

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет математики и компьютерных наук

УТВЕРЖДАЮ:



Проректор по учебной работе,
качеству образования – первый
проректор

Хагуров Т.А.

мая 2019 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
Б1.В.ДВ.01.02 МОДЕЛИ ЕСТЕСТВЕННЫХ ЯЗЫКОВ**

Направление подготовки	02.04.01 Математика и компьютерные науки
Направленность (профиль)	Вычислительная математика
Форма обучения	очная
Квалификация (степень) выпускника	магистр

Краснодар 2019

Рабочая программа дисциплины Б1.В.ДВ.01.02 Модели естественных языков составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 02.04.01 Математика и компьютерные науки

Программу составил:

Ю.М. Вишняков профессор, доктор техн.наук, профессор


подпись

Рабочая программа дисциплины утверждена на заседании кафедры вычислительной математики и информатики протокол № 13 «18» апреля 2019 г.

Заведующий кафедрой (разработчик) Гайденок С.В.


подпись

Рабочая программа дисциплины утверждена на заседании кафедры вычислительной математики и информатики протокол № 13 «18» апреля 2019 г.

Заведующий кафедрой (выпускающей) Гайденок С.В.


подпись

Утверждена на заседании учебно-методической комиссии факультета математики и компьютерных наук протокол № 2 «24» апреля 2019 г.

Председатель УМК факультета Титов Г.Н.


подпись

Рецензенты:

Профессор кафедры прикладной математики Кубанского государственного университета кандидат физико-математических наук доцент Кармазин В.Н.

Доктор экономических наук, кандидат технических наук, профессор кафедры компьютерных технологий и систем КубГАУ Луценко Е.В.

1 Цели и задачи изучения дисциплины (модуля).

1.1 Цель освоения дисциплины.

Указываются цели освоения дисциплины (или модуля), соотнесенные с общими целями ООП ВО.

Основная цель дисциплины Б1.В.ДВ.01.02 «Модели естественных языков» - дать студентам базовые знания по основным подходам к построению моделей и автоматической обработке естественных языков, а также приложениям в обработке текстовой информации на естественном языке, научить решать комплексные задачи в области проектирования систем обработки нечисловой информации.

1.2 Задачи дисциплины.

В результате освоения дисциплины должны быть решены следующие основные задачи. Студент должен:

- знать базовые сведения по основным подходам к построению моделей и автоматической обработке естественных языков, а также приложениям в обработке текстовой информации на естественном языке, приобрести навыки решения комплексных задач в области проектирования систем обработки нечисловой информации.
- уметь применять знания по основным подходам к построению моделей и автоматической обработке естественных языков, а также приложениям в обработке текстовой информации на естественном языке в области проектирования систем обработки нечисловой информации.
- владеть восприятием, анализом и обобщением информации в профессиональной области и выбором путей решения профессиональных задач на основе знаний и умений дисциплины «Модели естественных языков».

1.3 Место дисциплины (модуля) в структуре образовательной программы.

Дисциплина «Методы и средства автоматической обработки информации» относится к вариативной части Блока 1 "Дисциплины (модули)" учебного плана.

Дисциплины, являющиеся пререквизитом для дисциплины «Модели естественных языков»: «Основы компьютерных технологий», «Технологии программирования», «Компьютерная графика». Дисциплины, для которых курс «Модели естественных языков» является пререквизитом: «Теоретические основы и технологии информационного поиска».

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 общекультурных/ профессиональных компетенций *ПК-1 и ПК-2*.

