

**FULL TEXT SEARCHING AND CUSTOMIZATION IN THE NASA ADS ABSTRACT SERVICE.** G. Eichhorn, A. Accomazzi, C.S. Grant, M.J. Kurtz, E.A. Henneken, D.M. Thompson, S.S. Murray, *Harvard-Smithsonian Astrophysical Observatory, Cambridge, MA 02138, USA, (gei@cfa.harvard.edu).*

## Introduction

The NASA-ADS Abstract Service provides a sophisticated search capability for the literature in Astronomy, Planetary Sciences, Physics/Geophysics, and Space Instrumentation. The ADS is funded by NASA and access to the ADS services is free to anybody world-wide without restrictions. It allows the user to search the literature by author, title, and abstract text.

The ADS database contains over 3.6 million references, with 965,000 in the Astronomy/Planetary Sciences database, and 1.6 million in the Physics/Geophysics database. 2/3 of the records have full abstracts, the rest are table of contents entries (titles and author lists only). The coverage for the Astronomy literature is better than 95% from 1975. Before that we cover all major journals and many smaller ones. Most of the journal literature is covered back to volume 1. We now get abstracts on a regular basis from most journals. Over the last year we have entered basically all conference proceedings tables of contents that are available at the Harvard-Smithsonian Center for Astrophysics library. This has greatly increased the coverage of conference proceedings in the ADS.

The ADS also covers the ArXiv Preprints. We download these preprints every night and index all the preprints. They can be searched either together with the other abstracts or separately. There are currently about 260,000 preprints in that database.

In January 2004 we have introduced two new services, full text searching and a personal notification service called "myADS". As all other ADS services, these are free to use for anybody.

## Full Text Searching

In order to provide the full text search capability we have translated all 2.5 million scanned pages into text through Optical Character Recognition (OCR). While this process does not yield a completely accurate reproduction of the text in the original papers, it has allowed us to create a searchable interface for this body of full-text. This allows ADS users to locate any page in the archive which contains a particular word or combination of words. Even though this interface is still under development, we encourage our users to test it and give us feedback. An added benefit of the OCR process is that we are often able to automatically extract and process reference lists and abstracts from the scanned papers, thus adding to the ADS citation and abstract databases.

## "myADS" Notification service

"myADS" is a personalized notification service. This service helps with maintaining current awareness of the recent technical literature in astronomy and physics based on each individual subscriber's queries. It delivers the results weekly to subscribers in html format via e-mail. Each week

the "myADS" Update Service scans the literature added to the ADS in the past seven days, and creates custom lists of recent papers for each subscriber, formatted to allow quick reading and access. The user can specify a set of authors that should be searched, and two subject matter queries. For the subject matter queries, "myADS" returns the most recent articles on this subject, the most recent preprints, the most popular articles in this field, and the most cited ones. In addition to these queries, "myADS" also returns the most recent articles that cite any of the users articles.

## References/Citations

An increasingly requested part of the abstract service is the database of references and citations. We have over 13 million reference-citation pairs in the database and are continually increasing the number of citations in the system. This database allows the user to get citations to an article and the list of references in an article. Despite the large numbers, users should be reminded that the citations contained in the ADS are incomplete due to the partial coverage of journals used to build the citation database and our inability to match 100% of references (e.g. works in press, private communications, author typos and other errors). Anyone using the citations for analysis of publishing records should keep this in mind.

## Scanned Articles

The second major part of the ADS is the Article Service. We have scanned over 2.5 million pages in over 330,000 articles from 40 journals, 15 conference proceedings series and 60 individual conference proceedings. Most of the societies have given the ADS permission to scan their publications. This includes for instance Meteoritics and Planetary Sciences, the Minor Planet Bulletins, most of the LPI Special Publications, all the LPSC conference abstracts, several books published by the LPI, The Antarctic Meteorite Research proceedings series, and the DPS Meeting abstracts. published in the Bulletins of the AAS. Other scanned publications include the Astrophysical Journal, Astronomical Journal, Astronomy and Astrophysics, Monthly Notices of the Royal Astronomical Society and the Publications of the Astronomical Society of the Pacific, the Zeitschrift für Astrophysik (with permission from Springer Verlag), Solar Physics (with permission from Kluwer Academic Publishers), the Minor Planet Bulletin, Astronomische Nachrichten (with permission from Wiley), and the Journal of the Royal Astronomical Society of Canada, all back to volume 1.

## Historical Observatory Publications

We have been collaborating recently with a preservation project at the Wolbach Library of the Harvard-Smithsonian Center for Astrophysics and the Harvard Library to microfilm

the historical observatory literature. We are getting a copy of each microfilm and have it scanned. We have currently over 50 observatory publications on-line with about 300,000 scanned pages. We expect to eventually scan up to 1 million pages of these publications. The scanned observatory publications will be an important part of the ADS. This literature is difficult to find and not available in many libraries. Having it available through the ADS will be important for scientists at smaller universities and for many non-US users. The drawback of the scanned observatory publications is that we currently do not have meta-data (page numbers, article information) for these scans. We have developed a user interface to enter meta-data for these publications, and would welcome any help with this task.

If you have a published book that is relevant to Astronomy and that is either not copyrighted (e.g. all NASA publications), or for which you can obtain permission from the copyright holder for ADS to scan the book, please contact the first author. If you can provide us with an extra copy of the book (it needs to be cut in order to be scanned), we will be happy to scan it. Some books can be rebound after scanning. In that case we will rebind the book and return it after scanning.

#### **Mirror Sites**

In order to provide better access to the ADS from different countries we have 11 mirror sites of the ADS, located in France, Japan, Chile, Germany, the United Kingdom, China, India, Russia, Brazil, Argentina, and South Korea. Of these mirror sites six (France, Japan, China, India, Russia, and South Korea) host a complete mirror with all scanned articles, two (the UK and Argentina) have a partial article mirror, the others mirror the abstract service. We would welcome suggestions for other

mirror sites.

#### **Access Statistics**

The ADS is used extensively by users world-wide. The ADS Abstract Service is used by over 10,000 users per month who make more than 10 queries per month. The total number of users per month is about 60,000 (32,000 from outside the USA). These users make over 1 million queries and retrieve 60 million references and abstracts. In November 2001 there were users from 91 countries who queried the ADS.

The ADS Article Service is used by about 25,000 users per month (over 15,000 from outside the USA). They retrieve over 1.2 million pages of scanned articles. In November 2001 there were users from 78 countries who retrieved scanned articles from the ADS.

A complete description of the ADS is given by Accomazzi et al.(2000) [1], Eichhorn et al.(2000) [2], Grant et al.(2000) [3], and Kurtz et al.(2000) [4].

This work is funded by NASA under Grant NCC5-189.

#### **References**

- [1] Accomazzi, A., Eichhorn, G., Kurtz, M. J., Grant, C. S., & Murray, S. S. 2000, *A&AS*, 143, 85
- [2] Eichhorn, G., Kurtz, M. J., Accomazzi, A., Grant, C. S., & Murray, S. S. 2000, *A&AS*, 143, 61
- [3] Grant, C. S., Accomazzi, A., Eichhorn, G., Kurtz, M. J., & Murray, S. S. 2000, *A&AS*, 143, 111
- [4] Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., Murray, S. S., & Watson, J. M. 2000, *A&AS*, 143, 41