

USGS Database Analyzer

A program for downloading and manipulating USGS earthquake catalogs.

Instruction Manual

Introduction

The USGS used to publish their raw earthquake catalogs via FTP server which made it easy for anyone to download the entire catalog for research purposes. However changes to the USGS front end now only provide access to the catalog through a web interface via their earthquake map page. It is possible do a search of the catalog using any chosen parameters and receive a file with the requested data in various formats.

Unfortunately, this interface limits the number of records returned to a maximum of 20,000. The scope of the catalog at the time of this writing encompasses more than 3.6 million records. If one wishes to download large amounts or the entire catalog, this limitation makes it difficult to accomplish.

The solution, a 'bot' program which can automatically download the requested records in smaller chunks. The web interface is published and fully documented by the USGS, so it was just a matter of writing a program that can submit a properly formatted URL to the server and then locally store the returned data.

That's how this program was born.

The program has additional features for manipulating the downloaded files, such as converting it to a smaller binary format, saving to new formats or files, and applying filters to those operations. There are also the beginnings of tools for analyzing the records, such as binning by magnitude, both for the entire data set and on a by year basis.

IT NEEDS TO BE NOTED THAT THIS SOFTWARE IS IN IT'S INFANCY AND HAS NOT BEEN MADE FULLY ROBUST AGAINST USER ERROR. PLEASE BE CERTAIN OF YOUR SETTINGS AS IT MAY LEAD TO ERRONEOUS RESULTS OR POSSIBLY CAUSE THE PROGRAM TO CRASH. USE OF THE PROGRAM IS 100% YOUR RESPONSIBILITY AND NOT THAT OF ANYONE ELSE.

Legal Stuff

NOT YET DECIDED WHICH LICENSE TO USE. FOR NOW CONSIDER IT FULLY COPYRIGHTED AND DISTRIBUTED BY INVITATION ONLY, NOT TO BE REDISTRIBUTED.

THIS SOFTWARE IS "AS IS" AND TO BE USED AT YOUR OWN RISK.

SHIT HAPPENS.

Installation

The program is self contained and does not need to be 'installed' into Windows in order to use it. Just unzip the compressed files into the directory location of your choice. Make sure all files are present and in the same directory. There are no sub-directories.

You should have the following files. They all need to be together in the same directory.

cacert.pem
defaults
libcrypto-1_1.dll
libcurl.dll
libssh2.dll
libssl-1_1.dll
USGSAnalyzer.exe

To run the program, just double click on 'USGSAnalyzer.exe'. Or for convenience you could make a desktop icon pointing to that file.

To uninstall the program, just delete the files.

Quick Start

Let's download some data. First, please make sure you are connected to the internet.

Adjust the filters as follows: Start Date = 1/1/1980 and End Date = 12/31/1980. You can either type directly in the text box or click the down arrow to bring up a calendar.

Select 'earthquake' in the event types list.

Leave all other settings at their default.

Click on the 'To Memory' button in the Raw Download section. After a few seconds you should see in the status window what the filter settings were for the request, and a raw listing of the returned data. Since we left the request type to 'Count', what is returned is simply the count of how many earthquakes that are in the data base for 1980. I got 23698.

Now click on 'All To Files'. You will be prompted for where to save your data. Once you choose your location, the download process will begin. You will see information scrolling along as to your progress. It only took about a minute to download the data on a 1 megabit WiFi connection.

Look in your chosen directory. You should see 12 files, named 1980-01-01.csv through 1980-12-01.csv. Check them out in a text editor if you like.

If that's all you need, then read no further. For an introduction to the binary files, read on.

Click on 'CSV to BIN' in the Convert section. You will first be prompted for the directory where you just saved you CSV files. Then you will be prompted for the location and name of your upcoming new BIN file. I used the same directory, and named in 1980.bin. Once done the status display will show the progress of the conversion. It should be pretty quick.

Look in your chosen directory again. You should now see a file called 1980.bin.

To do something with this file, we first have to load it. Click on 'Load File' in the Binary Files section, selecting the file you just loaded. This loading process should be even faster yet.

