**Глава 4. Импорт всех файлов из папки**

Это продолжение перевода книги Кен Пульс и Мигель Эскобар. Язык М для Power Query. Главы не являются независимыми, поэтому рекомендую читать сначала.

[Предыдущая глава](http://baguzin.ru/wp/?p=20906) [Содержание](http://baguzin.ru/wp/?p=20852#content) Следующая глава

Power Query имеет специальную опцию для импорта всех файлов из папки. Эта глава посвящена импорту двоичных файлов, таких как TXT и CSV. Следующая глава посвящена импорту данных из нескольких книг Excel. Примеры файлов для этой главы разбиты на четыре подпапки: Begin, 2008 – More, 2009, 2010. Большинство скриншотов получено в Excel 365 (июль 2019). Упоминаются отличия, замеченные в Excel 2019.

Создайте новую книгу, пройдите на вкладку *Данные*, кликните *Получить данные* –> *Из файла* –> *Из папки*:

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.1. Меню импорта файлов из папки

В окне *Папка* выберите папку *Begin* (рис. 4.2), нажмите Ok.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.2. Выбор папки для импорта файлов

Откроется окно, в котором отражаются все файлы в папке и их тип:

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.3. Окно предварительного просмотра содержимого выбранной паки; в Excel 2019 кнопка *Преобразовать данные* называется *Изменить*

Нажмите *Преобразовать данные*, и список файлов загрузиться в редактор Power Query (рис. 4.4).

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.4. Окно загрузки файлов в редакторе Power Query

В этой главе вы будете объединять csv-файлы из четырех папок. Прежде чем идти дальше, вы должны отфильтровать список, чтобы оставить только csv-файлы. Если этого не сделать, а кто-то в будущем поместит в вашу папку Excel-файл, это может сломать загрузку.

Для начала застрахуйтесь от того, что расширение может быть набрано в верхнем CSV или нижнем csv регистре (поскольку текстовые фильтры чувствительны к регистру). Щелкните правой кнопкой мыши на заголовке столбца Extension (расширение) –> *Преобразование* –> *нижний регистр*. Далее щелкните стрелку фильтра в столбце Extension, пройдите по меню *Текстовые фильтры* –> *Равно*, и в открывшемся окне фильтрация строк, установите *равно csv* (рис. 4.5). Поскольку у вас в папке есть только csv-файлы, то кликнув на стрелку фильтра вы увидите *Выбрать все*. Не переживайте, в строке формул отразился правильный фильтр (см. строку формул на рис. 4.4). Обратите внимание, что в фильтре перед csv нужна точка.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.5. Установка фильтра для файлов загрузки

### Объединение файлов

Посмотрите на значки в первых трех столбцах текущего запроса:

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.6. Пиктограммы первых трех столбцов текущего запроса

В столбце *Name* отображается значок невыбранного фильтра, в столбце *Extension* – примененный фильтр. Но самым замечательным является значок в столбце *Content*. Эта пиктограмма появляется на столбцах, содержащих двоичные файлы, и при ее нажатии происходит объединение файлов. Сначала откроется окно предварительного просмотра (рис. 4.7), а после нажатия Ok, все файлы из папки Begin объединяться в один запрос, и будут импортированы в редактор Power Query (рис. 4.8). Любопытно, что в Excel 2019 при импорте был также добавлен столбец *Source.Name*, содержащий имя файла в каждой строке трансакции. Таким образом, можно было бы обойтись без части ухищрений, описанных ниже))

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.7. Окно предварительного просмотра *Объединить файлы*

Переименуйте столбцы, измените тип данных с использованием локали, удалите ошибки, переименуйте запрос в Transactions, кликните *Закрыть и загрузить* (подробнее см. [главу 2](http://baguzin.ru/wp/?p=20889)).

