Abstract

Controlled terminologies and ontologies play a central role in basic and clinical research, and they are a key component to the metadata that is used within caBIG and crucial for the semantic interoperability of clinical and research information. Training in the terminology domain is not widespread, however, and relatively few understand the importance and potential application of terminologies in scientific research.

We explain the fundamental components of terminologies and basic principles of their construction. We also present scientific use cases designed to illustrate the relevance of terminologies to clinical and basic science investigators. Examples include:

- How terminologies are used to prevent medication errors
- How terminologies are used to provide standard definitions for clinical diagnoses and adverse events
- How terminologies can be used to facilitate data set identification and integration

The Vocabulary Knowledge Center is a centralized resource for education and development in the terminology domain. We provide domain expertise, documentation, and training materials designed to facilitate the adoption and use of terminologies. The Vocabulary Knowledge Center also supports software tooling for collaborative terminology authoring, editing, and services infrastructure.

Examples of Terminology Use Cases

Preventing Medication Errors

Decision support systems capable of detecting drug-drug or drug-disease interactions are becoming commonplace. However, they universally depend upon having existing or prescribed drugs specified in terms that these decision support systems can recognize and process. Patient records frequently contain mentions of drugs from other providers, either off-formulary or using terms that do not correspond to local drug dictionaries, which can prevent decision support systems from recognizing the terms.

The LexEVS tools and resources support the RxNorm drug dictionary, which is evolving to become the universal drug dictionary for use in the United States. Leveraging this resource with rich cross-mappings to alternate drug names enables decision support systems to recognize and process drug names that would be otherwise unrecognized; thus potentially preventing serious errors.

Standardizing Clinical Diagnoses and Adverse Events

Understanding the safety and efficacy of health activities, and specifically detecting adverse events, requires that the representation of patient findings, symptoms, observations, and summary diagnoses be comparable and consistent. One mechanism for ensuring the comparability and consistency of patient event data is to aggregate this information into a parsimonious number of event categories, as in syndromic surveillance; however, this results in substantial loss of detail and information.

The LexEVS tools permit the mapping of many representation forms about patient events into a common terminology, such as SNOMED or CTCAE, using the rich synonymy and cross-links among clinical terminologies and classifications. Aggregation into higher-level concepts can be easily undertaken by utilizing the inherent hierarchies of clinical terminology systems.

Facilitating Data Set Identification and Integration

The proliferation of public data sets and databases has resulted in an unprecedented pool of information that can be mined by investigators. Frequently, however, the data in those systems is annotated with plain-text phrases that contain substantial variation in lexical form and term usage. In order to identify, integrate, and aggregate information from disparate data sets with confidence, researchers must accurately understand the meaning of the terms used for annotation.

The LexEVS tools enable researchers to establish and use common terminologies for data annotation, and to create mappings between terminologies that can be used for data set integration and aggregation. Annotating data with terms from terminologies allows the data to be interpreted accurately by other researchers, making secondary uses of the data possible.

Terminologies

Terminology is a formal discipline for studying the labeling and designation of concepts particular to one or more subject fields or domains for the purpose of documenting and promoting correct usage. It is focused on the meaning and conveyance of concepts, and their inter-relationships within a culture.

Vocabulary Knowledge Center

The Vocabulary Knowledge Center is an organization funded by the NCI and operated by Mayo Clinic that offers expertise and support related to the vocabulary domain and tools in the caBIG program.

Mayo Clinic has extensive experience with terminology tooling and management, and has in-depth technical knowledge of LexEVS, LexWiki, and NCI Protégé – the key tools in the caBIG vocabulary domain.

Terminology Infrastructure

LexEVS is a collection of programmable interfaces that allow users to access controlled terminologies hosted by the NCI Enterprise Vocabulary Services (EVS) Project. LexEVS is built on LexGrid, a system for terminology storage, and it provides standard tooling for loading and distributing vocabulary content.

LexWiki is a collaborative terminology authoring platform based on Semantic MediaWiki and related technology. It is designed to provide terminology users and subject matter experts the ability to browse and discuss terminology content and structure. BiomedGT is an instance of LexWiki.

NCI Protégé is a collection of Protégé plugins that provide a customized editing environment tailored to the needs of NCI terminologies, including additional capabilities that are not available in Protégé. NCI Protégé is used to edit BiomedGT content and the NCI Thesaurus.

Collaborative Authoring

LexWiki is a collaborative terminology authoring platform based on Semantic MediaWiki and related technology. It is designed to provide terminology users and subject matter experts the ability to browse and discuss terminology content and structure. BiomedGT is an instance of LexWiki.

Terminology Editing

NCI Protégé is a collection of Protégé plugins that provide a customized editing environment tailored to the needs of NCI terminologies, including additional capabilities that are not available in Protégé. NCI Protégé is used to edit BiomedGT content and the NCI Thesaurus.

Overview of Supported Tools

To learn more about these tools or the services that we offer, please contact the Vocabulary Knowledge Center:

https://cabig-vc.nci.nih.gov/VocabKC
vocabulary@mayo.edu

Summary

- Controlled terminologies formally define terms and the relationships between them.
- Terminologies can be used to consistently annotate information (e.g., as metadata) from disparate systems.
- Terminologies facilitate the computational processing of medical records that contain information about medications, clinical diagnoses, and adverse events.
- Data that is annotated using terms from controlled terminologies can be compared, integrated, and aggregated consistently and without loss of information.

The Vocabulary Knowledge Center supports an integrated set of software tools for terminology authoring, editing, and management.

The Vocabulary Knowledge Center offers expertise and support for topics in the vocabulary domain.

Acknowledgements

- The Vocabulary Knowledge Center is funded by the NCI through a contract with SAIC-Frederick.
- The Vocabulary Knowledge Center is a designated caBIG® Knowledge Center.