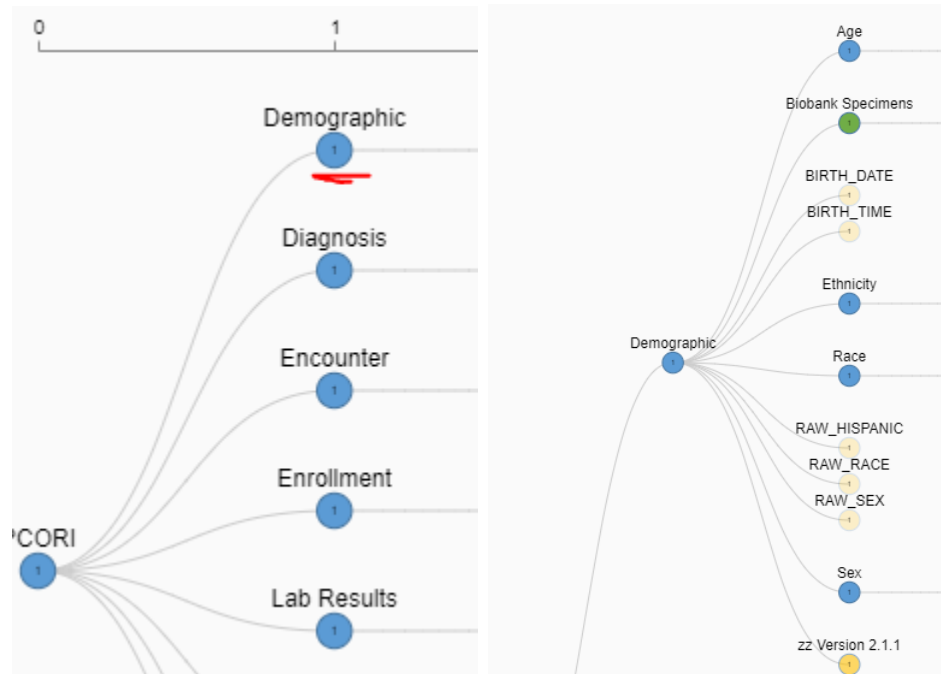# i2b2 SumTree Visualization

Basic interaction

This document serves as a guide for the usage of the i2b2 SumTree visualization. If you click on a container (blue) or folder (green), the underlying tree structure is revealed. For example:

Clicking on the "Demographic node" (left) will open up all the branches below it (right).

What is this visualization about?

i2b2 is a software program used by biomedical researchers to search large clinical databases for patients based on demographics (such as gender and age), diagnoses, laboratory tests, and other medical concepts. Collectively, these concepts are called the i2b2 "ontology", or "tree". Users select one or more concepts from the i2b2 tree to include in their queries. Hospitals and other healthcare institutions that install i2b2 typically include standard medical terminologies, such as "ICD-10" diagnoses and "RxNorm" medications, in their i2b2 tree. These contain hundreds of thousands of concepts, and they change regularly as new concepts are added and others are removed. This makes it difficult for researchers to understand what types of clinical data are available.

Because there are so many concepts, the entire expanded i2b2 tree is far too large to show on a computer screen. As a result, users can only see a tiny portion of the tree at any given time. To address this problem, we created a new "SumTree" visualization, which extends the traditional

tree view by graphically showing next to each concept summary statistics about the full subtree under that concept, such as the number of levels of depth deep below it. With SumTree, users can simultaneously see both the overall structure of the entire tree and details of individual concepts. This SumTree demo visualizes about 500,000 concepts under the root concept "PCORI". SumTree is 11 levels deep.

What do the symbols in the visualization mean?

Circles in SumTree represent individual concepts. You can double-click a circle to view its child concepts, as in a traditional tree view. However, to the right of each collapsed, i.e., not expanded, circle is a rectangle that summarizes its child concepts in a condensed format. To the right of that is another rectangle that summarizes the children of those child concepts. Additional rectangles are drawn until there are no further child concepts. A concept without any children is also called a "leaf". The total number of rectangles equals the depth of the subtree under the circle, with each rectangle summarizing the concepts at different levels.

What happens when you mouse over a square or circle?

If you mouse over a square or a circle, information about the data encoded by the square or circle will be displayed to you.

How do you use the legend on the left?

In the SumTree Legend, users select what they want the size, color, and opacity of the rectangles to represent. For example, the size can be either proportional to the number of concepts at the rectangle's level or proportional to the cumulative number of concepts at or below the rectangle's level. The count is displayed as a number above the rectangle. The color and opacity can indicate one of five concept attributes: (1) "Type" - whether the concept can be further expanded ("container" or "folder"); (2) "Status" - whether the concept is hidden by default or no longer available ("inactive"); (3) "Table" - the database table the concept references; (4) "Synonym" - whether the concept has the same meaning as another concept; and (5) "Metadata" - whether the concept has an associated value, such as a laboratory test which is expected to have a result. A rectangle can have multiple colors and opacities if it represents different kinds of concepts.

Original documentation:
http://weber.hms.harvard.edu/HealthcareSystemDynamics/

Terminology

General

Term	Definition
Tree	A tree is a type of network where concepts are related to each other in a hierarchical way.

Node	A node is a point in a network or a tree.
Parent	A parent is a node in a tree that has at least 1 node under it.
Child	A child is a node in a tree that has a connection to a higher-up node. All nodes in a tree except the root node are children of another node.
Root	A root is a node in a tree that is not a child of another node, and thus does not have a parent node.
Leaf	A leaf is a node in a tree that has no children.
Branch	A branch is a subset of nodes in a tree that can all be traced back to the same node that is directly attached to the root.
Level	A level is the step in a tree hierarchy where a certain node resides. The node(s) at level 0 is/are called the root. The i2b2 tree is 11 levels deep.

SumTree Specific

Term	Definition
Concept	Concepts are the building blocks of the SumTree visualization. Singular concepts are depicted as squares, while cumulated concepts are depicted as circles. Examples of concepts are demographics (such as gender and age), diagnoses, laboratory tests, and other medical terms.
Count	The count variable specifies whether concepts are counted "By Level" or "Cumulative".
Color	With the color option, you get to choose what variable of the tree is encoded by color. For example, "Type", the default choice, gives three distinct colors to nodes by their type. As a result, containers are blue, folders are green, and leaves are yellow. Similarly, when "Status" is chosen, the same colors are assigned but based on the status of the concepts rather than their type.
Folder	A container is a folder that is not allowed to be used in a query. Containers are typically used for concepts that match so many patients that queries would take a very long time to run.
Container	A container is a folder that is not allowed to be used in a query. Containers are typically used for concepts that match so many patients that queries would take a very long time to run.
Opacity	With the opacity option, you get to set the level of opacity for the nodes drawn in the visualization. For example, setting the opacity by "Status" encodes active nodes as fully opaque, hidden nodes as half-opaque, and inactive nodes as completely transparent. Similarly, setting this option to "Type" encodes containers as fully opaque, folders as half-opaque, and leaves as completely transparent.
Type	Types are "Container", "Folder", or "Leaf", see above.
Status	Status can be "Active", "Hidden", or "Inactive".
Table	Table Name - Encoding by Ontology table names. 5) what table it came from
Synonym	Encoding by if a node is a synonym of another node or not.
Metadata	Encoding by if a node has XML metadata associated with the concept.
Fixed Color	If this option is selected, all nodes will be drawn with the same color.
Fully Opaque	If this option is selected, all nodes will be drawn with the same opacity.