DataSeeker: A Web Interface to Chandra M&TA Data

Ryan S. Overbeck, Scott J. Wolk, Mark Cresitello-Dittmar, James G. Petreshock, Brad Spitzbart

Harvard-Smithsonian Center for Astrophysics, 60 Garden St. Cambridge, MA, 02138 USA.

Abstract. *Chandra*'s Monitoring and Trends Analysis (M&TA) team utilizes a series of database tables to store binned information about spacecraft configuration and condition over the life of the mission. The DATASEEKER was developed as an efficient means to specify a particular set of this data and have it conveniently downloaded in an usable format.

In *Chandra*'s brief life of a little over two years in orbit, there are already well over 10 GB of data stored in approximately 50 relational database tables for the Science Operation Team to sift through. DATASEEKER has made it possible to discriminately pull out small subsets of these data in a matter of minutes.

1. Introduction

The Monitoring and Trends Analysis (M&TA)[1] group at the *Chandra* X-ray Center[2] studies the on-orbit performance of the *Chandra* X-ray Observatory and the effects of current spacecraft state on *Chandra*'s ability to collect scientifically useful data. We perform this task by analyzing engineering and science data describing spacecraft state and configuration across the whole mission. The M&TA Databases (Wolk et al. 2001) are the final repository for this data.

The DATASEEKER application is the primary access point to the M&TA Databases. At any given time, *Chandra* science personnel are only interested in a small subset of M&TA data, whether it be the spacecraft's main body temperature during a particular observation, or *Chandra*'s altitude while its High Energy Transmission Grating (HETG) is inserted. DATASEEKER's goal is to provide *Chandra* science personnel with a simple yet powerful method of extracting discriminate subsets of M&TA data from the databases both interactively and automatically.

DATASEEKER's modularity is integral in achieving this goal. The databases were built as a set of Sybase[3] databases using an intuitive 3-tier architecture. DATASEEKER, written in Perl, uses DBI to connect to the database server and

[1] `http://cxc.harvard.edu/mta`

[2] `http://cxc.harvard.edu`

[3] `http://www.sybase.com`

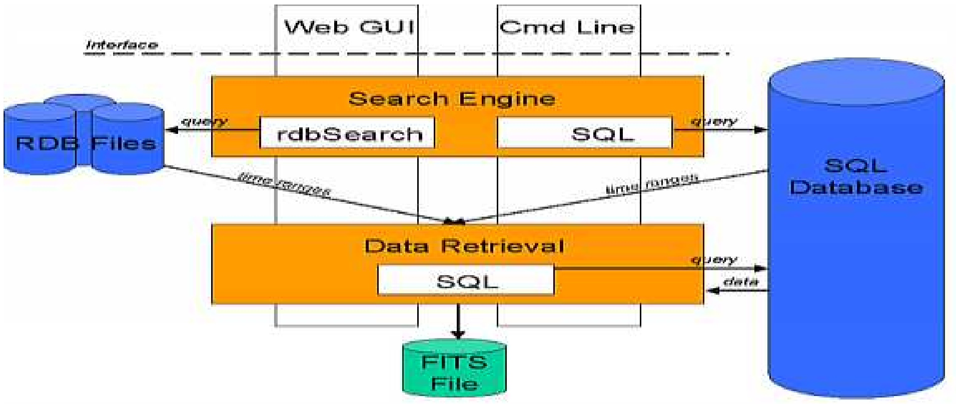

Figure 1. DATASEEKER's architecture.

execute Sybase SQL queries. It uses HTML to present an user-friendly 3-tier web interface, mirroring the underlying databases. By simply removing and replacing three modules, the same Perl code used to construct CGI DATASEEKER for the web is also used to construct command line DATASEEKER for scripts and other automated processes.

2. Architecture

2.1. Procedure

DATASEEKER's procedure is summarized in Figure 1. DATASEEKER first retrieves search criteria using either its web GUI or command line interface. Search criteria are then passed to the search engine.

