



Введение в анализ данных

Лекция 8.1

Шевцов Василий Викторович,
директор ДИТ РУДН, shevtsov_vv@rudn.university

Обработка внешних баз данных

Импорт внешних данных: Web, Access, Text.

Импорт с Web-страниц

Получение внешних данных

The screenshot displays the Microsoft Excel interface with the 'Данные' (Data) ribbon selected. The 'Получение внешних данных' (Get External Data) dropdown menu is open, showing various options for connecting to external data sources. The background shows a blank Excel worksheet with columns L and M visible.

Ресурсы:

- Получение внешних данных
- Создать запрос
- Показать запросы
- Из таблицы
- Последние источники
- Скачать & преобразовать
- Обновить все
- Подключения
- Свойства
- Изменить связь
- Сортировка
- Фильтр
- Очистить
- Повторить
- Дополнительно
- Текст по столбцам

Получение

- Из Access
- Из Интернета
- Из текста
- Из других источников
- Существующие подключения

С сервера SQL Server
Создание подключения к таблице SQL Server. Импорт данных в Excel как таблицы или отчета сводной таблицы.

Из служб аналитики
Создание подключения к кубу служб аналитики SQL Server. Импорт данных в Excel как таблицы или отчета сводной таблицы.

Из веб-канала данных ODATA
Подключение к веб-каналу данных ODATA. Импорт данных в Excel в виде таблицы или отчета сводной таблицы.

Из импорта данных XML
Открытие или сопоставление XML-файла в Excel.

Из мастера подключения к данным
Импорт данных в дополнительном формате с использованием мастера подключения к данным и OLEDB.

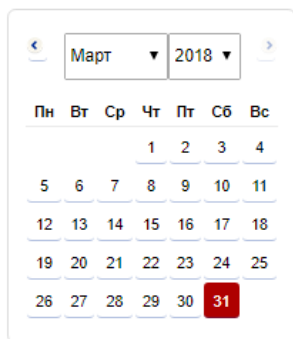
Из Microsoft Query
Импорт данных в дополнительном формате с использованием мастера запросов Microsoft Query и ODBC. Функциональность ограничена в целях совместимости с предыдущими версиями.

Готово

Из Интернета

http://cbr.ru/currency_base/daily.aspx?date_req=31.03.2018

Официальные курсы валют на заданную дату, устанавливаемые ежедневно



Центральный банк Российской Федерации установил с 31.03.2018 следующие курсы иностранных валют к рублю Российской Федерации без обязательств Банка России покупать или продавать указанные валюты по данному курсу

Цифр. код	Букв. код	Единиц	Валюта	Курс
036	AUD	1	Австралийский доллар	44,1169
944	AZN	1	Азербайджанский манат	33,6358
051	AMD	100	Армянских драмов	11,9426
933	BYN	1	Белорусский рубль	29,3290
975	BGN	1	Болгарский лев	36,0974
986	BRL	1	Бразильский реал	17,3267
348	HUF	100	Венгерских форинтов	22,5977

Из Интернета

Книга2 - Excel


Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Power Pivot Что вы хотите сделать?

Получение внешних данных

Создание веб-запроса

Адрес: http://cbr.ru/currency_base/daily.aspx?date_req=31.03.2018 Пуск

Щелкните значки ➡ таблиц, которые нужно выбрать, и нажмите кнопку "Импорт".

 **Центральный банк
Российской Федерации**

Официальные курсы валют на заданную дату,
устанавливаемые ежедневно

Март 2018

Пн	Вт	Ср	Чт	Пт	Сб	Вс
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Импорт Отмена

Готово

Из Интернета

Импорт данных ? X

Куда следует поместить данные?

☒ Имеющийся лист:

=SAS1

☐ Новый лист

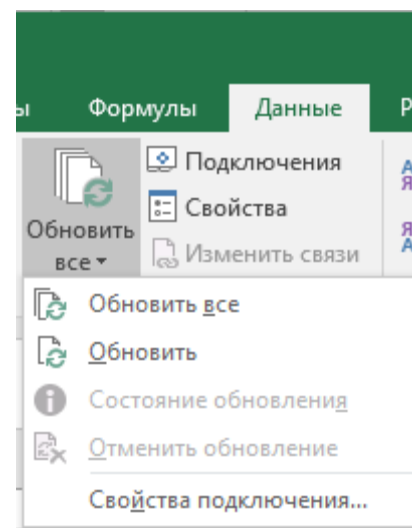
☐ Добавить эти данные в модель данных

Свойства... OK Отмена

	A	B	C	D	E	F
1	xЗакреть					
2	Интернет-приемная Ответы на вопросы					
3	RU					
4	EN					
5	База данных по курсам валют					
6						
7	Официальные курсы валют на заданную дату, устанавливаемые ежедневно					
8	Динамика официального курса заданной валюты					
9	Официальные курсы валют на заданную дату, устанавливаемые ежемесячно (до 11.01.2010)					
10	Курсы валют за период до 01.07.1992					
11	Официальные курсы Госбанка СССР					
12	Архив информационных сообщений					
13	Официальные курсы валют на заданную дату, устанавливаемые ежедневно					
14						
15	Центральный банк Российской Федерации установил с 31.03.2018 следующие курсы иностранных					
16	Цифр. код	Букв. код	Единиц	Валюта	Курс	
17	36	AUD	1	Австралийский доллар	44,1169	
18	944	AZN	1	Азербайджанский манат	33,6358	
19	51	AMD	100	Армянских драмов	11,9426	
20	933	BYN	1	Белорусский рубль	29,329	
<div> < > Лист1 Лист7 Лист8 Лист9 Лист5 Лист6 Лист2 Лист3 Лист4 </div>						
<div> Готово </div>						

Зачем?

- Не надо открывать стороннее приложение
- Не надо форматировать данные
- Таблицы при переносе могут не сохранить структуру
- Про большом объеме данных могут возникнуть проблемы с производительностью
- Можно один раз настроить источник, а потом только обновлять данные



Запросы

Создать запрос

Показать запросы

Из таблицы

Последние источники

Обновить все

Подключения

Свойства

Изменить св

- Из файла
 - Из книги
 - Из CSV
 - Из XML
 - Из текста
 - Из папки
- Из базы данных
- Из Azure
- Из других источников
- Объединить запросы

Поиск каталогов данных

Мои запросы к каталогу данных

Параметры источника данных...

Параметры запроса

Показать запросы

Из таблицы

Последние источники

Обновить все

Подключения

Свойства

Изменить связи

Сортировка

Фильтр

- Из файла
- Из базы данных
 - Из базы данных SQL Server
 - Из базы данных Microsoft Access
 - Из базы данных SQL Server Analysis Services
 - Из базы данных Oracle
 - Из базы данных IBM DB2
 - Из базы данных MySQL
 - Из базы данных PostgreSQL
 - Из базы данных Sybase
 - Из базы данных Teradata
- Из Azure
- Из других источников
- Объединить запросы

Создать запрос

Показать запросы

Из таблицы

Последние источники

Обновить все

Подключения

Свойства

Изменить связи

Сортировка

Фильтр

- Из файла
- Из базы данных
- Из Azure
- Из других источников
 - Из Интернета
 - Из списка SharePoint
 - Из канала OData
 - Из файла Hadoop (HDFS)
 - Из Active Directory
 - Из Dynamics 365 (интернет-версия)
 - Из Microsoft Exchange
 - Из Facebook
 - Из объектов Salesforce
 - Из отчетов Salesforce
 - Из ODBC
 - Пустой запрос
- Объединить запросы