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
1.	ПК-1	Способен демонстрировать фундаментальные знания математических и	базовые сведения фундаментальных знаний по моделям естественных языков	применять фундаментальные знания по моделям естественных языков в области проектирования систем обработки	владеть восприятием, анализом и обобщением информации в профессиональной области и

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
		естественных наук, программирования и информационных технологий	приложениям, приобрести навыки решения комплексных задач в области проектирования систем обработки нечисловой информации	нечисловой информации и в своей профессиональной деятельности	выбором путей решения профессиональных задач на основе фундаментальных знаний и умений дисциплины «Модели естественных языков»
2.	ПК-2	Способность проводить научные исследования, на основе существующих методов в конкретной области профессиональной деятельности	Способы и методы проведения научных исследований на основе существующих методов в конкретной области профессиональной деятельности	проводить научные исследования, на основе существующих методов в конкретной области профессиональной деятельности на основе знаний дисциплины	Способами и методами проведения научных исследований на основе существующих методов в конкретной области профессиональной деятельности

2. Структура и содержание дисциплины.

2.1 Распределение трудоёмкости дисциплины по видам работ.

Общая трудоёмкость дисциплины составляет 2 зач.ед. (72 часов), их распределение по видам работ представлено в таблице
(для студентов ОФО).

Вид учебной работы	Всего часов	Семестры (часы)
		А
Контактная работа, в том числе:		
Аудиторные занятия (всего):	30	30
Занятия лекционного типа	-	-
Лабораторные занятия	30	30
Занятия семинарского типа (семинары, практические занятия)	-	-
	-	-
Иная контактная работа:		
Контроль самостоятельной работы (КСР)	-	-

Вид учебной работы		Всего часов	Семестры (часы)
			А
Промежуточная аттестация (ИКР)		0,2	0,2
Самостоятельная работа, в том числе:			
Подготовка к текущему контролю		41,8	41,8
Контроль:			
Подготовка к экзамену		-	-
Общая трудоемкость	час.	72	72
	в том числе контактная работа	30,2	30,2
	зач. ед	2	2

2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы дисциплины, изучаемые в А семестре (*очная форма*)

№ разд ела	Наименование разделов	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Введение в предметную область естественных языков.	9	2	2		5
2.	Модели естественных языков: частотная, вероятностная, грамматическая.	9	2	2		5
3.	Морфологический разбор предложения. Морфологический разбор слова.	9	2	2		5
4.	Инструменты для работы с текстами язык Python. Библиотеки по обработке текстовых данных.	9	2	2		5
5.	Частотный анализатор текстовой информации. Критерии анализа.	9	2	2		5
6.	Подсчет уникальности текста на основе морфологического разбора слов.	9	2	2		5
7.	Парсеры. Морфологический парсер, синтаксический парсер. Яндекс парсер.	12,8	4	2		6,8
8.	Обработка неструктурированных текстов шаблонным методом.	18	2	2		5
	<i>Итого по дисциплине:</i>	71,8	14	16		41,8

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов дисциплины:

2.3.1 Занятия лекционного типа.

	Наименование раздела	Содержание раздела	Форма текущего контроля
1	2	3	4
1	Введение в предметную область естественных языков. Модели: частотная, вероятностная, грамматическая.	Введение в предметную область естественных языков. Модели: частотная, вероятностная, грамматическая.	Устный опрос Проверка СР
2	Синтаксические конструкции естественного языка.	Синтаксические конструкции естественного языка. Сентенциальная форма. Формализация.	Устный опрос Проверка СР
3	Морфология	Морфологический разбор предложения. Морфологический разбор слова.	Устный опрос Проверка СР
4	Инструменты для работы с текстами язык Python.	Библиотеки по обработке текстовых данных.	Устный опрос Проверка СР
5	Текстовый анализатор	Частотный анализатор текстовой информации. Критерии анализа.	Устный опрос Проверка СР
6	Подсчет уникальности текста на основе морфологического разбора слов.	Подсчет уникальности текста на основе морфологического разбора слов..	Устный опрос Проверка СР
7	Парсеры.	Парсеры. Морфологический парсер, синтаксический парсер. Яндекс парсер.	Устный опрос Проверка СР
8	Обработка неструктурированных текстов шаблонным методом.	Обработка неструктурированных текстов шаблонным методом.	Устный опрос Проверка СР

2.3.2 Занятия семинарского типа.

