

# 2. FORMULA: Domains and Models

Ethan Jackson, Nikolaj Bjørner and Wolfram Schulte  
Research in Software Engineering (RiSE), Microsoft Research

Dirk Seifert, Markus Dahlweid and Thomas Santen  
European Microsoft Innovation Center (EMIC), Microsoft Research

**FORMULA**

Modeling Foundations.



# i. Type Declarations

<http://research.microsoft.com/formula>

**FORMULA**

Modeling Foundations.





# Domains Encapsulate Abstractions

*Definition.*

$G \stackrel{\text{def}}{=} (V, E)$  such that  $E \subseteq V \times V$ .

*A domain contains ADTs and constraints.*

```
domain Digraph
{
    //// ADT Declarations
    //// CLP rules/queries
}
```



# Type Declarations - Constructors

Declares both an n-ary constructor and type with the same name.

$$con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$$

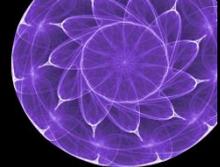
Body of declaration gives type constraints and labels for each argument.

$$V \quad ::= (\text{id}: \text{Integer}).$$
$$E \quad ::= (\text{src}: V, \text{dst}: V).$$
$$\text{path} ::= (\text{beg}: V, \text{end}: V).$$

Base types are order-sorted and include constants and:

NegInteger, PosInteger, Natural, Integer, Real,  
String, Boolean, Basic, Any.

The type of a constant is itself (denoting a singleton set containing the constant).



# Type Declarations - Enumerations

Declares a type naming a finite set of constants.

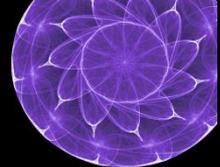
$$enm ::= \{ c_1, \dots, c_n \}.$$

Body contains constants.

```
Colors ::= { red, green, blue }.
```

```
Mixed ::= { 1, 2.5, true, "hello" }.
```

Enumerations can introduce new constants, like red. Numeric and string constants are predefined, as are the distinguished constants true/false.



# Type Declarations – Arbitrary Unions

Declares a type naming a union of terms.

$$unn ::= \tau_1 + \cdots + \tau_n.$$

Union types are not tagged and equivalent to the mathematical union.

GraphThing ::= V + E.

VertexOrColor ::= V + Colors.



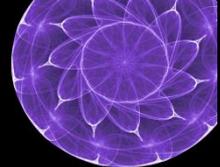
# Type Declarations – Recursive ADTs

Have enough ingredients to represent many structures.

```
Nil    ::= { nil }.  
VList ::= Nil + VList.  
VList ::= (head: V, tail: VList).
```

Unions can simulated multiple-inheritance and interfaces.

```
Colors ::= { red, green, blue }.  
V      ::= (id : Integer).  
ColV   ::= (id : Integer, col: Colors).  
IV     ::= V + ColV.  
E      ::= (src: IV, dst: IV).
```



# Subtyping

Types are subtype related if one denotes a subset of the other.

$$\tau' <: \tau \stackrel{\text{def}}{=} \llbracket \tau' \rrbracket \subseteq \llbracket \tau \rrbracket$$

Where  $\llbracket \tau \rrbracket$  is the set of terms accepted by  $\tau$ .

```
1 <: 1 U 2 <: PosInteger <: Natural  
<: Integer <: Real <: Basic
```

```
V <: GraphThing, V <: VertexOrColor,  
GraphThing <: Any, VertexOrColor <: Any
```

```
VLCon(V(Natural), nil) <: VList
```

