Astronomical Data Analysis Software and Systems VI
ASP Conference Series, Vol. 125, 1997
Gareth Hunt and H. E. Payne, eds.

Astronomical Information Discovery and Access: Design and Implementation of the ADS Bibliographic Services

A. Accomazzi, G. Eichhorn, M. J. Kurtz, C. S. Grant, S. S. Murray

Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138

Abstract. The NASA Astrophysics Data System integrates a wealth of scientific bibliographic and data resources—originally generated in multiple formats and available from multiple providers—in three discipline-oriented, centralized databases. Search and retrieval of the bibliographies and data sources is possible via a set of World Wide Web forms and interface programs that transparently link the ADS's resources to those of other data providers. The approach followed in designing the ADS system is offered as a paradigm for building flexible networked information and discovery systems. The rationale behind the current technical implementation and the planned enhancements of the system are also discussed.

1. Introduction

The design behind the Astrophysics Data System (ADS) bibliographic databases was mainly dictated by the desire for a powerful and discipline-oriented system featuring sophisticated search capabilities. The main considerations which shaped the final outcome of the system were: the advantages and disadvantages of using a commercial or publicly available RDBMS system versus a custom-build one; the quality and quantity of the data at hand versus the resources available to the project; and the tradeoff between search speed and simplicity on one hand and sophistication on the other.

General-purpose search engines and relational databases were used as part of the abstract service in the first implementation of the search engine, but they were eventually dropped in favour of a home-grown system as the desire for better performance and custom features grew with time (Accomazzi et al. 1995).

The heterogeneous nature of the bibliographic data that had to be entered into our database, and the need to effectively deal with the imprecision in it, lead us to design a system where a large set of discipline-specific interpretations are made. For instance, to cope with the different use of abstract keywords by the publishers, and to correct possible spelling errors and typos in text, sets of words have been grouped together as synonyms for the purpose of searching the databases. Also, many astronomical object names are translated in a uniform fashion when indexing and searching the database.

Because of the large number of features that we have been adding to the abstract service in the last few years, we had to strike a balance between simplicity of the user interface and the creeping featurism syndrome so commonly found in many user interfaces. To avoid overwhelming users with complex search pages, we have devised a design where the main search parameters are always visible within the top part of the screen, with more options to follow. Because of the very nature of the WWW, we have been able to create simpler HTML forms that have much of the additional functionality hidden from the user, and we now even allow users to create and customize their own search form according to their preferences.

In order to provide transparent access to our system from other WWW-based systems, we have provided access interfaces that use bibliographic codes (Schmitz et al. 1995)—or *bibcodes*, as referred to in the rest of this paper—as unique identifiers for references in our databases. Direct HTTP access to our CGI interface programs, and a high-level programming interface implemented as a library of Perl routines, are provided as hooks into our bibliographic search engine.

2. Database Search Interface

The ADS CGI interfaces implement a variety of possibly complex searches of the bibliographic databases, but searches can generally be divided in two classes: reference searches and concept searches.

2.1. Reference Searches

This type of interface allows users to lookup a particular publication or to browse a set of references published in a journal. Access to the program that implements this interface is available by retrieving the URL:

`http://adsabs.harvard.edu/cgi-bin/abs_connect?bibcode=`*bibcode*

where *bibcode* is either a fully qualified, 19-digit bibliographic code, a partial bibcode, or a bibcode pattern possibly containing metacharacters. Consider, for instance, the cases where *bibcode* is one of the following:

- `1996adass...5..558A`: the URL contains a fully-qualified bibcode, and therefore it refers to an individual paper published by the Author at the ADASS V conference in 1996.

- `1996adass...5`: the URL contains a bibcode stem (i.e., truncated bibcode), and will therefore generate the list of publications whose bibcodes begin with the string `1996adass...5`. This list consists of all the papers published at the ADASS V conference.

- `199?adass...?`: the URL consists of a bibcode pattern containing two instances of the "?" metacharacter which matches any single character. The set of references returned by the query will be the list of papers published in all ADASS conferences so far (`1992adass...1`, `1993adass...2`, `1994adass...3`, `1995adass...4`, `1996adass...5`, which currently happen to match the above regular expression.

