

Министерство науки и высшего образования Российской Федерации  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Филологический факультет

УТВЕРЖДЕНО:  
Декан филологического факультета  
И.В. Тубалова

Оценочные материалы по дисциплине

**Лингвистические базы данных**

по направлению подготовки

**45.03.03 Фундаментальная и прикладная лингвистика**

Направленность (профиль) подготовки:  
**«Фундаментальная и прикладная лингвистика»**

Форма обучения  
**Очная**

Квалификация  
**Бакалавр**

Год приема  
**2024**

СОГЛАСОВАНО:  
Руководитель ОП  
А.В. Васильева

Председатель УМК  
Ю.А. Тихомирова

## **1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами**

Целью дисциплины «Лингвистические базы данных» является:

- 1) привить навыки проектирования и использования реляционных баз данных;
- 2) сформировать навыки проектирования и разработки прикладных проектов с использованием современных СУБД

В результате курса формируются следующие компетенции:

ПК-4. Способен разрабатывать программный код при решении задач автоматической обработки текстов

ОПК-7. Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-7.2. Использует информационные ресурсы при решении стандартных задач профессиональной деятельности

ИОПК-7.3. Применяет современные технические средства и информационно-коммуникационные технологии для решения задач профессиональной деятельности

ИПК-4.1. Применяет способы формализации и алгоритмизации поставленных задач в сфере автоматической обработки текстов

## **2. Оценочные материалы текущего контроля и критерии оценивания**

К числу форм контроля, оценивающих уровень достижения компетенций по текущей дисциплине, относятся тесты, устные вопросы, выполнение практических заданий (ПК-4, ПК-4.1), итоговый проект (продукт (ПК-4, ПК-4.1, ОПК-7, ИОПК-7.2, ИОПК-7.3), код (ИПК-4.2), контрольная и самостоятельная работы (ИПК-4.1).

Итоговый проект включает блок вопросов по теоретической части изучаемой дисциплины (ПК-4, ПК-4.1, ОПК-7, ИОПК-7.2):

Пример практических задач:

1. Напишите цель и этапы реализации концепции своей базы данных, связанной с лингвистикой или экспериментальной лингвистикой.
2. Постройте концепцию ограничений и структуры своей базы данных
3. Постройте ег-диаграмму своей базы данных.
4. Создайте базу данных в MS SQL или PostgreSQL.
5. Наполните базу данных.
6. Выполните унарные операции.
7. Выполните бинарные операции.
8. Настройте уровни доступа к базе данных

Ответы:

Теоретические вопросы (ПК-4, ИПК-4.2):

Критерии оценивания ответа на теоретический вопрос

Оценка	Критерии
	<ol style="list-style-type: none"><li>1. Понимание и логика высказывания изученного материала</li><li>2. Представление взаимосвязей процесса и взаимосвязи теоретических модулей изучаемого предмета</li><li>3. Полнота данных ответов;</li></ol>

	<p>4. Аргументированность данных ответов;</p> <p>5. Правильность ответов на вопросы;</p>
«зачтено»	<p>Полно и аргументировано даны ответы по содержанию задания. Обнаружено понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только по учебнику, но и самостоятельно составленные. Изложение материала последовательно и правильно</p> <p>Ответы обучающегося удовлетворяют тем же требованиям, что и для оценки «отлично», но допускается 1-2 ошибки, которые сам же исправляет.</p> <p>Обучающийся демонстрирует знание и понимание основных положений данного задания, но: 1) излагает материал неполно и допускает неточности в определении понятий или формулировке правил; 2) не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; 3) излагает материал непоследовательно и допускает ошибки.</p>
«не зачтено»	<p>Демонстрация незнания ответа на соответствующее задание, допускаются ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагается материал; отмечаются такие недостатки в подготовке, которые являются серьезным препятствием к успешному овладению последующим материалом.</p>

Критерии оценивания практической работы:

Оценка	Критерии
	<ol style="list-style-type: none"> <li>1. Понимание и логика алгоритма работы</li> <li>2. Наличие или отсудив ошибок в коде</li> <li>3. Полнота решения практических задач</li> <li>4. Своевременность выполнения;</li> <li>5. Умения связать практический материал с теоретическим;</li> <li>6. Понимание базовых формул обработки естественного языка и программирования;</li> </ol>
«зачтено»	<p>Основные требования к решению практических задач выполнены. Продемонстрированы умение анализировать алгоритмы и находить оптимальное количество решений, умение работать с информацией, в том числе умение затребовать дополнительную информацию, необходимую для уточнения реализации алгоритма, навыки разработки программного кода;</p> <p>Основные требования к решению практических задач выполнены, но при этом допущены недочеты. В частности, недостаточно раскрыты навыки стиля, недостаточно комментариев</p> <p>Имеются существенные отступления от решения практических задач. В частности отсутствуют навык и умения моделировать решения в соответствии с заданием, представлять различные подходы к разработке алгоритмов, ориентированных на конечный результат</p>
«не зачтено»	<p>Задача не решена, обнаруживается существенное непонимание проблемы</p>

Итоговый проект (ПК-4, ПК-4.1, ОПК-7, ИОПК-7.2, ИОПК-7.3)

Итоговый проект включает создание базы данных по выбранной области лингвистики. К числу тем может относиться проект экспериментальной лингвистики, словари, языковые модели, тезаурусы, требующие разработки информационных систем. Проект должен быть выполнен в виде приложения с элементами управления во фронтенде или языке программирования R или Python.

Выбор тем текстов предусматривает возможность выбора индивидуальной образовательной траектории, связывающую проектную деятельность со смежными гуманитарными дисциплинами. Однако существуют ограничительные критерии реализации проекта, направленные на формирование вышеуказанных компетенций: 1. Сбор и структуризация данных. 2. Наличие двух классов в обучающем датасете, наличие числовых переменных. 3. Объем материала не менее 2 сущностей и 4 атрибутов.

Сбор и структуризация данных предусматривает свободу выбора темы. В случае, если выбор не был осуществлен, то предлагаются следующие рубрики:

1. Новостные сайты и пресс релизы (классификация и генерация текстов по рубрике и/или информационному агентству)
2. Отзывы к товарам или фильмам (Наличие 3 класса: позитивный, негативный, нейтральный отзывы)
3. Комментарии в социальных сетях
4. Стены сообществ «ВКонтакте» (классификация по рубриками)

Критерии оценивания практической работы:

Оценка	Критерии
	<ol style="list-style-type: none"><li>1. Сформулирована гипотеза проекта</li><li>2. Собран и структурирован датасет</li><li>3. Четкая логика реализации алгоритмов обработки естественного языка в коде;</li><li>4. Правильность ответов на вопросы;</li><li>5. Наличие структурированной презентации</li><li>6. Полнота проекта</li></ol>

«зачтено»	<p>Выполнены все требования к проекту: сформулирована гипотеза, создан прасер сайтов для сбора и структуризации информации, написан код для обработки естественного языка</p> <p>Выполнены все требования к составлению презентаций: дизайн слайдов, логика изложения материала, текст хорошо написан и сформированные идеи ясно изложены и структурированы</p> <p>Существуют незначительные ошибки в проекте, не влияющие на конечный результат. В частности, может быть низкий уровень формальных метрик, неточности в визуализации данных</p> <p>Основные требования к презентациям выполнены, но при этом допущены недочеты. В частности, имеются неточности в изложении материала; отсутствует логическая последовательность в суждениях; не выдержан объем презентации</p> <p>Существенные ошибки в коде, отсекает логика представления проекта, недостаточный объем датасета</p> <p>Имеются существенные отступления от требований к презентациям. В частности: тема освещена лишь частично; допущены фактические ошибки в содержании презентаций или при ответе на дополнительные вопросы.</p>
«не зачтено»	<p>Критические ошибки в коде, гипотеза не подвержена, недостаточное количество или отсутствие обучающей выборки</p> <p>Тема презентации не раскрыта, обнаруживается существенное непонимание проблемы</p>

### 3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

**Зачет в пятом семестре** состоит из трех частей.

**Первая часть** представляет собой тест из 20 вопросов, проверяющих ПК-4, ИПК-4.1, ОПК-7.2

Ответы на вопросы первой части даются путем выбора из списка предложенных, установления соответствия между объектами или формулирования однословного ответа на вопрос открытого типа.

