FieldPyculator - калькулятор полей с использованием выражений на Python для QGIS

Обсудить в форуме Комментариев — 11

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

Использование выражений на языке Python для операций с полями в QGIS

Одной из наиболее частых задач при работе с векторными данными является обработка атрибутивных полей. Не менее часто приходится производить массовое изменение атрибутивных данных, как например заполнение поля значениями, вычисленными на основе других столбцов слоя. Возможность вычисления поля для всех записей в слое - обязательный признак современной ГИС общего назначения, и от удобства использования этого инструмента напрямую зависит скорость обработки атрибутивной информации.

Ядро QGIS имеет достаточно продвинутый калькулятор полей. Но его основной минус - невозможность использования функций, отличных от тех, что заложены разработчиками. На этот случай, в других ГИС обычно предусматривается возможность использования скриптовых языков, для создания необходимых функций обработки (например ArcGIS Desktop позволяет в калькуляторе полей активировать режим, при котором вычисление полей можно производить с использованием JavaScript или VB Script).

Для исправления этого недостатка QGIS было разработано расширение, позволяющее производить вычисление полей с помощью интерпретатора языка Python и использовать практически все доступные для него библиотеки. Идея впервые была предложена Raymond Nijssen. Но к сожалению, его плагин FieldEvaluator до сих пор находится на стадии глубокой разработки и позволяет вычислять лишь выражения на Python, не давая возможности использовать всю мощь этого языка. Данное расширение, по сути, позволяет пользователю создавать сколь угодно сложный алгоритм для вычисления значений, и имеет чуть более удобный интерфейс. Будем надеяться, Раймонд не обидится за то что мы предложили еще одну реализацию его идеи.

Содержание

- 1 Получение и установка
- 2 Работа с расширением
- <u>3 Примеры работы</u>
 - о 3.1 Заполнение поля GUID'ами
 - о 3.2 Обработка текстовых полей
 - <u>3.3 Изменение строк, удовлетворяющих</u> условию
 - 3.4 Замещение JOIN'а с последующим присвоением значений
 - о 3.5 Работа с геометриями
- 4 Дополнительные ссылки
- <u>5 Контакты</u>

Получение и установка

Расширение доступно из официального репозитория.

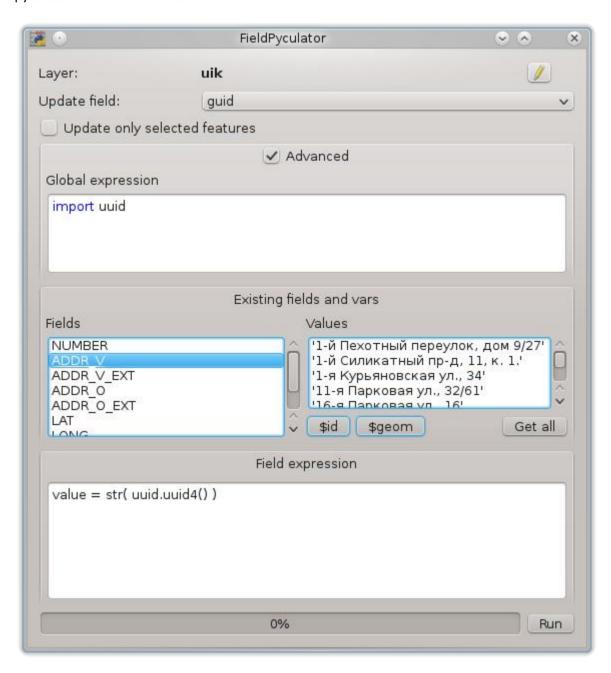
FieldPyculator разработан для QGIS версии 1.7.0 или старше и не имеет зависимостей.

Исходный код модуля можно получить выполнив команду:

git clone git://github.com/yellow-sky/FieldPyculator.git

Работа с расширением

После подключения и запуска расширения с помощью кнопки шили из меню Plugins->Field pyculator->Field pyculator появится главное окно.



