

БИБЛИОТЕКА ГЕОКОДИРОВАНИЯ ДЛЯ MICROSOFT EXCEL

М.С. Грибова

Научный руководитель: доцент, к.т.н. А.Л. Богданов

Национальный исследовательский Томский государственный университет,

Россия, г. Томск, пр. Ленина, 36, 634050

E-mail: maria_94_7@mail.ru

GEOCODING LIBRARY FOR MICROSOFT EXCEL

M.S. Gribova

Scientific Supervisor: Docent, Cand. Sci. A.L. Bogdanov

Tomsk State University, Russia, Tomsk, Lenin str., 36, 634050

E-mail: maria_94_7@mail.ru

***Abstract.** The task of developing the tool that allows an analyst who is analyzing the real estate market to use online geocoding services is considered. It is assumed that the analyst uses Microsoft Excel and geocoding service offered by Yandex, the most popular data analysis tool and the most popular Russian Internet company.*

Введение. Рассматривается задача разработки инструмента, позволяющего аналитику, занимающемуся анализом рынка недвижимости, использовать в своей работе доступные через Интернет сервисы геокодирования. Предполагается, что аналитик решает свои задачи в табличном процессоре Microsoft Excel – популярном средстве анализа данных, и использует сервис геокодирования компании Яндекс – одной из самых популярных российских интернет-компаний.

Описание основных понятий. Геокодирование – это процедура определения координат (долготы и широты) и сведений о географическом объекте по его адресу. Процедура, выполняющая обратное преобразование, называется обратным геокодированием. В настоящее время существует несколько интернет-провайдеров сервиса геокодирования: Яндекс [1], Google [2], OpenStreetMap [3], PickPoint [4]. Выбор компании Яндекс в качестве поставщика сервиса геокодирования обусловлен наилучшей поддержкой русского языка её поисковой машиной.

Сервис компании Яндекс, реализующий геокодирование, называется Яндекс. Карты [1]. Доступ к сервису предоставляется на платной и бесплатной основе. В обоих случаях требуется пройти процедуру регистрации и получить ключ доступа (API-key). Бесплатная версия сервиса допускает не более 25000 запросов в сутки.

Для обращения к сервису можно использовать HTTP-протокол [5]. Например, запрос координат главного корпуса Томского государственного университета имеет вид:

<https://geocode-maps.yandex.ru/1.x/?geocode=Россия,+Томск,+Ленина,+36>

Ответ сервис возвращает либо в формате XML, либо в формате JSON [6]. Ответ сервиса содержит координаты запрашиваемого объекта (долготу и широту) и его почтовый адрес.

Имея полученную информацию, аналитик может использовать её для построения эконометрических моделей с учётом расстояний между объектами.

Постановка задачи. Задачей текущей работы является разработка библиотеки функций, реализуемых в виде надстройки Microsoft Excel, которые аналитик сможет применять для обращения к сервису геокодирования и вычисления расстояний между объектами. К библиотеке функций предъявляется дополнительное требование: результаты ранее выполненных запросов должны сохраняться в кэше (локальной базе данных), что позволит уменьшить время выполнения операции геокодирования и сократить количество запросов к серверу.

Описание библиотечных функций. Реализованная библиотека содержит следующие функции:

1. getLocation(Description As String, Optional IgnoreCache As Boolean = False, Optional UpdateCache As Boolean = True)

Функция, возвращает массив, содержащий координаты и почтовый адрес объекта, заданного строкой описания Description. Параметр IgnoreCache является необязательным, при значении True предписывает выполнить запрос к серверу геокодирования, игнорируя содержимое кэша. По умолчанию IgnoreCache = False. Параметр UpdateCache является необязательным, при значении True предписывает обновить содержимое кэша обновлёнными данными, полученными от сервера геокодирования. По умолчанию UpdateCache = True.

2. getLat(Location), getLon(Location), getPostAddress(Location)

Функции извлекают соответственно широту, долготу и почтовый адрес объекта из параметра Location, который является результатом работы функции getLocation().

3. getURL2MAP(Lon As Double, Lat As Double), getURL2MAP(Description)

Функции возвращают адрес URL, который можно ввести в адресную строку браузера и перейти на страницу карты с меткой объекта, заданного соответственно координатами (Lon, Lat) или строкой описания (Description).

4. getDistance1(Lon1 As Double, Lat1 As Double, Lon2 As Double, Lat2 As Double), getDistance2(Description1 As String, Description2 As String)

Функции возвращают расстояние в метрах между двумя объектами заданными соответственно координатами и (Lon1, Lat1, Lon2, Lat2) или описаниями (Description1, Description2).

