Astronomical Data Analysis Software and Systems XII
ASP Conference Series, Vol. 295, 2003
H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds.

SkyDOT: A Publicly Accessible Variability Database, Containing Multiple Sky Surveys and Real-Time Data

D. Starr, P. Wozniak, W. T. Vestrand

Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract. SkyDOT (Sky Database for Objects in Time-Domain) is a Virtual Observatory currently comprised of data from the RAPTOR, ROTSE I, and OGLE II survey projects. This makes it a very large time domain database. In addition, the RAPTOR project provides SkyDOT with real-time variability data as well as stereoscopic information. With its web interface, we believe SkyDOT will be a very useful tool for both astronomers and the public.

Our main task has been to construct an efficient relational database containing all existing data, while handling a real-time inflow of data. We also provide a useful web interface allowing easy access to both astronomers and the public. Initially, this server will allow common searches, specific queries, and access to light curves. In the future we will include machine learning classification tools and access to spectral information.

1. Introduction

A new phase in optical astronomy research is occurring from increases in computing power and the financial feasibility of massive data collecting projects. Successful experiments such as micro lensing searches and galactic surveys have developed because of this availability of measurements. But, the rapid increase in these types of projects creates a need to organize and commonly collect their data. With such a resource then publicly available, it can be queried and used to its fullest potential. These collections of datasets are often called "virtual observatories."

Although there are currently dozens of variability sky surveys underway, the usefulness of their datasets is often limited by their frequency of observations, spatial coverage of the sky, and their availability to other astronomers. To this date, there are very few large sky surveys which are collecting high temporal-resolution variable datasets, while also providing its access to the public. SkyDOT will fill this niche by allowing general access to a large spatial database built from multiple sky surveys. One of which continually produces new data with a frequency of around once a minute.

2. Surveys Currently Used

Within the database, SkyDOT's component surveys will be similarly structured, making them compatible with it and each other. This makes integrating additional projects fairly easy, and multi-survey queries simpler to implement. Although, of the three currently available datasets, the OGLE-II project's relative simplicity requires a slightly less complicated table structure.

2.1. OGLE II

OGLE II is a galactic bulge variable star survey (Wozniak, Udalski et al. 2002) which was conducted with the 1.3 m Warsaw telescope at Las Champanas Observatory, Chile. Started in 1997, this 3 year project covers 11 square degrees and contains over 220,000 I-band light curves with magnitudes between 10.5 and 20.0. The 49 constituent fields contain between 200–300 measurement frames: although 10% of the objects recorded are spurious and most likely non-stellar.

2.2. ROTSE I

Our database will also contain a year of data from the Robotic Optical Transient Search Experiment I (ROTSE-I) (Wozniak, Akerlof et al. 2002), which operated for four years in Los Alamos, NM. This survey contains 644 fields, each being 8×8 degrees in size and sampled with between 300 and 40 frames, for declinations of +90 and −30 degrees, respectively. The total number of objects with time histories is expected to be 20 million and it is estimated that around 32,000 of these will be periodic variable stars.

2.3. RAPTOR

The Rapid Telescopes for Optical Response (RAPTOR) sky monitoring experiment (Vestrand et al. 2002) will be our database's third source of data. This LANL optical transient search project will run concurrently with SkyDOT, providing a real-time stream of photometric measurements. RAPTOR uses two identical telescopes, separated by 38 km, to stereoscopically view the same region of sky. After software processing, this binocular information can be used to separate potentially interesting objects, leaving the more distant, stellar data.

Each of the telescope pair has 4 85 mm cameras, which cover 1500 square degrees and reach a depth of 13th magnitude (Wren et al. 2002). They also contain a central 400 mm camera with a sensitivity to 16th magnitude. The combined pair's 10 cameras take 30–60 second exposures which are then immediately reduced, the sources extracted, and an object catalog is updated—before the next image is received. Transient objects found for a camera are then compared with concurrent data from the sister telescope, to determine whether the object is a possible celestial transient. If so, these interesting objects can be used to alert the rest of the project and possibly other astronomical institutions.

