

# Crossmatching Two External catalogs

Alan

19 Feb 2013

This is a simple cross-matching program that can be used to match catalogs. The program is not well optimized yet (for instance, TopCat<sup>1</sup>, a Java-based utility used to manipulate large catalogs can match two 500k catalogs in approximately 5 seconds, while this program takes longer than 10 minutes). However, this program is now a single command and requires no interaction with a GUI. The program is built on the healpix mapping algorithm.

## 1 Dependencies

You need to have HealPy and NumPy installed for your python installation. There are no other dependencies.

## Cross-Matching

### The matching process

The method used to match the data is a simple nearest neighbor cross-match on ra/dec, requiring the two matched galaxies to be nearest neighbors in both directions. For example if galaxy  $a$  from catalog  $A$  and galaxy  $b$  from catalog  $B$  are matched this implies that  $b$  is the closest object to the ra/dec of  $a$  in the entire catalog  $B$  and galaxy  $a$  is the closest object to the ra/dec of  $b$  in the entire catalog  $A$ . The external matching utility is powered by the Healpix mapping module for Python called healpy. Using the python module healpy, external catalogs can be quickly cross-matched.

### Operation of the program

Given two catalogs provided by the user, the program `crossmatch_external.py` will provide a matched catalog with the rownumber from catalog 1, the rownumber from catalog 2 and the separation in arcseconds for each object that matches between the two catalogs. The original catalogs should contain at a minimum two columns (ra and dec) additional columns are allowed, but they will

---

<sup>1</sup>see <http://www.star.bris.ac.uk/~mbt/topcat/> for the Topcat program description

be ignored. RA and DEC must be the first two columns in each catalog. `crossmatch_external.py` can be run at the command line. Here is the “help” string associated with the program:

Usage: `crossmatch-external.py` OPTIONS

This program will crossmatch two input catalogs of ra/dec coordinates and output the results. In general, you need to specify the input tables, and output file.

Options:

```
-h, --help           show this help message and exit
-f INCAT1, --first-cat=INCAT1
                     first input catalog
-s INCAT2, --second-cat=INCAT2
                     second input catalog
-d MAXDIST, --max-distance=MAXDIST
                     the maximum separation in arcsecs that any source may
                     have.
-o OUTFILE, --output-file=OUTFILE
                     output file for crossmatched catalog
-n NSIDE, --nside=NSIDE
                     nside parameter used to construct healpix map
```

A simple execution of the program would then look like this:

```
./crossmatch_external.py -f catalog1.txt -s catalog2.txt -o my_output.txt
```

This command will crossmatch the two catalogs (“catalog1.txt” and “catalog2.txt”) and place the output in the file “my\_output.txt” which can then be read into other programs. Note that the output file has several lines of metadata at the top. These lines are prefixed with a hash character “#” so that you can easily skip them when reading the file. The metadata contains the settings used to crossmatch the catalogs including the input catalog names, the maximum allowed separation, the nsides value used by healpix and the date of the matching.

This matching program should work on any computer that has python, healpy, and numpy installed. The only direct dependancies are numpy and healpy, but I think pyfits and matplotlib may be dependencies of healpy.