Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Смирнов Сергей Николаевич

Должность: врио ректора

Дата подписания: 18.10.2023 10:12:36 Уникальный программым программым программым рество науки и высшего образования Российской Федерации 69e375c64f7e975d4e8830e7 ФКБОУ5ВО «Тверской государственный университет»



Рабочая программа дисциплины (с аннотацией)

ОСНОВЫ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

Направление подготовки 02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

> Профиль подготовки Информатика и компьютерные науки

> > Для студентов 4-го курса Форма обучения - очная

> > > Составитель:

к.ф.-м.н. Б.Н.Карлов Карил в

I. Аннотация

1. Цель и задачи дисциплины:

Цель курса — ознакомить студентов с основными понятиями компьютерной лингвистики, с различными способами задания языков, с возможностью применения ЭВМ для обработки естественных языков.

2. Место дисциплины в структуре ООП

Дисциплина входит в раздел «Элективные дисциплины» части, формируемой участниками образовательных отношений, блока 1.

Предварительные знания и навыки. Знание курсов «Математическая логика и теория алгоритмов», «Теория автоматов и формальных языков».

Дальнейшее использование. Полученные знания используются для итоговой государственной аттестации, прохождении практики, а также в дальнейшей трудовой деятельности выпускников.

3. Объем дисциплины: 3 зачетных единицы, 108 академических часов, **в том** числе:

контактная аудиторная работа: лекции 30 часов, практические занятия 30 часов;

контактная внеаудиторная работа: контроль самостоятельной работы 0 часов, в том числе курсовая работа 0 часов;

самостоятельная работа: 48 часов, в том числе контроль 0 часов.

4. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Планируемые результаты освое-	Планируемые результаты обучения по
ния образовательной программы	дисциплине
(формируемые компетенции)	
ПК-1 Способен к поиску научно-	ПК-1.1 Знает основные приемы поиска
технической информации в обла-	научно-технической информации
сти теоретической и прикладной	ПК-1.2 Отбирает научно-техническую
информатики	информацию в соответствии с постав-
	ленной задачей
	ПК-1.3 Изучает и анализирует научно-
	техническую информацию на предмет их
	применимости для решения поставлен-
	ной задачей

Планируемые результаты освое-	Планируемые результаты обучения по	
ния образовательной программы	дисциплине	
(формируемые компетенции)		
ПК-2 Способен к анализу научно-	ПК-2.1 Классифицирует области ИКТ, к	
технических задач теоретической	которой относится поставленная задача	
и прикладной информатики	ПК-2.2 Анализирует известные методы	
	на предмет их применимости для реше-	
	ния поставленной задачей	
	ПК-2.3 Применяет типовые методы для	
	решения поставленной задачи	
	ПК-2.4 Анализирует полученные при ре-	
	шении задачи результаты	

- **5. Форма промежуточной аттестации и семестр прохождения:** зачёт в 7 семестре.
- 6. Язык преподавания: русский.

II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Учебная программа –			Контак	гная р	абота (ч	ac.)			
наименование разделов и тем			Всего	Ле	екции	_	ктиче- заня-	м. раб., зая ра-	Самостоя- тельная ра- бота, в том
	(час.)	Всего	в т.ч. практическая подготовка	Всего	в т.ч. прак- тическая подготовка	Контроль сам. раб., в т.ч. курсовая ра- бота	числе Контроль (час.)		
1	2	3	4	5	6	7	8		
Порождающие грам- матики	24	6		6			12		
Контекстно-зависимые грамматики	28	8		8			12		
Категориальные грам- матики	28	8		8			12		
Слабо контекстные грамматики	28	8		8			12		
Итого	108	30	0	30	0	0	48		

