

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)**

СОЦИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра теории и истории социологии

Автоматизированный контент-анализ текстов

Рабочая программа дисциплины

Название и код направления подготовки 39.03.01 «Социология»

Название направленности «Цифровая социология и социальная диагностика»

Квалификация выпускника (бакалавр)

Форма обучения (очная)

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2024

«Автоматизированный контент-анализ текстов»

Рабочая программа дисциплины

Составитель:

И.А. Лавров, к.соц.н., зам. директора Научного центра цифровой социологии «Ядов-центр», РГГУ

Ответственный редактор

Доктор социологических наук, проф. М.Б. Буланова

УТВЕРЖДЕНО

Протокол заседания Совета факультета

№ 3 от 19.03.2024

ОГЛАВЛЕНИЕ

1. Пояснительная записка

1.1 Цель и задачи дисциплины

1.2. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

1.3. Место дисциплины в структуре образовательной программы

2. . Структура дисциплины

3. Содержание дисциплины

4. Образовательные технологии

5. Оценка планируемых результатов обучения

5.1. Система оценивания

5.2. Критерии выставления оценки по дисциплине

5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

7. Материально-техническое обеспечение дисциплины

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

9. Методические материалы

9.1. Планы практических (семинарских, лабораторных) занятий

9.2. Методические рекомендации по подготовке письменных работ

9.3. Иные материалы

Приложения

Приложение 1. Аннотация дисциплины

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Цель курса: подготовить специалиста, обладающего знаниями и умениями автоматизированного контент-анализа текстов, а также владеющего профессиональными компетенциями, необходимыми в сфере деятельности социолога.

Задачи курса: изучить технику и технологии контент-анализа текстов; усвоить этапы работы с текстом; овладеть необходимыми инструментами автоматизированного контент-анализа; сформировать навыки организации и проведения контент-анализа текста в проблемном поле социо-гуманитарных дисциплин; развить умение использовать эти навыки в практической и исследовательской деятельности.

1.2. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ПК-4 Организация работы по сбору, обработке и анализу данных фундаментальных и прикладных социологических исследований	ПК-4.1 – Проведение социальной диагностики и подготовка сбора социологических данных ПК-4.2 – Сбор данных из первичных и вторичных источников с применением современной аппаратуры, оборудования ПК-4.3 – Обработка, анализ и представление собранных данных с применением информационных технологий	Знать: технику и технологии контент-анализа текстов. Уметь: организовать и провести контент-анализ текстов в проблемном поле социо-гуманитарных дисциплин. Владеть: необходимыми инструментами автоматизированного контент-анализа текстов.

1.3. Место дисциплины в структуре образовательной программы:

Дисциплина «Автоматизированный контент-анализ текстов» относится к вариативной части, формируемой участниками образовательных отношений по направлению «Социология».

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: «Социальная структура и стратификация», «Количественные методы», «Качественные методы», «История цифровой социологии», ознакомительной и производственной практик.

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: «Социология средств массовой информации и медиа-анализ», «Интеллектуальный анализ текстов», «Научное и исследовательское проектирование в цифровой социологии»; прохождения производственной и педагогической практик.

2. Структура дисциплины

Структура дисциплины для очной формы обучения

Общая трудоёмкость дисциплины составляет 3 з.е., 108 академических часов.

Объем дисциплины в форме очной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
6	Лекции	20
6	Семинары	40
Всего:		60

Объем дисциплины в форме самостоятельной работы обучающихся составляет 48 академических часов.

3. Содержание дисциплины

Тема 1. Введение в метод контент-анализа текстов

- Контент-анализ как социологический метод: понятие и область применения.
- История метода.
- Возможности и ограничения метода.
- Основные этапы контент-анализа.
- Смысловые единицы – индикаторы и категории.
- Виды контент-анализа.
- Эволюция метода контент-анализа

Тема 2. Документ, текст, дискурс как основные источники информации для контент-анализа

- Понятия документа, текста, дискурса. Синонимичность понятий и их основные отличия.
- Разновидности документов для анализа.
- Основные признаки текста.
- Типы и назначения текстов.
- Особенности организации текстов.
- Структура и закономерности функционирования текстов.
- Отличительные черты дискурса.
- Выбор источников для анализа.
- Методы преобразования информации в текст.

