Astronomical Data Analysis Software and Systems VII
ASP Conference Series, Vol. 145, 1998
R. Albrecht, R. N. Hook and H. A. Bushouse, eds.

The VizieR System for Accessing Astronomical Data

François Ochsenbein

Centre de Données Astronomiques, Strasbourg Observatory, France

Abstract. The recently reshaped VizieR[1] system, a unified query interface to an increasing number of catalogs (presently ~ 1500), is presented.

1. Historical Background

VizieR was first presented at the AAS meeting in early 1996 (Ochsenbein et al. 1995), as the result of a joint effort between CDS[2] and ESA-ESRIN[3] (the European Space Agency's Information Systems division) in order to provide the astronomical community with a dedicated tool for retrieving astronomical data listed in published catalogs and tables — a follow-up of the ESIS (European Space Information System) project.

Shortly after this first version, which has been effectively accessible since February 1996, new needs for performance and for standardisation led to basic changes in the system: the ASU[4] (Astronomical Standardized URL), resulting from discussions between several institutes, was adopted as the way to specify constraints in the new version of VizieR, which was introduced on 29 May 1997 — just in time for the distribution of the results of *Hipparcos* catalogs. The basic concept of ASU is a standardized way of specifying catalogs (as `-source=`*catalog_designation*), target positions (as `-c=`*name_or_position*`,rm=`*radius_in_arcmin*), output format (as `-mime=`*type*), and general constraints on parameters (as *column_name*=*constraint*).

Besides the adoption of this new protocol, the most visible changes in this new version of VizieR are an easy access to notes, and possibilities of navigation between the tables of a catalog.

The quantitative daily usage of VizieR is presently (September 1997) about 1,000 external requests from 75 different nodes; 1,000 different nodes effectively submitted queries to VizieR during the first 3 months of the new installation (June to August 1997); among all queries, about 40% of the hits concern the recent results of the *Hipparcos* and *Tycho* missions.

---

[1] http://vizier.u-strasbg.fr

[2] http://cdsweb.u-strasbg.fr

[3] http://www.esrin.esa.it/esrin/esrin.html

[4] http://vizier.u-strasbg.fr/doc/asu.html

2. How to Query in VizieR

The "standard query" in VizieR consists in a few steps:

1. Locate the interesting catalogs in the VizieR Service[5]. This page presents various ways of finding out the interesting catalog among this large set:
   (a) from one of their usual *acronyms*, like **GSC**, **HD**, **HIC**, etc...
   (b) from a set of words (author's names and/or words from the title of the catalog), and/or keywords attached to each catalog;
   (c) or by clicking in a Kohonen Self-Organizing Map[6], a map created by neural network techniques which tends to group in nearby locations those catalogs having similar sets of keywords. This technique is the same as the one used by Lesteven et al. (1996) to index the bibliography.
2. Once a catalog table – or a small set of catalog tables — is located (for instance the *Hipparcos* Catalog[7] resulting from the *Hipparcos* mission), *constraints* about the input and/or output can be specified, as:
   - constraints based on the celestial coordinates (location in the neighbourhood of a target specified by its actual coordinates in the sky, or its name, using SIMBAD[8] as a name resolver);
   - any other qualification on any of the columns of the table(s); the standard comparison and logical operators are available, detailed in the VizieR help pages[9];
   - which columns are to be displayed, and in which order the matching rows are to be presented.

   By pushing the appropriate buttons, it is for instance easy to get the list of *Hipparcos* stars closer than 5 parsecs to the Sun, ordered by their increasing distance[10].
3. Obtaining full details about one row is achieved by simply clicking in the first column of the result: for instance, the first row of the search for nearby stars described above leads to the VizieR Detailed Page with *Hipparcos* parameters and their explanations concerning Proxima Centauri[11].

---

[5] http://vizier.u-strasbg.fr/cgi-bin/VizieR

[6] http://vizier.u-strasbg.fr/cgi-bin/VizieR#Qkmap

[7] http://vizier.u-strasbg.fr/cgi-bin/VizieR?-source=I/239/hip_main

[8] http://simbad.u-strasbg.fr/Simbad

[9] http://vizier.u-strasbg.fr/cgi-bin/Help?VizieR/intro

[10] http://vizier.u-strasbg.fr/cgi-bin/VizieR?-source=I/239/hip_main&-sort=-Plx&Plx=%3e=200

[11] http://vizier.u-strasbg.fr/cgi-bin/VizieR-5?-source=I/239/hip_main&HIP=70890

4. Finally, there may be correlated data, like notes or remarks, references, etc.... In our example, Proxima Centauri is related to the α Cen system, which components can be viewed from the CCDM link appearing in the detailed page.

