

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

Филологический факультет

УТВЕРЖДЕНО:

Декан

И. В.Тубалова

Рабочая программа дисциплины

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

Томск – 2024

1. Цель и планируемые результаты освоения дисциплины (модуля)

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. Задачи освоения дисциплины

Задача дисциплины направлена на изучение проектирования и использования баз данных, формирование навыков проектирования и разработки прикладных лингвистических проектов с использованием современных СУБД, систематизирует знания о способах анализа, верификации и оценки полноты информации в ходе профессиональной деятельности

- Проектирование и разработка реляционных баз данных
- Создание и анализ ег-диаграмм
- Создание SQL-запросов: создание, изменение данных, манипуляция данными
- Разработка лингвистических приложений, реализуемых в СУБД MS SQL и/или PostgreSQL
- Бекенд: создание запросов, вывод данных через html, javascript, flask

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Семестр 6, зачет.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Введение в языкознание», «Общая фонетика», «Общая морфология», «Общий синтаксис», «Общая семантика», «Информационные технологии и

основы информационной культуры в лингвистике», «Информатика и основы программирования», «Квантитативные методы лингвистики», «Вероятностные модели».

6. Язык реализации

Русский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 2 з.е., 72 часа, из которых:

- лекции: 12 ч.;
- семинарские занятия: 0 ч.
- практические занятия: 22 ч.;
- лабораторные работы: 0 ч.

в том числе практическая подготовка: 22 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. Базы данных и их классификация

Тема 2. Основные понятия, связанные с лингвистическими информационными ресурсами. Классификация СБД и ИС.

Тема 3. Общие сведения о реляционной модели данных (РМД)

Тема 4. Структурная и целостная части РМД

Тема 5. Манипуляционная часть РМД

Тема 6. Язык структурированных запросов (SQL). DDL, DML

Тема 7. Оптимизация плана выполнения запросов. Индексирование данных

Тема 8. Проектирование БД. Обзор нотаций описания БД. CASE системы

Тема 9. Разработка хранимых функций, процедур, триггеров

Тема 10. Сравнение технологий доступа к данным.

Тема 11. Технологии клиент-сервер. Понятия тонкого и толстого клиентов.

9. Текущий контроль по дисциплине

Текущий контроль образовательной программы (темы, раздела, модуля) требованиям образовательных стандартов по направлениям подготовки/специальностям. Текущий контроль успеваемости обучающихся направлен на определение соответствия результатов обучения после освоения элемента по дисциплине проводится путем контроля посещаемости, проведения контрольных работ, тестов по лекционному материалу, разработки кода, выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр. Примерные задания текущего контроля:

– Продумайте концепцию вашей базы данных, это может быть словарь, корпус, база респондентов и т.п.

– Создайте ER-диаграмму вашей базы

– При помощи СУБД создайте базу данных:

```
CREATE DATABASE [Cook_demo]
```

```
GO
```

```
ALTER DATABASE [Cook_demo]
```

```
SET
```

```
ANSI_NULL_DEFAULT OFF,
```

```
ANSI_NULLS OFF,
```

```
ANSI_PADDING OFF,
```

```
ANSI_WARNINGS OFF,
```

```
ARITHABORT OFF,
```

```
AUTO_CLOSE OFF,
```

```

    AUTO_CREATE_STATISTICS ON,
    AUTO_SHRINK OFF,
    AUTO_UPDATE_STATISTICS ON,
    AUTO_UPDATE_STATISTICS_ASYNC OFF,
    COMPATIBILITY_LEVEL = 130,
    CONCAT_NULL_YIELDS_NULL OFF,
    CURSOR_CLOSE_ON_COMMIT OFF,
    CURSOR_DEFAULT GLOBAL,
    DATE_CORRELATION_OPTIMIZATION OFF,
    DB_CHAINING OFF,
    HONOR_BROKER_PRIORITY OFF,
    MULTI_USER,
    NESTED_TRIGGERS = ON,
    NUMERIC_ROUNDABORT OFF,
    PAGE_VERIFY CHECKSUM,
    PARAMETERIZATION SIMPLE,
    QUOTED_IDENTIFIER OFF,
    READ_COMMITTED_SNAPSHOT OFF,
    RECOVERY SIMPLE,
    RECURSIVE_TRIGGERS OFF,
    TRANSFORM_NOISE_WORDS = OFF,
    TRUSTWORTHY OFF
    WITH ROLLBACK IMMEDIATE
GO

