

Паспорт зачета

по дисциплине «Компьютерная лингвистика», 4 семестр

1. Методика оценки

Зачет проводится в устной форме, по билетам. Билет состоит из 2 вопросов и формируется по следующему правилу: вопросы выбираются из диапазона вопросов 1-28.

Таким образом, проверяется уровень сформированности компетенций и соотношенных с ними индикаторов, закрепленных за дисциплиной.

На зачете преподаватель вправе задавать студенту уточняющие и дополнительные вопросы из общего перечня (п. 4).

Форма билета для зачета

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Факультет АВТФ

Билет № _____

к зачету по дисциплине «Компьютерная лингвистика»

1. Вопрос 1.
2. Вопрос 2.

Утверждаю: зав. кафедрой ВТ _____ А.А. Якименко
(подпись)

(дата)

2. Уровни освоения компетенций и критерии оценки

Ответ на билет (тест) для зачета засчитывается на **продвинутом** уровне, если студент при ответе на вопросы проводит сравнительный комплексный анализ материала, выявляет проблемы, предлагает механизмы их решения, представляет количественные характеристики определенных процессов, приводит конкретные примеры, не допускает ошибок. Компетенции и соотношенные с ними индикаторы, закрепленные за дисциплиной, сформированы в полном объеме. Оценка составляет *от 18 до 20 баллов*.

Ответ на билет (тест) для зачета засчитывается на **базовом** уровне, если студент при ответе на вопросы формулирует основные понятия, дает характеристику процессов, явлений, проводит анализ причин, условий, способен представить количественные и качественные характеристики процессов. Компетенции и соотношенные с ними индикаторы, закрепленные за дисциплиной, содержат несущественные пробелы и сформированы на базовом уровне. Оценка составляет *от 14 до 17 баллов*.

Ответ на билет (тест) для зачета засчитывается на **пороговом** уровне, если студент

при ответе на вопросы дает определение основных понятий, может показать причинно-следственные связи явлений. Компетенции и соотнесенные с ними индикаторы, закрепленные за дисциплиной, содержат пробелы и сформированы на пороговом уровне. Оценка составляет *от 10 до 13 баллов*.

Ответ на билет (тест) для зачета считается **неудовлетворительным**, если студент при ответе на вопросы не дает определений основных понятий, не способен показать причинно-следственные связи явлений. Компетенции и соотнесенные с ними индикаторы, закрепленные за дисциплиной, не сформированы. Оценка составляет *менее 10 баллов*.

3. Шкала оценки

Зачет считается сданным, если сумма баллов по всем заданиям билета составляет от 10 до 20 баллов включительно. Сумма менее 10 баллов признается неудовлетворительным результатом промежуточной аттестации по дисциплине.

В общей оценке по дисциплине баллы за зачет учитываются в соответствии с правилами балльно-рейтинговой системы, установленными в НГТУ.

4. Вопросы к зачету по дисциплине «Компьютерная лингвистика»

1. Основные направления, понятия, задачи компьютерной лингвистики.
2. Принципы построения интеллектуальных систем автоматической обработки текстов. Основные компоненты и этапы их построения.
3. Морфологический анализ, этапы построения морфологических анализаторов (стемматизация, лемматизация, приписывание грамем, получение парадигм).
4. Основные типы морфологических анализаторов, алгоритмы их построения.
5. Синтаксический анализ, методы построения синтаксических анализаторов, построение деревьев составляющих и деревьев зависимостей.
6. Проблемы построения синтаксических анализаторов и методы их решения.
7. Реализация морфологического анализатора в поисковых системах, примеры систем.
8. Реализация синтаксического анализатора в поисковых системах, примеры систем.
9. Задача определения авторства текста и методы ее решения.
10. Задача определения эмоциональной окраски текста и методы ее решения.
11. Задача тематического моделирования и методы ее решения.
12. Основные программные средства (библиотеки и функции) языка R для решения задач анализа текстов.
13. Реализация в среде R частотного анализа текста, анализа n-грамм, примеры.
14. Реализация в среде R задачи определения настроения текста, примеры.
15. Реализация в среде R задачи тематического моделирования, примеры.
16. Закон Ципфа, закон Хипса, модель TF*IDF.
17. Классификация текстов на основе методов машинного обучения с учителем: метод наименьших квадратов (метод регрессионного анализа).
18. Классификация текстов на основе методов машинного обучения с учителем: метод ближайших соседей (метрический метод классификации).
19. Классификация текстов на основе методов машинного обучения с учителем: метод опорных векторов (линейный метод классификации).
20. Классификация текстов на основе методов машинного обучения с учителем: метод Байеса (вероятностный метод классификации).
21. Классификация текстов на основе методов машинного обучения с учителем: метод деревьев решений (логический метод классификации).

22. Оценка качества классификации текстов, основные меры и методы.
23. Кластеризация текстов на основе методов машинного обучения без учителя: иерархические методы.
24. Кластеризация текстов на основе методов машинного обучения без учителя: центроидные методы;
25. Кластеризация текстов на основе методов машинного обучения без учителя: вероятностные методы;
26. Кластеризация текстов на основе методов машинного обучения без учителя: методы на основе систем искусственного интеллекта.
27. Оценка качества кластеризации текстов, основные меры и методы.
28. Основные направления развития компьютерной лингвистики.