

Bibliographical Search : Keep a Critical Eye on Your Result

Pascal Dubois

*Centre de Données Astronomique de Strasbourg, Observatoire
Astronomique, 11 rue de l'Université, F-67000 Strasbourg, France*
dubois@astro.u-strasbg.fr

Abstract. Bibliographical databases become indispensable tools in astronomical research, but the search results depend on indexation rules that are specific for each database. Some of these database differences will be analysed and illustrated.

1. Introduction

At one time, our best astronomical tool for bibliographical searches was the series entitled *Astronomy and Astrophysics Abstract* (AAA). However, lately, their main disadvantage was the lengthy delay between the publication of the scientific paper and its appearance in the bibliography. Also, computer searches can be quicker than books. These inconveniences were partially compensated for with the on-line bibliographical database ARIBIB (based on AAA).

At present, we have several different bibliographical databases that are related to astronomy. None of them are exhaustive and it is necessary to know their characteristics in order to evaluate the search results. Besides the question of exhaustiveness, there is the problem of "hidden data", or rather, that information not retrieved by the requestor.

2. Hidden Data

When making a bibliographical search, for example, in the library, you know what was left out of the search (e.g. the French or Russian publications, the papers before a certain date, the observatories publications, etc.). However, when making a search in an electronic database, it is more difficult to know what has been excluded (e.g. which journals have not been indexed, have parameters been utilized, etc.)

2.1. Examples

These examples cannot be used for statistical purposes, but serve as ideas for further discussion. The interrogations of the databases were done on 18th June 2002.

For NGC 1566, SIMBAD will return 317 references if you use the default parameters. But there is a limitation, since the bibliographic search starts in

1983. In total, SIMBAD has 333 references for this object of these, 16 references were 'hidden' references since they existed prior to 1983.

Why was this database limited to starting in 1983 ? The reason is because SIMBAD made a systematic scan of 90 journals for galactic and extragalactic objects; and, previous to this time, the scan of the literature covered fewer journals and concentrated on stars.

ADS, which receives the bibliography made by SIMBAD, skips this problem and, by default, provides 333 references indexed by SIMBAD. If you select NED as the search database, you obtain 307 references, but if you set SIMBAD and NED as the databases, ADS will return 390 references. There were 57 hidden references when compared to the default summary. In SIMBAD and NED, there are 250 common references.

NED contributes references earlier than 1983, as well as references to catalogues of galaxies. SIMBAD adds recent references and the catalogue object names from the *Dictionary of Nomenclature of Celestial Objects*.

NED, SIMBAD and ADS collect some references from astronomical meetings and theses. For example, if you add 'NGC 1566' to ADS's 'Title Words' and 'Text Words' windows, the database will return a total of 410 references, twenty of these will not have previously been indexed in the other two databases. These references are primarily from books or proceedings. On the other hand, unlike NED and SIMBAD, ADS word search does not take into account the other names of an object (which is done by searching NED and SIMBAD), and thus, there may be other "hidden references". SIMBAD and NED are preferred when used for periodicals because results published in colloquia can become obsolete.

3. What Could We Find in Other Databases ?

We found the following results for a search for NGC 1566:

- INSPEC returned 64 references of which 12 are not retrieved by ADS (52 are in common).
- ARIBIB returned 32 references of which 4 are neither in ADS nor in INSPEC.
- Pascal Scitech returned 13 references (from 1987 to April 2002) all in ADS.
- Current Contents returned 6 references (from 2000 to April 2002) all in ADS.

Most all additional references came from some astronomical colloquia or meetings and were not referenced in ADS. Two references came from two journals quoted in ADS, but for which no references could be found. Two other references from *A & A* and *PASP*, should have been indexed by SIMBAD, but were not due to a misprint in the paper.

4. Remarks

In astronomy, most objects have different names and the databases, SIMBAD and NED, take this into account; however, this is not the case for databases that use the name of an object such as in a string of characters.

The examples used in this bibliographical study related to a single object. If we had accessed a the bibliography on a specific subject, the examples would be different, but the question of hidden data would remain.

5. Conclusion

If you do not use the SIMBAD and NED databases, most of the references in a bibliographical search will be missing. ADS, including this indexation, is nearly complete. Presently, only a small percentage of bibliographies could be added using to the other databases. A more detailed study of the completeness of ADS can be found in FAQ section.

The problem of exhaustivity and of hidden data is not restricted to bibliographical databases and you should keep this in mind for each online request.

Finally, for objects with many references, the main problem in using the retrieved data may be its large volume and the need for selection of the most relevant papers (see the paper by Lesteven and Dubois in these proceedings)!