

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Социологический факультет

УТВЕРЖДАЮ
(декан)

_____/Н.Г. Осипова/
« ____ » _____ 2019 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«Большие данные: возможности и перспективы»
«Big data: opportunities and prospects»

Уровень высшего образования:
Магистратура

Направление подготовки (специальность):

39.04.01 СОЦИОЛОГИЯ

Направленность (профиль) ОПОП:

«Современные технологии социологических исследований»

Форма обучения:

Очная

Рабочая программа рассмотрена и одобрена

На заседании Ученого Совета факультета

(протокол №__ от _____ 2019 г.)

Москва 2019

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 39.04.01 Социология (уровень магистратуры) в редакции приказа МГУ от _____ 2019 года.

Год (годы) приема на обучение: 2019.

1. Место дисциплины (модуля) в структуре ОПОП ВО: относится к вариативной части профессионального цикла, дисциплины по выбору, 3 семестр.

2. Входные требования для освоения дисциплины (модуля), предварительные условия (если есть): освоение дисциплин: «Технология социального измерения», «Технологические подходы к конструированию социологического исследования».

3. Результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников

Планируемые результаты обучения по дисциплине (модулю)

ЗНАТЬ: современные информационно-коммуникационные технологии в сфере профессиональной деятельности

ЗНАТЬ: Возможности и ограничения применения современных информационно-коммуникационные технологии при решении профессиональных задач

ЗНАТЬ: методы (соответственно: социологические, статистические, демографические и др.) анализа профессиональной информации

ЗНАТЬ: Современные информационные технологии, которые можно применить для эффективного решения поставленных задач при проведении, планировании и формировании социологических исследований

ЗНАТЬ: Возможности и интерфейс аналитических ресурсов, парсинга, построенных на основе технологии больших данных

УМЕТЬ: обосновывать необходимость разработки новые подходов и методов исследования применительно к задачам социологического исследования

УМЕТЬ: разрабатывать новые подходы и методы с учетом цели и задач исследования на основе современных теорий и моделей исследования

УМЕТЬ: анализировать и интерпретировать профессиональную (социальную, демографическую ...) информацию в соответствии с выбранным концептуальным подходом

УМЕТЬ: критически оценивать поступающую информацию, избегать автоматического применения стандартных приемов при решении исследовательских задач

УМЕТЬ: Использовать программное обеспечение для анализа данных в социологии

УМЕТЬ: применять информационные технологии для планирования, организации, проведения и мониторинга

УМЕТЬ: использовать аналитические ресурсы, парсинга на основе больших данных для решения задач при проведении, планировании и формировании социологических исследований (Компетенция СПК-1);

ВЛАДЕТЬ: навыками работы со специализированным профессиональным оборудованием и программным обеспечением качественных и количественных социологических исследований

ВЛАДЕТЬ: аналитическими навыками обработки информации, полученной в результате работы на платформах, построенных на основе больших данных

ВЛАДЕТЬ: инструментами платформ, построенных на основе больших данных

4. Формат обучения: очная.

5. Объем дисциплины (модуля) составляет 4 з.е., в том числе 28 академических часов, отведенных на контактную работу обучающихся с преподавателем из них 14лекционных

и 14семинарских часов, 116 академических часов на самостоятельную работу обучающихся.

6. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий:

Наименование и краткое содержание разделов и дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю)	Все го (часы)	В том числе				Форма текущего контроля успеваемости (опрос, тестирование, коллоквиум, контрольная работа, реферат и т.п.)
		Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, часы			Самостоятельная работа обучающегося, часы	
		Занятия лекционного типа*	Занятия семинарского типа*	Все го		
Тема 1. Введение. Большие данные: основные вопросы.	32	4	4	8	24	устный опрос; эссе; практические работы; выполнение проектов
Тема 2. Мультифункциональные сервисы для проведения анкетирования	28	2	2	4	24	устный опрос; эссе; практические работы; выполнение проектов внутрисеместровая аттестация №1
Тема 3. Парсинг. Ресурсы и технологии	32	4	4	8	24	устный опрос; эссе; практические работы; выполнение проектов внутрисеместровая аттестация №2
Тема4. Аналитические инструменты социальных сетей.	32	4	4	8	24	устный опрос; эссе; практические работы; выполнение проектов
Промежуточная аттестация (экзамен)					20	

