

T-REX is not Lord but Mate in Information Land

Tatyana Dorokhova and Nikolay Dorokhov

Astronomical Observatory of Odessa National University,
Park Shevchenko 65014 Odessa, Ukraine
tnd@te.net.ua

Abstract. T-Rex is a Thesaurus Building Project realized by R.M. and R.R. Shobbrook, in cooperation with librarians from different countries and was based on instructions from the International Astronomical Union, Commission 5. The Thesaurus is useful for librarians and scientists and especially useful for information retrieval and for larger astronomical database systems such as ESO's library system and NASA's ADS (Astrophysical Data System). We are presently translating the Astronomy Thesaurus into Russian and Ukrainian as exemplified by the thesaurus's Multi-Lingual Supplement. This work for ranking and classifying knowledge is particularly valuable when one considers the avalanche-like growth of World Wide Web (WWW).

1. Introduction

Today, Peter Boyce's slogan (Boyce 2001), "Own nothing, access everything" is very relevant and almost certainly will be more important in the future. Every scientist, writer, librarian and reader feels the avalanche of information in all subjects and topics. Based on this account, the slogan might be transformed to eventually read, "Efficient access is everything."

First and foremost, the information has to be analyzed and classified. A great number of systems exist that classify and regulate information, but one of the first and most well-arranged is the UDC (Universal Decimal Classification). One of its great advantages is its ability to use many languages independently. (Wilkins, 1998; Dorokhova, 1998).

However, with the development of informational and Internet technologies, the semantic approach becomes prevalent, and the access to the information is provided by vocabulary control. In turn, a developed structure of key words and thesauri is required. Many difficulties arise when there are different interpretations of the same concept, "a synonym expansion," or a penetration of foreign words in the language, etc. (Accomazzi et al. 1999) in their description of NASA's ADS architecture noticed: "As we have added more and more bibliographic references from historical and foreign sources, the non-English words in our database have been slowly but steadily increasing. As a result, we intend to merge the proper foreign words with each group of English synonyms in a systematic fashion." It should be remarked that if the ADS Abstract Service had gotten its beginnings at the conference "Astronomy from Large Data-bases",

held in Garching in 1987 (Kurtz et al. 1999), the initial investigations for T-REX project were made in 1986 (Shobbrook 1989). This demand and supply appeared almost simultaneously!

2. T-Rex is a Thesaurus Building Project

T-Rex is a Thesaurus Building Project realized by Robin and Robert Shobbrook from Anglo-Australian Observatory and Sydney University, respectively, who worked in cooperation with librarians from different countries, and in accordance with the instructions of the International Astronomical Union (Shobbrook 1989). Other librarians who collaborated on the project included:

Liz Bryson (CFHT, Hawaii, USA)
Ellen Bouton (NRAO, USA)
Brenda Corbin (US Naval Observatory, USA)
Marlene Cummins (University of Toronto, Canada)
Anne Fraser (Royal Observatory, Scotland)
Betty Holmstrom (Steward Observatory, USA)
Helen Knudsen (California Institute of Technology, USA)
Vagiswari Alladi (Indian Institute of Astrophysics, India)
Eugenia Gomez (ESO, Chile)
Lucenya Kedziora (Telescope National Facility, Australia)

and many other astronomers and librarians from different countries. As a final result, the hard copies of *The Astronomy Thesaurus* (Shobbrook et al. 1993) and the *Multi-Lingual Supplement* (Shobbrook et al. 1995) were published by Australia Image Output. At present, the very convenient electronic version is available at :

<http://msowww.anu.edu.au/library/thesaurus>

The Thesaurus was the voluntary work of great number of people: “The construction took many, many person/months if not years” (Cummins 2002). This work did not receive due attention from the astronomical community, nor financial support (Shobbrook 1989), so it was discontinued. On the title page of T-REX is written: “The Astronomy Thesaurus has not been updated since 1993/5 and there are no plans to do so in the foreseeable future.” Both Robyn and Bob Shobbrook have now retired from their institutions and would participate only as consultants in any future work.

Corbin (1999) and Cummins (1999), both early and active participants in the project, noted the importance and necessity of the Astronomy Thesaurus in the daily work of librarians and astronomers and regretted that this project has not been updated.

3. Other Thesauri and their Usefulness

There are many other thesauri but usually these exist for some utilitarian purpose. For example, the thesauri of keywords for *Astronomy and Astrophysics*, *The Astrophysical Journal* and *Monthly Notices of the Royal Astronomical Society* provide only brief lists of terms, which appear as topics in the journals.

