МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ,
СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ ИМ. ПРОФ. М.А. БОНЧ-БРУЕВИЧА»
(СПбГУТ)

Кафедра Информационных управляющих систем
(полное наименование кафедры)

СПбГУТ))) Документ подписан простой электронной подписью
Сертификат: 00fd759708ffd39703
Владелец: Бачевский Сергей Викторович
Действителен с 15.03.2021 по 14.03.2026

УТВЕРЖДАЮ
Первый проректор - проректор по учебной работе
_____ Г.М. Машков
«07» 07 2021 г.

Регистрационный №_21.02/59-Д

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Анализ больших данных
(наименование дисциплины)

образовательная программа высшего образования

09.03.02 Информационные системы и технологии
(код и наименование направления подготовки / специальности)

бакалавр
(квалификация)

Дизайн графических и пользовательских интерфейсов информационных систем
(направленность / профиль образовательной программы)

очная форма
(форма обучения)

Санкт-Петербург

Рабочая программа дисциплины составлена на основе требований Федерального государственного образовательного стандарта высшего образования по направлению (специальности) подготовки «09.03.02 Информационные системы и технологии», утвержденным приказом Министерства образования и науки Российской Федерации от 19.09.2017 № 926, и в соответствии с рабочим учебным планом, утвержденным ректором университета.

1. Цели и задачи дисциплины

Целью преподавания дисциплины «Анализ больших данных» является:

освоение современных технологии анализа данных, учитывающих как особенности чрезвычайных размерностей, так и различных концепции многомернои модели данных. Дисциплина базируется на изучении различных аспектов взаимодеиствия информационных ресурсов в семантическом вебе, новеиших протоколов по передаче и поиску информации, процедур семантического анализа контента. Дисциплина «Анализ больших данных» должна обеспечивать формирование фундамента подготовки будущих специалистов в области информационных технологии, а также, создавать необходимую базу для успешного изучения последующих специальных дисциплин учебного плана. Она должна способствовать развитию творческих способностеи студентов, умению формулировать и решать задачи в области изучаемои специальности, умению творчески применять и самостоятельно повышать свои знания.

Эта цель достигается путем решения следующих(ей) задач(и):

фундаментализации, интенсификации и индивидуализации процесса обучения путем внедрения и эффективного использования достижении в области информационных и сетевых технологии. В результате изучения дисциплины у студентов должны сформироваться знания, умения и навыки построения, сопровождения и эксплуатации инфокоммуникационных систем и сетеи, способах представления, анализа и поиска информации, обработки больших данных и визуализации результатов обработки. Дисциплина является первои дисциплинои, в которои студенты изучают принципы семантического анализа и его применения для решения практических задач в области информационных технологии. Она находится на стыке дисциплин, обеспечивающих базовую и специальную подготовку студентов.

2. Место дисциплины в структуре образовательной программы

Дисциплина «Анализ больших данных» Б1.О.15 является дисциплиной обязательной части учебного плана подготовки бакалавриата по направлению «09.03.02 Информационные системы и технологии». Изучение дисциплины «Анализ больших данных» основывается на базе знаний, умений и компетенций, полученных студентами в ходе освоения школьных курсов.

3. Перечень планируемых результатов обучения, соотнесенных с планируемыми результатами освоения образовательной программы

Процесс изучения дисциплины направлен на формирование следующих компетенций:

Таблица 1

№ п/п	Код компетенции	Наименование компетенции

1	ОПК-3	Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности
2	УК-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений

Индикаторы достижения компетенций

Таблица 2

ОПК-3.1	Знать: принципы, методы и средства решения стандартных задач профессиональной деятельности на основе информационной и библиографической культуры с применением информационнокоммуникационных технологий и с учетом основных требований информационной безопасности.
ОПК-3.2	Уметь: решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационнокоммуникационных технологий и с учетом основных требований информационной безопасности.
ОПК-3.3	Иметь навыки: подготовки обзоров, аннотаций, составления рефератов, научных докладов, публикаций и библиографии по научноисследовательской работе с учетом требований информационной безопасности.
УК-2.1	Знать: виды ресурсов и ограничений для решения профессиональных задач; основные методы оценки разных способов решения задач; действующее законодательство и правовые нормы, регулирующие профессиональную деятельность.
УК-2.2	Уметь: проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения; анализировать альтернативные варианты для достижения намеченных результатов; использовать нормативно-правовую документацию в сфере профессиональной деятельности.
УК-2.3	Владеть: методиками разработки цели и задач проекта; методами оценки потребности в ресурсах, продолжительности и стоимости проекта; навыками работы с нормативно-правовой документацией.

