1 Semistructured Merge

1.1 Early Concepts

1. A_L and A_R are the sets of all the nodes added by left and right, respectively
2. D_L and D_R are the sets of all the nodes deleted by left and right, respectively

1.2 Merge Algorithms

Algorithm 1: Merge Files

Input: l, b, r, o

```
1  if l.content = b.content then
2  |  o.content ← r.content;
3  else if b.content = r.content ∨ l.content = r.content then
4  |  o.content ← l.content;
5  else
6  |  L ← fileToTree(l);
7  |  B ← fileToTree(b);
8  |  R ← fileToTree(r);
9  |  M ← mergeTrees(L, B, R);
10 |  H ← getActiveHandlers();
11 |  foreach h ∈ H do
12 |  |  h.handle(M);
13 |  end
14 |  o.content ← treeToText(M);
15 end
```

Algorithm 2: Merge Trees

Input: L, B, R
Output: result of merging left, base and right trees

```
1  L.index = 1;
2  B.index = 2;
3  R.index = 3;
4  LB ← mergeNodes(L, B, 1);
5  M ← mergeNodes(LB, R, 2);
6  D ← D_L ∩ D_R;
7  removeRemainingBaseNodes(M, D);
8  mergeMatchedContent(M);
9  return M;
```

2 Renaming Handler

2.1 Early Concepts

2.1.1 Possibly renamed without body changes nodes

$R_{wobc}(T, B) = \{b \in B \mid (\neg\exists t \in T\,(t.id = b.id))) \wedge (\exists t \in T\,(t.body = b.body))\}$

2.1.2 Possibly deleted or renamed with body changes nodes

$DR_{wbc}(T, B) = \{b \in B \mid \neg\exists t \in T\,(t.id = b.id \vee t.body = b.body)\}$

Algorithm 3: Remove Remaining Base Nodes

Input: T, D

```
1  if D = ∅ then return;
2  foreach d ∈ D do
3  |  if T = d then
4  |  |  P ← T.parent;
5  |  |  P.children ← P.children − T;
6  |  |  return;
7  |  end
8  end
9  foreach t ∈ T.children do
10 |  removeRemainingBaseNodes(t, D);
11 end
```

Algorithm 4: Merge Matched Content

Input: T

```
1  function mergeMatchedContent(T)
2  |  foreach t ∈ T.children do
3  |  |  mergeMatchedContent(t);
4  |  end
5  |  if T.children = ∅ then
6  |  |  if SEPARATOR ∈ T.body then
7  |  |  |  l, b, r ← split(T.body, SEPARATOR);
8  |  |  |  l ← l − MARKER;
9  |  |  |  T.body ← textualMerge(l, b, r);
10 |  |  end
11 |  end
12 end
```

Algorithm 5: Merge Nodes

Input: A, B, step
Output: result of merging node A and node B

```
1  if A.type ≠ B.type ∨ A.id ≠ B.id then
2  |  return null;
3  end
4  M ← A;
5  if A.children = ∅ ∧ B.children = ∅ then
6  |  if MARKER ∈ A.body then
7  |  |  M.body ← A.body + B.body;
8  |  else if step = 1 then
9  |  |  M.body ← MARKER + A.body + SEPARATOR + B.body + SEPARATOR;
10 |  else if A.index = 1 then
11 |  |  M.body ← MARKER + A.body + SEPARATOR + SEPARATOR + B.body;
12 |  else
13 |  |  M.body ← MARKER + SEPARATOR + A.body + SEPARATOR + B.body;
14 |  end
15 |  return M;
16 else if A.children ≠ ∅ ∧ B.children ≠ ∅ then
17 |  foreach b ∈ B.children do
18 |  |  a ← find(a ∈ A.children → a.type = b.type ∧ a.id = b.id);
19 |  |  if a = null then
20 |  |  |  M.children ← M.children ∪ b;
21 |  |  |  if step = 1 then D_L ← D_L ∪ b ;
22 |  |  |  else A_R ← A_R ∪ b ;
23 |  |  else
24 |  |  |  if a.index = -1 then a.index ← A.index ;
25 |  |  |  if b.index = -1 then b.index ← B.index ;
26 |  |  |  if step = 1 ∧ a ∈ A_L then A_R ← A_R ∪ b ;
27 |  |  |  m ← mergeNodes(a, b, step);
28 |  |  |  M.children ← M.children ∪ m;
29 |  |  end
30 |  end
31 |  foreach a ∈ A.children do
32 |  |  b ← find(b ∈ B.children → a.type = b.type ∧ a.id = b.id);
33 |  |  if b = null then
34 |  |  |  ls ← leftSiblings(a);
35 |  |  |  rs ← rightSiblings(a);
36 |  |  |  M.children ← ls ∪ a ∪ rs;
37 |  |  |  if step = 1 then A_L ← A_L ∪ a ;
38 |  |  |  else D_R ← D_R ∪ a ;
39 |  |  end
40 |  end
41 |  return M;
42 end
43 return null;
```

