

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«Кубанский государственный университет»

Факультет компьютерных технологий и прикладной математики
Кафедра вычислительных технологий

УТВЕРЖДАЮ:

Проректор по учебной работе
качеству образования — первый
проректор

Хатуров Е.А.
подпись
« 27 » 04 2018



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ Б1.В.ДВ.10.02 "МЕТОДЫ ПОИСКА В WEB"

Направление
подготовки/специальность 02.03.02 **Фундаментальная информатика и**
информационные технологии
(код и наименование направления подготовки/специальности)

Направленность (профиль) /
специализация Вычислительные технологии
(наименование направленности (профиля) специализации)

Программа подготовки академический бакалавриат
(академическая /прикладная)

Форма обучения очная
(очная, очно-заочная, заочная)

Квалификация (степень) выпускника бакалавр
(бакалавр, магистр, специалист)

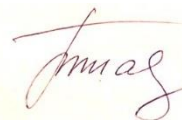
Краснодар 2018

Рабочая программа дисциплины Б1.В.ДВ.10.02 "МЕТОДЫ ПОИСКА В WEB" составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки

02.03.02 Фундаментальная информатика и информационные технологии Программу составил(а):

Приходько Татьяна Александровна, доцент, к. т. н.

Ф.И.О. , должность, ученая степень, ученое звание



подпись

Рабочая программа дисциплины Б1.В.ДВ.10.02 "МЕТОДЫ ПОИСКА В WEB" утверждена на заседании кафедры Вычислительных Технологий протокол № 7 «03» апреля 2018 г.

Заведующий кафедрой (разработчика)

Миков А. И.

фамилия, инициалы



подпись

Рабочая программа обсуждена на заседании кафедры Вычислительных Технологий протокол № 7 «03» апреля 2018 г.

Заведующий кафедрой (выпускающей)

Миков А. И.

фамилия, инициалы



подпись

Утверждена на заседании учебно-методической комиссии факультета Компьютерных Технологий и Прикладной Математики протокол № 1 от «20» апреля 2018 г

Председатель УМК факультета

Малыхин К. В.

фамилия, инициалы



подпись

Рецензенты:

Гаркуша О.В., доцент кафедры информационных технологий
ФБГОУ ВО «Кубанский государственный университет»,
кандидат физико-математических наук.

Зайков В.П. Ректор НЧОУ ВО «Кубанский институт информзащиты»
д.экон. наук, к.т.н., доцент.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1 Цель освоения дисциплины

Целью дисциплины «Методы поиска в WEB» является обучение передовым методам, моделям, средствам и технологиям поиска и компьютерной обработки информации.

Задачи дисциплины:

Дать **знания** о:

- истории и тенденциях развития информационно-поисковых систем, крупных ученых, участвовавших в их разработке,
- основных принципах обмена данными в глобальной сети Интернет;
- основных методах функционирования информационно-поисковых систем;
- основных современных инструментальных средствах их разработки;
- основных методах программирования поиска, как на стороне сервера, так и на стороне клиента.

Привить **умение**:

- использовать в проектируемых и эксплуатируемых информационных системах и технологиях современные средства Интернет-программирования;
- применять информационно-поисковые системы при нахождении в сети Интернет требуемой информации для проектно-конструкторской, организационно-управленческой и научно-исследовательской деятельности;
- разрабатывать информационно-поисковые системы для нахождения данных на стороне сервера или клиента.

1.2 Задачи дисциплины

В результате освоения дисциплины студент должен овладеть компетенциями ПК-1, в результате чего

знать: Методы эффективного поиска информации в сетевых источниках, приемы сбора, обработки и интерпретации данных современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.

уметь: Производить эффективный поиск необходимой информации, способствующей приобретению новых знаний и умений, в том числе, в новых областях знаний, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.

владеть: Навыками самостоятельного приобретения и использования в практической деятельности новых знаний и умений, в том числе, в новых областях знаний, навыками обработки и интерпретации данных современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.

1.3 Место дисциплины в структуре образовательной программы

Дисциплина «Методы поиска в WEB» является обязательной дисциплиной из вариативной части профессиональных дисциплин. Пререквизитами данной дисциплины являются дисциплины математического и естественнонаучного блоков ООП подготовки бакалавров: «Алгоритмы вычислительной математики», «Конструирование алгоритмов и структур данных», «Управление информацией», а также дисциплины "Распределенные системы и алгоритмы".

