Astrobrowse: A Multi-site, Multi-wavelength Service for Locating Astronomical Resources on the Web

T. McGlynn[1] and N. White

NASA/Goddard Space Flight Center, Greenbelt, MD 20771, Email: tam@silk.gsfc.nasa.gov

Abstract.
We report on the development of a Web agent which allows users to conveniently find information about specific objects or locations from astronomical resources on the World Wide Web. The HEASARC Astrobrowse agent takes a user-specified location and queries up to hundreds of resources on the Web to find information relevant to the given target or position. The current prototype implementation is available through the HEASARC and provides access to resources at the HEASARC, CDS, CADC, STScI, IPAC, ESO and many other institutions. The types of resources the user can get include images, name resolution services, catalog queries or archive indices.

The Astrobrowse effort is rapidly evolving with collaborations ongoing with the CDS and STScI. Later versions of Astrobrowse will use the GLU system developed at CDS to provide a distributable database of astronomy resources. The Astrobrowse agent has been written to be customizable and portable and is freely available to interested parties.

1. Introduction

The myriad astronomical resources now available electronically provide an unprecedented opportunity for astronomers to discover information about sources and regions they are interested in. However, many are intimidated by the very number and diversity of the available sites. We have developed a Web service, Astrobrowse[2], which makes using the Web much easier. The Astrobrowse agent can go and query many other Web sites and provide the user easy access to the results. In the next section we discuss the history and underlying philosophy of our Astrobrowse agent. The subsequent sections address the current implementation, status and future plans.

---

[1]Universities Space Research Association

[2]http://legacy.gsfc.nasa.gov

481

2.   Why Astrobrowse?

Astronomers wishing to use the Web in their research face three distinct problems:

Discovery Given the hundreds of Web sites available it is virtually impossible users to know of all the sites which might have information relevant to a given research project.

Utilization Even when users know the URLs of useful Web sites, each Web site has different formats and requirements for how to get at the underlying resources.

Integration Finally, when users have gotten to the underlying resources, the data are given in a variety of incompatible formats and displays.

As we began to design our Astrobrowse agent to address these problems we factored in several realizations: First, as we looked at the usage of our HEASARC catalogs we found that by about 20 to 1, users simply requested information by asking for data near a specified object or position. The particular ratio may be biased by the data and forms at our site, but clearly being able just to do position based searches would address a major need in the community.

Second, we saw that the CGI protocols are quite restrictive so that regardless of the appearance of the site, essentially all Web sites are queried using a simple keyword=value syntax. This commonality of interface presents a unique opportunity. Earlier X-windows forms that many data providers had created, and emerging technologies like Java do not share this.

Another consideration was that for a system to be successful, it should require only minimal, and preferably no effort, on the part of the data providers. We could not build a successful system if it mandated how other sites use their scarce software development resources.

Finally, and perhaps most important, we recognized that problem of integration is by far the most difficult to solve. Integrating results requires agreement on formats and names to a very low level. This is also an area which can require deep understanding of the resources provided so that it may appropriately be left to the astronomer. We would provide very useful service to users even if we only addressed the issues of discovery and utilization.

With these in mind, the outline of our Astrobrowse system was straightforward: Astrobrowse maintains a database which describes the general characteristics of each resource and detailed CGI `key=value` syntax of the Web page. It takes a given target position, and translates the query into the CGI syntax used at the various sites and stores the results. In current parlance, Astrobrowse is a Web agent which explodes a single position query to all the sites a user selects. Since very many, if not most, astronomy data providers have pages which support positional queries, Astrobrowse can access a very wide range of astronomy sites and services.

3. Implementation

The HEASARC Astrobrowse implementation has three sections: resource selection, where the user chooses the sites to be queried; query exploding where the positional query is sent to all of the selected resources; and results management, where Astrobrowse provides facilities for the user to browse the results from the various sites.

3.1. Resource Selection

Once the total number of resources available to an Astrobrowse agent grows beyond 10-20, it is clear that a user needs to preselect the resources to be queried. The current Astrobrowse implementation provides nearly a thousand resources. Querying all of them all of the time would strain the resources of some of the data providers and would also confuse the user. We currently provide two mechanisms for selecting resources. A tree of resources can be browsed and desired resources selected. Alternatively a user can search for resources by performing Alta-Vista-like queries against the descriptions of those resources. E.g., a user might ask for all queries which have the words 'Guide Star' in their descriptions. The user can then select from among the matching queries.

