

Werkcollege from 23 September
Astronomical databases - SDSS CasJobs

Originally created by Jarle Brinchmann, Leiden Observatory for IAC 2010 Modern Data Mining and Virtual Observatory

1 Astronomical Databases and Catalogs

Understanding the data you are using is crucial for answering questions you have, and the first question is - does this database/catalog suitable for my purposes?

Find following characteristics:

- selection criteria (wavelength, depth, completeness)
- spectral coverage and resolution
- spatial coverage and resolution

for following catalogs:

- a) The Bright Star Catalog Supplement
- b) A Spectrophotometric Atlas of Synthetic Standard Spectra (Pickles 1985)
- c) CCD H α and R photometry of 334 galaxies (James+, 2004)
- d) The ROSAT brightest cluster sample - I. (Ebeling+, 1998)

The focus of next tasks will be on SDSS data. For information about the SDSS, algorithms etc. you should go to: <http://www.sdss.org/dr7/>.

2 CasJobs

Create a CasJobs account - run one simple search and download the result. This is not a terribly challenging task but it is very useful to be familiar with CasJobs - and at first glance it can be a bit bewildering. Note that the schema for the SDSS database is available on <http://cas.sdss.org/dr7/en>

To set up an account or later log in: casjobs.sdss.org/CasJobs

- a) On the page - use the Create Account link on top to create your account. You can then create your first query. The best now is just to make a very simple one - choose the first 100 objects in SpecObj table (actually a view) that is within 0.001 of $z=0.04$:

```
SELECT TOP 100 z, ra, dec
FROM SpecObj
WHERE abs(z-0.04) < 0.001
```

(If you prefer you can of course run whatever query you want but make sure it does not take too much time and does not return too many data points).

- b) After the query has finished - or while it is running, click on the History link to be informed on its progress. This is a very convenient functionality btw - when you come back to CasJobs after some time away!
- c) Your search has ended up in MyDB - click the link for this and verify that there is a table there - click on the table name to get up an interface to look at it. Explore this and at the end click Download to download the finished table in your preferred format. If you download a CSV or FITS file you might want to use TOPCAT to explore the results - or R/Iraf/IDL/PDL etc if you prefer that.
- d) What is the difference in SQL implementation between MySQL and SQL Server you have noted from this query? How to implement the same feature in MySQL?

3 Stellar abundance distribution

Armed with a CasJobs account it is now relatively easy to do some simple searches. Let us have a look at the distribution of metallicity in stars. A simple query would be:

```
select floor(feha) as fehabin, COUNT(*) as number
FROM spparams
Where
  fehaerr > 0
GROUP BY floor(feha)
ORDER BY fehabin
```

- a) Try this out! It should only take 3-4s so you can easily run this in Quick mode. You should find a maximum of 106,810 stars if run on DR7.
- b) As you will notice, this approach bins the abundances in integers - very coarse. Use the notes from the lecture and bin this in bins of 0.2 in feha. The metal abundance that you find is the logarithm of the iron (Fe) abundance relative to the solar abundance. So a value of 0.0 means equal to the solar value.
What is really interesting are those stars that are very metal poor - say with $[Fe/H] < -3$. How many are there?
- c) How do these stars with $[Fe/H] < -3$ distribute with Galactic latitude? First find the necessary columns you need from the Schema. Then calculate the number of stars as a function of latitude, perhaps in bins of 10 degrees. Do it either all in SQL or do a search and download the table to Topcat [it is much faster to do it all inside CasJobs].
- d) The preceding does not take into account the general drop in stars with latitude so you want to do it relative to the total number of stars. This can again be done in SQL or in Topcat. Since it is perhaps a bit complex in SQL here is a hint:

```
select a.latitude, cast(a.num AS float)/cast(b.numall AS float) as
fraction -- Num is number per bin of low-metallicity stars
-- Numall is the number per bin of all stars
-- Latitude is the latitude bin and the same binning
-- binning is assumed in both tables below.
FROM (
  select <... What you did in c ...>
) as A JOIN
(
  select <... Do the same, but without a feha cut ...>
) as B ON a.latitude=b.latitude
ORDER BY a.latitude
```

4 Bright high-redshift objects

Let us return to the example from the lecture, namely to find bright, high-redshift objects. The query there was:

```
Select TOP 10 plate, mjd, fiberid, z
From SpecPhoto
Where
  petroMag_R < 14
AND
petroMagErr_R < 0.1
ORDER BY -z
```

We would not like to look at these to check whether they are reasonable or not. A quick way to do this is to request the URL for a page where the object can be inspected. This is done using a function called fGetUrlExpEq which constructs the URL to the explorer page (other functions exist, see the Schema):

Select TOP 10 plate, mjd, fiberid, z, fGetUrlExpEq(ra, dec)

- a) Modify the query above to return the URLs of the explorer page and check a few - do they look like real galaxies (extended objects) or quasars (essentially point sources) to you? The answer should be no!
- b) When you do searches in large databases it is crucial to reject rubbish - especially when you are interested in extreme/atypical objects. The search above has a cut on bad photometry - the second variable we use is redshift (z) - look at the Schema for SpecPhoto and suggest a column that might help in only selecting objects with good redshift determinations. There are several possibilities and if you work together it might be an idea to test a couple. Can you get rid of most of the rubbish detections?
- c) What stored function can be used to find a distance between objects in SDSS?

