Bidirectionalizing Graph Transformations

Soichiro Hidaka Zhenjiang Hu Kazutaka Matsuda Keisuke Nakano
Kazuhiro Inaba Hiroyuki Kato Tohoku University, Japan
National Institute of Informatics, Japan The University of
hidaka.hu,kinaba,kato}@nii.ac.jp kztg@kb.ecei.tohoku.ac.jp

Categories and Subject Descriptors D.3.2 [Programming Languages]: Language Classifications—Specialized application languages; E.1 [Data Structures]: Graphs and networks

General Terms Design, Languages

Keywords bidirectional transformation, view updating, graph query and transformation, structural recursion

1. Introduction

Bidirectional transformations (Czarnecki et al. 2009; Foster et al. 2005) provide a novel mechanism for synchronizing and maintaining the consistency of information between input and output. Despite many promising results on bidirectional transformations, these have been limited to the context of relational or XML (tree-like) databases. We challenge the problem of bidirectional transformations within the context of graphs, by proposing a formal definition of a well-behaved bidirectional semantics for UnCAL, i.e., a graph algebra for the known UnQL graph query language. The key to our successful formalization is full utilization of both the recursive and bulk semantics of structural recursion on graphs. We carefully refine the existing forward evaluation of structural recursion so that it can produce sufficient trace information for later backward evaluation. We use the trace information for backward evaluation to reflect in-place updates and deletions on the view to the source, and adopt the universal resolving algorithm for inverse computation and the narrowing technique to tackle the difficult problem with insertion. We prove our bidirectional evaluation is well-behaved. Our current implementation is available online and confirms the usefulness of our approach with nontrivial applications.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICFP’10, September 27–29, 2010, Baltimore, Maryland, USA.
Copyright © 2010 ACM 978-1-60558-794-3/10/09 ...$10.00

First, UnQL/UnCAL is a graph query language that has been well studied in the database community with a solid foundation and efficient implementation. It has a concise and practical surface syntax based on select-where clauses like SQL, and can be easily used to describe many interesting graph transformations. First, UnQL/UnCAL is a graph query language that has been well studied in the database community with a solid foundation and efficient implementation. It has a concise and practical surface syntax based on select-where clauses like SQL, and can be easily used to describe many interesting graph transformations.

Second, and more importantly, graph transformations in UnQL can be automatically mapped to those in terms of structural recursion in UnCAL, which can be evaluated in a bulk manner (Buneman et al. 2000); a structural recursion is evaluated by first processing in parallel on all edges of the input graph and
then combining the results. This bulk semantics significantly contributes to our bidirectionalization, providing a smart way of treating shared nodes and cycles in graphs and of tracing back from the view to the source.

Our main technical contributions are summarized as follows.

- We are, as far as we are aware, the first to have recognized the importance of structural recursion and its bulk semantics in addressing the challenging problem of bidirectional graph transformation, and succeeded in a novel two-stage framework of bidirectional graph transformation based on structural recursion. We demonstrate that graph transformations defined in terms of structural recursions (being suitable for optimization as have been intensively studied thus far (Buneman et al. 2000)) make backward evaluation easier.

- We give a formal definition of bidirectional semantics for UnCAL by (1) refining the existing forward evaluation so that it can produce useful trace information for later backward evaluation (Section 4), and (2) using the trace information to reflect in-place updates and deletions on the view to the source, and adopt the narrowing technique to tackle the difficult problem with insertion (Section 5). We prove our bidirectional evaluation is well-behaved.

- We have fully implemented our bidirectionalization presented in this paper and confirmed the effectiveness of our approach through many non-trivial examples, including all those presented in this paper and some typical bidirectional graph transformations in database management and software engineering. More examples and demos are available on our BiG project Web site.

We consider an operation based approach, which means that the user explicitly provides editing operations in terms of "rename", "delete", and "insert". Currently these operations are treated according to the order specified by users. It might be challenging to produce these operation sequences automatically from the states before and after user's modifications on the view, but it is beyond the scope of this paper.

The forward transformations we consider is based on UnCAL, which is bisimulation generic, meaning that the transformation can’t distinguish between graphs that are bisimilar. For example, it can’t extract “first child of a node”. Extending our model to cope with order is included in our future work.

Also note that backward transformation is not bisimulation generic, meaning that two results of updates that are bisimilar do not always lead to bisimilar source. However, this is not necessarily a limitation introduced by our bidirectionalization, since this asymmetry comes from the expressiveness of conditional expression in the original UnCAL graph algebra. Similar argument apply for isomorphic updates.

Outline We start with a brief review of the basic concept of a graph data model and the structural recursion of UnCAL in Section 2. Then, we clarify the bidirectional properties within our context and give an overview of our two-staged framework for bidirectionalizing graph transformations in Section 3. After explaining how to extend the forward evaluation of UnCAL with trace information in Section 4, we give a formal definition of bidirectional semantics for UnCAL and prove that it is well-behaved in Section 5. We discuss implementation issues in Section 6 and related work in Section 7. We conclude the paper in Section 8.

2. UnCAL: A Graph Algebra

We adopted UnCAL (Buneman et al. 2000), a well-studied graph algebra, as the basis of our bidirectional graph transformation. We will briefly review its graph data model and the core of UnCAL.

2.1 Graph Data Model

We deal with rooted, directed, and edge-labeled graphs with no order on outgoing edges. They are edge-labeled in the sense that all information is stored on labels of edges while labels of nodes serve only as a unique identifier without a particular meaning. UnCAL graph data model has two prominent features, markers and ε-edges. Nodes may be marked with input and output markers, which are used as an interface to connect them to other graphs. An ε-edge represents a shortcut of two nodes, working like the ε-transition in an automaton. We use Label to denote the set of labels and M to denote the set of markers.

Formally, a graph G, sometimes denoted by G(V,E,I,O), is a quadruple (V,E,I,O), where V is a set of nodes, E ⊆ V × Label is a set of edges, I ⊆ M × V is a set of pairs of an input marker and the corresponding input node, and O ⊆ V × M is a set of pairs of output nodes and associated output markers. For each marker m ∈ M, there is at most one node v such that (m,v) ∈ I. The node v is called an input node with marker m and is denoted by I(m). Unlike input markers, more than one node can be marked with an identical output marker. They are called output nodes. Intuitively, input nodes are root nodes of the graph (we allow a graph to have multiple root nodes, and for singly rooted graphs, we often use default marker ε to indicate the root), while an output node can be seen as a “context-hole” of graphs where an input node with the same marker will be plugged later. We write inMarker(G) to denote the set of input markers and outMarker(G) to denote the set of output markers in a graph G. In addition, we write label(ξ) to denote the label of the edge ξ.

Note that multiple-marker graphs are meant to be an internal data structure for graph composition. In fact, the initial source graphs of our transformation have one input marker (single-rooted) and no output markers (no holes). For instance, the graph in Figure 1(a) is denoted by (V,E,I,O) where V = {1, 2, 3, 4, 5, 6}, E = {1, a, 2), (1, b, 3), (1, c, 4), (2, a, 5), (3, a, 5), (4, c, 4), (5, d, 6)}, I = {{(ε, 1)}, and O = {}.

Value Equivalence between Graphs Two graphs are value equivalent if they are bisimilar. Please refer to (Buneman et al. 2000) for the complete definition. Informally, graph G1 is bisimilar to graph G2 if every node x1 in G1 has at least a bisimilar counterpart x2.

† This analogy would choose NFA rather than DFA, since we allow multiple outgoing edges with identical labels from a node.
in $G_2$ and vice versa, and if there is an edge from $x_1$ to $y_1$ in $G_1$, then there is a corresponding edge from $x_2$ to $y_2$ in $G_2$ that is a bisimilar counterpart of $y_1$, and vice versa. Therefore, unfolding a cycle or duplicating shared nodes does not really change a graph. This notion of bisimulation is extended to cope with $\varepsilon$-edges. For instance, the graph in Figure 1(b) is value equivalent to the graph in Figure 1(a); the new graph has an additional $\varepsilon$-edge (denoted by the dotted line), duplicates the graph rooted at node 5, and unfolds and splits the cycle at node 4. Unreachable parts are also disregarded, i.e., two bisimilar graphs are still bisimilar if one adds subgraphs unreachable from input nodes.