4. Объем дисциплины и виды учебной работы

Очная форма обучения

Таблица 3

Вид учебной работы		**Всего часов**	**Семестры**
			5
Общая трудоемкость	5 ЗЕТ	180	180
Контактная работа с обучающимися		68.35	68.35
в том числе:			
Лекции		26	26
Практические занятия (ПЗ)		22	22
Лабораторные работы (ЛР)		18	18
Защита контрольной работы			-
Защита курсовой работы			-
Защита курсового проекта			-
Промежуточная аттестация		2.35	2.35
Самостоятельная работа обучающихся (СРС)		78	78
в том числе:			

Курсовая работа		-
Курсовой проект		-
И / или другие виды самостоятельной работы: подготовка к лабораторным работам, практическим занятиям, контрольным работам, изучение теоретического материала.	78	78
Подготовка к промежуточной аттестации	33.65	33.65
Вид промежуточной аттестации		Экзамен

5. Содержание дисциплины

5.1. Содержание разделов дисциплины.

Таблица 4

№ п/п	Наименование раздела (темы) дисциплины	Содержание раздела	№ семестра		
			очная	очно-заочная	заочная
1	Раздел 1. Большие данные и методы их обработки	Цели, задачи и структура курса. Большие данные и информация. Способы представления, хранения и обработки. Задачи классификации, индексирования и поиска.	5		
2	Раздел 2. Технологии обработки больших данных	HDFS (hadoop distributed file system)системы для работы с большими данными. Hadoop-технология. Среда обработки RStudio. Библиотеки для работы с Big Data.	5		
3	Раздел 3. Технологии data mining	Статистическая обработка, сжатие, кластеризация, факторный анализ, парсинг и визуализация данных.	5		
4	Раздел 4. Интеллектуальныи анализ текстовои информации с использованием машинного обучения	Принципы и методы семантико-синтаксического анализа естественно-языковых текстов. Методы кластеризации текстов. Методы автоматизации реферирования (аннотирования) текстов. Методы извлечения фрагментов для аннотации.	5		

5.2. Междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами.

Таблица 5

№ п/п	Наименование обеспечиваемых (последующих) дисциплин
1	Интеллектуальные информационные системы и технологии

5.3. Разделы дисциплин и виды занятий.

Очная форма обучения

Таблица 6

№ п/п	Наименование раздела (темы) дисциплин	Лекции	Практ. занятия	Лаб. занятия	Семинары	СРС	Всего часов
1	Раздел 1. Большие данные и методы их обработки	4				12	16

2	Раздел 2. Технологии обработки больших данных	6	6	6		22	40
3	Раздел 3. Технологии data mining	8	8	6		22	44
4	Раздел 4. Интеллектуальныи анализ текстовои информации с использованием машинного обучения	8	8	6		22	44
	Итого:	26	22	18	-	78	144

6. Лабораторный практикум

Очная форма обучения

Таблица 7

№ п/п	Номер раздела (темы)	Наименование лабораторной работы	Всего часов
1	2	Анализ данных в среде RStudio	6
2	3	Кластеризация и визуализация данных	6
3	4	Изучение методов семантического анализа	6
		Итого:	18

7. Практические занятия (семинары)

Очная форма обучения

Таблица 8

№ п/п	Номер раздела (темы)	Наименование практических занятий (семинаров)	Всего часов
1	2	Исследование пакета обработки Hadoop	6
2	3	Исследование методов факторного анализа	8
3	4	Исследование методов машинного обучения	8
		Итого:	22

