

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.В.ДВ.01.02 Математическая и компьютерная
лингвистика

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

01.04.02 Прикладная математика и информатика

Направленность (профиль)

01.04.02.06 Прикладная математика и информатика в гуманитарных и
социально-экономических науках

Форма обучения

очная

Год набора

2021

Красноярск 2022

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили _____

к.ф.-м.н, доцент, Семенова Дарья Владиславовна; к.ф.-м.н., Доцент,

Куликов Владимир Русланович

должность, инициалы, фамилия

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Целью преподавания дисциплины является формирование у студентов твердых теоретических знаний и практических навыков применения методов прикладной математики и информатики, связанных с современными концепциями лингвистического моделирования и разработкой на основе лингвистических моделей технологий и программных средств обработки текстов на естественном языке.

1.2 Задачи изучения дисциплины

Задачами дисциплины являются освоение методов и алгоритмов компьютерной лингвистики применительно к задачам обработки текстовой информации в информационных системах, в том числе

- изучение основных теоретических положений, методов и моделей математической лингвистики;
- исследование особенностей семиотического и математического моделирования естественного языка и речи с целью перевода информации, содержащейся в неформализованном виде в тексте, на формализованный искусственный язык;
- изучение методики и методов структурно-вероятностного моделирования языка и речи, основных математико-статистических методов обработки лингвистической информации;
- разработка на основе лингвистических моделей технологий и программных средств для обработки текстов на естественном языке.

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

Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине
ПК-1: Способен разрабатывать и исследовать математические модели, методы и алгоритмы по тематике проводимых исследований	
ПК-1.1: Обладает достаточными фундаментальными теоретическими и практическими знаниями математических и естественных наук, основ программирования и информационных технологий для проведения в конкретной области профессиональной деятельности	подходы к проведению научных исследований в области прикладной математики и информатики; новые научные результаты и предысторию их появления; методики и методы структурно-вероятностного моделирования языка и речи, основные математико-статистические методы обработки лингвистической информации. применять классические и современные методы прикладной математики и информатики для решения задач компьютерной лингвистики; систематизировать научные результаты, выделять из них главное, и удалять второстепенное; самостоятельно выбирать эффективные методы

	<p>решения поставленных задач и разрабатывать новые методы для получения новых научных и прикладных результатов.</p> <p>методами математического моделирования лингвистических систем;</p> <p>навыками использования современных информационных технологий, конкретных программных продуктов и информационных ресурсов при моделировании лингвистических систем;</p> <p>навыками проведения научных исследований самостоятельно и в составе научного коллектива.</p>
<p>ПК-1.2: Решает научные задачи в соответствии с поставленной целью и в соответствии с выбранной методикой</p>	<p>основные понятия компьютерной лингвистики;</p> <p>основные методы анализа и синтеза получаемой информации по тематике курса;</p> <p>современные тенденции и направления в научных исследованиях, проводимых в мире, в математической и компьютерной лингвистике</p> <p>использовать возможности отечественных и зарубежных универсальных программных средств и систем для аналитической обработки данных;</p> <p>исследовать и разрабатывать математические модели, методы и алгоритмы по тематике проводимых исследований;</p> <p>анализировать новые возникающие проблемы и находить пути их решения.</p> <p>инструментальными средствами для решения задач компьютерной лингвистики;</p> <p>методами разработки и анализа концептуальных и теоретических моделей лингвистических систем;</p> <p>навыками выбора подходящего математического метода и программного обеспечения для решения прикладных задач.</p>

1.4 Особенности реализации дисциплины

Язык реализации дисциплины: Русский.

