баллов**;
- содержание реферата не соответствует заявленной теме или отсутствуют обязательные разделы – **0 баллов**.

1.4 Требования к оформлению текста реферата

Реферат является текстовым документом, поэтому оформление текста должно соответствовать ГОСТ 2.105-95.

1.4.1 Формат титульного листа реферата

МИНИСТЕРСТВО ОБР МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра Автоматизированных Систем Управления

РЕФЕРАТ

по дисциплине Технологии и решения для анализа больших данных
тему _____

Выполнил студент группы _____

Принял _____

Новосибирск 20-- г.

1.4.2 Требования к оформлению текста

Требования к оформлению текста реферата основаны на ГОСТ 2.105-95. ГОСТ 2.105-95 определяет общие требования к текстовым документам. Выдержки из него приведены только в учебных целях. Во всех других случаях необходимо ссылаться на официальный текст.

Текст реферата следует печатать, соблюдая следующие требования:

1. Текст набирается шрифтом Times New Roman размером (кеглем) 12 - 14, строчным, без выделения, с выравниванием по ширине.
2. Абзацный отступ должен быть одинаковым и равен по всему тексту 1,25 см.
3. Строки разделяются полуторным интервалом.
4. Поля страницы: верхнее и нижнее – 20 мм, левое не меньше 20 мм, правое - 10 мм.
5. Текст документа при необходимости разделяют на разделы и подразделы.
6. Разделы должны иметь порядковые номера в пределах всей пояснительной записки, обозначенные арабскими цифрами без точки и записанные с абзацевого отступа. Подразделы должны иметь нумерацию в пределах каждого раздела. Номер подраздела состоит из номеров раздела и подраздела, разделенных точкой. В конце номера подраздела точка не ставится. Разделы, как и подразделы, могут состоять из одного или нескольких пунктов.
7. Разделы, подразделы должны иметь заголовки. Пункты, как правило, заголовков не имеют. Заголовки должны четко и кратко отражать содержание разделов, подразделов.
8. Заголовки следует печатать с прописной буквы без точки в конце, не подчеркивая. Переносы слов в заголовках не допускаются. Если заголовок состоит из двух предложений, их разделяют точкой.
9. Содержанием, введение, заключение и список литературы не нумеруются.
10. Расстояние между заголовком и текстом при выполнении документа машинописным способом должно быть равно 3,4 интервалам. Расстояние между заголовками раздела и подраздела - 2 интервала.
11. Каждый раздел пояснительной записки рекомендуется начинать с нового листа (страницы).
12. Страницы пояснительной записки следует нумеровать арабскими цифрами, соблюдая сквозную нумерацию по всему тексту работ. Титульный лист включают в общую нумерацию страниц работ. Номер страницы на титульном листе не проставляют.
13. Материал, дополняющий текст пояснительной записки, допускается помещать в приложениях. Приложениями могут быть, например, графический материал, таблицы большого формата, расчеты, описания аппаратуры и приборов, описания алгоритмов и

программ задач, решаемых на ЭВМ и т.д. Приложение оформляют как продолжение данного документа на последующих его листах или выпускают в виде самостоятельного документа.

14. В тексте документа на все приложения должны быть даны ссылки. Приложения располагают в порядке ссылок на них в тексте документа, за исключением информационного приложения "Библиография", которое располагают последним.

15. Каждое приложение следует начинать с новой страницы с указанием наверху посередине страницы слова "Приложение" и его обозначения. Приложение должно иметь заголовок, который записывают симметрично относительно текста с прописной буквы отдельной строкой. Приложения обозначают заглавными буквами русского алфавита, начиная с А, за исключением букв Ё, З, Й, О, Ч, Ь, Ы, Ъ. После слова "Приложение" следует буква, обозначающая его последовательность. Допускается обозначение приложений буквами латинского алфавита, за исключением букв I и O. Если в документе одно приложение, оно обозначается "Приложение А".