Тема 3. Количественный и качественный контент-анализ

- Различия количественного и качественного контент-анализа.
- Качественные и количественные данные в контент-анализе и способы их получения.
- Программа исследования, категории анализа, единицы счета и процесс кодирования информации при качественном и количественном контент-анализе.
- Критерии оценки качественного и количественного исследования методом контент-анализа.

Тема 4. Анализ и интерпретация полученных данных

- Принципы аналитического описания и осмысления результатов.
- Первичное описание текстовых данных.
- Проверка надежности данных.

Тема 5. Концепции электронного текста как источника информации для автоматизированного контент-анализа

- Электронный текст как информационный ресурс. Основные концепции электронного текста.
- Жизненный цикл электронного текста.
- Основные источники электронных текстов. Академические, новостные, исторические и прочие системы-агрегаторы хранения электронных текстов. Поисковые системы для отбора электронных текстов.

- Способы извлечения информации из электронных текстов.

Тема 6. Подготовка (препроцессинг) электронных текстов к автоматизированную контент-анализу

- Процедуры подготовки (препроцессинга) электронного текста к автоматизированному контент-анализу.
- Лемматизация текста. Понятие лемматизации. Способы приведения словоформ к исходным формам слова. Существующие библиотеки, словари, лемматизаторы и парсеры для подготовки текста к автоматизированному контент-анализу.
- Понятие шаблона текста. Создание правил и категорий автоматизированного контент-анализа.
- Маркирование текста. Цели и задачи маркирования. Проблема субъективизма и скрытых смыслов при маркировании текста.

Тема 7. Классификация и кластеризация электронных текстов

- Понятие классификации текста. Автоматическая и ручная классификация электронного текста. Разновидности алгоритмов классификации текстов.
- Понятие кластеризации текста. Разновидности алгоритмов кластеризации текстов.
- Понятие корпуса текста. Расчет индикатора значимости темы в корпусе текста.
- Процессы оптимизации количества классов, кластеров и тем, использующихся в автоматизированном контент-анализе.

Тема 8. Сентимент-анализ и извлечение мнений

- Понятие сентимент-анализа или анализа тональности текста. Основные подходы к оценке тональности текста. Разработка тезауруса и меток тональности для сентимент-анализа.
- Понятие типологии мнений. Конструирование типологии мнений для сентимент-анализа в социальных сетях.
- Выявление социальных проблем при помощи сентимент-анализа социальных сетей.

Тема 9. Извлечение объектов, аспектов и фактов

- Понятие объекта в электронном тексте. Подходы к извлечению объектов из электронного текста.
- Понятие аспекта объекта в электронном тексте. Типология и классификация аспектов.
- Понятие факта в электронном тексте. Специфика разметки электронного текста для выделения групп объектов и взаимосвязи между ними.

Курс "Автоматизированный контент-анализ текстов" позволит студентам овладеть основными методами и инструментами для автоматизированного контент-анализа текстов, а также развить практические навыки в области анализа социогуманитарных текстов.

4. Образовательные технологии

<i>№ п/п</i>	<i>Наименование темы</i>	<i>Виды учебной работы</i>	<i>Образовательные технологии</i>
<i>1</i>	<i>2</i>	<i>3</i>	<i>5</i>
1.	<i>Введение в метод контент-анализа текстов</i>	<i>Лекция 1 Лекция 2 Семинар 1 Семинар 2 Семинар 3 Самостоятельная работа</i>	<i>Вводная лекция. Занятие –дискуссия. Консультация преподавателя.</i>
2.	<i>Документ, текст, дискурс как основные источники информации для контент-анализа</i>	<i>Лекция 3 Семинар 4 Семинар 5 Самостоятельная работа</i>	<i>Проблемная лекция Занятие-дискуссия Консультация преподавателя.</i>
3.	<i>Количественный и качественный контент-анализ</i>	<i>Лекция 4 Семинар 6 Семинар 7 Самостоятельная работа</i>	<i>Проблемная лекция Занятие-дискуссия Консультация преподавателя.</i>
4.	<i>Анализ и интерпретация</i>	<i>Лекция 5</i>	<i>Проблемная лекция</i>