5 Planetary nebulae

Let us next find all planetary nebulae in the SDSS.

- a) Find a table of Galactic nebulae in Vizier.
- b) Load this up in topcat, deselect all columns except the RAJ2000, DEJ2000 and the name column. Save this as a VOTable (this is necessary to avoid a limitation in CasJobs - the import doesn't allow null values on input so the table should only include non-null entries). Now use the Import menu in CasJobs to load the table into CasJobs.
- c) The next step is to match this to the SDSS. This can be done easily using a function:

```
SELECT p.RAJ2000 as ra, p.DEJ2000 as dec,  
dbo.fGetNearestObjIdEq(p.RAJ2000, p.DEJ2000,3.0) as objId into  
mydb.PNMatch from MyDB.PNe as p  
WHERE  
dbo.fGetNearestObjIdEq(p.RAJ2000, p.DEJ2000,3.0) IS NOT NULL
```

Here you can see that we use the fGetNearestObjIdEq function to find the closest match to a given PNe. We use a search radius of 3 arcsec and return the objectID of the closest object.

- d) Now match this to the PhotoPrimary view to get photometry for the objects, at least g, r, i.
- e) We would also like to examine the images of these matches - you can either create URLs as you learnt above, or you can simply print out a list with objID, ra, dec - copy this and paste it into the SDSS Imlist service. My recommendation would be the latter - do all your matches obviously correspond to a planetary nebula?

6 Brief visits to other services.

While the emphasis here has been on the SDSS CasJobs, there are many other useful databases around. Take a few minutes to explore these:

- a) The GALEX satellite works in the ultraviolet and they have an SQL service very similar to that of the SDSS <http://galex.stsci.edu/GR4/>.
- b) The 6dF survey is a galaxy redshift survey with an SQL interface. Try to find how many HIPASS objects there are in the survey with redshifts > 0.1 . I found 109 when I limited myself to objects with quality 3 or 4 and progID < 8 .
- c) The IPAC service at IRSA: <http://irsa.ipac.caltech.edu/> provides comprehensive access to many catalogues. Go for instance to Spitzer/General Catalog Queries - and select the SAGE survey. What are the limitations of the search screen you come to relative to a full SQL search do you think?
- d) A more traditional database interface is that for the debris disk database at

<http://www.roe.ac.uk/ukatc/research/topics/dust/identification.html>.

Explore this - comment on the flexibility of this relative to some of the other services you have tried.

7 Additional tasks

There are four additional tasks:

- I. Find asteroids (moving objects) in the SDSS.
- II. Find & match a catalogue of known exoplanets to SDSS and to one on habitability.

7.1 Problem 1 - Finding asteroids

The SDSS images are taken with some time between each filter - approximately 72s. The pixel size of the SDSS camera is 0.394 arcseconds/pixel and a moving object can then be detected if has moved between images in different filters - you will use this to construct a catalogue of potential asteroid candidates.

- a) Estimate what the minimum velocity on the sky is that you can detect - assume that the software can detect shifts of a 1/10th of a pixel. There is also an upper limit - very fast objects move so far between exposures that they are not recognised as re-observations of the same object. There are some of these - Near Earth Objects typically - but we ignore those here. Note that the standard units for motion on the sky for these objects is degrees/Day.
- b) Identify the necessary columns to search for moving objects in the PhotoObj table and write an SQL survey that find moving objects - be conservative and use 10 times the limit you found in a) as the minimum velocity. You probably want to limit yourself to objects with r-magnitude < 21.0 or so be sure the detections are reliable - or maybe even brighter. You might find it useful to know that to calculate the square of x is done using POWER(x, 2). I got about 174,000 asteroid candidates in my search, your results might differ depending on how you select objects. [be aware that this search will take \approx 1 hour or so].
- c) Run the query and determine the number of possible Asteroids. Download the file and create velocity-colour plots for the putative asteroids. Did you get any outliers?

If we were to continue this work we would need to match to a catalogue of proper motions because a study of asteroids is not very useful without orbital parameters. This can be done by matching to the ASTORB catalogue which is available through Vizier. We will not do that here but I note that this job has been done in what is called the SDSS MOC - the SDSS Moving Object Catalogue:

<http://www.astro.washington.edu/users/ivezic/sdssmoc/sdssmoc.html>

Also see Ivezić et al (2001) which is linked to from that page for more detail (also Juric et al (2002) and Parker et al (2008) among others).

7.2 Problem 2 - Exoplanets

- a) Download a VOTable of all known exoplanets from exoplanet.eu.
- b) Upload the table to CasJobs and match it to the Star table in CasJobs. How many stars have a counterpart in the SDSS? Look at the images of a few (5-10) - can you say something general about them? Do you think the SDSS data are useful for the majority of the stars.
- c) Find a catalogue on the habitability of exo-planets and match this to the full exoplanets table so that you keep all exoplanet candidates but when possible include the habitability information. This is effectively an OUTER JOIN. You can do this by using CasJobs.
- d) Finally take the resulting catalogue and create a mysql database & table with the information you gathered.