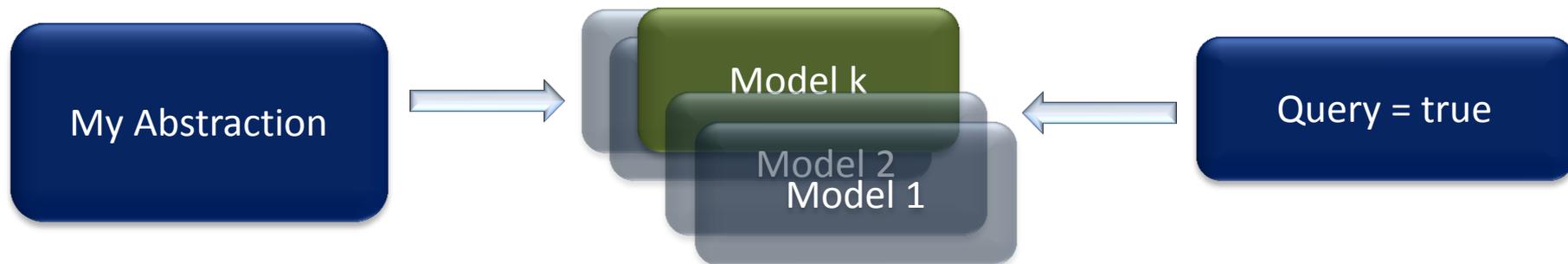


# Primitive/Derived Constructors

Some constructors represent the building blocks of an abstraction. Others are used to store information derived from these building blocks.

```
domain Digraph
{
  primitive V ::= (id: Integer).
  primitive E ::= (src: V, dst: V).
  path ::= (beg: V, end: V).
}
```

The primitive keyword indicates the building blocks and impacts modeling finding, because models can only contain primitives.

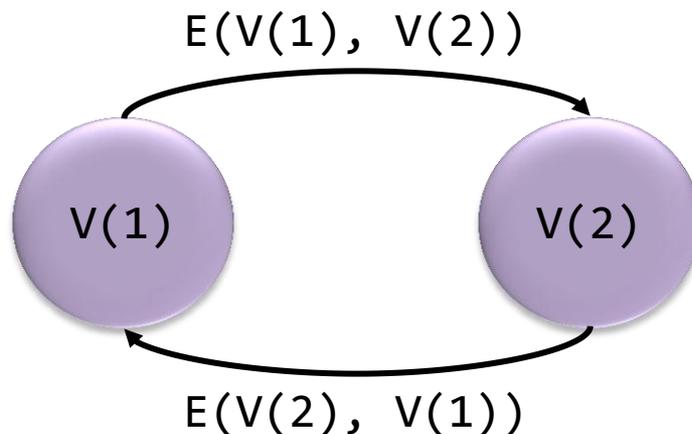




# Models are Sets of Primitives...

...that conform to a domain.

```
model G1 of Digraph
{
  V(1) V(2)
  E(V(1), V(2))
  E(V(2), V(1))
}
```



Can alias terms with the “is” keyword.

```
model G1p of Digraph
{
  v1 is V(1) v2 is V(2)
  E(v1, v2)
  E(v2, v2)
}
```

*Claims to conform to the Digraph domain* ▶

*A set of terms, which are applications of constructors.* ▶



# Compiler Checks Conformance

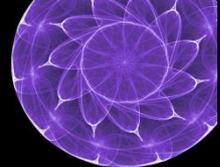
Compiler rejects this model, because it doesn't respect data type declarations.

```
✘ model G2 of Digraph
{
    V(E(V(1), V(2)))
}
```

Compiler rejects this model, because it contains a term that is not primitive:

```
✘ model G3 of Digraph
{
    path(V(1), V(2))
}
```

We will see more complex conformance constraints soon...



# Models Close Domains

Models extend a domain with facts, and the result is a closed logic program.

```
model G1p of Digraph
{
  v1 is V(1) v2 is V(2)
  E(v1, v2) E(v2, v2)
}
```

≡

Digraph U

```
{ V(1). V(2). E(V(1), V(2)). E(V(2), V(2)). }
```

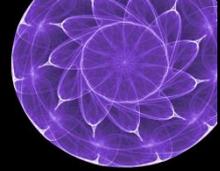
## ii. Rules

<http://research.microsoft.com/formula>

**FORMULA**

Modeling Foundations.