Поиск каталогов данных

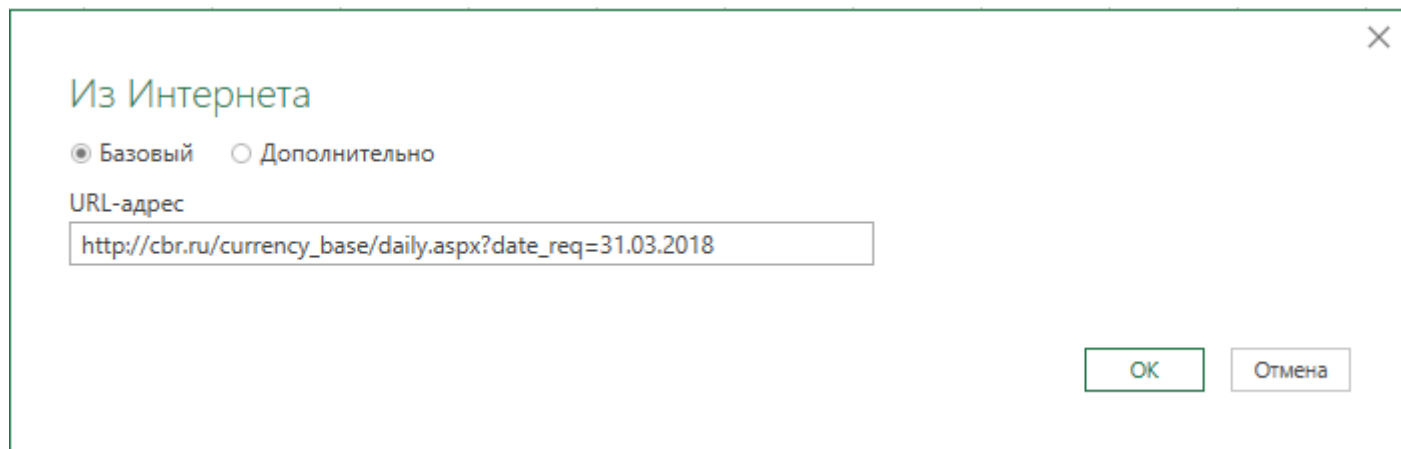
Мои запросы к каталогу данных

Параметры источника данных...

Параметры запроса



Запрос из Интернета



Из Интернета

☒ Базовый ☐ Дополнительно

URL-адрес

ОК Отмена

Запрос из Интернета

Навигатор

☐ Несколько элементов

Отобразить параметры ▾

http://cbr.ru/currency_base/daily.aspx?date_re...

Document

Table 0

Центральный банк Российской Федерации...

Представление таблицы Веб-представление

Центральный банк Российской Федерации установил с...

Цифр. код	Букв. код	Единиц	Валюта
36	AUD	1	Австралийский доллар
944	AZN	1	Азербайджанский манат
51	AMD	100	Армянских драмов
933	BYN	1	Белорусский рубль
975	BGN	1	Болгарский лев
986	BRL	1	Бразильский реал
348	HUF	100	Венгерских форинтов
410	KRW	1000	Вон Республики Корея
344	HKD	10	Гонконгских долларов
208	DKK	10	Датских крон
840	USD	1	Доллар США
978	EUR	1	Евро
356	INR	100	Индийских рупий
398	KZT	100	Казахстанских тенге
124	CAD	1	Канадский доллар
417	KGS	100	Киргизских сомов
156	CNY	10	Китайских юаней
498	MDL	10	Молдавских леев
934	TMT	1	Новый туркменский манат
578	NOK	10	Норвежских крон
985	PLN	1	Польский злотый

Загрузить ▾ Правка Отмена

Результат

	A	B	C	D	E
1	Цифр. код ▼	Букв. код ▼	Единиц ▼	Валюта ▼	Курс ▼
2	36	AUD	1	Австралийский доллар	44,1169
3	944	AZN	1	Азербайджанский манат	33,6358
4	51	AMD	100	Армянских драмов	11,9426
5	933	BYN	1	Белорусский рубль	29,329
6	975	BGN	1	Болгарский лев	36,0974
7	986	BRL	1	Бразильский реал	17,3267
8	348	HUF	100	Венгерских форинтов	22,5977
9	410	KRW	1000	Вон Республики Корея	53,935
10	344	HKD	10	Гонконгских долларов	72,9619
11	208	DKK	10	Датских крон	94,7263
12	840	USD	1	Доллар США	57,2649
13	978	EUR	1	Евро	70,5618
14	356	INR	100	Индийских рупий	87,9443
15	398	KZT	100	Казахстанских тенге	17,9829
16	124	CAD	1	Канадский доллар	44,5156
17	417	KGS	100	Киргизских сомов	83,5685
18	156	CNY	10	Китайских юаней	91,2719
19	498	MDL	10	Молдавских леев	34,8009
20	934	TMT	1	Новый туркменский манат	16,3848

Импорт из текстового файла

Файлы CSV

CSV (от англ. Comma-Separated Values — значения, разделённые запятыми) — текстовый формат, предназначенный для представления табличных данных.