**Graph Constructors** Figure 2 summarizes the nine graph constructors that are powerful enough to describe arbitrary (directed, edge-labeled, and rooted) graphs (Buneman et al. 2000):

\[ G ::= \{ \} \quad \{ \text{single node graph} \} \]
\[ | \{ l : G \} \quad \{ \text{an edge pointing to a graph} \} \]
\[ | G_1 \cup G_2 \quad \{ \text{graph union} \} \]
\[ | \varepsilon x := G \quad \{ \text{label the root node with an input marker} \} \]
\[ | \varepsilon y \quad \{ \text{a node graph with an output marker} \} \]
\[ | () \quad \{ \text{empty graph} \} \]
\[ | G_1 \oplus G_2 \quad \{ \text{disjoint graph union} \} \]
\[ | G_1 \otimes G_2 \quad \{ \text{append of two graphs} \} \]
\[ | \text{cycle}(G) \quad \{ \text{graph with cycles} \} \]

Here, $\{ \}$ constructs a root-only graph, $\{ l : G \}$ constructs a graph by adding edge $l$ pointing to the root of graph $G$, and $G_1 \cup G_2$ adds two $\varepsilon$-edges from the new root to the roots of $G_1$ and $G_2$. Also, $\varepsilon x := G$ associates an input marker, $\varepsilon x$, to the root node of $G$, $\varepsilon y$ constructs a graph with a single node marked with an output marker $\varepsilon y$, and () constructs an empty graph that has neither a node nor an edge. Further, $G_1 \oplus G_2$ constructs a graph by using a componentwise $(V, E, I$ and $O)$ union. $\cup$ differs from $\oplus$ in that $\cup$ unifies input nodes while $\oplus$ does not. $\oplus$ requires input markers of operands to be disjoint, while $\cup$ requires them to be identical. $G_1 \otimes G_2$ composes two graphs vertically by connecting the output nodes of $G_1$ with the corresponding input nodes of $G_2$ with $\varepsilon$-edges, and $\text{cycle}(G)$ connects the output nodes with the input nodes of $G$ to form cycles. Newly created nodes have unique identifiers. We will give this creation rule extended for our bidirectionalization in Section 4.1. The definition here is based on graph isomorphism (identical graph construction expressions results in identical graphs up to isomorphism), and they are, together with other operators, also bisimulation generic (Buneman et al. 2000), i.e., bisimilar result is obtained for bisimilar inputs.

UnCAL (Unstructured Calculus) is an internal graph algebra for the graph query language UnQL, and its core syntax is depicted in Figure 3. It consists of the graph constructors, variables, conditionals, and structural recursion. We have already detailed the data constructors, while variables and conditionals are self explanatory. Therefore, we will focus on structural recursion, which is a powerful mechanism in UnCAL to describe graph transformations.

A function $f$ on graphs is called a structural recursion if it is defined by the following equations\(^1\)

\[ f(\{ \} ) = \{ \} \]
\[ f(\{ s! : s@ \} ) = e @ f(s! \oplus f(s@)) \]
\[ f(s! \cup f(s@)) = f(s! \oplus f(s@)) \]

where the expression $e$ may contain references to variables $s!$ and $s@$ (but no recursive calls to $f$). Since the first and the third equations are common in all structural recursions, we write the structural recursion in UnCAL simply as

\[ f(sdb) = \text{rec}(\lambda(s! \oplus s@).e)(sdb) \]

Despite its simplicity, the core UnCAL is powerful enough to describe interesting graph transformation including all graph queries (in UnQL) (Buneman et al. 2000), and nontrivial model transformations (Hidaka et al. 2009). Some simple examples are given below.

**Example 1.** The graph equivalent to that in Figure 1(a) can be constructed as follows (though not uniquely).

\[ \varepsilon z @ \text{cycle}(e \oplus (\varepsilon z := \{ a : \{ a : \{ a : s_3 \} \} \} \cup \{ b : \{ a : s_3 \} \} \cup \{ c : s_2 \}) \]
\[ \oplus (\varepsilon z_1 := \{ d : \{ \} \}) \]
\[ \oplus (\varepsilon z_2 := \{ c : s_2 \}) \]

For simplicity, we often write $\{ l_1 : G_1 \cup \ldots \cup l_n : G_n \}$ to denote $\{ l_1 : G_1 \} \cup \ldots \cup \{ l_n : G_n \}$.

**Example 2.** The following structural recursion $a2b$ replaces edge label $a$ with $b$ and leaves other labels unchanged.

\[ a2b(sdb) = \text{rec}(\lambda(s! \oplus s@).e).f(s! \oplus e^2(s! \oplus (s@)^4(sdb)^5)) \]

(The superscripts are for identifying code positions, which will be important in Section 4; they can simply be ignored for now.) Here is an instance of an execution:

\[ a2b(\varepsilon z : a \oplus \varepsilon x) = \varepsilon z : b \]

where $\circ$ denotes the root of the graph.

\(^1\)Informally, the meaning of this definition can be considered to be a fixed point (though may not necessarily unique) over the graph, which is again defined by a set of equations using the three constructors $\{ \}, \cup$, and $\oplus$. For instance, the graph in Figure 1(a) can be considered to be the fixed point of the following equations:

\[ G_{\text{root}} = \{ a : G_3 \}, b : \{ a : G_3 \}, c : G_4 \]
\[ G_3 = \{ d : \{ \} \} \]
\[ G_4 = \{ c : G_4 \} \]
Example 5.

$\text{abstract parameter}$ graph in Figure 4(c).

Example 3.

can refer to outer variables (as multiple markers. For instance, the following function
tracts subgraphs that can be accessible by traversing two connected mutually recursive call, and $abab$ to consist of two mutual recursive
functions. The first is $\varepsilon \mathcal{Z}_1$, which, at each edge in the original graph, generates a new a edge pointing to the result of $\varepsilon \mathcal{Z}_2$ at the original
destination node. The second is $\varepsilon \mathcal{Z}_2$ that generates b edges pointing to the result of $\varepsilon \mathcal{Z}_1$ from its destination. The result of the whole
expression is defined to be the result of the $\varepsilon \mathcal{Z}_1$ at the root node of the argument graph. The following figure should be helpful. The dashed edges denote the edges that are unreachable from the output
root node.

$abab(\varepsilon d \rightarrow d b) = \varepsilon \\
\begin{array}{c}
\text{Figure 4. Bulk Semantics of Structural Recursion in UnCAL} \\
\text{Example 3. The following structural recursion $a2d_{\text{xc}}$ replaces all} \\
\text{labels a with d and removes edges labeled c.} \\
a2d_{\text{xc}}(\{db\}) = \\
\text{rec}(\lambda (sl, sg). \text{if sl = a then} \{d : \varepsilon^1\}^2 \text{else if sl = c then} \varepsilon : \varepsilon^2)^4 \text{else} \{\varepsilon : \varepsilon^3\}^6 (\{db\})^7$.

Applying the function $a2d_{\text{xc}}$ to the graph in Figure 1(a) yields the graph in Figure 4(c).

Example 4.
The following structural recursion $consecutive$ extracts subgraphs that can be accessible by traversing two connected edges of the same label.

$consecutive(\{db\}) = \\
\text{rec}(\lambda (sl, sg), \text{rec}(\lambda (sl', sg')). \text{if sl = sl' then} \text{result : sg'}^1 \text{else} \{\text{result : sg'}^2 \text{result : sg'}^3 \text{result : sg'}^3 \text{result : sg'}^4 (\{db\})^7)$.

