

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования

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

ИСТОРИКО-АРХИВНЫЙ ИНСТИТУТ

ФАКУЛЬТЕТ АРХИВОВЕДЕНИЯ И ДОКУМЕНТОВЕДЕНИЯ

Кафедра источниковедения

Рабочая программа дисциплины

МЕТОДЫ И ТЕХНОЛОГИИ РАБОТЫ С ТЕКСТАМИ

Направление подготовки 46.03.02 – Документоведение и архивоведение

Направленность – Электронные архивы и документы

Уровень квалификации выпускника – бакалавр

Форма обучения – очно-заочная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2021

МЕТОДЫ И ТЕХНОЛОГИИ РАБОТЫ С ТЕКСТАМИ

Рабочая программа курса

Составители:

д.и.н., доц. И.М. Гарскова

к.ф.-м.н., доц. С.В. Шпирко

УТВЕРЖДЕНО

Протокол заседания кафедры

№__4__ от__17.05.2021__

ОГЛАВЛЕНИЕ

| | |
|---|----|
| 1. Пояснительная записка | 4 |
| 1.1. Цель и задачи дисциплины | 4 |
| 1.2. Формируемые компетенции, а также перечень планируемых результатов обучения | 5 |
| 1.3. Место дисциплины в структуре основной образовательной программы | 6 |
| 2. Структура дисциплины | 7 |
| 3. Содержание дисциплины | 8 |
| 4. Образовательные технологии | 9 |
| 5. Оценка планируемых результатов обучения | 12 |
| 5.1. Система оценивания | 12 |
| 5.2. Критерии выставления оценок | 12 |
| 5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине | 13 |
| 6. Учебно-методическое и информационное обеспечение дисциплины | 16 |
| 6.1. Список источников и литературы | 13 |
| 6.2. Перечень ресурсов информационно-телекоммуникационной сети Интернет | 21 |
| 7. Материально-техническое обеспечение дисциплины | 22 |
| 8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов | 22 |
| 9. Методические материалы | 23 |
| 9.1. Планы практических (семинарских, лабораторных) занятий | 23 |
| 9.2. Методические рекомендации по подготовке письменных работ | 28 |
| Приложения | |
| Приложение 1. Аннотация дисциплины | 30 |

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Дисциплина «МЕТОДЫ И ТЕХНОЛОГИИ РАБОТЫ С ТЕКСТАМИ» входит в состав части, формируемой участниками образовательных отношений, блока дисциплин учебного плана бакалаврской программы «Электронные архивы и документы» по направлению подготовки 46.03.02 – Документоведение и архивоведение.

Цель дисциплины: дать студентам теоретические и практические знания о современных принципах анализа текстов, способах обработки, вторичного использования и извлечения скрытой информации, представленной в описательном (неструктурированном) виде, а также об информационных технологиях, на которых основан компьютеризованный анализ текстовой информации.

Задачи дисциплины:

- обзор междисциплинарных методов и технологий работы с текстами в гуманитарных исследованиях (в сфере истории, политологии, социологии, лингвистики), базирующихся на системном подходе и формализованных методах обработки текстовой информации;
- теоретические концепции анализа текстовой информации и их эволюция;
- специфика и анализ опыта применения методов и технологий работы с текстами в исторических исследованиях;
- характеристика контент-анализа как качественно-количественного метода анализа содержания;
- понятийный аппарат контент-анализа, его виды и этапы;
- изучение основных подходов к автоматизации контент-анализа, определение их достоинств и недостатков;
- знакомство с программным обеспечением компьютеризованного анализа текстовой информации.

1.2. Формируемые компетенции, а также перечень результатов обучения

| Компетенция (код и наименование) | Индикаторы компетенций (код и наименование) | Результаты обучения |
|--|---|---|
| ПК-1 Способность применять научно-методические основы документоведения и архивоведения в профессиональной сфере | ПК-1.3 Решает стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом требований информационной безопасности | <p><i>Знать:</i> основные методы и подходы, перспективные направления развития информационных технологий в области анализа текстов; зарубежный и отечественный опыт формализованного анализа текстовой информации исторических источников; основы проектирования текстовых баз данных; методологию и методику контент-анализа; основные характеристики и концепции электронного текста, принципы функционирования прикладного программного обеспечения, предназначенного для компьютеризованного анализа неструктурированной текстовой информации ; проводить структурную и семантическую разметку (маркирование) текстов; создавать полнотекстовые базы данных и выполнять информационный поиск ; методологию и методику контент-анализа ; специфику использования этих методов и технологий в различных гуманитарных науках</p> <p><i>Владеть:</i> современными информационными технологиями работы с электронными текстовыми архивами ; аппаратными и программными средствами перевода текстов в электронный формат ; методами маркирования текстов и информационного поиска в полнотекстовых системах; методами математической статистики, связанными с дескриптивной статистикой и анализом взаимосвязей, их графическим представлением ;</p> <p><i>Уметь:</i> работать со словарем текста; подсчитывать количественные характеристики совместной встречаемости лексических единиц и семантических категорий; визуализировать результаты анализа и</p> |

| | | |
|--|--|--|
| | | экспортировать их для дальнейшей обработки в статистические программы; сканировать текстовые материалы для перевода их в электронный формат; работать со словарем текста; подсчитывать количественные характеристики совместной встречаемости лексических единиц и семантических категорий; визуализировать результаты анализа и экспортировать их для дальнейшей обработки в статистические программы ; определять характеристики текстовых документов, которые с необходимостью отражали бы существенные стороны их содержания; проводить структурную и семантическую разметку (маркирование) текстов; создавать полнотекстовые базы данных и выполнять информационный поиск; работать со словарем текста; подсчитывать количественные характеристики совместной встречаемости лексических единиц и семантических категорий; визуализировать результаты анализа и экспортировать их для дальнейшей обработки в статистические программы |
|--|--|--|

1.3. Место дисциплины в структуре ООП

Дисциплина "Методы и технологии работы с текстами" относится к части, формируемой участниками образовательных отношений, блока дисциплин учебного плана бакалаврской программы «Электронные архивы и документы» по направлению подготовки 46.03.02 – Документоведение и архивоведение. Дисциплина реализуется кафедрой источниковедения факультета архивного дела историко-архивного института РГГУ. и адресована студентам 4 курса (7 семестр). Дисциплина реализуется кафедрой источниковедения факультета архивного дела историко-архивного института РГГУ.

