Similarity Estimation

XML Tree (Use Java DOM for this to parse?)

(<https://docs.oracle.com/javase/tutorial/jaxp/dom/readingXML.html>)

1. Canonical Solution parsed, each “level” becomes a list of nodes denoted by name, sorted alphabetically
2. Student solution parsed, each “level” becomes a list of nodes denoted by name, sorted alphabetically

Use an existing decision tree reduction algorithm: applying the algorithm to a tree will result in a final, irreducible tree that conveys exactly the same information as the original. Such a tree should be “optimal” so that comparison between two optimal trees lead to absolute differences. (Maybe base it upon <https://dl.acm.org/citation.cfm?id=96576>?)

Collapsed true guaranteed to be unique - Proof for uniqueness of tree collapsing from Dustin

Grading will be done by measuring the similarity of two optimally reduced decision trees in term of their semantic similarity. This can be done by measuring the percentage of times both trees produce the same result based on given input criteria. Depending on the size of the tree, we could run a sample space of all possible input permutations to each tree and compare outputs, assigning overall performance a percentage and then weighting the final grading based upon this. Intermediate steps and results could be saved along processing state, and areas where the paths diverged could be saved to a set that would be used to show where the student’s tree differed in decision making from the canonical solution (i.e. the website would show given input “x-y-z” at step “y” correct solution node: “” and your solution node: “”)

Baseline Grading: Existing set of nodes from the canonical solution. Begin by processing the student solution tree once and confirm presence/absence of all nodes. Assign a baseline grade based on performance here, with no weight towards extra nodes. Could either be having all nodes present begins the grade at 100% and deductions for extra and misplacements, or have the baseline grade be 50% and passing input cases give additional points.

Pruning?

