

Министерство культуры Российской Федерации

федеральное государственное бюджетное образовательное учреждение
высшего образования

«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ИНСТИТУТ КИНО И ТЕЛЕВИДЕНИЯ»



ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Е. В. САЗОНОВА
ректор

Сертификат: 00eec2e5b252a0885bc682f9fa99feef8b

Основание: УТВЕРЖДАЮ

Дата утверждения: 19 июня 2024 г.

Рабочая программа дисциплины

«Большие данные»

Наименование ОПОП: Интеллектуальные системы и технологии в
медиаискусстве

Направление подготовки: 09.03.02 Информационные системы и технологии

Форма обучения: очная

Факультет: медиатехнологий

Кафедра: аудиовизуальных систем и технологий

Общая трудоемкость дисциплины составляет 180 академ. час. / 5 зач.ед.
в том числе: контактная работа: 68,5 час.
самостоятельная работа: 111,5 час.

Вид(ы) текущего контроля	Семестр (курс)
активная работа на занятии	6
практикум (выполнение и защита лабораторных работ)	6
практикум (выполнение практических работ)	6
присутствие на занятии	6
Вид(ы) промежуточной аттестации, курсовые работы/проекты	Семестр (курс)
экзамен	6

Рабочая программа дисциплины «Большие данные» составлена:

— в соответствии с требованиями Федеральным государственным образовательным стандартом высшего образования — Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.02 Информационные системы и технологии (приказ Минобрнауки России от 19.09.2017 г. № 926)
— на основании учебного плана и карты компетенций основной профессиональной образовательной программы «Интеллектуальные системы и технологии в медиаискусстве» по направлению подготовки 09.03.02 Информационные системы и технологии

Составитель(и):

Сорокина И.В., доцент кафедры аудиовизуальных систем и технологий, канд. пед. наук

Рецензент(ы):

Горбунова И.Б., профессор кафедры информатизации образования ФГБОУ «Российский государственный педагогический университет им. А.И. Герцена» , д-р пед. наук

Рабочая программа дисциплины рассмотрена и одобрена на заседании кафедры аудиовизуальных систем и технологий

Рабочая программа дисциплины одобрена Советом факультета медиатехнологий

СОГЛАСОВАНО

Руководитель ОПОП

А.И. Ходанович

Начальник УМУ

С.Л. Филипенкова

УКАЗАННАЯ ЛИТЕРАТУРА ИМЕЕТСЯ В НАЛИЧИИ В БИБЛИОТЕКЕ ИНСТИТУТА ИЛИ ЭБС

Заведующий библиотекой Н.Н. Никитина

1. ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЙ РАЗДЕЛ

1.1. Цели и задачи дисциплины

Цель(и) дисциплины:

освоение современных технологии анализа данных, учитывающих как особенности чрезвычайных размерностей, так и различных концепции многомерной модели данных. Дисциплина базируется на изучении различных аспектов взаимодействия информационных ресурсов в семантическом вебе, новейших 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 способностей студентов, умению формулировать и решать задачи в области изучаемой специальности, умению творчески применять и самостоятельно повышать свои знания.

Задачи дисциплины:

фундаментализации, интенсификации и индивидуализации процесса обучения путем внедрения и эффективного использования достижения 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 сетей, способах представления, анализа и поиска информации, обработки больших данных и визуализации результатов обработки. Дисциплина является первой дисциплиной, в которой 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.

1.2. Место и роль дисциплины в структуре ОПОП ВО

Дисциплина относится к обязательной части Блока 1 «Дисциплины (модули)».

Дисциплина основывается на знаниях и умениях, приобретенных в ходе изучения предшествующих дисциплин/прохождения практик и взаимосвязана с параллельно изучаемыми дисциплинами:

Анализ данных и информационный поиск

Электротехника и электроника

Архитектура ПК и информационных систем

Научно-исследовательская работа (получение первичных навыков научно-исследовательской работы)

Высшая математика

Основы государственной культурной политики Российской Федерации

Управление данными

Физика

Ознакомительная практика

История информационных и коммуникационных технологий

Правовое обеспечение профессиональной деятельности

Физическая картина мира

Интернет вещей

Компьютерная геометрия и графика

Методы и средства проектирования информационных систем и технологий

Администрирование информационных систем в медиаискусстве

Алгоритмизация и программирование
Основы проектирования киберфизических систем в медиаискусстве
Информационные и коммуникационные технологии в медиаиндустрии
Теория информации, данные, знания
Основы мультимедиа
Освоение дисциплины необходимо как предшествующее для дисциплин и/или практик:
Выполнение и защита выпускной квалификационной работы
Подготовка к сдаче и сдача государственного экзамена
Методы информационно-аналитической работы в медиаискусстве
Научно-исследовательская работа
Преддипломная практика

1.3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОПОП ВО

Универсальные компетенции

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.

УК-2.2 — Выбирает оптимальные способы решения задач в рамках поставленной цели, составляет план реализации из действующих ресурсов и ограничений.