Для освоения дисциплины необходимы компетенции, сформированные в ходе изучения следующих дисциплин: "Информатика", "Информационные технологии", "Математические методы в исторических исследованиях", "Компьютерные технологии в исторических исследованиях" и "Источниковедение", которые должны быть прослушаны до начала изучения данной дисциплины.

В результате освоения дисциплины "Методы и технологии работы с текстами" формируются компетенции, необходимые для изучения следующих дисциплин: "Электронные архивы", "Публикация исторических источников в электронной форме" и "Электронные документы в составе историко-культурного наследия современной России".

2. Структура дисциплины

Структура дисциплины для очно-заочной формы обучения

Общая трудоемкость дисциплины составляет 4 зачетных единицы, 152 час., в том числе контактная работа обучающихся с преподавателем 48 час., самостоятельная работа обучающихся 104 час.

| № п/п | Раздел дисциплины/темы | Виды учебной работы (в часах) | | | | | | Формы текущего контроля успеваемости, форма промежуточной аттестации (по семестрам) |
|----------|--|----------------------------------|---------|-------------------------|-------------------------|-------------------------------|-----------------------------|--|
| | | Контактная | | | | Промежуточ- ная аттестация | Самостоятель- ная работа | |
| | | Лекции | Семинар | Практические занятия | Лабораторные занятия | | | |
| 1 | Тема 1. Введение | 2 | 4 | | | | 12 | доклад |
| 2 | Тема 2. Методология и методика работы с текстовыми докумен- тами | 2 | 4 | | | | 12 | дискуссия |
| 3 | Тема 3. Концепции электронного текста | 4 | 6 | | | | 10 | контрольная |
| 4 | Тема 4. Традиционные и формализованные методы анализа со- держания. Контент- анализ | 4 | 6 | | | | 30 | дискуссия |
| 5 | Тема 5. Компьютеризован- ный анализ содержа- ния текстов | 4 | 4 | | | | 20 | Доклад |
| 6 | Тема 6. Опыт приме- нения контент- анализа при изучении различных типов ис- точников | 4 | 4 | | | | 20 | дискуссия |
| | Зачет | | | | | | | защита проекта |
| | ИТОГО: | 20 | 28 | | | | 104 | 152 |

3. Содержание дисциплины

Тема 1. Введение

- Традиционные методы анализа текстов.
- Направления и эволюция междисциплинарных методов анализа текстовых источников.
- Возможности формализованных методов изучения содержания текстов. Словари и конкордансы.
- История возникновения контент-анализа.
- Современные методы и технологии информационного поиска и анализа информационных потоков в сетевом пространстве (Text Mining, Content Monitoring).

Тема 2. Методология и методика работы с текстовыми документами

- Перевод "бумажных" текстов в электронный формат: сканирование и оптическое распознавание.
- Концепция явной (выраженной) и скрытой (структурной) информации.
- Методы структурной и семантической разметки текстов.
- Языки разметки (маркирования) текста. Язык SGML.

Тема 3. Концепции электронного текста

- Электронный текст как информационный ресурс.
- Жизненный цикл электронного текста. Основные концепции электронного текста.
- Коллекции и архивы электронных текстов. Полнотекстовые системы.
- Системы с контролируемым и свободным словарем.
- Методы информационного поиска в полнотекстовых системах. Тезаурусы.

Тема 4. Традиционные и формализованные методы анализа содержания. Контент-анализ

- Суть методов контент-анализа. Основные этапы контент-анализа.
- Смысловые единицы – индикаторы и категории.
- Виды контент-анализа. Частотный анализ.
- Проблема объективности контент-анализа.
- Контент-анализ и традиционные методы исследования текстов.

Тема 5. Компьютеризованный анализ содержания текстов

- Знакомство с программным обеспечением MAXQDA.
- Подготовка текста и создание полнотекстовой базы данных.
- Создание системы категорий. Лексический поиск и разметка.
- Частотный анализ. Построение таблиц сопряженности.
- Визуализация результатов анализа.
- Экспорт данных и возможности последующей статистической обработки.

Тема 6. Опыт применения контент-анализа при изучении различных типов источников

- Особенности контент-анализа при изучении хроник, летописей, религиозных текстов.
- Применение контент-анализа к материалам прессы.
- Возможности и специфика применения контент-анализа к источникам личного происхождения.
- Применение приемов контент-анализа к изучению делопроизводственной документации.
- Контент-анализ протоколов, стенограмм, политических документов.

4. Образовательные технологии

Образовательные технологии направлены на формирование общекультурных и профессиональных компетенций и ориентированы как на традиционные формы обучения, так и на интерактивный режим аудиторных и внеаудиторных занятий, коллективное и индивидуальное общение студентов с преподавателем и консультации в режиме удаленного доступа (форумы, электронная почта).

Изложение материала является проблемно-ориентированным. Теоретическая часть курса включает знакомство с основными подходами к созданию полнотекстовых баз данных и электронных текстовых архивов, с направлениями, методами и технологиями анализа содержания текстовых источников и проходит в форме лекций и семинаров с компьютерной поддержкой презентаций (частично может осуществляться в дистанционном режиме обучения). Практическое освоение программного обеспечения проходит на базе компьютерного класса, укомплектованного соответствующими программами. Учебно-методические материалы по курсу, а также рабочие файлы для лабораторных работ размещены в Интернете.

