

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 25.10.2023 15:05:09
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Тверской государственный университет»



Утверждаю:
Руководитель ООП
С.М. Дудаков
25 октября 2023 г.

Рабочая программа дисциплины (с аннотацией)
ОСНОВЫ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

Направление подготовки
02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА
И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Направленность (профиль)
Информатика и компьютерные науки

Для студентов 4 курса
Очная форма

Составитель:
к.ф.-м.н. Карлов Б.Н. *Карлов*

Тверь, 2023

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: 2 зачетных единицы, 72 академических часа, 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: 12 часов, в том числе контроль 0 часов.

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

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

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
ПК-2 Способен к анализу научно-технических задач теоретической и прикладной информатики	ПК-2.1 Классифицирует области ИКТ, к которой относится поставленная задача ПК-2.2 Анализирует известные методы на предмет их применимости для решения поставленной задачей ПК-2.3 Применяет типовые методы для решения поставленной задачи ПК-2.4 Анализирует полученные при решении 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 – наименование разделов и тем	Всего (час.)	Контактная работа (час.)					Самостоятельная работа, в том числе Контроль (час.)
		Лекции		Практические занятия		Контроль сам. раб., в т.ч. курсовая работа	
		Всего	в т.ч. практическая подготовка	Всего	в т.ч. практическая подготовка		
1	2	3	4	5	6	7	8
Порождающие грамматики	15	6		6			3
Контекстно-зависимые грамматики	19	8		8			3
Категориальные грамматики	19	8		8			3
Слабо контекстные грамматики	19	8		8			3
Итого	72	30	0	30	0	0	12

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 КЗ-грамматик и неукорачивающих грамматик.
 - Линейно ограниченные автоматы. Эквивалентность КЗ-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 (КГЗ). Основные свойства КГЗ.
 - 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 – наименование разделов и тем	Вид занятия	Образовательные технологии
Порождающие грамматики	лекции, практические занятия	изложение теоретического материала, решение задач
Контекстно-зависимые грамматики	лекции, практические занятия	изложение теоретического материала, решение задач
Категориальные грамматики	лекции, практические занятия	изложение теоретического материала, решение задач
Слабо контекстные грамматики	лекции, практические занятия	изложение теоретического 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 обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Знать понятия порождающей грамматики, системы составляющих, связь систем составляющих с грамматиками	<p>Примеры вопросов к зачёту:</p> <ul style="list-style-type: none"> • 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. • Контекстно-зависимые грамматики. Эквивалентность КЗ-грамматик и неукорачивающих грамматик. • Линейно ограниченные автоматы. Эквивалентность КЗ-грамматик и ЛО-автоматов. • Нормальная форма Куроды. • Свойства замкнутости для класса КЗ-языков. • Пример рекурсивного языка, не являющегося КЗ-языком. 	<p>оценка 3 — знает определения основных понятий (КЗ-грамматика, ЛО-автомат, система составляющих), оценка 4 — кроме того знает основные свойства КЗ-грамматик, оценка 5 — кроме того знает доказательства перечисленных утверждений</p>

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	<ul style="list-style-type: none"> • 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. 	

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	<p>Примеры вопросов к зачёту:</p> <ul style="list-style-type: none"> • Индексные грамматики. • Гнездовые стековые автоматы. Эквивалентность индексных грамматик и гнездовых автоматов. • Понятие слабо контекстных грамматик. • 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-термов. 	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

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

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Уметь строить различные слабо контекстные грамматики и автоматы по описанию языка	<p>Примеры задач для контрольных работ:</p> <ul style="list-style-type: none"> • Постройте ТАГ-грамматику для языка $L = \{ a^i b^j c^k \mid 0 < i < j < k \}.$ • Постройте комбинаторную категориальную грамматику для языка $L = \{ w * w^{-1} * w \mid w \in \{ 0, 1 \}^* \}.$ • Постройте автомат с вложенными магазинами для языка $L = \{ a^{3i} b^{2i} c^i \mid i > 0 \}.$ • Постройте гнездовой автомат для языка $L = \{ a^k \mid k = 2^n \text{ для некоторого } n > 0 \}.$ 	<p>оценка 3 — умеет строить слабо контекстные грамматики по описанию языка, оценка 4 — кроме того умеет строить автоматы с вложенными магазинами по описанию языка, оценка 5 — кроме того умеет строить гнездовые автоматы по описанию языка</p>

