## Specification with Graph Rewriting Systems

## Characterization

- Formal specification of abstract data types
- Model-oriented specification
- Graphs as underlying data model
- Specification of read operations by graph tests, specification of write operations by graph rewrite rules
- Proof technique: induction
- Rapid Prototyping by generating code from the specification


## Specification of Software Systems

## Introductory Example

## Specification of Software Systems

## Representation of a list as a graph



## Example of a graph rewrite rule




## Example of a graph test



## Interface of the abstract data type List

| List creation and deletion |  |
| :--- | :--- |
| CreateList | Creation of a list |
| DeleteList | Deletion of a list |
| Write operations | Insertion into an empty list |
| InsertFirstElement | Insertion before an element |
| PreInsertElement | Insertion after an element |
| PostInsertElement | Deletion of an element |
| DeleteElement |  |
| Read operations | List empty? |
| IsEmpty | First element |
| GetFirstElement | Last element |
| GetLastElement | Next element |
| GetNextElement | Previous element |
| GetPreviousElement | Data of current element |
| GetData |  |

## Specification of Software Systems

## Cases of PreInsertElement



PreInsertLastElement

## Example of programming with graph rewrite rules



## Theoretical Foundations

## Graphs

## Definition: Directed, labeled graph

$G=(V, E, I)$ is a directed graph over label sets $L_{V}$ (labels for vertices) and $L_{E}$ (labels for edges) $\Leftrightarrow$
" $V$ is a set of nodes (node identifiers).
" $E \subseteq V \times L_{E} \times V$ is a set of labeled edges.
» I:V $\rightarrow L_{V}$ is a labeling function for nodes.

## Remarks:

- Nodes have identifiers, but not edges.
- Thus, there are no parallel edges with the same labels.
- Edges are binary relationships.
- Nodes and edges are typed (labeled).
- So far, neither nodes nor edges are attributed.


## Specification of Software Systems

## Example of a directed graph


$G=(V, E, I)$ with:

- $V=\{1,2,3,4,5\}$
- $E=\{(1$, First, 2), (1, Elem, 3), (1, Elem, 4), (1, Last, 5),
(2, Next, 3), (3, Next, 4), (4, Next, 5) \}
- $I=\{(1$, List $),(2$, Element), (3, Element), (4, Element), (5, Element) $\}$


## Partial graphs and subgraphs

## Definition: Partial graph

$G=(V, E, I)$ is a partial graph of $G^{\prime}=\left(V^{\prime}, E^{\prime}, I^{\prime}\right) \Leftrightarrow$
» $V \subseteq V^{\prime}$, i.e., the nodes of $G$ are also contained in $G^{\prime}$.
» $E \subseteq E^{\prime}$, i.e., the edges of $G$ are contained in $G^{\prime}$.
" $\left.I^{\prime}\right|_{v}=l$, i.e., the nodes of $G$ have the same labels in $G^{\prime}$.

## Definition: Subgraph

$G=(V, E, I)$ is a subgraph of $G^{\prime}=\left(V^{\prime}, E^{\prime}, I^{\prime}\right) \Leftrightarrow$
» $G$ is a partial graph of $G^{\prime}$.
» $E=\left\{\left(v_{1}, e l, v_{2}\right) \in E^{\prime} \mid v_{1}, v_{2} \in V\right\}$
$G$ contains all edges of $G^{\prime}$ whose sources and targets are common to $G$ and $G^{\prime}$.

## Example of partial graphs and subgraphs

Graph

$\qquad$


Subgraph


## Graph morphisms

## Definition: Graph morphism

A function $h: V \rightarrow V^{\prime}$ is a graph morphism from $G$ to $G^{\prime}$, i.e., $h: G \rightarrow G^{\prime} \Leftrightarrow$
» $\forall v \in V: l^{\prime}(h(v))=I(v)$, i.e., labels are preserved.
» $\forall\left(v_{1}, e l, v_{2}\right) \in E:\left(h\left(v_{1}\right), e l, h\left(v_{2}\right)\right) \in E^{\prime}$, i.e., edges are preserved.

