

Государственное образовательное учреждение
"Приднестровский государственный университет им. Т.Г. Шевченко"

Физико-технический институт

Факультет информатики и вычислительной техники

Кафедра программного обеспечения вычислительной техники

УТВЕРЖДАЮ

Заведующий кафедрой ПОВТ



С.Г. Федорченко

«29» августа 2024 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине
КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Направление подготовки
2.09.04.01 Информатика и вычислительная техника

Профиль подготовки
Информационное и программное обеспечение вычислительных систем

Квалификация (степень)

выпускника:

магистр

Форма обучения:

очная

Год набора:

2023 г.

Разработал:

к.п.н., доцент кафедры ПОВТ



/С.В. Помян

«29» августа 2024 г.

Тирасполь, 2024

Паспорт фонда оценочных средств по учебной дисциплине

1. В результате изучения дисциплины «Компьютерная лингвистика» у обучающегося должны быть сформированы следующие компетенции:

Категория (группа) компетенций	Код и наименование	Код и наименование индикатора достижения универсальной компетенции
<i>Профессиональные компетенции и индикаторы их достижения</i>		
-	ПК-1. Способен управлять развитием БД	ИД-1ПК-1 Знать: способы управления развитием БД ИД-2 ПК-1 Уметь: управлять развитием БД ИД-3ПК-1 Владеть: навыками управления развитием БД

2. Программа оценивания контролируемой компетенции:

Текущая аттестация	Контролируемые модули, разделы (темы) дисциплины их название	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
РУБЕЖНЫЙ КОНТРОЛЬ	Раздел 1 Компьютерная морфология и синтаксис Раздел 2 Компьютерное представление значений Раздел 3 Корпусная лингвистика	ПК-1	Презентация №1 Кейс-задача №1 Кейс-задача №2
РУБЕЖНАЯ АТТЕСТАЦИЯ	Раздел 4 Информационный поиск. Извлечение информации Раздел 5 Анализ тональности. Авторство текста		Кейс-задача №3 Кейс-задача №4 Итоговый тест
Промежуточная аттестация		Код контролируемой компетенции (или ее части)	Наименование оценочного средства
№1		ПК-1	Зачет

3. Показатели и критерии оценивания компетенции по этапам формирования, описание шкал оценивания

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
Первый этап	ИД-1ПК-1 Знать: способы управления развитием БД	Не знает	Знает основные понятия, но не знает современные способы управления развитием БД	Знает способы управления развитием БД, но допускает незначительные ошибки	Знает способы управления развитием БД

Этапы оценивания компетенции	Показатели достижения заданного уровня освоения компетенции	Критерии оценивания результатов обучения			
		2	3	4	5
Второй этап	ИД-2 ПК-1 Уметь: управлять развитием БД	Не умеет	Умеет применять базовые знания, но не умеет самостоятельно разрабатывать стратегию управления развитием БД	Умеет управлять развитием БД, но не умеет выбирать оптимальные решения	Умеет управлять развитием БД
Третий этап	ИД-3 ПК-1 Владеть: навыками управления развитием БД	Не владеет	Владеет основными навыками работы с БД, но не владеет современными навыками управления БД.	Владеет навыками управления развитием БД, но не владеет навыками использования оптимальных решений	Владеет навыками управления развитием БД

4. Шкала оценивания

Согласно Положению «О порядке организации аттестации в ИТИ ПГУ им. Т.Г. Шевченко, итоговая оценка представляет собой сумму баллов, полученных студентом по итогу освоения дисциплины (модуля):

Оценка в традиционной шкале	Оценка в 100-балльной шкале	Буквенные эквиваленты оценок в шкале 3Е (% успешно аттестованных)
5 (отлично)	88–100	А (отлично) – 88-100 баллов
4 (хорошо)	70–87	В (очень хорошо) – 80-87баллов
		С (хорошо) – 70-79 баллов
3 (удовлетворительно)	50–69	Д (удовлетворительно) – 60-69 баллов
		Е (посредственно) – 50-59 баллов
2 (неудовлетворительно)	0–49	Гх – неудовлетворительно, с возможной пересдачей – 21-49 баллов
		Г – неудовлетворительно, с повторным изучением дисциплины – 0-20 баллов