Now you can play with some of the other settings in the Binary Files section. Try the 'Bin by Mag' button first. Now you know how many of each whole magnitude bin there are in 1980. There appear to be a total of six quakes magnitude 7 or greater that year.

'Count' the quakes. Again we see there are 23698 on of them.

But now change some of the filter settings, such as magnitudes, or depths. Now see how many you get. If by chance you want to make a new file with some particular filter settings, you can save it either as a BIN or a single large CSV.

I know there isn't a lot in this program, but it was meant to serve a fairly singular purpose – to easily download large portions of or the entire USGS earthquake catalog, and to manipulate that data into usable formats.

But, there are plans for more.... :)

Menu Bar

File

- Quit – obvious

Defaults

- Load Defaults
- Save Defaults
- Reset Defaults
- Set Load Directory
- Set Save Directory

The 'defaults' here are the user settings that the program starts up with. Things like the filter parameters, check boxes, date ranges, etc...

If there are different settings you prefer to start with, you may save them. That way the program will always start up with your references.

While using the program you may want to go back to these settings, you may load them back.

Should you decide to revert back to the factory defaults, you may reset them.

You may also set your preferred default loading and saving directories. When the program prompts you where to load or save a file, it will default to your preferred locations, which can save you time navigating from the root directory every time.

Debug

- Verbose

This is off by default. Turning it on causes the program to display a lot of debugging information in the console window. Useful for the programmer who wrote this software when resolving issues.

Help

- About – obvious, but not much here at the moment.

Program Control

Quit

Quits the program. This can be invoked at most any time. Should this be interrupting a process in progress, the program should ask if you're sure, giving the user the opportunity to cancel quitting should this have been by mistake.

STOP!

Pressing this button will abort whatever function is in progress. Useful if you accidentally request a task which will take to long, or a long task where you realized you used the wrong settings.

Progress Bar

Just a little progress bar to show that the program is actually doing something. Used on most functions.

Status Window

Lots of stuff can be displayed here. Not only can you view data, listings of events, and data about the events, but also status messages, information about the progress of the current function, and of course, error messages.

How much status information and error messages are displayed is controlled by turning on "Verbose" in the debug menu. This is really only useful to the programmer when resolving problems.

The status window is a well featured text area. You can select and highlight information, allowing you to copy it the clipboard. You can also type in this area, although that may only be useful for you to jot down some notes. If you select everything and hit 'delete' on the keyboard, you can clear the display.

Nothing is saved from the display. All information is lost when the program is terminated.

TO DO/CONSIDER:

- saving display contents to file
- clear display with a button

Filters

Many filter options are available. NOTE: these filters are active on processes at all times except where a user selected check box is provided to enable/disable the filters.

Please note that there are default values which if used will effectively disable that filter when the request is made to the USGS server. The purpose is so that records which have left that data field empty will still be returned. Otherwise the USGS server would eliminate these records. This is most significant for older records. NOTE: This does not affect databases loaded into memory. Only for download requests to the USGS server.

All of these filter settings can be saved and reloaded from the main menu as defaults.

The available filters are,

Start Date – This is the start date of any process. The time assumed is always 00:00:00.000, midnight on the morning of the selected date.

End Date – The end date of any process. The time assumed is always 23:59:59.999, just shy of midnight of the night of the selected date.

The smallest time grain available is therefore one full day, 24hrs. If both dates are the same this is what you get. In other words, the date range is “inclusive”. To select all of January 1969, use 1/1/1969 and 1/31/1969. If you ended with 2/1/1969, you would also get events that occurred on February 2nd.

Minimum and Maximum Magnitude – The magnitude range you wish to work with. The default values are -99.0 and 10.0. Using a minimum magnitude of less than -99, or a maximum magnitude greater than 10 disables those filters for USGS download requests

Latitude and Longitude – North and South latitudes, and West and East longitudes. West longitudes are always given as negative values. The default values, respectively, which disable the USGS server filter are, 90.0, -90.0, -180.0, and 180.0.

