Astronomical Data Analysis Software and Systems X
ASP Conference Series, Vol. 238, 2001
F. R. Harnden Jr., F. A. Primini, and H. E. Payne, eds.

The Design of Solar Web, a Web Tool for Searching in Heterogeneous Web-based Solar Databases

Isabelle Scholl

Institut d'Astrophysique Spatiale, CNRS/Université Paris XI, Bâtiment 121, F-91405 ORSAY, France

Abstract. Today, scientists interested in solar data have access to a large variety of data collections (coming from space or ground-based observatories), but only available with different and heterogeneous software. Moreover, there is no common definition for keywords describing the data. Consequently, this makes the pre-analysis period long and difficult.

This paper presents the design and the architecture of the 'Solar Web Project' which aims to provide scientists interested in solar data with collaborative software for browsing distributed and heterogeneous solar databases using a unique web-based interface.

The project is developed at MEDOC (Multi-Experiment Data and Operations Centre), located at the Institut d'Astrophysique Spatiale (Orsay, France) which is one of the European Archives for SOHO (selected by ESA) and for TRACE data. JAVA and XML are technologies being used for designing and developing this tool.

1. Introduction

Today scientists interested in solar science are facing large amounts of data and many heterogeneous tools to identify and analyze them. Observations often arise from collaborative programs involving several observatories (space and/or ground-based). Cross-data analysis starts then with a heavy task: to identify and to retrieve data around several archive centers, each of them using different user interfaces, vocabularies, and semantics.

The tool developed for the SOHO archive[1] (Scholl 1998) at MEDOC (Scholl 1999) is a first step in the process of simplifying the access to heterogeneous data. SOHO is a spacecraft that has on-board 12 instruments, each of them having their own catalog and data in their own format. The software developed to browse the SOHO catalog and to retrieve files (including a web-based graphical user interface) was designed with open concepts in mind: this tool had to be generic enough to be able to easily manage 1) homogeneously all instrument catalogs and 2) other solar observations than those from SOHO (TRACE data are already available with this tool).

The Solar Web Project (Scholl 2001) ensues from: the acknowlegment that scientists have no unique or generic tool available for browsing several heteroge-

[1] `http://www.medoc-ias.u-psud.fr/archive/`

neous and worldwide distributed archives, the fact that an archive center should not store all mission data (data should be located near the relevant scientific experts), and the need to extend the SOHO archive access system to other solar archives in order to help scientists in their data search phase.

2. System Design

The aim of the 'Solar Web' tool is to provide users with one unique interface to access several heterogeneous and distributed archives. This assumes the development of a software with capabilities of sending multiple queries over the Internet, waiting for remote results before merging them and finally displaying them on one single view. But it also assumes that these archives are coordinated in order to standardize communications and catalogs.

The project is divided in 2 main parts: 1) scientific work: definition of a common model based on a standardized vocabulary for describing a 'Generic Solar Observation', and 2) technical work: software design and development.

The first step is to defined what a 'Generic Solar Observation' means. This phase consists in classifying a few parameters which are sufficient for the purpose of selecting the data. Around twenty keywords are grouped in 4 categories: General Parameters, Observational Technical Parameters, Scientific Parameters, and Data Management Parameters.

The purpose of this model is to help communications between existing archives. By using such a model, existing archives will be able to communicate in a standard way, even if some of them would probably not have all this information or will have to convert their own keywords to this model.

3. System Architecture

The architecture of the 'Solar Web' tool is based on a collaborative system. There are 2 main components:

- the 'Application Server'. Its role consists in accepting connections from clients and managing their queries (local and remote). It is installed under the HTTP server tree.
- the 'Web-User Interface' or client module.

The specificity of the 'Application Server' resides in its capability to communicate with other kinds of servers. It can communicate with databases using four different methods (see Figure 1):

1. the server directly accesses the local database (DB) using the JDBC API,
2. the server directly accesses registered remote databases that offer a direct JDBC access,
3. the server accesses existing cgi scripts that query their local databases (located on remote web-servers) using an HTTP GET method, and
4. the server communicates to any other 'Application Servers' that query their local databases.

Items 2 and 3 are free of developments on remotes sites.

