T-REX: A COOPERATIVE THESAURUS BUILDING PROJECT FOR THE INTERNATIONAL ASTRONOMICAL UNION

Robyn M. Shobbrook
Anglo-Australian Observatory
P.O. Box 296, Epping NSW 2121, Australia

I. INTRODUCTION

Astronomers and librarians have been experiencing difficulties in keeping up with the amount of published literature. The astronomer tries to keep abreast in his particular field and the librarian in the management, control and retrieval of scientific information. The 1980's have seen a revolution in the methods for information storage and retrieval and in particular the advent of the online database. The speed of processing information for storage has been embraced by all, however little thought has been given to *how* we shall achieve effective high precision recall of documents.

Many librarians firmly believe the best road to success in information retrieval from automated systems is provided by vocabulary control. Contrary to belief, free text or natural language searching alone does not lead to high precision recall. Consistency and integrity of the online catalogue can only be achieved with the addition of a controlled vocabulary. With today's technology it is possible to maintain the best of both worlds. The controlled vocabulary is used to index the major concepts of a given document over and above the natural language used within the document.

II. BACKGROUND TO THE PROJECT

The international project for the development of an astronomy thesaurus was undertaken at the request of the International Astronomical Union (IAU). The project was affectionately dubbed T-REX by the collaborators. It was begun in 1986 with an initial feasibility investigation. This attracted several volunteers to assist in the compilation of the first draft of preferred terms. The volunteers are librarians from Australia, United Kingdom, United States, India and Canada: each compiled terms for a nominated section of the alphabet. I have attempted to coordinate all efforts in producing the first draft list for submission to Commission 5 at this General Assembly. It is important for scientists and astronomy librarians to discuss the thesaurus, the need for it, and its implications; particularly in the efficient retrieval

of information in our libraries and for its use in online databases locally, nationally and internationally.

The American National Standard for "Guidelines for Thesaurus Structure, Construction and Use" (1.) defines a thesaurus as:

> "A thesaurus is a compilation of words and phrases showing synonymous, hierarchical and other relationships and dependencies, the function of which is to provide a standardized vocabulary for information storage and retrieval systems."

III. THE RATIONALE BEHIND THE THESAURUS PROJECT

I think we must agree there is a need to standardise the terminology in the field of astronomy and astrophysics. There is increasing specialisation in astronomy since no one can be a specialist in all areas. Librarians as the interface between the scientists and the stored information are expected to organise and retrieve promptly and efficiently documents and/or bibliographic references; we maintain that it is possible, but not likely, that we can do this effectively without some guide to the terminology.

Results of a questionnaire (Fig 1.) relating to the use of subject headings in astronomy libraries revealed that librarians are compiling their own *ad hoc* systems with little knowledge of the subject field or of the subject approach to information; indeed there are many libraries in this field without subject access to their collections. Therefore a thesaurus or guide to the terminology showing preferred terms together with the relationship of broader, narrower and related terms (BT, NT, RT's) will mean the achievement of some kind of standardisation which will be of enormous benefit in preparing search strategies for automated and manual library systems and in the allocation of keywords for scientific papers destined for publication.

It is important to realise that at this time there is no standardised list of terms in the field of astronomy and astrophysics although there are other specialist thesauri which attempt to cover the fringes of the subject e.g. NASA and INSPEC . Several of our major reference journals have their own keyword lists and I will not attempt to discuss their shortcomings for our purposes because they were designed to facilitate the compilation of the subject index to the journal itself. However it should be realised that the format of these subject keyterms in our major reference journals will NOT facilitate the retrieval of documents from any computer database system. Most commercial databases and in-house database software prefer the usage of terms with a unit concept and linked by boolean operators i.e.and, or, not. This is the format adopted for the IAU thesaurus.

IV. THE BENEFITS OF AN ASTRONOMY THESAURUS FOR THE SCIENTISTS AND FOR THE LIBRARIANS

i. The standardisation of terminology will lead to the improvement of astronomy information services worldwide particularly with the increased use of online databases and networking.

ii. Editors and writers of scientific papers will benefit with the provision of an internationally recognised list of key terms endorsed by the IAU. The 1st Dictionary of Nomenclature was the first step towards standardisation.

iii. With the application of computers in all areas of library management a standardised list of terms will help eradicate *ad hoc* systems, develop understanding and increased expertise in subject analysis, improve search strategies in online retrieval of information and thus save in terms of time and particularly online costs. Overall there will be an improvement in the identification, organisation, and retrieval of all types of documents related to astronomical research.

V. THE COSTS INVOLVED

Although the IAU requested the thesaurus, it is not its policy to fund such projects, but when it is accepted it will endorse and publish the work. Perhaps Commission 5 could review this policy because funding is urgently required to proceed with this project in terms of manpower and software to reach the next level in the development of the thesaurus.

Initially the costs of compiling a thesaurus are high in terms of time and labour. The unique method of international collaboration for this project has meant the costs of the T-REX project have been kept to a minimum for the preliminary draft. Staff time and labour have been met by each institution and volunteer. Future work will need to be accomplished as part of the normal day-to-day tasks of the librarians involved because thesauri are never really finished and regular maintenance or updating will need to be done.