Спецификация:

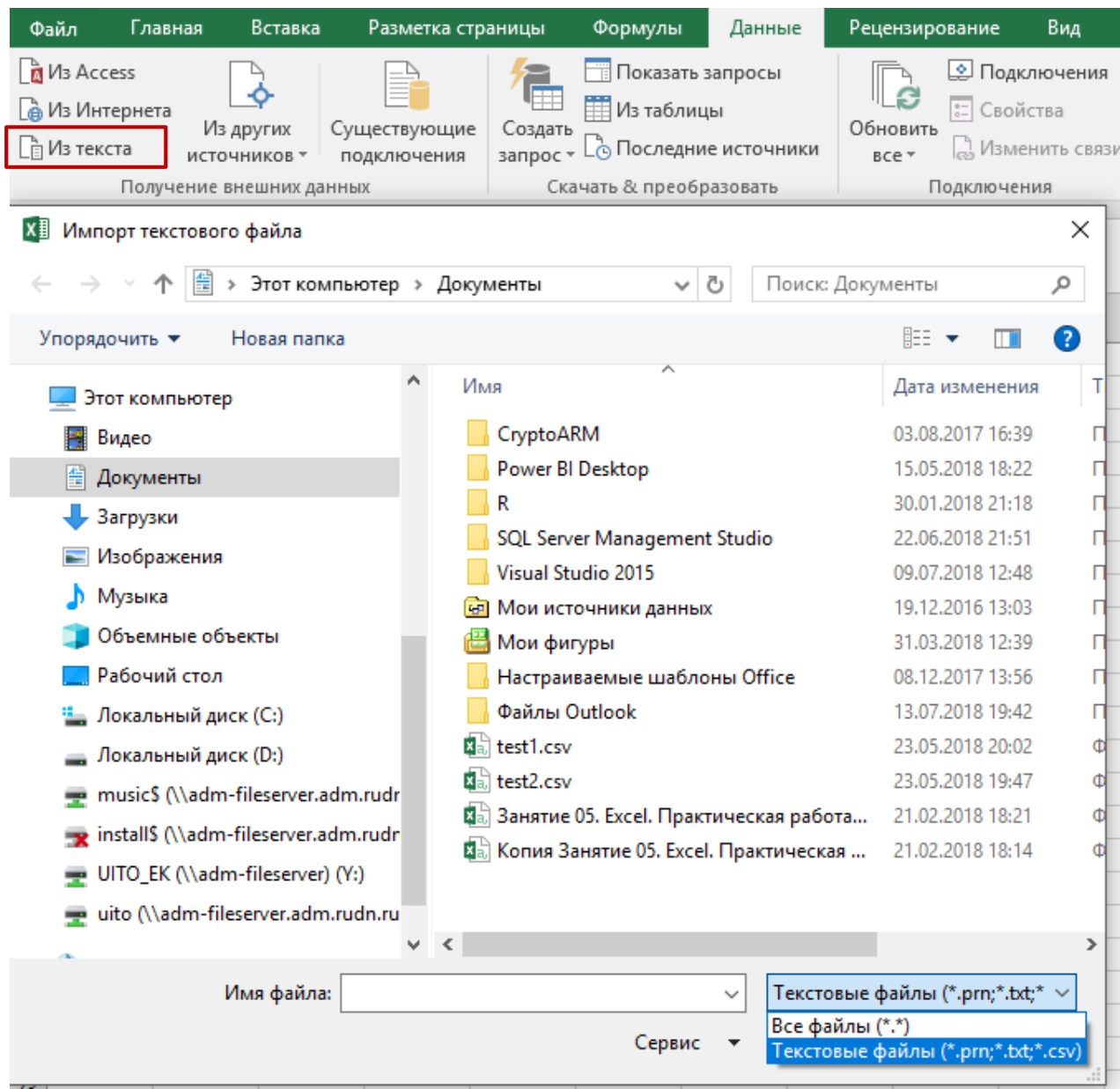
- Каждая строка файла — это одна строка таблицы.
- Разделителем (англ. delimiter) значений колонок является символ запятой (,). Однако на практике часто используются другие разделители, то есть формат путают с DSV и TSV.
- Значения, содержащие зарезервированные символы (двойная кавычка, запятая, точка с запятой, новая строка) обрамляются двойными кавычками ("). Если в значении встречаются кавычки — они представляются в файле в виде двух кавычек подряд.

Большинство программ понимают под CSV более общий формат DSV (англ. delimiter-separated values — значения разделённые разделителем), допускающий использование иных символов в качестве разделителя. В частности, в русской и других локалях запятая по умолчанию зарезервирована под десятичный разделитель. Поэтому как разделитель используется точка с запятой или табуляция (формат TSV).

Текстовые файлы

Текстовым файлам с разделителями CSV операционной системой назначается программа по умолчанию – MSExcel.

В списке фильтра файлов присутствуют стандартные текстовые файлы и все файлы (*.*)



Мастер текстов. Шаг 1

Мастер текстов (импорт) - шаг 1 из 3

Данные восприняты как список значений с разделителями.
Если это верно, нажмите кнопку "Далее >", в противном случае укажите формат данных.