Итого	144	28	116	
--------------	------------	-----------	-----	--

п/п	Раздел	Содержание (темы)
1	Тема 1. Введение. Большие данные: основные вопросы.	<p>Определения данных: философский и юридический аспекты. Большие данные в англоязычных источниках.</p> <p>Что делает данные «большими» и откуда берутся большие данные? Что способствовало появлению больших данных? Ценность больших данных. Успешные примеры использования больших данных. Начало работы с источниками больших данных: машинные данные, машиностроительные данные, большие данные, полученные от людей: неструктурированные данные, данные организаций и компаний: структурированные, но зашумленные, создаваемые организацией данные: выгоды от объединения с другими типами данных. Интеграция разнообразных данных. Наука о данных: получение ценности из больших данных. Создание стратегии больших данных. Пять компонентов науки о данных (v).</p> <p>Возможности использования инструмента больших данных при проведении социологических исследований.</p>
2	Тема2. Мультифункциональные сервисы для проведения анкетирования	<p>Обзор сервисов и платформ для создания и проведения анкетирования он-лайн, анализ результатов из перечня сервисов ресурсного обеспечения. Англоязычные, русскоязычные, смешанные ресурсы. Сущность, задачи, сравнительные характеристики. Условия и возможности применения. Возможности использования инструмента больших данных при проведении социологических исследований. Реализация.</p>

3	Тема 3. Парсинг. Ресурсы и технологии Инструменты таргетинга и поиска целевой аудитории.	Определение. Разновидности парсинга. Этапы парсинга. Преимущества технологии. Правовые нормы, применяемые к парсингу. Защита парсинга. Сервисы –парсеры из перечня ресурсного обеспечения. Сравнительная характеристика, возможности применения. Реализация. Понятие таргетированной рекламы. Поиск целевой аудитории. Англоязычные, русскоязычные, смешанные ресурсы. Возможности использования инструмента больших данных при проведении социологических исследований.
4	Тема 4. Аналитические инструменты социальных сетей.	Обзор. Область применения. Результаты применения. Сравнительные характеристики. Модели и технологии. Англоязычные, русскоязычные, смешанные ресурсы. Возможности использования инструмента больших данных при проведении социологических исследований. Реализация

7. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине (модулю)

7.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости.

Контрольные вопросы:

Тема 1.

1. Дайте определения данных: философский и юридический аспекты.
2. Что делает данные «большими» и откуда берутся большие данные?
3. Что способствовало появлению больших данных?
4. Ценность больших данных. Приведите примеры использования больших данных.
5. Перечислите источники больших данных.
6. Возможность интеграции разнообразных данных.
7. Создание стратегии больших данных.
8. Перечислите пять компонентов науки о данных (v).

Тема 2.

1. Перечислите инструменты для создания и проведения анкетирования .
2. Расширьте список инструментов.
3. Дайте характеристику инструментов.

4. Проведите сравнительный анализ: по сфере применения, технологии, доступности, удобстве использования, конечному результату и защите информации

Тема 3.

1. Сформулируйте определение парсинга.
2. Перечислите этапы парсинга.
3. Технологии парсинга.
4. Правовые нормы, применяемые к парсингу.
5. Защита информации.
6. Дайте определение понятия таргетированной рекламы.
7. Перечислите инструменты для парсинга.
8. Расширьте список инструментов.
9. Дайте характеристику инструментов.
10. Проведите сравнительный анализ: по сфере применения, технологии, доступности, удобстве использования, конечному результату и защите информации.

Тема 4.

1. Перечислите инструменты для аналитики социальных сетей.
2. Расширьте список инструментов.
3. Дайте характеристику инструментов.
4. Проведите сравнительный анализ: по сфере применения, технологии, доступности, удобстве использования, конечному результату и защите информации

Практические работы:

1. С помощью сервиса «Анкетолог» создайте опрос, соберите ответы, получите результаты, проведите анализ.
2. С помощью сервиса «SurveyMonkey» создайте опрос, соберите ответы, получите результаты, проведите анализ.
3. Выберите сервис для проведения анкетирования, наиболее «подходящий» Вам, проведите полный цикл от настройки анкеты до обработки результатов.
4. Преобразуйте веб-сайт в структурированные, полезные данные по URL с помощью Import.io.
5. Соберите web-данные с помощью Webhose.io
6. Соберите данные с помощью CloudScrape. Сохраните собранные данные в облаке, например, Google Drive .

