

База географических названий GeoNames

[Обсудить в форуме](#) Комментариев — 1

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

Сводные данные и примеры для скачивания

[GeoNames](#) - свободно распространяемая географическая база данных имен объектов (топонимики). Лицензия - [Creative Commons Attribution](#). На настоящий момент содержит более 260 тысяч точечных объектов по территории России и более чем 8 миллионов по всему миру. Все объекты разделены на 9 классов (однобуквенный код) или 645 подклассов (2-5 буквенный код) ([полный список](#)):

- административное деление (A)
- гидросеть (H): каналы, заливы, озера, колодцы, моря
- территории (L): парки, поля, военные базы, охраняемые территории
- населенные пункты (P)
- дорожная сеть (R)
- точки (S): здания, банки, мосты, остановки транспорта, пересечения, фермы, заводы
- природные объекты (T): пляжи, цирки, пики, острова
- подводные объекты (U): каньоны, впадины, рифы
- леса (V): отдельные деревья, массивы, плантации, степные участки.

Содержание

- [1 Получение данных](#)
- [2 Данные на РФ](#)
- [3 Поля базы данных](#)
- [4 Импорт](#)
- [5 Недостатки](#)

Получение данных

Исходно данные доступны бесплатно для [прямого скачивания](#) в виде текстовых файлов, кодировка UTF-8. Примерный размер запакованного файла (РФ) 8 Мб, распакованного файла 49.9 Мб, генерируются из основной базы данных ежедневно.

Получить подготовленные данные можно выбрав одну из ссылок в таблице. Если вы работаете в ГИС скорее всего вам нужны share-файлы в кодировке UTF-8. Нажав на дату, можно посмотреть историю обновлений.

Также возможно использование веб-сервисов GeoNames. GeoNames предоставляет целый ряд [веб-сервисов](#), позволяющих работать с его данными и получать результаты в форматах XML и JSON. Каждый объект базы GeoNames имеет свой стабильный URI.

Помимо англоязычных названий, GeoNames собирает данные на разных языках, сведения о высоте, населении и т.п. из различных источников. Все координаты хранятся в WGS84 (World Geodetic System 1984). Пользователи могут редактировать и добавлять новые названия с использованием вики-интерфейса.

Данные на РФ

Срез данных на 20 августа 2012 г. содержит 345089 (263821 на 03.2009) объекта на территории РФ. Из них:

- 188247 названий населенных пунктов (PPL)
- 53120 названий ручьев (STM)
- 15364 названий озер (LK)
- 8938 названий высот (MT)
- 5407 названий ЖД станций (RSTN)
- и т.д.

Из 188247 объектов типа населенный пункт, альтернативное (включая кириллическое, т.е. реально названий на кириллице меньше) название имеют 165887 (88%, 49% на 03.2009) объектов. Данные о населении имеют 4399 пунктов (2%).

Поля базы данных

Параметр	Описание	Формат
geonameid	идентификатор записи в БД geonames	int
name	название объекта (utf8)	varchar(200)
asciiname	название объекта (ascii)	varchar(200)
alternatenames	альтернативные названия, разделитель - запятая	varchar(5000) для SQL Server
latitude	широта в десятичных градусах (wgs84)	
longitude	долгота в десятичных градусах (wgs84)	
feature class	http://www.geonames.org/export/codes.html	varchar(10)
feature code	http://www.geonames.org/export/codes.html	
country code	2-символьный код страны согласно ISO-3166	2 characters
cc2	альтернативный код страны, разделитель - запятая, 2-символьный согласно ISO-3166	60 characters
admin1 code	fipscode (в будущем будет изменен на isocode), isocode для США и Швейцарии, см. admin1Codes.txt для обозначений кодов	varchar(20)
admin2 code	код для административного деления второго уровня, county в США, см. admin2Codes.txt	varchar(80)
admin3 code	код для административного деления третьего уровня	varchar(20)
admin4 code	код для административного деления четвертого уровня	varchar(20)
population		integer
elevation	в метрах	integer
gtopo30	средняя высота по области 30'x30' (900mx900m) в метрах	integer
timezone	идентификатор часового пояса (см. timeZone.txt)	
modification date	date of last modification in yyyy-MM-dd format	

Импорт

Подробно импорт данных Geonames в ESRI Shape описан в [отдельной статье](#). Здесь приведены полезные настройки, которые могут помочь сэкономить время.

Для импорта текстового файла GeoNames с заголовком, можно добавить в начало этого файла следующую строку.

```
geonameid      name      asciiname      altnames      lat      long      fclass      fcode
      ccode      cc2      acode1      acode2      acode3      acode4      pop      elev      gtopo      tz
      mdate
```

Содержание файла csvt:

```
"integer","string(255)","string(255)","string(255)","Real(10.7)","Real(10.7)","string(10)","string(10)","string(2)","string(60)","string(20)","string(80)","string(20)","string(20)","integer","integer","integer","string(255)","string(255)"
```

Содержание файла vrt:

```
<OGRVRTDataSource>
  <OGRVRTLayer name="ru">
    <LayerSRS>WGS84</LayerSRS>
    <SrcDataSource>ru.csv</SrcDataSource>
    <GeometryType>wkbPoint</GeometryType>
    <GeometryField encoding="PointFromColumns" x="long" y="lat"/>
  </OGRVRTLayer>
</OGRVRTDataSource>
```

Недостатки

Главным недостатком данных является использование в качестве источника топонимики картографической базы Google, что ведет к распространению копирайтов Google и на эти данные. В связи с этим данные не рекомендуется использовать в таких проектах, как OpenStreetMap ([подробнее](#)).

К другим недостаткам можно отнести:

- отсутствия возможность отфильтровать альтернативный названия на латинице от кириллических
- большое количество ошибок и устаревшей информации
- использование двойных кавычек в названиях НП, что приводит к ошибкам при импорте данных (необходима предварительная обработка, чтобы от них избавиться)

Ссылки по теме

- [База географических названий GNS](#)
- [Перевод названий населенных пунктов VMap0](#)
- [GeoNames на вики OpenStreetMap](#)

[Обсудить в форуме](#) Комментариев — 1

Последнее обновление: 2014-05-15 01:36

Дата создания: 27.05.2009

Автор(ы): [Максим Дубинин](#)