

Подготовка данных для сводной таблицы Excel

Если у вас, как и у большинства офисных работников, имеются терабайты разного рода данных, стекающихся со всех сторон, то вам наверняка потребуется все это каким-то образом обобщить, структурировать и подготовить так, чтобы можно было делать полезные и значимые выводы. Возможно, эти данные имеют вид бесконечных списков, в которых фиксируются отдельные акты продаж, расходы и выплаты; либо это перечни клиентов, поставщиков, сотрудников, пособий, заказов; возможно, это сведения о складских запасах или каталоги товаров, книг, коллекций; это могут быть бесчисленные множества накладных, счетов-фактур или других всевозможных документов. Кроме того, весьма вероятно, что все эти данные поступают к вам из самых разных источников и такой информационный поток не иссякает.

Когда-то давным-давно, когда я только начинал работать в сфере ИТ, сбор всех данных, а затем создание на их основе разного рода отчетов, графиков и диаграмм было — по большей части — делом команды ИТ-специалистов. В то время в каждом таком случае требовалось писать длинные и сложные программы, которые после своего запуска должны были открывать файлы данных, просматривать поочередно все содержащиеся в них записи, сортировать, фильтровать и, если требовалось, как-то обобщать, а затем, наконец, выдавать запрошенные конечными пользователями отчеты или диаграммы. Такие программы разрабатывались на различных языках программирования: COBOL, Fortran, Basic, Pascal, ColdFusion, VBA, dBase, FoxPro и др. Это был весьма трудоемкий процесс, чреватый ошибками и возникновением многих других проблем.

Для большинства компаний эти времена уже в далеком прошлом. Теперь это ваша обязанность (или ваших коллег и сотрудников) — собирать все данные из разных источников и, проведя соответствующий анализ, извлекать из них какие-то знания. От вас ожидается, что вы будете знать, как сортировать, фильтровать, обобщать данные, составлять диаграммы и отчеты, чтобы на

различного рода планерках и совещаниях представлять какие-то значимые выводы, которые следуют из этих данных. Ах да, кстати, ведь следующее такое совещание назначено на сегодня после обеда! Но вы же справитесь? Ведь это всего лишь рабочая рутина! Правда, от нее зависит ваша карьера, а может быть, и ваша дальнейшая жизнь...

Что же делать?

Не беспокойтесь, все будет хорошо. Эта книга призвана помочь вам разобраться в том, как работать с громадным объемом информации. Вам не нужно быть программистом, чтобы получать важные и значимые результаты от ее обработки. Из этой книги вы узнаете, как создавать и управлять сводными таблицами, то есть как пользоваться встроенным в Excel мощным инструментом создания отчетности. Инструмент сводных таблиц поможет вам обрабатывать огромные объемы данных, обобщать их практически любым способом, и при этом вам не потребуется помощь сотрудников отдела ИТ. Предпринимая описанные в этой книге действия, вы действительно сможете извлекать максимум пользы из своих наборов данных.

ЧТО ТАКОЕ ДАННЫЕ

Для целей этой книги понятие «данные» можно определить как список из строк, содержащих некие сведения или транзакции, которые каким-либо образом связаны между собой. Наборы данных могут представлять собой любое количество подобных списков. Это могут быть перечни имен клиентов, сотрудников, учителей, студентов, оценок, накладных, счетов-фактур, складских запасов, актов продаж, адресов, городов, стран; данные о кредиторской и дебиторской задолженностях, инвестициях и т. д. Данными могут быть любые перечни вещей, состояние которых требуется отслеживать. Это может быть короткий список в несколько строк, а может быть крупнейший, состоящий из сотен тысяч строк. Электронные таблицы Microsoft Excel в принципе могут содержать более 1 048 000 строк и более 16 000 столбцов. Еще большими по размеру базами данных, составленными из сотен миллионов строк, можно управлять с помощью инструмента Power Pivot, о котором говорится в главе 8 «Расширение возможностей сводных таблиц с помощью инструмента Power Pivot». Каждая строка рабочего листа Excel — это транзакция или запись. Каждый столбец — это поле.

Какими должны быть данные

Чтобы данные можно было использовать в сводной таблице, они должны быть определенным образом структурированы. При необходимости для такой очистки можно использовать пакет Power Query, который рассматривается далее

в этой главе, а также другие, более традиционные методы, встроенные в Excel. На рис. 1.1 показан пример «хороших», а на рис. 1.2 — «плохих» данных. Ниже приведены способы структурирования данных таким образом, чтобы они могли использоваться в сводной таблице.