Расшифровка уровня знаний, соответствующего полученным баллам, дается в таблице, указанной ниже

А	“Отлично” - теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.
В	“Очень хорошо” - теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом в основном сформированы, все предусмотренные программой обучения учебные задания выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному.
С	“Хорошо” - теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения

	ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками.
D	“Удовлетворительно” - теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки.
E	“Посредственно” - теоретическое содержание курса освоено частично, некоторые практические навыки работы не сформированы, многие предусмотренные программой обучения учебные задания не выполнены, либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному.
FX	“Условно неудовлетворительно” - теоретическое содержание курса освоено частично, необходимые практические навыки работы не сформированы, большинство предусмотренных программой обучения учебных заданий не выполнено, либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий.
F	“Безусловно неудовлетворительно” - теоретическое содержание курса не освоено, необходимые практические навыки работы не сформированы, все выполненные учебные задания содержат грубые ошибки, дополнительная самостоятельная работа над материалом курса не приведет к какому-либо значимому повышению качества выполнения учебных заданий.

5. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций при изучении учебной дисциплины в процессе освоения образовательной программы

5.1 Примерный перечень тематик презентации №1

1. Основные понятия математической лингвистики. Устройство систем автоматизированной обработки текстов. Основные этапы построения. Графематический и морфологический анализ.
2. Методы задания синтаксической структуры предложений. Системы составляющих. Деревья подчинения. Проблемы синтаксического анализа.
3. Принципы построения синтаксических анализаторов. Проект "Автоматическая Обработка Текста" (АОТ). Синтаксический парсер LINK. Применение морфологического и синтаксического анализа в поисковых системах.
4. Методы теоретического исследования семантики текстов. Лексические функции. Валентности слов. Теоретико-модельный подход.
5. Теоретико-множественные модели языка. Основные определения: отмеченные последовательности, контексты, дистрибутивные классы и др. Формализация понятий «часть речи» и «синтаксический тип». Формализация понятия «грамматический род». Формализация «категории падежа».
6. Представление знаний для компьютерной обработки. Тезаурусы и онтологии. WordNet. Общие принципы построения. Меры семантической близости.
7. Семантические сети. Фреймы. Формальные логические модели. Искусственные языки и нотации, применяемые в компьютерной лингвистике.
8. Корпусная лингвистика. Частотные методы в компьютерной лингвистике.
9. Теория речевых действий. Классификация речевых действий.
10. Модели и методы автоматической классификации и кластеризации текстовой информации. Иерархические и вероятностные подходы. Интеллектуальный анализ данных.

11. Автоматические системы извлечения информации. Алгоритмические основы. Принципы обработки неструктурированной и плохо структурированной информации. Тематическая индексация текстов.
12. Формальные методы определения автора текста. Лингвостатистические параметры. Статистические методы атрибуции. Авторский инвариант и лингвистические спектры. Применение методов кластеризации и классификации для установления авторства текстов.
13. Социальные сети. Направления исследований. Графовые модели анализа социальных сетей. Понятие центральности. Методы обнаружения сообществ и анализ связанных подгрупп. Модели динамики сети.
14. Методы обнаружения спама: вероятностные и статистические, байесовский классификатор.

5.2 Типовой вариант Кейс-задача №1.

Тема: Ознакомление с проектом «Автоматическая Обработка Текста» и WordNet

Цель кейс-задачи: Познакомиться с технологиями, примененными в проектах

Задание: описать основные приемы работы и рассмотреть возможность применения алгоритмов в магистерской диссертации

Типовое задание:

Проект "Автоматическая Обработка Текста". Синтаксический парсер LINK. Применение морфологического и синтаксического анализа в поисковых системах. Методы задания синтаксической структуры предложений. Системы составляющих. Деревья подчинения. Проблемы синтаксического анализа. Основные понятия математической лингвистики. Устройство систем автоматизированной обработки текстов. Основные этапы построения. Графематический и морфологический анализ. Формализация понятий «часть речи» и «синтаксический тип». Формализация понятия «грамматический род». Формализация «категории падежа». Представление знаний для компьютерной обработки. Тезаурусы и онтологии. WordNet. Общие принципы построения. Меры семантической близости

5.3 Типовой вариант Кейс-задача №2.

