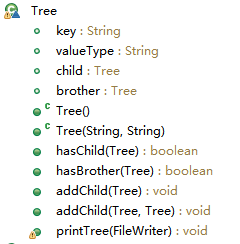
json转parquet schema部分：

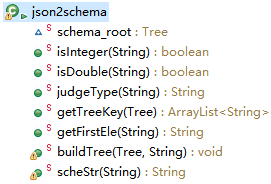
总体思路：

读取文件的每一条json记录，根据这些记录生成一棵孩子兄弟树。每个节点记录了两个值：key和valueType。key为json文件中的key，valueType为json文件中key所对应的value的类型。最后深度优先遍历该树，输出这个json文件的schema。

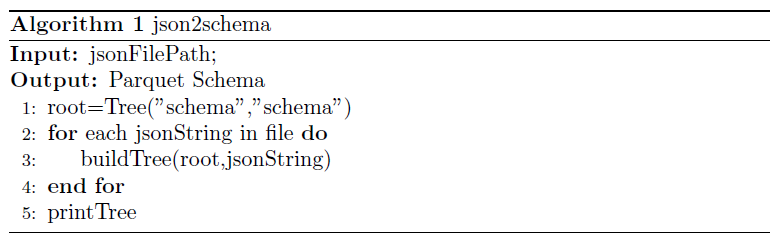
具体实现：

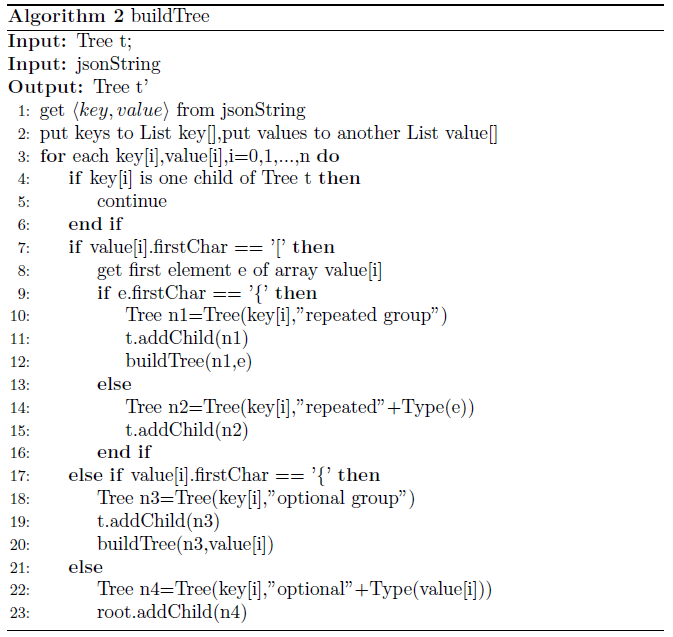
1、Tree类实现对树的一些操作，如加入孩子节点、判断是否有孩子节点、判断是否有兄弟节点等。printTree深度优先遍历Tree的每一个节点，并输出每个节点的valueType和key，组成schema。



2、json2schema类完成了将json文件转换为schema字符串的主要操作。

伪代码如下：





由于算法2中第4、5行，可以将之前没有出现过的<key,value>对考虑进来，所以生成的schema支持缺值操作；

第7行考虑了value为数组的情况，所以生成的schema可以支持多值情况；

第9行和第17行分别考虑了value为Object的数组的情况，以及value为Object的情况，并且在第12、20行嵌套调用了buildTree，所以生成的schema可以支持嵌套的情况。