

**Федеральное государственное автономное образовательное  
учреждение высшего образования  
«Московский физико-технический институт  
(национальный исследовательский университет)»**

**УТВЕРЖДЕНО**  
**Проректор по учебной работе**

**А.А. Воронов**

**по дисциплине:** **Рабочая программа дисциплины (модуля)**  
Анализ данных в истории науки

**программа аспирантуры:** Биологические науки

**курс:** учебно-научный центр гуманитарных и социальных наук  
1

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 48 час.

Всего часов: 78, всего зач. ед.: 2

Количество контрольных работ, заданий: 2

Программу составил: А.А. Костин, канд. филос. наук

Программа обсуждена на заседании учебно-научного центра гуманитарных и социальных наук 26.08.2022

## Аннотация

Курс рассчитан на аспирантов первого года обучения, желающих приобрести и углубить познания в истории науки и применении к этой области методов анализа текстовых данных.

Большие массивы текстовых данных и общая информационная перегруженность современных областей знания затрудняют работу с текстовой базой. Принципиальным этот вопрос является для областей знания, в которых текст является основным источником знания. Однако и такие эмпирические области, как физика и биология, требуют тщательного исследования источниковедческой базы, объем которой часто превышает человеческие возможности.

Вместе с тем, достижения в таких областях, как обработка естественного языка и глубокое обучение нейронных сетей, делают возможной продуктивную работу с текстами любого объема и сложности. Это и предполагается продемонстрировать на материале курса.

На примере исторических исследований науки рассматриваются основные методы обработки, анализа и визуализации данных: векторное представление текстовых данных, решение задач на их семантическое сходство, анализ и визуализация обработанных данных в виде сложных сетей.

## 1. Цели и задачи

### Цель дисциплины

предоставить слушателям теоретический и фактический материал, раскрывающий особенности различных областей истории науки.

### Задачи дисциплины

- формирование базовых знаний в области истории науки, сопоставление исторических источников, работа с различными вариантами концептуализации исторического материала;
- формирование у учащихся цифровых компетенций для выполнения ими исследований в области истории науки с применением современных методов анализа данных.

## 2. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- основные концепции истории науки;
- передовые достижения в области обработки естественного языка.

уметь:

- работать с историческими источниками с помощью методов анализа текстовых данных.

владеть:

- базовыми методами работы с неструктурированными текстовыми данными.

## 3. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

### 3.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Анализ данных в истории науки: введение в предметную область	1			1

2	Анализ данных: базовые модели и методы	6			15
3	Базовые концепции в историографии науки	6			6
4	Пример исследования: тематика Физтех-школы радиотехники и компьютерных технологий	2			2
5	Пример исследования: тематика Физтех-школы физики и исследований им. Ландау	2			2
6	Пример исследования: тематика Физтех-школы аэрокосмических технологий	2			2
7	Пример исследования: тематика Физтех-школы электроники, фотоники и молекулярной физики	2			2
8	Пример исследования: тематика Физтех-школы прикладной математики и информатики	2			2
9	Пример исследования: тематика Физтех-школы биологической и медицинской физики	2			2
10	Пример исследования: тематика Института нано-, био-, информационных, когнитивных и социогуманитарных наук и технологий	2			2
11	Пример исследования: тематика Физтех-школы бизнеса высоких технологий	2			2
12	Итоговая сборка: анализ выполнения цели и задач курса	1			10
Итого часов		30			48
Подготовка к экзамену		0 час.			
Общая трудоёмкость		78 час., 2 зач.ед.			