	<i>полученных данных</i>	<i>Семинар 8 Семинар 9 Самостоятельная работа</i>	<i>Занятие-дискуссия Консультация преподавателя.</i>
5.	<i>Концепции электронного текста как источника информации для автоматизированного контент-анализа</i>	<i>Лекция 6 Семинар 10 Семинар 11 Самостоятельная работа</i>	<i>Проблемная лекция Занятие-дискуссия Консультация преподавателя.</i>
6.	<i>Подготовка (препроцессинг) электронных текстов к автоматизированную контент-анализу</i>	<i>Лекция 7 Семинар 12 Семинар 13 Самостоятельная работа</i>	<i>Проблемная лекция Занятие-дискуссия Консультация преподавателя.</i>
7.	<i>Классификация и кластеризация электронных текстов</i>	<i>Лекция 8 Семинар 14 Семинар 15 Семинар 16 Самостоятельная работа</i>	<i>Проблемная лекция Занятие-дискуссия Консультация преподавателя.</i>
8.	<i>Сентимент-анализ и извлечение мнений</i>	<i>Лекция 9 Семинар 17 Семинар 18 Самостоятельная работа</i>	<i>Проблемная лекция Занятие-дискуссия Консультация преподавателя.</i>
9.	<i>Извлечение объектов, аспектов и фактов</i>	<i>Лекция 10 Семинар 19 Семинар 20 Самостоятельная работа</i>	<i>Проблемная лекция Занятие-дискуссия Консультация преподавателя.</i>

При реализации программы дисциплины «Автоматизированный контент-анализ текстов» используются: проблемный метод изложения лекционного материала, отработка теоретических знаний и практических навыков в ходе семинарских занятий.

Также используются такие образовательные технологии, как поиск материалов и справочной литературы через Интернет, игровые методики, связанные с освоением конкретных способов анализа образовательных проблем.

Самостоятельная работа студентов осуществляется под руководством преподавателя (консультативная помощь при выполнении учебно-исследовательских проектов), а также подразумевает индивидуальную работу студента в библиотеке в целях подготовки к каждому семинарскому занятию, работу в «поле» в целях проведения социологического исследования.

Самостоятельная работа студентов организуется с использованием свободного доступа к Интернет-ресурсам, в том числе с обращением к материалам (словарь, комментарии, тексты лекций и т.д.), размещенным на сайте социологического факультета РГГУ <http://soc.rsuh.ru/> и в базе данных архива социологического факультета.

Студентам предоставляется возможность выступить по одной из тем курса на студенческой конференции социологического факультета РГГУ в рамках ежегодных Гуманитарных чтений, разместить текст доклада на сайте социологического факультета РГГУ <http://soc.rsuh.ru/>.

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

5. Оценка планируемых результатов обучения

5.1. Система оценивания

Оценка знаний студентов складывается из *промежуточных* аттестаций (60%), а также *итоговой* аттестации в письменной форме (40%).

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль: - участие в дискуссии на семинаре	5 баллов	25 баллов

- самостоятельная работа	5 баллов	35 баллов
Промежуточная аттестация (коллоквиум)		40 баллов
Итого за семестр Зачет с оценкой		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (EuropeanCreditTransferSystem; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

5.2. Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	«отлично»/ «зачтено (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	<p>«хорошо»/ «зачтено (хорошо)»/ «зачтено»</p>	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	<p>«удовлетворительно»/ «зачтено (удовлетворительно)»/ «зачтено»</p>	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	<p>«неудовлетворительно»/ не зачтено</p>	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Текущий контроль

Формируемые компетенции и индикаторы их достижения: ПК-4 (ПК-4.1, ПК-4.2, ПК-4.3)

При оценивании готовности студента на семинарском занятии учитываются:

- Использование материалов лекции (0-1 балл);
- Использование рекомендованных к занятию источников и литературы (0-2 балла);
- Умение поддержать дискуссию по актуальным проблемам с использованием рекомендованных материалов (0-2 балла).

Промежуточная аттестация-коллоквиум (зачет)

При проведении промежуточной аттестации студент должен ответить на 3 вопроса.