Верхняя часть окна содержит название активного векторного слоя, выпадающий список для выбора редактируемого поля и кнопку начала/завершения редактирования активного слоя. Так же, с помощью флажка "Update only selected features" можно регулировать какие из записей будут обновлены - все или только выбранные с использованием стандартных инструментов QGIS.



Панель "Advanced" содержит поле для ввода "глобального" кода. Этот код будет выполнен один раз, перед циклом обработки каждой записи. В этом блоке удобно производить все подготовительные действия: импортировать необходимые модули, рассчитывать переменные, которые будут использоваться при

вычислении каждой из записей.



Панель "Existing fields and vars" содержит название всех полей редактируемого слоя. Двойной клик по элементу списка "Fields" приводит к добавлению выбранного поля в выражение для расчета, а так же заполнению списка "Values" урезанным набором значений выбранного поля. При нажатии кнопки "Get all" урезанный набор значений будет заменен полным, что может занять достаточно продолжительный период времени.

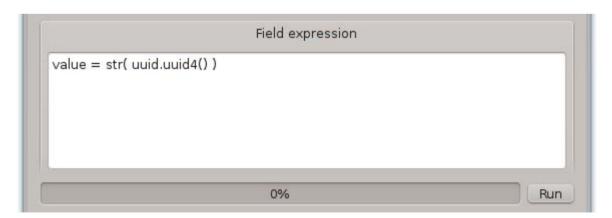
Так же на панели расположены кнопки добавления в выражение специальных переменных: "\$id" и "\$geom". При обработке выражения эти переменные будут содержать соответственно идентификатор обрабатываемой записи и её геометрию.



Код, вычисляющий значение редактируемого поля содержится в "Field Expression". Код может содержать функции, объявления переменных, и иметь сколь угодно сложную структуру. Но обязательным условием является объявление переменной "value" для всех веток выполнения кода. Значение этой переменной будет присвоено редактируемому полю.

Имена полей слоя необходимо заключать в угловые скобки (При добавлении из списка, скобки добавляются автоматически).

Кнопка 'Run' запускает процесс расчета.



Примеры работы

Для примеров воспользуемся результатами <u>проекта проверки УИКов г. Москвы</u>. Точечный слой можно скачать <u>тут</u>.

Заполнение поля GUID'ами

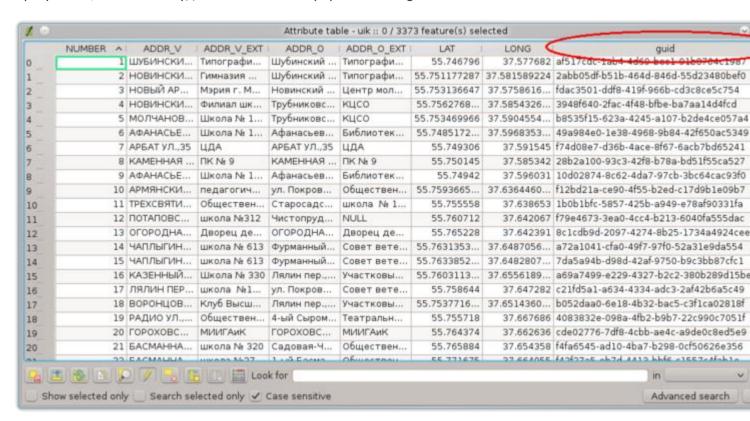
Допустим, нам необходимо каждому объекту слоя присвоить уникальный идентификатор, чтобы в дальнейшем иметь возможность связывать наш слой с данными, созданными на базе этого слоя (id типа int крайне ненадежная штука). Создадим новый столбец 'guid' типа string размером 36 символов. Запустив расширение и включив редактирование слоя, выбираем из выпадающего списка созданное поле 'guid'. В поле глобального кода добавляем код для импорта нужного нам модуля (для этого необходимо активировать переключатель 'Advanced'):

import uuid

Код, для вычисления поля выглядит следующим образом:

```
value = str( uuid.uuid4() )
```

