

The Status of Implementation of TBX

Alan Melby and Kara Warburton

Abstract

TBX (TermBase eXchange) is an OSCAR (<http://www.lisa.org/sigs/oscar/>) industry standard that has been submitted to ISO TC 37 as a candidate for becoming an ISO standard as well. One reason for TBX to be both an industry standard and ISO standard is to encourage its implementation by language technology developers and terminology database developers, both in the private and public sectors. The value of a terminology exchange format is directly related to how widely it is implemented as an import-export format for software that stores terminological information.

TBX is an XML application that is connected to various existing ISO standards. In particular, it qualifies as a Terminology Markup Language under TMF (ISO 16642). This means that it conforms to the abstract meta-model for terminological data collections that can be viewed as the skeleton of TMF. The meat on that skeleton is the set of data categories that are allowed. The 2002 version of TBX included a master set of data categories selected from ISO 12620:1999. It is anticipated that the revised version of TBX will refer to the Data Category Registry (DCR) associated with the upcoming revision of ISO 12620, which will include the data categories from ISO 12620:1999 among other sets of data categories.

It is well-known that not all terminological databases use the same data categories. This considerably complicates the requirements for a terminology exchange format. One possible approach is to specify a standard set of data categories and require every software application that conforms to the standard to implement only those data categories. Another approach is to define a core exchange format with a recommended set of data categories and a mechanism for specifying subsets of the suggested set, and when clearly documented and justified, additional data categories not in the recommended set. Each customized set of data categories then corresponds to a variant of the standard. This is the approach taken by TBX, and the mechanism for specifying subsets and supersets is called an eXtensible Constraint Specification (XCS). The presentation will describe the recently developed subset of TBX called TBX-Lite.

The focus of this presentation is to explore the status of the implementation of TBX in the real-world. The LISA (www.lisa.org) Terminology Special Interest Group is conducting (as of June 2007) a survey of developers of language technology products and other developers of terminology data, especially translation technology products, to find out who has or has not implemented TBX

(<http://www.lisa.org/standards/tbx/>) and why.

The survey is available at:

http://www.lisa.org/sigs/terminology/tbx_survey.html

Some objections to implementing terminology exchange are independent of TBX or any other terminology interchange format. For example, there is the objection that concept-based terminology is incompatible with natural language processing lexicons. There is also the objection that exchange cannot be perfect because various systems use different sets of data categories. The survey might reveal that these are the primary obstacles to implementation of TBX. On the other hand, it might reveal that entirely different issues are relevant. After a summary of the results of the survey, the audience will be invited to participate in a discussion of how to overcome the various objections raised by survey respondents.