VI. THE PROBLEMS AND HOW ASTRONOMERS CAN HELP

- Comprehension
    I hesitate to reveal that in my experience scientists generally have little knowledge or concept of what a thesaurus is. Unfortunately I strongly suspected that not many librarians do either so this project has been an education process for all involved, myself included. My interest in this project came about because I was in the process of converting the AAO card catalogue to a computer database since we had no previous subject access to our records.

Until confronted with retrieval problems many will not see the need or the relevance of a thesaurus; lack of knowledge of on-line databases in this field is one major problem. We are all in the process of re-education in this era of online catalogues. Unless we come to grips with improving input and access it will be a case of 'garbage in - garbage out'. Cochrane (2) an expert in subject indexing maintains "If adequate control is not built into the subject control mechanism before searching begins, the search product will be disappointing...the usefulness of what comes out of a file depends entirely on what has been put in and how it has been put in".

- Consistency
  Each,of the collaborators came to the project with their own ideas and perceptions of what was required, even though rules and a standard was adopted. Scientists need to assist us to iron out problems with terminolgy, for example synonyms in particular areas such as variable stars.

- Classification numbers
  Some difficulty has been encounted with the allocation of American Institute of Physics (AIP) classification numbers. This was mainly due to a term's abstract nature and/or lack of context. Help is needed from the scientist to allocate appropriate classification numbers.

VII. THE FUTURE

I hope T-REX is not destined to become extinct! A great deal of work has been done and still has to be done. I am sure Commission 5 did not have any idea what was involved when they requested the work to be done. A thesaurus IS necessary but I think some financial assistance should be provided for the continuation of this vital project. Most thesauri I have read about have taken five years to produce. With some financial assistance and some help from the scientists the next more refined version of T-REX might be ready by the time the next General Assembly comes around.

Librarians will be involved in the testing of the thesaurus over the next year. In this way we will identify inaccuracies and omissions and any anomalies associated with preferred term usage. It is inevitable that there will be national and local differences, but there must be an attempt to resolve these differences with the objective of producing a reliable, authoritative, internationally acceptable list of terms.

Meanwhile the terms can be sorted using the AIP classification numbers into specialisations. When sorted the lists can be sent to various subject specialists to assess the accuracy and completeness. I would like to emphasise that I am concerned that

not enough work has been done by scientists in the compilation of the list. Only *they* know the terminology well.

VIII. CONCLUSION

Remember the old cliche — 'Oh what a tangled web we weave'. It is especially relevant when we practice to retrieve.

IX. REFERENCES

1. American National Standards Institute, Inc., "Guidelines for thesaurus structure, construction and use". ANSI Z39.19-1980.

2. Cochrane, P.A., "Redesign of catalogs and indexes for improved online subject access: selected papers". Phoenix, Arizona, The Oryx Press, 1985.
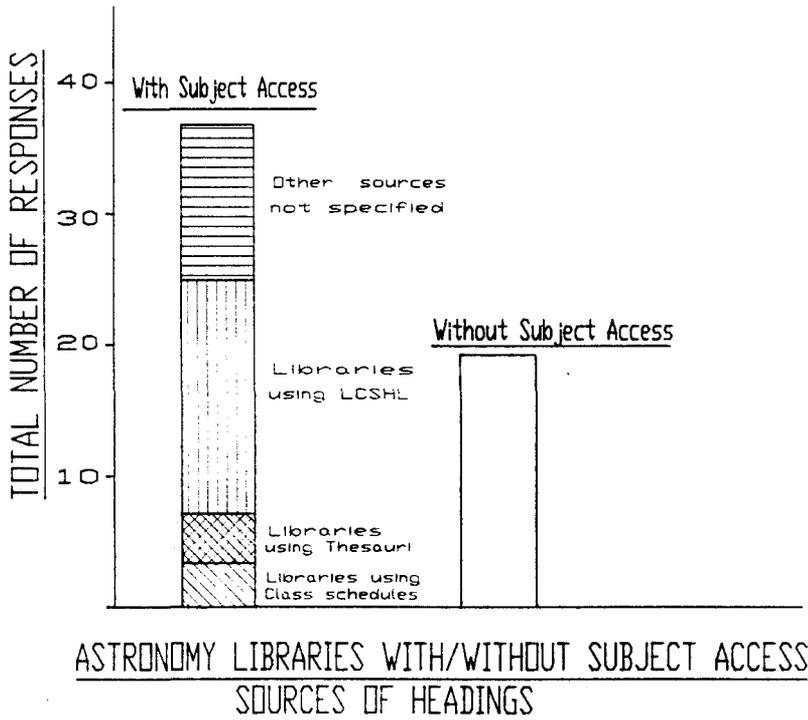

ASTRONOMY LIBRARIES WITH/WITHOUT SUBJECT ACCESS
SOURCES OF HEADINGS

Fig. 1.