

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Степыкин Николай Иванович
Должность: Заведующий кафедрой
Дата подписания: 02.02.2023 16:05:15
Уникальный программный ключ:
79cb37fa15c029eb9fe555478f21c47b73e92308

МИНОБРНАУКИ РОССИИ

Юго-Западный государственный университет

УТВЕРЖДАЮ:

Заведующий кафедрой

теоретической и прикладной

лингвистики



Н.И. Степыкин

«11» февраля 2022 г.

ОЦЕНОЧНЫЕ СРЕДСТВА

для текущего контроля успеваемости

и промежуточной аттестации обучающихся

по дисциплине

Технологии корпусной лингвистики

45.03.03 Фундаментальная и прикладная лингвистика

**Юго-Западный государственный университет
Кафедра теоретической и прикладной лингвистики**

Вопросы для собеседования по дисциплине «Технологии корпусной лингвистики»

Раздел (тема) дисциплины «Введение в корпусную лингвистику. История корпусной лингвистики»

1. Британский национальный корпус.
2. История лингвистических корпусов: от картотеки к корпусу.
3. Классификация (типология) корпусов.
4. Корпусная лингвистика: современное состояние.
5. Основные понятия корпусной лингвистики.
6. Направления в лингвистике, предвосхитившие появление корпусной лингвистики: от картотеки к корпусу
7. История создания лингвистических корпусов.

Раздел (тема) дисциплины «Понятие разметки»

1. Основные характеристики корпусов.
2. Репрезентативность корпусов.
3. Классификация корпусов по различным основаниям.
4. Особые типы корпусов.
5. Понятие разметки.
6. Разметка. Средства разметки корпусов
7. Лингвистическая разметка.
8. Экстралингвистическая разметка.

Раздел (тема) дисциплины «Национальные корпуса»

1. Стандартизация в корпусной лингвистике.
2. Обзор существующих корпусов различных типов.
3. Зарубежные национальные корпуса.
4. Специальные корпуса.

Раздел (тема) дисциплины «Лингвистические исследования на базе корпуса»

1. Корпусные исследования.
2. Пользователи корпусов.
3. Способы использования корпусов.
4. Лексикографические исследования, основанные на корпусах.

Критерии оценки:

2 балла выставляется обучающемуся, если его ответ полный, содержательный, развернутый, нормативно оформленный, наполненный языковыми примерами.

1 балла выставляется обучающемуся, если его ответ полон содержательно, но не наполнен языковыми примерами.

Юго-Западный государственный университет
Кафедра теоретической и прикладной лингвистики
Вопросы для зачета по дисциплине «Технологии корпусной лингвистики»

1. История лингвистических корпусов: от картотеки к корпусу (18 баллов)
2. Классификация (типология корпусов), основания типологии (18 баллов)
3. Корпусная лингвистика: история возникновения и современное состояние.
Корпусная лингвистика в России (18 баллов)
4. Обзор существующих корпусов различных типов (18 баллов)
5. Корпус как поисковая система. Структура и характеристики корпусов (18 баллов)
6. Методология компиляции корпусов (18 баллов)
7. Корпусоподобные интерфейсы между лингвистом и поисковыми системами
Интернет (18 баллов)
8. Лингвистические исследования, базирующиеся на корпусах (18 баллов)
9. Британский национальный корпус (18 баллов)
10. Национальный корпус русского языка (18 баллов)
11. Проблемы репрезентативности корпусов (18 баллов)
12. Понятие разметки. Типы разметок. Понятие гранулярности (18 баллов)
13. Формирование корпуса текстов (18 баллов)
14. Круг потенциальных пользователей лингвистических корпусов (18 баллов)
15. Получение конкордансов и словников (18 баллов)
16. Корпусная лингвистика и компьютерная лингвистика (18 баллов)
17. Требования к созданию корпуса (18 баллов)
18. Типы аннотирования в национальном корпусе (18 баллов)
19. Метаразметка как часть поискового аппарата корпуса (18 баллов)
20. Лингвистическое аннотирование в корпусе (18 баллов)
21. Выравнивание в параллельных корпусах (18 баллов)
22. Различные технологии разметки (18 баллов)

Критерии оценки:

18 баллов выставляется обучающемуся, если его ответ полный, содержательный, развернутый, нормативно оформленный, наполненный языковыми примерами.

12 баллов выставляется обучающемуся, если ответ содержательно соответствует теме, грамотно оформлен, но не достаточно развернут и наполнен языковыми примерами.

8 баллов выставляется обучающемуся, если его ответ полон содержательно, но не наполнен языковыми примерами.