№	Наименование раздела	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1.	Раздел 1-3	Определение основных требований для построения алгоритма обработки текстов.	Текущий опрос

2.	Радел 4-5	Разработка требований к универсальному протоколу алгоритму обработки текстов на естественном языке.	Текущий опрос
3.	Радел 6	Построение шаблонного парсера для разбора структурированного текста.	Текущий опрос
4.	Раздел 7-8	Анализ существующих парсеров. Яндекс Парсер.	Текущий опрос

2.3.3 Лабораторные занятия.

Лабораторные занятия - не предусмотрены

2.3.4 Примерная тематика курсовых работ (проектов)

Курсовые работы – не предусмотрены учебным планом дисциплины.

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Работа с материалами дисциплины, поиск и анализ литературы и электронных источников информации по заданной проблеме	1. Методические рекомендации по организации самостоятельной работы студентов утвержденные кафедрой вычислительной математики и информатики, протокол № 14 от 14.06.2017 г.
2	Знакомство с концепцией ООП	1. Акулич И.К., Математическое программирование в примерах и задачах, 3-е издание, издательство «Лань», 2011, 352 стр., http://e.lanbook.com , электронные ресурсы библиотеки КубГУ. 2. Федоров Д. Ю. Программирование на языке высокого уровня python : учебное пособие для прикладного бакалавриата / Д. Ю. Федоров. — М. : Издательство Юрайт, 2018. — 126 с. — https://biblio-online.ru/book/1EE056CF-F11A-4C18-8D33-40B703D49AC5
3	Подготовка к зачету	1. Методические рекомендации по организации самостоятельной работы студентов утвержденные кафедрой вычислительной математики и информатики, протокол № 14 от 14.06.2017 г.

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.

Сем естр	Вид занятия	Используемые интерактивные образовательные технологии	Количес тво часов
6	Лабораторные занятия	Кейс-метод «Понятие грамматики. Синтаксические конструкции естественного языка. Средства и фазы обработки естественного языка.»	2
		Метод проектов «Очистка текста, токенизация, парсинг»	2
		Метод проектов «Морфология. Морфологический разбор предложения. Морфологический разбор слова.»	2
		Метод проектов «Библиотеки по обработке текстовых данных языка Python»	2
		Метод проектов «Частотный анализатор текстовой информации. Критерии анализа.»	2
		Метод проектов «Парсинг. Яндекс парсер.»	2
		Метод проектов «Обработка неструктурированных текстов шаблонным методом.»	4
		Итого:	16

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций со студентом при помощи электронной информационно-образовательной среды ВУЗа.

4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации.

4.1 Фонд оценочных средств для проведения промежуточной аттестации.

Примеры теоретических вопросов:

1. Понятие естественного языка, виды моделей.
2. Что такое автоматическая обработка текстов естественных языков?
3. Виды и структура текстовых документов.
4. Лексический анализ документов. Выделение стоп-слов в документе.
5. Морфологический анализ слов. Модели словообразования языка. Разбор флексивных классов.
5. Морфологический анализ слов. Лемматизация и стеминг.
6. Статистические характеристики и особенности текстов.
7. Структура и организация словарей. Методы и алгоритмы построения.

8. Предсинтаксический и синтагматический анализ. Основные объекты, результаты и особенности.
9. Методы и способы установления синтагматических связей.
10. Применение морфологического и синтагматического анализа в задачах информационного поиска, классификации и кластеризации документов.
11. Особенности обработки текстов с использованием морфологических модулей (АОТ, stemka-yandex, lx-platform и др.).
12. Синтаксический анализ предложения. Подходы к решению задачи синтаксического анализа. Проблемы омонимии и синонимии при решении задачи синтаксического анализа.
13. Особенности обработки текстов с использованием синтаксических модулей (АОТ, lx-platform и др.).
14. Семантический анализ текстов. Общие модели представления семантической информации. Алгоритмы и правила семантического анализа предложений. Анализ и сравнение результатов работы систем семантического анализа текстовых документов.
15. Особенности обработки текстов с использованием семантических модулей (АОТ, lx-platform, ЭТАП и др.).