- Важно, чтобы заголовки столбцов (названия полей) отображались в самой верхней строке, которая также называется *строкой заголовков*.
- Для заголовков столбцов следует использовать только одну строку (верхнюю).
- Если полученные данные, которые вы хотите анализировать с помощью сводной таблицы, не содержат строки заголовков, то следует найти время на то, чтобы вручную вставить ее в самом верху и дать каждому столбцу осмысленное название.
- Строка заголовков не обязательно должна быть строкой номер 1 в электронной таблице Excel, но она должна быть первой строкой списка данных. Аналогично первый столбец не обязательно должен быть столбцом A этой таблицы, но он должен быть первым столбцом этого списка.
- Если над строкой заголовков имеются еще какие-то записи, то непосредственно перед ней должна быть хотя бы одна полностью пустая строка, а если перед первым столбцом списка данных существуют какие-то непустые столбцы, то перед ним должен находиться хотя бы один пустой столбец.
- Ниже строки заголовка не должно быть никаких полностью пустых строк до самого конца списка данных. Точно так же не должно быть полностью пустых столбцов до правой границы списка. Это требование обусловлено тем, что при первом создании сводной таблицы Excel сам автоматически определит диапазон строк и столбцов списка данных и остановится на следующей полностью пустой строке, которую обнаружит либо над, либо под текущей ячейкой, а также на следующем полностью пустом столбце, который найдет либо слева, либо справа от текущей ячейки.
- В разных местах списка данных могут быть пустые ячейки, но целая строка или целый столбец этого списка не должны быть пустыми.
- Каждая отдельная операция/транзакция должна занимать только одну строку в электронной таблице Excel.
- В каждом столбце должна быть информация только одного вида. Например, столбец с заголовком «Страна» должен содержать только названия стран.
- Содержимое всех ячеек одного столбца должно иметь одинаковый тип данных: например, только текст, только числа или только даты.
- В списке данных не должно быть объединенных ячеек, которые образуются, когда несколько ячеек сливаются в одну большую. Наличие объединенных ячеек не позволит создать сводную таблицу.

EmployeeID	FirstName	LastName	JobTitle	BirthDate	HireDate	Region
1	Michael	Henderson	Sales Representative	02.01.1989	02.01.2017	East
2	Pam	Scott	Vice President	04.02.1993	03.02.2018	North
3	Jim	Flanders	Internet Sales Representative	05.03.1998	04.03.2019	South
4	Andy	Peterson	Sales Representative	02.05.1987	05.04.2020	West
5	Dwight	Costanza	Sales Manager	02.09.2003	07.06.2023	Central
6	Angela	Brady	Sales Representative	05.08.2001	08.07.2023	East
7	Toby	Kramer	Sales Representative	05.12.1999	06.05.2010	North
8	Ryan	Pierce	Sales Coordinator	04.01.1987	04.01.2018	South
9	David	Honeycutt	Sales Representative	04.01.1987	04.01.2018	West
10	Oscar	Burns	Sales Manager	04.06.1998	03.05.2012	Central

Рис. 1.1. Пример «хорошей» структуры данных

	A	B	C	D	E	F	G	H	I	J
1	1	Michael	Henderson	Sales Representative	02.01.1989		02.01.2017	East		
2	2	Pam		President, Sales	04.02.1993		03.02.2018	North		
3	3	Jim		net Bot	05.03.1998		04.03.2019	South		
4	4	Andy		Representative	02.05.1987					
5	5	Dwight	Costanza	Sales Manager	02.09.2003					
6										
7	6	Angela	Brady	Sales Representative	05.08.2001		08.07.2023	East		
8	7	Toby								
9	8	Ryan								
10	9	David	Kramer				02.09.2022	West		
11	10	Oscar	Burns	Sales Manager			03.05.2012	Central		
12										
13										
14										
15										

Рис. 1.2. Пример «плохих» данных

ПРИМЕЧАНИЕ**Проверка наличия объединенных ячеек**

Ниже приводится быстрый способ проверить, есть ли в вашем списке данных объединенные ячейки:

1. Выделите какую-либо ячейку внутри диапазона данных.
2. Выделите весь диапазон данных, нажав одновременно клавиши Ctrl + A.
3. Перейдите на вкладку Главная.
4. Щелкните по значку Найти, как показано на рис. 1.3.
5. В диалоговом окне Найти и заменить щелкните по кнопке Параметры. Появится список параметров, как показано на рис. 1.4.
6. Далее щелкните по кнопке Формат. Появится диалоговое окно Найти формат, как показано на рис. 1.5.

