Pergamon

Vistas in Astronomy Vol. 40, No. 3, pp. 393–394, 1996

0083-6656/96 $32.00 + 0.00

PII: S0083-6656(96)00021-9

INDEXING: WHAT YOU FIND IS WHAT YOU GET

MICHAEL J. KURTZ
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, U.S.A.

Abstract— A progress report is given on the status of the NASA Astrophysics Data System Abstract Service as an index to the literature of astronomy. Copyright ©1996 Elsevier Science Ltd.

Without an organizing principle information is just noise. An index is an organizing principle over a field of information; it is closely related to classification. The NASA Astrophysics Data System Abstract Service is primarily an index to the astronomical literature.

The ADS Abstract Service allows sophisticated queries based on author name, name of astronomical object (via CDS/SIMBAD and NED), words in the title, and words in the abstract. Sophisticated vector-space matching techniques have not yet been implemented for natural language queries, only simple inverse log of word frequency weighting. Until now the simple techniques have provided satisfactory performance.

In May 1996 the abstract service had about 8000 different users, who made 160,000 queries, and retrieved 3,300,000 references and 70,000 pages of whole text. 70,000 queries were primary literature queries to the astronomy database.

61% of all these queries asked for references by author, 31% by words in the abstract, 20% by words in the title, and 7% by name of astronomical object. 15% of author queries (= 10% of all queries) add words to the query, while 25% of object queries (= 10% of all queries) add words.

Most word queries use three or fewer words, the number of word queries halves with every additional word above three, until 30 or more are used. Queries with more than 30 words are formed by the find similar abstracts relevance feadback feature, and account for 1.5% of all queries.

The ADS provides direct access to the whole text of several journals, soon to include all the major journals of astronomy. In May 5000 articles were printed, 18,000 individual pages were displayed on the screen, and an unknown number of links were made to the electronic Astrophysical Journal Letters. This exceeds the total number of papers read in the largest astronomy libraries by a factor of several. The article service is bearly a year old, and is only now beginning to have a substantial fraction of the major journals on-line; expansion by an order of magnitude can be expected in the next year, or sooner.

The usage patters (as a function of age of the journal) appears quite similar from journal to journal, save for a scaling factor (e.g. ApJ is used 2.5 times AJ). For the first three years readership of an article declines with age with a half life of 1 year, for the next 10 years the half life is about 2.5 years, afterward the use flattens out, and the half life appears to be about 10 years.

An examination of usage patterns, and the plans of other groups leads to a prioritized "wish list" for work on ADS.

First is increased access and coverage (more abstracts, more tables of contents, more whole text). Second, improve author queries (better author synonym lists, use of middle initial). Third, improve few word queries (improved synonym list, use of phrases, use of Boolean logic: and, or, not). Fourth, better access to data by object name, especially data in archives; we hope CDS will take the lead in this. Fifth, obtain the reference and citation data on articles. We are actively pursuing this currently. Finally, improve methods for many word queries (vector space techniques, better relevance feedback). These are necessary for the article notification service to be fully successful.

As the journals come on-line the article scanning phase of the ADS project will end (save for historical projects) and the ADS will be able to concentrate on its primary function, providing a rich set of powerful bibliographic and bibliometric tools to allow one to find and access astronomical literature and data within the context of a tightly interconnected set of networked astronomical service providers.

Acknowledgements

Partially supported by NASA, NCCW-0024. Collaborators on the ADS project are: Guenther Eichhorn, Alberto Accomazzi, Carolyn Grant, Joyce Watson, Elizebeth Bohlen, and Steve Murray