Изучение технологий и методов работы с текстовыми документами, а также обзор отечественного и зарубежного опыта работы с электронными текстами в гуманитарных исследованиях сочетает формы теоретического обучения и практических занятий с дискуссиями о возможностях и перспективах изучаемых методов в исторических исследованиях. В программе курса предусмотрено написание и обсуждение рефератов обзорного характера по материалам конкретно-исторических исследований из списка источников по курсу.

Интерактивность предполагает, что на семинарах студенты вместе с преподавателем обсуждают наиболее интересные темы и результаты исследований, а также собственные доклады и рефераты. Дискуссии, обсуждения и собеседования проводятся как в аудитории, так и в дистанционном режиме. Интерактивные формы обучения составляют около 35% аудиторной нагрузки по данной дисциплине.

Самостоятельная работа студентов строится на поиске и изучении информации, знакомстве с литературой, в том числе с научно-образовательными интернет-ресурсами, для подготовки к лекциям, семинарам, лабораторным работам, а также к текущей и промежуточной аттестации по курсу. Для работы с сетевыми ресурсами необходим доступ студентов в Интернет.

Существенной компонентой изучения дисциплины является использование авторского образовательного ресурса, размещенного на портале дистанционного образования (*E-learning*) исторического факультета МГУ им. М.В. Ломоносова и содержащего материалы для лекционных и семинарских занятий, самостоятельных работ, а также ссылки на полнотекстовые электронные ресурсы. В курсе также активно используется электронная библиотека Ассоциации "История и компьютер", содержащая электронные копии основных работ, рекомендованных студентам.

| № п/п | Наименование раздела | Виды учебной работы | Формируемые компетенции | Информационные и образовательные технологии |
|-------|--|--|-------------------------|--|
| 1 | 2 | 3 | 4 | 5 |
| 1. | Введение в проблематику | Лекции 1-2 Самостоятельная работа | ПК-4 | Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами |
| 2. | Методология и методика работы с текстовыми документами | Лекции 3-4 Семинар 1 Самостоятельная работа | ПК-4 | Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка доклада |
| 3. | Концепции электронного текста | Лекции 5-6 Семинар 2 Самостоятельная работа | ПК-4 | Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к дискуссии |
| 4. | Традиционные и формализованные методы анализа содержания. Контент-анализ | Лекции 7-8 Семинары 3-5 Самостоятельная работа | ПК-4 | Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к дискуссии |
| 5. | Компьютеризованный анализ содержания текстов | Лекции 9-10 Семинары 6-8 Лаб. работы 1-4 Самостоятельная работа | ПК-4 | Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Компьютерный практикум Подготовка к лаб. работам с использованием электронного курса лекций Подготовка к контрольной |
| 6. | Опыт применения контент-анализа при изучении различных типов источников | Лекции 11–12 Семинары 9-10 Самостоятельная | ПК-4 | Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка реферата |

| | | | | |
|--|--|--------|--|--|
| | | работа | | |
|--|--|--------|--|--|

5. Фонд оценочных средств для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

5.1. Система оценивания

| Форма контроля | Макс. количество баллов | |
|--------------------------|-------------------------|------------|
| | За одну работу | Всего |
| Текущий контроль: | | |
| – контрольная | 15 баллов | 30 баллов |
| – опрос | 5 баллов | 10 баллов |
| – реферат | 20 баллов | 20 баллов |
| Промежуточная аттестация | | 40 баллов |
| Итого за семестр | | 100 баллов |

5.2. Критерии выставления оценки по дисциплине

В качестве *текущего контроля* используются такие формы, как контрольные работы, доклады, рефераты, участие в дискуссиях и итоговое тестирование. Формой *промежуточной аттестации* является зачет.

Для получения зачета обучающийся в сумме должен набрать не менее 50 баллов из 100 возможных (итоговое тестирование – 25 баллов, контрольная работа – 25 баллов, доклад – 20 баллов, реферат – 20 баллов, дискуссии – две по 5 баллов).

Совокупный результат (максимум 100 баллов) конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; 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 |

| Баллы / Шкала ECTS | Оценка по дисциплине | Критерии оценки результатов обучения по дисциплине |
|--------------------|----------------------|---|
| 100–83 / A, B | 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 учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p> |
| 82–68 / C | зачтено | <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 | зачтено | <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> |

| | | |
|--------------|------------|---|
| | | Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный». |
| 49–0 / F, FX | Не зачтено | <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 для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется 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 (ПК-1-3)

Примерная тематика рефератов:

Обзор контент-аналитических исследований при изучении источников определенного вида:

- материалы прессы,
- источники личного происхождения,
- материалы массовой корреспонденции,
- документы государственных и политических организаций,
- летописи, хроники.

Для написания рефератов используются источники, выбранные из списка, приведенного в разделе "Учебно-методическое и информационное обеспечение дисциплины".

Требования к реферату. Объем реферата – не менее 10 тыс. знаков. В структуре реферата должны присутствовать:

- описание содержательной проблемы, поставленной в реферируемой работе,
- характеристика источниковой базы,
- характеристика подхода к созданию системы категорий и индикаторов,

- при использовании в работе выборочного метода – описание типа выборки,
- описание основных результатов исследования,
- оценка достоинств и недостатков работы.