For example, we have

$consecutive(\mathcal{A}) = \text{result}$

Note that the structural recursive definition of $consecutive$ uses graph parameter $sg'$ to achieve the transformation. Also note that structural recursions are allowed to be nested, and inner recursion can refer to outer variables (as $sl$ in the example). This enables us to express the join of multiple queries.

Example 5. Although the examples given so far are self-recursive, it is possible to simulate mutual recursion by returning graphs with multiple markers. For instance, the following function $abab$

$abab(\{db\}) = \varepsilon \mathcal{Z}_1 \uplus \text{rec}(\lambda (sl, sg). \text{if sl = a then} \{d : \varepsilon^1\}^2 \text{else if sl = c then} \varepsilon : \varepsilon^2)^4 \text{else} \{\varepsilon : \varepsilon^3\}^6 (\{db\})^7$.

changes all edges of even distances from the root node to a, and odd distance edges to b. We may consider the markers $\varepsilon \mathcal{Z}_1$ as a mutually recursive call, and $abab$ to consist of two mutual recursive
functions. The first is $\varepsilon \mathcal{Z}_1$, which, at each edge in the original graph, generates a new a edge pointing to the result of $\varepsilon \mathcal{Z}_2$ at the original
destination node. The second is $\varepsilon \mathcal{Z}_2$ that generates b edges pointing to the result of $\varepsilon \mathcal{Z}_1$ from its destination. The result of the whole
expression is defined to be the result of the $\varepsilon \mathcal{Z}_1$ at the root node of the argument graph. The following figure should be helpful. The dashed edges denote the edges that are unreachable from the output
root node.

$abab(\varepsilon d \rightarrow d b) = \varepsilon \\
\begin{array}{c}
2.3 \text{ Bulk Semantics of Structural Recursion} \\
By allowing $\varepsilon$-edges, we can evaluate a structural recursion in a bulk manner. Consider the structural recursion,

$\text{rec}(\lambda (sl, sg), e)$

which is to be applied to an input graph $G$. In bulk semantics, we apply body $e$ independently on every edge $(l, g)$ in $G$ where $l$ is the
label of the edge and $g$ is the graph that the edge is pointing to, then join the results with $\varepsilon$-edges (as in the $\uparrow$ constructor).

Recall the structural recursion $a2d_{\text{xc}}$ defined in Example 3. Applying it to the input graph in Figure 1(a) yields the graph in Figure 4(a), where each edge from i to j in the input graph leads to a subgraph containing a graph with an edge from $Eij$ to $Eij$ in the output graph (where the dotted edge denotes an $\varepsilon$-edge),
and these subgraphs are connected with $\varepsilon$-edges according to the original shape of the graph. If we eliminate all $\varepsilon$-edges as explained
in Section 3.2, we obtain a standard graph in Figure 4(c).

One distinct feature of bulk semantics is that the shape of the input graph is remembered through additional $\varepsilon$-edges, which will be
fully utilized in our later bidirectionalization.

3. Overview: Bidirectionalizing UnCAL

It is more challenging to bidirectionalize transformations on graphs than trees, because graphs may contain shared nodes or cycles. We shall demonstrate that the structural recursion in UnCAL can serve
as the basis to solve this problem. Although structural recursion was proposed within the context of query optimization, we will show that it plays a crucial role in our bidirectionalization.

3.1 Bidirectional Properties

Bidirectionalization is used to derive backward transformation from forward transformation. We approach the problem of bidirectionalization in graph transformation by providing a bidirectional semantics for UnCAL: forward semantics (forward evaluation) corresponds to forward transformation and backward semantics (backward evaluation) corresponds to backward transformation.

Before giving our bidirectional semantics for UnCAL, let us clarify the bidirectional properties that the forward and backward evaluations should satisfy. Let $F[e]_{\rho}$ denote a forward evaluation (get) of expression $e$ under environment $\rho$ to produce a view, and $B[e]_{\rho}(\rho, G')$ denote a backward evaluation (put) of expression $e$ under environment $\rho$ to reflect a possibly modified view $G'$ to the source by computing an updated environment. $\rho$ is a set of mappings with form $\{x \mapsto G\}$ with a graph (or label) $G$. The following are two important properties:

$F[e]_{\rho} = G$

$B[e]_{\rho}(\rho, G') = \rho'$

The $(\text{GETPUT})$ property states that unchanged view $G$ should give no change on the environment $\rho$ in the backward evaluation, while the $(\text{PUTGET})$ property states that if a view is modified to $G'$ which is in the range of the forward evaluation, then this modification can be reflected to the source such that a forward evaluation will produce the same view $G'$.

These two properties are essentially the same as those in (Foster et al. 2005). One problem with the $(\text{PUTGET})$ property is that it
needs to check whether a graph is in the range of forward evaluation, which is difficult to do in practice. To avoid this range check, we allow the modified view and the view obtained by backward evaluation followed by forward evaluation to differ, but require both views to have the same effect on the original source if backward evaluation is applied.

\[
B[e](p, G'') = R' \quad F[e]p' = G''
\]

The get in our (\text{WPUTGET}) can be considered as an amendment of the modified view \(G'\) to \(G''\). Certainly, if the (\text{PUTGET}) property holds, so does the (\text{WPUTGET}).

We say that a pair of forward and backward evaluations is well-behaved if it satisfies (\text{GETPUT}) and (\text{WPUTGET}) properties. In the rest of this paper, we will give a bidirectional evaluation (semantics) for UnCAL, and prove the following theorem, which is a direct consequence of Lemmas 2, 3, and 4 that will be discussed later.

**Theorem 1** (Well-behavedness). Our forward and backward evaluations are well-behaved, provided their evaluations succeed.

### 3.2 Two-Stage Bidirectionalization

Recall a\(a2d_{\text{xc}}\), which maps the source graph in Figure 1(a) to the view graph in Figure 4(c). The big gap between the source and the view makes it hard to reflect changes on the view to the source. Our idea to bridge this gap was to divide the forward evaluation into two easily handled stages:

- **Stage 1**: Forward evaluation (in the bulk semantics) with sufficient \(\varepsilon\)-edges, so that the output graph will have a similar shape to the input graph, making the later backward evaluation easier.
- **Stage 2**: Elimination of \(\varepsilon\)-edges to produce a usual view.

For a\(a2d_{\text{xc}}\), Stage 1 maps the source graph to the intermediate graph in Figure 4(a), and Stage 2 maps the intermediate graph to the view graph (Figure 4(c)).

Stage 1 satisfies the \(\varepsilon\)-marker preserving property: (1) No \(\varepsilon\)-edges are added or deleted, (2) Markers are not added, deleted, or changed and (3) Unreachable parts are not modified. This property is very important in our bidirectionalization, because it not only enforces the nine graph constructors so that they are invertible, but it also makes it easy to bidirectionalize structural recursion because there is a clear correspondence between the input and output graphs.

In the rest of this paper, we will focus on bidirectional graph transformation in Stage 1.

### 4. Traceable Forward Evaluation

An UnCAL expression usually specifies a forward evaluation mapping a graph database (which is just a graph) to a view graph (in Section 2). The main purpose of the present paper is to give backward evaluation (backward semantics), which specifies how to reflect view updates to the graph database. For this purpose, we have to detect how each node of the view is generated, particularly when it is constructed through connecting input/output markers and removing \(\varepsilon\)-edges, which are no longer in the view. To make the view more informative, viz., traceable, we enrich the original semantics of UnCAL by embedding trace information (like provenance traces (Cheney et al. 2008)) in all nodes of the view that possibly includes \(\varepsilon\)-edges. In this section, we explain what kind of trace information is embedded in the view, and extend the original semantics for UnCAL expressions to be evaluated into traceable views.