Учебная программа дисциплины

- 1. Порождающие грамматики
 - Определение порождающих грамматик. Иерархия Хомского.
 - Теорема о том, что грамматики типа 0 порождают в точности рекурсивно перечислимые языки.
 - Теорема о представлении рекурсивно перечислимых языков.
 - Алгоритмические проблемы для порождающих грамматик.
- 2. Контекстно-зависимые грамматики
 - Теорема об эквивалентности К3-грамматик и неукорачивающих грамматик.
 - Линейно ограниченные автоматы. Эквивалентность К3-грамматик и ЛО-автоматов.
 - Нормальная форма Куроды.
 - Свойства замкнутости для класса КЗ-языков.
 - Пример рекурсивного языка, не являющегося КЗ-языком.
 - Алгоритмические проблемы для КЗ-языков.
 - PSPACE-полнота проблемы принадлежности для КЗ-языков.
 - Деревья вывода и системы составляющих.
- 3. Категориальные грамматики
 - Классические категориальные грамматики. Эквивалентность классических категориальных грамматик и КС-грамматик.
 - Исчисление Ламбека и грамматики Ламбека.
 - Эквивалентность грамматик Ламбека и КС-грамматик.
 - Категориальные грамматики зависимостей (КГ3). Основные свойства КГ3.
 - Алгоритм анализа для КГЗ.
 - Деревья зависимостей. Связь деревьев зависимостей и систем составляющих.
- 4. Слабо-контекстные грамматики
 - Индексные грамматики.
 - Гнездовые стековые автоматы. Эквивалентность индексных грамматик и гнездовых автоматов.
 - Множественные КС-грамматики.
 - Линейные индексные грамматики.
 - Комбинаторные категориальные грамматики.
 - ТАГ-грамматики.
 - Головные грамматики.
 - Эквивалентность четырёх определений слабо-контекстных грамматик.
 - Автоматы со вложенными магазинами. Их эквивалентность ТАГ-грамматикам.
 - Применения комбинаторных категориальных грамматик для представления семантики в виде λ-термов.

III. Образовательные технологии

Учебная программа –	Вид занятия	Образовательные техноло-
наименование разделов и		ГИИ
тем		
Порождающие грамма-	лекции, практиче-	изложение теоретического
тики	ские занятия	материала, решение задач
Контекстно-зависимые	лекции, практиче-	изложение теоретического
грамматики	ские занятия	материала, решение задач
Категориальные грамма-	лекции, практиче-	изложение теоретического
тики	ские занятия	материала, решение задач
Слабо контекстные грам-	лекции, практиче-	изложение теоретического
матики	ские занятия	материала, решение задач

IV. Оценочные материалы для проведения текущей и промежуточной аттестации

Типовые контрольные задания и/или критерии для проверки индикатора ПК-1.1

Требования к	1	Показатели и кри-
обучающемуся	оценки знаний, умений, навыков	терии оценива-
		ния, шкала оцени-
		вания
Знать понятия	Примеры вопросов к зачёту:	оценка 3 — знает
порождающей	• Определение порождающих грамма-	определения ос-
грамматики,	тик. Иерархия Хомского.	новных понятий
системы со-	• Теорема о том, что грамматики типа	
ставляющих,	0 порождают в точности рекурсивно	ЛО-
связь систем	перечислимые языки.	автомат, система
составляющих	• Теорема о представлении рекур-	составляющих),
с граммати-	сивно перечислимых языков.	оценка 4 — кроме
ками	• Алгоритмические проблемы для по-	того знает основ-
	рождающих грамматик.	ные свойства КЗ-
	• Контекстно-зависимые грамматики.	грамматик,
	Эквивалентность КЗ-грамматик и не-	оценка 5 — кроме
	укорачивающих грамматик.	того знает доказа-
	• Линейно ограниченные автоматы.	тельства перечис-
	Эквивалентность КЗ-грамматик и	ленных утвержде-
	ЛО-автоматов.	ний
	• Нормальная форма Куроды.	
	• Свойства замкнутости для класса КЗ-	
	языков.	

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оцени-
	 Пример рекурсивного языка, не являющегося КЗ-языком. Алгоритмические проблемы для КЗ-языков. PSPACE-полнота проблемы принадлежности для КЗ-языков. Системы составляющих, размеченные системы составляющих, иерархизованные системы составляющих. Построение системы составляющих по дереву вывода. 	вания

Типовые контрольные задания и/или критерии для проверки индикатора ПК-1.2

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Знать понятие слабо-контекстных грамматик	 Примеры вопросов к зачёту: Индексные грамматики. Гнездовые стековые автоматы. Эквивалентность индексных грамматик и гнездовых автоматов. Понятие слабо контекстных грамматики. Множественные КС-грамматики. Линейные индексные грамматики. Комбинаторные категориальные грамматики. ТАГ-грамматики. Головные грамматики. Эквивалентность четырёх определений слабо-контекстных грамматик. Автоматы со вложенными магазинами. Их эквивалентность ТАГ-грамматикам. λ-термы. Применения комбинаторных категориальных грамматик для представления семантики в виде λ- 	оценка 3 — знает различные классы слабо контекстных грамматик, оценка 4 — кроме того знает различные свойства слабо контекстных грамматик, оценка 5 — кроме того знает доказательства перечисленных утверждений

Требования к	Типовые контрольные задания для	Показатели и кри-
обучающемуся	оценки знаний, умений, навыков	терии оценива-
		ния, шкала оцени-
		вания
	термов.	

