

Конвертация данных из CSV в SHP и обратно с OGR

[Обсудить в форуме](#) Комментариев — 11

Эта страница опубликована в основном списке статей сайта по адресу <http://gis-lab.info/qa/csv2shp.html>

Рассмотрена конвертация данных из формата CSV в формат ESRI shape и обратно.

Содержание

- [1 Введение](#)
- [2 Формат CSV и особенности драйвера OGR CSV](#)
- [3 Работа с кавычками](#)
- [4 CSVT - файл описания типов полей](#)
- [5 Конвертация данных из .csv в .shp](#)
 - [5.1 UTF-8](#)
 - [5.2 Геометрии в формате WKT](#)
 - [5.3 Ошибки и предупреждения](#)
- [6 Конвертация данных из .shp в .csv](#)
- [7 Примеры программных реализаций](#)

Введение

Довольно часто возникает задача конвертации файлов формата csv в shape-файлы. Иногда возникает и обратная задача – конвертация shape-файлов в csv. Две данные операции и будут рассмотрены в этой статье.

Эта статья ориентирована на пользователя знакомого с командной строкой. В случае, если необходимо сконвертировать в shape один-другой файл csv, то возможно проще воспользоваться плагином QGIS «Delimited text», однако обратную конвертацию в QGIS выполнить на данный момент невозможно.

Стоит отметить, что до выхода версии GDAL 1.6.0 возможности конвертировать шейп-файл в формат csv с сохранением информации о геометрических объектах не было, поэтому советуем обновить вашу версию, если это необходимо.

Прежде чем переходить к преобразованию данных из одного формата в другой, рассмотрим, что же представляет из себя непосредственно формат CSV и более детально остановимся на особенностях драйвера OGR CSV. В этом описании используется информация с [официального сайта библиотеки GDAL](#).

Формат CSV и особенности драйвера OGR CSV

OGR поддерживает чтение и запись табулированных данных (данных разделенных специальным символом-разделителем), хранящихся в текстовых файлах формата csv, являющегося распространенным форматом обмена табличными данными между различными программами. Файл данного типа можно легко создать вручную, например, с помощью любого текстового редактора.

Хотя на практике файл с разделителями может иметь любое расширение (csv,txt,...), драйвер OGR CSV поддерживает только файлы с расширением .csv. Данное требование объясняется тем, что при использовании утилит командной строки OGR не указывается формат входного файла и поэтому его формат определяется на основании расширения. Именем источника данных может быть либо единственный csv файл, либо каталог, содержащий несколько файлов. Чтобы каталог распознал как источник csv данных, необходимо, чтобы по крайней мере половина файлов в каталоге имела расширение .csv. Из каждого .csv файла в каталоге

генерируется один слой (одна таблица).

В файле csv каждая строка описывает отдельный объект в слое (таблице). Значения полей отделены друг от друга запятыми. В каждой строке должно быть минимум 2 поля. Строки отделены друг от друга разделителями в стиле DOS (CR/LF) или в стиле Unix (LF). Каждая строка должна иметь одинаковое количество полей. Начиная с версии GDAL 1.7.0, драйвер также поддерживает в качестве разделителей между полями – точку с запятой или табуляцию. Это автоопределение будет работать только в том случае, если в первой строке нет других потенциально возможных разделителей. В противном случае в качестве разделителя будет использована запятая.

Драйвер пытается обработать первую строку файла как список имен всех полей. Однако, если одно из полей имеет числовое значение – то подразумевается, что в первой строке содержатся данные, описывающие отдельный объект, а в качестве имен полей используются имена field_1, field_2 и т.д.

Драйвер OGR CSV поддерживает как чтение, так и запись. Поскольку файл формата csv содержит текстовые строки переменной длины, то чтение выполняется последовательно (строка за строкой). Слой OGR CSV никогда не имеет описания системы координат. Драйвер OGR CSV возвращает тип всех полей как строковый, если нет файла описания типов полей (дополнительный файл с расширением .csvt).

Пример (employee.csv):

```
ID,Salary,Name,Comments
132,55000.0,John Walker,"The ""big"" cheese."
133,11000.0,Jane Lake,Cleaning Staff
```

Встречаются текстовые файлы, часто называемые csv, включающие данные не разделенные запятыми, а с фиксированной шириной поля. Данный драйвер не поддерживает подобный формат. Если у вас такие данные, преобразуйте их в простой .csv формат, чтобы он стал понятен драйверу OGR CSV.

Работа с кавычками

Составные значения полей (например, содержащие запятые, кавычки) следует помещать в двойные кавычки. Если в значении поля содержатся кавычки, то их следует удвоить. Например:

```
"La"ke" -> "La""ke"
```

Наличие одинарной (не заключенной в двойные) кавычки в одном из полей "порвет" импорт CSV файла, импортируются данные только до этой строки (актуально для версии GDAL 1.9.1). Пример файла, в котором будет импортированы только 2 из 3 строк:

```
lat,long,Name
52.3637,32.2096,Walker
53.2712,39.2771,La"ke
58.31106,64.55456,Doe
```

CSVТ - файл описания типов полей

Поддерживаемые типы полей - Integer, Real, String, Date (YYYY-MM-DD), Time (HH:MM:SS+nn) и DateTime (YYYY-MM-DD HH:MM:SS+nn). Как уже отмечалось, файл с описанием типов полей должен иметь расширение .csvt и такое же имя, как и файл csv. Типы полей задаются в двойных кавычках через запятую в одну строку (например, "Integer","String"). Также возможно непосредственно указать ширину и точность каждого поля в скобках после его типа.