#### 4.1 Traceable Views

A view is obtained by evaluating an UnCAL expression with a database. Every node of the view originates in either a node of the database or a construct in the UnCAL expression, except when the node is generated through a structural recursion with a rec construct (in the bulk semantics). Recall that an expression rec\((\lambda(l, g, v_e))\) is evaluated by binding variables \(l\) and \(g\) in \(v_e\) to a part of the evaluation result of \(v_e\). In this case, a node in the view may originate not only in the whole rec expression but also a sub-expression in \(v_e\).

A traceable view is a view each node of which has information for tracing its origin. The information, called trace ID, is defined by

\[
\text{TraceID} ::= \text{SrcID} | \text{Code} \text{Pos Trace} \text{Marker} | \text{RecN} \text{Pos TraceID} \text{Marker} | \text{RecE} \text{Pos TraceID} \text{Edge},
\]

where \(\text{SrcID}\) ranges over identifiers uniquely assigned to all nodes of the database, \(\text{Pos}\) ranges over code positions in the UnCAL expression, \(\text{Marker}\) ranges over input/output markers, and \(\text{Edge} \) stands for \(\text{TraceID} \times \text{Label} \times \text{TraceID}\) with a set of labels \(\text{Label}\).

We now briefly explain the meaning of each trace ID. Let \(i\) be a trace ID of a node \(u\) in a traceable view. When \(i\) is a node identifier in \(\text{SrcID}\), node \(u\) originates in the node assigned by \(i\) in the database. When \(i\) is \(\text{Code}\) \(p\) \(\text{Pos}\) \(\text{Marker}\) with code position \(p\) and input marker \(\text{Marker}\), node \(u\) originates in the subexpression at \(p\) in the UnCAL expression. The marker \(\text{Marker}\) is only required when the subexpression is given by the or cycle construct. This is because these constructs yield as many \(\varepsilon\)-edges as input markers. When \(i\) is either \(\text{RecN} p l 0\) \(\text{Marker}\) or \(\text{RecE} p l 0\) \(\text{Marker}\) \(\text{Edge}\), node \(u\) is generated through the rec construct at the code position \(p\). \text{RecN}\ and \text{RecE}\ stand for what node and edge, respectively, of the argument of the recursion, the node originates in.

Let us explain these cases through an example where the UnCAL expression \(a2d_{\text{xc}}\) in Example 3 is applied to the database \(G_{\text{dc}}\) in Figure 1(a). The traceable view we want can be obtained from the graph \(G_{\text{view}}\) in Figure 4(a) by assigning trace IDs to all nodes. The trace ID assigned to node 1 in \(G_{\text{view}}\) is

\[\text{Figure 5. General } \varepsilon\text{-edge Elimination Procedure}\]
(RecN 7 1 z) because the node originates in node 1 of $G_{src}$ in $SrcID$, which is used as part of the argument of the ree construct at code position 7 in $a2d_{x,c}$. The trace ID assigned to node S12 in $G_{view}$ is (RecE 7 (Code 2) (1, a, 2)) because the node originates in the a-labeled edge from node 1 to 2 of $G_{src}$ in $Edge$ through the graph constructor $\{d : j\}$ at code position 2 in the ree construct at 7 in $a2d_{x,c}$. When the argument of the ree construct is also a ree expression, RecN and RecE in the trace ID are nested like (RecN p (RecE p' ... ) ... ) and (RecE p (RecE p' ... ) (RecN ... a, RecN ...)).

A traceable view is denoted by a quadruple $(V, E, I, O)$ just like an ordinary UnCAL graph. The only difference is that in traceable views, trace IDs are assigned to all nodes.

4.2 Enriched Forward Semantics

Traceable views can be computed by a simple extension of the original forward semantics of UnCAL so that tracing information is recorded when a node is created. Let $\rho$ denote an UnCAL subexpression $e$ at code position $p$. We write $\rho([z])$ for $G$ when $(z : z) \in \rho$. $\rho$ is naturally used as variable substitution in UnCAL expressions, e.g., $\rho e$ for an expression $e$. We inductively define the enriched forward semantics $F[e\|\rho]$ for each UnCAL construct of $e$.

Graph Constructor Expressions. The semantics of graph constructor expressions is straightforward according to the construction in Figure 2. For instance, we have

$$F[\{e_1 \cup e_2\}\|\rho] = F[e_1]\|\rho \cup F[e_2]\|\rho,$$

where $\cup\rho$ is a union operator for two graphs concerning position $p$ and is defined by

$$\begin{aligned}
&G_1 \cup\rho G_2 = \langle V \cup V_1 \cup V_2, E \cup E_1 \cup E_2, I, O_1 \cup O_2 \rangle \\
&V = \langle \{\text{Code } p \| \text{sm}, \text{v} \} \cup \{\text{sm}, \text{v} \} \in I_1 \cup I_2 \}
\end{aligned}$$

$I = \{\{\text{sm}, \text{Code } p \| \text{sm} \} \cup \{\text{sm} \in M \}$

We omit definitions for other constructor expressions.

Variable. A variable looks up its binding from environment $\rho$.

$$F[[\$v]\|\rho] = \rho[\$v]$$

Condition. The forward semantics of a condition is defined as

$$F[[\text{if } l_1 = l_2 \text{ then } e_1 \text{ else } e_2]\|\rho] = \begin{cases} F[e_1]\|\rho & \text{if } l_1 \rho = l_2 \rho \\ F[e_2]\|\rho & \text{otherwise} \end{cases}$$

It first evaluates the conditional expression $l_1 \rho = l_2 \rho$, and with the result it evaluates either the then branch or the else branch.

Structural Recursion. The semantics of a structural recursion is given by bulk semantics as reviewed in Section 2.3, which can be formally defined by

$$F[[\text{rec}(\lambda([l, \$g], e_0))(e_0)]\|\rho] = \text{compose}_{\text{rec}}(\text{fwd}\_\text{edge}\_\text{chage}(G_a, \rho, e_0, G_a, M))$$

where $M = \text{inMarker}(e_0) \cup \text{outMarker}(e_0)$, $G_a = F[e_0]\|\rho$.

where $\text{fwd}\_\text{edge}\_\text{chage}$ and $\text{compose}_{\text{rec}}$ are defined in Figure 6. Intuitively, $\text{fwd}\_\text{edge}\_\text{chage}$ evaluates the body expression $e_0$ at each edge $\rho$ of the argument graph $G_a$ obtained by evaluating $e_0 \rho$ and returns the set of result graphs. Then, $\text{compose}_{\text{rec}}$ glues all the graphs together along the structure of $G_a$ concerning code position $p$. Note that subgraph $(G, \rho)$ denotes the subgraph to which the edge $\rho$ is pointing in the graph $G$.

Example 6. We will now illustrate the semantics of ree through an example: the structural recursion $a2d_{x,c}$, which is defined with position information in Example 3, is applied to $G_{src}$ in Figure 1(a), and the traceable view is a graph similar to $G_{view}$ in Figure 4(a).