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.8. Объединенные данные в окне редактора Power Query

Убедитесь, что таблица в Excel включает данные за три месяца. Создайте сводную:

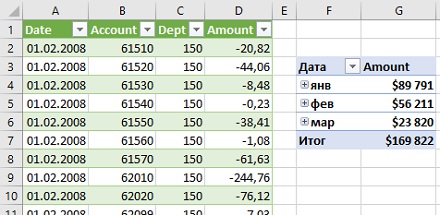


Рис. 4.9. Сводная таблица на основе объединенных данных из трех файлов из папки Begin

### Добавление новых файлов

Добавить новые файлы в запрос довольно просто:

* Перетащите папку 2008-More в папку Begin
* Вернитесь в Excel, перейдите на вкладку *Данные* (цифра 1 на рис. 4.10), кликните *Обновить* *всё* (2). Обратите внимание, что количество загруженных строк возросло (3). Кликните на сводной таблице правой кнопкой мыши и выберите Обновить (4). Сводная таблица включает теперь все файлы, содержащиеся в папке Begin. Даже несмотря на то, что ряд файлов размещено во вложенной папке.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.10. Добавление новых файлов в запрос

Перетащите папки 2009 и 2010 в папку Begin. Повторите операции обновления запроса и обновления сводной таблицы.

**Рекурсия.** Для добавления новых файлов в запрос не потребовалось перемещать файлы в корень папки Begin. Power Query по умолчанию использует рекурсию, т.е. проверяет корневую и все вложенные папки на наличие файлов с правильным расширением. Это позволяет сохранить файлы в отдельных папках, например, по годам (если это необходимо).

Если вы откроете редактор запросов и перейдете на шаг *Источник*, то увидите, что для каждого файла указан путь к папке. Если требуется, вы можете применить фильтры к этому столбцу, чтобы включить только корневую папку или исключить определенные папки.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.11. С помощью фильтрации можно включать в запрос только нужные папки

### Сохранение свойств файла при импорте

Допустим ваша бухгалтерская система настроена так, что экспортирует список транзакций в файл и присваивает ему имя месяца и года (например, Feb 2008.csv). Система не включает даты транзакций в файл, поскольку все они относятся к дате окончания месяца. При импорте нескольких файлов жизненно важно указать дату в каждой строке.

Вернитесь в редактор запросов. Перейдите на шаг *Отфильтрованные скрытые файлы1*. Выберите столбец *Name*, удерживая нажатой клавишу Ctrl, выберите столбец *Content*. Щелкните правой кнопкой мыши один из выделенных столбцов, выберите *Удалить другие столбцы.*

Power Query обрабатывая каждый шаг перед переходом к следующему. Это означает, что в отличие от Excel, он не требует наличия столбцов, в которых установлен фильтр. Учитывая это, и поскольку вы уже отфильтровали csv-файлы, вы можете удалить и этот столбец.

Во-первых, вы удалили все посторонние столбцы самым надежным способом. Это способ лучше, чем выделение столбцов, которые нужно удалить. Если в будущем в таблице появятся дополнительные столбцы, они всё равно будут удалены, так как команда фактически говорит6 «оставь только столбцы *Name* и *Content».* Во-вторых, из-за порядка, в котором вы выбрали столбцы, вы перевернули порядок, в котором они представлены в редакторе Power Query:

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.12. Два оставшихся столбца расположены в порядке обратном, к тому что был до команды *Удалить другие столбцы*

Следующая задача – преобразовать имена файлов в допустимые даты конца месяца:

* Щелкните правой кнопкой мыши столбец *Name*, выберите *Замена значений*
* Замените .csv на пусто
* Еще раз щелкните правой кнопкой мыши столбец *Name*, выберите *Замена значений*
* Замените ˽ (пробел) на ˽1,˽ (пробел, единица, запятая, пробел)

Имеем:

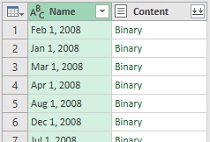


Рис. 4.13. Промежуточный результат преобразований

Теперь от первого числа месяца можно перейти к последнему дню месяца:

* Щелкните правой кнопкой мыши столбец *Name* –> *Тип изменения* –> *Дата*
* Перейдите на вкладку *Преобразование* –> *Дата* –> *Месяц* –> *Конец месяца*
* Щелкните правой кнопкой мыши столбец *Name* –> *Переименовать* –> *Date*

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 4.14. Преобразование первого числа месяца в последнее число месяца

Нет необходимости этот набор дат импортировать с использованием локали, так как Power Query правильно распознает полнотекстовую дату.