# Rules

For more complex constraints we need logic programs.

$$f(\vec{t}) \leftarrow f_1(\vec{t}_1), \dots, f_m(\vec{t}_m),$$
$$\mathbf{fail} \ g_1(\vec{s}_1), \dots, \mathbf{fail} \ g_n(\vec{s}_n),$$
$$c_1 \mathit{rel}_1 c'_1, \dots, c_o \mathit{rel}_o c'_o.$$

Rule construct new facts from the set of currently known facts  $K$ .  
Roughly, a rule creates new knowledge for every substitution  $\theta$  such that:

- (1) Find predicates hold:  $f_i(\vec{t}_i\theta) \in K$ .
- (2) Fail predicates hold:  $g_i(\vec{s}_i\theta) \notin K$ .
- (3) Constraint predicates hold:  $(c_i\theta, c'_i\theta) \in \mathit{rel}_i$ .

$$\mathit{rel}_i \in \{ =, !=, <, <=, >, >= \}$$

Order of rules/predicates doesn't matter.  
Finds occur in higher quantifier scope than fails.



# Edges and Paths

The following sets of rules are equivalent ways of defining the paths in a graph.

```
path(x, y) :- E(x, y).  
path(x, z) :- path(x, y), path(y, z).
```

```
path(x, z) :- path(x, y), path(y, z).  
path(x, y) :- E(x, y).
```

```
path(x, y) :- E(x, y).  
path(x, z) :- E(x, y), path(y, z).
```

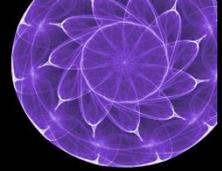
```
path(x, z) :- E(x, y), path(y, z).  
path(x, y) :- E(x, y).
```



# Digraphs Again

```
domain Digraph
{
  primitive V ::= (id: Integer).
  primitive E ::= (src: V, dst: V).
  path ::= (beg: V, end: V).

  path(x, y) :- E(x, y).
  path(x, z) :- path(x, y), path(y, z).
}
```



# Knowledge

The knowledge generated by a model is the least  $K^\infty$  (least fixpoint) where no rules extend  $K^\infty$ .

$$K_0 = \{ V(1), V(2), E(V(1), V(2)), E(V(2), V(2)) \}$$

►  $\text{path}(x, y) :- E(x, y).$

Let  $\theta(x) \mapsto V(1), \theta(y) \mapsto V(2)$  then  $E(x\theta, y\theta) \in K_0$

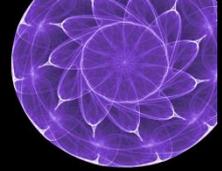
$$\{ \text{path}(x\theta, y\theta) \stackrel{\text{def}}{=} \text{path}(V(1), V(2)) \} \cup K_0 = K_1$$

Let  $\theta(x) \mapsto V(2), \theta(y) \mapsto V(2)$  then  $E(x\theta, y\theta) \in K_1$

$$\{ \text{path}(x\theta, y\theta) \stackrel{\text{def}}{=} \text{path}(V(2), V(2)) \} \cup K_1 = K_2$$

$$K_2 = K_0 \cup \{ \text{path}(V(1), V(2)), \text{path}(V(2), V(2)) \}$$

\*FORMULA computes the lfp bottom-up, like datalog instead of prolog.



# Knowledge

►  $\text{path}(x, y) :- \text{path}(x, y), \text{path}(y, z).$

Let  $\theta(x) \mapsto V(1), \theta(y) \mapsto V(2)$  then  $\text{path}(x\theta, y\theta) \in K_2$

$$\{ \text{path}(x\theta, y\theta) \stackrel{\text{def}}{=} \text{path}(V(1), V(2)) \} \cup K_2 = K_3$$

Let  $\theta(x) \mapsto V(2), \theta(y) \mapsto V(2)$  then  $\text{path}(x\theta, y\theta) \in K_3$

$$\{ \text{path}(x\theta, y\theta) \stackrel{\text{def}}{=} \text{path}(V(2), V(2)) \} \cup K_3 = K_4$$

$$K_4 = K_1 \cup \{ \text{path}(V(1), V(2)), \text{path}(V(2), V(2)) \} = K_1$$

Therefore  $K^\infty \stackrel{\text{def}}{=} K_1$  is the lfp, which is everything knowable from the model.



# More On Rules – Range Restriction

Rules can only do a finite amount of work at each step.