При оценивании ответа на вопросы по лекционному курсу учитывается:

- содержание лекционного курса не освоено, знание материала носит фрагментарный характер, наличие грубых ошибок в ответе (0-9 баллов);
- содержание лекционного курса освоено частично, допущено не более двух-трех недочетов (10-19 баллов);
- содержание лекционного курса освоено почти полностью, допущено не более одного-двух недочетов, но обучающийся смог бы их исправить самостоятельно (20-29 баллов);
- содержание лекционного курса освоено полностью, ответ построен по собственному плану (30-40 баллов).

Вопросы к коллоквиуму

- История возникновения контент-анализа;
- Контент-анализ как социологический метод: понятие и область применения;
- Основные этапы контент-анализа, понятие смысловых единиц, единиц счета и индикаторов;
- Документ, текст, дискурс. Их основные отличия. Основные признаки текста. Выбор источников для анализа и способы преобразования информации в текст;
- Количественный и качественный контент-анализ. Различия в категориях анализа, единицах счета и процессе кодирования и обработки информации;
- Основные принципы аналитического описания результатов проведенного контент-анализа и способы проверки надежности данных;
- Понятие электронного текста. Его ключевые характеристики и отличительные черты. Основные источники электронных текстов и способы извлечения информации из них;
- Лемматизация текста. Обоснование необходимости лемматизации текста перед проведением автоматизированного контент-анализа и способы её проведения;
- Основные принципы классификации текста. Возможности и ограничения автоматизированной классификации текста;
- Кластеризация текста. Преимущества и недостатки автоматизированной кластеризации текстов;
- Понятие сентимент-анализа или анализа тональности текста. Способы выявления социальных проблем при сентимент-анализе социальных сетей;
- Понятия объекта, аспекта и факта при автоматизированном контент-анализе текста. Способы и методы их извлечения из электронного текста.

Формируемые компетенции и индикаторы их достижения: ПК-4 (ПК-4.1, ПК-4.2, ПК-4.3)

1. Сущность и содержание классического контент-анализа текстов.
2. Сущность и содержание автоматизированного контент-анализа текстов.
3. Отличия, недостатки и преимущества количественной и качественной стратегии проведения контент-анализа.
4. Сущность и содержание препроцессинга электронных текстов к автоматизированную контент-анализу.

5. Основные принципы классификации и кластеризации электронных текстов.
6. Основные подходы к оценке тональности текста.
7. Сущность процесса извлечения объектов, аспектов и фактов из электронного текста.

Тематика докладов (для дискуссии на семинарах)

Формируемые компетенции и индикаторы их достижения: ПК-4 (ПК-4.1, ПК-4.2, ПК-4.3)

1. Какие особенности присущи автоматизированному контент-анализу текстов в настоящее время?
2. Каковы достоинства и недостатки автоматизированного контент-анализа текстов?
3. Какие области исследования может охватить автоматизированный контент-анализ текстов?
4. Каковы их отличительные особенности автоматизированного контент-анализа текстов от классического?
5. Какие инструменты используются в автоматизированном контент-анализе текстов? Каковы их особенности?
6. Какие исследовательские задачи могут быть решены с помощью автоматизированного контент-анализа текстов?
7. Каким образом автоматизированный контент-анализ текстов может помочь в научных исследованиях?
8. Какие трудности могут возникнуть при использовании автоматизированного контент-анализа текстов? Как их можно преодолеть?
9. Какие достоинства и недостатки у автоматизированного контент-анализ текстов по сравнению с другими методами анализа данных?
10. Какие требования предъявляются к исследователям, работающим в области автоматизированного контент-анализ текстов?

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

Источники

1.Федеральный закон "Об образовании в Российской Федерации" от 29.12.2012 N 273-ФЗ (последняя редакция) // http://www.consultant.ru/document/cons_doc_LAW_140174/

Литература

1. Обязательная.

Бурдые П. Социология социального пространства. - М.: Институт экспериментальной социологии, 2007.

Василенко Л.А., Мещерякова Н.Н. Социология цифрового общества. Томск: Изд-во Томского политехнического университета. 2021. 226 с.

Гидденс Э. Современность и самоидентификация: Самоанализ и социологическая теория. - М.: Праксис, 2005.