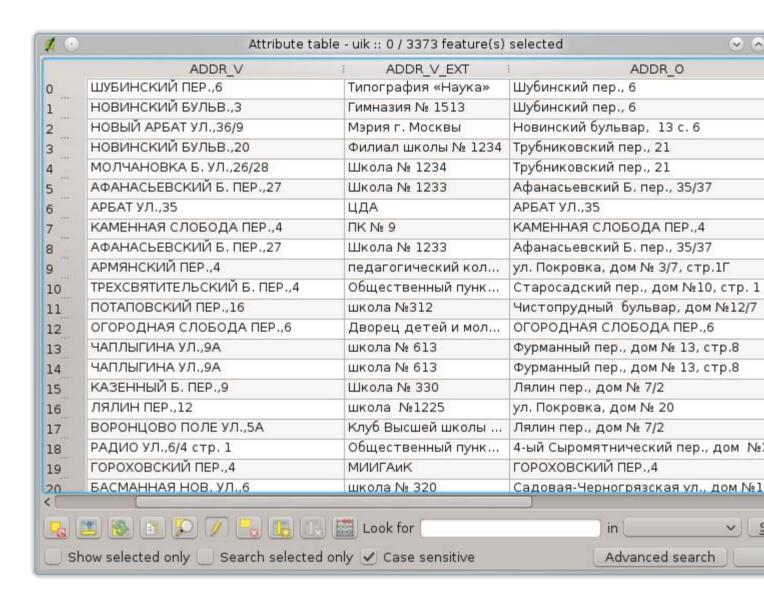
В результате, новое поле будет заполнено сгенерированными guid'ами:



Обработка текстовых полей

Работа с текстовыми атрибутами - одна из наиболее частых и в то же время запутанных. Связанно это с тем, что текстовая информация достаточно часто бывает неструктурированной.

Пример. Рассмотрим поля ADDR_O и ADDR_V слоя УИКов. Как можно заметить, в поле ADDR_O адреса записаны в произвольной форме, а различные формы записи перемешаны. Форма записи в поле ADDR_V выглядит более структурированной.



Давайте представим, что у нас нет поля ADDR_V, и попытаемся рассчитать его, использую поле ADDR_O. Создадим дополнительное поле, для записи результата: 'ADDR_O_STR' типа string и размером 250 символов. Пишем код обработки поля:

```
if <ADDR O>:
        #переводим в верхний регистр
       addr upper = <ADDR O>.upper()
        #удаляем вхождения ДОМ и №
       addr cleaned = addr upper.replace(u'ДOM','').replace(u'Nº','')
       #удаляем лишние запятые (все после первой) и пробелы после зпт
       parts = addr cleaned.partition(',')
       addr cleaned = parts[0]+','+parts[2].replace(',','').lstrip()
       #если УЛ. идет перед названием, то переносим
       if addr cleaned.startswith(u'УЛ.'):
           addr_cleaned = addr cleaned.replace(u'YT.', '')
           addr cleaned = addr cleaned.replace(',', u' УЛ.,')
        #меняем К. на КОРП. и С. на СТР.
       addr cleaned = addr cleaned.replace(u'K.', u'KOPH.').replace(u'C.', u'CTP.')
       #удаляем лишние пробелы
       value = addr cleaned.strip()
else:
       value = <ADDR O>
```

После выполнения получаем поле, готовое к интеграции и автоматизированной обработке:

		DR_O_	LAT :			
***	Шубинский пер., 6	Ти	55	37	af5	ШУБИНСКИЙ ПЕР.,6
44	Шубинский пер., 6	Ти	55	37	2a	ШУБИНСКИЙ ПЕР.,6
**	Новинский бульвар, 13 с. 6	Це	55	37	fda	НОВИНСКИЙ БУЛЬВАР,13 СТР
	Трубниковский пер., 21	K	55	37	39	ТРУБНИКОВСКИЙ ПЕР.,21
	Трубниковский пер., 21	K	55	37	b8	ТРУБНИКОВСКИЙ ПЕР.,21
4	Афанасьевский Б. пер., 35/37	Би	55	37	49	АФАНАСЬЕВСКИЙ Б. ПЕР.,35/3
	АРБАТ УЛ.,35	ЦДА	55	37	f74	АРБАТ УЛ.,35
	КАМЕННАЯ СЛОБОДА ПЕР.,4	ПК	55	37	28	КАМЕННАЯ СЛОБОДА ПЕР.,4
777	Афанасьевский Б. пер., 35/37	Би	55	37	10	АФАНАСЬЕВСКИЙ Б. ПЕР.,35/3
	ул. Покровка, дом № 3/7, стр.1Г	06	55	37	f12	ПОКРОВКА УЛ.,3/7 СТР.1Г
	Старосадский пер., дом №10, стр. 1	ш	55	37	1b	СТАРОСАДСКИЙ ПЕР.,10 СТР.
	Чистопрудный бульвар, дом №12/7	N	55	37	f79	ЧИСТОПРУДНЫЙ БУЛЬВАР,12
	ОГОРОДНАЯ СЛОБОДА ПЕР.,6	Д	55	37	8c	ОГОРОДНАЯ СЛОБОДА ПЕР.,6
	Фурманный пер., дом № 13, стр.8	Co	55	37	a7	ФУРМАННЫЙ ПЕР.,13 СТР.8
	Фурманный пер., дом № 13, стр.8	Co	55	37	7d	ФУРМАННЫЙ ПЕР.,13 СТР.8
	Лялин пер., дом № 7/2	Уч	55	37	a6	ЛЯЛИН ПЕР.,7/2
	ул. Покровка, дом № 20	Co	55	37	c2	ПОКРОВКА УЛ.,20
	Лялин пер., дом № 7/2	Уч	55	37	b0	ЛЯЛИН ПЕР.,7/2
	4-ый Сыромятнический пер., дом №3/5,	Te	55	37	40	4-ЫЙ СЫРОМЯТНИЧЕСКИЙ ПЕ
	ГОРОХОВСКИЙ ПЕР.,4	М	55	37	cd	ГОРОХОВСКИЙ ПЕР.,4
	Садовая-Черногрязская ул., дом №11/2	06	55	37	f4f	САДОВАЯ-ЧЕРНОГРЯЗСКАЯ У

Изменение строк, удовлетворяющих условию

Достаточно часто возникает необходимость обновить в слое только записи, удовлетворяющие определенному условию. Для выполнения этой операции с помощью стандартного калькулятора полей необходимо последовательно выполнить два действия: выбрать необходимые строки (с помощью расширенного поиска в таблице атрибутов) и при выполнении расчета в окне калькулятора активировать переключатель "Обновить только выделенные объекты". Такая двухходовка неудобна по многим причинам - стандартный механизм выбора может далеко не все, калькулятор так же ограничен стандартным набором операций, и при необходимости выполнить несколько подобных присвоений вероятность ошибиться в выборках вырастает. Попробуем выполнить эту операции с помощью плагина. Создадим в слое УИКов новое поле - "type" типа string размером 100 символов. Для всех школ и гимназий присвоим значение "school", для библиотек - "library", а всем остальным - "other".