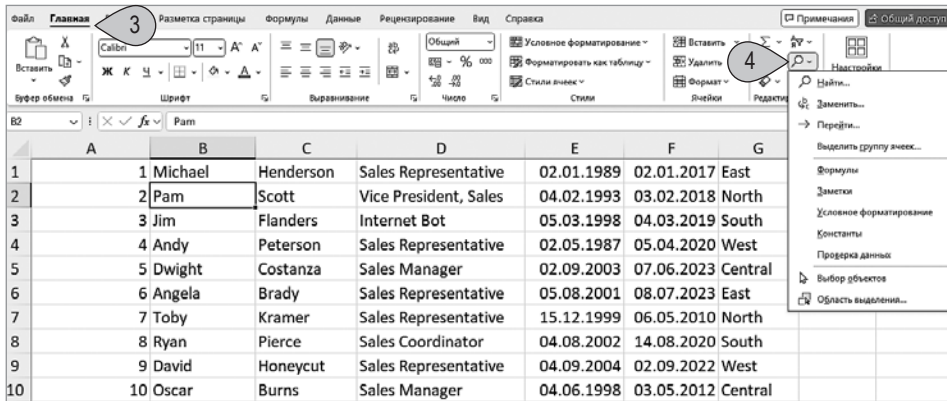


Рис. 1.3. Выбор значка Найти

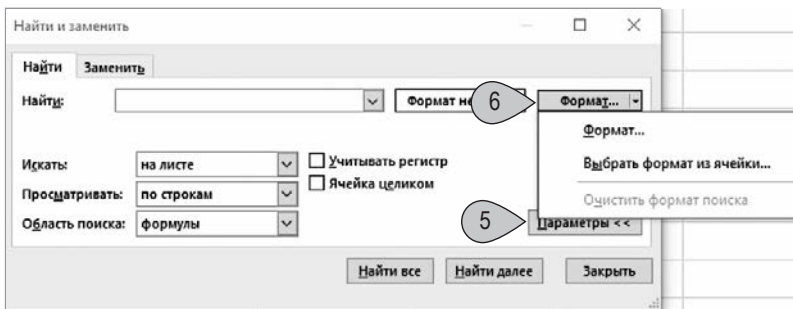


Рис. 1.4. Переход в следующее диалоговое окно по кнопке Параметры

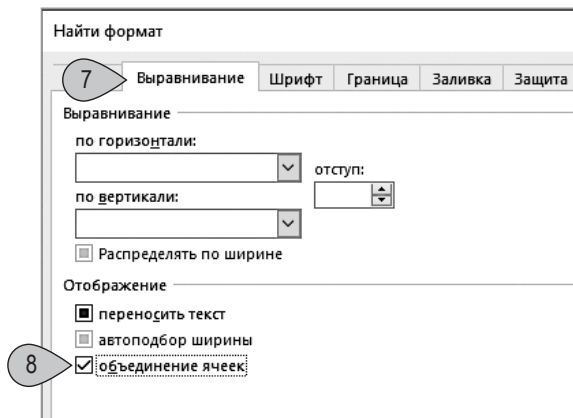


Рис. 1.5. Выбор вкладки Выравнивание в диалоговом окне Найти формат

7. Перейдите на вкладку Выравнивание.
8. Поставьте галочку рядом с опцией Объединение ячеек.
9. Щелкните по кнопке ОК.
10. Щелкните по кнопке Найти все. После этого будут найдены и выделены все объединенные ячейки в выбранном списке.
11. Закройте диалоговое окно Найти и заменить.
12. Если в списке данных были обнаружены объединенные ячейки, выделите этот список, переместите указатель мыши внутрь выделенной области, нажмите правую кнопку мыши, выберите пункт выпадающего меню Формат ячеек и кликните один или несколько раз по чекбоксу рядом с опцией Объединить ячейки, пока он не станет пустым. При этом все объединенные ячейки будут разделены.