Rule:  $h \text{ :- } B[K].$

          ▲          ▲

          Head      Body, a first order formula  
                          with free variable  $K$

Let  $sat(B, K)$  be the set of all satisfying substitutions for  $B$  given  $K$ .

Range restriction.

If  $|K|$  is finite then  $|sat(B, K)|$  must be finite.

\*Range restriction allows for an effective symbolic execution procedure to eliminate nested quantifiers and fixpoints.



# More On Rules – Range Restriction

Range restriction must be approximated.  
(Some range restricted programs are rejected.)

✗  $\text{path}(x, y) \text{ :- } V(x).$



Derives an infinite number of paths.  
Rule: Head variables must be bound on RHS.

✗  $\text{path}(x, y) \text{ :- } V(x), \text{ fail } V(y).$



Derives an infinite number of paths.  
Rule: Head variables must appear outside of fail scopes.

✗  $\text{path}(x, y) \text{ :- } V(x), y + 1 = x * x.$



Range restricted, but cannot prove it.  
Rule: Write directed equalities to make this obvious.



# More On Rules – Range Restriction

These programs are range restricted.

✓  $\text{path}(x, V(z)) \text{ :- } E(x, y), V(z) = y.$

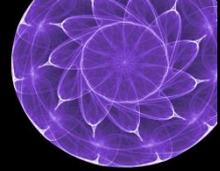


The compiler knows  $z$  is a subterm of  $y$  and  $y$  is range restricted because it is a subterm of  $E(x, y) \in K$ .

✓  $\text{path}(x, y) \text{ :- } V(x), y = x * x - 1.$



Because  $y$  is function of  $x$ , which is range restricted.



# More On Rules – Type Safety

Rules must construct well-typed terms regardless of the model used to close the program.

Rule:  $h :- B[K]$ .

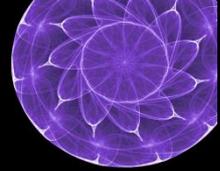
And:  $x \in B$

Then an over-approximation of the values  $x$  can take is:

$$values(x) \stackrel{\text{def}}{=} \bigcup_K \{ \theta(x) \mid \theta \in sat(B, K) \}$$

Type inference assigns  $x$  a type  $\tau_x$  which over approximates  $values(x)$  by a regular set:

$$values(x) \subseteq \llbracket \tau_x \rrbracket$$



# More On Rules – Type Safety

A rule is type safe if the inferred type of the head is a subtype of the declared type.

$$f(\vec{t}[\vec{x}]) :- B.$$

The inferred type of the head is:

$$f(\vec{t}[x_i \setminus \tau_{x_i}])$$

The inferred type of the head must satisfy:

Type safety.

$$f(\vec{t}[x_i \setminus \tau_{x_i}]) <: f$$

A rule is also rejected if the body cannot be satisfied:

$t : \perp$  for some term in B.

$p : false$  for some predicate in B.



# More On Rules – Type Safety

The following rules are not type safe

✗  $\text{path}(x, V(y)) \text{ :- } V(x), y = x / 2.$



A world with an odd numbered vertex gives  $y$  a fractional value, so  $y : \text{Real}$ . Therefore  $\text{path}(x, V(y)) : \text{path}(V, V(\text{Real}))$ .

✗  $\text{path}(x, y) \text{ :- } E(x, V(y)), x = y.$



Bug here, because  $x : V$  and  $y : \text{Integer}$ . But  $x = y$  implies  $x, y : V \cap \text{Integer} \equiv \perp$ .

✗  $\text{path}(V(x), p) \text{ :- } V(x), V(y), p = \text{path}(V(x), V(y)).$



Because  $\text{path}(x, p) : \text{path}(V, \text{path})$ .



# More On Rules – Type Safety

These rules are type safe.

Colors ::= { red, green, blue }.

V ::= (id : Integer).

ColV ::= (id : Integer, col: Colors).

IV ::= V + ColV.

E ::= (src: IV, dst: IV).

cpath ::= (beg: ColV, end: ColV).

✓ cpath(x, y) :- E(x, y), x.col = y.col.

Because only colored vertices have col arguments.

cpath(x, y) : cpath(ColV, ColV).