16. Текст каждого приложения, при необходимости, может быть разделен на разделы, подразделы, пункты, подпункты, которые нумеруют в пределах каждого приложения. Перед номером ставится обозначение этого приложения. Приложения должны иметь общую с остальной частью документа сквозную нумерацию страниц.

17. В пояснительной записке после листа с заданием помещают содержание, включающее номера и наименования разделов и подразделов с указанием номеров листов (страниц). Содержание включают в общее количество листов, но страницы содержания не нумеруют.

18. Слово "Содержание" записывают в виде заголовка (симметрично тексту) с прописной буквы. Наименования, включенные в содержание, записывают строчными буквами, начиная с прописной буквы.

19. В конце пояснительной записки допускается приводить список литературы, которая была использована при её составлении. Выполнение списка и ссылки на него в тексте - по ГОСТ 7.05-2008. Список литературы включают в содержание документа.

20. В тексте пояснительной записки допускается использовать сокращения слов, определённые в ГОСТ 2.316-68. Все условные буквенные обозначения или знаки должны соответствовать принятому действующему законодательству и государственным стандартам.

21. В пояснительной записке следует применять стандартизованные единицы физических величин, их наименования и обозначения в соответствии с ГОСТ 8.417-81.

Оформление таблиц

1. Таблицы применяют для лучшей наглядности и удобства сравнения показателей. Название таблицы, при его наличии, должно отражать ее содержание, быть точным, кратким. Название следует помещать над таблицей.
2. При переносе части таблицы на ту же или другие страницы название помещают только над первой частью таблицы.
3. Цифровой материал, как правило, оформляют в виде таблиц в соответствии с рисунком 1.

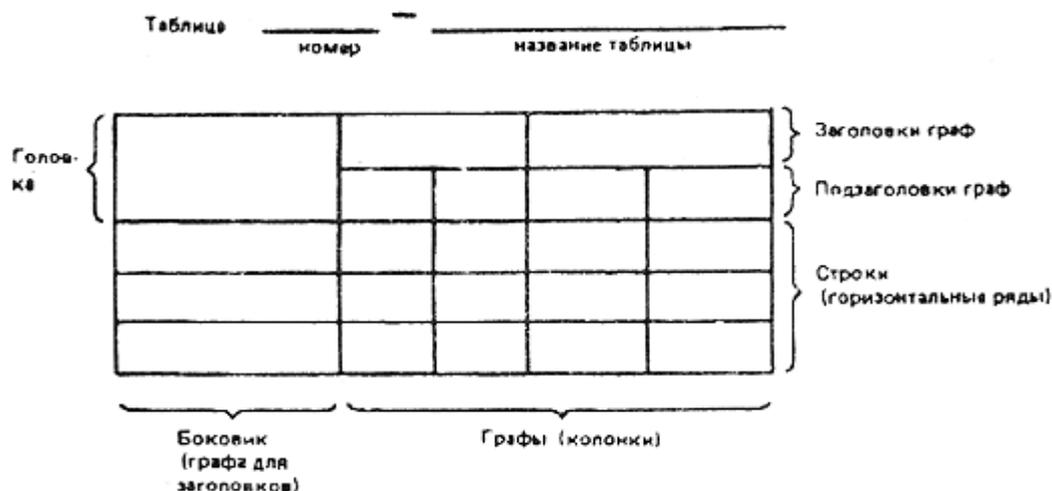


Рисунок 1

4. Таблицы, за исключением таблиц приложений, следует нумеровать арабскими цифрами сквозной нумерацией.
5. Таблицы каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением перед цифрой обозначения приложения. Если в документе одна таблица, она должна быть обозначена "Таблица 1" или "Таблица В.1", если она приведена в приложении В.
6. Допускается нумеровать таблицы в пределах раздела. В этом случае номер таблицы состоит из номера раздела и порядкового номера таблицы, разделенных точкой.
7. На все таблицы документа должны быть приведены ссылки в тексте документа, при ссылке следует писать слово "таблица" с указанием ее номера.
8. Заголовки граф и строк таблицы следует писать с прописной буквы, а подзаголовки граф - со строчной буквы, если они составляют одно предложение с заголовком, или с прописной буквы, если они имеют самостоятельное значение. В конце заголовков и подзаголовков таблиц точки не ставят. Заголовки и подзаголовки граф указывают в единственном числе.