Maximum and Minimum Depth – Limits the depth range to look at. This is useful for looking for deep focus quakes. The default values which turn off the USGS server filters are 1000km and -10km. Note that since we are talking depth, then negative values actually mean above sea level. There are events in the record with negative values.

Event Type – This is a predetermined list of already known event types in the USGS catalog. You may select one type at a time. “all” of course includes everything. The list is approximately in the order of greatest to least number of events. Useful for looking at only earthquake events, or studying other seismic events.

Not that this list is updated by the “List Types” function in the binary files section. If the binary file loaded only has earthquakes in it, “List Types” will cause this list to be reset to only have “all” and “earthquakes”. You would have to reload your defaults or reset defaults to get the list back.

TO DO/CONSIDER:

individual and/or group reset buttons – but may result in cluttered user interface

Raw Download

This is where you can download information and event records from the USGS server.

Note that when downloading from the server, the filters are applied. This way you can limit your request to only the exact information you desire.

CAUTION: Requests that are too large or take the server too long to process will fail. The server will return nasty error messages and those will display in the status window. Remember, you can always “STOP!” something in progress if you realize you made an error on your filter selections that may result in a server problem.

Text / CSV – This controls which format the returned data comes back in. The format of these files is described in the appendix.

Count / Query – These are the two fundamental types of requests to the USGS server.

“Count” simply returns the number of events found that meet the search criteria.

“Query” returns the actual data records themselves.

CAUTION: The USGS server limits requests to 20,000 records maximum. You can request a count for more than this and the server will tell you how many records there are. But if you do a query for more than the 20k limit, the server will return an error.

To Memory – The requested data is returned to program memory and displayed once fully received. It is not stored to a file, nor is there an option to. It is simply so you can view small amounts of data, or to simply find out how many records there are.

Show Memory – Simply displays whatever data is currently in the memory buffer.

CAUTION: The 20,000 record limit is quite large. Although the status window is capable of displaying that much data, it will take a loooong time to do so.

All To Files – This is the real gold nugget of the program. This function will download the event data from the USGS server and save it to files. The user will be prompted for the save location.

Remember that the filters are applied to this function. That way you can limit the request to only that data which you desire.

Downloads are broken up into one month segments, and as such are saved the same way. For example, if you request a date range of 1/1/1969 to 12/31/1969, the data will be requested by the month and saved into 12 separate CSV files, named with the first date of the month. For example “1969-01-01.csv”.

The program is smart enough to ask how many records are being downloaded first. If a particular month returns more than 20,000 records, then the request is broken down into 1 week chunks, and the file names will reflect the first date of those chunks.

Convert

Data conversion tools.

Right now this is limited.

CSV to BIN – Convert a directory full of CSV files into one single BIN file. The user is prompted for load directory and save file locations. NOTE: ALL CSV files in the selected directory will be converted.

BIN to CSV – NOT CURRENTLY IMPLEMENTED. Intent is to convert a BIN file to a single CSV file. This can be accomplished the 'Binary Files' section by first loading the binary, then saving it to CSV. See the manual section of "Binary Files" for more info.

Using Filters – Allows the user to apply filters to the conversion process. For example, you have a CSV files with all event types, but only want a BIN containing a specific event type, such as nuclear explosions, or earthquakes.

NOTE: the binary format this software makes and uses contains less information than either the text or csv formats that the USGS returns the data in. The binary format is fully documented in the appendix so that it can be used by other programs. But here is a quick list of the data in that format,

- date/time is ISO8601 format
- date/time in Julian decimal format
- latitude
- longitude
- depth
- magnitude
- event type

All other data is discarded in the conversion process.

TO DO/CONSIDER:

- select which files to convert rather than default to all
- implement BIN to CSV – to one single CSV or to multiple.

Binary Files

These functions allow you to work with files of the binary format this program can produce.

NOTE: the binary format this software makes and uses contains less information than either the text or csv formats that the USGS returns the data in.

Load File – Pretty self explanatory.

Save to BIN – Allows the user to save the loaded binary data. The usefulness of this is that the filters are applied to this process. Some examples would be to save only deep focus quakes, or only sonic booms, or events in a specific region on the planet defined by latitude and longitude.