Дисциплина (модуль) реализуется с применением ЭО и ДОТ

URL-адрес и название электронного обучающего курса: <https://e.sfu-kras.ru/course/view.php?id=11068>.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	е
		1
Контактная работа с преподавателем:	1,06 (38)	
занятия лекционного типа	0,53 (19)	
практические занятия	0,53 (19)	
Самостоятельная работа обучающихся:	1,94 (70)	
курсовое проектирование (КП)	Нет	
курсовая работа (КР)	Нет	
Промежуточная аттестация (Экзамен)	1 (36)	

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

		Контактная работа, ак. час.							
№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа		Занятия семинарского типа				Самостоятельная работа, ак. час.	
				Семинары и/или Практические занятия		Лабораторные работы и/или Практикумы			
		Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС
1. Математическая лингвистика									
	1. Лингвистическое моделирование. Основные задачи и проблемы анализа естественно-языковых текстов. Задачи математической лингвистики.	2							
	2. Модели векторного представления слов			3					
	3. Концептуальное моделирование: алгебраический подход	4							
	4. Построение решеток замкнутых множеств и базисов импликаций			4					
	5. Вероятностное тематическое моделирование текстовых коллекций	7							
	6. Решение задач по теме «Вероятностное тематическое моделирование текстовых коллекций»			12					
2. Компьютерная лингвистика									

1. Автоматическое реферирование и аннотирование текста.	2							
2. Корпусная лингвистика	2							
3. Машинный перевод	2							
4. Исследовательский проект							50	
5. Разработка алгоритмов, тестирование и отладка программ							10	
6. Изучение изданий из списка основной и дополнительной литературы, а также изучение публикаций по данной предметной области в сети Интернет							10	
7.								
Всего	19		19				70	

4 Учебно-методическое обеспечение дисциплины

4.1 Печатные и электронные издания:

1. Айвазян С. А. 2011. т.1(М.: Маркет ДС).
2. Айвазян С. А. 2011. т.3(М.: Маркет ДС).
3. Загоруйко Н. Г. Прикладные методы анализа данных и знаний: монография(Новосибирск: Институт математики СО РАН).
4. Загоруйко Н. Г. Вычислительные системы: Вып. 117. Анализ данных в экспертных системах: сборник научных трудов(Новосибирск: Институт математики СО АН СССР).
5. Загоруйко Н. Г., Скоробогатов В. А. Вычислительные системы: Вып. 126. Экспертные системы и распознавание образов: сборник научных трудов(Новосибирск: Институт математики СО АН СССР).
6. Загоруйко Н. Г. Вычислительные системы: Вып. 150. Анализ последовательностей и таблиц данных: сборник научных трудов (Новосибирск: Институт математики СО РАН).
7. Загоруйко Н. Г. Вычислительные системы: Вып. 166. Обнаружение эмпирических закономерностей: сборник научных трудов(Новосибирск: Институт математики СО РАН).
8. Маккинли У. Python и анализ данных(Москва: ДМК Пресс).
9. Vandelanote L., Davidse K., Gentens C., Kimps D. Recent Advances in Corpus Linguistics. Developing and Exploiting Corpora(Amsterdam: Rodopi).
10. Антонио Д., Суджит П. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow(Москва: ДМК Пресс).
11. Семенова Д.В., Быкова В.В. Математическая и компьютерная лингвистика: [учеб-метод. материалы к изучению дисциплины для ...01.04.02.06 - Прикладная математика и информатика в гуманитарных и социально-экономических науках](Красноярск: СФУ).
12. Семенова Д.В. Математические методы анализа данных и распознавания образов: [учеб-метод. материалы к изучению дисциплины для ...01.04.02.06 Прикладная математика и информатика в гуманитарных и социально-экономических науках](Красноярск: СФУ).

4.2 Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства (программное обеспечение, на которое университет имеет лицензию, а также свободно распространяемое программное обеспечение):

1. Семинарские занятия проводятся в компьютерном классе с удаленным доступом к сети Интернет с установленным программным обеспечением:
2. Visual Studio,

3. система компьютерной вёрстки TeX,
4. Python, R (The R Project for Statistical Computing)
5. Visual Studio Code
6. MS Office,
7. Adobe Acrobat.

4.3 Интернет-ресурсы, включая профессиональные базы данных и информационные справочные системы:

1. научная библиотека СФУ <http://bik.sfu-kras.ru/>;
2. поисковые системы: Google или Яндекс.

5 Фонд оценочных средств

Оценочные средства находятся в приложении к рабочим программам дисциплин.

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

Используется проектор (для лекций или семинаров), слайды мультимедийных презентаций и компьютеры с предустановленным программным обеспечением и доступ в Интернет.