7. Соберите данные с нескольких страниц с помощью различных инструментов по выбору из перечня ресурсного обеспечения (сервисы-парсеры и сервисы-аналитики) и экспортируйте в один из форматов CSV, XML, JSON и SQL.
8. Соберите данные с помощью Spinn3r из блогов или новостных лент.
9. Соберите информацию с web-страницы с помощью ресурса OutWit Hub
10. Выделите пять платформ для аналитики социальных сетей, проведите исследование по контенту, используя различные подходы к запросу: по ключевому слову, хештегу, адресу. Постройте отчет.

Темы эссе, рефератов:

1. Возможности использования технологий больших данных для проведения социологических исследований. Проработанность темы в англоязычных странах (на английском языке)
2. Большие данные в современном мире. Обзор англоязычных источников (на английском языке).
3. Большие данные современные подходы к сбору и хранению информации (на английском языке).
4. Технологии больших данных в современном мире (на английском языке).

7.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации.

Вопросы для подготовки к экзамену.

1. Дайте определения данных: философский и юридический аспекты.
2. Что делает данные «большими» и откуда берутся большие данные?
3. Что способствовало появлению больших данных?
4. Ценность больших данных. Приведите примеры использования больших данных.
5. Перечислите источники больших данных.
6. Возможность интеграции разнообразных данных.
7. Создание стратегии больших данных.
8. Перечислите пять компонентов науки о данных (v).
9. Перечислите инструменты для создания и проведения анкетирования.
10. Расширьте список инструментов.
11. Дайте характеристику инструментов.
12. Проведите сравнительный анализ: по сфере применения, технологии, доступности, удобстве использования, конечному результату и защите информации.
13. Сформулируйте определение парсинга.
14. Перечислите этапы парсинга.
15. Технологии парсинга.
16. Правовые нормы, применяемые к парсингу.

17. Защита информации.
18. Дайте определение понятия таргетированной рекламы.
19. Перечислите инструменты для парсинга.
20. Расширьте список инструментов.
21. Дайте характеристику инструментов.
22. Проведите сравнительный анализ: по сфере применения, технологии, доступности, удобстве использования, конечному результату и защите информации.
23. Перечислите инструменты для аналитики социальных сетей. Расширьте список инструментов.
24. Дайте характеристику инструментов.
25. Проведите сравнительный анализ: по сфере применения, технологии, доступности, удобстве использования, конечному результату и защите информации.
26. Оцените возможность использования ресурсов для проведения анкетирования в социологическом исследовании.
27. Оцените возможность использования сервисов-парсеров в социологическом исследовании.
28. Оцените возможность использования аналитических инструментов в социологическом исследовании.

Примерная тематика исследовательских проектов.

1. Возможности аппарата BigData в социологии (на английском языке).
2. Задачи социологии, решаемые с помощью больших данных(на английском языке).
3. Готовые решения BigData для социологов-исследователей (на английском языке).
4. По теме научной диссертации с помощью аналитических инструментов проведите аналитику социальной сети: ВКонтакте, Инстаграмм(на английском языке).
5. По теме научной диссертации проведите аналитику видеохостинга Ютуб(на английском языке).
6. Изучение основ программирования на языке Python.
7. Системы моделирования и управления большими данными(на английском языке).
8. Моделирование больших данных. MapReduce(на английском языке).
9. Работа с моделями данных(на английском языке).
10. Основы для систем больших данных и программирования. Hadoop(на английском языке).

8. Ресурсное обеспечение:

8.1. Учебно-методическое и информационное обеспечение дисциплины

Научная библиотека МГУ

Электронные библиотеки

Электронная библиотечная система IPRBOOKS

а) основная литература:

1. Адлер Ю.П. Статистическое управление процессами. «Большие данные» [Электронный ресурс]: учебное пособие/ Адлер Ю.П., Черных Е.А.— Электрон. текстовые данные.— Москва: Издательский Дом МИСиС, 2016.— 52 с.— Режим доступа: <http://www.iprbookshop.ru/64199.html>.— ЭБС «IPRbooks»
2. Виктор Майер-Шенбергер Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим [Электронный ресурс]/ Виктор Майер-Шенбергер, Кеннет Кукьер— Электрон. текстовые данные.— Москва: Манн, Иванов и Фербер, 2014.— 220 с.— Режим доступа: <http://www.iprbookshop.ru/39150.html>.— ЭБС «IPRbooks»
- 3.Хименко В.И. Случайные данные. Структура и анализ [Электронный ресурс]/ Хименко В.И.— Электрон. текстовые данные.— Москва: Техносфера, 2017.— 424 с.— Режим доступа: <http://www.iprbookshop.ru/84706.html>.— ЭБС «IPRbooks»
- 4.Воронова Л.И. Big Data. Методы и средства анализа [Электронный ресурс]: учебное пособие/ Воронова Л.И., Воронов В.И.— Электрон. текстовые данные.— Москва: Московский технический университет связи и информатики, 2016.— 33 с.— Режим доступа: <http://www.iprbookshop.ru/61463.html>.— ЭБС «IPRbooks»