Рабочий пример (5 полей в исходном CSV):

```
"Real (10.7) ", "Real (10.7) ", "Integer (5) ", "String (255) ", "Integer (5) "
```

Конвертация данных из .csv в .shp

Одна из существенных возможностей OGR - извлечение пространственной информации из файла csv . В случае, если csv файл содержит поля в которых хранятся X и Y координаты точек, то конвертировать такой файл csv в shape-файл можно посредством использования промежуточного драйвера VRT. Описание этого драйвера и всех его параметров доступно на [официальном сайте](#).

Рассмотрим следующий файл csv (test.csv):

```
Latitude,Longitude,Name
48.1,0.25,"First point"
49.2,1.1,"Second point"
47.5,0.75,"Third point"
```

Создадим соответствующий файл vrt (test.vrt):

```
<OGRVRTDataSource>
  <OGRVRTLayer name="test">
    <LayerSRS>WGS84</LayerSRS>
    <SrcDataSource>test.csv</SrcDataSource>
    <GeometryType>wkbPoint</GeometryType>
    <GeometryField encoding="PointFromColumns" x="Longitude" y="Latitude"/>
  </OGRVRTLayer>
</OGRVRTDataSource>
```

и выполним команду:

```
ogr2ogr output test.vrt
```

Будет создана папка с названием output, а в нее добавлен файл test.shp.

Если же видоизменить команду:

```
ogr2ogr output.shp test.vrt
```

то shape-файл с именем output.shp будет создан в текущей папке.

UTF-8

Если исходные данные находятся в кодировке UTF-8, необходимо явным образом указать, что выходные данные также будут в UTF-8, с помощью опции `-lco ENCODING=UTF-8`, вот так:

```
ogr2ogr -lco ENCODING=UTF-8 output.shp test.vrt
```

Геометрии в формате WKT

В случае, если файл csv (test.csv) содержит поле в котором хранится описание геометрии объекта в формате WKT:

```
GeomField,Code
"POLYGON((1 2,3 4, 2 1))",8
"POLYGON((11 21,31 41, 21 11))",9
```

то чтобы конвертировать его в shape-файл нужно отредактировать файл vrt (test.vrt):

```
<OGRVRTDataSource>
  <OGRVRTLayer name="test">
    <SrcDataSource relativeToVRT="1">test.csv</SrcDataSource>
    <GeometryType>wkbUnknown</GeometryType>
    <GeometryField encoding="WKT" field="GeomField"/>
  </OGRVRTLayer>
</OGRVRTDataSource>
```

Если поле с WKT описанием геометрии в файле csv имеет имя WKT, например такой:

то можно обойтись без использования драйвера VRT. В этом случае конвертация из csv в shape выполняется командой:

Таким образом, что бы получить CSV файл, который легко перевести в shape-файл, нужно правильно заполнить одну из колонок csv файла WKT-описанием геометрии.

Длина названий полей в CSV должна быть не больше 10 символов, иначе остальные будут отброшены. Будет также выведено следующее сообщение:

Длина строковых значений по умолчанию - 80 символов, если у вас строка длиннее, она будет обрезана. Будет также выведено следующее сообщение:

Чтобы избежать этого - создайте файл *.csvt (см. выше).

Конвертация данных из шейп файла также выполняется утилитой ogr2ogr. Опции создания csv файла:

- Пример конвертации шейп файла в csv:

```
ogr2ogr -f CSV output input.shp -lco GEOMETRY=AS_WKT -lco CREATE_CSVT=YES
```

В результате выполнения данной команды в текущей директории будет создан каталог output и в него помещен файл input.csv, содержащий поле с WKT-представлением геометрических объектов, а также файл input.csvt с описанием типов полей csv файла.

Отметим, что ключ -f используется при конвертировании шейп файла в csv в то время как при конвертации csv в шейп данный ключ не используется. Это происходит потому, что выходной формат по умолчанию «ESRI Shapefile», что равносильно строке -f «ESRI Shapefile».

Примеры программных реализаций

[Perl](#) - обрабатывает все файлы в папке (csv), создает vrt и генерирует shape-файлы каждый в своей папке. Условие - во всех файлах lat и long столбцы должны называться одинаково. Протестировано для точечных слоев на Suse 11.

[Python](#) - обрабатывает все файлы в папке (csv), создает vrt и генерирует shape-файлы.

[Обсудить в форуме](#) Комментариев — 11

Последнее обновление: 2014-05-15 01:40

Дата создания: 20.09.2009

Автор(ы): [Денис Рыков](#)