Типовые контрольные задания и/или критерии для проверки индикатора ПК-1.3

-	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оцени-
		вания
различные слабо кон- текстные грам- матики и авто- маты по описа- нию языка	Примеры задач для контрольных работ: • Постройте	грамматики по описанию языка, оценка 4 — кроме того умеет строить автоматы с вложенными магазинами по описанию языка,

Типовые контрольные задания и/или критерии для проверки индикатора ПК-2.1

Требования к	Типовые контрольные задания для	Показатели и кри-
обучающемуся	оценки знаний, умений, навыков	терии оценива-
		ния, шкала оцени-
		вания
Знать понятие	Примеры вопросов к зачёту:	оценка 3 — знает
категориаль-	• Классические категориальные грам-	понятия классиче-
ной грамма-	матики. Эквивалентностьклассиче-	ской категориаль-
тики, дерева	ских категориальных грамматик и	ной грамматики,
зависимостей,	КС-грамматик.	дерева зависимо-
связь деревьев	• Исчисление Ламбека и грамматики	стей, оценка 4 —

Требования к	Типовые контрольные задания для	Показатели и кри-
обучающемуся	оценки знаний, умений, навыков	терии оценива-
		ния, шкала оцени-
		вания
зависимостей с	Ламбека.	кроме того знает
категориаль-	• Эквивалентность грамматик Лам-	другие типы кате-
ными грамма-	бека и КС-грамматик.	гориальных грам-
тиками	• Категориальные грамматики зависи-	матик, оценка 5
	мостей (КГЗ). Основные свойства	— кроме того
	КΓ3.	знает доказатель-
	• Алгоритм анализа для КГЗ.	ства перечислен-
	• Деревья зависимостей. Проективные	ных утверждений
	и непроективные деревья. Построе-	
	ние деревьев зависимостей с помо-	
	щью категориальных грамматик.	

Типовые контрольные задания и/или критерии для проверки индикатора ПК-2.2

Требования к обучающемуся	1	Показатели и критерии оценивания, шкала оценивания
тики по описанию языка,	Примеры задач для контрольных работ: • Постройте классическую категориальную грамматику для языка $L = \{ a^i b^j c^k \mid i, j, k \geqslant 0, i = j $ или $j = k $ }. • Постройте категориальную грамматику зависимостей для языка $L = \{ w \in \{ a, b, c, d \}^+ \mid w _a + w _c \geqslant w _b + w _d$ и все символы a стоят правее c }. • Докажите в исчислении Ламбека L секвенцию $p/(q \setminus (r \cdot (r \setminus q))) \rightarrow (p/(r \setminus q))/(q \setminus r)$. Примеры тем для самостоятельной работы: • Напишите программу, которая по классической категориальной грамматике и предложению на русском	оценка 3 — кроме того умеет стро- ить категориаль- ные грамматики по описанию языка, оценка 4 — кроме того умеет обосновывать правильность по- строения, оценка 5 — кроме того умеет выводить секвенции в исчислении Ламбека

Требования к	Типовые контрольные задания для	Показатели и кри-
обучающемуся	оценки знаний, умений, навыков	терии оценива-
		ния, шкала оцени-
		вания
	языке строит все его размеченные де-	
	ревья зависимостей.	

Типовые контрольные задания и/или критерии для проверки индикатора ПК-2.3

Требования к	Типовые контрольные задания для	Показатели и кри-
обучающемуся	оценки знаний, умений, навыков	терии оценива-
		ния, шкала оцени-
		вания
Уметь строить	Примеры задач для контрольных ра-	оценка 3 — умеет
неукорачиваю-	бот:	строить неукора-
щие грамма-	• Постройте неукорачивающую грам-	чивающие грамма
тики и ли-	матику, порождающую следующий	тики по описанию
нейно-ограни-	нейно-ограни- язык:	
ченные авто-	$L = \{ a^i b^j c^k \mid 0 < i \le j, j > k > 0 \}.$	кроме того умеет
маты по описа- • Постройте линейно-ограниче		строить линейно-
нию языка	автомат, распознающий следующий	ограниченные ав-
	язык:	томаты по описа-
	$L = \{ a^k \mid k=3^n \text{ для некоторого } n > 0 \}.$	нию языка,
	Примеры тем для самостоятельной	
	работы:	того умеет обос-
	• Напишите программу, которая по	новывать пра-
	КЗ-грамматике и предложению на	вильность постро-
	русском языке строит все его разме-	ения
	ченные системы составляющих.	

