

Vorgehen beim Aufbau eines semantisch fundierten Data Dictionaries

Heller B., Lippoldt K., Kühn K., Löffler M.

Universität Leipzig

Institut für Medizinische Informatik, Statistik und Epidemiologie

Domainspezifische Data Dictionaries bilden die einheitliche Begriffsbasis für Anwendungen (papier- und rechnerbasiert) eines spezifischen Bereiches (z.B. Klinische Studien). Das Ziel ist die semantisch konsistente Wiederverwendbarkeit von Begriffen und Regeln in verschiedenen Anwendungen. Die Erstellung eines bereichsspezifischen Data Dictionaries gliedert sich in mehrere Schritte:

- (1) Sammlung von Begriffen mit Volltextbeschreibung,
- (2) semantische Analyse,
- (3) Erstellung eines formalen Modells,
- (4) Implementation in einer logikbasierten Repräsentationssprache.

Die Sammlung relevanter Begriffe umfasst die Erhebung von Definitionsvarianten in verschiedenen Kontexten, Quellen der Begriffsbeschreibungen sowie Relationen. Die verteilte Erhebung und Beschreibung der Begriffe und der Relationen innerhalb eines medizinischen Kompetenznetzwerkes wird durch das selbstentwickelte internetbasierte Software-Werkzeug OntoBuilder unterstützt. Die standardisierte Eingabe von Begriffen und deren natürlich-sprachlichen Beschreibungen im Hinblick auf die spätere Analyse erfolgt dezentral durch Fachpersonal und Experten auf dem entsprechenden Gebiet anhand eines entwickelten Leitfadens. Dieser Leitfaden enthält lexikalische und semantische Regeln zur Definition von Begriffen und Relationen. Der OntoBuilder unterstützt darüber hinaus auf Basis eines Nutzer- und Rollenkonzeptes die Konsensfindung einheitlicher Begriffsdefinitionen in natürlicher Sprache. Die in Volltext vorliegenden Begriffsbeschreibungen werden im zweiten Schritt hinsichtlich funktionaler, inhaltlicher und Granularitätsaspekte analysiert. Der Fokus hierbei liegt auf der Separierung generischer und domainspezifischer Konzepte sowie der Definition kontextabhängiger Eigenschaften und Relationen. Die semantisch aufbereiteten Begriffe werden in einem dritten Schritt in einem formalen Modell repräsentiert. Für ein flexibles handhabbares Data Dictionary sind hierbei folgende Anforderungen zu stellen: dynamischer Aufbau (Vermeidung statischer hierarchischer Konzeptbeziehungen), Trennung unterschiedlicher Sichten, einheitliche Partitionierungskriterien. Die beschriebene Vorgehensweise wird in der Domäne der Klinischen Studien im Rahmen der Projekte Kompetenznetz Maligne Lymphome und SEPNET angewendet.

Literaturauswahl:

- [1] Burgun A., Bogenreider, O. (2001): Aspects of the Taxonomic Relation in the Biomedical Domain. *Formal Ontology in Information Systems (FOIS)*: 222-233.
- [2] Heller, B., Löffler, M. (1999): Telematik und rechnerbasiertes Qualitätsmanagement in einem Kommunikationsnetzwerk. Teilprojekt 2 Kompetenznetz Maligne Lymphome. Antrag zur Förderung eines medizinischen Netzwerkes bei malignen Lymphomen; Universität Leipzig.
- [3] Miller, G. A. et. al. (1993): Introduction to Word Net: An On-line Lexical Database.
- [4] Rector, A. L. (1995): Coordinating Taxonomies: Key To Re-usable Concept Representations. Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95), Pavia, Italy, Springer: 17-28.
- [5] Rector, A. L. (1994): Compositional Models of Medical Concept: Towards Re-Usable Application - Independent Medical Terminologies (PS), in Barahona, P., Christiansen, J. P (eds.) *Knowledge and Decision in Health Telematics*, IOS Press: 109 - 114.