Девятко И.Ф. Методы социологического исследования. Екатеринбург: Издательство Уральского университета. 1998. 191 с.

Добринская Д.Е. Введение в цифровую социологию. М.: МАКС Пресс. 2022. 144 с.

Кольцова О.Ю. Автоматизированные методы анализа больших массивов интернет-текстов // Онлайн-исследования в России: тенденции и перспективы. М.: Издательство МИК. 2016.

Котарбинский Т. Логика и методология науки. - М.: Прогресс, 1987.

Крыштановская О.В. Бесконтактная социология: новые формы исследований в цифровую эпоху // Цифровая социология. 2018. №1. С. 5-8.

Лавров И.А. Цифровая социология и современные методы изучения политической элиты // Вестник Университета. 2019. №6. С. 173-179.

Лурия А. Развитие психики. - М.: Изд-во Академии наук СССР, 1959.

Манн Г. Основы теории социальных систем. - М.: Канон-плюс, 2007.

Мерло-Понти М. Феноменология восприятия. - М.: Издательство Московского университета, 2006.

Миллс Р. Воображаемое. Социологическое воображение. - М.: Республика, 2003.

Ницевич В.Ф. Цифровая социология: теоретико-методологические истоки и основания // Цифровая социология. 2018. №1. С. 18-28.

Поппер К. Логика и рост научного знания. - М.: Прогресс, 1983.

Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.

Таршис Е.Я. Контент-анализ: принципы методологии. М.: ЛЕНАРД. 2021. 174 с.

Шалак В.И. Современный контент-анализ. 2004. М.: Омега-Л. 272 с.

Lupton D. Digital sociology. Routledge. 2014. 236 p.

1. Дополнительная.

Бурдые П. Культурные ориентиры. – М.: Прогресс, 1997.

Бурдые П. Социальное пространство: поля и практики. – М.: Институт экспериментальной социологии, 2007.

Вирно П. Грамматика множества. К анализу форм современной жизни. М.: Ад Маргинем Пресс. 2016. 144 с.

Журавлева Е.Ю. Социология в сетевой среде: к цифровым социальным исследованиям // Социологические исследования. 2015. № 8. С.25-33.

Каган М.С. Социология знания. – М.: Институт экспериментальной социологии, 2005.

Лапина-Кратасюк Т., Запорожец О., Возьянов А.. Сети города: Люди. Технологии. Власти. М: Новое литературное обозрение. 2021. 576 с.

Лурия А. Развитие психики. – М.: Изд-во Академии наук СССР, 1959.

Back L. Live sociology: social research and its futures // The Sociological Review. No. 60. 2012. pp. 18-39.

Beer D. Punk sociology. Springer. 2014. 76 P.

Latzko-Toth G., Bonneau C., Millette M. Small data, thick data: Thickening strategies for trace-based social media research // The SAGE handbook of social media research methods. SAGE Reference. 2017. С. 199-214.

Parizer E. The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think. NY: Penguin Books. 2012. 304 p.

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

Профессиональные базы данных и информационно-справочные системы.

Национальная электронная библиотека (НЭБ) www.rusneb.ru

ELibrary.ru Научная электронная библиотека www.elibrary.ru

Электронная библиотека Grebennikon.ru www.grebennikon.ru

Cambridge University Press

ProQuest Dissertation & Theses Global

SAGE Journals
Taylor and Francis
JSTOR
Kribrum

Иные ресурсы

Всероссийский институт изучения общественного мнения. - [http:// www.wciom.ru](http://www.wciom.ru)

Институт социологии РАН – [http:// www.isras.rssi.ru](http://www.isras.rssi.ru).

Институт социально-политических исследований – [http:// www.ispr.ras.ni](http://www.ispr.ras.ni).

Московский общественный научный фонд – <http://www.mpsf.org/ndexr.html.ru>

Фонд «Общественное мнение» - <http://www.fom.ru>

Центр независимых социологических исследований – <http://www.indepsocres.spb.ru>

Центр социологических исследований Министерства образования и науки – http://www.ingormika.ru/win-dows/goscom/cinogra/n/socio/first_pg.html.

Центр социологических исследований МГУ – <http://www.opinio.msu.ru>.

