720. Best Practices for Ontology Construction

Summary

i2b2 uses **user-controlled vocabularies** to compose complex queries running on the clinical data. Such a vocabulary could be refefred to as a **concept scheme** consisting of concepts. **Concepts** have at least a natural language **label** and some **code** that unambiguously identifies it. This code is also used in an association to the medical facts. So, queries of concepts relate to the clinical data. The universe of all vocabularies used in an i2b2 project is called **i2 b2 ontology**.

A catchy architecture for the i2b2 ontology is of great importance for the usability of the whole i2b2 project. All clinical data (or facts) are stored in one central database table, independent of their origin, structure, source system or describing entity. It's up to the user to build a query by selecting the right parameters out of the provided i2b2 ontology. Therefore, the navigational hierarchy needs to be arranged in way that required concepts will be found easily, even in very large or very complex vocabularies.

Furthermore, when mapping source data to the i2b2 data model, there is often more than one way to represent the characteristics of the source data elements. This is especially true for integrating data from heterogeneous sources or from different research projects when several conceptual models need to be harmonized (**late mapping**).

Best practices for i2b2 ontology construction will include:

- Creating an optimal hierarchy for users
- Naming concepts and folders
- Using concepts vs. using modifiers
- Using XML blobs
- Partitioning of the i2b2 ontology tables
- Miscellaneous