

## New Capabilities of the ADS Abstract and Article Service

G. Eichhorn, A. Accomazzi, C.S. Grant, M.J. Kurtz and S.S. Murray  
*Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge,*  
*MA 02138, Email: [gei@cfa.harvard.edu](mailto:gei@cfa.harvard.edu)*

**Abstract.** The ADS abstract service at: <http://adswww.harvard.edu> has been updated considerably in the last year. New capabilities in the search engine include searching for multi-word phrases and searching for various logical combinations of search terms. Through optimization of the custom built search software, the search times were decreased by a factor of 4 in the last year.

The WWW interface now uses WWW cookies to store and retrieve individual user preferences. This allows our users to set preferences for printing, accessing mirror sites, fonts, colors, etc. Information about most recently accessed references allows customized retrieval of the most recent unread volume of selected journals. The information stored in these preferences is kept completely confidential and is not used for any other purposes.

Two mirror sites (at the CDS in Strasbourg, France and at NAO in Tokyo, Japan) provide faster access for our European and Asian users.

To include new information in the ADS as fast as possible, new indexing and search software was developed to allow updating the index data files within minutes of receipt of time critical information (e.g., IAU Circulars which report on supernova and comet discoveries).

The ADS is currently used by over 10,000 users per month, which retrieve over 4.5 million references and over 250,000 full article pages each month.

### 1. Introduction

The Astrophysics Data System (ADS<sup>1</sup>) provides access to almost 1 million references and 250,000 scanned journal pages (Eichhorn 1997). These can be accessed from the World Wide Web (WWW) through a sophisticated search engine (Kurtz et al. 1993), as well as directly from other data centers through hyperlinks or Perl scripts (Eichhorn 1996). Our references in turn link to other data and information sources. This cross-linking between different data systems provides the user with the means to find comprehensive information about a given subject.

---

<sup>1</sup><http://adswww.harvard.edu>

## 2. New Search Features

### 1. Complex Query Logic

The search system allows the user to specify complex queries in two forms:

- (a) Simple logic: This allows the user to specify that certain words must appear in the selected reference (+word) or must not appear in the select reference (-word). Phrases of multiple words can be specified by enclosing the words in double quotes. An example is in Figure 1 in the title field:

+“black hole” -galaxies +=unstable

This query searches for references that contain the phrase “black hole”, but not the word “galaxies” or its synonyms (like “galaxy” or “galactic”). They must also contain the word “unstable”. The ‘=’ before “unstable” turns off the automatic synonym replacement. It is not sufficient for a reference to contain a synonym of “unstable”.

- (b) Full boolean logic: This allows the user to use “AND”, “OR”, “NOT”, and parentheses for grouping to build complex logical expressions. An example is in Figure 1 in the abstract text field:

(“black hole” or “neutron star”) and (“globular cluster” or binary) and not (“cataclysmic variable” or CV)

This expression searches for references that contain one of the expressions “black hole” or “neutron star” as well as either the expression “globular cluster” or the word binary, but neither the expression “cataclysmic variable” nor the word CV.

### 2. Object Queries

- (a) Extra-Solar system objects: By selecting SIMBAD and/or NED above the input field for object names, the user can query for objects outside the solar system.
- (b) Solar system objects: By selecting LPI, a small database of meteorite and lunar sample names can be queried.
- (c) IAU objects: By selecting IAU, the database of names that have appeared in IAU Circulars is searched.

## 3. Indexing

- 1. Preprint Database: We are now indexing the preprints from the Los Alamos preprint server on a regular basis. Every night after the preprint server is updated, we automatically retrieve the articles from the preprint server and index them into a separate database. This database can be searched through the same interface as our other databases (Astronomy, Instrumentation, and Physics/Geophysics).
- 2. Mirror Sites: The mirror sites and the associated updating procedures are described by Accomazzi et al. (this volume).

3. Quick Updates: The quick updates allow us to quickly enter new data into the database. This was mainly developed to enable us to include IAU Circulars into the ADS within minutes after publication. Normally a full indexing of our database takes more than one day. Quick updates append new index sections to the original ones and link these new sections to the existing ones. The searches are not noticeably slowed down by these additional links. Every two to three weeks we re-index the complete database to include these additional sections in the main index.

#### **4. New User Interface Features**

We implemented WWW cookies in the ADS user interface. WWW cookies are a system that allows us to identify individual users uniquely, even without knowing who they are. This allows us to customize our response to individual users. Users can for instance select which mirror sites they prefer for some external services, how to print articles by default, and how the pages of the system should look (font sizes, colors, etc). The system also remembers through the cookie mechanism which tables of contents have been retrieved by the user for several different journals. Lastly, we can send out one-time messages to users. The ADS remembers which message has already been sent to a user. The user database is of course completely confidential and will not be made available to any other system.

#### **5. Future Plans**

Two major projects that will require software development in the future are the OCRing (Optical Character Recognition) and indexing of the historical literature once we have scanned it, and the parsing of the scanned articles for references.