## Definition: Graph isomorphism

A graph morphism $h: G \rightarrow G^{\prime}$ is a graph isomorphism $\Leftrightarrow$
" $h: V \rightarrow V^{\prime}$ is injective and surjective
» $h: V \rightarrow V^{\prime}$ induces a function $h^{\prime}: E \rightarrow E^{\prime}$, which must be injective and surjective, as well.

## Set-theoretic graph operations (1)

## Definition: Union of graphs

Let $G$ and $G^{\prime}$ be directed graphs, $\|_{\vee \cap v^{\prime}}=\left.I^{\prime}\right|_{V \cap v}$ :
» $G \cup G^{\prime}=G^{\prime \prime}=\left(V^{\prime \prime}, E^{\prime \prime}, I^{\prime \prime}\right)$ with
$\Rightarrow V^{\prime \prime}=V \cup V^{\prime}$
$\Rightarrow \forall v \in V^{\prime \prime}: l^{\prime \prime}(v)=\underline{\text { if }} v \in V$ then $I(v)$ else $I^{\prime}(v)$ end
$\Rightarrow E^{\prime \prime}=E \cup E^{\prime}$
» $G \oplus G^{\prime}$ disjoint union of $G$ and $G^{\prime}$ :
Rename nodes of $G$ or $G^{\prime}$ such that $V \cap V^{\prime}=\varnothing$, and apply the graph union defined above:
$G \oplus G^{\prime}=G \cup G^{\prime}$.

## Set-theoretic graph operations (2)

## Definition: Difference of graphs

Let $G$ and $G^{\prime}$ be directed graphs, $\left.I\right|_{v \cap v}=\left.I^{\prime}\right|_{v \cap v^{\prime}}$ :
» $G \backslash G^{\prime}=G^{\prime \prime}=\left(V^{\prime \prime}, E^{\prime \prime}, l^{\prime \prime}\right)$ with
$\Rightarrow V^{\prime \prime}=V \backslash V^{\prime}$
$\Rightarrow l^{\prime \prime}=\| V^{\prime}$
$\Rightarrow E^{\prime \prime}=E \backslash E^{\prime}$ (without deletion of dangling edges)
» $G \backslash G^{\prime}=G^{\prime \prime}=\left(V^{\prime \prime}, E^{\prime \prime}, I^{\prime \prime}\right)$ with
$\Rightarrow V^{\prime \prime}=V \backslash V^{\prime}$
$\Rightarrow I^{\prime \prime}=\| V^{\prime}$
$\Rightarrow E^{\prime \prime}=\left(E \backslash E^{\prime}\right) \cap\left(V^{\prime \prime} \times L_{E} \times V^{\prime \prime}\right)=E \cap\left(V^{\prime \prime} \times L_{E} \times V^{\prime \prime}\right)$
(with deletion of dangling edges)

## Set-theoretic graph operations (3)

## Definition: Intersection of graphs

Let $G$ and $G^{\prime}$ be directed graphs, $\left.I\right|_{V \cap V^{\prime}}=\left.I^{\prime}\right|_{V \cap v^{\prime}}$ :
» $G \cap G^{\prime}=G^{\prime \prime}=\left(V^{\prime \prime}, E^{\prime \prime}, I^{\prime \prime}\right)$ with
$\Rightarrow V^{\prime \prime}=V \cap V^{\prime}$
$\Rightarrow l^{\prime \prime}=\| V^{\prime \prime}$
$\Rightarrow E^{\prime \prime}=E \cap E^{\prime}$

## Graph rewrite rules

## Definition: Graph rewrite rule

A graph rewrite rule is a triple $r=(L, K, R)$ with:
» $L$, the left-hand side of $r$, is a graph.
» $R$, the right-hand side of $r$, is a graph.
» $K=L \cap R$ is the gluing graph.

## Remarks

- Elements of $L$ but not of $K$ are deleted.
- Elements of $R$ but not of $K$ are inserted.
- Elements of $K$ are preserved.
- $K$ is called gluing graph because it is used for the embedding of new nodes of $R$.


