Subject Fields in Termbases - Their Design, Use and Representation

Kara Warburton¹

¹University of Illinois at Urbana-Champaign, 707 S. Mathews Ave., Urbana, Illinois, 61801, USA

Abstract

Subject fields play an essential role in terminological resources by allowing for the creation of semantically-based subdivisions in addition to acting as a conceptual boundary for the principle of univocity. However, due to the lack of guidelines and standards, their application in termbases risks being ad-hoc, which reduces their effectiveness in achieving these goals. ISO TC/37 has published a technical specification (TS) aimed to increase the rigour of subject-field use and the interoperability of the data. This paper describes some issues and challenges relating to subject-fields in termbases and how the TS may resolve them.

Keywords

Terminology, TBX, subject fields, domains.

1. Introduction

Classification is a widely-used ordering mechanism, indispensable for instance in information and library science [7, 5]. Philosophers such as Aristotle, taxonomists such as Carl Linnaeus, and documentalists such as Melvil Dewey established principles for the classification of knowledge into categories that are widely used today. It is no surprise then that terminological entries are frequently organized into categories. These categories can be based on semantic properties, or criteria of a more administrative nature such as institutional departments, clients, and so forth. In the former case, the most common type of categorization is referred to as domains or subject fields.

2. Subject fields in Terminology

The notion of subject fields is critical to terminology theory and practice. According to convention, terms designate concepts that belong to a language for special purposes (LSP) (as opposed to language for general purposes or LGP) [8], and an LSP is the language used by specialists in a subject field [1]. For many scholars, adherence to a subject field is a requirement for a linguistic unit to be deemed a term [6, 2]. Indeed, specifying the subject field that a term belongs to is often considered mandatory for terminological description [6, 1, 7, 5].

Univocity, a key principle in classical terminology theory, may also depend on subject fields. According to this principle, a term should have only one meaning. But we maintain that univocity is only achievable if it is applied within the scope of a subject field. This is because "identical" lexical units occur in different subject fields with different meanings (homonyms, homographs) (for example, "port" the strong wine and "port" the computer connection). Consequently, univocity has been defined with domain-specificity as its scope [2].

Scholars have also noted that subject fields should be organized in a hierarchical structure, to include sub-fields and even finer divisions [1, 2, 5]. Figure 1 provides an example of a three level system showing the top level Education, followed by child levels, three of which are further divided into subordinate values.

¹st International Conference on "Multilingual digital terminology today. Design, representation formats and management systems", June 16 – 17, Padova, Italy



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)





3. Challenges

3.1.1. Lack of guidelines and of a universal subject-field classification system

Guidelines, standards, and representation models for subject fields are lacking in the literature. Given the importance of subject fields which we have demonstrated, this is surprising if not troubling. Consequently, the use of subject fields in today's termbases varies considerably. Some termbases use none at all, others feature a flat list of values². Each termbase that features subject fields employs a unique set, different even from that of other termbases that cover the same or similar spheres of knowledge. The lack of a universal subject-field classification system represents a major obstacle to the interoperability of terminological databases.

3.1.2. Difficulties when assigning subject field values to concepts

Deciding which subject field a concept "belongs" to is another challenge. The choice is not always obvious, and terminologists often rely purely on intuition. Under these conditions, subject-field assignments will not be reliable, which raises questions as to the effectiveness of subject fields as a classificatory mechanism.

There is also the question of whether a concept can be assigned to more than one subject field. Here, terminologists disagree; some say yes, others no. However, if a subject-field value sets a boundary enabling the term to be univocal, then one would assume that it is confined to this subject field. This leads to the possibility that, if a terminologist feels inclined to select two subject fields, perhaps it is their "parent" that should be assigned instead. These are philosophical questions worthy of further debate.

3.1.3. Lack of models for representing subject fields

ISO Technical Committee 37, Sub-committee 3, has published a standard for representing terminological resources in an XML markup format, *ISO 30042: TermBase eXchange (TBX)*. TBX also constitutes a model framework for designing a termbase. However, subject fields and their representation is not addressed in any substantive manner. They are loosely modelled in plain text fields (with therefore no control over permissible values), and there is no facility for establishing a taxonomic structure. The standard merely stipulates that subject fields are to be represented in a <descrip> element at the concept level, for example:

² In the full version of this paper to be submitted for publication, some examples will be provided.

<descrip type="subjectField">Nuclear power</descrip>

4. The response of ISO TC 37

To address the TBX limitations, in 2021 the committee published a Technical Specification (TS) that provides guidelines for subject fields as well as for concept relations (another important feature of termbases for which guidelines are lacking): *ISO/TS 24634 - TBX-compliant representation of concept relations and subject fields*. In the following paragraphs, we summarize the contents of this TS.

4.1. Constraints

The TS specifies the following constraints relating to subject fields. The aim is to increase interoperability.

1. The content of the subject-field data category shall be a picklist (closed list of values). These values form the organization's subject field classification system.

2. Whenever possible, an existing public subject field classification system should be adopted, such as EuroVoc or Lenoch.

3. The name and source of the subject-field classification must be declared in the TBX header.

4. The full subject-field classification system should be described, either in the backmatter of the TBX document instance, or through an XML namespace. Within this description, the scope, or meaning, of subject-field values, should also be defined. This aims to facilitate a more reliable assignment of subject-field values to concept entries.

4.2. XML representation

An XML model for representing subject-field classification systems is provided in the TS. The model includes some markup adopted from the RDF-based SKOS.

5. Conclusion

The ISO TS should help to increase the interoperability of termbases. However, it will only have an effect if its provisions are adopted by termbase administrators. The uptake of ISO TC37 standards, however, has been slow in the past. Furthermore, full interoperability will not be achieved without a universal classification of subject fields. Whether that is a realistic goal remains open to debate.

6. References

- [1] M. Teresa Cabre, Terminology Theory, Methods, and Applications, John Benjamins Publishing Co., Amsterdam, 1999.
- [2] R. Dubuc, Manuel Pratique de Terminologie, Linguatech, Montreal, 1992.
- [3] International Organization for Standardization, ISO 30042 TermBase eXchange (TBX), Geneva, 2019.
- [4] International Organization for Standardization, ISO/TS 24634 TBX-compliant representation of concept relations and subject fields, Geneva, 2021.
- [5] A. Rey, Essays on Terminology, John Benjamins Publishing Co., Amsterdam, 1995.
- [6] G. Rondeau, Introduction à la terminologie, Centre Educatif et Culturel Inc., Montreal, 1981.
- [7] J. Sager, A Practical Course in Terminology Processing, John Benjamins Publishing Co., Amsterdam, 1990.
- [8] W. Teubert, Language as an economic factor: the importance of terminology, in G. Barnbrook, P. Danielsson, M.Mahlberg (Eds.), Continuum, London, 2005, pp. 96-106.