Знает: действующие правовые нормы

Умеет: выбирать оптимальные способы решения задач в рамках поставленной цели

Владеет: навыками анализа задач профессиональной деятельности, составления планов реализации

Общепрофессиональные компетенции

ОПК-1 — Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности;

ОПК-1.2 — Выявляет естественно-научную сущность проблем, возникающих в процессе решения задач инженерной деятельности, привлекает для их устранения соответствующий физико-математический аппарат.

Знает: основные естественнонаучные и общеинженерные понятия

Умеет: выявлять естественно-научную сущность проблем, возникающих в процессе решения задачи инженерной деятельности

Владеет: навыками работы с физико-математическим аппаратом для решения задач профессиональной деятельности

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;

ОПК-3.1 — Использует информационно-коммуникационные технологии для поиска, хранения, обработки, анализа и представления в требуемом формате информации из различных источников и баз данных.

Знает: принципы построения современных информационных систем

Умеет: использовать современные ресурсы и информационные технологии; анализировать и интерпретировать информацию, содержащуюся в отчетности и использовать полученные сведения для задач профессиональной деятельности

Владеет: методами практического использования современных информационно-коммуникационных технологий с учетом основных требований информационной безопасности

2. СТРУКТУРА, ОБЪЕМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ) В КОМПЕТЕНТНОСТНОМ ФОРМАТЕ

2.1. Структура и трудоемкость учебной дисциплины

Общая трудоемкость дисциплины составляет 180 академ. час. / 5 зач.ед.

в том числе: контактная работа: 68,5 час.
самостоятельная работа: 111,5 час.

Вид(ы) промежуточной аттестации, курсовые работы/проекты	Семестр (курс)
экзамен	6

Распределение трудоемкости по периодам обучения:

Семестр	6	Итого
Лекции	16	16
Практические	16	16
Лабораторные	32	32
Консультации	2	2
Самостоятельная работа	78	78
Самостоятельная работа во время сессии	33,5	33,5
Итого	177,5	177,5

2.2. Содержание учебной дисциплины

Тема 1. Большие данные. Основы анализа.

Цели, задачи и структура курса. Экосистема больших данных. Задачи и стандарты.

Тема 2. Технологии хранения и обработки больших данных

Системы распределенного хранения и обработки данных. Фреймворки. Hadoop. HDFS (hadoop distributed file system). MapReduce. Языки программирования и интегрированные среды разработки. Язык программирования R. Среда разработки RStudio. Библиотеки для работы с большими данными.

Тема 3. Интеллектуальный анализ данных

Классификация. Кластеризация. Поиск ассоциативных правил. Визуализация.

Тема 4. Интеллектуальный анализ текстовых данных

Извлечение ключевых понятий. Парсинг. Классификация и кластеризация текстов. Аннотирование текстов.

3. РАСПРЕДЕЛЕНИЕ ЧАСОВ ПО ТЕМАМ И ВИДАМ УЧЕБНОЙ РАБОТЫ

№ п/п	Наименование раздела, (отдельной темы)	Лекции	Лекции с использованием ДОТ	Лабораторные работы	Практические занятия	Практические с использованием ДОТ	Индивидуальные занятия	Итого
1	Большие данные. Основы анализа.	4	0	8	4	0	0	16
2	Технологии хранения и обработки больших данных	4	0	8	4	0	0	16
3	Интеллектуальный анализ данных	4	0	8	4	0	0	16
4	Интеллектуальный анализ текстовых данных	4	0	8	4	0	0	16
	ВСЕГО	16	0	32	16	0	0	64

4. ЛАБОРАТОРНЫЕ ЗАНЯТИЯ

№ п/п	Наименование лабораторных работ	Трудоемкость (час.)
1	Тема: «Большие данные. Основы анализа.».	2
2	Тема: «Большие данные. Основы анализа.».	2
3	Тема: «Большие данные. Основы анализа.».	2
4	Тема: «Большие данные. Основы анализа.».	2
5	Тема: «Технологии хранения и обработки больших данных.».	2
6	Тема: «Технологии хранения и обработки больших данных.».	2
7	Тема: «Технологии хранения и обработки больших данных.».	2
8	Тема: «Технологии хранения и обработки больших данных.».	2
9	Тема: «Интеллектуальный анализ данных.».	2
10	Тема: «Интеллектуальный анализ данных.».	2
11	Тема: «Интеллектуальный анализ данных.».	2
12	Тема: «Интеллектуальный анализ данных.».	2
13	Тема: «Интеллектуальный анализ текстовых данных.».	2

14	Тема: «Интеллектуальный анализ текстовых данных».	2
15	Тема: «Интеллектуальный анализ текстовых данных».	2
16	Тема: «Интеллектуальный анализ текстовых данных».	2

5. ПРАКТИЧЕСКИЕ ЗАНЯТИЯ (СЕМИНАРЫ)

№ п/п	Тематика практических занятий (семинаров)	Трудоемкость (час.)
1	Тема: «Большие данные. Основы анализа.».	2
2	Тема: «Большие данные. Основы анализа.».	2
3	Тема: «Технологии хранения и обработки больших данных».	2
4	Тема: «Технологии хранения и обработки больших данных».	2
5	Тема: «Интеллектуальный анализ данных».	2
6	Тема: «Интеллектуальный анализ данных».	2
7	Тема: «Интеллектуальный анализ текстовых данных».	2
8	Тема: «Интеллектуальный анализ текстовых данных».	2

6. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Оценивание результатов обучения по дисциплине осуществляется в ходе текущего контроля и промежуточной аттестации с использованием балльно-рейтинговой системы.