Знания, получаемые при изучении методов поиска в WEB, используются при изучении и других дисциплин профессионального цикла учебного плана бакалавра (параллельное и распределенное программирование, Введение в мультиагентные системы, дисциплины вариативной части), а также при работе над выпускной квалификационной работой.

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы.

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих **профессиональных компетенций**:

- ПК-1: способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.

Таблица 1. Профессиональные компетенции студента

Компетенция	знать	уметь	владеть
ПК-1	Методы эффективного поиска информации в сетевых источниках, приемы сбора, обработки и интерпретации данных современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.	Производить эффективный поиск необходимой информации, способствующей приобретению новых знаний и умений, в том числе, в новых областях знаний, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.	навыками самостоятельного приобретения и использования в практической деятельности новых знаний и умений, в том числе, в новых областях знаний, навыками обработки и интерпретации данных современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.

4. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 4 зач.ед. (144 часа), их распределение по видам работ представлено в таблице (для студентов ОФО)

Вид учебной работы	Всего часов	Семестры (часы)			
		8			
Контактная работа в том числе:					
Аудиторные занятия (всего):	54,3	54,3			
В том числе:					
Занятия лекционного типа	16	16			
Занятия семинарского типа (семинары, практ. занятия)					
Лабораторные занятия	32	32			
Иная контрольная работа	6	6			
Промежуточная аттестация (ИКР)	0,3	0,3			
Самостоятельная работа (всего)					
В том числе:					

Курсовая работа					
Проработка учебного (теоретического) материала	20	20			
Выполнение индивидуальных заданий (подготовка сообщений, презентаций)	5	5			
Реферат					
Подготовка к текущему контролю	20	20			
Контроль:					
Подготовка к экзамену:	44,7	44,7			
Общая трудоемкость	час	144	144		
	в т.ч. контактная работа	54,3	54,3		
	зач. ед.	4	4		

2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.
Разделы дисциплины, изучаемые в _8_ семестре (очная форма)

№	Наименование разделов	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	КСР	ЛР	
1	2	3	4	5	6	7
1.	Раздел 1. Аналитика в сети Интернет	39	8		16	15
2.	Раздел 2. Методологии сбора данных из сетевых источников	27	4	6	8	15
3.	Раздел 3. Типы информационных систем. Устройство и принцип работы поисковых систем.	27	4		8	15
4.	Подготовка к экзамену	45				
5.	<i>Итого по дисциплине:</i>	144	16	6	32	45

Примечание: Л – лекции, КСР – контрольные и самостоятельные работы, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов дисциплины:

2.3.1 Занятия лекционного типа

№ раздела	Наименование раздела	Содержание раздела	Форма текущего контроля
1	2	3	4
1	Раздел 1. Аналитика в сети Интернет	Тема 1. Генезис сети Интернет. Тема 2. Структура WEB, Deep WEB. Тема 3. Системы управления контентом.	ЛР

2	Раздел 2. Методология сбора и анализа данных из сетевых источников	Тема 4. Технологии извлечения знаний из WEB - WEB-mining. Тема 5. Понятие <i>data scraping</i> или «срезание данных с поверхности». Классификация способов извлечения информации из WEB-источников. Тема 6. Модели информационного поиска.	ЛР
3	Раздел 3. Типы информационных систем. Устройство и принцип работы поисковых систем.	Тема 7. Типология, структура и функция информационных систем. Системы переработки информации. Типы информационных систем. Уточнение структуры информационных систем. Информационные системы Интернета. Тема 8. Устройство и принцип работы поисковых систем. Автоматическое индексирование. Семантический вэб. Искусственный интеллект. Разработка ИПТ. Отраслевой тезаурус. Тема 9. Способы хранения больших данных в WEB	ЛР