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 (модулю) 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 (модуля).

5.1 Основная литература:

1. Трофимов, В. В. Алгоритмизация и программирование [Электронный ресурс]: учебник для академического бакалавриата / В. В. Трофимов, Т. А. Павловская ; под ред. В. В. Трофимова. - Москва: Юрайт, 2018. - 137 с. – <https://biblio-online.ru/book/B08DB966-3F96-4B5A-B030-E3CD9085CED4>.

2. Рудинский, И.Д. Технология проектирования автоматизированных систем обработки информации и управления [Электронный ресурс] / И.Д. Рудинский. — Электрон. дан. — Москва : Горячая линия-Телеком, 2011. — 304 с. — Режим доступа: <https://e.lanbook.com/book/5191>.

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечных системах «Лань» и «Юрайт».

5.2 Дополнительная литература:

1. Математическая обработка информации [Электронный ресурс] : учебник и практикум для академического бакалавриата / М. Ю. Глотова, Е. А. Самохвалова. - 2-е изд., испр. и доп. - Москва : Юрайт, 2018. - 347 с. - <https://biblio-online.ru/book/915C18E7-1D7F-405B-A1B5-4717E978EDC9>.

6. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля).

1. Электронный каталог Научной библиотеки КубГУ
<http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотечная система "Университетская библиотека ONLINE"
<http://biblioclub.ru/>
3. Электронная библиотечная система издательства "Лань"
<https://e.lanbook.com/>
4. Электронная библиотечная система «Юрайт» <http://www.biblio-online.ru>
5. Электронная библиотечная система «ZNANIUM. COM» www.znanium.com
6. Электронная библиотечная система «BOOK.ru» <https://www.book.ru>

7. Методические указания для обучающихся по освоению дисциплины (модуля).

Обязательными для самостоятельной работы студентов являются:

- разбор и самостоятельное изучение теоретического материала по конспектам лекций и имеющейся литературе;
- подготовка и настройка собственной компьютерной техники к работе;
- подготовка к лабораторным занятиям.

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.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.

8.1 Перечень информационных технологий.

При изучении студентами дисциплины используются следующие технологии:

- технологии проблемного обучения (проблемные лекции, проводимые в форме диалога, решение учебно-профессиональных задач на практических занятиях;
- игровые технологии («интеллектуальные разминки», «мозговые штурмы»);

- интерактивные технологии (проведение лекций диалогов, коллективное обсуждение различных подходов к решению той или иной профессиональной задачи);
- информационно-коммуникативные образовательные технологии (моделирование изучаемых явлений, презентация учебных материалов) и элементы технологий проектного обучения.

Для выполнения моделирования необходим пакет прикладных программ для работы с текстами и презентациями, а также программы из п 8.2.

Лекционные занятия по ряду тем проводятся преподавателем как проблемные в форме диалога. На практических занятиях используются и «интеллектуальные разминки», элементы дискуссий, коллективное обсуждение решений задач и моделей изучаемых явлений, подготовленных студентами к занятию и т.д.

8.2 Перечень необходимого программного обеспечения.

1. NetBeans (доступен по открытой лицензии);
2. PyCharm (доступен по открытой лицензии);
3. Free Pascal (доступен по открытой лицензии);
4. Lazarus.

8.3 Перечень информационных справочных систем:

1. Электронная справочная система для разработчиков – MSDN (<https://msdn.microsoft.com/ru-ru>).
2. Образовательный портал для освоения базовых навыков программирования GeekBrains (www.geekbrains.ru).
1. Справочно-правовая система «Консультант Плюс» (<http://www.consultant.ru>)
2. Электронная библиотечная система eLIBRARY.RU (<http://www.elibrary.ru/>)

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю).

