

2016

# Address Standardizer

Komal Sorathiya  
*Grand Valley State University*

Follow this and additional works at: <http://scholarworks.gvsu.edu/cistechlib>

---

## Recommended Citation

Sorathiya, Komal, "Address Standardizer" (2016). *Technical Library*. Paper 237.  
<http://scholarworks.gvsu.edu/cistechlib/237>

This Project is brought to you for free and open access by the School of Computing and Information Systems at ScholarWorks@GVSU. It has been accepted for inclusion in Technical Library by an authorized administrator of ScholarWorks@GVSU. For more information, please contact [scholarworks@gvsu.edu](mailto:scholarworks@gvsu.edu).

# Address Standardizer

By

Komal Sorathiya

April 28, 2016

# Address Standardizer

By

Komal Sorathiya

April 28, 2016

A project submitted in partial fulfillment of the requirements for the

degree of

Master of Science

in

Computer Information Systems

at

Grand Valley State University

April 28, 2016

## **ACKNOWLEDGEMENTS**

My special thanks to Professor Jonathan Engelsma for all of his great commitment to this project and my career. It has been a tremendous pleasure to meet him and work under his supervision. He helped me a lot on every aspect of this project. Sometimes, I cannot find solution for problem even by research. He helped me out for all difficulties.

Also Adriana Paz, who is a Senior Programmer at GVSU's Johnson Center helped me for my project and provide me guidance for the project. Jeremy Pyne, GIS manager at GVSU's Johnson Center help me to decide flow of project. He also gives me an idea regarding requirement of project.

# TABLE OF CONTENTS

Abstract.....	5
Introduction.....	6
Background, Related Work and Requirements.....	8
Technology Required.....	9
Implementation and Results.....	12
Learned new things.....	16
Conclusions and Future Work .....	17
Bibliography .....	18

## **ABSTRACT**

I am working in Johnson Center at GVSU. Majority projects from Johnson center are using Geocodes for address for projects. Geocode is useful to display address in Google map. We can have geocode for address only if address is correct. At work, we got address from client for school, house location, organization location etc. But, sometimes people made mistakes by typing wrong address in form, then we will not get geocode for that address. And, we cannot display that address in map. So, we need to find that all address from master address file for geocode. For single address we need to search in whole file. This task is time consuming. I got an idea from Johnson center for my project. I decided to make one tool which can convert wrong address to correct address and it will find geocode for that address. Tool will find Geocodes for around 100 address in 1 minute instead of couple of hours. It is a tool to conversion of incorrect address to correct address and find Geocodes for that address and provides output in the form of excel file.

## INTRODUCTION

Address standardizer is a tool to conversion of incorrect address to correct address and find Geo-codes for that address. It standardizes address specifically for Kent County.

Address will be divided into 6 parts. House Number, Pre-direction, Street name, Street type, Suffix direction, zip code. I am going to use Master table with all correct address for making incorrect address to correct one.

For standardization I used following condition to implement algorithm,

1. If street name with spelling mistake, then we can make assumption by character matching in order.
2. If Suffix direction is missing, we can find address by house number. If House number is same on both side, then we can check on zip code.
3. If Street type with that street name is not exists, then we can change that to the street name with different street type.
4. If Zip code is wrong, then we can find closer match of zip code, because we know that this zip code does not exists in the specific County.
5. If House number is wrong for example 14000, then we can make assumption that this address is junk address because road ends at House number 1500.

We have to check in master address table and assign x and y co-ordinate for that address.

There are main three steps for the tool.

1. We need to upload one csv file to the server, file should contain list of rows with field names like id, address, city, zip.
2. Then, tool will take that file from the server and copy to the database. After that, it will copy value from one database to the another database table. Another table has all other fields like pre direction, house number, street type, street name, street direction, street

type. It will split address into all mentioned field. Then, it applies algorithm to find geocodes.

3. This stage will display output in the tabular form. So, we can see what geocode is generated by tool for each row. After that, there is a link to download that data in excel format. We can download excel file from that button.