ЛР – лабораторные работы

Расшифровка тем дисциплины

- Генезис сети Интернет.** История создания Сети. Развитие электрических и электронных средств связи. ARPANET. Всемирная паутина. Развитие интернет в XXI веке. Организационная структура Интернета. Схема адресации в сети Интернет. Модель BOW TIE. Понятия и различия WEB 2.0- WEB 4.0.
- Структура WEB, Deep WEB.** Невидимый WEB, его возможности и характеристики. Инструменты и технологии работы в невидимом WEB.
- Системы управления контентом.** Проблемы, возникающие при поддержании актуальности информации на сайте. Определение CMS. Краткое описание CMS. Динамический и статический сайты. Характеристика контента. Создание контента. Управление автоматизированными деловыми процессами. Распространение контента. Персонализация и глобализация контента. Критерии классификации систем управления контентом. Простая CMS. Шаблонная CMS. Профессиональная CMS. Универсальная CMS. Функциональные и технологические возможности систем управления контентом. Требования к системам управления контентом. Вопросы, решаемые при выборе системы управления контентом.
- Технологии извлечения знаний из WEB - WEB-mining.** Определение понятий WEB Mining и Data Mining? Отличия между ними. Задачи и этапы извлечения знаний из WEB. Направления WEB-mining: Извлечение Web-контента (Web Content Mining); Извлечение Web-структур (Web Structure Mining); Исследование использования Web-ресурсов (Web Usage Mining)
- Понятие *data scraping* или «срезание данных с поверхности».** Понятие бизнес-аналитического решения. Анализ журнала посещаемости сайта. Заказные статистические исследования. Определение профиля сайта. Определение перечня сайтов, посещаемых вашей аудиторией. Определение целевой аудитории сайта. Типы посетителей сайтов. Модели поведения посетителей сайта. Пользователи Интернет магазинов.
- Модели информационного поиска.** Булева модель, векторная модель, вероятностная модель, гибридная модель. Математические особенности обработки информации разными моделями. Сферы их применения.

7. **Типология, структура и функция информационных систем.** Системы переработки информации. Типы информационных систем. Уточнение структуры информационных систем. Информационные системы Интернета.
8. **Устройство и принцип работы поисковых систем.** Понятие поисковой системы. Принципы работы поисковых систем, которые нужно учитывать при продвижении сайта. Виды поисковых роботов. Порядок индексации сайтов. Порядок поисковой выдачи. Принципы алгоритмов выдачи поисковой системы Яндекс и Google. Выбор ключевых слов для продвижения сайта. Типы запросов по частотности. Типы запросов по степени конверсии. Понятие семантического ядра. Создание семантического ядра. Выбор ключевых страниц сайта. Распределение семантического ядра. Анализ сайтов конкурентов. Расчет сложности продвижения сайта. Выбор основной стратегии поискового продвижения сайта.
9. **Способы хранения данных в WEB.** Требования к хранилищам данных, OLTP и OLAP системы. Нереляционные базы данных.

2.3.2. Занятия семинарского типа

Занятия семинарского типа – не предусмотрены.

2.3.3. Лабораторные занятия

№ работы	№ раздела дисциплины	Наименование лабораторных работ	Форма текущего контроля
1	1	Вводное занятие. Настройка необходимого ПО и среды разработки.	
2	1	Классификация интернет-сервисов. Составление запросов по теме дипломной работы, выполнение поиска в открытых и закрытых сетевых источниках, сравнение эффективности поиска с помощью различных инструментов.	Отчет по лабораторной работе
3	1	Системы управления контентом. Обсуждение преимуществ и недостатков различных CMS, особенностей разработки WEB-ресурсов с их помощью.	Отчет по лабораторной работе
4	2	Технологии извлечения знаний из WEB – WEB Mining. Понятие <i>data scraping</i> или «срезание данных с поверхности». Используя любой из приведенных либо найденных вами способов извлечения информации с web страниц, разработать программу по сбору информации методами Web-scrapingа и продемонстрировать результат ее работы.	Отчет по лабораторной работе Контрольная
5	3	Модели информационного поиска. Устройство и принцип работы поисковых систем. Определение и анализ характеристик выбранной поисковой системы: Google, Yandex, Rambler, Yahoo, Bing, AltaVista.	Отчет по лабораторной работе
6	3	Устройство хранения данных в WEB.	Устный опрос

2.3.3 Примерная тематика курсовых работ (проектов)

Учебным планом не предусмотрены.

2.3.4 Расчетно-графические задания

Учебным планом не предусмотрены.

2.4. Самостоятельное изучение разделов дисциплины

Раздел 1. Историческое развитие Интернет и сервисов.

Раздел 2. Виды, структуры и функции систем управления контентом.

Раздел 3. Устройство и принцип работы поисковых систем. Инструментарий для сбора интернет-статистики.

2.5 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Раздел 1. Историческое развитие Интернет и сервисов.	Источники основной и дополнительной литературы
2	Раздел 2. Виды, структуры и функции систем управления контентом.	Источники основной и дополнительной литературы
3	Раздел 3. Устройство и принцип работы поисковых систем. Инструментарий для сбора интернет-статистики.	Приходько Т.А. Лекции (презентации) по дисциплине «Методы поиска в WEB».