Типовые контрольные задания и/или критерии для проверки индикатора ПК-2.4

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Уметь строить формальное представление синтаксиса и семантики предложений на естественных языках	Примеры задач для контрольных работ: • Постройте размеченную иерархизованную систему составляющих для предложения «К востоку от боровых озёр лежат громадные мещёрские болота — мшары». • Постройте размеченное дерево зависимостей для предложения «Для точной диагностики заболеваний внутренних органов человека рентген незаменим». • Дано предложение «Мастер внимательно осматривал станок». Его словам сопоставлены следующие категории и λ-термы: — мастер — NP: М — внимательно — С: В — осматривал — (((S \ NP) \ C) / NP): λхуz.Оzxy — станок — NP: С Сократите категории до S и упростите получающийся λ-терм. На первом шаге примените к первой категории правило (> T), а на втором шаге примените к третьей и четвертой категориям правило (>).	оценка 3 — умеет строить системы составляющих и деревья зависимостей для предложений, оценка 4 — кроме того умеет строить размеченные иерархизованные системы составляющих и размеченые деревья зависимостей, оценка 5 — кроме того умеет представлять семантику в виде λ-термов

V. Учебно-методическое и информационное обеспечение дисциплины

1. Рекомендованная литература

а) Основная литература

- [1] Волосатова, Т.М. Информатика и лингвистика [Электронный ресурс]: Учебное пособие/Волосатова Т.М., Чичварин Н.В. Электрон. дан. М.: НИЦ ИНФРА-М, 2016. 196 с.: 60х90 1/16. (Высшее образование: Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-16-010977-0 Режим доступа: https://znanium.com/catalog/document?id=422587
- [2] Марченков, С.С. Конечные автоматы [Электронный ресурс]: учеб. пособие

- Электрон. дан. Москва: Физматлит, 2008. 56 с. Режим доступа: https://e.lanbook.com/book/59510. Загл. с экрана.
- [3] Короткова, М.А. Задачник по курсу "Математическая лингвистика и теория автоматов": учебное пособие для вузов [Электронный ресурс]: учеб. пособие / М.А. Короткова, Е.Е. Трифонова. Электрон. дан. Москва: НИЯУ МИФИ, 2012. 92 с. Режим доступа: https://e.lanbook.com/book/75843. Загл. с экрана.

б) Дополнительная литература

[4] Федосеева, Л.И. Основы теории конечных автоматов и формальных языков [Электронный ресурс]: учеб. пособие / Л.И. Федосеева, Р.М. Адилов, М.Н. Шмокин. — Электрон. дан. — Пенза: ПензГТУ, 2013. — 136 с. — Режим доступа: https://e.lanbook.com/book/62703. — Загл. с экрана.

2. Программное обеспечение

Наименование помещений	Программное обеспечение
_	цензиями): Linux Kubuntu, KDE, TeXLive, TeXStudio, LibreOffice, GIMP, Gwenview, ImageMagick, Okular,

3. Современные профессиональные базы данных и информационные справочные системы

- [1] 3 EC «ZNANIUM.COM» http://www.znanium.com
- [2] ЭБС «Университетская библиотека онлайн» https://biblioclub.ru
- [3] 3 EC IPRbooks http://www.iprbookshop.ru
- [4] ЭБС «Лань» http://e.lanbook.com
- [5]ЭБС «Юрайт» https://urait.ru
- [6]ЭБС ТвГУ http://megapro.tversu.ru/megapro/Web
- [7] Научная электронная библиотека eLIBRARY.RU (подписка на журналы) https://elibrary.ru/projects/subscription/rus_titles_open.asp
- [8] Репозитарий ТвГУ http://eprints.tversu.ru

4. Перечень ресурсов информационно-телекоммуникационной

сети «Интернет», необходимых для освоения дисциплины

- [1] Natural Language Processing, http://www.learnerstv.com/Free-Computer-Science-Video-lectures-ltv676-Page1.htm
- [2] Московский центр непрерывного математического образования, http://www.mccme.ru/

VI. Методические материалы для обучающихся по освоению дисциплины

Примеры задач для подготовки к контрольным работам

1. Постройте гнездовой стековый автомат для языка

$$L = \{a^k \mid k=n^2$$
 для некоторого $n > 0\}$

и обоснуйте его правильность.

2. Постройте комбинаторную категориальную грамматику для языка

 $L = \{w\#w^{-1} \mid w - \text{правильное скобочное слово}\}$

и обоснуйте ее правильность.