Тема: Корпусная лингвистика

Цель кейс-задачи: Познакомиться с корпусами русского и английского языка, библиотеками обработки текстов

Задание: описать основные части речи русского языка ,разметку их в OpenCorpora и НКРЯ, привести примеры омонимии в корпусах.

Типовое задание:

Рассмотреть библиотеки автоматической обработки текстов, стемминг и лемматизацию. Привести примеры применения библиотек автоматической обработки текста, привести примеры снятия омонимии.

5.4 Типовой вариант Кейс-задача № 3.

Тема: Алгоритмы работы с текстом

Цель кейс-задачи: Познакомиться и разобрать алгоритмы и технологии работы с текстовой информацией

Задание: описать основные алгоритмы работы с текстовой информацией и рассмотреть возможность применения алгоритмов в магистерской диссертации

Типовое задание:

Классификация речевых действий. Модели и методы автоматической классификации и кластеризации текстовой информации. Иерархические и вероятностные подходы. Интеллектуальный анализ данных. Автоматические системы извлечения информации. Алгоритмические основы. Принципы обработки неструктурированной и плохо структурированной информации. Тематическая индексация текстов. Применение методов кластеризации и классификации для установления авторства текстов. Графовые модели анализа социальных сетей. Понятие центральности. Методы обнаружения. Методы обнаружения спама: вероятностные и статистические, байесовский классификатор.

5.5 Типовой вариант Кейс-задача №4.

Тема: Анализ тональности. Авторство текстов.

Цель кейс-задачи: Познакомиться с основными принципами определения тональности и авторства текста

Задание: Рассмотреть алгоритмы определения анализа тональности текста. Подход с использованием правил и словарей. Анализ тональности текста в целом и по аспектам. Методы определения авторства текста, формальные методы определения авторства текста.

Типовое задание:

Рассмотреть библиотеки автоматической, позволяющие производить анализ тональности текста, разработать программу, определяющую тональность текстов (например, отзывов на фильмы или продукты), метод определения тональности выбрать самостоятельно (с использованием словарей или правил). Описать методы определения авторства текста, формальные методы определения авторства текста.

5.6 Типовой тест промежуточной аттестации

1. Объяснение фактов языка является задачей

- a) теоретической лингвистики
- b) прикладной лингвистики
- c) математической лингвистики
- d) компьютерной лингвистики

2. Большие банки хранения информации, которыми пользуются специалисты по гуманитарным наукам, являются достижением

- a) теоретической лингвистики
- b) компьютерной лингвистики
- c) прикладной лингвистики
- d) математической лингвистики

3. Компьютерная лингвистика является одним из направлений

- a)Теоретической лингвистики
- b)Когнитивной лингвистики
- c)Прикладной лингвистики
- d)Формальной лингвистики

4. Компьютерная лингвистика изучает

- a)Проявления культуры народа, которые отразились в языке
- b)Язык как общий когнитивный механизм, когнитивный инструмент репрезентации информации
- c)Использование языка в коммуникации и сопровождающее его невербальное поведение
- d)Лингвистические основы информатики и аспекты моделирования языка и мышления в компьютерной среде при помощи компьютерных программ

5. Интересы компьютерной лингвистики лежат в области

- a)создания общих искусственных языков для представления информации
- b)создания Искусственного Интеллекта и решения вопросов машинного перевода
- c)решение вопросов роли языка в обществе и воздействия общества на язык
- d)решения вопросов функционирования стилей языка

6. Компьютерная лингвистика как особое направление оформилось

- a)В конце 19 в.
- b)В начале 19 в.
- c)В 60-х годах 20 в.
- d)В 20-х годах 20 в.