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

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Знать понятие категориальной грамматики, дерева зависимостей, связь деревьев зависимостей с категориальными грамматиками	<p>Примеры вопросов к зачёту:</p> <ul style="list-style-type: none"> • Классические категориальные грамматики. Эквивалентность классических категориальных грамматик и КС-грамматик. • Исчисление Ламбека и грамматики Ламбека. • Эквивалентность грамматик Ламбека и КС-грамматик. • Категориальные грамматики зависимостей (КГЗ). Основные свойства КГЗ. • Алгоритм анализа для КГЗ. 	<p>оценка 3 — знает понятия классической категориальной грамматики, дерева зависимостей, оценка 4 — кроме того знает другие типы категориальных грамматик, оценка 5 — кроме того знает доказатель-</p>

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	<ul style="list-style-type: none"> • Деревья зависимостей. Проективные и непроективные деревья. Построение деревьев зависимостей с помощью категориальных грамматик. 	ства перечисленных утверждений

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

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Уметь строить категориальные грамматики по описанию языка, строить деревья зависимостей	<p>Примеры задач для контрольных работ:</p> <ul style="list-style-type: none"> • Постройте классическую категориальную грамматику для языка $L = \{ a^i b^j c^k \mid i, j, k \geq 0, i = j \text{ или } j = k \}$. • Постройте категориальную грамматику зависимостей для языка $L = \{ w \in \{ a, b, c, d \}^+ \mid w _a + w _c \geq w _b + w _d \}$ и все символы a стоят правее c. • Докажите в исчислении Ламбека L секвенцию $p/(q \setminus (r \cdot (r \setminus q))) \rightarrow (p/(r \setminus q))/(q \setminus r)$. <p>Примеры тем для самостоятельной работы:</p> <ul style="list-style-type: none"> • Напишите программу, которая по классической категориальной грамматике и предложению на русском языке строит все его размеченные деревья зависимостей. 	оценка 3 — кроме того умеет строить категориальные грамматики по описанию языка, оценка 4 — кроме того умеет обосновывать правильность построения, оценка 5 — кроме того умеет выводить секвенции в исчислении Ламбека

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

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Уметь строить неукорачивающие грамматики и линейно-ограниченные автоматы по описанию языка	<p>Примеры задач для контрольных работ:</p> <ul style="list-style-type: none"> • Постройте неукорачивающую грамматику, порождающую следующий язык: $L = \{ a^i b^j c^k \mid 0 < i \leq j, j > k > 0 \}.$ • Постройте линейно-ограниченный автомат, распознающий следующий язык: $L = \{ a^k \mid k=3^n \text{ для некоторого } n > 0 \}.$ Примеры тем для самостоятельной работы: • Напишите программу, которая по КЗ-грамматике и предложению на русском языке строит все его размеченные системы составляющих. 	оценка 3 — умеет строить неукорачивающие грамматики по описанию языка, оценка 4 — кроме того умеет строить линейно-ограниченные автоматы по описанию языка, оценка 5 — кроме того умеет обосновывать правильность построения

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	<p>Примеры задач для контрольных работ:</p> <ul style="list-style-type: none"> • 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». 	оценка 3 — умеет строить системы составляющих и деревья зависимостей для предложений, оценка 4 — кроме того умеет строить размеченные иерархизованные системы составляющих и размеченные деревья

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	<ul style="list-style-type: none"> • Дано предложение «Мастер внимательно осматривал станок». Его словам сопоставлены следующие категории и λ-термы: <ul style="list-style-type: none"> – мастер — NP : M – внимательно — C : V – осматривал — (((S \ NP) \ C) / NP) : $\lambda x y z . O z x y$ – станок — NP : C <p>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 ($>$).</p>	<p>зависимостей, оценка 5 — кроме того умеет представлять семантику в виде λ-термов</p>