8. Примерная тематика курсовых проектов (работ)

Рабочим учебным планом не предусмотрено

9. Самостоятельная работа

Очная форма обучения

Таблица 9

№ раздела дисцип-лины	Содержание СРС	Форма контроля	Всего часов
1	Изучение вопросов раздела 1	Тест	12
2	Изучение вопросов раздела 2	Тест	22
3	Изучение вопросов раздела 3	Тест	22

4	Изучение вопросов раздела 4	Тест	22
		Итого:	78

10. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

Для самостоятельной работы по дисциплине рекомендовано следующее учебно-методическое обеспечение:

- Положение о самостоятельной работе студентов в Санкт-Петербургском государственном университете телекоммуникаций им. проф. М.А. Бонч-Бруевича;
- рекомендованная основная и дополнительная литература;
- конспект занятий по дисциплине;
- слайды-презентации и другой методический материал, используемый на занятиях;
- методические рекомендации по подготовке письменных работ, требования к их содержанию и оформлению (реферат, эссе, контрольная работа) ;
- фонды оценочных средств;
- методические указания к выполнению лабораторных работ для студентов;

11. Фонд оценочных средств для проведения промежуточной аттестации обучающихся

Фонд оценочных средств разрабатывается в соответствии с локальным актом университета "Положение о фонде оценочных средств" и является приложением (Приложение А) к рабочей программе дисциплины.

Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине включает в себя:

- перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы;
- описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания;
- типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы;
- методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Для каждого результата обучения по дисциплине определяются показатели и критерии оценки сформированности компетенций на различных этапах их формирования, шкалы и процедуры оценивания.

12. Перечень основной и дополнительной литературы, необходимой для освоений дисциплины

12.1. Основная литература:

1. Филиппов, Феликс Васильевич.

Обработка информации в среде RStudio : [Электронный ресурс] : учебное пособие / Ф. В. Филиппов, А. Н. Губин ; рец.: А. В. Калинин, В. Л. Литвинов ; Федеральное агентство связи, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования "Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича". - СПб. : СПбГУТ, 2016. - 86 с. : ил. - 472.85 р.

2. Губин, Александр Николаевич.
   Основы SPARQL : [Электронный ресурс] : учебное пособие / А. Н. Губин, Ф. В. Филиппов ; рец.: В. В. Спиридонов, Д. В. Волошинов ; Федер. агентство связи, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования "Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича". - СПб. : СПбГУТ, 2017. - 75 с. : ил. - 435.64 р.
3. Филиппов, Феликс Васильевич.
   Моделирование нейронных сетей глубокого обучения : [Электронный ресурс] : учебное пособие / Ф. В. Филиппов ; рец.: А. В. Шевченко, Т. В. Матюхина ; Федеральное агентство связи, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования "Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича". - СПб. : СПбГУТ, 2019. - 79 с. : ил. - 491.69 р.

12.2. Дополнительная литература:

1. Филиппов, Феликс Васильевич.
   Программирование на языке R : [Электронный ресурс] : практикум / Ф. В. Филиппов ; рец. А. В. Шевченко ; Федеральное агентство связи, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования "Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича". - СПб. : СПбГУТ, 2018. - 78 с. : ил. - 1233.55 р.

13. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

- www.sut.ru
- lib.spbgut.ru/jirbis2_spbgut

14. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем.

14.1. Программное обеспечение дисциплины:

- Open Office

- Google Chrome

14.2. Информационно-справочные системы:

- ЭБС iBooks (https://ibooks.ru)
- ЭБС Лань (https://e.lanbook.com/)
- ЭБС СПбГУТ (http://lib.spbgut.ru)

15. Методические указания для обучающихся по освоению дисциплины

15.1. Планирование и организация времени, необходимого для изучения дисциплины

Важным условием успешного освоения дисциплины «Анализ больших данных» является создание системы правильной организации труда, позволяющей распределить учебную нагрузку равномерно в соответствии с графиком образовательного процесса. Большую помощь в этом может оказать составление плана работы на семестр, месяц, неделю, день. Его наличие позволит подчинить свободное время целям учебы, трудиться более успешно и эффективно. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Все задания, включая вынесенные на самостоятельную работу, рекомендуется выполнять непосредственно после соответствующего аудиторного занятия (лекции, практического занятия), что способствует лучшему усвоению материала, позволяет своевременно выявить и устранить «пробелы» в знаниях, систематизировать ранее пройденный материал, на его основе приступить к овладению новыми знаниями и навыками.