Типы данных, которые можно использовать в сводных таблицах Excel

Данные для сводной таблицы могут поступать из разных источников. Они могут быть как в виде файла Excel, так и в других форматах. Внешние для Excel данные придется тем или иным образом импортировать и разместить их в рабочем листе Excel. Однако если в таблице более 1 048 000 строк, то их придется импортировать с помощью инструмента Power Pivot, о чем мы поговорим в главе 8 «Расширение возможностей сводных таблиц с помощью инструмента Power Pivot». Ниже перечислены наиболее распространенные источники данных, которые возможно импортировать в Excel, а затем внести в сводную таблицу:

- рабочая книга Excel;
- текстовый или CSV-файл;
- набор данных в формате XML;
- набор данных в формате JSON;
- файл PDF (этот исходный формат не является желательным, но его использование в принципе возможно, особенно если ничего другого нет);
- база данных SQL Server;
- база данных Access;
- база данных Oracle;
- база данных IBM DB2;
- база данных MySQL;
- база данных PostgreSQL;
- база данных Sybase;
- база данных Teradata;
- база данных SAP HANA;

- данные из Azure;
- данные из PowerBI;
- данные из Dataverse;
- потоки данных;
- список SharePoint;
- Microsoft Exchange;
- Microsoft Dynamics 365;
- Salesforce;
- веб-сайт, содержащий таблицу с данными;
- канал OData;
- файл Hadoop;
- источник данных ODBC;
- источник данных OLEDB.

Существуют и другие возможные источники данных, но выше перечислены основные, которые встречаются на практике. Кроме того, большинство широко используемых баз данных совместимы с ODBC, OLE или с ними обеими, так что почти каждая популярная база данных может служить источником, из которого данные можно импортировать в Excel и затем использовать для формирования сводной таблицы. Если вашей базы данных нет в вышеприведенном списке и она не совместима с ODBC или OLE, то все равно есть большая вероятность того, что в ней предусмотрена возможность экспорта данных либо в текстовый/CSV-файл, либо даже прямо в таблицу Excel. Поэтому в вашем распоряжении все равно должен быть способ преобразования ваших данных в формат Excel. Используя недавно усовершенствованный инструмент Power Query, вы можете импортировать непосредственно в Excel данные практически из любых источников. Power Query — это мощное встроенное в Excel средство, которое позволяет импортировать и структурировать данные из других источников, а также различными способами приводить информацию к нужному виду. Некоторые из этих способов будут продемонстрированы в данной главе.

Работа с данными формата Excel

Вполне вероятно, что у вас уже есть данные в существующей рабочей книге Excel, которые можно вставить в сводную таблицу. Они могут быть представлены как в виде неформатированной таблицы неочищенных данных — это так называемые *исходные (первичные, «сырые») данные*, или же могут быть структурированы в виде форматированной таблицы (таблицы Excel). На рис. 1.6 показан пример неформатированной, а на рис. 1.7 — форматированной таблицы Excel.

	A	B	C	D	E	F	G	H
1	EmployeeID	FirstName	LastName	JobTitle	BirthDate	HireDate	Region	
2	1	Michael	Henderson	Sales Representative	02.01.1989	02.01.2017	East	
3	2	Pam	Scott	Vice President, Sales	04.02.1993	03.02.2018	North	
4	3	Jim	Flanders	Internet Bot	05.03.1998	04.03.2019	South	
5	4	Andy	Peterson	Sales Representative	02.05.1987	05.04.2020	West	
6	5	Dwight	Costanza	Sales Manager	02.09.2003	07.06.2023	Central	
7	6	Angela	Brady	Sales Representative	05.08.2001	08.07.2023	East	
8	7	Toby	Kramer	Sales Representative	15.12.1999	06.05.2010	North	
9	8	Ryan	Pierce	Sales Coordinator	04.08.2002	14.08.2020	South	
10	9	David	Honeycut	Sales Representative	04.09.2004	02.09.2022	West	
11	10	Oscar	Burns	Sales Manager	04.06.1998	03.05.2012	Central	
12								
13								