## **BACKGROUND, RELATED WORK AND REQUIREMENTS**

Established in 1992, the Dorothy A. Johnson Center for Philanthropy is an academic center within the College of Community and Public Service at Grand Valley State University<sup>1</sup>. Like many academic centers, we offer original research and expertise on the latest theories in our field. Yet when we talk about “doing good,” we mean more than theory<sup>1</sup>. The Johnson Center puts research to work, with and for professionals across the country<sup>1</sup>.

Geocode is used to display address in map for many applications developed by Johnson Center. User needs to find correct address and geocode from master addresses manually. Master addresses is collection of correct address which has geocode present for each address. Geocode is nothing but x and y co-ordinate to display or locate address in map.

There are many projects which are using this geocodes. If we get address from any organization, then we need to find geocode for that address. Sometimes, we get wrong address from organization like spelling mistake, wrong zip code etc. Then, we need to correct that address on the basis of some criteria. There is a lots of work to do manually. It will require time to do all this kind of manual work.

For all their need, I decide to create one tool which can help to correct the incorrect address and find geocode for all possible correct address. Therefore, we can display that address in map using geocode generated by tool. This process will save lots of time and helps task to be done in much easier way.

## TECHNOLOGY REQUIRED

For implementation of project, I used different kind of technology and software. The technology and software are following:

1. JQUERY:

jQuery is a fast, small, and feature-rich JavaScript library. It makes things like HTML document traversal and manipulation, event handling, animation, and Ajax much simpler with an easy-to-use API that works across a multitude of browsers. With a combination of versatility and extensibility, jQuery has changed the way that millions of people write JavaScript<sup>2</sup>.

2. ECLIPSE:

Eclipse is famous for our Java Integrated Development Environment (IDE), but C/C++ IDE and PHP IDE are pretty cool too. We can easily combine language support and other features into any of its default packages, and the Eclipse Marketplace allows for virtually unlimited customization and extension<sup>3</sup>.

3. JAVA:

Java is a general-purpose computer programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA)<sup>4</sup>.

4. JSP:

Java Server Pages (JSP) is a technology that helps software developers create dynamically generated web pages based on HTML, XML, or other document types. Released in 1999 by Sun Microsystems, JSP is similar to PHP and ASP, but it uses the Java programming language<sup>4</sup>.

5. ORACLE DATABASE:

An Oracle database is a collection of data treated as a unit. The purpose of a database is to store and retrieve related information<sup>5</sup>.

6. ORACLE CLIENT:

Oracle Instant Client enables development and production deployment of Oracle Database applications. It is used for popular languages and environments including Node.js, Python and PHP, as well as providing access for OCI, OCCI, JDBC, ODBC and Pro\*C applications<sup>6</sup>.

7. SQL DEVELOPER:

Oracle SQL Developer is a free integrated development environment that simplifies the development and management of Oracle Database in both traditional and Cloud deployments. SQL Developer offers complete end-to-end development of your PL/SQL applications, a worksheet for running queries and scripts, a DBA console for managing the database, a reports interface, a complete data modeling solution, and a migration platform for moving your 3rd party databases to Oracle<sup>6</sup>.

8. APACHE TOMCAT:

The Apache Tomcat<sup>®</sup> software is an open source implementation of the Java Servlet, Java Server Pages, Java Expression Language and Java WebSocket technologies. The Java Servlet, Java Server Pages, Java Expression Language and Java WebSocket specifications are developed under the Java Community Process<sup>7</sup>.