	Вид работ	Материально-техническое обеспечение дисциплины (модуля) и оснащенность
1.	Лекционные занятия	Лекционная аудитория, специально оборудованная мультимедийными демонстрационными комплексами, учебной мебелью
2.	Семинарские занятия	Специальное помещение, оснащенное учебной мебелью, презентационной техникой (проектор, экран, ноутбук) и соответствующим программным обеспечением (ПО).
3.	Групповые (индивидуальные) консультации	Помещение для проведения групповых (индивидуальных) консультаций, учебной мебелью, оснащенное презентационной техникой (проектор, экран, ноутбук) и соответствующим программным обеспечением
4.	Текущий контроль, промежуточная аттестация	Помещение для проведения текущей и промежуточной аттестации, оснащенное учебной мебелью, персональными компьютерами с доступом к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду организации
5.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-

		образовательную среду университета
--	--	------------------------------------

РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.ДВ.01.01 Методы и средства автоматической обработки информации по направлению подготовки 02.04.01 – Математика и компьютерные науки Профиль: Вычислительная математика – магистр,
подготовленную кафедрой вычислительной математики и информатики КубГУ

Рабочая программа по дисциплине «Методы и средства автоматической обработки информации» разработана в соответствии с установленным образовательным стандартом и охватывает все базовые вопросы дисциплины.

Рабочая программа содержит следующие разделы: цели и задачи освоения дисциплины, место дисциплины в структуре ООП ВО, требования к результатам освоения дисциплины, структура и содержание дисциплины, распределение видов учебной работы и их трудоемкости по разделам дисциплины, содержание разделов дисциплины, содержание самостоятельной работы студентов, образовательные технологии, оценочные средства для контроля успеваемости, учебно-методическое и информационное обеспечение дисциплины.

Содержание рабочей программы соответствует уровню подготовленности студентов к изучению данной дисциплины. Успешность изучения обеспечивается предшествующей подготовкой студентов по ряду математических дисциплин профессионального цикла.

Для осмысления разделов и тем предусмотрено выполнение лабораторных работ, что позволяет не только закрепить теоретические знания, но и обеспечить возможность проведения промежуточного контроля знаний по теоретической и практической части дисциплины.

Рабочая программа содержит список рекомендуемой основной и дополнительной литературы, который способствует более глубокому и целенаправленному изучению дисциплины.

В целом, представленная рабочая программа может быть рекомендована при изучении вышеуказанной дисциплины.

Доктор экон. наук, канд. техн. наук, профессор кафедры компьютерных технологий и систем КубГАУ



Луценко Е.В.

РЕЦЕНЗИЯ

на рабочую программу дисциплины

Методы и средства автоматической обработки информации
по направлению подготовки 02.04.01 – Математика и компьютерные науки,
профиль Вычислительная математика (магистр), подготовленную кафедрой
вычислительной математики и информатики КубГУ

Рабочая программа по дисциплине «Методы и средства автоматической обработки информации» разработана в соответствии с установленным образовательным стандартом и охватывает все базовые вопросы дисциплины.

Рабочая программа содержит следующие разделы: цели и задачи освоения дисциплины, место дисциплины в структуре ООП ВО, требования к результатам освоения дисциплины, структура и содержание дисциплины, распределение видов учебной работы и их трудоемкости по разделам дисциплины, содержание разделов дисциплины, содержание самостоятельной работы студентов, образовательные технологии, оценочные средства для контроля успеваемости, учебно-методическое и информационное обеспечение дисциплины.

Содержание рабочей программы соответствует уровню подготовленности студентов к изучению данной дисциплины. Успешность изучения обеспечивается предшествующей подготовкой студентов по ряду математических дисциплин профессионального цикла.

Для осмысления разделов и тем предусмотрено выполнение лабораторных работ, что позволяет не только закрепить теоретические знания, но и обеспечить возможность проведения промежуточного контроля знаний по теоретической и практической части дисциплины.

Рабочая программа содержит список рекомендуемой основной и дополнительной литературы, который способствует более глубокому и целенаправленному изучению дисциплины.

В целом, представленная рабочая программа может быть рекомендована при изучении вышеуказанной дисциплины.

Профессор кафедры прикладной математики
Кубанского государственного университета
кандидат физико-математических наук доцент



Кармазин В.Н.