PART 3. SEARCHING FOR ASTRONOMICAL INFORMATION

https://doi.org/10.1017/S0252921100002943 Published online by Cambridge University Press

WORDS FOR SEARCHING - KEY WORDS

L. D. Schmadel
Astronomisches Rechen-Institut
Moenchhofstr. 12-14
D-6900 Heidelberg 1
Federal Republic of Germany

ABSTRACT. The importance of key words for information retrieval processes in astronomical bibliographies and data bases is stressed. Construction principles of astronomical vocabularies are discussed with special regard to the practice at 'Astronomy and Astrophysics Abstracts'.

I. INTRODUCTION

Astronomy and astrophysics are for some reasons unique cases in the framework of the exact sciences. Astronomers only can observe their research objects and cannot handle them in a laboratory. Most astronomical observations cannot be repeated so that all results must be preserved permanently. For an astronomer, the passage of time makes a repetition of an observation impossible. Therefore, the need for a somewhat condensed schedule of the entire astronomical literature throughout the world was already recognized in the early 18th century.

The need for a suitable indexing system arises from the large number of documents in the field of astronomy and astrophysics. Actually, we are faced with more than 20,000 papers, books, reports and other documents in these areas (see Fig. 1). The growth rate (Davoust and Schmadel, 1987) is very roughly proportional to the time and it seems from our counts that the material published doubles every 15 to 20 years. In the 19 years from 1969 to 1987 'Astronomy and Astrophysics Abstracts' abstracted and indexed more than 300,000 documents which were produced by more than 600,000 authors.

The question arises how one can find a special document or - more precisely - a selection of papers with similar characteristics out of this huge bulk of material. It is obviously very easy to pick up a certain paper with given bibliographic standard informations. The situation changes completely, however, if one is interested to find out all documents with one or more 'common' features. For this purpose one has to assign some characteristic indexing terms to a document which describes a certain facet of the scientific content. By means of these indexing terms it will be possible to do a single-minded retrieval work.

https://doi.org/10.1017/S0252921100002943 Published online by Cambridge University Press

The use of proper key words greatly facilitates this task.

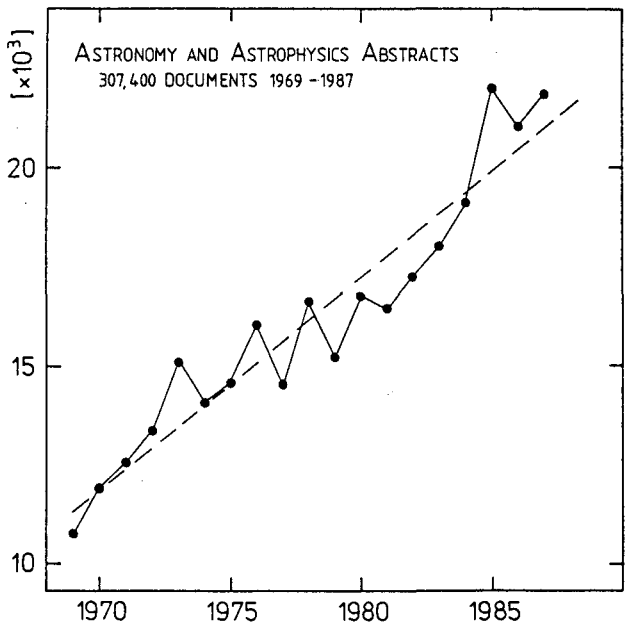

Fig. 1: Trend of the publishing activity of astronomers

II. CLASSICAL BIBLIOGRAPHIC ITEMS

A simple query for reference purposes could be accomplished by the exploitation of some standard bibliographic items (Schmadel, 1985): title of the document, name(s) of the author(s), or a precise source declaration. It is easily possible to find any document on the bases of formatted fields which is defined by means of these parameters. It is, however, impossible to compile a 'package' of documents with one or more scientific aspects in this way. For this purpose one has to apply a higher query level (Adorf and Busch, 1987) by using certain descriptive characteristics of a paper. We have three possibilities to realize this.

The construction of subdivisions to the entire material according to scientific principles - classificatory key words - establishes only a very coarse break down. 'Astronomy and Astrophysics Abstracts' provide more than 100 categories for classification aims. The attachment of a publication to only one category, however, is a difficult undertaking and it is - by no means - a well-defined procedure with unique results. A retrieval which is only based on these classificatory key words undoubtedly will yield too many references. The extensive use of cross

https://doi.org/10.1017/S0252921100002943 Published online by Cambridge University Press

references to further categories concerning some different aspects of a document is necessary to limit the query result.

If a paper deals with a special astronomical 'object' and if we mention the object designation in our list of indexing terms, then we are easily able to retrieve the paper. This procedure, however, will only yield the desired result if one uses a 'standardized' form of object designations. We all know about the numerous difficulties in the designation area with the danger of acronym confusions and even with the detrimental effects of purely spelling differences. A retrieval by means of object designations is, of course, impossible in all cases of purely theoretical papers. Under these circumstances there only exists the possibility of looking for certain phrases or descriptors which characteristically should appear in a paper dealing with a special field of knowledge.