First, $G_{src}$ is bound to a variable $\$db$. Then, $\text{fwd}\_\text{edge}\_\text{chage}$ generates a set of pairs of an edge and a ‘local result’ for each edge in $G_{src}$. The local result is obtained by evaluating the body of ree under $\rho = \{\$db \mapsto G_{src}\} \cup \{\$/L, \$g \mapsto G\}$ with the label $L$ of the edge and a subgraph $G$ reachable from the edge. For example, as the local result for edge $G_{src}$ (Code 2, d, Code 1) with input node Code 2 and output node Code 1 is generated because the subexpression $(d : z)\}^2$ is used due to $L = a$. The function $\text{compose}_{\text{rec}}$ glues all pairs of an edge and a local result after adding RecN or RecE to their nodes. For example, regarding a pair of edge $\rho = (3, a, 5)$ and its local result containing edge (Code 2, d, Code 1), the set $\text{RecE}$ contains edge (Code 7 (Code 2) $\rho$, d, RecE 7 (Code 1) $\rho$) where $7$ is the code position of the concerned ree, while set $\text{RecE}$ contains edge (RecE 7 (Code 2, 3, $\rho$, $\rho$, RecE 7 (Code 2) $\rho$) and (RecE 7 (Code 1) $\rho$, $\rho$, RecE 7 (Code 2) $\rho$) due to $(\rho, \rho, \rho) \in O$ and $(\rho, \rho, \rho) \in O$. The former corresponds to the edge from $S35$ to $E35$ of $G_{view}$ and the latter corresponds to two edges from $3$ to $35$ and from $35$ to $5$ of $G_{view}$. In this example, $E_c$ is an empty set since $G_{src}$ has no e-edges. The sets $\text{RecN}$ and $\text{O}_{\text{RecE}}$ of input and output nodes are obtained with $I = \{(\rho, 1)\}$ and $O = \emptyset$, respectively, which are those of $G_{src}$. Hence, $\text{RecN} = \{\rho, \rho, \text{RecE} 7 (1, 5)\}$ and $\text{O}_{\text{RecE}} = \emptyset$ because $M = \text{inMarker}(e_0) \cup \text{outMarker}(e_0) = \{\rho\}$. Here, "\cdot" denotes Skolem function (Buneman et al. 2000) that satisfies $(\rho, \rho, \rho) = \rho\cdot(\rho, \rho, \rho)$ (associativity) and $\rho\cdot\rho$ = $\rho\cdot\rho$ = $\rho\cdot\rho$ (left and right identity).

More concretely, if the source graph is $s = \{(1, 2), a2d_{x,c}(s)\}$ gives the graph

which is bisimilar to the graph $\ldots$

5. Backward Evaluation of UnCAL

With traceable views and the $e$-marker preserving property (Section 3) on the modification of such views, backward evaluation (in Stage 1) turns out to be simpler for two reasons.

- First, the graph constructors become invertible. For instance, if $G = G_1 \cup G_2$, $G$ is modified to $G'$, but the modification is $e$-marker preserving; then, we can follow tracing information, $e$-edges, and marker information to uniquely decompose $G'$ to $G_1'$ and $G_2'$ such that $G_1' \cup G_2' \equiv G$ holds.\footnote{$G_1(V_1, E_1, I_1, O_1) \equiv G(V_2, E_2, I_2, O_2)$: the exact equivalence of two graphs, is defined by $V_1 = V_2 \land E_1 = E_2 \land I_1 = I_2 \land O_1 = O_2$.}

We will write this...
decomposition as $\text{decomp}_{G_1 \cup G_2}$, and applying it to $G'$ will give $(G_1, G_2)$.

- Second, backward evaluation of a structural recursion $\text{rec}(e)$ is reduced to that of its body $e$ (followed by result gluing), because of the bulk semantics of structural recursion.

Backward evaluation greatly depends on what updates are allowed on the view. We allow the following three general updates on our edge-labeled graphs: (1) in-place updates as modification of edge labels, (2) deletion of edges, and (3) insertion of edges or a subgraph rooted at a node. And we accept a sequence of these updates on the view and reflect them to the source. In the rest of this section, we shall explain the respective backward evaluation for these updates on views.

5.1 Reflection of In-place Updates

In this section, we formally define backward semantics for UnCAL, where only in-place updates are considered.

Recall that backward semantics $B[e](\rho, G')$ is used to compute a new environment from the original input environment $\rho$ and the modified view $G'$. Like forward semantics, backward semantics can be defined inductively over the construction of expression.

5.1.1 Backward Evaluation of Simple Expressions

Graph Constructor Expressions. Since each constructor is reversible and is associated with a decomposition function, we can decompose the views of constructor expressions so as to define the backward semantics inductively. For example, we have

$$B[(e_1 \cup e_2)](\rho, G') = B[e_1](\rho, G_1') \uplus_B B[e_2](\rho, G_2')$$

where $G_1' = \text{F}[e_1][\rho]$,

$$G_2' = \text{F}[e_2][\rho]$$

$$\text{decomp}_{G_1' \cup G_2'}(G')$$

Unlike Foster et al. (2005), we have variable binding, and therefore multiple environments produced by backward evaluation of the operands are merged by $\uplus_B$ defined below, using an approach similar to that in Liu et al. (2007), which deals with variable bindings.

$$\langle \rho_1 \uplus_B \rho_2 \rangle = \left\{ \begin{array}{ll}
(\rho \mapsto mg(G, G_1, G_2)) & \quad (\rho \mapsto G) \in \rho_1, \\
(\rho \mapsto G) & \quad (\rho \mapsto G) \in \rho_2
\end{array} \right.$$

where $mg(G, G_1, G_2) =

$$G_1$$ if $G_2 = G \lor G_1 = G_2$

$$G_2$$ if $G_1 = G$

FAIL otherwise

$\uplus_B$ unifies each binding by $mg$. If only the binding on the left hand side is modified ($G_2 = G$), or both are consistently updated

(5.1.2 Backward Evaluation of Structural Recursion)

Due to the traceable bulk forward evaluation of structural recursion $\text{rec}$ and the $\varepsilon$-marker preserving property that retains similarity in shape between input and output graphs, backward semantics can easily be defined as

$$B[\text{rec}((\lambda(l, g). c_1) \cup c_2)](\rho, G') = B[c_1](\rho, G_1') \uplus_B B[c_2](\rho, G_2')$$

where $G_1' = \text{F}[c_1][\rho]$,

$$G_2' = \text{F}[c_2][\rho]$$

This definition is easy to understand if we note duality with the definition of its forward semantics. Backward semantics first decomposes through $\text{decomp}_{\text{rec}}$ the modified result graph $G'$ into pieces of graphs, which is intuitively an inverse operation of $\text{compose}_{\text{rec}}$. For every non-$\varepsilon$ edge $\zeta \in E_\zeta$ in the source argument graph, the decomposition extracts (possibly modified) subpart $G'_\zeta$ of $G'$, which originates at the result $G_\zeta$ of the forward computation on the edge. Then, in $\text{bwd}_\text{edge}$, we carry out backward computation of the body expression $c_1$ on each edge and compute the updated environment $\rho'_1$. Finally, these environments are merged into the updated environment $\rho'$ of the whole expression. The merge function does two pieces of work. First, by combining the information $\rho'_1(\lambda(l, g))$ and $\rho'_1(\varepsilon)$ from the updated environments (and $\varepsilon$-edges existing in the edges $E_\zeta$ of the source argument graph), it computes the modified argument graph $G'_\zeta$. Then, we inductively carry out backward
evaluation on the argument expression $e_a$ to obtain another updated environment $\rho'_s$. This $\rho'_s$ and all other $\rho'_s$ are merged into $\rho'$.

Let us explain in more detail the definition of $\text{decomp}_{\text{rec}}$, which is the key point of the backward evaluation.

The function first extracts from result graph $G'$ nodes $V'_s$ and edges $E'_s$ that belong to each edge $\zeta$ by matching trace ID $\text{RecE}_{\zeta}$. Note that if there are nodes that have been freshly inserted into the view and we also require these nodes to have this structure, so that these nodes are also passed to the backward evaluation of the recursion body. Input and output nodes with marker $\text{sm}$ are recovered by selecting those pointed to from/to “hub” nodes having structure $\text{RecN}_{\zeta}$. Top-level constructors of trace ID $E_{\text{sm}}$ are erased so that we can inductively compute the backward image from the body expression.

Example 7. Recall the simple example in Example 3 where the source is $s = \{(\text{a}, \text{b}), (\text{b}, \text{c})\}$, and $\text{a2d}_{\text{xc}}(s)$ gives the graph $G$. If the graph $G$ is modified to $G'$ where the edge label $l$ is updated to $l'$, then $\text{B}(\text{a2d}_{\text{xc}})(\{\text{s} \mapsto s\}, G')$ returns binding $\{\text{s} \mapsto s'\}$ where $s' = \{(\text{a}, \text{b}), (\text{b}, \text{c})\}$. Therefore, the in-place update of the change on the view graph is reflected to the source.

