

ISO/TC37述語および言語資源経営標準とその応用

崔 紀鮮

言語資源の一つである専門用語は特定知識分野の概念の表現のための言葉の集合である。これから、専門用語は非言語的概念表現、例えば、グラフ記号や数式なども含めている。専門用語は専門用語編集学的方法によって編集されて、辞書やデータベースとして現れる。言語資源（単一、多言語）、例えば音声データ、テキストコーパス、語彙コーパスとその処理方法にも標準化が必要である。ISO/TC37は、この為の四つの副委員会が標準化作業とその関連プロジェクトを進めている。現在、TC37の新しい分野としてeCommerceの為の多言語品物分類体系の原則の検討、SC1では専門用語作業の原則と方法、SC2はSC1の応用として専門用語編集及び辞書編集、SC3はそのコンピュータ応用、SC4は言語資源経営を担当している。以下、その詳しいリストを挙げる。

ISO/TC37 Terminology and Language Resource Management and its Application

Key-Sun CHOI¹

Terminology as a type of language resource is a set of terms that represent concepts of a specific knowledge domain. Increasingly, terminologies also include non-linguistic concept representations such as graphical symbols, formulae, etc. Terminologies are compiled according to terminographical methods and presented as dictionaries, databases, etc. Standardization is also needed for other language resources (mono- and multilingual), e.g. speech data, written (full) text corpora, lexical (general language) corpora and their processing methods. ISO/TC37 comprises of four subcommittees which have the following standards and undergoing projects:

TC37 possible working item

[PWI] "Basic principles of multilingual product classification for electronic commerce"

SC1 (Principles and methods)

[ISO 860:1996] "Terminology work – Harmonization of concepts and terms"

[ISO 1087-1:2000] "Terminology work – Vocabulary – Part 1: Theory and application"

[ISO 704: 2000/Rev] "Terminology work – Principles and methods"

[ISO AWI22134] "Terminology in sociolinguistic application"

SC2 (Terminography and Lexicography)

[ISO/DIS 12615] "Bibliographic references and source identifiers for terminology work"

[ISO FDIS 639-1] "Codes for the representation of names of languages – Alpha-2 codes"

[ISO 639-2: 1999] "Codes for the representation of names of languages – Alpha-3 codes"

[ISO 12616] "Translation-oriented terminography"

[ISO 15188] "Project management guidelines for terminology standardization"

[ISO 12199:2000] "Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet"

[ISO 1951: 1997] "Lexicographical symbols and typographical conventions for use in terminography"

[ISO 10241:1992 Rev.: ISO/WD 10241] "International Terminology Standards – Preparation and Layout"

[PWI 22128] "Quality criteria for terminology products"

SC3 (Computer applications)

[ISO 1087-2:2000] "Terminology work – Vocabulary – Part 2: Computer applications"

[ISO 6156: 1987] "Magnetic tape exchange format for terminological/lexicographical records (MATER)"

[ISO 12200: 1999] "Terminology – Computer applications – MARTIC – Negotiated Interchange"

[ISO/TR 12618: 1994] "Computational aids in terminology – Creation and use of terminological databases and text corpora"

[ISO/PWI TR 12618] "Computational aids in terminology – Design implementation and use of terminology management systems"

[ISO 12620-1999] "Terminology – Computer applications – Data categories"

[ISO/PWI 12620-1] "Computer applications in terminology – Data categories – Part 1: Model for description and procedures for maintenance of data category registries for language resources"

[ISO/PWI 12620-2] "Computer applications in terminology – Data categories – Part 2: Terminological data categories"

[ISO/DIS 16642] "Computer applications in terminology – Terminological mark-up framework (TMF)"

SC4 (Language resource management)

[NWI 21829] "Terminology of language resource management"

[PWI] "Linguistic annotation framework"

[Task force] "Metadata for multimodal and multilingual information"

¹ KAIST (Korea Advanced Institute of Science and Technology), Daejeon 305-701 Korea (NHK STRL invited researcher until February/2003), kschoi@cs.kaist.ac.kr