4 балла выставляется обучающемуся, есть его ответ содержательно верен, но не наполнен языковыми примерами и по своему оформлению не соответствует языковым нормам.

Юго-Западный государственный университет
Кафедра теоретической и прикладной лингвистики
Комплект заданий для аудиторных контрольных работ по дисциплине
«Технологии корпусной лингвистики»

Раздел (тема) дисциплины «Понятие лингвистического корпуса»
Аудиторная контрольная работа № 1

1. Дайте определение терминам:
 - корпусная лингвистика
 - корпусный менеджер
 - конкорданс

2. Проблемы традиционного сбора и обработки информации ...
 - А) занимают длительное время
 - Б) работа выполняется большим количеством людей
 - В) информация труднодоступна исследователям из других стран

3. Корпусная лингвистика занимается ...
 - А) разработкой программ для создания корпусов
 - Б) разработкой принципов построения и использования корпуса данных
 - В) рассмотрением истории и опыта создания корпусов

4. Основные черты современного корпуса текстов ...
 - А) машиночитаемый вид, репрезентативность, простота
 - Б) унифицированность, машиночитаемый вид, репрезентативность
 - В) репрезентативность, машиночитаемый вид, статичность

5. Соотнесите объект и предмет корпусной лингвистики.
 1. Объект
 2. Предмет
 - А) теория и практика использования массивов данных
 - Б) корпусы текстов
 - В) составление конкордансов

6. После создания первых электронных корпусов ученые поставили целью разработку нового типа корпуса: _____.

7. Подъем современной корпусной лингвистики связан с развитием технологий ...
 - А) сравнительно-исторического языкознания
 - Б) морфологии
 - В) педагогики
 - Г) истории
 - Д) стилистики
 - Е) лексикографии

8. Составителем одного из наиболее полных библейских конкордансов является:
 - А) Отто Есперсен
 - Б) Александр Круден
 - В) Рэндольф Квирк

9. Первый корпус компьютеризированных текстов - _____
Его объем: _____
Количество текстов: _____

10. После возникновения первого корпуса текстов обсуждалось ...
 - А) решаемые корпусом задачи

- Б) процедура обработки данных
- В) принципы отбора данных

11. Соотнесите ученых, проекты и нововведение.

I. Чарльз Фриз	1. Международный компьютерный архив современного английского языка (ICAME) (1977)	а. Учет авторских прав при составлении корпуса
II. Рэндольф Квирк	2. Структура английского языка (The structure of English) (1952)	б. Начало создания доэлектронного корпуса
III. Джеффри Лич, Нельсон Фрэнсис, Ян Свартвик	3. «Обзор употребления <i>английского языка</i> » (The Survey of English Usage) (1959)	в. Первое использование разговорных материалов как источника данных

12. Как отдельное направление науки о языке корпусная лингвистика оформилась с _____ годов.

Раздел (тема) дисциплины «Создание корпусов» Аудиторная контрольная работа № 2

1. Дайте определение понятию «репрезентативность».
2. Какой корпус не соответствует критерию репрезентативности?
 - А. Брауновский корпус
 - Б. НКРЯ
 - В. Корпус современного американского английского
3. По типу объекта корпуса могут быть:
 - А) отдельного автора
 - Б) универсальные
 - В) для отдельной цели
4. Для соответствия критерию репрезентативности объем современного национального корпуса текстов должен быть не менее _____ словоупотреблений?
5. Методология построения универсального корпуса основана на принципе:
 - А. индукции
 - Б. дедукции
6. По типу языковых данных корпуса бывают:
 - А. устные
 - Б. письменные
 - В. разговорные
 - Г. смешанные
 - Д. литературные
 - Е. речевые
7. Корпус LOB относится к корпусам:
 - А. одноязычным
 - Б. многоязычным
 - В. двуязычным
8. По характеру разметки корпуса могут быть:
 - А. морфологическими и просодическими
 - Б. синтаксическими и мониторными
 - В. риторическими и синтаксическими
 - Г. семантическими и публицистическими
9. Характеристика НКРЯ:
 - А. смешанный, многоязычный, многоцелевой, исследовательский, динамический

- Б. письменный, многоязычный, многоцелевой, исследовательский, динамический
- В. письменный, многоязычный, многоцелевой, иллюстративный, динамический
- Г. смешанный, одноязычный, многоцелевой, исследовательский, динамический
- Д. смешанный, многоязычный, многоцелевой, исследовательский, статический

10. Пропорция стоимости отбора слов из устных источников к стоимости отбора слов из письменных источников в 1990-е годы

- А. 10:60 Б. 10:100 В. 10:40 Г. 10:50 Д. 10:80

11. «Корпус устной речи американского варианта английского языка» включал записи ...