ROMIR – Monitoring – www.romir.ru

Федеральный образовательный портал – экономика, социология, менеджмент – <http://www.ecsocman.edu.ru>

Информация – онлайн о социологических исследованиях – <http://www.socresonline.org.uk/socresonline/2/3/9.html>.

Журнал социологии и социальной антропологии - <http://www.soc.pu.ru>: 8101 publications/jssa.

Социологический журнал - <http://win.www.nir.ru/socio/sci-publ/socjour.html>.

Социологические исследования - http://www.isras.rssi.ru/R_Socis.html.

Социология от А до Я - <http://www.giasnet.ru/-asch/sociolo-dy/>.

Электронное социологическое обозрение - <http://www.sociologica.ru>

АО «Крибрум» - исследовательская и технологическая компания - <https://kribrum.ru/>

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: учебные аудитории, оснащённые компьютером и проектором для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows

2. Microsoft Office

Профессиональные полнотекстовые базы данных:

1. Национальная электронная библиотека (НЭБ) www.rusneb.ru
2. **ELibrary.ru** Научная электронная библиотека www.elibrary.ru
3. Электронная библиотека Grebennikon.ru www.grebennikon.ru
4. Cambridge University Press
5. ProQuest Dissertation & Theses Global
6. SAGE Journals
7. Taylor and Francis
8. JSTOR

Информационные справочные системы:

1. Консультант Плюс
2. Гарант
3. PolyAnalyst
4. «Крибрум» - программно-аппаратный комплекс мониторинга и анализа данных от АО «Крибрум» - аналитического центра изучения социальных медиа. Ссылка: <https://kribrum.ru/>

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
 - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
 - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
 - письменные задания оформляются увеличенным шрифтом;
 - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
 - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
 - письменные задания выполняются на компьютере в письменной форме;

- экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата:
- лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
- письменные задания выполняются на компьютере со специализированным программным обеспечением;
- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
 - в печатной форме увеличенным шрифтом;
 - в форме электронного документа;
 - в форме аудиофайла.
- для глухих и слабослышащих:
 - в печатной форме;
 - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - в печатной форме;
 - в форме электронного документа;
 - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
 - устройством для сканирования и чтения с камерой SARA CE;

- дисплеем Брайля PAC Mate 20;
- принтером Брайля EmBrailleViewPlus;
- для глухих и слабослышащих:
 - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
 - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - передвижными, регулируемые эргономическими партами СИ-1;
 - компьютерной техникой со специальным программным обеспечением.

9. Методические материалы

9.1. Планы семинарских занятий

Семинар 1-3. Тема: Введение в метод контент-анализа текстов

Вопросы:

1. Какая область применения у контент-анализа в качестве социологического метода исследования?
2. Какие вы можете назвать преимущества и ограничения метода?
3. Какие вы можете назвать основные этапы при проведении контент-анализа?
4. Какие существуют разновидности контент-анализа?
5. Что такое индикаторы и категории контент-анализа?

Список источников и литературы:

Основная литература

1. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.
2. Таршис Е.Я. Контент-анализ: принципы методологии. М.: ЛЕНАРД. 2021. 174 с.

Дополнительная литература

1. Шалак В.И. Современный контент-анализ. 2004. М.: Омега-Л. 272 с.
2. Девятко И.Ф. Методы социологического исследования. Екатеринбург: Издательство Уральского университета. 1998. 191 с.

Семинар 4-5. Документ, текст, дискурс как основные источники информации для контент-анализа

Вопросы:

1. Чем отличаются друг от друга документ, текст и дискурс?
2. Какие основные признаки текста вы можете назвать?
3. Какие методы преобразования информации в текст вы знаете?

Список источников и литературы:

Основная литература

1. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.
2. Таршис Е.Я. Контент-анализ: принципы методологии. М.: ЛЕНАРД. 2021. 174 с.

Дополнительная литература

1. Шалак В.И. Современный контент-анализ. 2004. М.: Омега-Л. 272 с.
2. Девятко И.Ф. Методы социологического исследования. Екатеринбург: Издательство Уральского университета. 1998. 191 с.