9. HTML:

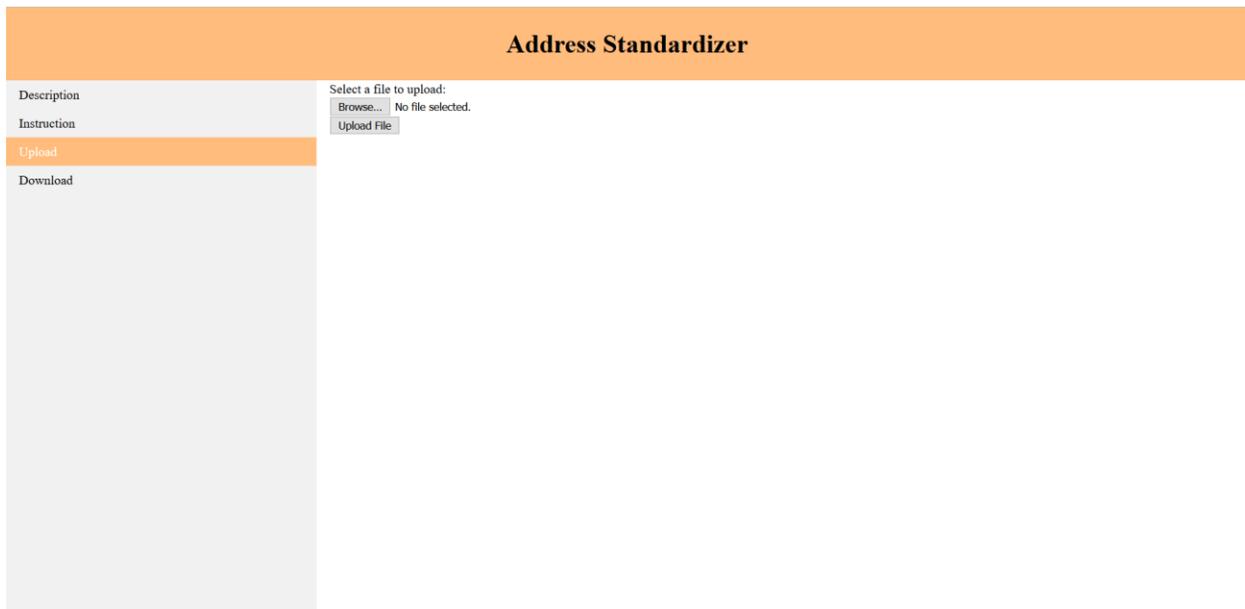
HyperText Markup Language, commonly abbreviated as HTML, is the standard markup language used to create web pages. Along with CSS, and JavaScript, HTML is a cornerstone technology used to create web pages, as well as to create user interfaces for mobile and web applications<sup>4</sup>.

## 10. CSS:

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language. Although most often used to set the visual style of web pages and user interfaces written in HTML and XHTML, the language can be applied to any XML document, including plain XML, SVG and XUL, and is applicable to rendering in speech, or on other media. Along with HTML and JavaScript, CSS is a cornerstone technology used by most websites to create visually engaging webpages, user interfaces for web applications, and user interfaces for many mobile applications<sup>4</sup>.

## IMPLEMENTATION AND RESULTS

There are mainly four modules in project.

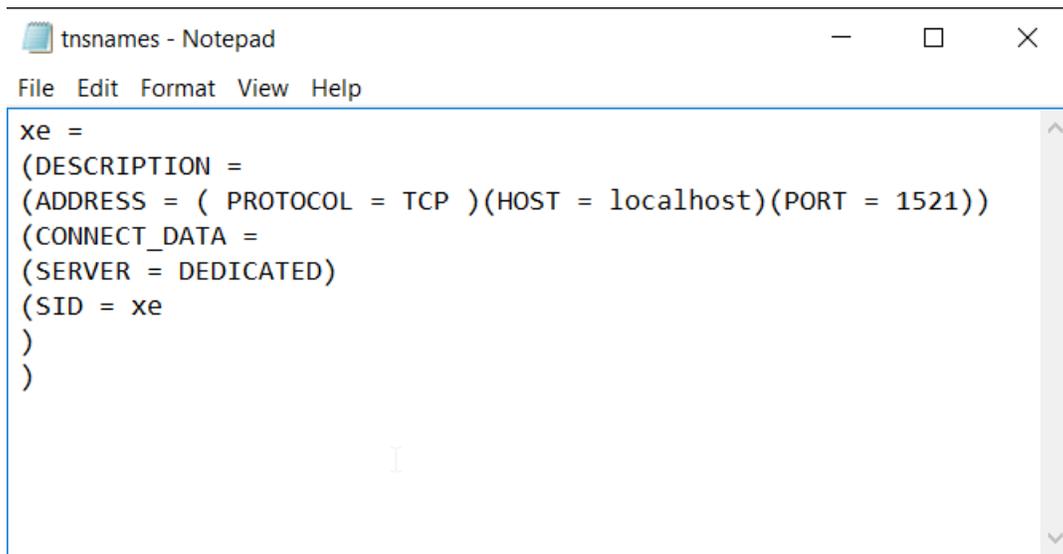


We can take reference from the above image to understand working process of all modules. In the Description module, there is for the short description about the project. In the Instruction module, there is instruction to walk through in tool. If we click on upload link from the project, then we can see screen same like above image.