7. Компьютерная лингвистика как прикладная дисциплина выделяется

- a)По использованию компьютерных средств обработки языковых данных
- b)По новым методам анализа языка
- c)По использованию математических методов исследования языковых
- d)По привлечению языкового материала древних языков

8. К основным направлениям компьютерной лингвистики не относится

- a)Теория коммуникации
- b)Компьютерная лексикография
- c)Машинный перевод
- d)Электронная лингводидактика

9. Компьютерная лингвистика необходима филологу для

- a)Приобретения умений и навыков использования компьютерных технологий в научно-исследовательской работе
- b)Получения необходимых знаний в профессиональной деятельности
- c)Создания компетентной инновационной филологической среды
- d)Создания имиджа современного перспективного ученого-филолога

10. Информационные технологии в лингвистике - это

- a)Последовательность действий при решении лингвистической задачи
- b)Совокупность законов, методов и средств получения, хранения, передачи и преобразования информации о языке с помощью компьютера
- c)Программа, расширяющая возможности операционной системы(например, анти-вирусные программы, архиваторы и т.д.)
- d)Совокупность систематизированных сведений, фактов, событий в электронном варианте

11. Лингвистический алгоритм – это

- a) Последовательность действий при решении лингвистической задачи
- b) Совокупность систематизированных сведений, фактов, событий в электронном варианте
- c) Программа, расширяющая возможности операционной системы (например, антивирусные программы, архиваторы и т.д.)
- d) Совокупность законов, методов и средств получения, хранения, передачи и преобразования информации о языке с помощью компьютера

12. К инструментальным средствам в рамках компьютерной лексикографии относятся

- a) компьютерные картотеки
- b) базы данных
- c) программы обработки текста
- d) бумажные аналоги словарей

13. Система машинного перевода включает в себя

- a) корпуса языков
- b) методы и приемы анализа
- c) двуязычные словари
- d) искусственные языки

14. Утилита – это

- a) Последовательность действий при решении лингвистической задачи
- b) Программа, расширяющая возможности операционной системы (например, антивирусные программы, архиваторы и т.д.)
- c) Совокупность систематизированных сведений, фактов, событий в электронном варианте
- d) Совокупность законов, методов и средств получения, хранения, передачи и преобразования информации о языке с помощью компьютера

15. База данных – это

- a) Последовательность действий при решении лингвистической задачи
- b) Совокупность законов, методов и средств получения, хранения, передачи и преобразования информации о языке с помощью компьютера
- c) Совокупность систематизированных сведений, фактов, событий в электронном варианте
- d) Программа, расширяющая возможности операционной системы (например, антивирусные программы, архиваторы и т.д.)

16. Дистанционное обучение от традиционных форм обучения отличают следующие характерные черты

- a) Гибкость (возможность заниматься в удобное для себя время, в удобном месте и темпе)
- b) Параллельность (параллельное с профессиональной деятельностью обучение, т.е. без отрыва от производства)
- c) Охват (одновременное обращение ко многим источникам учебной информации (электронным библиотекам, банкам данных, базам знаний и т.д.) большого количества обучающихся)
- d) Использование опыта традиционной методики проведения занятий

17. Язык программирования представляет собой

- a) Искусственный язык, предназначенный для обработки информации с помощью компьютера
- b) Компьютерная программа, преобразующая текст из бумажного варианта в электронный текстовый файл
- c) Специальный искусственный язык для коммуникации людей разных народов (например, эсперанто)
- d) Кодированный язык символов для передачи информации

18. Искусственные языки – это

- a) Фортран, Паскаль, Бейсик и др.
- b) Знаковые системы, создаваемые для использования в тех областях науки и техники, где применение естественного языка ограничено или невозможно
- c) Агглютинативные языки
- d) Флективные языки

19. Информационная система – это

- a) хранилище информации, снабженное процедурами ввода, поиска, размещения и выдачи информации
- b) набор компьютерных программ, дающих возможность производить обработку различных данных
- c) электронная база данных
- d) совокупность средств для поиска сведений из различных информационных источников

20. Большие банки хранения информации, которыми пользуются специалисты по гуманитарным наукам, являются достижением

- a) теоретической лингвистики
- b) компьютерной лингвистики
- c) сопоставительной лингвистики
- d) математической лингвистики

21. Автоматическое аннотирование текста представляет собой

- a) Составление реферата текста с помощью компьютера
- b) Составление аннотации текста с помощью компьютера
- c) Составление гипертекста и ввод его в контекст
- d) Составление рецензии на текст с помощью компьютера