9. Таблицы слева, справа и снизу, как правило, ограничивают линиями. Разделять заголовки и подзаголовки боковика и граф диагональными линиями не допускается.
10. Заголовки граф, как правило, записывают параллельно строкам таблицы. При необходимости допускается перпендикулярное расположение заголовков граф.
11. Головка таблицы должна быть отделена линией от остальной части таблицы.
12. Высота строк таблицы должна быть не менее 8 мм.
13. Таблицу, в зависимости от ее размера, помещают под текстом, в котором впервые дана ссылка на нее, или на следующей странице, а при необходимости, в приложении к документу.
14. Допускается помещать таблицу вдоль длинной стороны листа документа.
15. Если строки или графы таблицы выходят за формат страницы, ее делят на части, помещая одну часть под другой или рядом, при этом в каждой части таблицы повторяют ее головку и боковик. При делении таблицы на части допускается ее головку или боковик заменять соответственно номером граф и строк. При этом нумеруют арабскими цифрами графы и (или) строки первой части таблицы.
16. Слово "Таблица" указывают один раз слева над первой частью таблицы, над другими частями пишут слова "Продолжение таблицы" с указанием номера (обозначения) таблицы.

Оформление иллюстраций

1. Количество иллюстраций должно быть достаточным для пояснения излагаемого текста. Иллюстрации могут быть расположены как по тексту документа (возможно ближе к соответствующим частям текста), так и в конце его. Иллюстрации должны быть выполнены в соответствии с требованиями стандартов ЕСКД и СПДС. Иллюстрации, за исключением иллюстраций приложений, следует нумеровать арабскими цифрами сквозной нумерацией. Если рисунок один, то он обозначается "Рисунок 1".
2. Иллюстрации каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением перед цифрой обозначения приложения. Например - Рисунок А.3.
3. Допускается нумеровать иллюстрации в пределах раздела. В этом случае номер иллюстрации состоит из номера раздела и порядкового номера иллюстрации, разделенных точкой. Например - Рисунок 1.1.
4. При ссылках на иллюстрации следует писать "... в соответствии с рисунком 2" при сквозной нумерации и "... в соответствии с рисунком 1.2" при нумерации в пределах раздела.

5. Иллюстрации, при необходимости, могут иметь наименование и пояснительные данные (подрисующий текст). Слово "Рисунок" и наименование помещают после пояснительных данных и располагают следующим образом: Рисунок 1 – Схема алгоритма.

Сноски

1. Если необходимо пояснить отдельные данные, приведенные в документе, то эти данные следует обозначать надстрочными знаками сноски.
2. Сноски в тексте располагают с абзацного отступа в конце страницы, на которой они обозначены, и отделяют от текста короткой тонкой горизонтальной линией с левой стороны, а к данным, расположенным в таблице, в конце таблицы над линией, обозначающей окончание таблицы.
3. Знак сноски ставят непосредственно после того слова, числа, символа, предложения, к которому дается пояснение, и перед текстом пояснения.
4. Знак сноски выполняют арабскими цифрами со скобкой и помещают на уровне верхнего обреза шрифта (подобно этому: печатающее устройство²)
5. Нумерация сносок отдельная для каждой страницы.

1.4.3 Образец оформления списка литературы

Требования к оформлению списка литературы основаны на ГОСТ 7.1–2003. ГОСТ 7.1–2003 определяет общие требования и правила составления библиографического описания документа, его части или группы документов: набор областей и элементов библиографического описания, последовательность их расположения, наполнение и способ. Выдержки из него приведены только в учебных целях. Во всех других случаях необходимо ссылаться на официальный текст.