```

```

ALTER DATABASE [Cook_demo]
    COLLATE Cyrillic_General_CI_AS
GO

```

```

ALTER DATABASE [Cook_demo]
    SET DISABLE_BROKER
GO

```

Создайте свою базу данных

– Заполните существующую базу и добавьте свои данные:

```

CREATE TABLE [dbo].[Автор рецепта] (
    [ID_Автора] [int] IDENTITY,
    [Фамилия] [varchar](50) NOT NULL,
    [Имя] [varchar](50) NOT NULL,
    [Отчество] [varchar](50) NULL,
    [Пол] [char](1) NOT NULL,
    [Дата рождения] [date] NULL,
    [ФИО] AS (
        CONCAT(
            [Фамилия], ' ',
            LEFT([Имя],1), ' ',
            ''+LEFT([Отчество],1)+' '
        )
    ),
    CONSTRAINT [PK_Автор рецепта] PRIMARY KEY CLUSTERED ([ID_Автора])
WITH (FILLFACTOR = 100),
    CONSTRAINT [CKC_Автор рецепта - Дата рождения] CHECK(
        [Дата рождения]>='1700.01.01' AND [Дата рождения]<=GetDate()

```

```

),
CONSTRAINT [CKC_Автор рецепта - Пол] CHECK (
    [Пол] IN ('Ж', 'М')
)
)
ON [PRIMARY]

```

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет проводится в письменной и устной форме по выбранному проекту. Проект предполагает логическое изложение теоретического блока с привязкой к практической деятельности и проверяет уровень сформированности следующих компетенций: ПК-4, ОПК-7, ИОПК-7.2, ИОПК-7.3, ИПК-4.1.

Примерный перечень теоретических вопросов

1. Понятие информационной системы, БД и их классификация.
2. Определение системы баз данных (СБД) и её назначение.
3. Основные этапы проектирования БД.
4. Трехуровневая архитектура БД.
5. Доступ к данным в трехуровневой архитектуре.
6. Моделирование предметной области. Модель сущность-связь: основные понятия и методы. Этапы моделирования Назначение модели. Свойства связей.
7. Графические нотации представления ER модели данных.
8. Понятие РМД. Основные концепции и термины. Фундаментальные свойства отношений. Понятие потенциального, первичного и альтернативного ключей.
9. Структурная часть реляционной модели данных (РМД).
10. Целостностная часть РМД. Виды ограничений целостности. Возможный и первичный ключи отношений, внешние ключи.
11. Манипуляционная часть РМД. Эквивалентность абстрактных реляционных языков.
12. Реляционная алгебра. Операции объединения, пересечения, разности, произведения, присвоения.
13. Реляционная алгебра. Операции выборки, создания проекций, деления.
14. Реляционная алгебра. Операция соединения (естественное соединение, тета-соединение, внешнее соединение).
15. Язык SQL. Структура запроса на выборку. Команды SELECT, FROM, WHERE. Использование операторов сравнения, логических операторов, операторов IN, BETWEEN, LIKE в команде WHERE.
16. Язык SQL. Структура запроса на выборку. Команда SELECT. Исключение избыточных данных в результирующих отношениях.
17. Язык SQL. Структура запроса на выборку. Упорядочивание выходных результатов.
18. Язык SQL. Структура запроса на выборку. Группировка данных: предложения GROUP BY и HAVING.
19. Язык SQL. Организация многотабличных запросов: естественное соединение, тета-соединение, внешнее соединение, соединение таблицы с самой собой.
20. Язык SQL. Структура запросов с подзапросами. Некоррелированные подзапросы. Использование DISTINCT, IN и агрегатных функций в подзапросах.
21. Структура запросов с подзапросами. Коррелированные подзапросы. Сравнение коррелированных подзапросов и запросов на соединение.
22. Язык SQL. Комбинирование результирующих таблиц. Создание запросов на объединение, пересечение и разность.