Система университетского обучения основывается на рациональном сочетании нескольких видов учебных занятий (в первую очередь, лекций и практических занятий), работа на которых обладает определенной спецификой.

15.2. Подготовка к лекциям

Знакомство с дисциплиной происходит уже на первой лекции, где от студента требуется не просто внимание, но и самостоятельное оформление конспекта. При работе с конспектом лекций необходимо учитывать тот фактор, что одни лекции дают ответы на конкретные вопросы темы, другие – лишь выявляют взаимосвязи между явлениями, помогая студенту понять глубинные процессы развития изучаемого предмета, как в истории, так и в настоящее время.

Конспектирование лекций – сложный вид вузовской аудиторной работы, предполагающий интенсивную умственную деятельность студента. Конспект является полезным тогда, когда записано самое существенное и сделано это самим обучающимся. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Целесообразно вначале понять основную мысль, излагаемую лектором, а затем записать ее. Желательно запись осуществлять на одной странице листа или оставляя поля, на которых позднее, при самостоятельной работе с конспектом, можно сделать дополнительные записи, отметить непонятные места.

Конспект лекции лучше подразделять на пункты, соблюдая красную строку.

Этому в большой степени будут способствовать вопросы плана лекции, предложенные преподавателям. Следует обращать внимание на акценты, выводы, которые делает лектор, отмечая наиболее важные моменты в лекционном материале замечаниями «важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек, подчеркивая термины и определения.

Целесообразно разработать собственную систему сокращений, аббревиатур и символов. Однако при дальнейшей работе с конспектом символы лучше заменить обычными словами для быстрого зрительного восприятия текста. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть теоретическим материалом.

15.3. Подготовка к практическим занятиям

Тщательное продумывание и изучение вопросов плана основывается на проработке пройденного материала (материала лекций, практических занятий), а затем изучения обязательной и дополнительной литературы, рекомендованной к данной теме.

Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы практикума, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Необходимо понимать, что невозможно во время аудиторных занятий изложить весь материал из-за лимита аудиторных часов, и при изучении дисциплины недостаточно конспектов занятий. Поэтому самостоятельная работа с учебниками, учебными пособиями, научной, справочной литературой, материалами периодических изданий и Интернета является наиболее эффективным методом получения дополнительных знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у студентов свое отношение к конкретной проблеме.

15.4. Рекомендации по работе с литературой

Работу с литературой целесообразно начать с изучения общих работ по теме, а также учебников и учебных пособий. Далее рекомендуется перейти к анализу монографий и статей, рассматривающих отдельные аспекты проблем, изучаемых в рамках курса, а также официальных материалов и неопубликованных документов (научно-исследовательские работы, диссертации), в которых могут содержаться основные вопросы изучаемой проблемы.

Работу с источниками надо начинать с ознакомительного чтения, т.е. просмотреть текст, выделяя его структурные единицы. При ознакомительном чтении закладками отмечаются те страницы, которые требуют более внимательного изучения. В зависимости от результатов ознакомительного чтения выбирается дальнейший способ работы с источником. Если для разрешения поставленной задачи требуется изучение некоторых фрагментов текста, то используется метод выборочного чтения. Если в книге нет подробного оглавления, следует обратить внимание ученика на предметные и именные указатели.

Избранные фрагменты или весь текст (если он целиком имеет отношение к теме) требуют вдумчивого, неторопливого чтения с «мысленной проработкой» материала. Такое чтение предполагает выделение: 1) главного в тексте; 2) основных аргументов; 3) выводов. Особое внимание следует обратить на то, вытекает тезис из аргументов или нет. Необходимо также проанализировать, какие из утверждений автора носят проблематичный, гипотетический характер и уловить скрытые вопросы.