б) дополнительная литература:

1. Билл Фрэнкс Революция в аналитике [Электронный ресурс]: как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики/ Билл Фрэнкс— Электрон. текстовые данные.— Москва: Альпина Паблишер, 2017.— 320 с.— Режим доступа: <http://www.iprbookshop.ru/58563.html>.— ЭБС «IPRbooks»
2. Федин Ф.О. Анализ данных. Часть 2. Инструменты Data Mining [Электронный ресурс]: учебное пособие/ Федин Ф.О., Федин Ф.Ф.— Электрон. текстовые данные.— Москва: Московский городской педагогический университет, 2012.— 308 с.— Режим доступа: <http://www.iprbookshop.ru/26445.html>.— ЭБС «IPRbooks».
3. Чубукова И.А. Data Mining [Электронный ресурс]/ Чубукова И.А.— Электрон. текстовые данные.— Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016.— 470 с.— Режим доступа: <http://www.iprbookshop.ru/56315.html>.— ЭБС «IPRbooks»
4. Воронов В.И. Data Mining - технологии обработки больших данных [Электронный ресурс]: учебное пособие/ Воронов В.И., Воронова Л.И., Усачев В.А.— Электрон. текстовые данные.— Москва: Московский технический университет связи и информатики, 2018.— 47 с.— Режим доступа: <http://www.iprbookshop.ru/81324.html>.— ЭБС «IPRbooks»
5. Семичастный И.Л. Информационно-коммуникационные технологии. Часть 1 [Электронный ресурс]: конспект лекций для студентов ОУ «бакалавр» направлений подготовки 38.03.01 «Экономика», 38.03.02 «Менеджмент», 38.03.03 «Управление персоналом», 38.03.04 «Государственное и муниципальное управление» очной формы обучения/ Семичастный И.Л.— Электрон. текстовые данные.— Донецк: Донецкий государственный университет управления, 2016.— 123 с.— Режим доступа: <http://www.iprbookshop.ru/62360.html>.— ЭБС «IPRbooks»
6. Гасумова С.Е. Информационные технологии в социальной сфере [Электронный ресурс]: учебное пособие для бакалавров/ Гасумова С.Е.— Электрон. текстовые данные.— М.: Дашков и К, 2019.— 311 с.— Режим доступа: <http://www.iprbookshop.ru/85325.html>.— ЭБС «IPRbooks»
7. Журавлева Т.Ю. Информационные технологии [Электронный ресурс]: учебное пособие/ Журавлева Т.Ю.— Электрон. текстовые данные.— Саратов: Вузовское образование, 2018.— 72 с.— Режим доступа: <http://www.iprbookshop.ru/74552.html>.— ЭБС «IPRbooks»

8. Современные компьютерные технологии [Электронный ресурс]: учебное пособие/ Р.Г. Хисматов [и др.].— Электрон. текстовые данные.— Казань: Казанский национальный исследовательский технологический университет, 2014.— 83 с.— Режим доступа: <http://www.iprbookshop.ru/62279.html>.— ЭБС «IPRbooks»
9. Персова М.Г. Современные компьютерные технологии [Электронный ресурс]: конспект лекций/ Персова М.Г., Соловейчик Ю.Г., Домников П.А.— Электрон. текстовые данные.— Новосибирск: Новосибирский государственный технический университет, 2014.— 80 с.— Режим доступа: <http://www.iprbookshop.ru/45025.html>.— ЭБС «IPRbooks»

Перечень информационных технологий

Интернет-ресурсы:

1. Журнал «Социс. Социологические исследования» <http://www.nir.ru/socio/skipubl/socis.htm> -
2. Журнал «Вестник Московского университета. Серия 18. Социология и политология» <http://www.vestnik.socio.msu.ru/>
3. Российская государственная библиотека <http://www.rsl.ru/>
4. Научная электронная библиотека <http://www.elibrary.ru>
5. Национальная электронная библиотека <http://www.nel.nns.ru/>
6. Федеральная служба государственной статистики РФ <http://www.gks.ru>
7. Статистическая служба Европейского союза <http://ec.europa.eu>
8. Электронная библиотека МГУ имени М.В.Ломоносова <http://www.nbmgu.ru/publicdb/>
<http://www.fgosvo.ru>
9. Портал ФГОС ВО
10. Coursera. Введение в большие данные. <https://www.coursera.org/learn/big-data-introduction>
<https://www.coursera.org/specializations/big-data>