Расположение литературы в списке

Список использованных источников и литературы следует составлять в следующем порядке:

- официальные документы
официальные акты, документы общественных организаций, политических партий, постановления (в обратном - хронологическом порядке: вначале новые, затем принятые ранее)
- исторические источники
 - а) неопубликованные источники : архивные документы, рукописные материалы
 - б) опубликованные источники
 - статистические материалы, мемуары, дневники, письма.
 - научная литература

(Литература приводится в алфавите авторов и названий книг).

! Не следует отделять книги от статей. (Сведения о произведениях одного автора должны быть собраны вместе).

- справочная литература
 - иностранная литература
- (помещается после работы на русском языке, через пробел)
- библиографические указатели
 - описание электронных ресурсов.

Примеры библиографических ссылок

Книга одного автора

Семенов В. В. Философия: итог тысячелетий. Философская психология [Текст] / В. В. Семенов ; Рос. акад. наук, Пушин. науч. центр, Ин-т биофизики клетки, Акад. проблем сохранения жизни. – Пушино : ПНЦ РАН, 2000. – 64 с. – ISBN 5-201-14433-0.

Мюссе Л. Варварские нашествия на Западную Европу [Текст] : вторая волна / Люсьен Мюссе ; перевод с фр. А. Тополева ; [примеч. А. Ю. Карчинского]. – СПб. : Евразия, 2001. – 344 с. – ISBN 5-8071-0087-5.

Книга двух авторов

Бочаров И. Н. Кипренский [Текст] / Иван Бочаров, Юлия Глушакова. – 2-е изд., знач. доп. – М. : Молодая гвардия, 2001. – 390 с. – ISBN 5-235-02408-7.

Книга трёх авторов

Перроун П. Д. Создание корпоративных систем на базе Java 2 Enterprise Edition [Текст] : рук. разработчика : [пер. с англ.] / Поль Дж. Перроун, Венката С. Р. «Кришна», Р. Чаганти. – М. [и др.] : Вильямс, 2001. – 1179 с. – Перевод изд.: Building Java Enterprise systems with J2EE / Paul J. Perrone, Venkata S. R. (Krishna), R. Chaganti. Indianapolis. – ISBN 5-8459-0168-5.

Книга более 3 авторов

Культура доверия: мусульманско-христианско-иудейская традиция/ С.Боробенков [и др].- Казань: Школа, 2007.- 138с.

Учебники, учебные пособия

Агафонова Н. Н. Гражданское право [Текст] : учеб. пособие для вузов / Н. Н. Агафонова, Т. В. Богачева, Л. И. Глушкова ; под. общ. ред. А. Г. Калпина ; авт. вступ. ст. Н. Н. Поливаев ; М-во общ. и проф. образования РФ, Моск. гос. юрид. акад. – Изд. 2-е, перераб. и доп. – М. : Юристъ, 2002. – 542 с. – ISBN 5-7975-0223-2.

Материалы конференций

«Воспитательный процесс в высшей школе России», межвузовская науч.-практическая конф. (2001 ; Новосибирск). Межвузовская научно-практическая конференция «Воспитательный процесс в высшей школе России», 26–27 апр. 2001 г. [Текст] : [посвящ. 50-летию НГАВТ : материалы] / редкол.: А. Б. Борисов [и др.]. – Новосибирск : НГАВТ, 2001. – 157 с. – В надзаг. : Мэрия г. Новосибирска, Новосиб. обл. отд-ние Междунар. ассоц. по борьбе с наркоманией и наркобизнесом, Новосиб. гос. акад. вод. трансп.

Законодательные материалы

Российская Федерация. Конституция (1993). Конституция Российской Федерации [Текст] : офиц. текст. – М. : Маркетинг, 2001. – 39с. – ISBN 5-94462-025-0.

Правила

Правила безопасности при обслуживании гидротехнических сооружений и гидромеханического оборудования энергоснабжающих организаций [Текст] : РД 153-34.0-03.205–2001: утв. М-вом энергетики Рос. Федерации 13.04.01 : введ. в действие с 01.11.01. – М. : ЭНАС, 2001. – 158с. – В надзаг.: АО «ЕЭС России». – ISBN 5-93196-091-0.

