Application of UML for Terminological Concept Modeling in Accordance with ISO 704 and 1087

Hanne Erdman Thomsen

Abstract

In the information technology field it is becoming common to involve terminologists in the phase of concept modeling. Usually a first issue to be faced in such work is the recognition of two separate phases: terminological concept modeling and conceptual data modeling, and of the differences between them. Next terminological principles (ISO 704 and 1087-1) have to be introduced, as they are not always known to the IT developers, and the notation for terminological concept modeling has to be decided upon. In this connection UML (Unified Modeling Language) may be chosen, although it has certain shortcomings for terminology work.

In Denmark, terminologists from the DANTERM Centre at CBS (Copenhagen Business School) have been involved in concept modeling at national level in two institutions: The National Board of Health and The National IT and Telecom Agency. In both cases, terminological principles and working methods were unknown until terminologists were involved, and in both cases UML was proposed as the notation for modeling. In the case of the IT and Telecom Agency, the notation for terminological concept modeling was changed to that introduced in ISO 1087 and 704.

In the presentation, I will focus on the proposal of some enhancements to UML that would make it directly applicable for terminology work. The main issue is how to represent subdivision criteria in a way that will capture their close relationship with characteristics, and hence also how to represent characteristics. This part of the presentation will be based primarily on the work in SC1/WG5 (Concept modeling in terminology work) where we are currently working on a Technical Report - TR 24156: Guidelines for applying concept modeling in terminology work. This work is not well known outside the working group, but merits a presentation to a wider audience of the TC37 community. I will also draw on experience from the CBS project CAOS (Computer-Aided Ontology Structuring). This is a research and development project concerned with the development of a UML-based tool for terminological concept modeling.

Another important point is the introduction of terminological principles as described in ISO 704 and 1087-1, which are generally not known in the information technology community at large. For example, in the IT and Telecom Agency, we have been presented with a ‘definition’ consisting of a UML diagram, and in the case of written definitions, it is difficult to make the distinction between the definition proper and supplementary, encyclopedic knowledge.