Примерная тематика докладов:

- Междисциплинарная природа методов анализа текстовых источников.
- Методы анализа в задачах определения авторского стиля и установления авторства текстов.
- Методы анализа в задачах построения генеалогии списков древних текстов.
- Лингвистические методы формализованного анализа текстов. Словари и конкордансы.
- Документальные информационно-поисковые системы.
- Методы информационного поиска в полнотекстовых системах. Тезаурусы.
- Методы и технологии информационного поиска и анализа информационных потоков в сетевом пространстве.
- История возникновения контент-анализа.
- Типичные ошибки при проведении контент-анализа.

Вопросы к зачету:

- Эволюция методов анализа нарративных источников.
- Направления использования методов и технологий работы с текстами в исторических исследованиях.
- Понятие явной (выраженной) и скрытой (структурной) информации.
- Контент-анализ как междисциплинарный метод работы с текстами. Основные понятия контент-анализа. Смысловые единицы и единицы счета.
- Виды контент-анализа. Этапы контент-анализа.

- Частота встречаемости смысловых единиц, частота их связанности, "знак" высказывания.
- Проблема объективности в контент-анализе.
- Электронный текст как информационный ресурс. Коллекции и архивы электронных текстов. Полнотекстовые базы данных и информационные системы.
- Программное обеспечение компьютеризованного анализа текстовой информации.

Примеры вариантов контрольных заданий

- Контрольная работа по теме 5 (в качестве источников используются файлы с текстами В. Пуришкевича, Ф. Юсупова, В. Малькова – см. список источников).
1. Подготовка текстового файла к преобразованию в текстовую базу данных.
 - ✓ Провести предварительную структурную разметку текста в текстовом процессоре.
 - ✓ Импортировать размеченный документ.
 - ✓ Сформировать систему категорий.
 - ✓ Сохранить файл (проект) в формате MAXQDA.
 2. Семантическая разметка текста.
 - ✓ Открыть текстовую базу данных (проект).
 - ✓ Импортировать созданный на лабораторных работах словарь "стоп-слов".
 - ✓ Дополнить список "стоп-слов", используя словарь текста. Сохранить словарь "стоп-слов". Проверить результаты работы, анализируя облако тэгов.
 - ✓ Пользуясь словарем текста, сформировать систему категорий.
 - ✓ Провести автоматическую семантическую разметку текста по каждой категории с помощью лексического поиска по индикаторам.

- ✓ Сохранить списки индикаторов. Сохранить проект.
- 3. Частотный анализ текста.
 - ✓ Открыть текстовую базу данных (проект).
 - ✓ Построить матрицу "документы-коды", визуализировать результат.
 - ✓ Построить матрицу сопряженности категорий.
 - ✓ Экспортировать матрицы в MS Excel.
 - ✓ Преобразовать абсолютные частоты в относительные.
 - ✓ Построить диаграмму распределения категорий на множестве документов.
- 4. Статистическая обработка результатов контент-анализа в программе Statistica.
 - ✓ Открыть текстовую базу данных (проект).
 - ✓ Для группы выбранных документов экспортировать матрицы "документы-коды" в MS Excel.
 - ✓ Преобразовать абсолютные частоты в относительные.
 - ✓ Экспортировать таблицы MS Excel в программу Statistica.
 - ✓ Построить матрицу корреляции категорий.
 - ✓ Провести классификацию документов с помощью кластерного анализа.

6. Учебно-методическое и информационное обеспечение дисциплины:

6.1. Список источников и литературы

а) основная литература:

- Историческая информатика / Под ред. Л.И. Бородкина, И.М. Гарсковой. М., 1996. С. 107–142.
- Иудин А.А., Рюмин А.М. Контент-анализ текстов: компьютерные технологии. Учебное пособие. Нижний Новгород: ННГУ, 2010. 37 С.
- Захаров В.П. Информационные системы (документальный поиск). Учебное пособие. Спб, 2002. С. 6–26, 39–42.

- Ковальченко И.Д., Бородкин Л.И. Современные методы изучения исторических источников с использованием ЭВМ. М.: МГУ, 1987. С. 11–35.
- Количественные методы в исторических исследованиях. Учебное пособие / Отв. ред. И.Д. Ковальченко. М., Высшая школа, 1984. С. 345–367.
- Миронов Б.Н. История в цифрах. Л.: Наука, 1991. С. 14–29.
- *Петров А.Н. Компьютерный анализ текста. Историография метода // Круг идей: модели и технологии исторической информатики. Труды III конференции Ассоциации "История и компьютер". М., 1996. С. 255–277.
- *Тихонов В.И. Проблемы категоризации при контент-анализе // Круг идей: модели и технологии исторической информатики. Труды III конференции Ассоциации "История и компьютер". М., 1996.

б) дополнительная литература

- Бородкин Л.И. Контент-анализ и проблемы изучения исторических источников // Математика в изучении средневековых повествовательных источников. М., 1986. С. 8–30.
- Бородкин Л.И. Многомерный статистический анализ в исторических исследованиях. М., 1986.
- Когнитивные исследования за рубежом. (Идеи и методы искусственного интеллекта в изучении политического мышления) / Под ред. В.М. Сергеева. М., 1990. 267 С.
- Луков В.Б., Сергеев В.М. Опыт моделирования мышления исторических деятелей: Отто фон Бисмарк, 1866–1876 // Вопросы кибернетики. Логика рассуждений и ее моделирование. М., 1983. С. 65–90.
- Мангейм Дж.Б., Рич Р.К. Политология. Методы исследования. М., 1997. 544 С.
- Математика в изучении средневековых повествовательных источников. М., 1986. 151 С.