✓ cpath(x, y) :- E(x, y), x.col = y.col, x.col = red.

cpath(x, y) : cpath(ColV(Integer, red), ColV(Integer, red)).

## iii. Queries

<http://research.microsoft.com/formula>

**FORMULA**

Modeling Foundations.





# Query Definitions

Query definitions are special rules where the LHS is a constant.  
Query constants are separate from data type constants.

$$G \stackrel{\text{def}}{=} (V, E) \text{ such that } E \subseteq V \times V.$$



How do we encode this constraint?

`undec1Vertex := E(V(x), _), fail V(x).`



Query name



Query operator



Fresh “don’t care” variable

`undec1Vertex := E(_, V(y)), fail V(y).`

Multiple query definitions are allowed;  
they permit the same query to be derived in several different ways.



# Undeclared Vertices

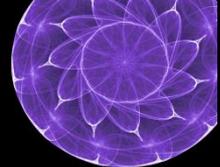
```
model GUndecl of Digraph
{ V(1), E(V(1), V(10)) }
```

$$K_0 = \{ V(1), E(V(1), V(10)) \}$$

► undeclVertex := E(\_, V(y)), fail V(y).

Let  $\theta(\_) \mapsto V(1), \theta(y) \mapsto 10$  then  $E(\_ \theta, V(y\theta)) \in K_0$  and  $V(y\theta) \notin K_0$ .

undeclVertex  $\in K_1$



# Conforms Query

We must tell the compiler that undeclared vertices are disallowed.  
This is done using the special `conforms` query.

```
conforms := !undeclaredVertex.
```



Queries are treated like Boolean variables on the RHS.

Every domain has a `conforms` query, even if it isn't explicitly defined.  
All conformance constraints must appear in the `conforms` query.



# Digraphs Yet Again

```
domain Digraph
{
  primitive V ::= (id: Integer).
  primitive E ::= (src: V, dst: V).
  path ::= (beg: V, end: V).

  path(x, y) :- E(x, y).
  path(x, z) :- path(x, y), path(y, z).

  undeclVertex := E(V(x), _), fail V(x).
  undeclVertex := E(_, V(y)), fail V(y).
  conforms := !undeclVertex.
}
```



# Legal Specifications - Termination

In addition to range restriction and well-typedness, FORMULA must be terminating.

Termination allows for a bottom-up fixpoint procedure and generalization to symbolic execution.

## Termination

The knowledge generated by an arbitrary model must be finite, i.e. a finite fixpoint must exist.

✗  $\text{path}(V(y), V(y)) \text{ :- path}(V(x), \_), y = x + 1.$

\*For those familiar with backwards-chaining LP, this requires an adjustment. Currently, FORMULA does not perform any termination analysis.



# Legal Specifications - Stratification

Not all logic programs have unique interpretations.

Stratification is a syntactic constraint on `fail`, which guarantees unique interpretations.

FORMULA programs must be stratified. This means specifications are in a fragment of FPL where model finding is easier to encode into constraint solvers.

Stratification

There is no subset of rules producing an f-term by failing on an f-term.

\*The compiler checks stratification.



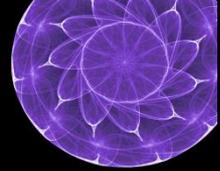
# Legal Specifications - Stratification

This specification is not stratified.

```
domain DigraphNotStrat
{
  primitive V ::= (id: Integer).
  primitive E ::= (src: V, dst: V).
  path ::= (beg: V, end: V).

  path(x, y) :- E(x, y).
  ✘ path(x, x) :- path(x, y), fail path(y, y).
}

model Problem of DigraphNotStrat
{
  V(1) V(2)
  E(V(1), V(2))
  E(V(2), V(1))
}
```



# Legal Specifications - Stratification

Consider computing the knowledge of `Problem`.

$$K_2 = \{\dots, path(V(1), V(2)), path(V(2), V(1))\}$$

► `path(x, x) :- path(x, y), fail path(y, y).`

Let  $\theta(x) \mapsto V(1), \theta(y) \mapsto V(2)$  then

`path(x $\theta$ , y $\theta$ )`  $\in K_2$  and `path(y $\theta$ , y $\theta$ )`  $\notin K_2$

$$\{ path(x\theta, x\theta) \stackrel{\text{def}}{=} path(V(1), V(1)) \} \cup K_2 = K_3$$