The search engine performs all searches by finding time ranges which match the input criteria. The search engine is built with the capability of searching both RDB files and Sybase databases. RDB[4] is an ASCII table format. To search the RDB files, it must prepare and call the rdbSearch tool, a program written in C++ for fast search times. To search the databases it must prepare and execute appropriate SQL statements. The Perl DBI[5] module is used as the intermediary between Perl and the Sybase server. The resulting time ranges are merged and passed to the data retrieval engine.

The data retrieval engine builds and executes SQL queries to retrieve all requested data within the matching time ranges. Data are merged and placed in a single FITS binary table as rows are retrieved.

2.2. Data Resources

DATASEEKER's main data resources are the M&TA Databases (Wolk et al. 2001) contained in the *Chandra* Data Archive (Rots et al. 2000). The databases

[4] `http://hea-www.harvard.edu/MST/simul/software/docs/rdb.html`

[5] `http://dbi.symbolstone.org`

organize all M&TA data into a uniform format. All tables have a "time" column measured in seconds, and all table records lie on 300 second intervals. The databases are used for both search and data retrieval.

DATASEEKER allows for a secondary data resource in the form of RDB files. This exception is provided for rapid implementation and prototyping of new data tables. These files will be consumed by the M&TA average databases over the next year.

3. Interface

DATASEEKER offers two interfaces: a CGI web GUI, and a command line script. The web interface provides the user with physically meaningful values with which to query the databases. It's a dynamic interface allowing quick addition of new criteria and presenting the user with a minimized set of information at each stage. Command line DATASEEKER is ideal for longer queries and inclusion into scripts and other automated tasks.

3.1. Web GUI

The web interface uses the three tier query definition procedure presented by Wolk et al. (2001):

$$Primary \rightarrow Secondary \rightarrow Tertiary.$$

Primary criteria deal with main spacecraft information (eg. time range, spacecraft attitude and altitude, observation ID). All primary criteria are presented on the first DATASEEKER web page.

Secondary criteria deal with particular spacecraft hardware (eg. power supplies on/off, radiation flags, calibration modes). All secondary criteria are discrete. Most have bi-level values and all reside in the configuration database, one of several M&TA Databases. There are currently 63 mnemonics in 6 tables available as secondary criteria. An RDB table is used to map HTML input form elements to SQL database pathnames. This table keeps track of each mnemonic's pathname, possible values, and description. This allows for quick insertion of new secondary criteria.

Tertiary criteria describe the final output table contents and use the average databases in the M&TA Databases. There are over 350 columns from 35 tables in 8 databases to choose from. Tertiary criteria definition is split into two pages. The first page offers a selection of tables from which to retrieve data. The second page allows the user to choose individual columns from the selected tables. The structure of the databases used for tertiary criteria are maintained by another RDB file. This table is generated periodically by reading directly from the M&TA Databases. This way new tertiary criteria appear in DATASEEKER almost immediately after realization in the databases.

DATASEEKER maintains criteria selections between tiers to allow progressive refinement of a query. All criteria are passed to the main body of code to perform data search and retrieval, and a link to the final FITS file is provided.

3.2. Command Line

The command line interface is a one-step version of DATASEEKER. It is generated using the same Perl code files as the CGI DATASEEKER.

Criteria can be listed either as options on the command line or in an ASCII file, allowing simple inclusion into other tools. One such tool, mta_db_get[6], already exists.

4. Current State and Future Developments

Currently, the databases contain a little over one year's worth of data in 43 tables. DATASEEKER is capable of performing most common queries on these data in a matter of minutes. Particularly complicated queries can last over two hours. Most future changes will be to the underlying databases.

Efforts will be made to make DATASEEKER itself more flexible. Currently, all queries are logical AND. We hope to add logical OR and NOT to the arsenal. Also, an administrative interface will be added so different database servers can be accessed as needed. Currently, all users need to log into DATASEEKER using their own account for the M&TA Databases, however Several methods are under review to maintain the present level of security without having to supervise multiple database accounts.

References

Rots, A. et al. 2000, in ASP Conf. Ser., Vol. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 184.

Wolk, S. J. et al. 2001, this volume, [P-72].

Wolk, S. J. et al. 2000, in ASP Conf. Ser., Vol. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 453.

---

[6]`http://asc.harvard.edu/mta_days/mta_db/IDL/`