Accessing Astronomical Data over the WWW using *dat*OZ

Patricio F. Ortiz

Department of Astronomy, University of Chile, Casilla 36-D, Santiago, Chile, Email: ortiz@das.uchile.cl

Abstract.
The Department of Astronomy of the University of Chile hosts a number of astronomical databases created with *dat*OZ (Ortiz, 1997). This is a site in which databases are accessed interactively by the use of an HTML interface and CGI database engines. Data can be retrieved by the users in a highly flexible way, from lists of user-selected quantities, to customized plots, including any additional multimedia information available for the database elements. The latest additions to the system point in two directions: a) to allow access to combinations of database variables for different purposes, and b)to allow the retrieval and further correlation of information stored in different *dat*OZ's databases, by means of a second tool named **Cat***MERGER*. Another tool lets users search for objects in all catalogs created with *dat*OZ at the same time.

The most recent development is the creation of a new version of *dat*OZ capable of handling catalogs such as USNO's A1.0 with 488 million objects. Specially designed indexing and data storage techniques allow the catalog size to be reduced by half.

1.   Introduction

Electronic sharing of scientific data is a concept which has been around for at least the last fifteen years. The first efforts pointed in the direction of creating sites accessible via ftp. Pioneer work at the Strasbourg CDS (SIMBAD), and by NASA (IPAC's NED, GSFC's ADC ftp site) amongst others, have set the ground to share data in a more efficient way amongst astronomers than the traditional printed, journals.

The WEB gives scientists a much richer way of sharing information. We can now have quite a number of things done by transferring just a few bytes between our computers and host computers running httpd (the HTTP daemon) by invoking CGI (common gateway interface) "scripts". The applications are almost unlimited, and we see more and more of them at astronomical sites around the world.

Scientists are usually interested in just a minor percentage of the data kept in a catalog to pursue their research. We might need a few "columns" of a catalog, and maybe just a small subset of the catalog's elements. There are currently a few efforts in the world pointing in that direction, VizieR at CDS, Wizard at NASA, Pizazz at NCSA Skycat at STScI, Starlink at the UK, *dat*OZ,

at the University of Chile, and possibly others. All of them should point to create a uniform method of data retrieval.

In the following sections an overview of the features of databases created with *dat*OZ will be presented. Detailed information can be found in the "user's manual" for the system, where static and dynamic databases are discussed.

2.   What is *dat*OZ and what does *dat*OZ offer?

*dat*OZ is a computer tool which creates the interface and the source code capable of handling a specific catalog (Ortiz 1997). The source code is written in C, with some specific routines written in Perl. The data is stored in machine format, with some degree of indexation for fast access, a key for large catalogs. The modified version of *dat*OZ for large catalogues fully indexes the data-file, introduces compressed variables and reduces the number of significant figures for RA and Dec without degradation of the positions.

Each database element can have associated several "multimedia" files (.gif, .jpg, .ps, spectra-like data files, and a note-pad for extra annotations), and supports a kind of variable which allows hyper-links to other sources of information. The system was built with the capability of accessing interfaces in more than one language; something particularly useful if this system is to be accessed by people in countries where English is not the official language.

*dat*OZ offers the user a database with fast and flexible data retrieval and visualization capabilities; it also offers uniformity, as all databases look alike and behave in the same way. These databases can receive requests from related tools developed with the purpose of combining information amongst databases created with *dat*OZ. Links to these tools are found in the homepage.

*dat*OZ's home page is: **http://146.83.9.18/datoz_t.html**

3.   *dat*OZ Retrieval Modes

*dat*OZ' databases have a flexible retrieval access mode also known as Advanced Access Mode (**AAM**), which lets the user define the quantities he/she wants to retrieve amongst the "natural" database variables or by using mathematical expressions formed with these variables. The opposite of the **AAM** is the Basic Access Mode (**BAM**) which restricts operations to the values of the database variables.

The reason for implementing the flexible access mode has to do with the fact that catalogs are usually created with one purpose in mind, and only certain quantities become the catalog's variables. A typical case is when the catalog contains "variables" for **V**, **(B-V)**, **(V-R)**, and **(V-I)**, but if the user needs the magnitude of an object in the **I** band he/she would normally have to retrieve **V** and **V-I** and then perform the difference by him/herself. The **AAM** lets the user specify the quantity to get as **V - (V-I)**, i.e., **I**. The way to express this quantity is using Reverse Polish Notation (RPN) (complete details are found on the manual and HTML help pages).