Lemma 2 (Well-behavedness for In-Place Updates). If output graphs are modified by in-place updates on edges, then for any expression $e$, the two evaluations $\mathcal{F}[e]$ and $\mathcal{B}[e]_{\text{inc}}$ form a well-behaved bidirectional transformation, if they succeed.

Proof. This statement can be proved by induction on the structure of $e$. For the base case where $e$ is a variable, it clearly holds. Considering the inductive case, (1) if $e$ is a constructor expression, it holds because each constructor is reentrant within our context, (2) if $e$ is a condition, its backward evaluation is reduced to that on either its true branch or its false branch, so the statement holds by induction, and (3) if $e$ is a structural recursion, by bulk semantics, its backward computation is reduced to its body expression, so the statement holds by induction.

5.2 Reflection of Deletion

Deletion in a view is reflected as deletion of the corresponding part by the source by trace IDs. Suppose we want to delete the edge labeled $d$ in the view of Example 7. Since both endpoints of the edge have trace IDs of the form $\text{RecE}_{\zeta}$, we can see that the selected edge has been generated due to the existence of the source edge $(1, a, 2)$, which is the "corresponding part" to be deleted in the source.

In general, for a labeled edge $\zeta = (u, l, v)$ with $l \neq e$, its corresponding edge $\text{corr}(\zeta)$ is defined as:

\[
\text{corr}((u, l, v)) = (u, l, v) \quad \text{if } u, v \in \text{SrcID}
\]

\[
\text{corr}(\text{RecE}_{\rho \ z, l} \ c) = \begin{cases} \text{corr}((u, l, v)) & \text{if } \text{corr}((u, l, v)) \neq \text{Fail} \\ \text{corr}(\zeta) & \text{if } \text{corr}((u, l, v)) = \text{Fail} \\ \text{Fail} & \text{otherwise.} \end{cases}
\]

Here, \text{Fail} means failure on finding the corresponding edge. The first case means that if the edge $\zeta$ is a copy of an edge in the source, then $\zeta$ itself is the corresponding edge. The second and the third cases are for when $\zeta$ is a result of some structural recursion. According to the forward semantics of $\text{rec}$ in Figure 6, the non-$e$ edge $\zeta$ must have the form $\text{RecE}_{\rho \ z, l} \ c$ for some $p, u, v, \text{and another non}-e \text{ edge } \zeta'$. This means that $\zeta$ consists of an edge $(u', l, v')$ originating from an evaluation of a recursion-body at $\zeta'$. Hence, for this case, we first recursively trace the corresponding source of $(u, l, v)$, and if this fails, then try that of $\zeta'$. In other cases, \text{corr} fails to find the corresponding source, because it must be the case that $u$ has a trace ID of the form $\text{Code}_{\rho}$ meaning that the edge is not derived from the source but from an \text{UnCAL} expression.

Let $\text{db}$ be the source graph, $G_{\text{view}}$ be the view produced by $\mathcal{F}[e]_{\rho}$ from a forward computation of expression $e$ with environment $\rho$, and $G_{\text{view}}$ be a graph from $G_{\text{view}}$ with a set of edges $D_{\text{out}} = \{\zeta_1, \ldots, \zeta_n\}$ removed. Our backward evaluation $\mathcal{B}[e]_{\rho}(\zeta, G_{\text{view}})$ consists of the following three steps.

1. Compute the set of source edges $D_{\text{in}} = \{\text{corr}(\zeta) \mid \zeta \text{ is not an } e\text{-edge}\}$.

2. If $\text{Fail} \in D_{\text{in}}$, backward evaluation fails. If it is obtained successfully without failure, compute $G_{\text{src}}' = \rho[\text{db} \leftarrow D_{\text{in}}]$. Here, $G - E$ denotes removal of the edges in the set $E$ in graph $G$.

3. Return $\rho' = \rho[\text{db} \leftarrow G_{\text{src}}']$ as the result if $\mathcal{F}[e]_{\rho'} = G_{\text{view}}$, and fail otherwise.

Lemma 3 (Well-behavedness for Deletion). If output graphs are modified by edge deletion, then for any expression $e$, the two evaluations $\mathcal{F}[e]_{\text{dec}}$ and $\mathcal{B}[e]_{\text{dec}}$ form a well-behaved bidirectional transformation, if they succeed.

Proof. The \text{GetPut} property is clear because of the fact that $D_{\text{in}} = \emptyset$ if $D_{\text{out}} = \emptyset$. For the \text{WPutGet} property, it holds because the third step actually does this check.
5.3 Reflection of Insertion

Reflection of insertion is much more complicated than that of inplace-updating and deletion. This is because there are no corresponding edges in the source for the freshly inserted edges in the view, which requires us not only to create new information but also to add it to a proper location in the source graph.

Our idea was to use the Universal Resolving Algorithm (URA) (Abramov and Glück 2002), a powerful method of inversion computation, to derive a right inverse of the forward evaluation, and use the distributive property of structural recursion

\[ \text{rec}(e)(S_{g1} \cup S_{g2}) = \text{rec}(e)(S_{g1}) \cup \text{rec}(e)(S_{g2}) \]

to properly reflect insertion to the source.

In this section, we shall give our algorithm for this reflection, before we highlight how URA can be used to derive the right inverse.

5.3.1 Insertion Reflection with Right Inverse

We assume the monotonicity of insertion in that an insertion on the view is translated to an insertion on the source rather than other updating operations. The monotonicity comes from the absence of isEmpty (Buneman et al. 2000) in our core UnCAL. We only consider insertion on the view graph produced by forward computation of a variable expression or a structural recursion. For the case of a variable, this reflection is done in the same way as in Section 5.1.1. Insertion for structural recursion, the basic computation unit in UnCAL, needs to be carefully designed. In the following, we will focus on structural recursion, omitting other cases for simplicity.

Before giving our reflection algorithm, we should clarify the meaning of right inverse. In general, a function \( h \) is said to be a right inverse computation if it returns \( G \) within our context, for an expression \( e \) allows similar treatment for every node.

\[ F[\text{rec}(e)(\$db)](\rho')_2 = \{ \text{def. of } \rho' \} \]

\[ F[\text{rec}(e)(\$db)](\rho' \cup \rho_1(\$db))_2 \]

\[ F[\text{rec}(e)(\$db)](\rho_2) \]

\[ \{ \text{partial application} \} \]

\[ \{ \text{structural recursion property} \} \]

\[ \{ \text{forward evaluation} \} \]

\[ \{ e \text{ does not contain free variable} \} \]

\[ G_{\text{view}} \cup F[\text{rec}(e)(\$db)](\rho_1) \]

\[ G_{\text{view}} \cup G' \]

**Proof.** First, the (GETPUT) property clearly holds because \( \rho \) is returned when no insertions occur. Next, we prove the (WPGET) property by using the following calculation.

\[ F[\text{rec}(e)(\$db)](\rho')_2 = \{ \text{partial application} \} \]

\[ F[\text{rec}(e)(\$db)](\rho')_2 = \{ \text{def. of } \rho' \} \]

\[ F[\text{rec}(e)(G_{\text{src}} \cup \rho_1(\$db))](\rho')_2 \]

\[ \{ \text{structural recursion property} \} \]

\[ F[\text{rec}(e)(G_{\text{src}} \cup \text{rec}(e)(\rho_1(\$db)))](\rho')_2 \]

\[ \{ \text{forward evaluation} \} \]

\[ F[\text{rec}(e)(G_{\text{src}})](\rho'_2) \]

\[ \{ \text{e does not contain free variable} \} \]

\[ G_{\text{view}} \cup F[\text{rec}(e)(\$db)](\rho_1) \]

\[ G_{\text{view}} \cup G' \]

\[ \square \]

It is worth noting that we have simplified our discussion in both the above algorithm and lemma by making it a requirement that \( e \) in \( \text{rec}(e)(\$db) \) does not contain any free variables. With this requirement, our forward and backward evaluation satisfies the stronger (PUTGET) property. In fact, it is acceptable to relax this condition by allowing \( e \) to contain other free variables and the initial \( \rho \) contains binding of other variables. Then, right inversion will produce \( \rho' \) that will be used to update all variable bindings in addition to \$db. In this case, \( F[\text{rec}(e)(G_{\text{src}})](\rho')_2 \) may produce a graph that is different from the original view \( G_{\text{view}} \). In any case, this different graph will not have an additional effect on the source when we apply backward evaluation to this new graph. Therefore, (WPGET) always holds.