### 3.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

#### Семестр: 2 (Весенний)

##### 1. Анализ данных в истории науки: введение в предметную область

###### 1.1. История науки и анализ данных: области пересечения

###### 1.2. Постановка цели и задач курса

##### 2. Анализ данных: базовые модели и методы

###### 2.1. Знакомство с рабочей средой и языком программирования: Google colaboratory, основы синтаксиса Python

###### 2.2. Основы обработки естественного языка

###### 2.3. Нейросетевые модели глубокого обучения: word2vec

###### 2.4. Нейросетевые модели глубокого обучения: Transformers

###### 2.5. Анализ и визуализация сложных сетей: библиотеки и алгоритмы

##### 3. Базовые концепции в историографии науки

- 3.1. Базовые концепции истории науки: интернализм
- 3.2. Базовые концепции истории науки: экстернализм
- 3.3. Социологический анализ истории науки
- 3.4. Антропологический поворот в истории науки
4. Пример исследования: тематика Физтех-школы радиотехники и компьютерных технологий
  - 4.1. История радиотехники: сборка и анализ корпуса
  - 4.2. Анализ текстовых данных: работа над онтологией предметной области
  - 4.3. Построение графа семантического сходства
5. Пример исследования: тематика Физтех-школы физики и исследований им. Ландау
  - 5.1. История квантовых вычислений: сборка и анализ корпуса
  - 5.2. Анализ текстовых данных: работа над онтологией предметной области
  - 5.3. Построение графа семантического сходства
6. Пример исследования: тематика Физтех-школы аэрокосмических технологий
  - 6.1. История аэродинамики: сборка и анализ корпуса
  - 6.2. Анализ текстовых данных: работа над онтологией предметной области
  - 6.3. Построение графа семантического сходства
7. Пример исследования: тематика Физтех-школы электроники, фотоники и молекулярной физики
  - 7.1. История фотоники: сборка и анализ корпуса
  - 7.2. Анализ текстовых данных: работа над онтологией предметной области
  - 7.3. Построение графа семантического сходства
8. Пример исследования: тематика Физтех-школы прикладной математики и информатики
  - 8.1. История математики: сборка и анализ корпуса
  - 8.2. Анализ текстовых данных: работа над онтологией предметной области
  - 8.3. Построение графа семантического сходства
9. Пример исследования: тематика Физтех-школы биологической и медицинской физики
  - 9.1. История биологии: сборка и анализ корпуса
  - 9.2. Анализ текстовых данных: работа над онтологией предметной области
  - 9.3. Построение графа семантического сходства
10. Пример исследования: тематика Института нано-, био-, информационных, когнитивных и социогуманитарных наук и технологий
  - 10.1. История ядерной физики: сборка и анализ корпуса
  - 10.2. Анализ текстовых данных: работа над онтологией предметной области
  - 10.3. Построение графа семантического сходства
11. Пример исследования: тематика Физтех-школы бизнеса высоких технологий
  - 11.1. Экономическая история: сборка и анализ корпуса
  - 11.2. Анализ текстовых данных: работа над онтологией предметной области
  - 11.3. Построение графа семантического сходства
12. Итоговая сборка: анализ выполнения цели и задач курса

- 12.1. Сравнительный анализ результатов тематических сборок
- 12.2. Работа над графом знаний предметных областей Физтеха

#### **4. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)**

Компьютер и мультимедийное оборудование (проектор, звуковая система) с подключением к сети «Интернет»

#### **5. Перечень рекомендуемой литературы**

Основная литература

1. Измерение и анализ случайных процессов [Текст] , Random data: analysis and measurement procedures/Дж. Бендат, А. Пирсол , -М., Мир, 1974

Дополнительная литература

#### **6. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)**

Открытая электронная библиотека “Проект Гутенберга”: <https://www.gutenberg.org/>  
Архив работ Исаака Ньютона: [newtonproject.ox.ac.uk](http://newtonproject.ox.ac.uk)

#### **7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)**

На занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

#### **8. Методические указания для обучающихся по освоению дисциплины (модуля)**

Лекционные часы составляют 45% от требуемого для итоговой аттестации времени. Присутствие на занятиях должно сопровождаться участием в дискуссии и ответами на вопросы. Отсутствие посещений будет отражаться на итоговой оценке.

Итоговая аттестация будет на 55% зависеть от решения итогового задания, основанного на подготовленном исследовании. Во время ответа могут быть заданы вопросы, обсуждавшиеся на лекциях, даже если они отсутствуют в учебниках, следовательно, участие в занятиях является необходимым условием для успешной аттестации.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

**программа аспирантуры:** Биологические науки  
Физтех-школа Биологической и Медицинской Физики  
учебно-научный центр гуманитарных и социальных наук

**курс:** 1

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

**Разработчик:** А.А. Костин, канд. филос. наук

## 1. Показатели оценивания компетенций

В результате изучения дисциплины «Анализ данных в истории науки» обучающийся должен:

### знать:

- основные концепции истории науки;
- передовые достижения в области обработки естественного языка.

### уметь:

- работать с историческими источниками с помощью методов анализа текстовых данных.

### владеть:

- базовыми методами работы с неструктурированными текстовыми данными.

## 2. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Проведение собственного исследования по одной из тем курса

## 3. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Базовые концепции истории науки: интернализм
2. Концепция истории науки К. Поппера
3. Базовые концепции истории науки: экстернализм
4. Концепция истории науки Б.М. Гессена
5. Концепция истории науки Л. Флека
6. Социологический анализ истории науки
7. История науки и социология научного знания Д. Блура
8. Антропологический поворот в истории науки
9. Революция в химии: анализ материально-практического компонента научного исследования в статье «Научный факт и научная теория» А.Л. Никифорова

Пример билета:

1. Концепция истории науки Б.М. Гессена
2. Концепция истории науки Л. Флека

## Критерии оценивания

Оценка «отлично» (10,9,8) выставляется обучающемуся, если он полно излагает материал, дает правильное определение основных понятий; обнаруживает понимание материала, может обосновать свои суждения, демонстрирует: обоснованные примеры, глубокое и прочное усвоение программного материала, последовательные, грамотные и логически излагаемые ответы, свободное владение материалом.

Оценка «хорошо» (7,6,5) выставляется обучающемуся, если демонстрируются: знание программного материала, грамотное изложение без существенных неточностей в ответе на вопрос, правильное изложение теоретических концепций и определений, способность приводить релевантные примеры.

Оценка «удовлетворительно» (4,3) выставляется обучающемуся, если демонстрируются: усвоение основного материала, при ответе допускаются неточности, при ответе даются недостаточно правильные формулировки, нарушается последовательность в изложении программного материала, имеются затруднения в приведении релевантных примеров.

Оценка «неудовлетворительно» (2,1) выставляется обучающемуся, если демонстрируются: незнание программного материала, при ответе возникают ошибки, затруднения при выполнении практических работ, а также за отсутствие ответа.

#### **4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности**

Формой дифференцированного зачета является письменное тестирование – ответ студентов на вопросы билета путем выбора правильного варианта ответа из нескольких предложенных.