In the next year we plan to scan several major journals back to volume 1. We will OCR these scans and make the OCR'd text available to selected researchers.

The other major project is the parsing of the references from the scanned literature. This will allow us to update and expand the reference and citation lists that are already available. This will be a very difficult task and there is no time line yet as to when we will be able to get useful data from that project.

**Acknowledgments.** This work was funded by NASA under grant NCCW 00254.

#### **References**

- Eichhorn, G., 1997, *Astroph. & Space Sci.*, 247, 189  
Kurtz, M.J., et al. 1993, in *ASP Conf. Ser.*, Vol. 52, *Astronomical Data Analysis Software and Systems II*, ed. R. J. Hanisch, R. J. V. Brissenden & Jeannette Barnes (San Francisco: ASP), 132

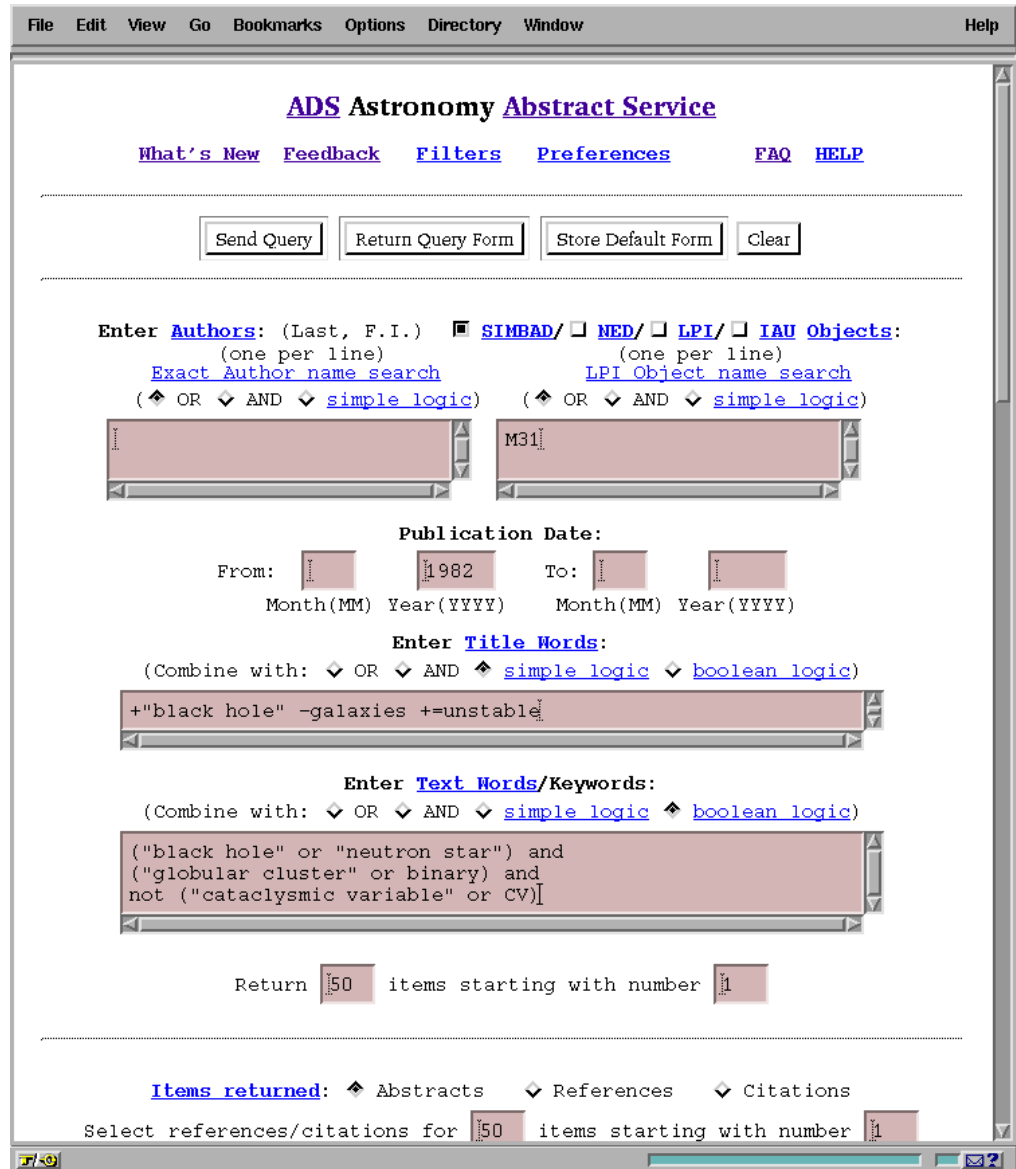


Figure 1. ADS Query page with examples of logical query constructs.

Eichhorn, G., et al. 1996, in ASP Conf. Ser., Vol. 101, Astronomical Data Analysis Software and Systems V, ed. George H. Jacoby & Jeannette Barnes (San Francisco: ASP), 569

Accomazzi, A., et al., 1998, this volume