Технологии bigdata

Apache Hadoop <https://hadoop.apache.org/>

Teradata: Business Analytics, Cloud Analytics & Enterprise <https://www.teradata.com>

IBM <https://www.ibm.com/developerworks/ru/library/bd-mapreduce/index.html>

Инструменты для создания анкет и опросов онлайн

1. Survey Monkey <https://ru.surveymonkey.com/>
2. Survio <https://www.survio.com/ru/anketnaja-sistema-funkcii>

3. Google forms <https://docs.google.com/forms/u/0/>
4. Let's test <https://letstest.ru/>
5. ProProfs <https://www.proprofs.com/>
6. Анкетолог <https://anketolog.ru>

Теоретическое описание парсинга и этапов: <https://www.seonews.ru/glossary/parsing/>

Инструменты парсинга:

1. Import.io <https://www.import.io>
2. Webhose.io <https://webhose.io>
3. dexi.io <https://dexi.io>
4. Scrapinghub <https://scrapinghub.com>
5. ParseHub <https://www.parsehub.com/>
6. VisualScaper <http://visualscaper.blogspot.com>
7. Spinn3r [Spinn3r](https://spinn3r.com)
8. 80legs <http://80legs.com/>
9. Scraper <http://scraper.com/>
10. OutWit Hub <https://www.outwit.com/products/hub/>
11. FMiner <http://www.fminer.com>
12. Content Grabber <https://sequentum.com>
13. Mozenda <https://www.mozenda.com>
14. Connotate <https://www.connotate.com>

Аналитические инструменты

1. Popsters <https://popsters.ru/>
2. AgoraPulse <https://www.agorapulse.com/>
3. Hootsuite <https://hootsuite.com/>
4. quintly <https://www.quintly.com/>
5. Simply Measured <https://sproutsocial.com/>
6. Socialbakers <https://www.socialbakers.com/>
7. Audiense <https://audiense.com/>
8. TweetReach <https://tweetreach.com/>
10. NetBase <https://www.netbase.com/>
11. Oktopost <https://www.oktopost.com/>
13. Tailwind <https://www.tailwindapp.com>
14. Keyhole <https://keyhole.co/>
15. Brand24 <https://brand24.com/>
16. Brandwatch <https://www.brandwatch.com/>
17. SumAll <https://sumall.com/>
18. Cyfe <https://www.cyfe.com/>
19. Klear <https://klear.com/>
20. Peakfeed <https://www.peakfeed.com/>
21. Sprout Social <https://sproutsocial.com/>
22. Followerwonk <https://followerwonk.com/>
23. TrendSpottr Signal <http://trendspottr.com/>
24. SimilarWeb <https://www.similarweb.com/>
25. BrandsEye <https://www.brandseye.com/>
26. Short Stack <https://www.shortstack.com/>
27. Talkwalker <https://www.talkwalker.com/>
28. Social Status <https://www.socialstatus.io/>

29. **Lucidya**<https://lucidya.com/>
30. **CX Social**<https://www.clarabridge.com/>
31. **Minter.io**<https://minter.io/>
32. **Foller.me**<https://foller.me/>
33. **nTuitive.social**<https://ntuitive.social/>
34. **SocialRank**<https://socialrank.com/>
35. **Brand Analytics** <https://br-analytics.ru>
36. **Alloka**<https://alloka.ru/>
37. **Repometr** <https://repometr.com/>
38. **AppMetrica** <https://appmetrica.yandex.ru>

8.2. Описание материально-технического обеспечения:

Для проведения образовательного процесса требуется аудитория с трансформируемым пространством, оборудованная компьютером и проектором, необходимыми для демонстрации презентаций. Компьютерный класс. Обязательное программное обеспечение – MSOffice.

9. Язык преподавания.

Английский.

10. Преподаватель (преподаватели).

Гончарова Инна Владимировна, доцент, кандидат пед. Наук

11. Автор (авторы) программы

Гончарова Инна Владимировна, доцент, кандидат пед. наук

Прончев Геннадий Борисович, доцент, кандидат физ-мат.наук

12. Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП указано в Общей характеристике ОПОП.