5. setAPIKey(Key As String), getAPIKey()

Функции устанавливают новое и возвращают текущее значение ключа доступа соответственно.

Описание работы кэша. Кэш представляет собой хранилище данных, реализованное в виде базы данных Microsoft Access, и библиотеки доступа к хранилищу. ER-модель хранилища данных представлена на рис. 1. Хранилище состоит из двух таблиц: ADDRESS, которая хранит описания объектов (адреса, которые использовал аналитик на рабочем листе таблицы Excel) и LOCATION, которая хранит ответы сервера геокодирования. Библиотека доступа к хранилищу кэша реализует алгоритм, блок схема которого представлена на рис. 2.

Пример использования. В целях демонстрации возможностей реализованной библиотеки на рис. 3 представлены примеры работы основных функций getLocation() и getDistance1().

Заключение. В данной работе реализована библиотека доступа к сервису геокодирования компании Яндекс в виде надстройки Microsoft Excel. Уникальной особенностью библиотеки является наличие кэша, который позволяет существенно уменьшить время выполнения операции геокодирования

и сократить количество запросов к серверу. Разработанная библиотека будет использована для построения, анализа и исследования свойств эконометрических моделей цен жилой недвижимости.

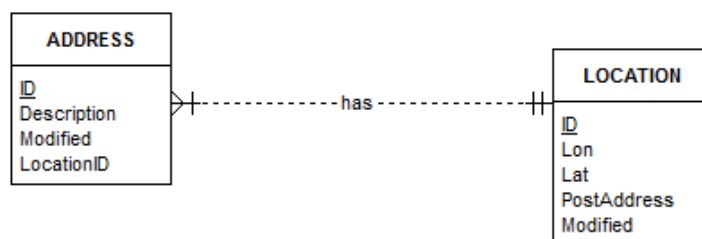


Рис. 1. ER-модель хранилища данных кэша

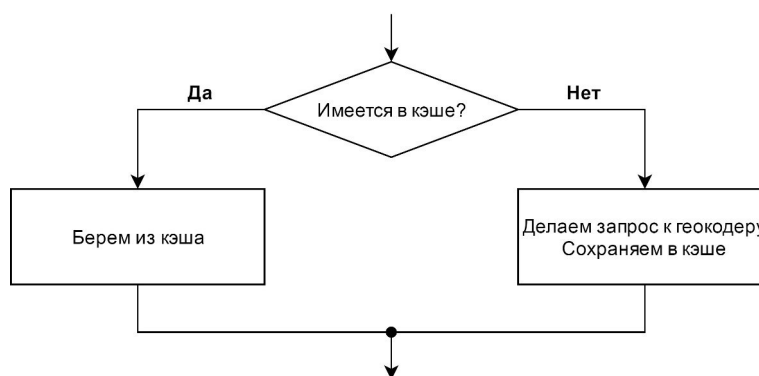


Рис. 2. Обобщенная схема работы с хранилищем данных кэша

	A	B	C	D
1	Исходный адрес	Долгота	Широта	Почтовый адрес
2	Томск, Ленина 36	84,947586	56,469478	Россия, Томск, проспект Ленина, 36
3	Томск, Ленина 30	84,950173	56,465405	Россия, Томск, проспект Ленина, 30
4				
5	Расстояние в метрах:	480,10098		
6				

Рис. 3. Пример работы библиотеки функций

СПИСОК ЛИТЕРАТУРЫ

1. Геокодирование Яндекс.Карт [Электронный ресурс]. – Режим доступа: <https://tech.yandex.ru/maps/doc/geocoder/desc/concepts/About-docpage>. – 15.03.17.
2. Геокодирование Google Maps Geocoding API [Электронный ресурс]. – Режим доступа: <https://developers.google.com/maps/documentation/geocoding>. – 15.03.17.
3. Геокодирование OpenStreetMap [Электронный ресурс]. – Режим доступа: <http://nominatim.openstreetmap.org>. – 15.03.17.
4. Геокодирование PickPoint [Электронный ресурс]. – Режим доступа: <https://pickpoint.io>. – 15.03.17.
5. Параметры HTTP-запроса к геокодеру Яндекса [Электронный ресурс]. – Режим доступа: https://tech.yandex.ru/maps/doc/geocoder/desc/concepts/input_params-docpage. – 15.03.17.
6. Ответ сервиса геокодирования Яндекс [Электронный ресурс]. – Режим доступа: https://tech.yandex.ru/maps/doc/geocoder/desc/concepts/response_structure-docpage. – 15.03.17.