Another advantage of the **AAM** is that the user can impose constraints on math expressions formed with "numerical" variables. Let's assume that, from
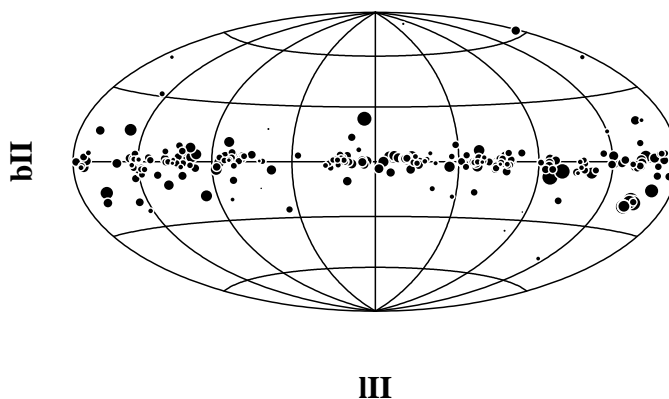
bII

lII

Figure 1.   Distribution of O stars in the galaxy.

the same catalog, we need sources brighter than 17 in the **I** band, but as we saw, **I** is not one of the catalog variables. We can define the expression to use as a constraint as: $\boxed{\text{V V-I -}}$ and impose it to be $\boxed{< 17}$.

One of the most important pluses of the **AAM** is that it lets the user submit a file with a list of the names of the objects which he/she needs to get information for, and get whatever data is stored in the database for those objects only. This is particularly useful when we deal with catalogs with a large number of objects, such as Hipparcos, or ROSAT for example. The files are transfered from the computer where the user is running **Netscape** to the database's host computer and then analyzed locally.

4.   *dat*OZ Visualization Tools

Visualization is the key to getting most of the science out of the data stored in a catalog. It allows quick checking of determined properties, and it is also a valuable teaching tool. *dat*OZ provides its databases with the capability to use any numerical variable to generate fully customized PostScript plots.

There are several types of plots that can be obtained: **histograms**, **scatter plots**, **scatter plots with variable symbol size**, **scatter plots with error bars**, **annotated scatter plots** (where the user prints the value of a variable instead of a symbol), **All-sky, equal area projection plots**, and **line plots**.

Plots can be viewed on the spot with `ghostview`, `ghostscript`, or any other PostScript previewer, and/or saved to disk using any of the ways Netscape provides.

5.   Searching for Neighbors with *dat*OZ

The word "neighbor" must be understood in a broad sense here. It could mean, of course, objects "close" in the sky, or it could apply to closeness in any numerical variable or expression.

The neighbor search operation is extremely valuable to explore the content of a catalog for diverse purposes. Besides using elements in the catalog as reference points, it is possible to provide a list of coordinates and ask which objects in the catalog are located near these reference points. The submission of files with lists of reference positions is an extremely valuable task to quickly search for matches. This feature is one of the most used in the catalogs installed in our system.

6.   Conclusions

*dat*OZ has proved to be a very powerful tool to handle astronomical information, at the present time, the following catalogs are implemented in our Department (Cerro Calán): **Galactic Globular Clusters**, Harris W.E. Astron. J. 112, 1487 (1996). **QSO's and AGNs** Veron-Cetty M.P., et al. (1996). **ROSAT Bright Sources Catalog** Max-Planck-Institut fuer extraterrestrische Physik, Garching (1996). **Catalog of Principal Galaxies**, Paturel et al. 1989 **Third Revised Catalog (RC3)**, de Vaucouleurs G. et. al. (1991). **AAT Equatorial photometric calibrators**, Boyle et al.'s (1995). **A Catalogue of Rich Clusters of Galaxies**, Abell G.O., Corwin Jr. H.G., Olowin R.P. (1989) **IRAS Catalog of Point Sources**, V2.0 Joint IRAS Science W.G. IPAC (1986). **Hipparcos Main Catalog** ESA, (1997). **Spectrophotometric Standards used at CTIO MK Spectral Classifications** - Twelfth General Catalog Buscombe W., and Foster B. E. (1995). **Northern Proper Motions**, **Positions and Proper Motions, SOUTH**, Bastian U., and Roeser S. (1993). **Catalog of 558 Pulsars**, Taylor J.H. et al. (1983). **Line Spectra of the Elements**, Reader J., and Corliss Ch.H. (1981). **USNO's A1.0 − 488 million stars**, Monet, D. (1996).

The near future will bring big survey works (Sloan, 2DF, 2MASS, AXAF, and EROS, to name a few). Making the data available to the community will be crucial to increase these projects' science impact. The current trend of having data available on CD's or ftp sites may discourage an important number of scientists to explore or analyze the data, tools like *dat*OZ will make the data readily available to anyone.

7.   Acknowledgments

References

Ortiz, P. F. 1997, in Proceedings of the Fifth workshop on Astronomical Data Analysis, Erice, 1996.