## Example of a graph rewrite rule (1)

production DeleteElement $=$


## Example of a graph rewrite rule (2)



## Application of a graph rewrite rule

(Direct) derivation
A graph $G^{\prime}$ is derivable from a graph $G$ by a rule $r=(L, K, R) \Leftrightarrow$
» There is an isomorphism $h: L \rightarrow G_{L}$, where $G_{L}$ is a partial graph of $G$ which determines the location of application of $r$.
» Nodes and edges of $G_{\llcorner }$not appearing as images of $K$ are deleted:
$H=G \backslash(h(L) \backslash h(K))$
» Nodes and edges of $R$ which do not belong to $K$ are inserted:
$G^{\prime}=H \oplus h^{\prime}(R \backslash K)$, where
$\Rightarrow h^{\prime}$ maps nodes of $R \backslash K$ such that they do not occur in $H$,
$\Rightarrow$ Context edges with sources or targets from $K$ are transferred to the respective nodes in $h(K)$.

## Other variants of graph rewrite rules

- $h$ need only be a morphism $\Rightarrow$
» Higher flexibility by identification of graph elements
» Danger of undesired effects of rule applications
- $G_{L}$ must even be a subgraph $G \Rightarrow$
» Larger left-hand sides
» Larger rule sets
- No edges from deleted nodes to context nodes (Dangling Edge Condition) $\Rightarrow$
» Exclusion of undesired side effects
» Large left-hand sides
- Empty gluing graph $\Rightarrow$
» Embedding of nodes of the right-hand side must be specified explicitly by embedding rules


## Example of the application of a graph rewrite rule



Host graph



## Graph rewriting systems

## Definition: Graph rewriting system

A graph rewriting system is a tuple $g s=\left(L_{V}, L_{E}, R, S\right)$ with:
» $L_{V}$ : finite set of node labels
» $L_{E}$ : finite set of edge labels
» $R$ : finite set of rules $r=(L, K, R)\left(L, K, R\right.$ graphs over $\left.L_{V}, L_{E}\right)$
" $S$ : start graph (over $L_{V}$ and $L_{E}$ )

## Definition: Derivability

Let $g s=\left(L_{V}, L_{E}, R, S\right)$ be a graph rewriting system. A graph $G$ is derivable from the start graph $S$ using the rule set $R g s \Leftrightarrow$
There is a sequence $G_{1} \ldots G_{n}$ with

$$
S \rightarrow G_{1} \ldots \rightarrow G_{n}=G
$$

## Proof techniques

## Proof by induction

A predicate $p$ is proved for all derivable graphs as follows ( $n$ : length of derivation):
$1 n=0: p$ holds for the start graph $S$.
$2 n \rightarrow n+1$ : Let $G$ be a graph which may be derived in $n$ steps from the start graph.
Induction assumption: $p$ holds for $G$.
Induction conclusion: $p$ holds for all graphs $G^{\prime}$ which may be derived from $G$ by some rule $r$ (in one derivation step). (Induction conclusion has to be proved for all rules $r$ and all potential locations of application.)

## Examples of (provable) properties of list graphs

- Each list has at most one First element.
- Each list has at most one Last element.
- The First element does not have a predecessor.
- The Last element does not have a successor.
- Each element has at most one predecessor.
- Each element has at most one successor.
- ...


## Specification with PROGRES

## What is PROGRES?

- PROGRES = PROgrammed GRaph REwriting Systems
- Multiparadigmatic Specification language, based on graph rewriting
» Object-oriented modeling of graph schemata
» Declarative definition and incremental evaluation of derived attributes
» Rule-based and visual specification of graph tests and graph transformations
» Imperative and non-deterministic programming
- Development environment for the construction of specifications
» Syntax-aided editor
» Static analyses
» Interpreter
» Code generator


## Components of PROGRES



## Example

Tools for programming-in-the-large


## Explanations

- There are three types of modules (fm, ado, adt).
- The function module (fm) Main exports a set of functions; it has exactly one body variant.
- This variant imports the modules UserInterface and Files.
- The data object module (ado) UserInterface realizes a data store and its access operations (in exactly one variant).
- The interface of module UserInterface imports from the module Files the type File.
- The data type module (adt) Files exports a data type with operations for creating and manipulating an arbitrary number of instances; it has two variants.
- Properties of variants are defined through a set of attributes (ws for Window System and OS for Operating System).
- Attributes either have concrete values (like UNIX or MSDOS) or are undefined.