- Методы количественного анализа текстов нарративных источников. М., 1983. 132 С.
- Таршис Е.Я. Перспективы развития метода контент-анализа // Социология: методология, методы, математические модели, 2002, №15.
- Федотова Л.Н. Анализ содержания: социологический метод изучения средств массовой коммуникации. М.: Научный мир, 2001.
- Хьетсо Г., Густавссон С., Бекман Б., Гил С. Кто написал "Тихий Дон"? (Проблема авторства "Тихого Дона"). М., Книга, 1989. 320 С.
- Krippendorf K. Content Analysis. An Introduction to its Methodology. London: Sage, 2004. 2-nd edition.
- Kuckartz U. Qualitative Text Analysis. A Guide to Methods, Practice and Using Software. London: Sage, 2014.
- Laswell H.D., Pool I. de S. The Comparative Study of Symbols. Stanford, Calif., 1952. 286 P.
- Neuendorf K.A. The Content Analysis Guidebook. London: Sage Publications, 2002. 301 P.
- Weber R. Basic Content Analysis. Newbury Park. Calif., 1990. 51 P.

в) источники

- Архангельская И.Д. К вопросу изучения периодической печати методами контент-анализа // Методы количественного анализа текстов нарративных источников. М., 1983. С. 11–26.
- *Блинова О.Н. Опыт изучения экономических программ политических партий России 1990-х гг.: методика применения контент-анализа на примере программ КПРФ (1993, 1995) и АПР (1994) // Круг идей: историческая информатика в информационном обществе / Отв. редакторы Л.И. Бородкин, В.Н. Владимиров и И.Ф. Юшин. М., 2001. С. 212–235.

- Бородкин Л.И., Григорьева Ю.Г., Петров А.Н., Селунская Н.Б. Становление российского парламентаризма начала XX века (материалы к изучению I Государственной Думы и методы их анализа). М., 1996. 284 С.
- Брагина Л.М. Методика количественного анализа философских трактатов эпохи Возрождения // Математические методы в историко-экономических и историко-культурных исследованиях. М., 1977. С. 280–298.
- Брагина Л.М. Опыт исследования философского трактата 15 века методом количественного анализа // Математические методы в исторических исследованиях. М., 1972.
- Буховец О.Г. Математика в исследовании общественного сознания: крестьянские приговоры и наказания 1905-1907 гг. // Число и мысль. Вып. 9. М., 1986. С. 32–51.
- Буховец О.Г. Социальные конфликты и крестьянская ментальность в Российской Империи начала XX века: новые материалы, методы, результаты. М., 1996. 400 С.
- Давыдов М.А. Современники глазами А.П. Ермолова // Число и мысль. Вып. 9, М., 1986.
- Деопик Д.В. Опыт количественного анализа древней восточной летописи Чуньцю // Математические методы в историко-экономических и историко-культурных исследованиях. М., 1977. С. 144–190.
- *Ибрагимова Д.Х. Перестройка, рыночные ориентиры, сельский менталитет: многомерный анализ массовой корреспонденции в центральные органы печати // Круг идей: модели и технологии исторической информатики. Труды III конференции Ассоциации "История и компьютер". М., 1996.
- Кобринский А.Л. Проблемы государственного строительства в Российской Федерации (по материалам стенограмм пленарных заседаний Государственной Думы 1993–1995 гг.). М., 2001. 216 С.
- *Кобринский А.Л. Формула российской государственности: компьютеризованный анализ дебатов Государственной Думы первого созыва // Круг

идей: историческая информатика на пороге XXI века. Труды VI конференции Ассоциации "История и компьютер". М.–Чебоксары, 1999.

- Коханова Е.А. Пресса интернациональных бригад в годы гражданской войны в Испании (1936-1939 гг.): анализ электронного ресурса // Историческая информатика. 2013. №1. URL: http://kleio.asu.ru/2013/1/hcsj-12013_61-70.pdf.
- Красильников С.А. Периодическая печать Сибири в период Октября в системе политической пропаганды: К постановке проблемы и методов исследования // Книжное дело Сибири и Дальнего Востока в годы строительства социализма. Новосибирск, 1984. С. 56–74.
- Маджаров А.С. К вопросу о применении контент-анализа к источникам личного происхождения // Проблемы источниковедения и историографии истории Восточной Сибири. Иркутск, 1982. С. 16–38.
- Мазырин В.М. Применение методов контент-анализа к материалам прессы // Количественные методы в гуманитарных науках. М., 1981. С. 49–54.
- Мальков В.М. Раскулачивание. Как это было. Воспоминания о прожитом и пережитом с 1920 года по 1939 год. Екатеринбург, 1996.
- Малькова В.К. Применение контент-анализа для изучения интернационального и национального в республиканской прессе // Статистика в этнографии. М., 1985. С. 60–76.
- *Маркевич А.М. Солдатские письма во ВЦИК и Петросовет в 1917 г. К методике анализа // Круг идей: историческая информатика на пороге XXI века / Отв. ред. Л.И. Бородкин, Ю.П. Смирнов и И.Ф. Юшин. М. – Чебоксары, 1999. С. 151–173.
- Минаева А.П. "«Наивные» мемуары В.М. Малькова: опыт контент-анализа" [Электр. ресурс] // Фольклор и постфольклор: структура, типология, семиотика: [сайт]. Режим доступа: <http://www.ruthenia.ru/folklore/minaeva1.htm>.