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Семестр	Вид занятия (Л, ПР, ЛР)	Используемые интерактивные образовательные технологии	Количество часов
---------	-------------------------	---	------------------

8	Л	Компьютерные презентации и обсуждение	16
	ЛР	Разбор конкретных ситуаций (задач), тренинги по решению задач, компьютерные симуляции (программирование алгоритмов), коллективные проекты	32
Итого:			48

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

4. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

4.1 Фонд оценочных средств для проведения текущего контроля

Фонд оценочных средств дисциплины состоит из средств текущего контроля выполнения лабораторных работ, коллективных проектов, средств для итоговой аттестации (экзамена).

Оценка успеваемости осуществляется по результатам:

- выполнения лабораторных работ, коллективных проектов;
- ответа на экзамене (для выявления знания и понимания теоретического материала дисциплины).

Примеры экзаменационных билетов:

Экзаменационный билет содержит 2 вопроса и задачу.

Экзаменационный билет № ____

1. Опишите структуру, пропорции, охарактеризуйте размеры и динамику WEB.
2. Классификация способов извлечения информации из WEB-источников.
3. Индивидуальное задание.

Перечень вопросов, которые выносятся на экзамен в семестре 8

1. Опишите структуру, пропорции, охарактеризуйте размеры и динамику WEB.
2. Понятие «Сильной связности» WEB-графа, типы его узлов. Какому функциональному закону подчиняются сети «тесного мира»?
3. Закономерности и ограничения модели Bow Tie.
4. Понятие WEB 2.0.
5. Deep WEB. Какие ресурсы его составляют. Какими средствами его можно исследовать.
6. Понятия Web Mining и Web Analytics. Этапы аналитики в соответствии со стандартом CRISP-DM.
7. Задачи Data Mining. Направления Data Mining.
8. Понятие и задачи Web Content Mining.
9. Перечислите и охарактеризуйте средства WEB scraping.
10. Методы Text Mining в приложении к специфике WWW.

11. Методологии Web Graph Mining для подхода Web Structure Mining.
12. Основные задачи Web Usage Mining, средства их решения, назначение кластерного анализа в контексте Web Usage Mining.
13. Классификация способов извлечения информации из WEB-источников.
14. Задачи Web-scraping, механизм его работы. Разновидность методов Web-scraping.
15. Этапы работы поисковой системы. Компоненты поискового движка.
16. Как работают алгоритмы индексирования. Необходимость ранжирования и задачи машинного обучения в приложении к информационному поиску.
17. Охарактеризуйте модели информационного поиска.
18. Изложите подробно принцип булевой модели информационного поиска (ИП), возможные средства оптимизации запроса.
19. Суть векторной и вероятностной моделей ИП, их достоинства и недостатки.
20. Назовите и кратко охарактеризуйте этапы нормализации текста перед индексацией.
21. Перечислите и дайте краткую характеристику методов лингвистического анализа.
22. Способы хранения словарей. Способы нечеткого поиска.

Критерии оценивания

Оценка «отлично»:

- 1) по теоретическим вопросам даны точные формулировки алгоритмов, теорем и правильные доказательства; точные определения математических объектов и ясные и правильные определения объектов, характеризующихся неформализованными понятиями;
- 2) по практической части приведены достоверные результаты исследования и даны подробные пояснения. Практическая часть курса выполнена полностью.

Оценка «хорошо»:

- 1) по теоретическим вопросам – при ответе на один вопрос даны точные формулировки алгоритмов; даны ясные и правильные определения объектов, характеризующихся неформализованными понятиями; при ответе на второй вопрос имеются неточности формулировки алгоритмов или пробелы в аргументации ответа; недостаточно точные определения или неясные и не совсем правильные определения объектов, характеризующихся неформализованными понятиями;
- 2) по практической части приведены достоверные результаты исследования и даны подробные пояснения. Практическая часть курса выполнена не менее чем на 80%.

Оценка «удовлетворительно»:

- 1) по теоретическим вопросам – при ответе на оба вопроса имеются неточности формулировки алгоритмов; недостаточно точные определения объектов или неясные и не совсем правильные определения объектов, характеризующихся неформализованными понятиями;
- 2) По практической части приведены достоверные результаты исследования, практическая часть курса выполнена не менее чем на 60%.