- 3. Постройте автомат с вложенными магазинами для языка $L = \{a^n b^n c^n d^n \mid n > 0\}.$
- 4. Постройте К3-грамматику для языка $L = \{a^n \mid n$ составное число $\}$.
- 5. Постройте размеченную иерархизованную систему составляющих и размеченное дерево зависимостей для предложения «Тощая торговка вяленой воблой торчала среди ящиков».
- 6. Обобщенной К3-грамматикой (ОК3-грамматикой) называется порождающая грамматика $G = (N, \Sigma, P, S)$, правила которой имеют вид $\xi A \eta \to \xi \alpha \eta$, где ξ , η , $\alpha \in (\Sigma \cup N)^*$, $A \in N$. Докажите, что любой рекурсивно перечислимый язык порождается некоторой ОК3-грамматикой.
- 7. Докажите, что класс языков типа 0 замкнут относительно тасовки: $TAC(L_1, L_2) = \{ x_1y_1...x_ny_n \mid x_i, y_i \in \Sigma^*, x_1...x_n = x, y_1...y_n = y, x \in L_1, y \in L_2 \}.$

Требования к рейтинг контролю (7 семестр)

Контрольная работа 1. Темы: порождающие грамматики, К3-грамматики, системы составляющих, деревья зависимостей. Пример задания:

1. Постройте неукорачивающую грамматику, порождающую следующий язык:

$$L = \{ a^n b^m a^n b^m \mid m, n > 0 \}.$$

- 2. Постройте размеченную иерархизованную систему составляющих и размеченное дерево зависимостей для предложения «Лебеди стадом летели из холодной стороны в тёплые земли».
- 3. Докажите, что класс языков типа 0 замкнут относительно правого деления:

 $L_1/L_2 = \{ w \in \Sigma^* \mid \text{существует слово } x \in L_2 \text{ такое, что } wx \in L_1 \}.$

За решение каждой задачи выставляется максимум 10 баллов.

Самостоятельная работа 1. Темы: К3-грамматики, системы составляющих,

деревья зависимостей. Пример задания:

Напишите программу, которая по системе составляющих строит согласованное с ней дерево зависимостей.

За решение задачи выставляется максимум 20 баллов.

Контрольная работа 2. Темы: категориальные грамматики, слабо-контекстные грамматики. Пример задания:

1. Постройте линейную индексную грамматику для языка

$$L = \{ w*w^{-1}*w \mid w \in \{ 0, 1 \}^* \}.$$

2. Постройте гнездовой автомат для языка

$$L = \{ a^n b^{k+m} \mid n \geqslant 0, m \geqslant 0, k = n^2 \}.$$

3. Разрешим использовать в гнездовых автоматах ещё три типа команд: Π_x перейти к левому брату, если метка ребра, входящего в текущую вершину, равна x; Π_x — перейти к правому брату, если метка ребра, входящего в текущую вершину, равна x; $U_{x,y}$ — изменить метку текущего ребра с x на y. Докажите, что любой язык типа 0 распознаётся расширенным гнездовым автоматом.

За решение каждой задачи выставляется максимум 10 баллов.

Самостоятельная работа 2. Темы: слабо-контекстные грамматики. Пример задания:

Напишите программу, которая по линейной индексной грамматике и слову проверяет, выводимо ли слово в этой грамматике.

За решение задачи выставляется максимум 20 баллов.

Общая сумма В сумме за все задачи выставляет не более 100 баллов.

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

Для аудиторной работы

·	Материально-техническое оснащение помеще-
ний	ний
ская обл., г. Тверь, пер.	Набор учебной мебели, экран проектор.
Садовый, д. 35)	

Для самостоятельной работы

Наименование помещений	Материально-техническое оснащение помещений
	Набор учебной мебели, доска маркерная, компьютер, сервер (системный блок), концентратор сетевой.

Наименование	иенование Материально-техническое оснащение помещений	
помещений		
(170002, Твер-		
ская обл., г.		
Тверь, пер. Садо-		
вый, д. 35)		

VIII. Сведения об обновлении рабочей программы дисциплины

√ <u>0</u> /π	Обновленный раздел рабочей программы дисциплины	Описание внесённых изменений	Дата и протокол заседания кафедры, утвердившего изменения
1	I. Аннотация. IV. Оценочные материалы для проведения текущей и промежуточной аттестации	Изменения в учебные планы и в рабочие программы дисциплин, формирующих новые/измененные компетенции в соответствии с приказом Минобрнауки России от 26.11.2020 г. № 1456	Протокол № 7 за- седания ученого совета от 30.12.2021 года
2	V. Учебно-методическое и информационное обеспечение дисциплины	Внесены изменения в программное обеспечение	От 29.09.2022 года, протокол № 2 ученого совета факультета
3	VII. Материально-техниче- ское обеспечение	Внесены изменения в материально-техническое обеспечение аудиторий	От 29.09.2022 года, протокол № 2 ученого совета факультета