## Desired tool functionality

- Syntax-aided editing of module graphs (context free correctness is enforced, e.g. variant import do not emanate from module interfaces)
- Checking of context sensitive constraints (e.g., no import cycles)
- Completely automatic or semi-automatic configuration of system variants satisfying given properties (Attention: Requires consistent selection of variants. In the example, there is only one consistent configuration for $\mathrm{WS}=\mathrm{x}$ and $\mathrm{OS}=$ UNIX.)


## Attributed Graphs and Graph Schemata




## PROGRES schema editor



## Definition of attributes

a A graph schema may be specified both textually and graphically.

- Attributes and attribute evaluation rules may be defined in a textual (sub-)view.
- Intrinsic attributes receive their values by explicit assignment and may be initialized with a constant.
- Derived attributes are calculated from the attribute values of related nodes with the help of a (directed) equation.
- Related nodes are all nodes which may be reached via a path, e.g.: » self: Returns the current node
» -e-> and <-e--, respectively : Traversal of edges of type e (in forward and backward direction, respectively)
» Concatenation:p1 \& p2



## Textual definition of a graph schema (2)

```
edge type has : SPECIFICATION [1:1] -> REALIZATION [0:n];
    (* A 'SPECIFICATION' 'has' an arbitrary number of *)
    (* 'REALIZATIONS', but a 'REALIZATION' belongs to a *)
    (* uniquely defined 'SPECIFICATION'. *)
node class COMPLEX is a UNIT
    redef derived
        Size = addSize( 0, all self.-contains-> );
            (* The 'Size' of a complex unit is the sum of *)
            (* the 'Size' of all its children. *)
end;
node class ATOM is a UNIT
    intrinsic
        File : file; (* Pointer to externally stored file. *)
    redef derived
        Size = size( self.File );
        (* 'Size' is the length of the attached File *) Edge type
end;
edge type contains : COMPLEX [1:1] -> ATOM [0:n];
    (* A 'COMPLEX' may contain 0 to n 'ATOM' nodes.
    (* Conversely, an 'ATOM' node is contained in exactly
    (* one'COMPLEX'. *)
```


## Textual definition of a graph schema (3)

```
node class SYSTEM is a SPECIFICATION, COMPLEX
    redef derived
        Size = addSize( 0, all self.-contains-> );
            (* Resolves inheritance conflict of attribute *)
            (* definitions in 'SPECIFICATION' and 'COMPLEX' *)
            (* by preferring definition in 'SPECIFICATION'. *)
end;
node class CONFIGURATION is a REALIZATION, COMPLEX end;
    (* A 'CONFIGURATION' is a set of variants of module *)
    (* realizations which fulfill all required properties. *)
node class MODULE is a SPECIFICATION, ATOM
    redef derived
        Size = size( self.File ) + max( 0, all self.-has->.Size );
            (* Resolves inheritance conflict of attribute *)
            (* definitions in 'SPECIFICATION' and 'COMPLEX' *)
            (* by building the sum of both definitions which *)
            (* are in conflict to each other. *)
end; (* A 'SYSTEM' contains a set of 'MODULES' which *)
    (* have 'VARIANTS' as their realizations. *)
node class VARIANT is a REALIZATION, ATOM end;
```


## Attribute redefinition conflicts



## Path expressions and restrictions

- A path expression is
" a derived relation between nodes (edges are intrinsic relations):
$v_{1}=p=>v_{2} \Leftrightarrow$ There is a path $p$ from $v_{1}$ to $v_{2}$
» a function on node sets:
$p(V)=\left\{v_{2} \mid \exists v_{1} \in V: v_{1}=p=>v_{2}\right\}$
- A restriction is a "unary path expression", i.e., a set of nodes is restricted to those elements which satisfy a certain condition.
- Path expressions and restrictions may be specified:
» textually
» graphically