- Миронов Б.Н. Статистическая обработка ответов на сенатскую анкету 1767 г. о причинах роста хлебных цен // Математические методы в исторических исследованиях. М., 1972.
- Миронов Б.Н. Формализация и генерализация содержания массовых источников: на материалах анкеты о ярмарках 1779 г. // Вспомогательные исторические дисциплины. Т. XIII. Л, 1982.
- *Поршнева О.С., Поршнев С.В. К вопросу об атрибуции текстов записей солдатских разговоров С.З. Федорченко // Информационный Бюллетень Ассоциации "История и компьютер". М. 2002. № 30.
- *Поршнева О.С., Поршнев С.В. К характеристике менталитета народных масс России: революция 1917г. в фокусе массового сознания (опыт статистического анализа писем рабочих, крестьян и солдат в центральные органы Советов рабочих и солдатских депутатов) // Круг идей: историческая информатика на пороге XXI века. Труды VI конференции Ассоциации "История и компьютер". М.–Чебоксары, 1999. С. 119–150.
- Пуришкевич В.М. Дневник [Электронный ресурс] // Документы XX века. Всемирная история в Интернете: [сайт]. Режим доступа: <http://doc20vek.ru/node/1431>.
- Раздина Н.В. Газета "За индустриализацию" как источник для изучения промышленной политики государства в годы первых пятилеток (опыт контент-анализа редакционных статей) // Историческая информатика. 2013. №1. URL: http://kleio.asu.ru/2013/1/hcsj-12013_41-60.pdf.
- *Рябухин И.В. Контент-анализ запросной деятельности в Государственной Думе Российской Империи, 1906–1912 гг. // Информационный бюллетень АИК. 2010. №36.
- Селунская В.М., Миронова Т.П. Опыт применения контент-анализа при изучении общественного сознания доколхозного крестьянства // Перестройка в исторической науке и проблемы источниковедения и специальных исторических дисциплин. Киев, 1990.

- *Тихонов В.И. Кто достоин избирательных прав? Типология и контент-анализ заявлений лишенцев // Круг идей: макро- и микроподходы в исторической информатике. Труды 5-й конференции Ассоциации "История и компьютер". М. – Минск, 1998. Т. 1. С. 140–177.
- Юсупов Ф.Ф. Конец Распутина. Париж, 1927. URL: http://www.rummuseum.ru/lib_yu/yusup_rasput00.php.

з) программное обеспечение

1. MS Office
2. MAXQDA
3. ABBY FineReader

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимый для освоения дисциплины

д) ресурсы Интернета

- полнотекстовые базы данных и электронные библиотеки в Интернете:
<http://www.vehi.net/> – библиотека "Вехи",
<http://www.gumer.info/> – библиотека Гумер и др.
- электронные справочники и руководства по методике исследований:
<http://academic.csuohio.edu/kneuendorf/content/> – The Content Analysis Guidebook Online
<http://users.ox.ac.uk/~ctitext2/index.html> – CTI Centre for Textual Studies и др.
- центры дистанционного образования:
<http://www.hist.msu.ru/VLE/> – портал дистанционного образования исторического факультета МГУ им. М.В.Ломоносова.

7. Материально-техническое обеспечение дисциплины

1. Компьютерный класс с установленным программным обеспечением MAXQDA, ABBY FineReader
2. Подключение компьютеров к Интернету.

- Состав программного обеспечения (ПО), современных профессиональных баз данных (БД) и информационно-справочные систем (ИСС)

- **Перечень БД и ИСС (к п. 6.2)**

| №п /п | Наименование |
|-------|--|
| 1 | Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus |
| 2 | Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis |
| 3 | Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru |

- **Состав программного обеспечения (ПО) (к п. 7)**

| №п /п | Наименование ПО | Производитель | Способ распространения (лицензионное или свободно распространяемое) |
|-------|-----------------------------|------------------|---|
| 1 | Adobe Master Collection CS4 | Adobe | лицензионное |
| 2 | Microsoft Office 2010 | Microsoft | лицензионное |
| 3 | Windows 7 Pro | Microsoft | лицензионное |
| 6 | SPSS Statistics 22 | IBM | лицензионное |
| 7 | Microsoft Share Point 2010 | Microsoft | лицензионное |
| 8 | SPSS Statistics 25 | IBM | лицензионное |
| 9 | Microsoft Office 2013 | Microsoft | лицензионное |
| 10 | ОС «Альт Образование» 8 | ООО «Базальт СПО | лицензионное |
| 11 | Microsoft Office 2013 | Microsoft | лицензионное |
| 12 | Windows 10 Pro | Microsoft | лицензионное |
| 13 | Kaspersky Endpoint Security | Kaspersky | лицензионное |
| 14 | Microsoft Office 2016 | Microsoft | лицензионное |
| 15 | Visual Studio 2019 | Microsoft | лицензионное |
| 16 | Adobe Creative Cloud | Adobe | лицензионное |
| 17 | Zoom | Zoom | лицензионное |

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 EmBraille ViewPlus;
- 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 практических занятий (с указанием тематических разделов и количества часов)

Тема 2. Методология и методика работы с текстовыми документами (2 час.)

1. Концепция явной (выраженной) и скрытой (структурной) информации.
2. Методы структурной и семантической разметки текстов.
3. Языки разметки (маркирования) текста. Язык SGML.
4. Выступления студентов с докладами по методологии и методике работы с текстовыми документами

Контрольные вопросы:

- Кодирование грамматических классов слов.
- Кодирование разночтений в версиях списков текста.
- Системы структурной разметки текста.

Литература:

- Историческая информатика / Под ред. Л.И. Бородкина, И.М. Гарсковой. М., 1996. С. 107–142.
- Мангейм Дж.Б., Рич Р.К. Политология. Методы исследования. М., 1997. С. 269–291.

Тема 4. Контент-анализ (10 час.)

1. Дискуссия по методике исследования на примере работы А.П. Минаевой "«Наивные» мемуары В.М. Малькова: опыт контент-анализа".
2. Дискуссия по процедуре формирования категорий и индикаторов – обсуждение результатов самостоятельной работы студентов по материалам анализа текстов В. Пуришкевича и Ф. Юсупова.

Контрольные вопросы:

- Причины субъективности результатов контент-анализа.
- Рекомендации к построению системы категорий и индикаторов.
- Два способа построения категорий: сверху вниз и снизу вверх.
- Чем различаются понятия "смысловая единица" и "единица счета"?