Other similar programs and HTML forms extend these capabilities by allowing selections based on publication date ranges and journals (see, for instance, the ADS Table of Contents Query Form[1]).

2.2. Concept Searches

Searches based on the identification of a set of references which are relevant to a particular topic or "concept" are implemented in a similar fashion. Because references are structured entities having several attributes (or "fields"), a fielded search is one in which one or more fields are to be searched and one or more terms to be searched for are specified for each field.

Currently the ADS Astronomy database allows users to search by author name, astronomical object name, keywords,[2] words in the title, and words in the abstract text. The general URL syntax for searching for terms in a particular field is `http://adsabs.harvard.edu/cgi-bin/abs_connect?`*field=words* where *field* is the name of the field to be searched and *words* represents the expression to be searched for. For instance, to find the list of all papers published by the Author in the ADS Astronomical Database, one would access the URL `http://adsabs.harvard.edu/cgi-bin/abs_connect?author=accomazzi`.

When specifying more than a single word to be searched in a particular field, the interface allows the user to select whether the resulting list is to include references which contain a subset of the search terms, which search terms must be present, and which should be excluded. When specifying words to be searched in separate fields, the user may choose how the lists of references resulting from the individual field searches should be combined, using a logic similar to the one applied for combining references generated from individual words within a field.

The ADS abstract service search form has many more features and settings that can be set customized, including restricting the search to be performed only on a particular journal or body of literature (e.g., searching on refereed journals only). One immediate application of this is that it provides users with several up-to-date indexes into subsets of the astronomical literature. For instance, to search for all the publications appearing in the ADASS conference series that mention ADS in their abstract, one would simply call the `abs_connect` script with the arguments: `text=ADS&jou_pick=YES&ref_stems=adass`.

3. Links to Bibliographic Resources

One of the most successful features of the design behind the current WWW software agents is that they allow users to transparently browse information available on the Internet via the selection of hyperlinks. In particular, this has created a de-facto standard interface and protocol for accessing network resources available from different institutions, thus becoming the glue between the services provided by different astronomical data centers.

---

[1] http://adsabs.harvard.edu/toc_service.html

[2] Keyword searching is currently not available from the main abstract service search page because of the lack of a uniform and consistent keywording system for the current references.

The ADS databases currently maintain for each bibliography a set of links to both local and network-accessible resources. The following hyperlinks provide interconnectivity between the ADS and other institutions:

- Electronic article links, which point to the full-text electronic version of the current reference, when available from the original publisher or from the ADS article service.
- Data links, which point to the list of electronic datasets published with the article, allowing retrieval of each of them. Currently these resources are available from the following institutions: CDS, NCSA/AIDL, GCIP.
- Object links, which point to the list of objects cited in the article, and available from the SIMBAD database. Available from the CDS.

The relationship between the ADS system and the data centers mentioned above is reciprocal, in the sense that they, in turn, provide hyperlinks from their databases to bibliographic resources available in the ADS, when appropriate.

4. Conclusions

The popularity and usefulness of the ADS bibliographical services is due, in large part, to its discipline-specific features and to the synergy created by several data centers adopting a common language and protocol to link their resources. This cooperation provides astronomers an ever-growing wealth of information and resources that are transforming the way they perform their research.

Because of the size and completeness of its databases, the NASA Astrophysics Data System has become a clearinghouse for astronomical bibliographic resources, and the ADS abstract service has become the bridge between networked resources available from different institutions and societies.

Acknowledgments. This work is funded by the NASA Astrophysics Program under grant NCCW-0024.

References

Accomazzi, A., Grant, C. S., Eichhorn, G., Kurtz, M. J., & Murray, S. S. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 36

Schmitz, M., Helou, G., Dubois, P., LaGue, C., Madore, B., Corwin Jr., H. G., & Lesteven, S. 1995, in Information & On-line Data in Astronomy, ed. D. Egret & M. A. Albrecht (Dordrecht: Kluwer Acad. Publ.), 271