With this idea, we shall propose an algorithm in which (PUTGET) property is satisfied without any additional requirements. The idea is to utilize the Trace ID information, as will be discussed later.

5.3.2 Improving Insertion Reflection

The method above satisfies the (PUTGET) property only if the variables of \( e \) are disjoint from the variables bound in the initial environment \( \rho \). However, in general, since a transformation may have multiple variable references, more effort is required to achieve the (PUTGET) property. We tackle the problem by first locating where we insert a graph by using trace IDs, and then applying the URA algorithm (to be described later) to find what graph should be inserted.

Consider the transformation \( a2d_{xc} \) and the view in Example 6. Suppose we want to insert a graph \( G_{\text{sins}} \) rooted at the view node \( v = \text{RecN 7 2 } \). Where should some graph be inserted into the source to reflect this insertion? The answer is that we must insert a graph rooted at the source node 2 because there would be no edge from \( v \) in the view unless there were an edge from 2 in the source according to the bulk semantics of structural recursion. Now, our next task is to find what graph should be inserted under the source node 2. That is, we hope to find \( G_{\text{sins}} \) such that the following holds.

\[ a2d_{xc}(1 \rightarrow 2) \rightleftharpoons G_{\text{sins}} \]

\[ \text{RecN 7 1 } \]

\[ \text{RecE 7 (Code 4) (1, b, 2) } \]

\[ \text{RecE 7 (Code 3) (1, b, 2) } \]

\[ \text{RecE 7 (Code 1) (1, a, 2) } \]

\[ \text{RecN 7 2 } \]
URA can help us to find such $G_{\text{ sins}}$ for $G_{\text{ vins}}$. For example, if $G_{\text{ vins}}$ is $\{b : \{\}\}$, then URA returns $G_{\text{ sins}} = \{b : \{\}\}$. If $G_{\text{ vins}}$ is $\{d : \{\}\}$, then URA returns one of the possibilities, $G_{\text{ sins}} = \{a : \{\}\}$ or $G_{\text{ sins}} = \{d : \{\}\}$, depending on the search method used in URA. According to the soundness and the completeness of URA, the reflection by URA is always correct in the sense that (PUTGET) holds, and moreover URA always returns a $G_{\text{ sins}}$ if such $G_{\text{ sins}}$ exists. Of these, soundness is the key to insertion-reflection satisfying (PUTGET) for general UnCAL transformations.

In summary, our insertion-reflection algorithm is as follows.

1. Let $v$ be a node under which we want to insert a graph $G_{\text{ vins}}$.
2. By using the $tr$ function in Figure 8, we find the source node $u = tr(v)$ under which we insert a graph to reflect the insertion.
3. Let $G_{\text{ vview}}$ be a graph obtained from the view by adding $\varepsilon$-edge from $v$ to $G_{\text{ vins}}$.
4. We find a graph $G_{\text{ sins}}$ connected from $u$ by an $\varepsilon$-edge, by applying URA for $G_{\text{ vview}}$.
5. We return a graph $G_{\text{ src}}$ obtained from the source by adding an $\varepsilon$-edge from $u$ to $G_{\text{ sins}}$.

The soundness of the insertion-reflection algorithm is directly derived from the soundness of URA.

Lemma 5 (Soundness of Insertion). Our insertion-reflection algorithm satisfies (PUTGET).

Note that we use URA for $G_{\text{ vview}}$ instead of $G_{\text{ vins}}$. Thus, URA rejects any insertion of $G_{\text{ sins}}$ that violates (PUTGET).

In addition, our insertion-reflection algorithm is complete in the sense that, if there exist some source insertions to reflect the view insertion under some conditions, the algorithm will find one of them.

Lemma 6 (Completeness of Insertion). Let $v$ be a node such that $tr(v) \neq \text{FAIL}$. For any source graph $G$, we can insert any graph into its view if there exists a source insertion that reflects the view insertion and $v$ still occurs in the view of the insertion-reflected source.

Recall that we only consider insertion on the view graph produced by forward computation of a variable expression or a structural recursion, which is expressed by $tr(v) \neq \text{FAIL}$. This lemma can be proved using the property of trace IDs stating that, to insert a graph rooted at view node $v$, we must insert a graph rooted at source node $tr(v)$. By induction on the trace ID of $v$, we can show that, if there is an edge from $v$, it must be the case that there is an edge from $(tr(v)$, which is implied by the property of trace IDs. Note that $G_{\text{ vins}}$ has no edge to the original view. However, this is not a restriction since if there is a crossing edge pointing to a subgraph of the original view, we can duplicate the subgraph and integrate it to $G_{\text{ vins}}$ so that the edge can be eliminated.

5.3.3 Right Inverse Computation by URA

Recall that the right inverse computation of an expression $e$ is to take a graph $G_{\text{ vview}}$ and return a $\rho$ such that $F[e][\rho] = G_{\text{ vview}}$. We adopt the universal resolving algorithm (URA) (Abravam and Glück 2002), a powerful and general inversion mechanism, to compute $\rho$. The basic idea behind URA is to search on a perfect process tree (Glück and Klimov 1993), which represents all possible computations of an expression, and to find a computation path that produced the result.

Our right inverse computation consists of three steps.

1. It lazily enumerates possible evaluation paths by symbolic computation called needed narrowing (Antoy et al. 1994)††.
2. From the generated evaluation paths, it constructs a table of input/output pairs of computations.
3. If there is a pair in the table whose output is $G_{\text{ view}}$, it generates a substitution (environment) from the path and returns it as the result.

Example 8. As a simple example, let us see how we find $\rho$ such that

$$F[a2d_xc(x)\rho] = G_{\text{ view}}$$

where $G_{\text{ view}} = \{d : \{\}\}$. We search $\rho$ by symbolic evaluation of $a2d_xc(x)$. To evaluate $a2d_xc(x)$, we unfold $x$ and recursively evaluate $a2d_xc$, i.e., a structural recursion. There are many ways to instantiate $x$ such as

$$x \mapsto \{\}, x \mapsto \{(l_1 : x_1), x \mapsto \{(l_1 : x_1, l_2 : x_2)\}$$

If we choose $x \mapsto \{\}$, the computation finishes, yielding a table consisting of an input/output pair $(\{\}, \{\})$. Since this table does not contain a pair whose output is $G_{\text{ view}}$, we continue searching. Assume that we choose $x \mapsto \{(l_1 : x_1)\}$. Then $a2d_xc(x)$ is unfolded to (if $l_1 = a$ then $\{d : x\}$ else if $l_1 = c$ then $\{x : x\}$ else $(l_1 : x))$ @ $a2d_xc(x_1)$. As evaluation gets stuck here because of a free variable $l_1$ in the if condition, we find a suitable $l_1$ to resume the evaluation. If we choose $l_1 \mapsto a$, then the expression is reduced to $\{d : x\} @ a2d_xc(x_1)$ and input/output pair $(\{a : \{\}\}, \{d : \{\}\})$ is obtained by choosing $x_1 \mapsto \{\}$. Since $G_{\text{ view}} = \{d : \{\}\}$, we gather all bindings along this computation and return the following environment as the result.

$$\{x \mapsto \{a : \{\}\}\}$$

Figure 9 shows part of a perfect process tree in our right-inverse computation: the left is the tree and the right is a table of a pair of input/output graph templates (it is more general than a pair of input/output graph instances, as we discussed above). Note this tree is a variant of SLD-resolution trees (Gluck and Sorensen 1994).