Оценка «неудовлетворительно»:

отсутствуют удовлетворительные ответы на два или более вопроса экзаменационного билета. Практическая часть курса выполнена менее чем на 60%.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья

предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

5.1. Основная литература:

1. Сеница С.Г. Веб-программирование и веб-сервисы – учебное пособие, КубГУ, 2013. (28 экз. в библиотеке КубГУ).
2. Щербаков, А. Интернет-аналитика: поиск и оценка информации в вебресурсах : практическое пособие / А. Щербаков. - Москва : Книжный мир, 2012. - 78 с. [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=89693>
3. Жуковский, О.И. Информационные технологии и анализ данных : учебное пособие / О.И. Жуковский ; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). - Томск : Эль Контент, 2014. - 130 с. : схем., ил - Библиогр.: с. 126. [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=480500>

5.2. Дополнительная литература:

1. Артемов, А.В. Мониторинг информации в интернете : учебнометодическое пособие / А.В. Артемов ; Межрегиональная Академия безопасности и выживания.- Орел : МАБИВ, 2014. - 160 с. [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=428606>
2. Бычков, И.В. Инфраструктура информационных ресурсов и технологии создания информационно-аналитических систем территориального управления / И.В. Бычков. - Новосибирск : Издательство Сибирского отделения Российской Академии Наук, 2016. - 240 с. - ISBN 978-5-7692-1478-3 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=467655>
3. Маннинг, К. Д. Введение в информационный поиск / К. Д. Маннинг, П. Рагхаван, Х. Шютце. – Пер. с англ. – М. : ООО “И.Д. Вильямс”, 2011. – 528 с.

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечных системах

- ЭБС Издательства «Лань» <http://e.lanbook.com> ,
 2. ЭБС «Университетская библиотека онлайн» www.biblioclub.ru ,
 3. ЭБС «Юрайт» <http://www.biblio-online.ru> ,
 4. ЭБС «ZnaniUM.COM» www.znanium.com,
 5. ЭБС «BOOK.ru» <https://www.book.ru>.

5.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. **Web Page Scraping using Java**, ресурс: *URL*: <http://half-wit4u.blogspot.ru/2011/01/web-scraping-using-java-api.html>
2. **Web Page Scraping with Java**, ресурс: *URL*: <https://www.packtpub.com/books/content/creating-sample-web-scrapers>
3. **Jaunt Java Web Scraping & JSON Querying** , ресурс: jaunt-api.com
4. **Web scraping с Node.js**, ресурс: *URL*: <http://www.webdesignmagazine.ru/internet-technology/other/web-scraping-s-node-js/>
5. **Web-scraping средствами R, примеры**. Ресурс: *URL*: kek.ksu.ru/EOS/WM/WebScraping.docx
6. Chang, C.-H., Kayed, M., Girgis, M. R., and Shaalan, K. F. 2006. A survey of web information extraction systems. *IEEE Trans. on Knowl. and Data Eng.* 18,10, 1411-1428.

6. Методические указания для обучающихся по освоению дисциплины

По курсу предусмотрено проведение лекционных занятий, на которых дается основной систематизированный материал, лабораторных работ, контрольной работы, зачета и экзамена.

Важнейшим этапом курса является самостоятельная работа по дисциплине с использованием указанных литературных источников и методических указаний автора курса.

Виды и формы СР, сроки выполнения, формы контроля приведены выше в данном документе.

Для лучшего освоения дисциплины при защите ЛР студент должен ответить на несколько вопросов из лекционной части курса.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю)

7.1 Перечень информационных технологий.

- Проверка домашних заданий и консультирование посредством электронной почты.
- Использование электронных презентаций при проведении лекций и практических занятий.

7.2 Перечень необходимого программного обеспечения

1. OS Windows, MS Office
2. Антивирусное ПО
3. NetBeans или Eclipse IDE, JDK.
4. Программы для демонстрации и создания презентаций («Microsoft Power Point»).

7.3 Перечень информационных справочных систем:

1. Электронная библиотечная система eLIBRARY.RU
(<http://www.elibrary.ru/>)

8. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

№	Вид работ	Материально-техническое обеспечение дисциплины (модуля) и оснащенность
1.	Лекционные занятия	Лекционная аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) и соответствующим программным обеспечением (ПО) PowerPoint. ауд. 129, 131, А305.
2.	Лабораторные занятия	Лаборатория, укомплектованная специализированными техническими средствами обучения – компьютерный класс, с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета. (лаб. 102-106.).
3.	Групповые (индивидуальные) консультации	Аудитория, (кабинет) – компьютерный класс
4.	Текущий контроль, промежуточная аттестация	Аудитория, приспособленная для письменного ответа при промежуточной аттестации.