# Mining Patterns in Source Code using Tree Mining Algorithms

- Context:
  - Software source codes have structure
  - Regularities in source code provide useful information for software engineers
- Challenges:
  - Output a large number of patterns
  - Not useful patterns
- Approach:
  - Constraint-based maximal tree mining

```
protected void updateEnabledState() {
   if (getView() != null) {
      setEnabled(...);
   } else {
      setEnabled(false);
    }
}
```

Fig.1. Source code

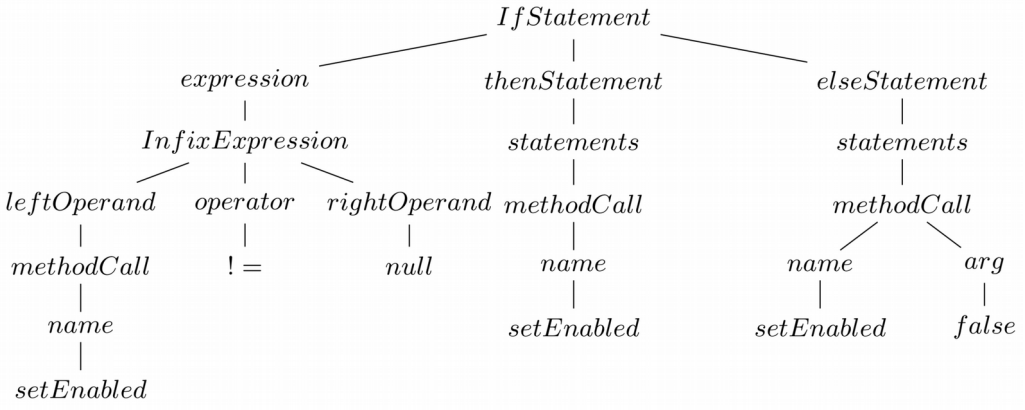
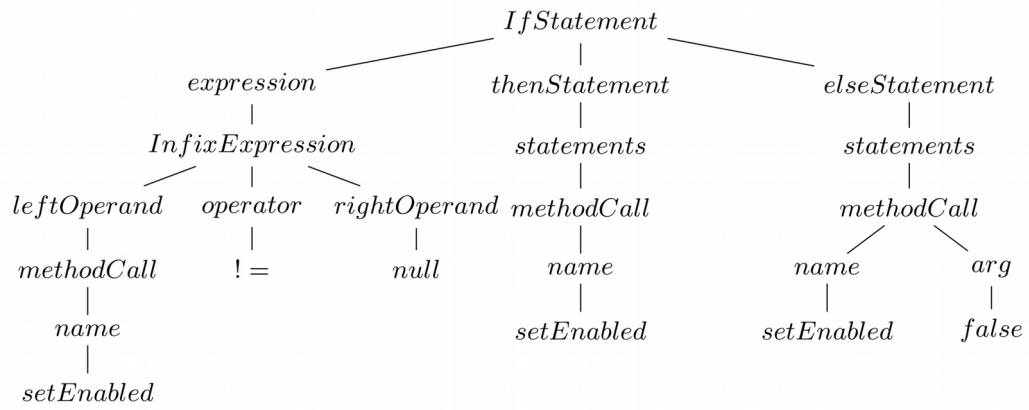



Fig.2. Abstract syntax tree

# Mining Patterns in Source Code using Tree Mining Algorithms

- Constraint-based maximal tree mining
  - Minimum size constraints
  - Constraints on labels
  - Constraints on leafs
  - Obligatory children




- Maximal subtree mining algorithm
  - Step 1: find patterns under constraints
  - Step 2: grow found patterns as large as possible, and return maximal patterns among these large patterns
- Remaining challenges:
  - Scalability