Стандарты

ГОСТ Р 517721–2001. Аппаратура радиоэлектронная бытовая. Входные и выходные параметры и типы соединений. Технические требования [Текст]. – Введ. 2002–01–01. – М. : Изд-во стандартов, 2001. – IV, 27 с.

Сборник стандартов

Система стандартов безопасности труда : [сборник]. – М. : Изд-во стандартов, 2002. – 102, [1] с. : ил. ; 29 см. – (Межгосударственные стандарты). – Содерж.: 16 док. – 1231 экз.

Промышленные каталоги

Оборудование классных комнат общеобразовательных школ [Текст] : каталог / М-во образования РФ, Моск. гос. пед. ун-т. – М. : МГПУ, 2002. – 235 с.

Машина специальная листогибочная ИО 217М [Текст] : листок-каталог : разработчик и изготовитель Кемер. з-д электромонтаж. изделий. – М., 2002. – 3 л.

Многотомные издания

Казьмин В. Д. Справочник домашнего врача [Текст]. В 3 ч. Ч. 2. Детские болезни / Владимир Казьмин. – М. : АСТ : Астрель, 2002. – 503 с. – ISBN 5-17-011143-6

Депонированные научные работы

Разумовский В. А. Управление маркетинговыми исследованиями в регионе [Текст] / В. А. Разумовский, Д. А. Андреев ; Ин-т экономики города. – М., 2002. – 210 с. – Деп. в ИНИОН Рос. акад. наук 15.02.02, № 139876.

Социологическое исследование малых групп населения [Текст] / В. И. Иванов [и др.] ; М-во образования Рос. Федерации, Финансовая академия. – М., 2002. – 110 с. – Библиогр.: с. 108–109. – Деп. в ВИНТИ 13.06.02, № 145432.

Состояние и перспективы развития статистики печати Российской Федерации [Текст] : отчет о НИР (заключ.) : 06-02 / Рос. кн. палата ; рук. А. А. Джиго ; исполн.: В. П. Смирнова [и др.]. – М., 2000. – 250 с. – Инв. № 756600.

Диссертации

Вишняков И. В. Модели и методы оценки коммерческих банков в условиях неопределенности [Текст] : дис. ... канд. экон. наук : 08.00.13 : защищена 12.02.02 : утв. 24.06.02 / Вишняков Илья Владимирович. – М., 2002. – 234 с. – 04200204433.

Электронные ресурсы

Библиографические ссылки составляют как на электронные ресурсы в целом (электронные документы, базы данных, порталы, сайты, веб - страницы и т.д., так и на составные части электронных ресурсов.

Пашенный Н.Л. Императорское Училище Правоведения и Правоведы в годы мира, войны и смуты [Электронный ресурс] / созд. и дизайн: В. Рогге. [М.]. Электрон. дан. URL: <http://genrogge.ru/isj/index.htm> , свободный. Яз.рус. (дата обращения 17.01.2009).

Статья из сборника

Двинянинова Г. С. Комплимент : Коммуникативный статус или стратегия в дискурсе [Текст] / Г. С. Двинянинова // Социальная власть языка : сб. науч. тр. / Воронеж. межрегион. ин-т обществ. наук, Воронеж. гос. ун-т, Фак. романо-герман. истории. – Воронеж, 2001. – С. 101–106.

Боголюбов А. Н. О вещественных резонансах в волноводе с неоднородным заполнением [Текст] / А. Н. Боголюбов, А. Л. Делицын, М. Д. Малых // Вестн. Моск. ун-та. Сер. 3, Физика. Астрономия. – 2001. – № 5. – С. 23–25.

Статья из газеты

Михайлов С. А. Езда по-европейски [Текст] : система платных дорог в России находится в начал. стадии развития / Сергей Михайлов // Независимая газ. – 2002. – 17 июня.