23. Язык SQL. Операторы языка манипулирования данными: DELETE, UPDATE, INSERT.

24. Язык SQL. Средства определения схемы базы данных. Общая структура, этапы определения таблицы, определение столбцов.

25. Язык SQL. Средства определения схемы базы данных. Общая структура, этапы определения таблицы, ограничительные условия на таблицу.

26. Операция соединения отношений. Примеры с использованием реляционной алгебры и решения с использованием средств языка SQL.

Примеры практических задач:

2. С помощью языка SQL разработать запрос согласно ER модели с использованием ограничения.

3. С помощью языка SQL разработать запрос согласно ER модели с использованием группировки.

4. С помощью языка SQL разработать запрос согласно ER модели с использованием соединения.

5. С помощью языка SQL разработать запрос на модификацию таблицы

6. С помощью языка SQL разработать запрос на удаление данных из таблицы.

7. С помощью средств реляционной алгебры составить запрос согласно ER модели.

Результаты зачета определяются оценками «зачтено», «не зачтено».

Критерии зачета обусловлены логической демонстрацией приобретенных компетенций в соответствии с текущей программой. Демонстрация предусматривает уверенное использование терминологии, понимание и корректное использование математического аппарата, предусматривает корректность написания кода, его понимание и корректное использование в нем математических методов. Отметка «зачтено» выставляется за счет демонстрации полученных компетенций в практиках, домашних работах и итоговом задании: уверенное владение и понимание работы кода, знание и демонстрация в практике теоретических основ баз данных. Минимальный порог зачета составляет 55 баллов, ниже 55 – «не зачтено»

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=2631>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) План семинарских / практических занятий по дисциплине.

1. Создание схемы БД

2. Создание БД (MS SQL, PostgreSQL)

3. Выполнение запросов наполнения БД

4. Манипуляция БД

5. Объединение: join, left join, right join, inner join

6. Переменные в БД

7. Условия

8. Создание интерфейса БД (SQL, python, HTML, CSS, flask)

г) Методические указания по проведению лабораторных работ.

Подготовка к проведению лабораторных работ начинается в начале теоретического изложения изучаемой темы и продолжается по ходу её изучения при освоении материала на занятиях в рамках практических заданий и работе над ним в ходе самостоятельной подготовки дома и в библиотеках. Для качественного выполнения лабораторных работ студентам необходимо:

1) повторить теоретический материал по конспекту и учебникам;

2) ознакомиться с описанием лабораторной работы;

3) в специальной тетради для лабораторных работ записать название и номер работы, перечень необходимого программного обеспечения, подготовить алгоритм или код;

4) выяснить цель работы, четко представить себе поставленную задачу и способы её достижения, продумать ожидаемые результаты опытов;

5) ответить устно или письменно на контрольные вопросы по изучаемой теме или решить ряд задач;

6) изучить порядок выполнения лабораторной работы. Подготовить среду выполнения кода к работе. После проверки правильности алгоритма работы программы преподавателем можно начинать выполнение лабораторной работы.

д) Методические указания по организации самостоятельной работы студентов.

Формы самостоятельной работы студентов разнообразны. Они включают в себя:

– изучение и систематизацию практических и теоретических примеров в рамках выполнения текущих заданий по предмету;

– изучение учебной, научной и методической литературы, материалов периодических изданий с привлечением электронных средств официальной, статистической, периодической и научной информации;

– подготовку докладов и презентаций, написание программного кода и его отладка;

– участие в работе студенческих конференций, комплексных научных исследованиях.