22. Автоматическое реферирование текста – это

- a) Составление реферата текста с помощью компьютера
- b) Составление краткого содержания текста, включая цели, задачи, методы и результаты исследования
- c) Составление аннотации текста с помощью компьютера
- d) Составление гипертекста и ввод его в контекст

23. Основным инструментарием компьютерной лингвистики являются

- a) Сценарии
- b) Семы
- c) Актанты
- d) Фреймы

24. Фреймы представляют собой

- a) Концептуальные структуры для процедурного представления знаний

- b) Концептуальные структуры для декларативного представления знаний
- с) Структуры знаний, фиксирующие представления о возможных действиях, ведущих к достижению определенной цели
- d) Структуры языкового уровня, имеющие идентифицирующие семантические признаки

25. Сценарии являются

- a) Структурами языкового уровня, имеющими идентифицирующие семантические признаки
- b) Структуры знаний, фиксирующие представления о возможных действиях, ведущих к достижению определенной цели
- с) Концептуальные структуры для декларативного представления знаний
- d) Концептуальными структурами для процедурного представления знаний

26. Мультимедийные обучающие программы представляют собой

- a) Программы составления на компьютере аннотации текста
- b) Программы составления на компьютере реферата текста
- с) Программы, использующие текст, звук, графическое и видео-изображения
- d) Программы распознавания и синтеза речи

27. Кому принадлежат слова: «Современная цивилизация есть цивилизация словаря»

- a) В. Виноградову
- b) А. Рею
- с) А. Франсу
- d) Н. Хомскому

28. Компьютерная (машинная) лексикография занимается

- a) решением проблем автоматической обработки текста
- b) решением проблем создания и функционирования электронных словарей разных типов
- с) решением вопросов функционирования большого массива текстов языков
- d) решением задач распознавания текста и моделирования языка

29. Отражая состояние цивилизации, закрепляя итоги человеческого познания, словарь выполняет

- a) важную культурную функцию
- b) важную коммуникативную функцию
- с) важную дискурсивную функцию
- d) важную функциональную роль

30. Словарь представляет собой

- a) Фонетическую систему языка
- b) Словарный состав языка
- с) Справочник, содержащий определенным образом систематизированные языковые единицы и информацию о них
- d) Грамматическую систему языка

31. Информационная система – это

- a) совокупность средств для поиска сведений из различных информационных источников
- b) хранилище информации, снабженное процедурами ввода, поиска, размещения и выдачи информации

- с)набор компьютерных программ, дающих возможность производить обработку различных данных
- d)электронная база данных

32.Электронный словарь – это

- a)компьютерная база данных, содержащая особым образом закодированные словарные статьи
- b)Совокупность текстов, имеющая лингвистическую разметку
- с)Справочник на бумажном носителе с различной информацией о слове
- d)словарь в компьютере или другом электронном устройстве, работающий благодаря комплексу сложнейших программ и алгоритмов

33.По признаку «характер носителя словарной информации» все словари делятся на

- a)бумажные
- b)электронные
- с)толковые
- d)фразеологические

34.Тезаурусы представляют собой

- a)разновидность словарей, включающих все словоупотребления в тексте с указанием на источник
- b)энциклопедические справочники мультимедийного типа
- с)лингвострановедческие словари языков
- d)особую разновидность словарей общей или специальной лексики, построенные по тематическому принципу (от понятия к слову)

35.К инструментальным средствам в рамках компьютерной лексикографии относятся

- a)базы данных
- b)компьютерные картотеки
- с)программы обработки текста
- d)бумажные аналоги словарей

36.Основными достоинствами электронных словарей для изучающих иностранные языки являются

- a)возможность активного использования при изучении иностранного языка
- b)скорость поиска информации
- с)возможности копирования информации в больших объемах
- d)удобная компактная форма

37.Недостатками электронных словарей являются

- a)ЭС требуют наличие сложных программ и компьютерных средств
- b)Информацию электронных словарей легко удалить и потерять
- с)Возможность размещения многих тысяч томов словарей на небольшом носителе
- d)Зависимость ЭС от заряда аккумулятора