Код вычисления поля:

```
addr_ext_lower = <ADDR_V_EXT>.lower()
value = 'other'
if u'школа' in addr_ext_lower or u'гимназия' in addr_ext_lower:
    value = 'school'
if u'библиотека' in addr_ext_lower:
    value = 'library'
```

Результат выполнения:

Ø 💿	Attribute table - uik :: 0 / 3373 feature(s) selected							
120	NUMBER ^		ADDR_V_EXT	A TOTAL CONTRACTOR OF THE PARTY	ADDR_O_EXT			
0	1	ШУБИНСКИ	Типография «Наука»	Шубинский	Типографи	55.74679		
1	2	новински	Гимназия № 1513	Шубинский	Типографи	55.75117728		
2	3	НОВЫЙ АР	Мэрия г. Москвы	Новинский	Центр мол	55.75313664		
3	4	новински	Филиал школы № 1234	Трубниковс	кцсо	55.7562768.		
4	5	молчанов	Школа № 1234	Трубниковс	кцсо	55.75346996		
5	6	АФАНАСЬЕ	Школа № 1233	Афанасьев	Библиотек	55.7485172.		
6	7	АРБАТ УЛ.,35	ЦДА	АРБАТ УЛ.,35	ЦДА	55.74930		
7	8	КАМЕННАЯ	ΠK № 9	КАМЕННАЯ	ΠK № 9	55.75014		
8	9	АФАНАСЬЕ	Школа № 1233	Афанасьев	Библиотек	55.7494		
9	10	АРМЯНСКИ	педагогический коллед	ул. Покров	Обществен	55.7593665.		
10	11	трехсвяти	Общественный пункт ох	Старосадс	школа № 1	55.75555		
11	12	потаповс	школа №312	Чистопруд	BubliK	55.76071		
12	13	огородна	Дворец детей и молоде	огородна	Дворец де	55.76522		
13	14	ЧАПЛЫГИН	школа № 613	Фурманный	Совет вете	55.7631353.		
14	15	ЧАПЛЫГИН	школа № 613	Фурманный	Совет вете	55.7633852.		
15	16	КАЗЕННЫЙ	Школа № 330	Лялин пер.,	Участковы	55.7603113.		
16	17	ЛЯЛИН ПЕР	школа №1225	ул. Покров	Совет вете	55.75864		
17	18	воронцов	Клуб Высшей школы эко	Лялин пер.,	Участковы	55.7537716.		
18	19	РАДИО УЛ.,	Общественный пункт ох	4-ый Сыром	Театральн	55.75571		
19	20	гороховс	МИИГАИК	гороховс	МИИГАИК	55.76437		
20	21	БАСМАННА	школа № 320	Садовая-Ч	Обществен	55.76588		
21	22	БАСМАННА	школа №27	1-ый Басма	Обществен	55.77167		
22	23	ТОКМАКОВ	детская музыкальная ш	ТОКМАКОВ	детская му	55.76631		
23	24	АПТЕКАРСК	Библиотека №61	АПТЕКАРСК	Библиотек	55.76806		
			3 (23) (23) (28)					
			Look fo	ri (
Shor	w selected only	/ Search se	lected only 🗹 Case sensitive					

Так, за один проход мы сделали то, что стандартным калькулятором можно сделать за 5 шагов.

Замещение JOIN'а с последующим присвоением значений

По сути этот пример является развитием предыдущего. Допустим нам необходимо заменить значения одного из полей в строгом соответствии с какой либо таблицей (Обычно такая потребность возникает, когда необходимо заменить числовой код типа объекта на название этого типа). Стандартными функциями QGIS это можно сделать следующим образом:

- 1) Создать таблицу, с двумя полями код и его название
- 2) Добавить эту таблицу в проект и выполнить JOIN с нужным слоем
- 3) В калькуляторе полей присвоить новое значение обрабатываемому полю, равное соответствующему значение из присоединенной таблицы

Если же количество уникальных значений этого поля не так велико, то можно упростить процедуру, написав достаточно компактный код. Для примера переведем строчный тип, созданный в предыдущем примере, в числовой (иногда ведь приходится производить и обратную операцию :). Откроем окно плагина, из выпадающего списке выберем поле "type". В списке Fields выбираем это же самое поле и нажимаем кнопку "Get all", чтобы получить все значения этого поля.