Самостоятельная работа приобщает студентов к научному творчеству, поиску и решению актуальных современных проблем.

Примеры самостоятельной работы студентов:

– 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
```

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' )
Создайте аналогичный для своей базы данных

```

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- Базы данных: Учебное пособие / О.Л. Голицына, Н.В. Максимов, И.И. Попов. - 2-е изд., испр. и доп. - М.: Форум: ИНФРА-М, 2009. - 400 с.: ил.; 60х90 1/16. - (Профессиональное образование). (переплет) ISBN 978-5-91134-098-8 - Режим доступа: <http://znanium.com/catalog/product/182482>
- Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем: учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — М. : ИД «ФОРУМ» : ИНФРА-М, 2016. — 368 с. — (Высшее образование). - Режим доступа: <http://znanium.com/catalog/product/556449>
- Эрик Редмонд, Джим. Р. Уилсон - Семь баз данных за семь недель. Введение в современные базы данных и идеологию NoSQL - Издательство "ДМК Пресс" - 2013 - 384с. - ISBN: 978-5-94074-866-3 - Текст электронный // ЭБС ЛАНЬ - URL: <https://e.lanbook.com/book/58690>

– Дейт, К. Дж. Введение в системы баз данных = An introduction to Database systems / К. Дж. Дейт ; Введение в системы баз данных, 8-е издание.: Пер. с англ. — М.: Издательский дом "Вильямс", 2005. — 1328 с

б) дополнительная литература:

- Грабер, Мартин. SQL : справочное руководство: пер. с англ. / М. Грабер. — М. : Лори, 2006. — 354 с.
- Грофф, Джеймс . SQL : Энциклопедия : пер. с англ. / Д. Р. Грофф, П. Н. Вайнберг. — 3-е изд. — СПб. : Питер, 2003. — 896 с.

– Кузнецов, Сергей Дмитриевич. Основы баз данных : курс лекций : учебное пособие / С. Д. Кузнецов ; Интернет-Университет информационных технологий. — М. : Интернет-Университет информационных технологий, 2005. — 484 с.

– Вебер Джим, Эфрем Эмиль - Графовые базы данных: новые возможности для работы со связанными данными - Издательство "ДМК Пресс" - 2016 - 256с. - ISBN: 978-5-97060-201-0 - Текст электронный // ЭБС ЛАНЬ - URL: <https://e.lanbook.com/book/90122>

ресурсы сети Интернет:

- открытые онлайн-курсы
- Журнал «Эксперт» - <http://www.expert.ru>
- Официальный сайт Федеральной службы государственной статистики РФ - www.gsk.ru
- Общероссийская Сеть КонсультантПлюс Справочная правовая система. <http://www.consultant.ru>
- Официальный сайт Flask: <https://flask.palletsprojects.com/en/2.0.x/>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

– язык программирования R (RStudio) и Python;

– СУБД MS SQL (<http://www.microsoft.com/sqlserver/ru/ru/default.aspx>), PostgreSQL (<https://www.postgresql.org/>)

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>

– ЭБС Лань – <http://e.lanbook.com/>

– ЭБС Консультант студента – <http://www.studentlibrary.ru/>

– Образовательная платформа Юрайт – <https://urait.ru/>

– ЭБС ZNANIUM.com – <https://znanium.com/>

– ЭБС IPRbooks – <http://www.iprbookshop.ru/>

в) профессиональные базы данных (*при наличии*):

– Университетская информационная система РОССИЯ – <https://uisrussia.msu.ru/>

– Единая межведомственная информационно-статистическая система (ЕМИСС) – <https://www.fedstat.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Лаборатории, оборудованные компьютерами (не ниже i3, RAM 8Gb), проектором

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешенном формате («Актру»).

15. Информация о разработчиках

Степаненко Андрей Александрович, НИ Томский государственный университет, ассистент кафедры общей, компьютерной и когнитивной лингвистики ТГУ.