Рис. 1.6. Неформатированная таблица неочищенных данных

	A	B	C	D	E	F	G	H
1	EmployeeID	FirstName	LastName	JobTitle	BirthDate	HireDate	Region	
2	1	Michael	Henderson	Sales Representative	02.01.1989	02.01.2017	East	
3	2	Pam	Scott	Vice President, Sales	04.02.1993	03.02.2018	North	
4	3	Jim	Flanders	Internet Bot	05.03.1998	04.03.2019	South	
5	4	Andy	Peterson	Sales Representative	02.05.1987	05.04.2020	West	
6	5	Dwight	Costanza	Sales Manager	02.09.2003	07.06.2023	Central	
7	6	Angela	Brady	Sales Representative	05.08.2001	08.07.2023	East	
8	7	Toby	Kramer	Sales Representative	15.12.1999	06.05.2010	North	
9	8	Ryan	Pierce	Sales Coordinator	04.08.2002	14.08.2020	South	
10	9	David	Honeycut	Sales Representative	04.09.2004	02.09.2022	West	
11	10	Oscar	Burns	Sales Manager	04.06.1998	03.05.2012	Central	
12								
13								

Рис. 1.7. Форматированная таблица

Любая из них будет прекрасным источником для создания сводной таблицы, хотя для этой цели форматированная таблица имеет свои преимущества. Среди них можно выделить следующие.

- Диапазон форматированной таблицы автоматически расширяется при вставке в нее новых столбцов или строк, которые будут перенесены и в сводную. С другой стороны, если в список данных, не являющийся форматированной таблицей, были добавлены какие-то строки или столбцы, то диапазон данных, используемый для создания сводной таблицы, придется расширять вручную.
- Форматированные таблицы можно использовать для создания единой сводной из нескольких листов путем объединения по общим полям. Этот прием рассматривается в главе 7 «Создание сводной таблицы из нескольких электронных таблиц» и в главе 8 «Расширение возможностей сводных таблиц с помощью инструмента Power Pivot».
- Форматированная таблица почти во всех случаях уже структурирована нужным для использования в сводной таблице образом. Список данных

в неформатированной таблице может потребовать очистки перед использованием в сводной таблице.

- Отформатированная таблица автоматически удалит все объединенные ячейки, которые ранее имелись в списке, и не позволит объединять какие-либо находящиеся внутри себя ячейки.
- Сортировка и фильтрация данных в форматированных таблицах облегчаются благодаря наличию выпадающих списков в верхней части каждого столбца.
- Заголовки в отформатированной таблице будут оставаться в верхней части каждого столбца по мере прокрутки вниз. В неформатированных таблицах этого не происходит, если не использовать встроенную в Excel команду **Закрепить области**, к которой можно получить доступ на вкладке **Вид**.
- Отформатированная таблица готова к использованию в Power Pivot.
- Отформатированная таблица дает подсказки при создании формул.

Вы можете отформатировать таблицу, выполнив следующие действия¹:

1. Щелкните по любой ячейке в списке ваших данных.
2. Перейдите на вкладку **Главная**.
3. Щелкните по опции **Форматировать как таблицу**, имеющейся на ленте инструментов Excel. Откроется галерея стилей таблиц, как показано на рис. 1.8.

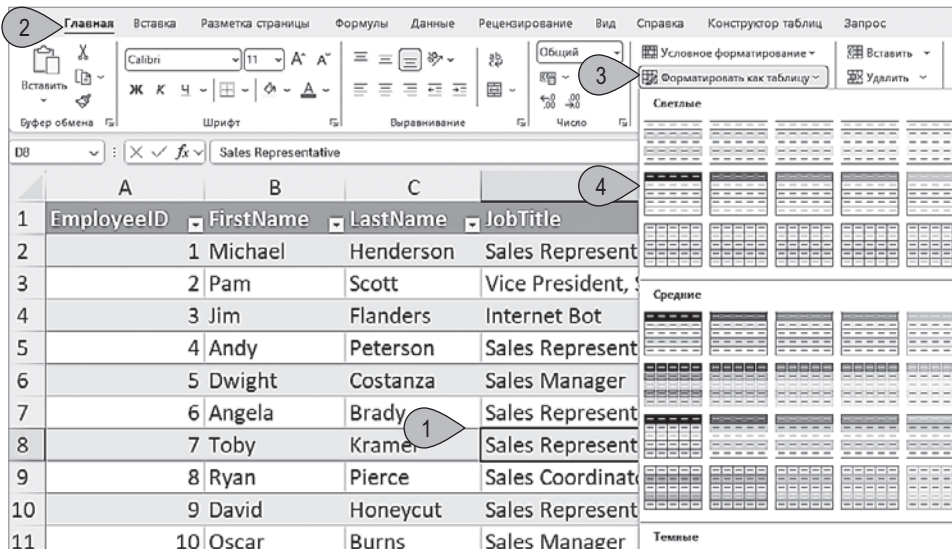


Рис. 1.8. Форматирование данных в виде таблицы

¹ Если какие-то опции или значки на панели инструментов не видны, то следует развернуть окно Excel во весь экран. — *Примеч. ред.*