Понятно, что умение таким образом работать с текстом приходит далеко не сразу. Наилучший способ научиться выделять главное в тексте, улавливать проблематичный характер утверждений, давать оценку авторской позиции – это сравнительное чтение, в ходе которого студент знакомится с различными мнениями по одному и тому же вопросу, сравнивает весомость и доказательность аргументов сторон и делает вывод о наибольшей убедительности той или иной позиции.

Если в литературе встречаются разные точки зрения по тому или иному вопросу из-за сложности прошедших событий и правовых явлений, нельзя их отвергать, не разобравшись. При наличии расхождений между авторами необходимо найти рациональное зерно у каждого из них, что позволит глубже усвоить предмет изучения и более критично оценивать изучаемые вопросы. Знакомясь с особыми позициями авторов, нужно определять их схожие суждения, аргументы, выводы, а затем сравнивать их между собой и применять из них ту, которая более убедительна.

Следующим этапом работы с литературными источниками является создание конспектов, фиксирующих основные тезисы и аргументы. Можно делать записи на отдельных листах, которые потом легко систематизировать по отдельным темам изучаемого курса. Другой способ – это ведение тематических тетрадей-конспектов по одной какой-либо теме. Большие специальные работы монографического характера целесообразно конспектировать в отдельных тетрадях. Здесь важно вспомнить, что конспекты пишутся на одной стороне листа, с полями и достаточным для исправления и ремарок межстрочным расстоянием (эти правила соблюдаются для удобства редактирования). Если в конспектах приводятся цитаты, то непременно должно быть дано указание на источник (автор, название, выходные данные, № страницы). Впоследствии эта информации может быть использована при написании текста реферата или другого задания.

Таким образом, при работе с источниками и литературой важно уметь:

- сопоставлять, сравнивать, классифицировать, группировать, систематизировать информацию в соответствии с определенной учебной задачей;
- обобщать полученную информацию, оценивать прослушанное и прочитанное;
- фиксировать основное содержание сообщений; формулировать, устно и письменно, основную идею сообщения; составлять план, формулировать тезисы;
- готовить и презентовать развернутые сообщения типа доклада;
- работать в разных режимах (индивидуально, в паре, в группе), взаимодействуя друг с другом;
- пользоваться реферативными и справочными материалами;
- контролировать свои действия и действия своих товарищей, объективно оценивать свои действия;
- обращаться за помощью, дополнительными разъяснениями к преподавателю,

другим студентам;

- пользоваться лингвистической или контекстуальной догадкой, словарями различного характера, различного рода подсказками, опорами в тексте (ключевые слова, структура текста, предваряющая информация и др.);
- использовать при говорении и письме перифраз, синонимичные средства, слова-описания общих понятий, разъяснения, примеры, толкования, «словотворчество»
- повторять или перефразировать реплику собеседника в подтверждении понимания его высказывания или вопроса;
- обратиться за помощью к собеседнику (уточнить вопрос, переспросить и др.);
- использовать мимику, жесты (вообще и в тех случаях, когда языковых средств не хватает для выражения тех или иных коммуникативных намерений).

15.5. Подготовка к промежуточной аттестации

При подготовке к промежуточной аттестации целесообразно:

- внимательно изучить перечень вопросов и определить, в каких источниках находятся сведения, необходимые для ответа на них;
- внимательно прочитать рекомендованную литературу;
- составить краткие конспекты ответов (планы ответов).

16. Материально-техническое обеспечение дисциплины

Таблица 10

№ п/п	Наименование специализированных аудиторий и лабораторий	Наименование оборудования
1	Лекционная аудитория	Аудио-видео комплекс
2	Аудитории для проведения групповых и практических занятий	Аудио-видео комплекс
3	Компьютерный класс	Персональные компьютеры
4	Аудитория для курсового и дипломного проектирования	Персональные компьютеры
5	Аудитория для самостоятельной работы	Компьютерная техника
6	Читальный зал	Персональные компьютеры