38.Корпус языка – это

- a)совокупность текстов одного автора, систематизированная в хронологическом порядке
- b)совокупность текстов на одном языке в электронном варианте
- с)совокупность электронных данных, обработанная специальной компьютерной программой

d)совокупность текстов в электронном виде, снабженная лингвистической разметкой

39.Процесс или результат приписывания текстам и их компонентам специальных меток называется

- a)лингвистической разметкой
- b)машинным переводом
- c)моделированием общения
- d)автоматическим синтезом речи

40.Различают следующие виды лингвистических разметок

- a)логическая
- b)морфологическая
- c)синтаксическая
- d)семантическая

41.Применение национальных корпусов языков связано с тем, что

- a)корпус – богатый источник (база данных) для создания словарей
- b) на корпусе проверяются системы автоматической обработки текста и- различные лингвистические теории.
- c)корпуса появились относительно недавно
- d)корпуса представляют кодированную информацию, недоступную любому пользователю

42.Информационно-справочная система, основанная на собрании русских текстов в электронной форме с обязательной лингвистической разметкой представляет собой

- a)Национальный корпус немецкого языка
- b)Национальный корпус русского языка
- c)Британский Национальный Корпус
- d)Национальный корпус чешского языка

43.Машинный перевод – это

- a)процесс составления сложных компьютерных программ и алгоритмов для распознавания речи
- b)составление аннотации текста с помощью компьютерных программ
- c)выполняемое на компьютере действие по преобразованию текста на одном естественном языке в эквивалентный по содержанию текст на другом языке, а также результат такого действия
- d)выполняемое на компьютере действие по проверке правописания

44.Системы машинного перевода представляют собой

- a) Комплекс специальных сложнейших программ и алгоритмов, специальные автоматические словари входного и выходного языков
- b)Комплекс персональных ЭВМ, связанных по локальной сети
- c)Комплекс методов и приемов представления знаний в компьютерной среде
- d)Комплекс методов и приемов моделирования общения

45.Система машинного перевода включает в себя

- a)корпуса языков
- b)методы и приемы анализа
- c)двуязычные словари
- d)искусственные языки

46.Компьютерная (машинная) лексикография занимается

- а) решением проблем создания и функционирования электронных словарей разных типов
- б) решением вопросов функционирования большого массива текстов
- с) решением проблем автоматической обработки текста
- д) решением задач распознавания текста и моделирования языка

47. Первые программы компьютерного перевода появились

- а) после возникновения прикладной лингвистики
- б) после появления электронных словарей
- с) после изучения возможностей машинного перевода
- д) после возникновения компьютера

48. Широкое распространение компьютерного перевода в 80-х годах прошлого века было обусловлено

- а) его экономической выгодой
- б) достижениями прикладной лингвистики
- с) полным и всесторонним изучением возможностей машинного перевода
- д) созданием первых корпусов языков

49. Основной недостаток компьютерного перевода – это

- а) невозможность получить полноценный перевод
- б) дороговизна времени работы компьютера
- с) отсутствие альтернативного варианта перевода
- д) невозможность оперативно его использовать

50. Компьютер не может заменить переводчика

- а) при переводе текстов научного характера
- б) при переводе художественных текстов
- с) при бытовом общении
- д) при переводе публицистических текстов

51. Существуют два принципиально разных подхода к построению алгоритмов машинного перевода:

- а) основанный на лексико-семантической системе языка
- б) основанный на синтаксисе
- с) основанный на правилах
- д) основанный на статистике (статистический)

52. Наибольшие трудности для компьютерного перевода представляют

- а) синтаксические особенности языка
- б) лексические особенности языка
- с) семантические особенности языка
- д) стилистические особенности языка

53. Широкому распространению компьютерных двуязычных словарей способствовало

- а) совершенствование систем компьютерного перевода
- б) повсеместное изучение иностранных языков
- с) развитие Internet
- д) развитие социальных сетей

54. Алгоритм компьютерного перевода представляет собой

- а)последовательность действий по одновременной обработке данных нескольких языков
- б)действия по грамматической, фонетической и лексической обработке текста на одном языке
- с)действия по грамматической, фонетической и лексической обработке текста на двух (трех) языках
- д)последовательность однозначно и строго определенных действий над текстом для нахождения переводных соответствий в данной паре языков