### Объединение свойств файла с двоичным содержимым

На этом этапе вы готовы объединить даты окончания месяца со всеми строками внутри двоичных файлов. Проблема в том, что стандартное нажатие кнопки объединить двоичные файлы (см. рис. 4.6) выбросит столбец Date, над которым вы так усердно работали. Поэтому вам нужно извлечь содержимое каждого двоичного файла:

* Перейдите на вкладку *Добавление столбца* –> *Настраиваемый столбец*
* Введите формулу: =Csv.Document([Content]). Внимание! Формулы Power Query чувствительны к регистру.
* Нажмите Ok

Если вы не хотите вводить длинное имя поля, вы можете просто дважды щелкнуть его в списке полей при создании формулы пользовательского столбца. Это поместит поле в формулу в обрамлении квадратных скобок. В результате появится новый столбец под названием *Пользовательская*. В каждой строке этого столбца содержится таблица (Table):

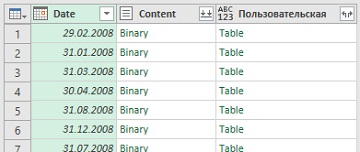


Рис. 4.15. В запрос добавлен пользовательский столбец

Csv.Document() – это функция, которую можно использовать для преобразования содержимого двоичного файла в таблицу. [Content] – просто ссылка на имя столбца. Любопытно, что функции Txt.Document() не существует. Для преобразовать содержимое текстового файла в таблицу, также используете Csv.Document().

Чтобы увидеть, что находится внутри любого из этих двоичных файлов, щелкните пустое пространство рядом со словом *Table* (1 на ри. 4.16). Откроется окно предварительного просмотра (2). Если вы щелкните на само слово *Table*, Power Query попытается запустить очередной шаг навигации. Кликните *Отмена*.

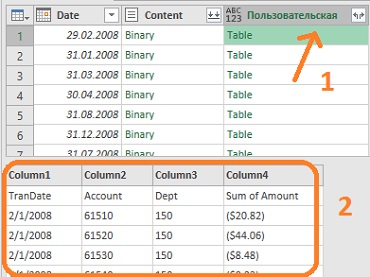


Рис. 4.16. Предварительный просмотр таблицы

Правой кнопкой мыши кликните заголовок столбца *Content* –> *Удалить*. Нажмите маленький двуглавый символ стрелки в правом верхнем углу столбца *Пользовательский*:

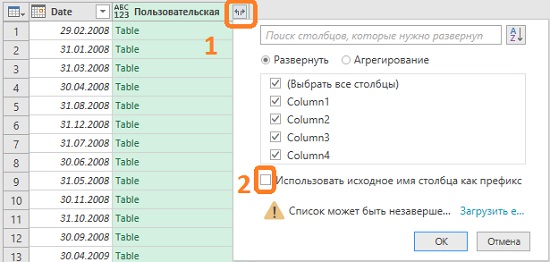


Рис. 4.17. Окно выбора столбцов *Table*, которые нужно развернуть

Это диалоговое окно позволяет выбрать, какие столбцы следует развернуть. Если некоторые столбцы вам не нужны, снимите галку. Вы уже видели в предварительном просмотре, что каждый из четырех столбцов содержит полезную информацию, поэтому сохраните их все. Снимите флажок *Использовать исходное имя столбца как префикс*. Если вы оставите его, Power Query добавит префикс *Пользовательская* к названию каждого столбца. Так что они будут иметь названия *Пользовательская*.*Столбец1*, *Пользовательская*.*Column2*, и так далее. В нашем случае, это не так важно, так как на следующем шаге вы выполните операцию *Использовать первую строку в качестве заголовков*.

