Adding catalog data to object lists using the VO

Author: M. Demleitner
Software: TOPCAT

This brief tutorial shows you how to quickly add proper motions and 2MASS photometry to (almost) any object list using the Virtual Observatory. The protocol to do that is called TAP ("table access protocol") and allows transferring data and queries to database servers. In the example, we will be using TOPCAT as a client of that protocol. There is no lockin to that: There are libraries and other tools allowing an integration of TAP operations into arbitrary workflows – that's what standards are about.

Describe your data to TOPCAT

Whatever you can load into TOPCAT can work here. To play through the worst case, let's assume you just have a load (millions are possible) of RA/Dec values in one of the text files we astronomers so adore. If you don't have a file of your own handy, we have prepared a sample file for you at http://docs.g-vo.org/pmadd-testdata.txt.

Load your data into TOPCAT (manually set format to ASCII in TOPCAT's load dialog if necessary). In our example, TOPCAT doesn't yet know it's dealing with positions. To tell it, bring up the columns info window by selecting Views/Column Info from the menu.

The easiest way to tell TOPCAT we have a position here is just editing the column names to be `ra` and `dec`. Double-click into table cells to edit them in TOPCAT.

Close the columns info window.

Select the service and table to match against

You'd normally use the registry to figure out a TAP-accessible table that contains the physics you want – we have a separate tutorial showing you that. For this tutorial, let's assume you've already figured in some way the data you're after is available from the table `ppmxl.main` within the service with the access URL http://dc.g-vo.org/tap.

In TOPCAT, select VO/TAP Query (I ususally click the little pin button in the top left to make sure the TAP window stays open, which is nice when editing and trying queries).

Near the bottom, in the field labelled TAP URL, enter the service's access URL as given above and hit **Enter Query**.

TOPCAT scwitches now to the "Enter Query" tab of your window. At the top, there is a dropdown menu with a list of tables that this service offers. Select `ppmxl.main` from this list; note how you can browse the table metadata within the TAP window.

Write the query

Since crossmatches to uploaded data are so frequent, TOPCAT can generate you a sample query to start from; under the Examples button (right above the large text field near the bottom of the TAP dialog) select "Upload join". A query adapted to joining the current topcat table with the TAP table select appears in the query field. It should look pretty much like this::

```
SELECT
    TOP 1000
    *
    FROM ppmxl.main AS db
    JOIN TAP_UPLOAD.t1 AS tc
```

```
    ON 1=CONTAINS(POINT('ICRS', db.raj2000, db.dej2000),
                  CIRCLE('ICRS', tc.ra, tc.dec, 5./3600.))
```

Don't worry if you don't understand all of it now – that can wait a bit. Suffice it to say that it's fairly standard SQL[1]

Edit it a bit to better fit your needs: Instead of `TOP 1000`, say `TOP 5000` or something appropriate to the size of your input data (but remember that an input line can match multiple objects, so be generous here: TOPCAT and the TAP servers shouldn't be scared by a million objects).

The `*` means: Get all columns. That's more than we care to see here. We want our columns – that's `tc.*`, the `tc` being the alias given in the line starting with `JOIN` –, and the proper motions and 2MASS magnitudes from the database, the names of which you can get from the table browser in the TAP window. So: replace the `*` with::

```
  tc.*, pmra, pmde, jmag, hmag, kmag
```

(you could have written `db.pmra` and so on here, but as long as the names are unique, you can skip the table name).

Finally, you may want to change the match radius. That's the `5./3600.`, and it's in degrees. 5 arcsec should be ok here.

Then hit **OK** and enjoy, plot, or devour your results.

Marginalia

The metadata we added above is of course extremely shoddy. For it to be more understandable outside of TOP-CAT, you would add UCDs[2] to the column metadata: For that you'd select Display/UCD from the window and enter `pos.eq.ra;meta.main` in the UCD field for the RA column and correspondingly `pos.eq.dec;meta.main` for the Dec column.

The query as written returns one line per *match*, i.e., input objects that have no counterpart in the database will be missing in the output; and indeed, you'll see that for the 4000 input objects (which, really, are typically fairly weak infrared sources if you have to know), only about 1500 output lines result.

If you want non-matching input objects in the output, with NULLs wherever database information would be, write `RIGHT OUTER JOIN` instead of `JOIN` ("right" because your table is the right operand in the join expression). If you used the sample data set, you'll see that 253 objects had multiple matches in PPMXL.

More on TAP and ADQL in GAVO's little course at http://docs.g-vo.org/adql

---

[1] "Structured Query Language", the de-facto standard for communicating with database systems since the seventies. In this incarnation, it's actually ADQL, where AD stands for "Astronomical Data", but that makes no real difference.

[2] "Unified Content Descriptors", standard labels for physical entities; you might have seen those on VizeR