Оценочные средства в полном объеме представлены в Фонде оценочных средств по дисциплине «Большие данные».

Предусмотрены следующие формы и процедуры текущего контроля и промежуточной аттестации:

Вид(ы) текущего контроля	Семестр (курс)
активная работа на занятии	6
практикум (выполнение и защита лабораторных работ)	6
практикум (выполнение практических работ)	6
присутствие на занятии	6
Вид(ы) промежуточной аттестации, курсовые работы/проекты	Семестр (курс)
экзамен	6

6.1. Оценочные средства для входного контроля (при наличии)

Входной контроль отсутствует.

6.2. Оценочные средства для текущего контроля успеваемости

6.3. Оценочные средства для промежуточной аттестации

Перечень вопросов к экзамену:

Экосистема больших данных.

Задачи и стандарты.

Системы распределенного хранения и обработки данных.

Фреймворки.

Hadoop.

HDFS (hadoop distributed file system).

MapReduce.

Языки программирования и интегрированные среды разработки.

Язык программирования R.

Среда разработки RStudio.

Библиотеки для работы с большими данными.

Классификация.

Кластеризация.

Поиск ассоциативных правил.

Визуализация.

Извлечение ключевых понятий.

Парсинг.

Классификация и кластеризация текстов.

Аннотирование текстов.

6.4. Балльно-рейтинговая система

Оценка успеваемости с применением балльно-рейтинговой системы заключается в накоплении обучающимися баллов за активное, своевременное и качественное участие в определенных видах учебной деятельности и выполнении учебных заданий в ходе освоения дисциплины.

Конкретные виды оцениваемой деятельности	Количество баллов за 1 факт (точку) контроля	Количество фактов (точек) контроля	Баллы (максимум)
Обязательная аудиторная работа			
Практикум (Выполнение практических работ)	2	8	16
Активная работа на занятии	6	1	6
Присутствие на занятии	1	32	32
Практикум (Выполнение и защита лабораторных работ)	1	16	16
ИТОГО в рамках текущего контроля	70 баллов		
ИТОГО в рамках промежуточной аттестации	30 баллов		
ВСЕГО по дисциплине за семестр	100 баллов		

Итоговая оценка по дисциплине выставляется на основе накопленных баллов в ходе текущего контроля и промежуточной аттестации в соответствии с таблицей:

Система оценивания результатов обучения по дисциплине

Шкала по БРС	Отметка о зачете	Оценка за экзамен, зачет с оценкой
85 – 100	зачтено	отлично
70 – 84		хорошо
56 – 69		удовлетворительно
0 – 55	не зачтено	неудовлетворительно

7. УЧЕБНО-МЕТОДИЧЕСКОЕ, ИНФОРМАЦИОННО-ТЕХНИЧЕСКОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Обучающиеся из числа лиц с ограниченными возможностями здоровья обеспечиваются печатными и (или) электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

7.1. Литература

1. Форман, Д. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; пер. с англ. Соколовой А. - Москва :Альпина Пабли., 2016. - 461 с. ISBN 978-5-9614-5032-3. - Текст : электронный. – Режим доступа: по подписке.
<https://znanium.com/catalog/product/551044>

7.2. Интернет-ресурсы

1. Google Trends: Google Trends показывает частоту ввода определенных поисковых терминов на нескольких языках по отношению к общему количеству поисковых запросов по всему миру.
2. <https://bigdataanalyticsnews.com/>

7.3. Лицензионное и свободно распространяемое программное обеспечение

Microsoft Office
Microsoft Windows

7.4. Профессиональные базы данных и информационные справочные системы

Электронный каталог библиотеки СПбГИКиТ. <https://www.gukit.ru/lib/catalog>
Электронная библиотечная система издательства «ЛАНЬ». <http://e.lanbook.com>
Электронная библиотечная система «Айбукс-ру». <http://ibooks.ru>

7.5. Материально-техническое обеспечение

Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы
Учебная аудитория	Рабочее место преподавателя, оборудованное компьютером и мультимедийным проектором. Рабочие места обучающихся. Доска (интерактивная доска) и/или экран.
Помещение для самостоятельной работы обучающихся	Рабочие места обучающихся оборудованные компьютерами с подключением к сети «Интернет» и доступом в электронную информационно-образовательную среду института.

8. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Важным условием успешного освоения дисциплины 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, лабораторной работы), 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, работа на которых обладает определенной спецификой.

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.

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 информация 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, словаописания 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).

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).