We focus here on the construction of key words or free descriptors as a central resource for retrieval purposes and we comment on some problems involved.

III. PRINCIPLES AND PROBLEMS

The assignment of key words to a given document will allow to consider a lot of diverse aspects. This procedure, however, strongly depends on the availability of a sufficient number of descriptors. Furthermore, a vocabulary of astronomical terms has to be widely accepted by the astronomical community. Based on our experience in the classification and indexing area at 'Astronomy and Astrophysics Abstracts' we introduced a list of free descriptors some ten years ago. It was our aim to compile a descriptor framework which should also serve as a first approximation of an astronomical thesaurus, i.e. a hierarchical list of relevant terms. The number of entries of such a key word list is an important supposition for the dissemination and world-wide acceptance. We found from a detailed analysis of the subject index entries in our AAA volumes that only about 1,500 descriptors will be sufficient to characterize all important aspects of an astronomical document. It has also been shown that the use of single key words is in many cases not adequate to describe a given fact precisely.

The descriptors can be roughly divided into three main categories:

(1) descriptors for classes of objects (c.f. the terms 'Minor Planets', 'Delta Cephei Stars', 'Spiral Galaxies',
(2) key words, describing attributes or properties of celestial objects such as 'Masses', 'Diameters', 'Rotation', or 'Chemical Composition',
(3) very general descriptors like 'Models' or 'Formation' and very specialized ones (c.f. 'Kirkwood Gaps' or 'FK5 System'). Beyond that we find in this third class descriptors for methods (c.f. 'VLB Interferometry') or astronomical instruments (c.f. 'Bolometers' or 'Ritchey-Chretien-Telescopes'). This is a rather heterogeneous category.

https://doi.org/10.1017/S0252921100002943 Published online by Cambridge University Press

Our counts have revealed that the great majority of key words is composed of members from the first two categories. This means that the connection of an object class with a special property yields a fairly good description of an aspect of the content of a paper.

It is important to notice that we preferably use pairs of key words which, mathematically spoken, means the application of the Boolean AND connection. Let us study an example: The key word 'Orbits', for instance, is a fairly general descriptor which can be attributed to a large variety of astronomical objects ranging from the planets over binary stars to, say, galaxies in clusters. A retrieval using only the term 'Orbits' therefore will yield a large number of papers with completely different scientific aims. Only the addition of a second explaining key word will limit the retrieval target area. If we combine, for instance, the term 'Orbits' with the term 'Minor Planets' then we have a good description of the fact that the document in question deals with problems of dynamical astronomy in the solar system. This combination procedure yields the result that we are able to describe the most important facets of a paper with only a few such key word pairs. At AAA we usually limit ourselves to the assignment of not more than five pairs. We demonstrated by test runs that this method is sufficient for a single-minded literature search.

Actually, we use at AAA a vocabulary of 2,343 single-term items. Bearing the unwished dispersion effects of very general as well as of very specific terms in mind, we have flagged these components in our vocabulary. This means that these signed descriptors do not appear as entries in the AAA subject indexes.

Some difficulties in the construction of key word lists should be summarized briefly. One problem is given by the use of plurals as well as singulars for the same item. We should consider both forms as possible. Further differences between published descriptor lists are given by slightly different arrangements of descriptor components. There obviously is no great difference between, for instance, the terms 'Space Distribution' and 'Distribution in Space' and we should standardize the shorter version throughout. For the implementation of vocabularies into a retrieval software one should also limit the number of characters to a certain maximum. One should furthermore limit the extent of usuable characters to capital and small letters and to apostrophes and hyphens - subscripts, superscripts and numbers should be avoided. It is clear, too, that spelling differences (c.f. 'CCD' versus 'C.C.D.') should be avoided as far as possible.

One major point of importance is given by the use of synonymous items. The terms 'Minor Planets' and 'Asteroids' exactly have the same meaning as does, for instance, the phrases reference frames and reference systems. We propose to accept only one version out of many existing synonymous descriptors.

Any vocabulary is a living structure which has to be improved with the development of our science. Ten years ago, the term 'Cosmic Strings' was unknown and it will surely happen that some 'classic' phrases will pass away in the near future.

IV. CONCLUSION

Efforts are being made to construct a commonly agreed compilation of descriptors and to install an astronomical thesaurus thereof. This will greatly facilitate not only the work of librarians and of the people from the abstracting services and data bases, but it will be of value to all scientists who are forced to pursue literature retrieval for their own research work. IAU Commission 5 therefore urges all authors, editors, and publishers of astronomical documents to frequently use standardized key words. This additional information is an essential means for proper indexing and is thereby an important supposition for the further and wider spread of scientific ideas and results.

REFERENCES

Adorf, H.-M., Busch, E.K.: 1988, 'Astronomy from Large Databases: Scientific Objectives and Methodological Approaches', A. Heck, F. Murtagh (Eds.), p. 143.

Davoust, E., Schmadel, L.D.: 1987, Publ. Astron. Soc. Pac., 99, 700.

Schmadel, L.D.: 1985, Bull. Inf. Cent. Donneés Stellaires, No. 28, 95.

https://doi.org/10.1017/S0252921100002943 Published online by Cambridge University Press