Раздел, глава

Малый А. И. Введение в законодательство Европейского сообщества [Текст] / Ал. Малый // Институты Европейского союза : учеб. пособие / Ал. Малый, Дж. Кемпбелл, М. О'Нейл. – Архангельск, 2002. – Разд. 1. – С. 7–26.

Глазырин Б. Э. Автоматизация выполнения отдельных операций в Word 2000 [Текст] / Б. Э. Глазырин // Office 2000 : 5 кн. в 1 : самоучитель / Э. М. Берлинер, И. Б. Глазырина, Б. Э. Глазырин. – 2-е изд., перераб. – М., 2002. – Гл. 14. – С. 281–298.

2. Критерии оценки

- Работа считается **не выполненной**, если выполнены не все части РГЗ(Р), отсутствует анализ объекта, диагностические признаки не обоснованы, аппаратные средства не выбраны или не соответствуют современным требованиям, оценка составляет 0 баллов.
- Работа считается выполненной **на пороговом** уровне, если части РГЗ(Р) выполнены формально: анализ объекта выполнен без декомпозиции, диагностические признаки недостаточно обоснованы, аппаратные средства не соответствуют современным требованиям, оценка составляет 10 баллов.
- Работа считается выполненной **на базовом** уровне, если анализ объекта выполнен в полном объеме, признаки и параметры диагностирования обоснованы, алгоритмы разработаны, но не оптимизированы, аппаратные средства выбраны без достаточного обоснования, оценка составляет 20 баллов.
- Работа считается выполненной **на продвинутом** уровне, если анализ объекта выполнен в полном объеме, признаки и параметры диагностирования обоснованы, алгоритмы разработаны и оптимизированы, выбор аппаратных средств обоснован, оценка составляет 30 баллов.

3. Шкала оценки

В общей оценке по дисциплине баллы за РГЗ учитываются в соответствии с правилами балльно-рейтинговой системы, приведенными в рабочей программе дисциплины, следующим образом: общая оценка по дисциплине складывается из баллов, полученных на экзамене (максимум 40), и баллов за выполнение практической части (максимум 30) и РГЗ (максимум 30).

Таблица соответствия баллов, традиционной оценки и буквенной оценки ECTS приведена ниже.

Диапазон баллов рейтинга	Оценка ECTS	Традиционная шкала оценки
96 -100	A+	отлично зачтено
93 - 96	A	
90 - 93	A-	
86 - 89	B+	
83 - 86	B	хорошо
80 - 83	B-	
76 - 79	C+	
73 - 76	C	
70 - 73	C-	
66 - 69	D+	удовлетворительно зачтено
63 - 66	D	
60 - 63	D-	
50 - 59	E	
25 - 49	FX	неудовлетворительно не зачтено
0 - 24	F	

4. Примерный перечень тем РГЗ(Р)

1. Большие данные и бизнес-аналитика
2. Технологии обработки больших данных
3. Обзор решений от производителей ПО в области больших данных
4. Методы и платформы интеграции данных
5. Организации хранилищ больших данных
6. ETL - функции и процессы
7. Технологии многомерных баз данных

8. Миварные базы данных
9. Шаблоны интеграции корпоративных информационных систем
10. Задачи с интенсивным использованием данных
11. Анализ сообществ в социальных сетях
12. Методы и технологии обработки неструктурированных данных
13. Обзор архитектуры Hadoop
14. Нетрадиционные модели данных not only SQL (NOSQL)
15. Представление знаний
16. Консолидация данных
17. Технологии Business Intelligence
18. Параллельные машины баз данных
19. Парадигмы и манифесты баз данных
20. Управление большими данными и анализ потоковых данных
21. Технология MapReduce
22. Технологии Data Mining
23. Современные технологии высокопроизводительных вычислений
24. Грид-системы: проблемы и ограничения
25. Решения ORACLE для работы с большими данными
26. Модели данных для больших данных
27. Роль языков программирования Python и R в аналитике больших данных
28. Интеграция Hadoop с традиционными хранилищами данных
29. Большие данные и машинное обучение
30. Технология HIVE