Семинар 6-7. Количественный и качественный контент-анализ

Вопросы:

1. Чем отличаются друг от друга количественный и качественный контент-анализ?
2. Что будет являться единицей счета в количественном контент-анализе? А в качественном?
3. В чем отличие процесса кодирования информации для качественного и количественного контент-анализа?

Список источников и литературы:

Основная литература

1. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.
2. Таршис Е.Я. Контент-анализ: принципы методологии. М.: ЛЕНАРД. 2021. 174 с.

Дополнительная литература

1. Шалак В.И. Современный контент-анализ. 2004. М.: Омега-Л. 272 с.
2. Девятко И.Ф. Методы социологического исследования. Екатеринбург: Издательство Уральского университета. 1998. 191 с.

Семинар 8-9. Анализ и интерпретация полученных данных

Вопросы:

1. Назовите основные принципы аналитического описания и осмысления результатов, полученных в ходе контент-анализа?
2. Каким образом возможно осуществить проверку надежности полученных в ходе контент-анализа данных?

Список источников и литературы:

Основная литература

1. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.
2. Таршис Е.Я. Контент-анализ: принципы методологии. М.: ЛЕНАРД. 2021. 174 с.

Дополнительная литература

1. Шалак В.И. Современный контент-анализ. 2004. М.: Омега-Л. 272 с.
2. Девятко И.Ф. Методы социологического исследования. Екатеринбург: Издательство Уральского университета. 1998. 191 с.

Семинар 10-11. Концепции электронного текста как источника информации для автоматизированного контент-анализа

Вопросы:

1. Что такое электронный текст? В чем его отличия от иных форм текста?
2. Какие основные источники электронных текстов вы можете назвать?
3. Каким образом осуществляется извлечение информации из электронных текстов?

Список источников и литературы:

Основная литература

1. Кольцова О.Ю. Автоматизированные методы анализа больших массивов интернет-текстов // Онлайн-исследования в России: тенденции и перспективы. М.: Издательство МИК. 2016.
2. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.

Дополнительная литература

1. Василенко Л.А., Мещерякова Н.Н. Социология цифрового общества. Томск: Изд-во Томского политехнического университета. 2021. 226 с
2. Крыштановская О.В. Бесконтактная социология: новые формы исследований в цифровую эпоху // Цифровая социология. 2018. №1. С. 5-8.

Семинар 12-13. Подготовка (препроцессинг) электронных текстов к автоматизированную контент-анализу

Вопросы:

1. Какие процедуры подготовки текста к автоматизированному контент-анализу вы можете назвать?
2. Что такое лемматизация текста? Каким образом она осуществляется?
3. Какие основные проблемы может испытывать исследователь в процессе маркировки текста?

Список источников и литературы:

Основная литература

1. Кольцова О.Ю. Автоматизированные методы анализа больших массивов интернет-текстов // Онлайн-исследования в России: тенденции и перспективы. М.: Издательство МИК. 2016.
2. Добринская Д.Е. Введение в цифровую социологию. М.: МАКС Пресс. 2022. 144 с.

Дополнительная литература

1. Василенко Л.А., Мещерякова Н.Н. Социология цифрового общества. Томск: Изд-во Томского политехнического университета. 2021. 226 с
2. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.

Семинар 14-16. Классификация и кластеризация электронных текстов

Вопросы:

1. Что такое классификация текста? Какие методы классификации текста вы знаете?
2. Как вы понимаете, что такое корпус текста?
3. Опишите процесс расчета индикатора значимости темы в корпусе текста?
4. Что такое кластеризация текста? Какие разновидности алгоритмов кластеризации текстов вы можете назвать?
5. Каким образом обеспечить оптимизацию количества классов, кластеров и тем для автоматизированного контент-анализа?

Список источников и литературы:

Основная литература

1. Кольцова О.Ю. Автоматизированные методы анализа больших массивов интернет-текстов // Онлайн-исследования в России: тенденции и перспективы. М.: Издательство МИК. 2016.
2. Добринская Д.Е. Введение в цифровую социологию. М.: МАКС Пресс. 2022. 144 с.

Дополнительная литература

1. Василенко Л.А., Мещерякова Н.Н. Социология цифрового общества. Томск: Изд-во Томского политехнического университета. 2021. 226 с
2. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.