Save to CSV – Same idea as above but to USGS CSV format. NOTE: since binary files contain less information than the USGS CSV format, the resulting CSV format will have a lot of empty fields. But the format is compatible.

Bin by Mag – Scans the entire loaded database (filters applied) and counts all quakes within whole magnitude bins.

Mag Bin by Year – Same as above, but breaks the bins down into whole years for the time span selected by the filters.

This is useful for getting a sense of the completeness of the catalog for different magnitudes over time.

List – Simply outputs a list of the loaded quakes limited by the filters. WARNING: Can result in a lot of returned data. Try doing a 'Count' first to make sure you don't list too much.

Count – Only provides a count of how many of the loaded events meet the filter criteria. Useful to do before listing the events.

List Types – This will scan the loaded data base for all the different event types and a tally of how many of each there are. CAUTION: This function will repopulate the Event Types in the Filters section. If the loaded database only has earthquakes in it, running this function will repopulate the filter with only 'all' and 'earthquake'.

Stats – will provide some basic metadata about the currently loaded data. Not much here yet.

Appendix A

USGS text and CSV data formats

Perhaps the best way to show the USGS format for these files is to simply show the header that the USGS server returns with each request. This header is saved in the files by this software when downloading these formats to disk ('All To Files').

Text format:

```
#EventID|Time|Latitude|Longitude|Depth/km|Author|Catalog|Contributor|ContributorID|  
MagType|Magnitude|MagAuthor|EventLocationName
```

CSV format:

```
time,latitude,longitude,depth,mag,magType,nst,gap,dmin,rms,net,id,updated,place,type,horiz  
ontalError,depthError,magError,magNst,status,locationSource,magSource
```

See the USGS website for more details.

<https://earthquake.usgs.gov/earthquakes/search/>

<https://earthquake.usgs.gov/fdsnws/event/1/>

Appendix B

USGSAnalyzer program binary file format

This is a full description of the binary data format that this program produces and works with. It has less information than what either the text or CSV formats returned by the USGS contain, but this helps to speed up some processes, especially when dealing with the entire USGS catalog.

This software is written in C++11 and the quake data is stored in a struct as follows,

```
struct QUAKEDATA {
    char    date[31];
    double  JD;
    double  lat;
    double  lon;
    double  depth;
    double  mag;
    char    type[31];
};
```

The record length is therefore 104 bytes. The binary file is headerless and raw. It simply contains a straight binary save of each record. The binary file's length is therefore always a multiple of 104.

For other languages, here's some details of each 104 byte record.

BYTES	DATA TYPE	DESCRIPTION
0...31	NULL terminated char string	Event date in ISO8601 format
32...39	floating point double	Decimal Julian date of event date
40...47	floating point double	Latitude
48...55	floating point double	Longitude
56...63	floating point double	Depth
64...71	floating point double	Magnitude
72...103	NULL terminated char string	Event type

Note that since this software was compiled for Windows on x86 architecture, the “double” data type is little endian 64 bit IEEE 754 double precision binary floating point format.

The NULL terminated character strings are in ASCII format.

The ISO8601 format as used by the USGS is formatted as “YYYY-MM-DDTHH:MM:SS.SSSZ”.

Appendix C

About and References

This software is written in C++11 using Code::Blocks as the IDE and compiled with GNU GCC. Developed on Windows 10.

The programmer is completely self taught thanks to amazing open source tools plus wonderful and free reference and tutorial websites such as,

<http://codeblocks.org/> - Code::Blocks Integrated Development Environment

<https://www.wxwidgets.org/> - Cross platform GUI library

<https://curl.haxx.se/libcurl/> - Multiprotocol file transfer library

<https://www.learncpp.com/> - C++ learning tutorials, reference

<http://www.cplusplus.com/> - C and C++ reference, and tutorials

<https://en.cppreference.com/> - C and C++ reference

USGS resources used to develop this software,

<https://earthquake.usgs.gov/earthquakes/search/> - Search the USGS catalog

<https://earthquake.usgs.gov/fdsnws/event/1/> - USGS web service API documentation