NAME

gshhg − Extract data tables from binary GSHHG or WDBII data files

SYNOPSIS

gshhg *binaryfile.b* [**−A***min*] [**−G**] [**−I***id*] [**−L**] [**−N***level*] [**−Qe|i**] [**−bo**binary] [**−do**nodata] [**−o**flags]

Note: No space is allowed between the option flag and the associated arguments.

DESCRIPTION

gshhg reads the binary coastline (GSHHG) or political boundary or river (WDBII) files and writes an ASCII (or binary; see **−b**) listing to standard output. It automatically handles byte−swabbing between different architectures. Optionally, only segment header info can be displayed. The header info has the format *ID npoints hierarchical−level source area f_area west east south north container ancestor*, where hierarchical levels for coastline polygons go from 1 (shoreline) to 4 (lake inside island inside lake inside land). Source is either W (World Vector Shoreline) or C (CIA World Data Bank II); lower case is used if a lake is a river−lake. The *west east south north* is the enclosing rectangle, *area* is the polygon area in kmˆ2 while *f_area* is the actual area of the ancestor polygon, *container* is the ID of the polygon that contains this polygon (−1 if none), and *ancestor* is the ID of the polygon in the full resolution set that was reduced to yield this polygon (−1 if full resolution since there is no ancestor). For line data the header is simply *ID npoints hierarchical−level source west east south north*. For more information about the file formats, see TECHNICAL INFORMATION below.

REQUIRED ARGUMENTS

binaryfile.b

GSHHG or WDBII binary data file as distributed with the GSHHG data supplement. Any of the 5 standard resolutions (full, high, intermediate, low, crude) can be used.

OPTIONAL ARGUMENTS

−A*min* Only output information for the polygon if its area equals or exceeds *min* [Default outputs all polygons].

−G Write output that can be imported into GNU Octave or Matlab by ending segments with a NaN−record.

−I*id* Only output information for the polygon that matches *id*. Use **−Ic** to get all the continents only [Default outputs all polygons]. See below for the *id* of the largest polygons.

−L Only output a listing of polygon or line segment headers [Default outputs headers and data records].

−N Only output features whose level matches the given *level* [Default will output all levels].

−Qe|i Control what to do with river−lakes (river sections large enough to be stored as closed polygons). Use **−Qe** to exclude them and **−Qi** to exclude everything else instead [Default outputs all polygons].

−bo[*ncols*][*type*] (more ...)
Select native binary output.

−do*nodata* (more ...)
Replace output columns that equal NaN with *nodata*.

−o*cols*[,...] (more ...)
Select output columns (0 is first column).

EXAMPLES

To convert the entire intermediate GSHHG binary data to ASCII files for Octave/Matlab, run

```
gmt gshhg gshhs_i.b --IO_SEGMENT_MARKER=N > gshhs_i.txt
```

To only get a listing of the headers for the river data set at full resolution, try

```
gmt gshhg wdb_rivers_f.b -L > riverlisting.txt
```

To only extract lakes, excluding river−lakes, from the high resolution file, try

```
gmt gshhg gshhs_h.b -Ee -N2 > all_lakes.txt
```

SPECIFIC POLYGONS

None of the polygons have any name information associated with them (i.e., the metadata does not contain this information). However, here are the largest polygons:

ID	Landmass
0	Eurasia
1	Africa
2	North America
3	South America
4	Antarctica (AC grounding line)
5	Antarctica (AC ice line)
6	Australia
7	Greenland
8	New Guinea
9	Borneo
10	Madagascar
11	Baffin Island
12	Indonesia

TECHNICAL INFORMATION

Users who wish to access the GSHHG or WDBII data directly from their custom programs should consult the gshhg.c and gshhg.h source code and familiarize themselves with the data format and how various information flags are packed into a single 4−byte integer. While we do not maintain any Octave/Matlab code to read these files we are aware that both MathWorks and IDL have made such tools available to their users. However, they tend not to update their code and our file structure has evolved considerably over time, breaking their code. Here, some general technical comments on the binary data files are given. **GSHHG**: These files contain completely closed polygons of continents and islands (level 1), lakes (level 2), islands−in−lakes (level 3) and ponds−in−islands−in−lakes (level 4); a particular level can be extracted using the **−N** option. Continents are identified as the first 6 polygons and can be extracted via the **−Ic** option. The IDs for the continents are Eurasia (0), Africa (1), North America (2), South America (3), Antarctica (4), and Australia (5). Files are sorted on area from large to small. There are two sub−groups for level 2: Regular lakes and the so−called "river−lakes", the latter being sections of a river that are so wide to warrant a polygon representation. These river−lakes are flagged in the header (also see **−Q**). All five resolutions are free of self−intersections. Areas of all features have been computed using a Lambert azimuthal equal−area projection centered on the polygon centroids, using WGS−84 as the ellipsoid. GMT use the GSHHG as a starting point but then partition the polygons into pieces using a resolution−dependent binning system; parts of the world are then rebuilt into closed polygons on the fly as needed. For more information on GSHHG processing, see Wessel and Smith (1996). **WDBII**. These files contain sets of line segments not necessarily in any particular order. Thus, it is not possible to extract information pertaining to just one river or one country. Furthermore, the 4 lower resolutions derive directly from the full resolution by application of the Douglas−Peucker algorithm (see gshhg_dp), hence self−intersections are increasingly likely as the resolution is degraded. Note that the river−lakes included in GSHHG are also duplicated in the WDBII river files so that each data set can be a stand−alone representation. Users who wish to access both data sets can recognize the river−lakes features by examining the header structure (see the source code for details); they are also the only closed polygons in the WDBII river file. There are many levels (classes) in the river file: River−lakes (0), Permanent major rivers (1), Additional major rivers (2), Additional rivers (3), Minor rivers (4), Intermittent rivers −− major (6), Intermittent rivers −− additional (7), Intermittent rivers −− minor (8), Major canals (10), Canals of lesser importance (11), and Canals −− irrigation type (12). For the

border file there are three levels: National boundaries (1), Internal domestic boundaries (2), and international maritime boundaries (3). Individual levels or classes may be extracted via **−N**.

REFERENCES

Douglas, D. H., and T. K. Peucker, 1973, Algorithms for the reduction of the number of points required to represent a digitized line of its caricature, *Can. Cartogr., 10*, 112−122.

Gorny, A. J., 1977, *World Data Bank II General User GuideRep. PB 271869*, 10pp, Central Intelligence Agency, Washington, DC.

Soluri, E. A., and V. A. Woodson, 1990, World Vector Shoreline, *Int. Hydrograph. Rev., LXVII(1)*, 27−35.

Wessel, P., and W. H. F. Smith, 1996, A global, self−consistent, hierarchical, high−resolution shoreline database, *J. Geophys. Res., 101(B4)*, 8741−8743.

SEE ALSO

gmt

COPYRIGHT