Семинар 17-18. Сентимент-анализ и извлечение мнений

Вопросы:

1. Что такое сентимент-анализ? С какой целью он осуществляется?
2. Какие преимущества и недостатки ручной разработки тезауруса и меток тональности для проведения сентимент-анализа вы можете назвать?
3. Что такое типология мнений? Каким образом она конструируется?

Список источников и литературы:

Основная литература

1. Кольцова О.Ю. Автоматизированные методы анализа больших массивов интернет-текстов // Онлайн-исследования в России: тенденции и перспективы. М.: Издательство МИК. 2016.
2. Добринская Д.Е. Введение в цифровую социологию. М.: МАКС Пресс. 2022. 144 с.

Дополнительная литература

1. Василенко Л.А., Мещерякова Н.Н. Социология цифрового общества. Томск: Изд-во Томского политехнического университета. 2021. 226 с
2. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.

Семинар 19-20. Извлечение объектов, аспектов и фактов

Вопросы:

1. Что такое объект в электронном тексте? Какими способами осуществляется его извлечение?
2. Что такое аспект объекта? Какая существует классификация аспектов объекта при проведении автоматизированного контент-анализа
3. В чем заключается специфика разметки электронного текста для выделения групп объектов и взаимосвязи между ними?

Список источников и литературы:

Основная литература

1. Кольцова О.Ю. Автоматизированные методы анализа больших массивов интернет-текстов // Онлайн-исследования в России: тенденции и перспективы. М.: Издательство МИК. 2016.
2. Добринская Д.Е. Введение в цифровую социологию. М.: МАКС Пресс. 2022. 144 с.

Дополнительная литература

1. Василенко Л.А., Мещерякова Н.Н. Социология цифрового общества. Томск: Изд-во Томского политехнического университета. 2021. 226 с
2. Семёнова А.В. Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения. М.: ИС РАН. 2010. 324 с.

9.2. Методические указания.

Тематика семинарских занятий органично дополняет тематику лекционного курса. На примере изучения конкретных проблем, конкретизируется и иллюстрируется аналитический материал, получаемый студентами в ходе лекций.

Вопросы, изучаемые студентами, позволяют наиболее полно представить проблемное поле современной социологии.

При подготовке к семинарским занятиям студенты должны анализировать источники и обязательную литературу по конкретным вопросам, поставленным к занятию. В случае, если материал, представленный источниками, недостаточен для ответа на вопросы, рекомендуется привлекать дополнительную литературу, указанную в списке к каждому занятию.

На практических занятиях студенты должны грамотно вести дискуссию, аргументировано излагать свое мнение в ходе защиты своей позиции.

Литература, отобранная для работы студентов, доступна, в основном сосредоточена в научной библиотеке РГГУ; отвечает потребностям изучения актуальных проблем социологии образования; содержит как классическую, так и современную литературу.

АННОТАЦИЯ ДИСЦИПЛИНЫ

Дисциплина «Автоматизированный контент-анализ текстов» реализуется на социологическом факультете кафедрой теории и истории социологии.

Цель дисциплины: подготовить специалиста, обладающего знаниями и умениями автоматизированного контент-анализа текстов, а также владеющего профессиональными компетенциями, необходимыми в сфере деятельности социолога.

Задачи дисциплины: курса: изучить технику и технологии контент-анализа текстов; усвоить этапы работы с текстом; овладеть необходимыми инструментами автоматизированного контент-анализа; сформировать навыки организации и проведения контент-анализа текста в проблемном поле социо-гуманитарных дисциплин; развить умение использовать эти навыки в практической и исследовательской деятельности.

Дисциплина направлена на формирование следующих компетенций студента:

ПК-4. Организация работы по сбору, обработке и анализу данных прикладных социологических исследований. Полностью освоивший программу курса студент должен:

1. Знать: технику и технологии контент-анализа текстов;
2. Уметь: организовать и провести контент-анализ текстов в проблемном поле социо-гуманитарных дисциплин;
3. Владеть: необходимыми инструментами автоматизированного контент-анализа текстов.

По дисциплине предусмотрена промежуточная аттестация в форме зачета с оценкой.

Общая трудоемкость освоения дисциплины составляет 3 зачетные единицы.