There are thesauri comprised of a broader scope of terms, (e.g., the NASA or INSPEC thesauri). Since the NASA Thesaurus (at www.sti.nasa.gov) is very comprehensive, it is continuously updated. However, it includes many terms that are superfluous for astronomy and astrophysics. In addition, the international CD ROM version of the NASA Thesaurus costs 374 (US) dollars and its data are not online for public access.

The IAU Thesaurus contains more than 3,300 primary terms (Shobbrook & Shobbrook 1992), but it is not a simple list of terms. The Thesaurus also supplies a hierarchy of words and phrases and includes their relationship "which is to provide a standardised vocabulary for information storage and retrieval systems." (Shobbrook & Shobbrook 1992).

The IAU Thesaurus is not necessarily required in a small library; however, for the network of libraries such as ESO libraries, it provides the keyword searches for literature (Grothkopf, Treunann & Gomez 2001). The IAU Thesaurus and the Multi-Lingual Supplement are also valuable as a supply of universal keywords for astronomical journals from different countries.

The keyword data for building and development of a literature retrieval service for ADS were received from the T-REX project (Kurtz et al. 1999). The authors (Accomazzi et al. 1999) who constructed ADS as the most suitable and favorable retrieval system mention: "It is the extensive work that has gone into compiling such a list that makes searches in the ADS so powerful. To give an idea of the magnitude of the task, it should suffice to say that currently the synonyms database consists of over 55 000 words grouped into 9,266 sets. Over the years, the clustering of terms in synonym groups has incorporated data from different sources, including the "Multi-Lingual Supplement to the Astronomy Thesaurus." The Multi-Lingual Supplement also appears to have been necessary for the purpose of solving the penetration of foreign words into ADS's vocabulary.

Since the Multi-Lingual Supplement to the Astronomy Thesaurus has not included the terms based on the Cyrillic alphabet, we decided to continue the translation of the Astronomy Thesaurus into Russian and Ukrainian. With the Multi-Lingual Supplement as our guide, we hope the project will make available new literature for this Thesaurus of Knowledge.

Acknowledgments. We thank Robyn and Bob Shobbrook and Marlene Cummins for their interest in the work and for helpful discussions.

References

- Accomazzi, A., Eichhorn, G., Kurtz, M.J., et al., 1999 *Astron. Astrophys. Suppl. Ser.* 143, 85
- Boyce, P.B., 2001, in *Proc. Fiesole-Oxford Workshop*, Oxford, UK
(<http://www.aas.org/~pboyce/epubs/Oxford-2000/index.html>)
- Corbin, B.G., 1998, in *ASP Conf. Series*, Vol. 153, Library and Information Services in Astronomy III, eds. U. Grothkopf, H. Andernach, S. Stevens-Rayburn & M. Gomez, (San Francisco: ASP), 27
- Cummins M., 1998, in *ASP Conf. Series*, Vol. 153, Library and Information Services in Astronomy III, eds. U. Grothkopf, H. Andernach, S. Stevens-Rayburn & M. Gomez, (San Francisco: ASP), 101
- Cummins M., 2002, private communication
- Dorokhova, T.N., 1998, in *ASP Conf. Series*, Vol. 153, Library and Information Services in Astronomy III, eds. U. Grothkopf, H. Andernach, S. Stevens-Rayburn & M. Gomez, (San Francisco: ASP), 269
- Grothkopf, U., Treumann, A., Gomez, M.E., 2001, *The Messenger*, 105, 45
- Kurtz, M.J., Eichhorn, G., Accomazzi, A., et al., 1999, *Astron. Astrophys. Suppl. Ser.* 143, 41
- Shobbrook, R.M., 1989, in *LISA I proc.*, ed. G.A. Wilkins & S. Stevens-Rayburn, US Naval Observatory Pub., 72
- Shobbrook R.M., Shobbrook R.R., 1992, *PASAu*, 10, 134
- Shobbrook, R.M., Shobbrook R.R. et al., 1993, *The Astronomy Thesaurus*, eds. R.M. Shobbrook & R.R. Shobbrook, Epping: Anglo-Australian Observatory
- Shobbrook R.M., et al., 1995, *Multi-Lingual Supplement to The Astronomy Thesaurus*, Epping: Anglo-Australian Observ.
- Wilkins G.A., 1998, in *ASP Conf. Series*, Vol. 153, Library and Information Services in Astronomy III, eds. U. Grothkopf, H. Andernach, S. Stevens-Rayburn & M. Gomez, (San Francisco: ASP), 317