Формат исходных данных

Укажите формат данных:

☒ с разделителями — значения полей отделяются знаками-разделителями

☐ фиксированной ширины — поля имеют заданную ширину

Начать импорт со строки: 1

Формат файла: 866 : Кириллица (DOS)

☒ Мои данные содержат заголовки

Предварительный просмотр файла C:\Users\v.shevtsov\Documents\test2.csv.

1	"", "mpg", "cyl", "disp", "hp", "drat", "wt", "qsec", "vs", "am", "gear", "carb"
2	"Mazda RX4", 21, 6, 160, 110, 3.9, 2.62, 16.46, 0, 1, 4, 4
3	"Mazda RX4 Wag", 21, 6, 160, 110, 3.9, 2.875, 17.02, 0, 1, 4, 4
4	"Datsun 710", 22.8, 4, 108, 93, 3.85, 2.32, 18.61, 1, 1, 4, 1
5	"Hornet 4 Drive", 21.4, 6, 258, 110, 3.08, 3.215, 19.44, 1, 0, 3, 1

Отмена < Назад Далее > Готово

Выбрать для файлов с разделителями (csv)

Выбрать для файлов из редакторов типа Лексикон

Если в предварительном просмотре текст не читается, сменить кодировку

Изучить предварительный просмотр на предмет заголовков

Мастер текстов. Шаг 2

Мастер текстов (импорт) - шаг 2 из 3

В этом диалоговом окне можно установить разделители для текстовых данных. Результат выводится в окне образца разбора.

Символом-разделителем является:

- ☐ знак табуляции
- ☐ точка с запятой
- ☒ запятая
- ☐ пробел
- ☐ другой:

☐ Считать последовательные разделители одним

Ограничитель строк:

Образец разбора данных

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1

Отмена < Назад Далее > Готово

При выборе разделителя в предварительном просмотре виден результат

Обычно достаточно значения по умолчанию. Касается специфических форматов, записанных в одну строчку

Можно использовать любой один символ

На этом этапе доступна опция завершения работы мастера, но следующий шаг очень важный!

Мастер текстов. Шаг 3

Мастер текстов (импорт) - шаг 3 из 3

Данное диалоговое окно позволяет задать для каждого столбца формат данных.

Формат данных столбца

☒ общий
☐ текстовый
☐ дата: ДМГ
☐ пропустить столбец

Общий формат является наиболее универсальным. Числовые значения автоматически преобразуются в числа, даты — в даты, а все прочие значения — в текст.

[Подробнее...](#)

Образец разбора данных

Текст	Общий	Общий	Общий	Общий	Общий	Общий	Общий	Общий	Общий	Общий	Общий	Общий
	mpg	cyl	dis	hp	drat	wt	qsec	vs	am	gear	carb	
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4	
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4	
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1	
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1	

Отмена < Назад Далее > Готово

Дополнительная настройка импорта текста

Настройка определения числовых данных

Разделитель целой и дробной части: ,

Разделитель разрядов:

Примечание. Числа отображаются в соответствии с настройкой элемента "Язык и стандарты" панели управления.

[Сброс](#) ☒ Знак минус в конце отрицательных чисел

[OK](#) [Отмена](#)

Для каждого столбца необходимо выбрать тип данных. Это связано с особенностью Excel переводить данные из общего формата в определенный тип. Возникают проблемы с числовыми идентификаторами и датами. Для данных, полученных из других локализаций можно сменить десятичный разделитель.

Мастер текстов. Результат

Импорт данных

Выберите способ представления данных в книге.

☒ Таблица

☐ Отчет сводной таблицы

☐ Сводная диаграмма

☐ Только создать подключение

Куда следует поместить данные?

☒ Имеющийся лист:

= \$A\$1

☐ Новый лист

☐ Добавить эти данные в модель данных

Свойства... OK Отмена

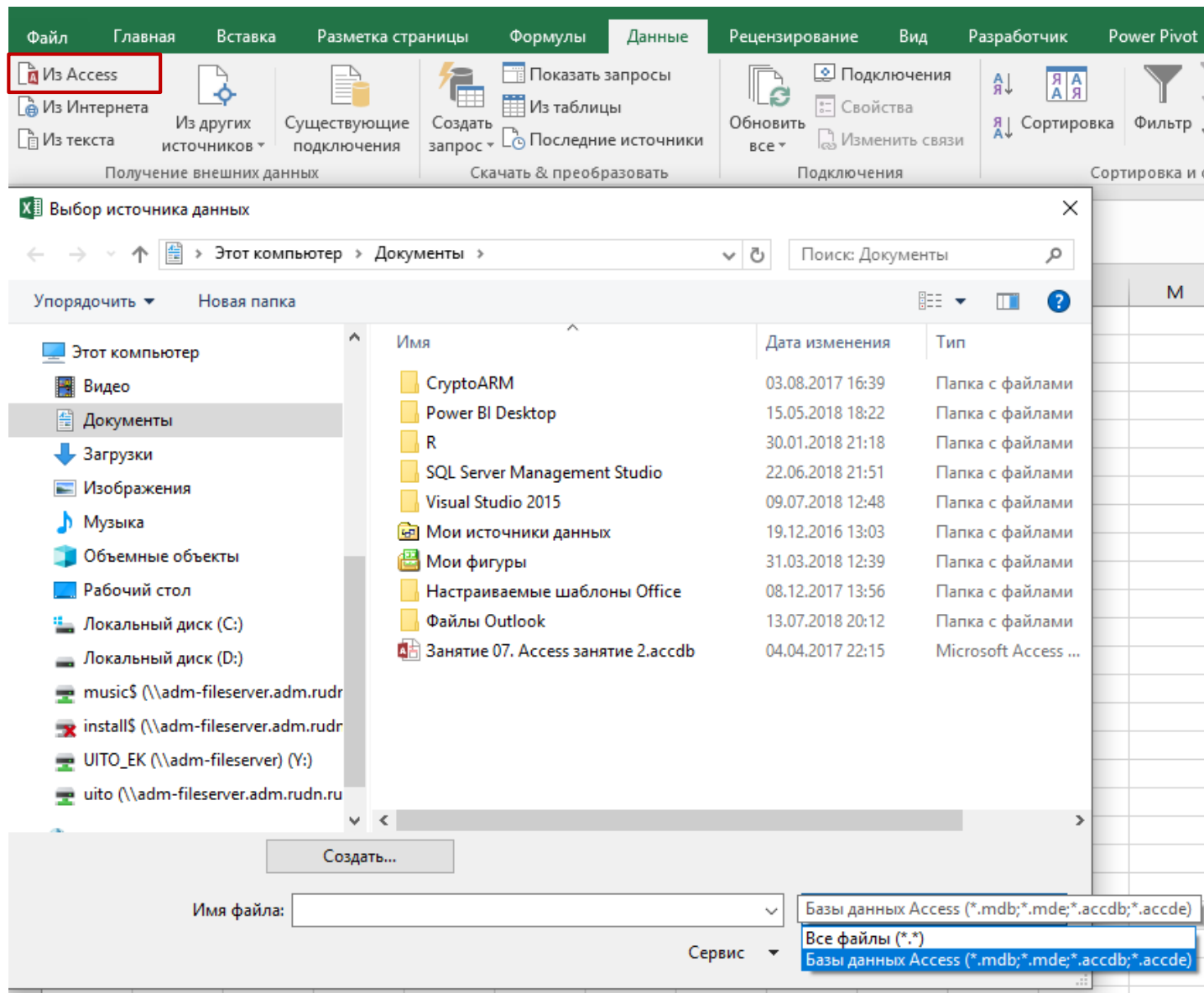
	A	B	C	D	E	F	G	H	I	J	K	L
1		mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
2	Mazda RX4	21	6	160	110	3,9	2,62	16,46	0	1	4	4
3	Mazda RX4 Wag	21	6	160	110	3,9	2,875	17,02	0	1	4	4
4	Datsun 710	22,8	4	108	93	3,85	2,32	18,61	1	1	4	1
5	Hornet 4 Drive	21,4	6	258	110	3,08	3,215	19,44	1	0	3	1
6	Hornet Sportabout	18,7	8	360	175	3,15	3,44	17,02	0	0	3	2
7	Valiant	18,1	6	225	105	2,76	3,46	20,22	1	0	3	1
8	Duster 360	14,3	8	360	245	3,21	3,57	15,84	0	0	3	4
9	Merc 240D	24,4	4	146,7	62	3,69	3,19	20	1	0	4	2
10	Merc 230	22,8	4	140,8	95	3,92	3,15	22,9	1	0	4	2
11	Merc 280	19,2	6	167,6	123	3,92	3,44	18,3	1	0	4	4
12	Merc 280C	17,8	6	167,6	123	3,92	3,44	18,9	1	0	4	4
13	Merc 450SE	16,4	8	275,8	180	3,07	4,07	17,4	0	0	3	3
14	Merc 450SL	17,3	8	275,8	180	3,07	3,73	17,6	0	0	3	3
15	Merc 450SLC	15,2	8	275,8	180	3,07	3,78	18	0	0	3	3
16	Cadillac Fleetwood	10,4	8	472	205	2,93	5,25	17,98	0	0	3	4
17	Lincoln Continental	10,4	8	460	215	3	5,424	17,82	0	0	3	4
18	Chrysler Imperial	14,7	8	440	230	3,23	5,345	17,42	0	0	3	4