It should be noted that the usage of the ASU protocol allows to write anywhere in a text (like it is done in this short article) a hyperlink to the result of a query: for instance, all parameters from the *Hipparcos* catalog for the star HIP 12345 can be pointed to by a call to VizieR with parameters: `-source=I/239/hip_main&HIP=12345`[12]; or a pointer to all *Tycho* stars closer than 0.5° to Sirius, ordered by their increasing distance to the brightest star, can be written by a call to VizieR with parameters: `-source=I/239/tyc_main&-c=Sirius,rm=30&-sort=_r`[13]

3. VizieR Structure

VizieR is based on the usage of a relational DBMS: the data tables are stored as relational tables, and a set of *META* tables — a structure which was called *Reference Directory* in the previous version of VizieR — contains the necessary descriptions of all items:

- *METAcat* is a table providing the description of *catalogs* (a *catalog* is a set of related tables, like a table of observations, a table of mean values, a table of references, ...); *METAcat* details the authors, reference, title, etc... of each stored catalog. There are presently $\sim$ 1500 rows in this table.
- *METAtab* is a table providing the description of each *table* stored in VizieR: title, number of rows, how to access the actual data, the equinox and epoch of the coordinates, etc... There are presently $\sim$ 3500 rows in this table, *i.e.* an average of $2\frac{1}{3}$ tables per catalog.
- *METAcol* is a table providing the description of each *column* stored in VizieR: labels, units, datatypes, etc... There are presently $\sim$ 45000 rows in this table, *i.e.* an average of 13 columns per table.
- about 10 more tables exist in the system to detail other parameters, like the definitions of keywords, the acronyms associated with the catalogs, the notes, etc...

4. The VizieR Feeding Pipeline

It is of course not possible to enter all details describing the $\sim$ 45000 columns by hand: the VizieR feeding pipe-line is fully automatic using as input the standardized description[14] of the catalogs shared by the Astronomical Data Centers,

[12] http://vizier.u-strasbg.fr/cgi-bin/VizieR?-source=I/239/hip_main&HIP=12345

[13] http://vizier.u-strasbg.fr/cgi-bin/VizieR?-source=I/239/tyc_main&-c=Sirius,rm=30&-sort=_r

[14] http://vizier.u-strasbg.fr/doc/catstd.htx

and used for the description of the electronic tables published by a large fraction of the major astronomical journals (A&A, ApJ, AJ, PASP).

The addition of a new catalog into VizieR consists in two steps: **(1)** adding into the *META* tables all items describing the new catalog: catalog title, authors, table captions, etc... and details about every column of each table; and **(2)** converting the data of each original (ASCII) file making up the catalog into a relational table. In this step, the existence of data (the so-called *NULL values*) is carefully tested, and some important parameters like astronomical positions, magnitudes and color indexes are converted to uniform units. We however take care of storing, as much as possible, the catalogs in their original form, and for instance the coordinates are stored in their original equinox and epoch.

5. Access to the Very Large Catalogs

Very Large Catalogs are defined here as catalogs made of more than 10^7 objects — a size which can hardly be managed by the existing DBMs. Dave Monet's (1977) USNO-A1.0 catalog[15] gathering 488,006,860 sources is a typical example: it consists originally in a set of 10 CDROMs (about 6Gbytes) with 12-bytes binary records (positions, magnitudes, and a couple of flags).

This catalog was losslessly compressed by grouping small regions of the sky, allowing to store only position offsets instead of full-range values positions: the resulting catalog occupies only 3.4Gbytes, allowing therefore faster access since the queries are heavily i/o-limited: on a Sparc-20 (72MHz), the average search time (radius of 2.5') is less than 0.1s, and the whole 488×10^6 objects are tested in about 40 minutes (i.e., $5\mu s$ per object).

6. VizieR developments

The current new developments include: *(a)* more *connectivity* between catalogs, with Simbad, and more *remote connectivity* with external databases, Observatory Archives, and other search engines; the *GLU* (Fernique et al. 1998) will most likely be extensively used for this purpose; *(b)* the creation of shared indexes on fundamental parameters like celestial positions in order to allow queries directed to a large number of tables; *(c)* a possibility to submit large lists for queries; and *(d)* a facility to present the results in graphical form.

References

Fernique, P., Ochsenbein, F., & Wenger M. 1998, this volume

Lesteven, S., Poinçot, P., Murtagh, F. 1996, Vistas in Astronomy 40, 395

Monet, D. 1997, "The USNO A1.0 catalog", http://psyche.usno.navy.mil/pmm

Ochsenbein, F., Genova, F., Egret, D., Bourekeb, I., Sadat, R., Ansari, S.G., & Simonsen, E. 1996, Bull. American Astron. Soc. 187 #9103

[15] http://vizier.u-strasbg.fr/cgi-bin/VizieR?-source=USNO-A1.0