2.2 Match Algorithm

Algorithm 6: Match Algorithm

Input: L, B, R, M

Output: Set of quadruples (l, b, r, m) consisting of the base node b and its corresponding left node l, right node r and merge node m

```
1  matches ← ∅;
2  foreach b in possiblyRenamedOrDeletedBaseNodes(L, B, R) do
3  |  l ← getCorrespondentNode(b, L);
4  |  r ← getCorrespondentNode(b, R);
5  |  m ← getMergeNode(l, r, M);
6  |  matches ← matches ∪ (l, b, r, m);
7  end
8  return matches
9  function possiblyRenamedOrDeletedBaseNodes(L, B, R)
10 |  return DR_wbc(L, B) ∪ DR_wbc(R, B) ∪ R_wobc(L, B) ∪ R_wobc(R, B);
11 end
12 function getCorrespondentNode(b, T)
13 |  t ← findFirst(t ∈ T → t.id = b.id);
14 |  if t = null then
15 |  |  t ← findFirst(t ∈ T → t.body = b.body);
16 |  end
17 |  if t = null then
18 |  |  t ← findFirst(t ∈ T → t.body ≈ b.body ∧ ( t.id.name = b.id.name∨
   |  |    t.id.params = b.id.params));
19 |  end
20 |  if t = null then
21 |  |  t ← findFirst(t ∈ T → t.body = substring(b.body) ∨ b.body = substring(t.body));
22 |  end
23 |  return t;
24 end
25 function getMergeNode(l, r, M)
26 |  if l ≠ null then
27 |  |  return find(m ∈ M → m.id = l.id);
28 |  end
29 |  if r ≠ null then
30 |  |  return find(m ∈ M → m.id = r.id);
31 |  end
32 |  return null;
33 end
```

2.3 Handle Algorithms

Algorithm 7: Check References and Merge Methods Variant

Input: (l, b, r, m), M

```
1  if singleRenamingOrDeletion(l, b, r) then
2  |  m.body = textualMerge(l, b, r);
3  |  removeUnmatchedNode(l, r, m, M);
4  else if l.id ≠ r.id then
5  |  m.body = conflit(l, b, r);
6  |  removeUnmatchedNode(l, r, m, M);
7  else if l.body ≠ r.body then
8  |  if newReferenceTo(l) ∨ newReferenceTo(r) then
9  |  |  m.body = conflict(l, b, r);
10 |  |  removeUnmatchedNode(l, r, m, M);
11 |  else
12 |  |  m.body = textualMerge(l, b, r);
13 |  |  removeUnmatchedNode(l, r, m, M);
14 |  end
15 end
16 function singleRenamingOrDeletion(l, b, r)
17 |  return l.id = b.id ∨ r.id = b.id;
18 end
19 function removeUnmatchedNode(l, r, m, M)
20 |  if l.id = m.id ∧ r.id ≠ m.id then
21 |  |  removeNode(r, M);
22 |  end
23 end
```

Algorithm 8: Merge Methods Variant

Input: (l, b, r, m), M

```
1  m.body = textualMerge(l, b, r);
2  removeUnmatchedNode(l, r, m, M);
3  function removeUnmatchedNode(l, r, m, M)
4  |  if l.id = m.id ∧ r.id ≠ m.id then
5  |  |  removeNode(r, M);
6  |  end
7  end
```

Algorithm 9: Check Textual and Keep Both Methods Variant

Input: (l, b, r, m), M

```
1  if singleRenamingOrDeletion(l, b, r) then
2  |  if textualMergeHasConflictInvolvingSignature(b) then
3  |  |  m.body = conflict(l, b, r);
4  |  |  removeUnmatchedNode(l, r, m, M);
5  |  end
6  else if l.id ≠ r.id ∧ l.body = r.body then
7  |  m.body = conflict(l, b, r);
8  |  removeUnmatchedNode(l, r, m, M);
9  end
10 function singleRenamingOrDeletion(l, b, r)
11 |  return l.id = b.id ∨ r.id = b.id;
12 end
13 function removeUnmatchedNode(l, r, m, M)
14 |  if l.id = m.id ∧ r.id ≠ m.id then
15 |  |  removeNode(r, M);
16 |  end
17 end
```

Algorithm 10: Keep Both Methods Variant

Input: (l, b, r, m), M

```
1  if singleRenamingOrDeletion(l, b, r) ∧ hasConflict(m) then
2  |  removeConflict(m);
3  end
4  function singleRenamingOrDeletion(l, b, r)
5  |  return l.id = b.id ∨ r.id = b.id;
6  end
```