## Specification of Software Systems

## Textual path expressions

| Operator | Semantics |
| :--- | :--- |
| $p \& q$ | Concatenation of $p$ and $q$ |
| $p$ or $q$ | Connection by $p$ or $q$ |
| $p$ and $q$ | Connection by $p$ and $q$ |
| $p$ but not $q$ | Connection by $p$ but not by $q$ |
| $[p \mid q]$ | Connection by $p$, else by $q$ |
| $p+$ | Transitive closure |
| $p^{*}$ | Reflexive and transitive closure |
| $\{p\}$ | Maximal iteration |

## Examples of textual path expressions

```
path needs : ATOM [0:n] -> MODULE [0:n] =
    (* The path 'needs' connects any variant or module to its imports. *)
        ( instance of MODULE & =moduleNeeds=> )
    or ( instance of VARIANT & =variantNeeds=> )
end;
path moduleNeeds : MODULE [0:n] -> MODULE [0:n] = -m_uses-> end;
path variantNeeds : VARIANT [0:n] -> MODULE [0:n] =
    -v_uses-> or ( <-has- & instance of MODULE & -m_uses-> )
end;
path dependsOn : MODULE [0:n] -> MODULE [0:n] =
    (* connects module to its interface imports & imports of its variants. *)
    ( self or -has-> ) & =needs=>+
end;
```


## Example of a graphical path expression



## Specification of Software Systems

## Graph Rewrite Rules

## Specification of Software Systems

## Composition of graph rewrite rules

```
production P ( parameter list ) =
    Left-hand side with:
    - single nodes, set nodes, negative/optional nodes
    - positive and negative edges between pairs of nodes
    - positive and negative paths between pairs of nodes
    - node restrictions
    ::=
Right-hand side with:
    - single nodes, set nodes, optional nodes
    - edges between pairs of nodes
fold Specified nodes may be identified;
condition Conditions on attributes of left-hand side nodes;
embedding Embedding rules for nodes of right-hand side;
transfer Attribute assignments for nodes of right-hand side;
return Assignments to out parameters;
end;
```


## Elements of left-hand sides and right-hand sides



## Example for parameters, negative nodes, attribute transfer, etc.

```
production CreateModule(MName : string; InterfaceDescription : file;
                            MType : type in MODULE; out NewM : MODULE) =
    |-----------------------------------------------
    ::=
    |
(* Creates a module with
name 'MName'if the 'System
                                does not already contain a
                                module with this name. The
                                'Mtype' parameter is a
                        'FunctionModule' or
                                'ADTModule' or 'ADOModule'.
    *)
    _ondition '2.Name = MName; (* Conditions only for normal nodes *)
end;
restriction name(UName : string) : UNIT = valid (self.Name = UName) end;
```


## Example of optional set nodes

production DeleteModule(Module : MODULE) =


## Example of a negative path



## Embedding rules

| Rule | Semantics |
| :---: | :---: |
| copy -e-> from `n to m' & Copy outgoing e edges from node ` n of the left-hand side to node $\mathrm{m}^{\prime}$ of the righthand side |  |
| remove -e-> from 'n | Delete outgoing e edges from node ` n |
| redirect -e-> from ' n to m ' | Redirect outgoing e edges from ' n to m ' |
| <-e- instead of -e-> | Analogously for incoming rather than outgoing edges |
| -e-> as -f-> instead of -e-> | Relabeling |
| -e-> as <-f- statt -e-> | ... and change of direction |

## Example of embedding rules



## PROGRES rule editor



## Example of a graph test

```
test UnresolvedImport(InConfig : Config; out MSet : MODULE [1:n]) =
```



```
end;
(* Returns all modules which are needed by some variant already
selected but for which no variant is included yet in the
configuration. *)
```


## Searching for subgraphs

- Complexity of naive implementation: $o\left(n^{k}\right)$ (for each of $k$ nodes in $L$ there are $n$ candidates in $G$ ).
- Heuristics (sketch):
» Start at those nodes which are fixed by input parameters.
» Extend the match by nodes which may be determined in a unique way (incoming or outgoing edges of cardinality 1 ).
» Process remaining candidate sets by increasing cardinality.
» Process set nodes at the end.