3.2. Query Exploding

The heart of Astrobrowse is the mechanism by which it takes the position or target specified by the user and then transforms this information into a query against the selected resources. For each resource the Astrobrowse database knows the syntax of the CGI text expected, and especially the format of the positional information, including details like whether sexagesimal or decimal format is used and the equinox expected for the coordinates. The current system uses a simple Perl Web query agent and spawns a separate process for each query.

3.3. Results Management

Astrobrowse takes the text returned from each query and caches it locally. If the query returns HTML text then all relative references in the HTML – which presumably refer to the originating site and thus would not be valid when the file is retrieved from the cache – are transformed into absolute references.

Our Astrobrowse interface uses frames to provide a simple mechanism where the user can easily switch among the pages returned. A number of icons return the status of each request, and allow the user to either delete a page which is no longer of interest, or to display it in the entire browser window.

3.4. The Astrobrowse Database

A database describing Astronomy Web sites is central to the functioning of Astrobrowse. For each resource, a small file describes the CGI parameters and provides some descriptive information about the resource. The file is human-readable and can be generated manually in a few minutes if one has access to the HTML form being emulated. We also provide a page on our Astrobrowse server to automatically build these files so that users can submit new resources to be accessed by our agent.
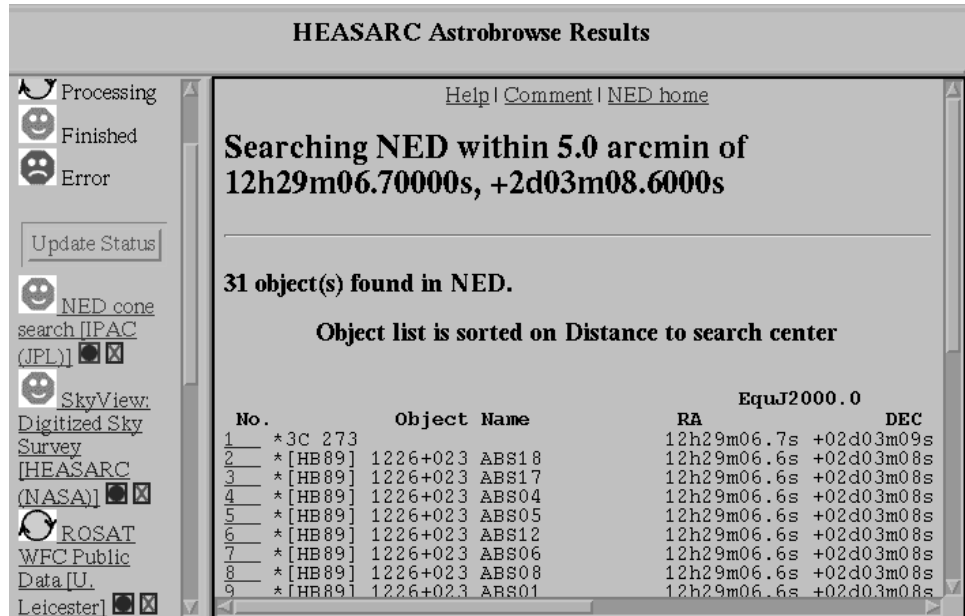
Figure 1.   An Astrobrowse Screen Shot.

4.   Future Plans

We believe the current Astrobrowse provides a convincing proof-of-concept for an astronomy Web agent and is already a very useful tool but we anticipate many changes in the near term. Among these are:

- Access to virtually all astronomical Web pages supporting positional queries. We expect to support about 3,000 to 10,000 resources compared to the current approximately 1,000 resources.

- Providing a few uniform keys for resources so that users can select data conveniently by waveband and data type.

- Conversion of the database to formats using the CDS's GLU format and use of the GLU system to manage distributed access and updates to the resource database.

- Substantial experimentation in the user interfaces to select resources and display results.

In the longer term we hope that Astrobrowse can be expanded beyond the limits of positional searches for astronomical resources and become the basis for tools to help integrate astronomy, space science and planetary data.