При нажатии кнопки Ok (рис. 4.17) все столбцы разворачиваются, но Power Query сохраняет столбец Date (в отличие от функции объединения двоичных файлов, которая удалила его). Еще раз подчеркну. Это произошло потому, что мы развернули не двоичные файлы, а пользовательские таблицы.

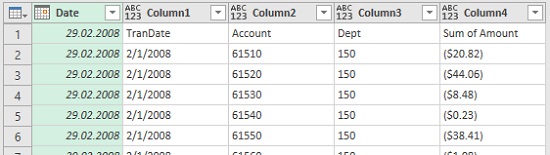


Рис. 4.18. При разворачивании таблиц *Table* столбец *Date* сохраняется

Повысьте заголовки. Обратите внимание, что первый столбец *Date* превратился в 29.02.2008. Переименуйте его в *Month End*. Удалите шаг *Столбец расширенной таблицы1*. В первом примере он выполнял функцию разворачивания двоичных файлов. Остальные шаги должны продолжить измененную первую часть запроса (переименование, измененный тип с локалью, удаление ошибок).

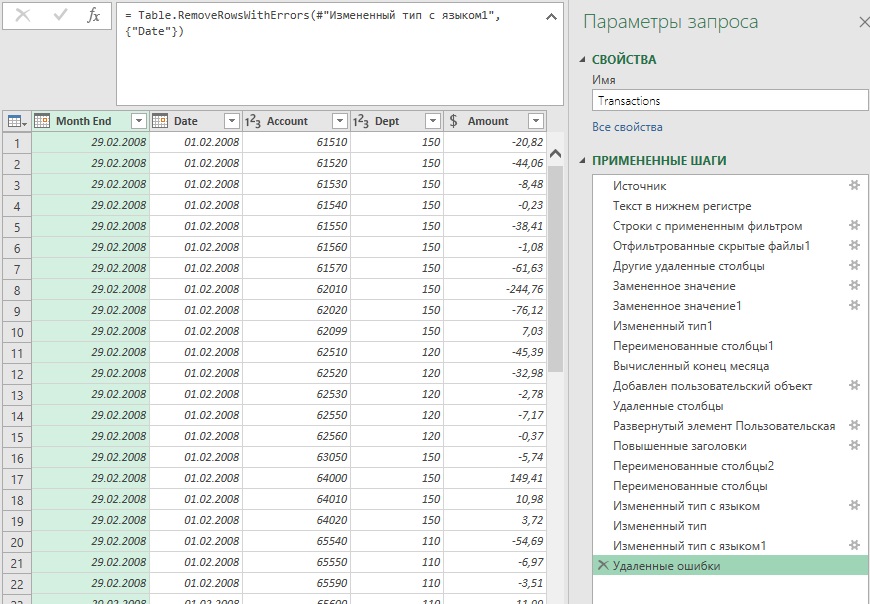


Рис. 4.19. Все файлы, импортированные из папки, сохранили дату на основе свойств файла

Измените имя запроса на *ManualCombine*. На вкладке *Главная* кликните *Закрыть и загрузить.* Если вы обновите сводную таблицу, то увидите, что она соответствует полученной ранее (см. рис. 4.10), с той лишь разницей, что появилось новое поле Month End.

**Резюме.** Импорт отдельных файлов, как описано в [главе 2](http://baguzin.ru/wp/?p=20889), как правило, является стартовым, когда вы только осваиваете работу с Power Query. Этот подход хорошо работает если набор данных стабилен, или требования к преобразованию различных файлов отличаются. Если же вы предполагаете, что решение на основе Power Query может масштабироваться, используйте подход, описанный в этой главе. Даже если сегодня в папке будет только один файл. Зато потом вам просто нужно будет поместить новый файл в папку и нажать *Обновить*. Иногда немного предвидения может значительно облегчить будущие работы.