- А. записи фильмов
- Б. записи университетских лекций
- В. записи психоаналитических сеансов
- Г. записи для аудиокурсов
- Д. записи бытовых диалогов

12. Для установления соответствий между текстом оригинала и его переводами используют метод:

- А. картирования
- Б. языка-посредника
- В. автоматического выравнивания

Раздел (тема) дисциплины «Использование корпусов»

Аудиторная контрольная работа № 3

1. Подготовка «технологического» описания корпуса осуществляется на этапе ...

- А. поступления текстов
- Б. преобразования в машиночитаемую форму
- В. анализа и предварительной обработки текстов
- Г. конвертирования и графематического анализа
- Д. разметки текстов

2. Объем создаваемого корпуса текстов в принятых единицах зависит от

- А. конечного пользователя
- Б. количества жанров, представленных в нем
- В. целей создания

3. При изучении частоты употребления буквосочетаний и звуков единиц корпуса текстов будет ...

- А. буква
- Б. лемма
- В. текст

4. Дайте определение *токенизации*.

5. Осуществите стемминг предложения “A rising, indignant feeling seized him”.

6. Осуществите лемматизацию предложения “The worst mistake is giving up on your dreams”.

7. Проведите парсинг предложения «*Фонетический барьер в вербальной коммуникации обусловлен особенностями произношения слов деловыми партнерами*».

8. В результате работы программ автоматического синтаксического анализа фиксируются...
- А. связи между словами и словосочетаниями
 - Б. грамматические характеристики
 - В. структура лексем
9. Проблема морфологической неоднозначности возникает при анализе слова ...
- А. дом
 - Б. печь
 - В. броня
10. Просодическая разметка показывает ...
- А. оговорки
 - Б. паузы и повторы
 - В. ударение и интонацию
11. Лексико-семантические теги для прилагательных сгруппированы по...
- А. мереологии
 - Б. таксономии
 - В. служебному статусу
12. Предметная область текста относится к ... факторам разметки текста.
- А. внешним
 - Б. смешанным
 - В. внутренним

Раздел (тема) дисциплины «Лингвистические исследования на базе корпуса»

Аудиторная контрольная работа № 4

1. Функции корпусного менеджера
- А. построение KWIC и конкордансов
 - Б. исправление орфографических ошибок в тексте запроса
 - В. транслитерация букв с одного языка на другой
 - Г. поддержка различных форматов данных
2. Модель информационного запроса включает в себя
- А. операторы конъюнкции
 - Б. требование на сортировку результатов
 - В. оператор трансформации
 - Г. шаблоны
3. При отрицательном фильтре запроса
- А. совпадающие с запросом строки выдаются в конкордансном списке
 - Б. совпадающие с запросом строки удаляются из конкордансного списка
 - В. удовлетворяющие запросу позиции в конкордансе выделяются цветом
4. Какие результаты выдадутся для сложного запроса !regular verb: snow ?
5. Как называется явление, когда строка или часть строки успешно сопоставилась с выражением?
- А. регулярное выражение
 - Б. совпадение
 - В. положительный фильтр
6. К управляющим символам относятся...
- А. «»
 - Б. /
 - В. *
 - Г. ?

7. Запишите запрос для результатов «рок, рак», используя простые символы.
8. Какие результаты можно получить для запроса «13+4»?
- А. 14
 - Б. 134
 - В. 1333334
 - Г. 431
9. Какой простой символ является оператором альтернативы?
- А. *
 - Б. ||
 - В. |
 - Г. ^
10. Какой корпус обновляется 2 раза в год?
- А. Британский национальный корпус
 - Б. Национальный корпус русского языка
 - В. Корпус современного американского английского
11. Упсальский машинный фонд русского языка был создан в ...
- А. России
 - Б. Швеции
 - В. Швейцарии
 - Г. Норвегии
12. «Звуковой» корпус современного русского языка повседневного общения называется ...
- А. Один речевой день
 - Б. Один разговорный день
 - В. Один языковой день
 - Г. Один звуковой день

Критерии оценки:

7-10 баллов выставляется обучающемуся, если доля правильных ответов составляет 70-100%;

4-6 баллов выставляется обучающемуся, если доля правильных ответов составляет 31-69%;

1-3 балла выставляется обучающемуся, если доля правильных ответов составляет 10-30%.