Последним вопросом мастера является место расположения данных. Обратите внимание: в исходных данных в качестве десятичного разделителя использовалась точка. Благодаря настройкам мастера точка меняется на локальную запятую и данные отражаются корректно.

Импорт из Access

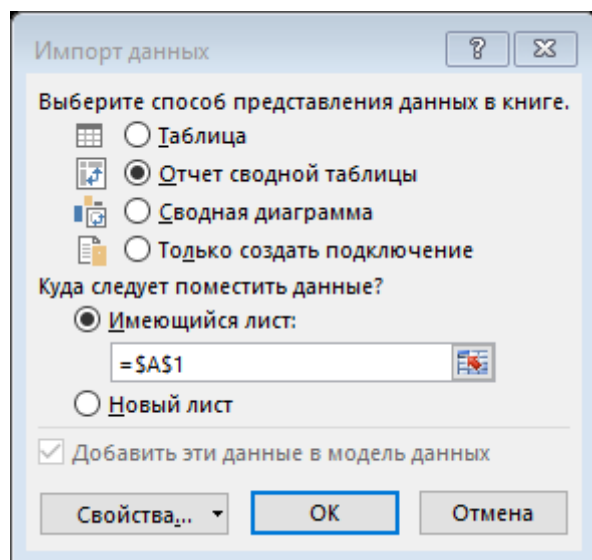
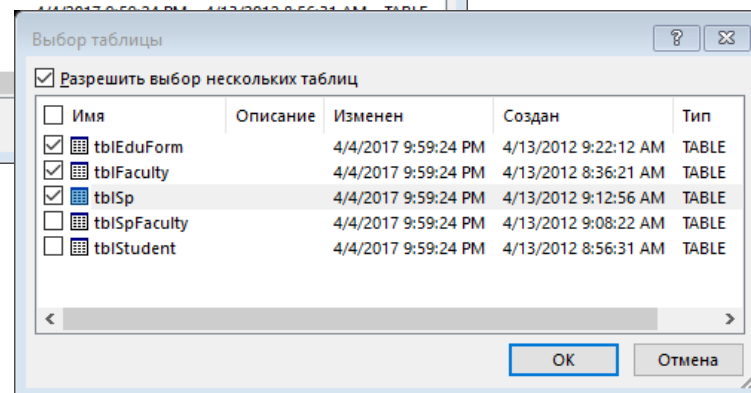
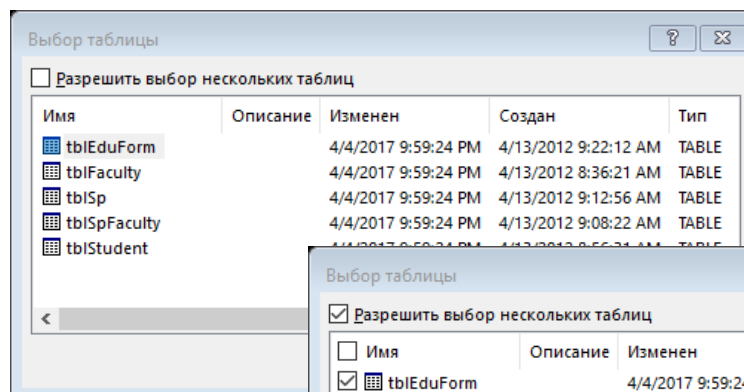
Импорт из Access

В списке
фильтра
файлов
расширения для
старый
форматов (mdb,
mde) и новых
форматов
(accdb, accde)



Импорт из Access

Microsoft Office Access – полноценная система управления базами данных, содержащая в своем составе множество объектов, в т.ч. таблицы, поэтому при импорте предлагается выбор из списка таблиц



Последнее диалоговое окно очень важное, Оно определяет способ представления данных в книге Excel:

- **Таблица** – все таблицы будут импортированы на отдельные листы и переведены в формат Таблица данных
- **Отчет сводной таблицы** – Excel используется как средство аналитики и генератор отчетов
- **Только создать подключение** – создается только подключение к файлу Access, которым можно воспользоваться по мере необходимости

Импорт из Access. Свойства подключения

Свойства подключения

Имя подключения: Занятие 07. Access занятие 2

Описание:

Использование | Определение

Обновление экрана

Последнее обновление:

☐ Фоновое обновление

☐ Обновлять каждые 60 мин

☐ Обновление при открытии файла

☐ Удалить данные из внешнего диапазона перед сохранением книги

☒ Обновлять это подключение по команде "Обновить все"

Форматирование сервера OLAP

При использовании этого подключения загружать с сервера следующие форматы:

☐ Формат чисел ☐ Цвет заливки

☐ Начертание ☐ Цвет текста

Детализация OLAP

Максимальное число извлекаемых записей:

Язык

☐ При загрузке данных и ошибок использовать язык интерфейса Office

OK Отмена

Свойства подключения

Имя подключения: Занятие 07. Access занятие 2

Описание:

Использование | **Определение**

Тип подключения: База данных Microsoft Access

Файл подключения: C:\Users\v.shevtsov\Documents\Занят... Обзор...

☐ Всегда использовать файл подключения

Строка подключения: Provider=Microsoft.ACE.OLEDB.12.0;User ID=Admin;Data Source=C:\Users\v.shevtsov\Documents\Занятие 07. Access занятие 2.accdb;Mode=Share Deny Write;Extended Properties='Microsoft Access Database Engine 12.0 OLEDB;Custom database ...'

☐ Сохранить пароль

Тип команды: Коллекция таблиц

Текст команды: "tblEduForm", "tblFaculty", "tblISp"

Службы Excel: Параметры проверки подлинности...

Изменить запрос... Параметры... Экспорт файла подключения...

OK Отмена

В свойствах подключения можно настраивать параметры обновления, параметры подключения, перечень объектов (при необходимости его корректировки) и т.д.

Импорт из Access. Результаты

	A	B	C	D	E	F
1	StudentID ▾	LastName ▾	FirstName ▾	SecondName ▾	Birth ▾	Sex ▾
2	168649	Антудова	Агуник	Александровна		Ж
3	168653	Ануфриева	Александра	Александровна		Ж
4	168658	Ануфриева	Александра	Александровна		Ж
5	168663	Ануфриева	Александра	Александровна		Ж
6	168664	Ануфриева	Александра	Александровна		Ж
7	168677	Апаршина	Алёна	Александровна		Ж
8	168711	Аралова	Анастасия	Александровна		Ж
9	168726	Арбузова	Анастасия	Александровна		Ж
10	168727	Аргал	Анастасия	Александровна		Ж
11	168735	Арефьева	Анастасия	Александровна		Ж
12	168757	Аринкина	Анжела	Александровна		Ж
13	168761	Аристова	Ани	Александровна		Ж
14	168766	Аркатова	Анна	Александровна		Ж
15	168830	Артемова	Анна	Александровна		Ж
16	168845	Артемяева	Анна	Александровна		Ж
17	168849	Артыкова	Армине	Александровна		Ж
18	168859	Арунова	Валерия	Александровна		Ж
19	168860	Арустамова	Валерия	Александровна		Ж
20	168876	Арутюнян	Виктория	Александровна		Ж

Результаты импорты в виде таблицы и в виде сводной таблицы. В источниках сводной таблицы присутствуют все импортированные таблицы и связи (Анализ - Отношения)

Поля сводной таблицы

АКТИВНАЯ

BCE

Выберите поля для добавления в отчет:

Поиск

tblEduForm2

☐ EduFormID
 ☐ EduForm

tblFaculty2

☐ FacultyID
 ☐ Faculty

tblSp2

☐ SpID
 ☐ Sp

tblSpFaculty1

☐ SpFacultyID

Перетащите поля в нужную область:

ФИЛЬТРЫ

СТОЛБЦЫ

СТРОКИ

Σ ЗНАЧЕНИЯ

☐ Отложить обновление макета

ОБНОВИТЬ

fx

Поля, элементы и наборы ▾

fx

Средства OLAP ▾

Отношения

Вычисления

Спасибо за внимание!



Шевцов Василий Викторович

shevtsov_vv@rudn.university
+7(903)144-53-57