Пишем глобальный код:

```
'library' : 1000,
'other' : 2000,
'school' : 3000
```

В достаточно простой форме мы определяем соответствие кодов и их названий. Если бы мы делали присвоение названий, то каждая строчка выглядела бы наоборот:

```
#пример!!!
table = {
1000 : 'library',
2000 : 'other',
3000 : 'school'
}
```

И код для вычисления поля:

```
value = table[<type>]
```

После выполнения, строчный код стал числовым (а вернее псевдочисловым, так как мы не создавали новое поле с типом int, а записали числа в существующее строчное поле):

	NUMBER	^	ADDR_V	ADDR_V_EXT	ADDR_O	ADDR_O_EXT	LAT
		1	ШУБИНСКИ	Типография «Наука»	Шубинский	Типографи	55.74679
77 44	-	2	новински	Гимназия № 1513	Шубинский	Типографи	55.7511772
		3	НОВЫЙ АР	Мэрия г. Москвы	Новинский	Центр мол	55.7531366
44		4	новински	Филиал школы № 1234	Трубниковс	кцсо	55.7562768
		5	молчанов	Школа № 1234	Трубниковс	кцсо	55.7534699
		6	АФАНАСЬЕ	Школа № 1233	Афанасьев	Библиотек	55.7485172
		7	АРБАТ УЛ.,35	ЦДА	АРБАТ УЛ.,35	ЦДА	55.74930
#		8	КАМЕННАЯ	ΠK № 9	КАМЕННАЯ	ΠK № 9	55.7501
44		9	АФАНАСЬЕ	Школа № 1233	Афанасьев	Библиотек	55.749
		10	АРМЯНСКИ	педагогический коллед	ул. Покров	Обществен	55.7593665
		11	трехсвяти	Общественный пункт ох	Старосадс	школа № 1	55.7555
***		12	потаповс	школа №312	Чистопруд	BubliK	55.7607
		13	огородна	Дворец детей и молоде	огородна	Дворец де	55.7652
		14	ЧАПЛЫГИН	школа № 613	Фурманный	Совет вете	55.7631353
		15	ЧАПЛЫГИН	школа № 613	Фурманный	Совет вете	55.7633852
4		16	КАЗЕННЫЙ	Школа № 330	Лялин пер.,	Участковы	55.7603113
		17	ЛЯЛИН ПЕР	школа №1225	ул. Покров	Совет вете	55.7586
7		18	воронцов	Клуб Высшей школы эко	Лялин пер.,	Участковы	55.7537716
		19	РАДИО УЛ.,	Общественный пункт ох	4-ый Сыром	Театральн	55.7557
		20	ГОРОХОВС	МИИГАиК	ГОРОХОВС	МИИГАИК	55.7643
***	:	21	БАСМАННА	школа № 320	Садовая-Ч	Обществен	55.7658
**		22	БАСМАННА	школа №27	1-ый Басма	Обществен	55.7716
	:	23	ТОКМАКОВ	детская музыкальная ш	ТОКМАКОВ	детская му	55.7663
44		24	АПТЕКАРСК	Библиотека №61	АПТЕКАРСК	Библиотек	55.7680
***					I maranes I		-

Переменная \$geom дает доступ к геометрии обрабатываемой строки. Переменная содержит объект типа QgsGeometry, работа с которым описана в <u>PyQGIS documentation</u>. Так же можно многое почерпнуть в <u>QGIS API</u>.

Простой пример. Необходимо обновить поля LAT и LONG в слое УИКов, после очередного обновления геометрий. Открываем форму расширения, из выпадающего списка полей выбираем поле LAT, запускаем редактирование. В поле кода вычисления поля пишем и запускаем обработку:

```
value = $geom.asPoint().y()
```

Не завершая редактирование, выбираем второй столбец - LONG, код обработки меняем на:

```
value = $geom.asPoint().x()
```

Запускаем обработку и завершаем редактирование.

В данном случае мы знаем что тип слоя - точечный, соответственно мы можем выполнить преобразование asPoint(), и получить координаты x() и y().