Источники и литература:

- Минаева А.П. "«Наивные» мемуары В.М. Малькова: опыт контент-анализа" [Электр. ресурс] // Фольклор и постфольклор: структура, типология, семиотика: [сайт]. Режим доступа:
<http://www.ruthenia.ru/folklore/minaeva1.htm>.
- Пуришкевич В.М. Дневник [Электронный ресурс] // Документы XX века. Всемирная история в Интернете: [сайт]. Режим доступа:
<http://doc20vek.ru/node/1431>.

- Юсупов Ф.Ф. Конец Распутина. Париж, 1927. URL:
http://www.rummuseum.ru/lib_yu/yusup_rasput00.php.
- Мангейм Дж.Б., Рич Р.К. Политология. Методы исследования. М., 1997. С. 269–291.

Тема 5. Компьютеризованный контент-анализ (10 час.)

1. Обсуждение результатов выполнения контрольной работы с программой MAXQDA с разбором методических и технических ошибок.

Контрольные вопросы:

- Типичные ошибки при построении системы категорий и индикаторов.
- Ошибки при структурной и семантической разметке текста.
- Что в программах контент-анализа понимается под контекстом?
- Различные варианты определения совместной встречаемости категорий.
- Абсолютные и относительные частоты при визуализации и интерпретации результатов статистики текста.

Литература:

- Ковальченко И.Д., Бородкин Л.И. Современные методы изучения исторических источников с использованием ЭВМ. М.: МГУ, 1987. С. 11–35.
- Миронов Б.Н. История в цифрах. Л.: Наука, 1991. С. 14–29.
- Историческая информатика / Под ред. Л.И. Бородкина, И.М. Гарсковой. М., 1996. С. 133–141.
- Петров А.Н. Компьютерный анализ текста. Историография метода // Круг идей: модели и технологии исторической информатики. Труды III конференции Ассоциации "История и компьютер". М., 1996. С. 255–277.

Тема 6. Опыт применения контент-анализа при изучении различных типов источников (10 час.)

1. Защита рефератов

Планы лабораторных работ и методические указания по организации и проведению

Указания

Лабораторные работы выполняются с помощью специализированного программного обеспечения MAXQDA11. Коллекция текстов включает стенограммы дебатов I Государственной Думы (1906 г.) по законопроекту о равноправии, дневник В.М. Пуришкевича, воспоминания Ф. Юсупова, материалы "круглого стола" по теме "Историк, источник и Интернет" и др. тексты. Коллекция файлов доступна в Интернете на сайте исторического факультета МГУ (<http://www.hist.msu.ru/Departments/Inf/Stud/index.html>).

Тема 5. Компьютеризованный контент-анализ (работа с материалами круглого стола и стенограммами I Государственной Думы) (10 час.).

1. Знакомство с программой MAXQDA:
 - установка программы,
 - интерфейс, главное меню, языковые настройки,
 - документы, классы и группы документов, коды (категории) и фрагменты.
2. Организация данных:
 - допустимые форматы данных,
 - структурная разметка текстов и импорт,
 - создание первого кода (категории).
 - сохранение проекта в формате MAXQDA.
3. Работа со словарем текста:
 - частотный словарь,
 - создание словаря "стоп-слов",
 - импорт словаря "стоп-слов".
 - использование облака тэгов.
4. Разработка системы категорий:

- выявление значимых слов с наибольшей частотой,
 - контекстный поиск потенциальных категорий,
 - стандартный и расширенный лексический поиск и разметка,
 - сохранение списков индикаторов,
 - редактирование системы категорий,
 - автоматическое кодирование фрагментов текста.
5. Работа с фрагментами текста, частотный анализ:
- активация документов и кодов,
 - просмотр и редактирование фрагментов,
 - матрица "документы-коды", возможности визуализации.
6. Анализ сопряженности категорий:
- пересечение категорий,
 - другие варианты сопряженности категорий,
 - уровни пересечения (близости) кодов,
 - матрица связи кодов, возможности визуализации.
7. Работа с переменными:
- переменные документов,
 - создание переменных путем преобразования кодов,
 - переменные кодов (категорий),
 - статистика переменных и ее визуализация,
 - активация документов и кодов по значениям переменных,
 - типология статистики количественных переменных по значениям качественных (категориальных) переменных.
8. Статистическая обработка результатов контент-анализа в табличном процессоре и программе Statistica:
- экспорт результатов в MS Excel,
 - "взвешивание" абсолютных частот и переход к процентам,
 - использование графических возможностей MS Excel,
 - экспорт из MS Excel в программу Statistica,

- статистическая обработка данных: классификация документов (авторов), корреляция кодов.

Контрольные вопросы:

- Что такое препроцессинг?
- Какие форматы текстовых файлов можно импортировать в MAXQDA?
- Как создать группу документов?
- Виды сопряженности категорий.
- Графические возможности программы MAXQDA.
- Можно ли категорию преобразовать в переменную документа?
- Возможности статистики переменных в программе MAXQDA.

Литература:

- Гарскова И.М. Компьютеризованный контент-анализ. Учебно-методическое пособие. М., 2015.

9.2. Методические рекомендации по подготовке письменных работ

Самостоятельная работа нацелена на расширение теоретических знаний и практических навыков на основе чтения рекомендуемой литературы, знакомства с научно-образовательными тематическими интернет-ресурсами.

Для подготовки к лекциям, семинарам, лабораторным и контрольным работам, докладам, рефератам, а также к промежуточной аттестации студенты используют списки источников и литературы, материалы лекций и презентаций по данным темам, список тем докладов и рефератов, вопросов к зачету, планы семинаров и лабораторных работ и списки контрольных вопросов к ним, примеры контрольных заданий.

Для подготовки к лабораторным и контрольным работам и выполнения практических заданий на компьютере используется программное обеспечение MAXQDA.