Let  $\theta(x) \mapsto V(2), \theta(y) \mapsto V(1)$  then

`path(x $\theta$ , y $\theta$ )`  $\in K_3$  and `path(y $\theta$ , y $\theta$ )`  $\in K_3$  ✘

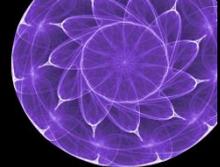
But, if the second substitution had been chosen first, then the result would be different. There are two least fixpoints:

$$K_3 = K_2 \cup \{ path(V(1), V(1)) \}$$

$$K'_3 = K_2 \cup \{ path(V(2), V(2)) \}$$

# vi. More Language Details

<http://research.microsoft.com/formula>



# Shortcuts - Closed

Constraints, such as no undeclared vertices, are very common. FORMULA provides annotations that automatically introduce the appropriate rules and modify conforms.

The closed annotation; subterms must be known (in  $K^\infty$ ).

$$[Closed(lbl_{d_1}, \dots, lbl_{d_m})]con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$$

The constraint is trivially satisfied for constants, because all constants are known.

Unnecessary  
constraint

```
► [Closed(id)]  
primitive V ::= (id: Integer).  
[Closed(src, dst)]  
primitive E ::= (src: V, dst: V).
```



# Shortcuts - Unique

The unique annotation; any two con-terms with the same arguments on  $\{d_1, \dots, d_m\}$  must have the same arguments on  $\{e_1, \dots, e_o\}$ .

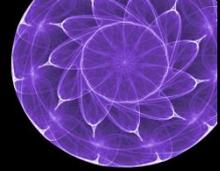
$$\left[ \text{Unique}(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o}) \right]$$

$con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$

For colored vertices there is a unique color for every vertex id:

```
[Unique(id -> col)]  
ColV    ::= (id : Integer, col: Colors).
```

```
✘ model GBad of ColDigraph  
  {  
    ColV(1, red)  
    ColV(1, blue)  
  }
```



# Shortcuts - Total

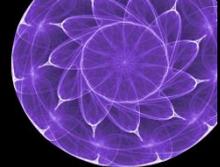
The total annotation; there must be a con-term for all possible combinations of legal arguments  $\{d_1, \dots, d_m\}$ .

$$\begin{aligned} & [Total(lbl_{d_1}, \dots, lbl_{d_m})] \\ con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n). \end{aligned}$$

This means there must be a vertex term for every possible integer.

$$\begin{aligned} & [Total(id)] \\ \times \text{primitive } V ::= (id: Integer). \end{aligned}$$

This use of the annotation is illegal, because a model cannot have an infinite number of terms.



# Shortcuts - Total

If Total is used with Closed, then the compiler knows only a finite number of elements are accepted.

```
primitive V ::= (id: Integer).  
[Total(src, dst)][Closed(src, dst)]  
primitive E ::= (src: V, dst: V).
```

✓ `model GGood of Digraph`

```
{  
  v is V(1) v is V(2)  
  E(v1, v1) E(v1, v2)  
  E(v2, v1) E(v2, v2)  
}
```

✗ `model GBad of Digraph`

```
{  
  v is V(1) v is V(2)  
  E(v1, v1) E(v1, v2)  
  E(v2, v1)  
}
```

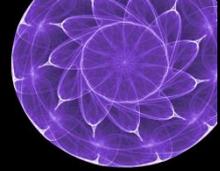


# More Shortcuts - Function

These annotations form the building blocks for other common constraints.

$$\left[ \textit{Function}(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o}) \right]$$
$$\textit{con} ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$$

$$\left[ \textit{Unique}(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o}) \right]$$
$$\left[ \textit{Total}(lbl_{d_1}, \dots, lbl_{d_m}) \right]$$
$$\textit{con} ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$$



# More Shortcuts - Injection

$[Injection(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o})]$   
 $con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$