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.: 60x90 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
Ауд. 201а (компьютерная лаборатория ПМиК) (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Перечень программного обеспечения (со свободными лицензиями): Linux Kubuntu, KDE, TeXLive, TeXStudio, LibreOffice, GIMP, Gwenview, ImageMagick, Okular, Skanlite, Google Chrome, KDE Connect, Konversation, KRDC, KTorrent, Thunderbird, Elisa, VLC media player, PulseAudio, KAppTemplate, KDevelop, pgAdmin4, PostgreSQL, Qt, QtCreator, R, RStudio, Visual Studio Code, Perl, Python, Ruby, clang, clang++, gcc, g++, nasm, flex, bison, Maxima, Octave, Dolphin, HTop, Konsole, KSystemLog, Xterm, Ark, Kate, Kcalc, Krusader, Spectacle, Vim.

3. Современные профессиональные базы данных и информационные справочные системы

- [1] ЭБС «ZNANIUM.COM» <http://www.znanium.com>
- [2] ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru>
- [3] ЭБС 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] Репозиторий 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

Важной составляющей данного раздела РПД являются требования к рейтинг-контролю с указанием баллов, распределенных между модулями и видами работы обучающихся.

Максимальная сумма баллов по учебной дисциплине, заканчивающейся

зачетом, по итогам семестра составляет 100 баллов (50 баллов - 1-й модуль и 50 баллов - 2-й модуль).

Студенту, набравшему 40 баллов и выше по итогам работы в семестре, в экзаменационной ведомости и зачетной книжке выставляется оценка «зачтено». Студент, набравший до 39 баллов включительно, сдает зачет.

Распределение баллов по модулям устанавливается преподавателем и может корректироваться.

Примеры задач для подготовки к контрольным работам

1. Постройте гнездовой стековый автомат для языка

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

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 КЗ-грамматику для языка $L = \{a^n \mid n \text{ — составное число}\}$.

5. Постройте размеченную иерархизованную систему составляющих и размеченное дерево зависимостей для предложения «Тощая торговка вяленой воблой торчала среди ящичков».

6. Обобщенной КЗ-грамматикой (ОКЗ-грамматикой) называется порождающая грамматика $G = (N, \Sigma, P, S)$, правила которой имеют вид $\xi A \eta \rightarrow \xi a \eta$, где $\xi, \eta, a \in (\Sigma \cup N)^*$, $A \in N$. Докажите, что любой рекурсивно перечислимый язык порождается некоторой ОКЗ-грамматикой.

7. Докажите, что класс языков типа 0 замкнут относительно тасовки:

$$\text{TAC}(L_1, L_2) = \{x_1 y_1 \dots x_n y_n \mid x_i, y_j \in \Sigma^*, x_1 \dots x_n = x, y_1 \dots y_n = y, x \in L_1, y \in L_2\}.$$

Требования к рейтинг контролю (7 семестр)

Контрольная работа 1. Темы: порождающие грамматики, КЗ-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: КЗ-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

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

3. Разрешим использовать в гнездовых автоматах ещё три типа команд: L_x — перейти к левому брату, если метка ребра, входящего в текущую вершину, равна x ; P_x — перейти к правому брату, если метка ребра, входящего в текущую вершину, равна x ; $I_{x,y}$ — 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

Наименование помещений	Материально-техническое оснащение помещений
Ауд. 308 (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Набор учебной мебели, экран проектор.

Для самостоятельной работы

Наименование помещений	Материально-техническое оснащение помещений
Ауд. 201а (компьютерная лаборатория ПМиК)	Набор учебной мебели, доска маркерная, компьютер, сервер (системный блок), концентратор сетевой.

Наименование помещений	Материально-техническое оснащение помещений
(170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	

VIII. Сведения об обновлении рабочей программы дисциплины

№ п/п	Обновленный раздел рабочей программы дисциплины	Описание внесённых изменений	Дата и протокол заседания кафедры, утвердившего изменения
1			
2			
3			
4			