

Jason II Annotation Database Extraction (JADE v.1) Utility

C. R. Young
young@mbari.org
young@darwin.ucsc.edu

JADE is a utility written in *c* that extracts sample and observational data from the Virtual Van database for Jason II and creates several databases based on the user-defined (highest level) annotation tags. The extracted databases are *.csv files (comma-delimited) that can be opened in many spreadsheet programs (e.g., Excel). JADE supplies event numbers for each log entry that can be used as references to extract framegrabs or other metadata from the Virtual Van database. A user-supplied Jason II dive number lookup table may be used to specify the dive number of entries in the extracted databases. Formatting of the output is flexible, and can be specified by a user-supplied format file.

History:

05/16/05: Version 1. Extracts sample lines from the database
05/19/05: Version 1. Added Jason dive lookup table
05/21/05: Version 1. Added formatting option for Lat/Lon (decimal min., deg. min. sec.)
05/27/05: Version 1. Reformatted output to coincide (roughly) with R2K DMS
05/28/05: Version 1. Works with user-defined annotation tags
06/01/05: Version 1. Added format input file and Other.event output file

Getting started

You first need to get the full cruise database from Virtual Van (VV). To do this:

1. Open VV and select the cruise that you're interested in. On the right of the VV page is a link to export "**all**" **cruise data**; click this link. It is **important** to get data for the whole cruise and not the data for the current dive. If you get data only for the current dive, then the event numbers will be wrong!
2. The data will open in a new browser window. Tell your browser to **show the source code** for the data.
3. This will open a text file with the data in it. **Save this text file** to your hard drive.
4. If you're using a Mac, then you need to **change the endline characters** of ALL input files to UNIX-brand endline characters. (This can be done using a text utility like BBEdit.)

Input files

You will need 2–4 input files depending on whether you use a dive lookup table or the formatting options. The first file is the data file that you saved in the previous step. The example data file is "divedata.txt". You may name this file anything that you want. You will also need a file telling JADE the annotation tags that you used. The tag file must be named "tags.txt". If you choose to use the dive lookup table, then you will need a file called "divedates.txt", and if you choose to suppress some of the information in the VV database, then you will need a file called "JADE_format.txt".

Input file (required)

This file can have any name. It contains the raw VV database. If you follow the instructions above, no further manipulations should be necessary.

tags.txt (required)

This file contains the highest-level annotation tags that you are using in VV. These are the tags that look something like “BIO.Samples” or “GEO.Structure”. JADE will find lines in the larger database that match these tags. The tags can be in any order with one exception. The first tag must contain the watch list (SRV.Changewatch). This is because when using the quality control option, Jade inserts lines from the first tag into all files. If the first tag listed is something other than the watch list, then the resulting files will be a mess. If the quality control option is not being used, then the order does not matter. This file should end on a blank line (the program will crash otherwise).

divedates.txt (optional)

This file contains the dive numbers and the dates and times of the beginning and end of each dive. The format for each line is:

<dive#> <month(begin)> <day(begin)> <hour(begin)> <month(end)> <day(end)> <hour(end)>

All tokens, <...>, must be whole numbers (no decimals or hr:min). The tokens must be separated by white space (space or tab). This file must end on a blank line. If this file is not present, then JADE will know that you don't want to use a dive lookup table.

JADE_format.txt (optional)

This file contains two lines. The first line corresponds to the names of the data fields in the VV database. If you want to rename these fields for the output files, do so on this line. There must be 14 strings on this line separated by white space (space or tab). The second line should contain 14 integers (zero or one). These are switches to turn on (1) or off (0) printing of the data fields labeled in the row above.

The data fields are:

Class: You probably don't want this. It will be “EVT” for everything that you can extract.

Date/Time

Latitude

Longitude

x

y

Heading

Pitch

Roll

Altitude

Depth

Bathymetry

Type: You probably don't want this either.

Event: Second-level tags and comments

If this file is not present, then all data fields will be printed.

Options

JADE allows two options for quality control and conversion of coordinates to decimal-minute (DM) or degree-minute-second (DMS) formats. The options are specified at runtime from the command-line. For example, when you run the program you will type:

```
“./JADE <option 1> <option 2> <infile>”
```

The options are as follows:

Option 1

Do you want to use the quality control option? You answer “0” for no and “1” for yes. If you answer yes, then entries containing the times when people change watches or are relieved are displayed in each output file.

Option 2

Which coordinate format would you like to output? You answer “0” for decimal degrees, “1” for decimal minutes, or “2” for degrees, minutes, seconds. Note: decimal degrees format is ALWAYS output unless you turn the fields off using the format-specification file. The other coordinate formats are conversions from the raw data added for convenience.

Running the program

All of the files listed above and the JADE executable should be in the same directory. Use the terminal to get to that directory. Then type: “./JADE <option 1> <option 2> <infile>” at the command prompt, where <infile> is the name of the file containing the VV annotation data. The data fields that will be printed, the annotation tags that were read by the program and the dive lookup table (if used) are output to the terminal. You should check to see if the program read your files correctly.

Output files

Output files are named according to the tags listed in the file tags.txt. There will be one more output file than there are tags (JADE automatically creates the file “Ann.Other.events.csv” that contains all of the events that do not fall under the tags list. The output files are comma-delimited files (*.csv) that can be opened in most spreadsheet programs (e.g., Excel). The files contain all annotation entries that the VV database contains under each tag.

If the quality control option is turned off (zero), then the names will be:

```
“Ann.<tag name>.csv”
```

If the quality control option is turned on (one), then the file names will be:

```
“QC.<tag name>.csv”
```

The distinction is made so that the program can be run with and without the QC option in the same directory and no files will be overwritten.

Other information and suggestions

JADE only checks the first eight characters in the tag. This feature can be as useful as it is annoying. For example, if you specify two tags as “BIO.vent1” and “BIO.vent2”, then JADE will see them as the same tag. You WILL have two files named “Ann.BIO.vent1.csv” and

“Ann.BIO.vent2.csv”, but both files will contain all events for “BIO.vent1” and “BIO.vent2”. You can specify the tag as “BIO.vent” and you have a single file, “Ann.BIO.vent.csv”, with all of the entries for both tags. This can be helpful if you want to keep from having to scroll a lot while annotating (the VV field size limits the number of second-level tags to about ten before you have to start scrolling to find them). However, if you use a tag shorter than eight characters, such as “BIO.Vnt”, then there will be no events in the file “Ann.BIO.vnt.csv”. Therefore, two things need to be kept in mind when specifying your tags for VV:

1. No tags shorter than eight characters (including the three character prefix and dot)
2. Two tags that are the same for all of the first eight characters will be combined

The number of characters of the tags that are examined by JADE can be altered by redefining CHECKTAGSIZE and recompiling the program.

JADE ignores commas in the comment field, so it is OK to use commas in the comments field of VV.

Try to not enter information without a predefined tag while annotating. All of this information ends up in the file “Ann.Other.events.csv” and can be difficult to find. Also, misspellings or non-standard terminology makes searching for information difficult or impossible.

Compilation

Place the source file (“JADE.c”) in the directory that you want the executable. Navigate into that directory using the terminal. At the command prompt, type:

```
gcc JADE.c -o JADE
```