$[Function(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o})]$   
 $[Unique(lbl_{e_1}, \dots, lbl_{e_o} \rightarrow lbl_{d_1}, \dots, lbl_{d_m})]$   
 $con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$



# More Shortcuts - Surjection

$[Surjection(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o})]$   
 $con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$

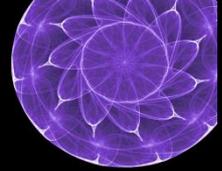
$[Function(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o})]$   
 $[Total(lbl_{e_1}, \dots, lbl_{e_o})]$   
 $con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$



# More Shortcuts - Bijection

$[Bijection(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o})]$   
 $con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$

$[Injection(lbl_{d_1}, \dots, lbl_{d_m} \rightarrow lbl_{e_1}, \dots, lbl_{e_o})]$   
 $[Injection(lbl_{e_1}, \dots, lbl_{e_o} \rightarrow lbl_{d_1}, \dots, lbl_{d_m})]$   
 $con ::= (lbl_1: \tau_1, \dots, lbl_n: \tau_n).$



# Digraphs, Finally

```
domain Digraph
{
  primitive V ::= (id: Integer).
  [Closed(src, dst)]
  primitive E ::= (src: V, dst: V).
  path ::= (beg: V, end: V).

  path(x, y) :- E(x, y).
  path(x, z) :- path(x, y), path(y, z).
}
```



# Shortcuts – Reusing Bodies

Multiple rules with the same head can be written as one rule with comma-separated heads.

$$\text{path}(x, y), \text{path}(y, x) \text{ :- } E(x, y).$$


Computes symmetric  
closure of E.

Multiple rules with the same and different bodies can be written as one rule with semi-colon-separated bodies.

$$\text{path}(x, z) \text{ :- } E(x, z); \text{path}(x, y), \text{path}(y, z).$$


Each semicolon is a  
new scope



# Generalizations – “Is”

The FORMULA syntax also supports some generalizations of classic logic programming.

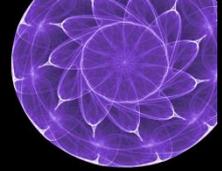
```
GraphThing ::= V + E.
```

```
threePath := x is GraphThing, y is GraphThing,  
             x.dst = y.src, x.src != y.dst.
```

The variables  $x$  and  $y$  range over all known terms conforming to the type `GraphThing`

```
SameCol(v1, v2) := v1 is ColV(_, col), v2 is ColV(_, col).
```

Can also use `is` to bind a variables to find/fail predicates.



# Generalizations – “In”

Attaches a type membership constraint to a variable.

```
Mixed ::= { 1, 2.5, true, "hello" }.
```

```
v1 := V(x), x in Mixed.
```



```
Integer  $\cap$  Mixed  $\equiv$  1
```



# Generalizations – Queries and Boolean Ops

Queries can be used only on the RHS of other query definitions.  
When this occurs, they are combined using the usual Boolean connectives  
instead of LP.

$$q ::= !(q1 \mid q2) \& q3.$$

▲            ▲            ▲  
not          or          and

Query definitions should be oriented, i.e. no circular-definitions.

✗  $q1 ::= !q2. \quad q2 ::= !q1.$

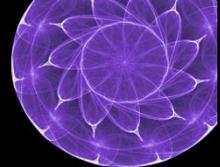
# vi. Partial Models

<http://research.microsoft.com/formula>

**FORMULA**

Modeling Foundations.



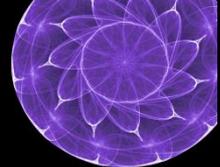


# Partial Models

Partial models allow targeted model finding.  
They are essentially lower bounds on the sort of models we want to find.

```
partial model P of Digraph
{
  E(_, V(3)) E(x, y) E(y, z)
}
  ▲                ▲
fresh variable    named variable
```

The model finder must return a solution containing these terms under some substitution.



# Solving – Cardinality Inference

The model finder tries to assign consistent cardinalities to primitives so that the Closed, Total, and Unique annotations can be satisfied.

$$\#E > 0 \Rightarrow \#V > 0.$$

This problem is undecidable (non-linear Diophantine inequalities), but solved up to some maximum cardinalities and greedily minimized.

