722. Naming concepts and folders

Length of labels

i2b2 allows labels that are supposed to be shown to the end users to have a length of 2000 characters. This would be equivalent to nearly one page of text and is of course unsuitable for the limited space on a standard computer screen.

It's better to limit the designation of a concept or folder to **40 characters at maximum**, most can be even shorter. But **be precise**: it's better to label a concept "Date of admission to the ICU" rather than "ICU admission" or "admission date" because every concept should be understandable on its own.

Order in the navigation tree

In the i2b2 navigation tree, folders are displayed before leafs. It can be disruptive when clinical forms contain some data elements with codelists (folders on level A) and code list items (leafs on level A+1) while other data elements have continuous values (leafs on level A). In this case, it is advisable to represent the latter with an **"auxiliary" folder** on level A and a leaf on level A+1.

Another problem is that i2b2 sorts all folders alphabetically, therefore mixing up the natural order which might result in misinterpretation ("blood drawn?" >> "date"). Since there is no attribute like "orderNumber" in i2b2, it's preferable to generate a numerical **tree-digit prefix** for all folders and leafs.

Naming of codes

There are no hard rules on how to designate concept codes, but because they are supposed to be unique, it is good practice to have a **combination of several prefixes and one suffix**. Prefixes are separated by the token "|" and the suffix by ":". The prefixes represent (certain parts of) the hierarchy and prevent reusing a code accidentally. The suffix represents the value of the property.

> ⓘ An good code for the concept *male* as a characteristic of *sex* in *demographics* would be DEM|SEX:MALE.

All relevant columns (C_BASECODE, CONCEPT_CD, MODIFIER_CD) have a length of 50 characters. The 50 chars should be shared equally between the prefixes and the suffix.

Codes are usually not shown to the user but are an affiliation of concept and fact. Therefore, it could be any unique combination of alphanumeric characters. In a number of scenarios, however, it makes perfect sense to use a "meaningful" code:

- you use a vocabulary (zip codes, country codes)
- you use a medical terminology (LOINC etc.)

> ⓘ In some of these cases, you should add the terminology code also to the label included in brackets, for instance for ICD10:C92.0 use a label "Acute myeloid leukemia C92.0."