55.При машинном переводе осуществляется следующая последовательность действий

- а)ввод текста и поиск входных словоформ
- б)перевод идиоматических словосочетаний
- с)окончательный грамматический анализ
- д)синтез выходных словоформ и предложения в целом на выходном языке

56.Качество машинного перевода зависит

- а)эффективности программного обеспечения
- б)возможностей самого компьютера
- с)объема словаря
- д)лексических и синтаксических соответствий языков

57.Первые программы компьютерного перевода появились

- а)после возникновения прикладной лингвистики
- б)после появления электронных словарей
- с)после изучения возможностей машинного перевода
- д)после возникновения компьютера

58.Использование электронных двуязычных словарей в помощь человеку-переводчику

- а)Считается машинным переводом
- б)Не считается машинным переводом
- с)Считается поверхностным(грубым) машинным переводом
- д)Не является перспективным видом деятельности

59.Различают следующие виды машинного перевода

- а)Информативный (поверхностный) перевод
- б)Профессиональный (автоматизированный) перевод
- с)Персональный перевод
- д)Сопутствующий перевод

60.Система машинного перевода включает в себя

- а)корпуса языков
- б)двуязычные словари
- с)методы и приемы анализа
- д)искусственные языки

61.Востребованность машинного перевода объясняется

- а)возросшим интересом общества к инновациям
- б)разработками в области компьютерной лексикографии

- с) ролью самой практики перевода в современном мире как необходимого условия обеспечения межкультурной коммуникации
- д) появлением разработок в области корпусной лингвистики

62. Лингвистический алгоритм – это

- а) Совокупность систематизированных сведений, фактов, событий в электронном варианте
- б) Последовательность действий при решении лингвистической задачи
- с) Программа, расширяющая возможности операционной системы (например, антивирусные программы, архиваторы и т.д.)
- д) Совокупность законов, методов и средств получения, хранения, передачи и преобразования информации о языке с помощью компьютера

5.7 Вопросы к зачету по дисциплине «Компьютерная лингвистика»

1. Основные понятия вычислительной (компьютерной) лингвистики.
2. Устройство систем автоматизированной обработки текстов.
3. Основные этапы построения. Графематический и морфологический анализ.
4. Методы задания синтаксической структуры предложений. Системы составляющих. Деревья подчинения. Проблемы синтаксического анализа.
5. Принципы построения синтаксических анализаторов. Проект "Автоматическая Обработка Текста" (АОТ).
6. Синтаксический парсер LINK. Применение морфологического и синтаксического анализа в поисковых системах.
7. Методы теоретического исследования семантики текстов. Лексические функции. Валентности слов. Теоретико-модельный подход.
8. Теоретико-множественные модели языка. Основные определения: отмеченные последовательности, контексты, дистрибутивные классы и др.
9. Формализация понятий «часть речи» и «синтаксический тип». Формализация понятия «грамматический род». Формализация «категории падежа».
10. Представление знаний для компьютерной обработки. Тезаурусы и онтологии.
11. Семантические сети. Фреймы. Формальные логические модели. Искусственные языки и нотации, применяемые в компьютерной лингвистике.
12. Корпусная лингвистика. Частотные методы в компьютерной лингвистике.
13. Корпусы языков. НКРЯ, OpenCorpora, структура, омонимия, примеры.
14. Теория речевых действий. Классификация речевых действий.
15. Модели и методы автоматической классификации и кластеризации текстовой информации. Иерархические и вероятностные подходы. Интеллектуальный анализ данных.
16. Автоматические системы извлечения информации.
17. Алгоритмические основы. Принципы обработки неструктурированной и плохо структурированной информации. Тематическая индексация текстов.
18. Формальные методы определения автора текста.
19. Авторский инвариант и лингвистические спектры. Применение методов кластеризации и классификации для установления авторства текстов.
20. Социальные сети. Направления исследований. Графовые модели анализа социальных сетей.
21. Методы обнаружения сообществ и анализ связанных подгрупп. Модели динамики сети.

22. Методы обнаружения спама: вероятностные и статистические, байесовский классификатор.