Примерный тест

#### 1. Базы данных SQL

1. Какой язык используется для работы с базами данных SQL?

- A. - Java
- B. - SQL+
- C. - Python
- D. - HTML

2. Какая команда используется для создания новой таблицы в базе данных SQL?

- A. INSERT
- B. DELETE
- C. CREATE TABLE+

- D. - UPDATE
- SQL? 3. Какая функция используется для выборки данных из таблицы в базе данных
- A. - ADD  
B. - SELECT+  
C. - MODIFY  
D. - REMOVE

- SQL? 4. Какой оператор используется для обновления данных в таблице базы данных
- A. INSERT  
B. DELETE  
C. CREATE  
D. UPDATE+
5. Какая команда используется для удаления таблицы из базы данных SQL?
- A. SELECT  
B. DROP TABLE+  
C. ALTER TABLE  
D. - DELETE TABLE

## 2. Корпусная лингвистика

1. Что такое корпус в лингвистике?
- A. - Набор текстов, используемый для исследования языка+  
B. - Языковая единица, состоящая из звуков  
C. - Коллекция грамматических правил  
D. - Метод анализа лексической семантики
2. Какие типы корпусов существуют?
- A. - Социолингвистические и прагматические  
B. - Академические и профессиональные  
C. - Синхронные и диахронные+  
D. - Структурные и семантические
3. Что такое лемматизация в корпусной лингвистике?
- A. - Анализ грамматической структуры предложений  
B. - Изучение синтаксических отношений между словами  
C. - Приведение слова к его базовой форме+  
D. - Извлечение ключевых слов из текста
4. Какая метрика используется для измерения коллокаций в корпусной лингвистике?
- A. - Частотность  
B. - Средняя длина предложений  
C. - PMI (Pointwise Mutual Information)+  
D. - Коэффициент корреляции
5. Что такое конкорданс в корпусной лингвистике?
- A. - Список с контекстами, в которых встречается заданное слово+  
B. - Группа слов, образующая лексическую единицу  
C. - Метод анализа семантической роли слов в предложении  
D. - Статистическая мера сходства между текстами

### Критерии оценивания тестирования:

Оценка	Критерии
	1. Полнота выполнения тестовых заданий; 2. Своевременность выполнения; 3. Правильность ответов на вопросы; 4. Самостоятельность тестирования

«зачтено»	<p>Выполнено более 85 % заданий предложенного теста, в заданиях открытого типа дан полный, развернутый ответ на поставленный вопрос</p> <p>Выполнено более 70 % заданий предложенного теста, в заданиях открытого типа дан полный, развернутый ответ на поставленный вопрос; однако были допущены неточности в определении понятий, терминов и др.</p> <p>Выполнено более 54 % заданий предложенного теста, в заданиях открытого типа дан неполный ответ на поставленный вопрос, в ответе не присутствуют доказательные примеры, текст со стилистическими и орфографическими ошибками.</p>
«не зачтено»	<p>Выполнено не более 53 % заданий предложенного теста, на поставленные вопросы ответ отсутствует или неполный, допущены существенные ошибки в теоретическом материале (терминах, понятиях).</p>

#### 4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

Первая часть. Решение тестовых задач:

##### Тест №1

1. Какой язык используется для работы с базами данных SQL?
  - A. - Java
  - B. - SQL+
  - C. - Python
  - D. - HTML
2. Какая команда используется для создания новой таблицы в базе данных SQL?
  - A. INSERT
  - B. DELETE
  - C. CREATE TABLE+
  - D. - UPDATE
3. Какая функция используется для выборки данных из таблицы в базе данных SQL?
  - A. - ADD
  - B. - SELECT+
  - C. - MODIFY
  - D. - REMOVE
4. Какой оператор используется для обновления данных в таблице базы данных SQL?
  - A. INSERT
  - B. DELETE
  - C. CREATE
  - D. UPDATE+
5. Какая команда используется для удаления таблицы из базы данных SQL?
  - A. SELECT
  - B. DROP TABLE+
  - C. ALTER TABLE
  - D. - DELETE TABLE
6. Какой оператор используется для добавления новой записи в таблицу базы данных SQL?
  - A. INSERT INTO+

- B. UPDATE
  - C. DELETE FROM
  - D. CREATE
7. Какая функция используется для подсчета количества записей в таблице базы данных SQL?
- a. COUNT
  - b. SUM
  - c. AVG
  - d. SELECT COUNT(\*) FROM+
8. Какая команда используется для изменения структуры таблицы в базе данных SQL?
- A. - INSERT
  - B. - DELETE
  - C. - ALTER TABLE+
  - D. - UPDATE
9. Какой оператор используется для установки условия в SQL запросе?
- a. ORDER BY
  - b. WHERE+
  - c. GROUP BY
  - B. - HAVING
10. Какая функция используется для нахождения максимального значения в столбце базы данных SQL?
- A. - MIN
  - B. - AVG
  - C. - MAX+
  - D. - SUM
11. Какая команда используется для удаления данных из таблицы базы данных SQL?
- A. REMOVE
  - B. ERASE
  - C. DROP
  - D. x DELETE FROM+
12. Какой оператор используется для объединения данных из нескольких таблиц в базе данных SQL?
- A. x JOIN+
  - B. MERGE
  - C. UNION
  - D. GROUP BY
13. Какая функция используется для нахождения среднего значения в столбце базы данных SQL?
- A. - MAX
  - B. - MIN
  - C. - x AVG+
  - D. - SUM
14. Какой оператор используется для сортировки данных в базе данных SQL?
- A. SELECT
  - B. ORDER
  - C. SORT
  - D. x ORDER BY+
15. Какая команда используется для изменения данных в таблице базы данных SQL?
- A. UPDATE VALUES
  - B. MODIFY
  - C. UPDATE SET+

- D. - ALTER
16. Какой оператор используется для ограничения числа возвращаемых строк в базе данных SQL?
- A. - LIMIT
  - B. - TOP+
  - C. - FIRST
  - D. - ROWS
17. Какая функция используется для нахождения суммы значений в столбце базы данных SQL?
- A. MAX
  - B. MIN
  - C. AVG
  - D. SUM+
18. Какой оператор используется для добавления нового столбца в таблицу базы данных SQL?
- A. INSERT COLUMN
  - B. ADD COLUMN
  - C. CREATE COLUMN
  - D. x ALTER TABLE ADD COLUMN+
19. Какая команда используется для создания индекса в базе данных SQL?
- A. - CREATE INDEX
  - B. - ADD INDEX
  - C. - MAKE INDEX
  - D. - x CREATE INDEX+
20. Какой оператор используется для удаления столбца из таблицы базы данных SQL?
- A. REMOVE COLUMN
  - B. - DELETE COLUMN
  - C. - DROP COLUMN
  - D. - x ALTER TABLE DROP COLUMN+

Вторая часть. Решение практических задач:

Задача 1

Дан корпус текстов в формате csv

При помощи языка программирования R или Python наполните базу данных

Задача 2

– Изучите код, представленный на практике (определение количества):

```
SELECT
```

```
*
```

```
FROM
```

```
[Рецепты] P
```

```
SELECT
```

```
[Всего] = COUNT(*),
```

```
[Всего с авторами] = COUNT([ID_Автора]),
```

```
[Задействованных категорий] = COUNT(DISTINCT [Название категории])
```

```
FROM
```

```
[Рецепты] P
```

Задача 3. Создайте аналогичный для своей базы данных

Изучите код, представленный на практике (определение количества):

```
SELECT
```

```

*
FROM
  [Автор рецепта] AP
WHERE
  AP.[Дата рождения] > '1959-01-01'