Для линейных объектов мы можем выполнять операцию расчета длины - \$geom.length(), а для полигональных расчета площади - \$geom.area().

Рассмотрим более сложный пример. Допустим мы хотим проверить адреса в слое УИКов с помощью обратного геокодинга (Не будем обращать внимание на уже существующие поля LAT и LONG,тем более что геометрии могли быть отредактированы). Воспользуемся арі, предоставляемое сервисом Nominatim, а так же модулем json стандартной библиотеки рython для разбора ответа сервера. Для сохранения результатов создадим новое поле ADDR_OSM типа string размером 250 символов. Будьте осторожны! Обработка большого количества записей с обращением к интернет сервисам может занять большое количество времени! Особенно для медленных соединений!

Глобальный код:

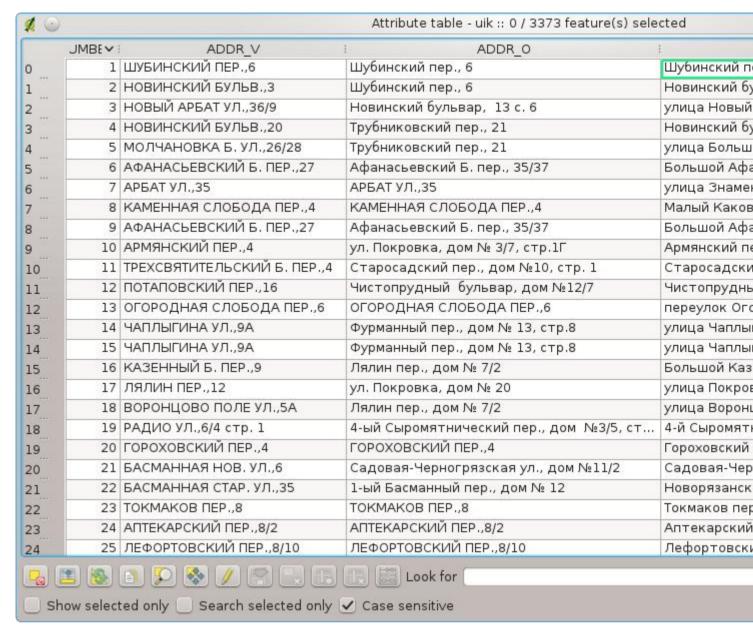
```
import json
import sys
import urllib2
url =
'http://nominatim.openstreetmap.org/reverse?format=json&lat=%f&lon=%f&zoom=18&addressde
tails=1'
```

Код вычисления поля:

```
try:
```

```
f = urllib2.urlopen ( url % ($geom.asPoint().y(),$geom.asPoint().x()))
    resp_str = unicode(f.read(), 'utf-8')
    resp = json.loads(resp_str)
    if resp.has_key('address'):
        addr = resp['address']
        if addr.has_key('pedestrian'):
            value = addr['pedestrian']
    else:
        value = addr['road']
    if addr.has_key('house_number'):
        value = value+', '+addr['house_number']
    else:
        value = 'Not found'
except:
    value = 'Error: '+unicode(sys.exc_info()[1])
```

Результат:



Дополнительные ссылки

Python:

• Python documentation

PyQGIS:

- PvQGIS Cookbook
- Перевод PyQGIS Cookbook

Другое:

- QGIS API
- Shapely Python библиотека для работы с геометриями
- Работа с векторными данными при помощи OGR и Python
- Создание скрипта на Python для ГИС GRASS
- Обратный геокодинг от Google

Контакты

Если вы нашли ошибку или у вас есть предложения по улучшению расширения - напишите на <u>email</u> или в <u>баг трекер на GitHub</u>.

Обсудить в форуме Комментариев — 11

Последнее обновление: 2014-05-14 23:49

Дата создания: 12.01.2012 Автор(ы): <u>Никулин Евгений</u>