If we click on the browse button, then one screen will appear for selection of file. From that we have to select csv file with address list. File should be with four field names like id, address, city, zip. After selection of file, name of file will appear right next to browse button. Then you have to click on upload button to upload that csv file with data in server. The file will be uploaded to the Apache Server.

From server, the file will be copied automatically to the database with the help of batch file. I have created one batch file. The batch file is in the server. Batch file has sqlloader command to copy data from csv file to table in oracle database. But for connection to database, we need to

provide all information related to oracle connection in the tnsnames.ora file. That file is located in oracle client home location. Oracle client will helpful to communicate between oracle database and the application. Following is the image for tnsnames.ora file. We can see that there is information related to oracle database connection in that file.



```
tnsnames - Notepad
File Edit Format View Help
xe =
(DESCRIPTION =
(AADDRESS = ( PROTOCOL = TCP )(HOST = localhost)(PORT = 1521))
(CONNECT_DATA =
(SERVER = DEDICATED)
(SID = xe
)
)
```

Now, database table has record which we uploaded using the application. There are couple of oracle procedure in the database which will help to implement algorithm for standardization of data.

Data will be copied from that temporary table which has four field in that like id, address, city, state, zip. That table name is Address\_1. We can look into following image for the better understanding of table structure in database.

SYSTEM.ADDRESS_1	
ID	VARCHAR2 (20 BYTE)
STREETADDRESS	VARCHAR2 (200 BYTE)
CITY	VARCHAR2 (100 BYTE)
STATE	VARCHAR2 (100 BYTE)
ZIP	VARCHAR2 (50 BYTE)

SYSTEM.ADDRESS_2	
ADDRESS_ID	VARCHAR2 (20 BYTE)
STREETADDRESS	VARCHAR2 (200 BYTE)
CITY	VARCHAR2 (100 BYTE)
STATE	VARCHAR2 (100 BYTE)
ZIP	VARCHAR2 (50 BYTE)
HN	VARCHAR2 (100 BYTE)
PD	VARCHAR2 (100 BYTE)
PT	VARCHAR2 (100 BYTE)
SN	VARCHAR2 (200 BYTE)
ST	VARCHAR2 (100 BYTE)
SD	VARCHAR2 (100 BYTE)
TEMP_SA	VARCHAR2 (200 BYTE)
PPN	VARCHAR2 (100 BYTE)
X	VARCHAR2 (100 BYTE)
Y	VARCHAR2 (100 BYTE)

TEMP\_ADDRESS5\_IDX1 (TEMP\_SA)

SYSTEM.MASTER_ADDRESS	
ADDRESS	VARCHAR2 (100 BYTE)
ADDR_HN	VARCHAR2 (100 BYTE)
ADDR_PD	VARCHAR2 (100 BYTE)
ADDR_PT	VARCHAR2 (100 BYTE)
ADDR_SN	VARCHAR2 (100 BYTE)
ADDR_ST	VARCHAR2 (100 BYTE)
ADDR_SD	VARCHAR2 (100 BYTE)
ADDR_ZIP	VARCHAR2 (100 BYTE)
ADDR_CITY	VARCHAR2 (100 BYTE)
PPN	VARCHAR2 (100 BYTE)
X	VARCHAR2 (100 BYTE)
Y	VARCHAR2 (100 BYTE)

Then, we data records will be copied to Address\_2 table from Address\_1 table.

1. Find\_housenumber oracle procedure from database will grab house number from the address.
2. Find\_predirection oracle procedure from database will take pre-direction from the address like south, north, west, east.
3. Find\_streettype oracle procedure from database will take street-type from the address like drive, ave, rd etc.
4. Find\_streetdirection oracle procedure from database will grab street-direction from the address like SW, NW, SE, NE.
5. Find\_geocodes oracle procedure from database will compare address from Master\_address table to Address\_2 table. Master\_address table has around 221923 Records with correct addresses and geocodes. If address matches, then it will return geocode and it will update the records in Address\_2 table.