```

```
GO
```

```

SELECT
*
FROM
  [Автор рецепта] AP
WHERE
  AP.[Отчество] <> NULL

```

```
GO
```

```

SELECT
*
FROM
  [Автор рецепта] AP
WHERE
  AP.[Отчество] LIKE 'Иван%'

```

```
GO
```

```

SELECT
*
FROM
  [Автор рецепта] AP
WHERE
  AP.[Фамилия] LIKE (AP.[Имя] + '%')

```

```
GO
```

```

SELECT
*
FROM
  [Автор рецепта] AP
WHERE
  ( AP.[Отчество] LIKE '%вич' AND AP.[Пол] = 'Ж' )
  OR
  ( AP.[Отчество] LIKE '%вна' AND AP.[Пол] = 'М' )

```

Задача 4. Создайте аналогичный для своей базы данных

Задача 5. Постройте бинарные и унарные запросы по примеру запроса SQL (объединение сущностей, проекция)

```

SELECT dbo.Lemma.lemtype, dbo.Lexeme.lex, dbo.Lemma.suffix, dbo.Lemma.tag,
dbo.Lemma.descr

```

```

FROM dbo.Lemma
JOIN dbo.Lexeme ON dbo.Lexeme.lexid = Lemma.lexid
WHERE dbo.Lemma.lem = 'домик'

```

Задача 5. Создайте проекцию, где отчество респондента заканчивается на «вич»

```

select * from dbo.[Автор рецепта]

```

```

update [Автор рецепта]
set [Пол] = 'М'
--where [Отчество] like '%вич'
where [ID_Автора] = 15
go

```

Задача 6. При помощи команды concat создайте сущность, которая формирует объект ФИО из атрибутов респондента: фамилия, имя, отчество. Результат должен содержать Фамилию и первые буквы имени и отчества, разделенных точками

```

SELECT
  P.[Название рецепта],
  [ФИО]=CONCAT(
    AP.[Фамилия], ' ',
    LEFT(AP.[Имя],1), ' ',
    '+'LEFT(AP.[Отчество], 1)+'.'
  )

```

Задача 7. Создайте объект, добавляющий файлы

```

INSERT
  INTO Литература ([Тип научного издания], Название_Экземпляра, [ID_Автора
экземпляра], [Название университета], [Год издания], Вложение, [Ключевые слова]
)

```

```

SELECT 'Статья', 'Эмоционально-психологические характеристики человека в
зеркале собственно образных слов русского и английского языков', 35, 'Национальный
исследовательский Томский политехнический университет', ' ', BulkColumn, 'собственно
образное слово, лексико-семантическая группа, национально-культурная специфика'

```

```

FROM OpenRowSet
(
  BULK N'C:\Users\Домашний\Desktop\БД\статьи\184.pdf',
  SINGLE_BLOB
) AS Файл

```

```

UNION ALL
FROM OpenRowSet
(
  BULK N'C:\Users\Домашний\Desktop\БД\статьи\192-197.pdf',
  SINGLE_BLOB
) AS Файл

```

Задача 8. При помощи языка программирования Python создайте соединение с базой данных

```

conn = pyodbc.connect('Driver={SQL Server};'
                      'Server=HOME\MSSQLSERVER01;'
                      'Database=Cookbook;'
                      'Trusted_Connection=yes;')

```

```

cursor = conn.cursor()
if conn:
  print ("Connected Successfully")

```

```

else:
  print ("Connection Not Established")

```

Задача 9. Выполните запросы к базе в скрипте python

```

result = conn.cursor()
cursor.execute("""select * from dbo.[Автор рецепта]""")
for row in cursor:
  print(row)

```

Задача 10. Сохраните полученные результаты запроса в виде таблицы.

Задача 11. Выполните запрос SQL из любого офисного приложения. Например, MS SQL. полученный результат сохраните в таблице.

### **Информация о разработчиках**

Степаненко Андрей Александрович, старший преподаватель кафедры общей, компьютерной и когнитивной лингвистики