Самостоятельная работа может выполняться студентом в библиотеке, в компьютерном классе или дома.

| Вид работы | Содержание | Трудоемкость | Рекомендации |
|--|---------------------------------------|-----------------|--|
| Подготовка к лекциям и семинару по темам 1-2 | Тема 1 (см. тематический план) | 10 час. | См. список литературы, интернет-ресурсы, планы семинарских занятий и контрольные вопросы, |
| Подготовка доклада по теме 2 | Тема 2 (см. тематический план) | 10 час. | См. темы докладов, список литературы, материалы лекции по данной теме и интернет-ресурсы |
| Подготовка к лекциям и семинару по теме 3 | Тема 3 (см. тематический план) | 10 час. | См. список литературы, интернет-ресурсы, планы семинарских занятий и контрольные вопросы, |
| Подготовка к дискуссии по теме 3 | Тема 3 (см. тематический план) | 10 час. | См. планы семинарских занятий и контрольные вопросы, список источников и литературы, материалы лекций по данной теме |
| Подготовка к лекциям и семинару по теме 4 | Тема 4 (см. тематический план) | 10 час. | См. список литературы, интернет-ресурсы, планы семинарских занятий и контрольные вопросы, |
| Подготовка к дискуссии по теме 4 | Тема 4 (см. тематический план) | 10 час. | См. планы семинарских занятий и контрольные вопросы, список источников и литературы, материалы лекций по данной теме |
| Подготовка к лекциям и семинарам по теме 5 | Тема 5 (см. тематический план) | 10 час. | См. список литературы, интернет-ресурсы, планы семинарских занятий и контрольные вопросы, |
| Подготовка к лаб. работам по теме 5 | Тема 5 (см. тематический план) | 10 час. | См. список литературы, интернет-ресурсы, планы лаб. работ и контрольные вопросы, |
| Подготовка к контрольной по теме 5 | | 10 час. | См. планы лабораторных работ и контрольные вопросы, список источников и примеры контрольных заданий |
| Подготовка к лекциям и семинарам по теме 6 | Тема 6 (см. тематический план) | 10 час. | См. список литературы, интернет-ресурсы, планы семинарских занятий и контрольные вопросы, |
| Подготовка реферата по теме 6 | Тема 5 (см. тематический план) | 10 час. | См. темы рефератов, список источников и литературы, материалы лекции по данной теме и интернет-ресурсы |
| Промежуточная аттестация | Письменный зачет | 22 час. | См. вопросы к зачету, список литературы, материалы лекций и интернет-ресурсы |
| Итого по курсу | | 132 час. | |

Аннотация

Дисциплина "Методы и технологии работы с текстами" входит в состав части, формируемой участниками образовательных отношений, блока дисциплин учебного плана бакалаврской программы «Электронные архивы и документы» по направлению подготовки 46.03.02 – Документоведение и архивоведение и адресована студентам 3 курса (5 семестр). Дисциплина реализуется кафедрой источниковедения факультета архивоведения и документоведения Историко-архивного института РГГУ.

Цель курса: дать студентам теоретические и практические знания о современных принципах анализа текстов, способах обработки, вторичного использования и извлечения скрытой информации, представленной в описательном (неструктурированном) виде, а также об информационных технологиях, на которых основан компьютеризованный анализ текстовой информации.

Задачи курса: обзор междисциплинарных методов и технологий работы с текстами в гуманитарных исследованиях (в сфере истории, политологии, социологии, лингвистики), базирующихся на системном подходе и формализованных методах обработки текстовой информации; теоретические концепции анализа текстовой информации и их эволюция; специфика и анализ опыта применения методов и технологий работы с текстами в исторических исследованиях; характеристика контент-анализа как качественно-количественного метода анализа содержания; понятийный аппарат контент-анализа, его виды и этапы; изучение основных подходов к автоматизации контент-анализа, определение их достоинств и недостатков; знакомство с программным обеспечением компьютеризованного анализа текстовой информации.

Дисциплина направлена на формирование следующих компетенций выпускника: владение тенденциями развития информационно-документационного и обеспечения управления архивного дела (ПК 1.3).

В результате освоения дисциплины обучающийся должен:

Знать основные методы и подходы, перспективные направления развития информационных технологий в области анализа текстов; специфику использования этих методов и технологий в различных гуманитарных науках; зарубежный и отечественный опыт формализованного анализа текстовой информации исторических источников; методологию и методику контент-анализа; основные характеристики и концепции электронного текста, принципы функционирования прикладного программного обеспечения, предназначенного для компьютеризованного анализа неструктурированной текстовой информации; основы проектирования текстовых баз данных. Уметь сканировать текстовые материалы для перевода их в электронный формат; определять характеристики текстовых документов, которые с необходимостью отражали бы существенные стороны их содержания; проводить структурную и семантическую разметку (маркирование) текстов; создавать полнотекстовые базы данных и выполнять информационный поиск; работать со словарем текста; подсчитывать количественные характеристики совместной встречаемости лексических единиц и семантических категорий; визуализировать результаты анализа и экспортировать их для дальнейшей обработки в статистические программы. Владеть аппаратными и программными средствами перевода текстов в электронный формат; методами маркирования текстов и информационного поиска в полнотекстовых системах; методами математической статистики, связанными с дескриптивной статистикой и анализом взаимосвязей, их графическим представлением; современными информационными технологиями работы с электронными текстовыми архивами.

Рабочей программой предусмотрены следующие виды контроля: текущий контроль успеваемости в форме контрольных, участия в дискуссиях, а также подготовки докладов и рефератов; промежуточная аттестация в форме зачета с оценкой.

Общая трудоемкость освоения дисциплины составляет 4 зачетных единицы (152 часа).