To use URA effectively for our right inverse computation of UnCAL, we define a small-step semantics for UnCAL such that a perfect process tree can be constructed though these small steps. The only non-standard feature of this semantics is that we use memoization to avoid infinite loops probably caused by cycles in the source graph (See Appendix of (Hidaka et al. 2010) for details).

In addition, we provide a Dijkstra-searching strategy to enumerate all the possible evaluation paths so that a solution can always be found if one exists. The two heuristics we use to design the cost function are:

- We use a (weighted) size of graphs (to be inserted into the source) as a cost function in the Dijkstra-search.
- For the weighted size, the depth (the length of the path) has more weight than the width (the number of paths). This strategy works nicely for consecutive in Example 4.

Moreover, we show that a suitable binding to continue evaluation of conditional expressions can easily be found for our core UnCAL, because the conditional part of a conditional expression is in the simple form of $l_1 = l_2$.

†† The same notion is called driving (Glück and Klimov 1993; Glück and Sorensen 1994) in (Abravam and Glück 2002).
6. Implementation and Experiments

The prototype system has been implemented and is available on our BiG project Website. In addition to all the examples in Buneman et al. (2000) and in this paper, we have tested three non-trivial examples, demonstrating its usefulness in software engineering and database management.

- **Customer2Order**: A case study in the textbook on model-driven software development (Pastor and Molina 2007).
- **PIM2PSM**: A typical example of transforming a platform independent object model to a platform specific object model.
- **Class2RDB**: A non-trivial benchmark application for testing the power of model transformation languages (Bezivin et al. 2005).

All of these have demonstrated the effectiveness of our approach in practical applications.

In our implementation, we carefully treat \(\varepsilon\)-edges introduced during operations related to markers, and retrieval of edges or nodes of interest, which greatly affect performance. Poor treatment would hinder large-scale UnQL queries to evaluate in bidirectional mode\(^{11}\) in a reasonable amount of time. Speed-up of several orders of magnitude has been achieved since our initial implementation due to the above and the following optimizations.

**Reduction in number of \(\varepsilon\)-edges** As mentioned in the UnQL paper (Buneman et al. 2000), \(\varepsilon\)-edges are generously generated during evaluation, especially in rec. This slows the evaluation process due to the increase in input size. Removing \(\varepsilon\)-edges during evaluation has no harm on forward semantics because of bisimulation equivalence. However, since \(\varepsilon\)-edges play an important role in backward evaluation, they are not freely omitted in our bidirectional settings. Moreover, a straightforward implementation of the removal algorithm (Buneman et al. 2000) may introduce additional edges, which may harm backward evaluation. Toward prudently removing \(\varepsilon\)-edges that are suitable for backward evaluation, our \(\varepsilon\)-removal algorithm glues source and destination nodes of \(\varepsilon\) as long as bisimulation equivalence is not violated.

**Optimization by fusion transformation** Note that the backward evaluation of \(\text{rec}(e_1) \circ \text{rec}(e_2)(e_3)\), a composition of structural recursions, requires to generate intermediate result of backward transformation, which is very expensive. This can be avoided by fusing the two structural recursions into one. We have implemented this based on the fusion rule (Buneman et al. 2000): if \(e_1(t, G)\) does not depend on \(G\) then \(\text{rec}(e_1)(\text{rec}(e_2)(e_3)) = \text{rec}(\text{rec}(e_1) \circ e_2)(e_3)\). With auxiliary rewriting rules such as \(e_1 \cdot e_2 = e_1\) for \(e_2\) that produces no output nodes, 30\% and 50\% of CPU time reductions are respectively achieved for forward and backward execution in Customer2Order composed with selection, 30\% and 65\% reductions for simpler examples that appeared in the evaluation for unidirectional transformation (Hidaka et al. 2009). These experiments are for in-place updates, but similar reduction could be achieved for other updates.

7. Related Work

Bidirectional transformation has been discussed as view updating problem in the database community. Bancilhon and Spyrota (1981) proposed a general approach to the view updating problem. They introduced an elegant solution based on the concept of a constant complement view that captures the information in the view but not in the original database. Their idea was not only applied to relational databases (Hegner 1990; Lechtenbörger and Vossen 2003) but also to tree structures (Matsuda et al. 2007). Constant complement views satisfy very strong bidirectional properties at the sacrifice of the number of reflectable updates. Although such strong properties are nice for some applications (Hegner 1990), they are too strong for our purpose, i.e., model transformation in software engineering. Recent work by Fegaras (2010) propagates updates on XML views created from relational databases. It supports duplicates but detects view side effects at both compile and run time.

In the area of programming languages, view updating has been studied as bidirectional transformation. Foster et al. (2005) proposed the first linguistic approach to solving this problem. They developed some domain specific languages to support the development of bidirectional transformation on strings and trees. Bohannon et al. (2006) applied these techniques to relational databases, making use of functional dependencies in relations to correctly propagate updates. However, their approach is limited to strings, trees and relations, and is difficult to apply to graph transformation due to graph-specific features such as circularity and sharing.

Within the context of software engineering, there has been several works on bidirectional model (graph) transformation (Ehrig et al. 2005; Jouault and Kurtev 2005; OMG 2005; Schürr and Klär 2008; Stevens 2007), which can deal with kinds of graph structures. However, they lack a clear formal bidirectional semantics and there has not yet been any powerful method of bidirectionalization that can be used to automatically derive backward model transformations from forward model transformations, so that both transformations can form a consistent bidirectional model transformation.

The concept of structural recursion is not new and has been studied in both the database (Breazu-Tannen et al. 1991) and the functional programming communities (Sheard and Fegaras 1993). However, most of these have focused on structural recursion over lists or trees instead of graphs. Examples include the higher order function \textit{fold} (Sheard and Fegaras 1993) in ML and Haskell, and the generic computation pattern called \textit{catamorphism} in programming algebras (Bird and de Moor 1996). UnCAL (Buneman et al. 2000)

\(^{11}\)Note that we preserve every result of forward computation in the bidirectional mode.
demonstrates that the idea of structural recursion can be extended to graphs, but the original focus was on the optimization of query fusion rather than bidirectionalization.

Our work was greatly inspired by interesting work on efficient graph querying (Buneman et al. 2000; Sheng et al. 1999). Unlike trees, graphs involve subtle issues on their representation and equivalence. The use of bisimulation and structural recursion in (Buneman et al. 2000) opens a new way of building a framework for both declarative and efficient graph querying with high modularity and composability. This motivated us to extend the framework from graph querying to graph transformation and apply it to model transformation (Hidaka et al. 2009). This work is a further step in this direction to extend it from unidirectional model transformation to bidirectional model transformation.

8. Concluding Remarks

This paper reports our first attempt toward solving the challenging problem of bidirectional transformation on graphs. We show that structural recursion on graphs and its unique bulk semantics play an important role not only in query optimization, which has been recognized in the database community, but also in automatic derivation of backward evaluation, which has not been recognized thus far. As far as we are aware, the bidirectional semantics of UnCAL proposed in this paper is the first complete language-based framework for general graph transformations.

Future work includes extending the framework from unordered graphs to ordered graphs, introducing graph schemas to provide structural information for more efficient bidirectional computation, an efficient algorithm for checking updatability, and more practical applications of the system for bidirectional model transformation in software engineering.

Acknowledgments

We thank Mary Fernandez who kindly provided us with the SML source codes of an UnQL system. We thank Fritz Henglein and James Cheney, and anonymous reviewers for their thorough comments on earlier versions of the paper. The research was supported in part by the Grand-Challenging Project on “Linguistic Foundation for Bidirectional Model Transformation” from the National Institute of Informatics, Grant-in-Aid for Scientific Research No. 22300012, No. 20500043, and No. 20700035.

References