Last, you can now see data records in Download module. Upload module link is indicated in Blue. Download module link is indicated in Red. So you need to click on Download module link which is indicated in red, then you can see data records with geocode. If you want to download that data records in excel file, then you can click on Export to Excel button which is located at the right side in green.

Address Standardizer												
Description											Export to Excel	
Instruction												
Upload												
Download												
Address	City	State	Zip	House #	Pre direction	Street Name	Street Type	Street Direction	PPN	X	Y	
14650 TYRONE AVE NW	KENT CITY	MI	49330	14650		TYRONE	Ave	NW	410127100016	-85.7310067	43.2315913	
107 WYNSONG CT NW	KENT CITY	MI	49330	107		WYNSONG	Court	NW	410133154008	-85.74825538	43.21338013	
156 WYNDONG CT NW	KENT CITY	MI	49330	156		WYNDONG	Court	NW	null	null	null	
17822 RED PINE DR NW	KENT CITY	MI	49330	17822		RED PINE	Drive	NW	410101200011	-85.68029438	43.29190009	
17720 RED PINE DR NW	KENT CITY	MI	49330	17720		RED PINE	Drive	NW	null	null	null	
680 22 MILE RD NW	KENT CITY	MI	49330	680		22 MILE	Road	NW	410101100043	-85.68892949	43.29147515	
710 22 MILE RD NW	KENT CITY	MI	49330	710		22 MILE	Road	NW	null	null	null	
17555 RED PINE DR NW	KENT CITY	MI	49330	17555		RED PINE	Drive	NW	null	null	null	
2175 18 MILE RD NW	KENT CITY	MI	49330	2175		18 MILE	Road	NW	410127100013	-85.72516091	43.2336831	

## LEARNED NEW THINGS

- ORACLE DATABASE INSTALLATION:

For the installation of oracle database, I downloaded couple of different version of database and I tried to install one by one. At the end, I can successfully install the require version of oracle database for application. But you need to remember username and password which asked by software at the time of installation. Otherwise, it will make thing difficult, you need to reset password for user.

- ORACLE CLIENT CONFIGURATION:

I need oracle client to be installed in my system for application to be work when it's in server. I don't need oracle client If I am running apache server from my local computer. So, I didn't install oracle client. Before installation of oracle client, I can successfully upload data records to the database from data file present in server. But after installation of oracle client, it was difficult. So, I researched something relevant to that whole process, and I came to know that I need to set proper configuration information relevant to oracle connection in tnsnames.ora file.

- It was hard time when I was trying to connect to GVSU network, I have windows 10 installed in my computer. So, network connection is not work normally in my system. I have to install some software to connect to the GVSU network. I watched youtube video for the instruction to connect to network. But it did not work. I also used another system to connect to network, but it did not work again. At the end, I called to IT helpdesk, and they told me that I need to install software called Pulse Secure. Then, I can successfully connect to the GVSU network.

## CONCLUSION AND FUTURE WORK

With the help of all technology and software, I successfully created tool to generate geocodes for the addresses. Now, we can download that file with all data records by single click on Export to Excel link. This application will save lots of time; it will do task of couple of hours in 1 minute. So, for sure, we can use that time to some other important work, instead of looking into individual address and finding that its correct or not. Application will automatically copy geocodes values of relevant record to data field. Therefore, we do not need to worry about copy all that data manually and one by one and paste that to the output file. However, there are still some points which we can improve or implement for the application.

- Multiuser functionality will allow multiple user to upload file at the same time to the server, and let application to generate proper output file with geocodes in data records.
- Multi-extension file support is a feature in that we can upload with different types of extensions like csv, excel etc.

## BIBLIOGRAPHY

1. <http://johnsoncenter.org/about/>
2. <https://jquery.com/>
3. <https://eclipse.org/ide/>
4. <https://en.wikipedia.org>
5. <https://docs.oracle.com>
6. <http://www.oracle.com/>
7. <http://tomcat.apache.org/>