4. Выберите один из стилей. Появится диалоговое окно **Форматирование таблицы**.
5. Убедитесь, что установлен флажок **Таблица с заголовками**.
6. Щелкните по кнопке **ОК**.

Теперь ваш список представляет собой отформатированную таблицу.

ИМПОРТ ДАННЫХ В EXCEL ИЗ ВНЕШНИХ ИСТОЧНИКОВ

Иногда возникает ситуация, когда нужно импортировать в Excel данные из каких-либо внешних источников. Инструмент **Power Query** позволяет производить импорт данных из файлов самых разных типов и баз данных различных форматов. При импорте автоматически создается таблица, куда заносятся данные в новом формате; на основе этой таблицы можно затем легко создать сводную. В следующем разделе мы рассмотрим шаги, необходимые для импорта данных из наиболее часто встречающихся источников. Работа с другими источниками, которые мы здесь не упоминаем, проводится аналогичным образом.

ПРИМЕЧАНИЕ

Для получения доступа к внешнему источнику данных вам может потребоваться помощь ИТ-отдела. При работе с такими источниками соблюдайте установленные правила и следуйте процедурам, принятым в вашей компании.

Импорт данных из текстового/CSV-файла

Текстовые/CSV-файлы — это простые неотформатированные файлы с текстовой информацией. Они являются универсальным источником данных, поскольку практически все программы для работы с базами данных и электронными таблицами позволяют импортировать данные из текстового/CSV-файла и/или экспортировать в него. CSV-файл — это особый тип текстового файла. Сокращение CSV расшифровывается как *comma-separated values*, то есть каждое поле данных в таком файле отделяется от других запятыми. Хотя CSV-файлы являются наиболее популярным форматом для хранения текстовых данных, в качестве источников последних могут использоваться файлы и других видов.

Желательно, чтобы названия полей (заголовки столбцов) были в первой строке текстового файла, служащего источником данных. Но в принципе это необязательно. Все это можно добавить вручную после успешного импорта данных в Excel. Также желательно, чтобы исходный текстовый файл имел какую-то узнаваемую структуру. Поля данных должны быть разделены каким-либо символом,

обычно одним из следующих: пробел, запятая, точка с запятой, двоеточие или знак табуляции. С другой стороны, файл, содержащий текст в свободной форме, например фрагмент из книги или журнальной статьи, скорее всего, не будет подходящим источником данных, которые можно было бы использовать в сводной таблице.

ПРИМЕЧАНИЕ

Напоминаем, что файлы примеров для этой книги можно загрузить с сайта издательства «Питер». QR-код для скачивания находится в разделе «От издательства». В следующем примере используется файл *ProductDescriptions.csv* (Описание продукции), входящий в состав набора файлов примеров.

Ниже описаны шаги по импорту данных из текстового/CSV-файла.

1. Откройте новую пустую рабочую книгу и перейдите на вкладку **Данные**.
2. Щелкните по опции **Получить данные** и выберите пункты **Из файла** ► **Из текстового/CSV-файла** в открывшемся меню, как показано на рис. 1.9. Появится диалоговое окно **Импорт данных**.



Рис. 1.9. Импорт данных

3. Перейдите в папку, где находится ваш текстовый/CSV-файл, и щелкните по нему левой кнопкой мыши.
4. Щелкните по кнопке **Импорт**. В новом окне появятся данные из выбранного вами файла, как показано на рис. 1.10.
5. Возможно, вам придется изменить символ разделителя полей, но обычно Excel правильно его определяет. Если все же вам придется его заменить на какой-то другой, попробуйте использовать каждый символ из выпадающего списка разделителей. Таким перебором добейтесь, чтобы данные выглядели пригодными для их разнесения по столбцам. В CSV-файле в качестве разделителя обычно используется запятая.
6. Щелкните по кнопке **Загрузить**.