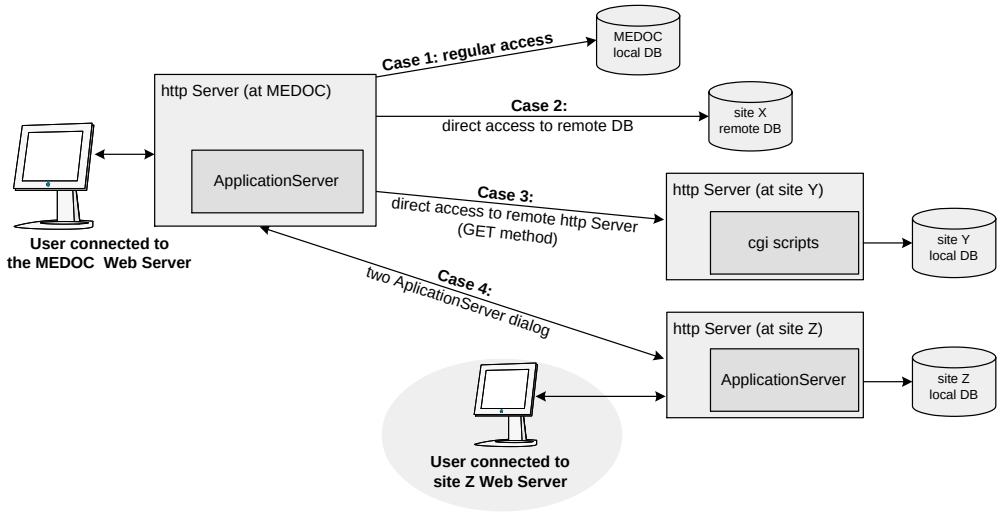

Figure 1. General System Architecture.

4. Software Architecture

This software is based on a 3-tier and distributed architecture. It has been designed to be open to new kinds of data (see Figure 2).

On the server side, the 'Application Server' fulfils the following roles:

- A Session System which manages sessions with clients using a unique ID
- A Query System composed of:
  - a Query Generator (QG) preparing SQL queries for a specific DB. Each DB structure is implemented in dedicated classes and provided by the QG as an API, and
  - a Query Processor which runs queries, gets answers, formats results with a DataBase Manager for local DB access and a Query Router for remote DB access.
- A Cache System which manages buffers associated to clients

On the client side, the applet is dedicated to the powerful query capture and result visualization. The client (excepted for the dialog with the server) and the query generator are the only business specific components. These two modules will be available as an API of the generic software to implement extensions.

Exchanges between the main server and remote applications are made using either a native application protocol to simplify development, or XML with a standardization objective: the 'Generic Solar Observation' is implemented as an XML DTD.

Security is a major concern in this project. Permitting direct remote accesses is always a potential security risk for any organization. A secure protocol for communications between remote servers is currently under study.

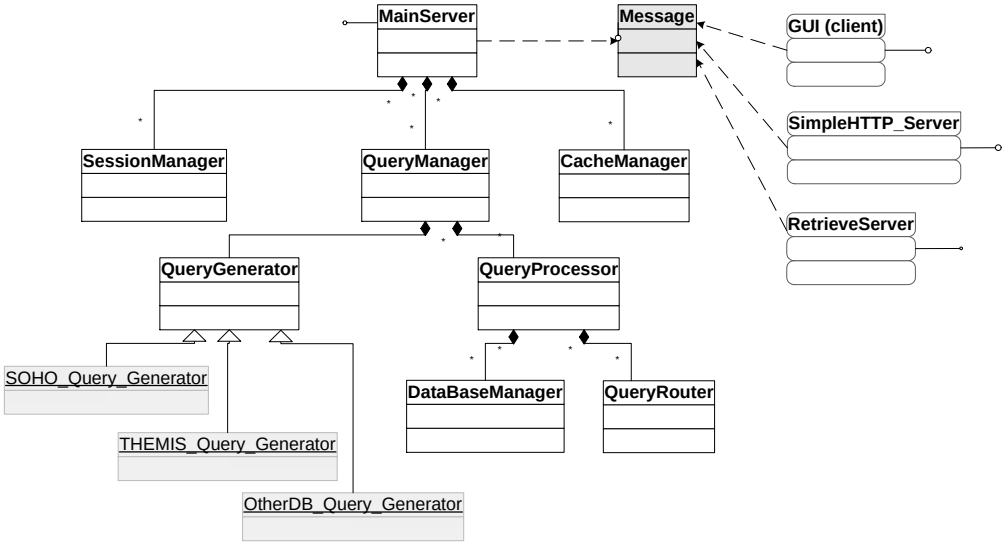

Figure 2. Software Architecture.

5. Conclusion

'Solar Web' is currently a stand alone development effort from MEDOC. Such another project exists, the Virtual Solar Observatory (Hill 2001), but it is still at a very early stage. Discussions are in progress with potential partners. Tests will be made with Bass2000 (Themis data, at Tarbes, France) to experiment with the communication between MEDOC and an existing remote web-server. The status of the project can be found at `http://www.medoc-ias.u-psud.fr/archive/solarweb/`. We are open to suggestions for new collaborations.

References

Scholl, I., et al. 1998, in ASP Conf. Ser., Vol. 172, Astronomical Data Analysis Software and Systems VIII, ed. David M. Mehringer, Raymond L. Plante, & Douglas A. Roberts (San Francisco: ASP), 253

Scholl, I. 1999, in Proceedings 8th SOHO Workshop, (ESA Publ. SP-446) (Paris 1999), 611

Hill, F. 2001, in ASP Conf. Ser., Vol. 225, Virtual Observatories of the Future, ed. Robert J. Brunner, S. George Djorgovski, & Alexander Szalay (San Francisco: ASP), in press

Scholl, I. 2001, in ASP Conf. Ser., Vol. 225, Virtual Observatories of the Future, ed. Robert J. Brunner, S. George Djorgovski, & Alexander Szalay (San Francisco: ASP), in press