## Specification of Software Systems

## Control Structures

## Control structures: motivation and properties

- Composition of graph tests into complex queries (which do not modify the host graph)
- Composition of graph rewrite rules (and graph tests) into complex transactions
- ACID properties of transactions (and graph rewrite rules):
» Atomic: either complete execution or no modification of the host graph
» Consistent: consistency-preserving transformation
» Isolated: isolation in multi-user mode
» Durable: persistent
- Additional property: non-determinism
- Failure of executing an operation results in backtracking


## Overview of control structures

| Control structure | Semantics |
| :---: | :---: |
| $p \& q$ | Sequence |
| $p$ and $q$ | $p$ und $q$ in any order |
| $p$ or $q$ | Non-deterministic choice |
| choose p1 else p2 ... end | Try p1, else p2 ... |
| loop p end | Loop (execute p as long as possible) |
| for all n do p end | Execute p for all nodes n |
| use v : ... do p end | Block with declaration of variables |

## Example of a transaction with backtracking

```
transaction CreateConfig(CName : string; CProps : string [0:n]) =
    use ReqProps := CProps do
        InitConfig(CName, ReqProps, out ReqProps)
        & loop
            ResolveImport(CName, ReqProps, out ReqProps) (* see p. 56 *)
        end
        & not UnresolvedImportExists(CName) (* similarly to p. 57 *)
    end
end;
```


## Specification of Software Systems

## Rule for initializing a configuration

production InitConfig(CName : string; CPropsIn : string; out CPropsOut : string) =


$$
\text { condition } 3 . \text { Props in CPropsIn; }
$$

$$
\text { transfer } 4^{\prime} . \text { Name }:=\text { CName; }
$$

$$
\text { return CPropsOut }:=\text { merge(`3.Props, CPropsIn); }
$$

end;

## Example (1)



## Example (2)



## Example (3)

ResolveImport("C", \{\}, out ReqProps = \{OS:MSDOS\})


## Example (4)

- ResolveImport fails for the module UserInterface because \{OS:UNIX\} in $\{O S: M S D O S\}$ does not hold.
- Loop terminates successfully, but the subsequent test UnresolvedImportExists reveals an unresolved import.
- As a result of backtracking, the previous selection of the variant of Files is revised (slide 68 shows the situation after selection of another variant).
- Finally, a variant of UserInterface may be selected successfully (slide 69).


## Example (5)

## ResolveImport("C", \{\}, out ReqProps = \{OS:UNIX\})



## Specification of Software Systems

## Example (6)

ResolveImport("C", \{OS:UNIX\}, out ReqProps = \{WS : X , OS:UNIX\})


## Specification of Software Systems

## Summary

## Advantages of specifying with graph rewrite rules

- Graphs are an appropriate data model for a large set of applications.
- Even complex data structures with a high number of consistency constraints may be represented as graphs.
- Complex graph transformations may be specified declaratively by graph rewrite rules.
- Visual programs composed of graph rewrite rules are easy to comprehend.
- The specification is operational, code generation is supported (for rapid prototyping).


## Disadvantages of specifying with graph rewrite rules

- Generality is constrained: commitment to a specific data model.
- For simple data types, graphs and graph transformations are an "overkill".
- Potential efficiency problems (subgraph search is NP-complete).
- In case of PROGRES:
» Very expressive, but also very complex language.
» Specifying-in-the-large not completely elaborated.


## Literature

- G. Rozenberg (Ed.): Handbook of Graph Grammars and Computing by Graph Transformation, Vol. 1: Foundations, World Scientific (1997) Collection of papers on the theory of graph rewriting systems
- H. Ehrig et al. (Eds.): Handbook of Graph Grammars and Computing by Graph Transformation, Vol. 2: Applications, Languages and Tools, World Scientific (1999)
Applications of graph rewriting systems, including e.g. chapters on PROGRES and its applications
- A. Schürr: Operationales Spezifizieren mit programmierten Graphersetzungssystemen, Deutscher Universitätsverlag, Wiesbaden (1991)

PROGRES language definition (in German)

- A. Schürr, B. Westfechtel: Graphgrammatiken und

Graphersetzungssysteme, script for the corresponding lecture, Aachener Informatik-Berichte 92-15 (1992)
Survey of the most important approaches concerned with graph rewriting