Besides the telescope pair, RAPTOR also has a separate large field "Patrol" telescope, as well as a rapid response spectroscopic telescope. The patrol telescope can record up to 16th magnitude objects and will independently cover the local sky in about 2–3 days. This will provide roughly 30 million objects with time histories. Our 0.3 m Ritchey-Chretien rapid response telescope is equipped

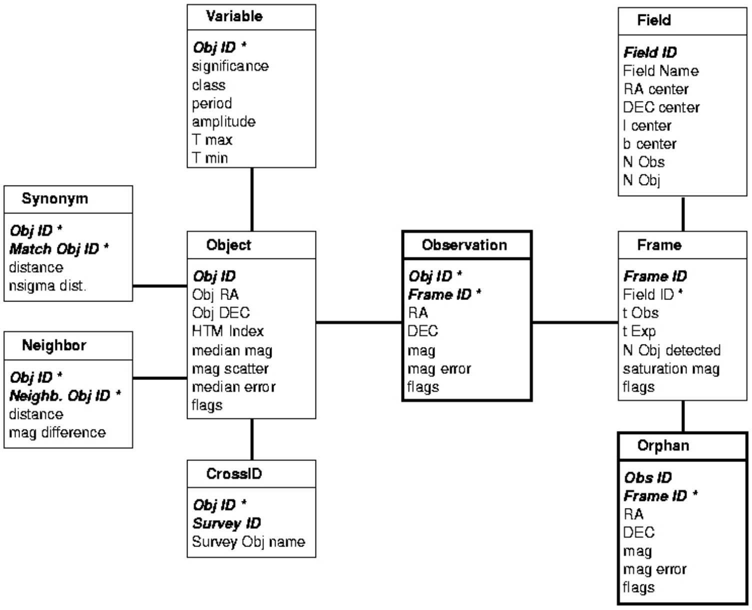

Figure 1. Database Schema.

with a transmission grating, allowing low resolution spectroscopy. This will be used to respond to our alerts and other selected objects.

3. Database Design

Our database is built using PostgreSQL, a free, open source, relational system that also has object oriented capabilities. Although an object oriented model is most natural for using classes containing parameters and their transformations, performance problems were reported (Szalay et al. 2002). For our task, the relational system is more successful.

The database is structured around the "Observation" main event table which contains all measurements. This table is indexed with two primary keys (Obj ID & Frame ID) creating access points for the "Object" and "Frame" dimension tables. Most searches utilize the object information in tables surrounding the dimension tables. When a query is made, the observed data is retrieved through the dimension table's link to the observation table. This snowflake structured database results in a fairly compact system. Although a star shaped scheme with more redundancy of data would allow faster general queries, we chose to conserve disk space since much of our searches will be done with an optimized index for spatiality.

Spatial queries are expected to be our most common searches. So we need an efficient method for finding objects within a circle on a sphere. PostgreSQL contains a few tree searching methods, but they do not pertain to 2D searches of high resolution positions. For this reason, we implemented Hierarchical Triangular Mesh (HTM) indexing, developed at Johns Hopkins University (Kunszt et al. 2000).

The HTM search algorithm partitions a sphere up into triangles, which are recursively divided into finer triangles. The algorithm identifies every triangle with a position related HTM ID. So, a triangle with a specific position and resolution will contain within its HTM ID the shorter IDs of its parent triangles. HTM IDs are also B-tree indexed, allowing efficient searches for objects within a triangle. Thus a circular query will return large and small triangles within the circle, which can then be searched for the individual object IDs they contain.

Tables surrounding the "Object" and "Frame" dimension tables allow a variety of searches. The "Variable" table contains typical periodic characteristics, as well as computer generated classes: created through machine learning algorithms. The "Synonym" table is used to relate an object which resides in multiple fields. This effect is fairly common in sky surveys especially in fields near the poles. Additionally, RAPTOR's overlapping cameras also compound the effect. This often leads to duplicate information, which may lead to improvements in data quality for that object. To search using the data from all three surveys, the "Cross ID" table is joined to the query. This then links an object to its counterpart in the other survey, if it exists.

On the other side of the primary observation data table is the "Frame" dimension table. It is related to the field table by the field ID index. The field Table is intended for system diagnostics and is not to be used in normal database queries. On the other hand, the "Orphan" table contains objects that are temporally unique and could not be matched with existing data. These objects, when processed using information from RAPTOR's stereoscopic system, can produce interesting alerts.

4. Summary

SkyDOT will be a powerful tool for astronomers. Its planned combination of three datasets makes it one of the largest variability databases. Plus, it will be able to include additional surveys, when they become available. The synoptic information from RAPTOR's binocular telescopes provides a wealth of information that might be used for intra-solar system searches. SkyDOT is also unique because its database will be updated by the minute with data from RAPTOR's real-time pipeline.

References

Kunszt, P. Z., Szalay, A. S., Csabai, I. & Thakar, A. R. 2000, in ASP Conf. Ser., Vol. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 141

Szalay, A. S. et al. 2002, Proc. of ACM SIGMOD Int. Conf.

Vestrand, W. T. et al. 2002, Proc. SPIE, 4845

Wren, J. et al. 2002, Proc. SPIE, 4845

Wozniak, P. R., Akerlof, C. et al. 2002, Bull. Am. Astron. Soc., 200

Wozniak, P. R., Udalski, A. et al. 2002, Acta Astronomica, 52, 129