Added two vertices with unknown IDs. ▶ `partial model P of Digraph`

```
{  
  V(_) V(  
  E(_, V(3)) E(x, y) E(y, z)  
}
```

The partial model is automatically extended with consistent number of non-ground terms.



# Solving – Symbolic Execution

The program is symbolically executed under the extended partial model and the target query is solved using Z3.

```
partial model P of Digraph
```

```
{
```

```
  V(_) V(_)
```

```
  E(_, V(3)) E(x, y) E(y, z)
```

```
}
```


$$\theta(_0) \mapsto -78$$
$$\theta(_1) \mapsto 3$$
$$\theta(x), \theta(y), \theta(z) \mapsto V(3)$$

```
model P_1 of Digraph at "../ICTAC.4m1"
```

```
{
```

```
  V(-78)
```

```
  V(3)
```

```
  E(V(3), V(3))
```

```
}
```

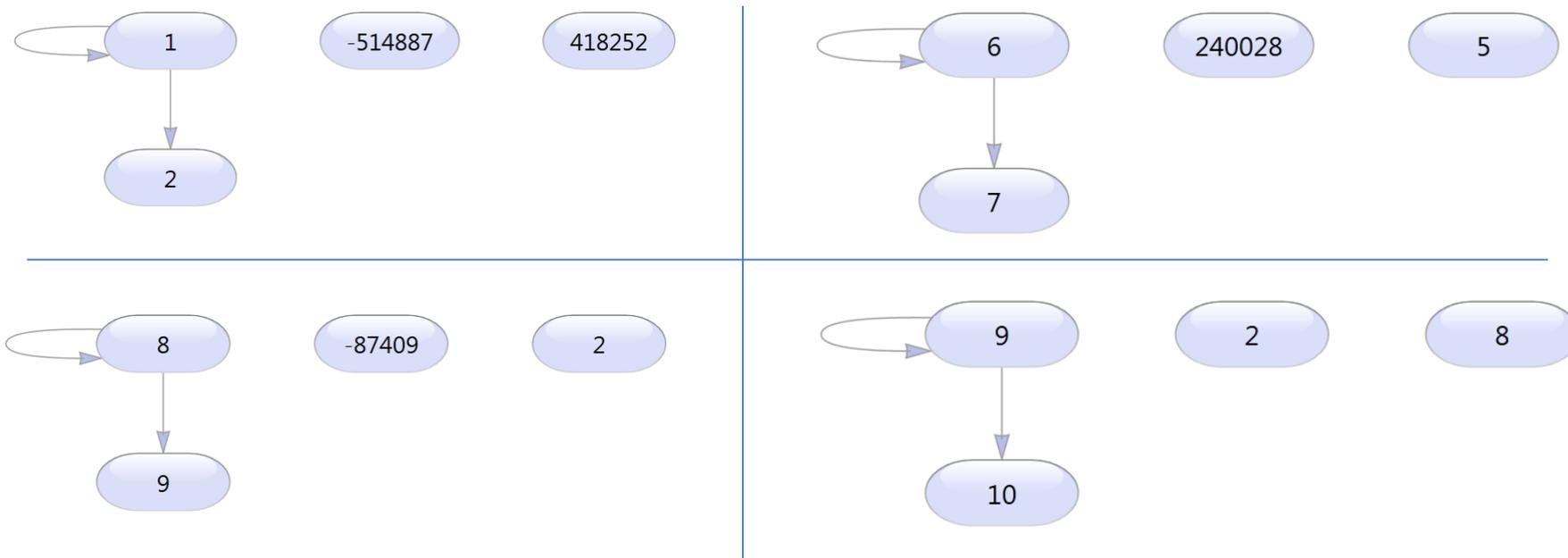
Each solution provides a substitution that is used to convert a partial model into a complete model.



# Solving – Search By Point Exclusion

There are two modes of searching for multiple solutions: **point exclusion** and **symmetry exclusion**.

Point exclusion prevents the solver from returning any solution that has already been seen, and then asks for another one.



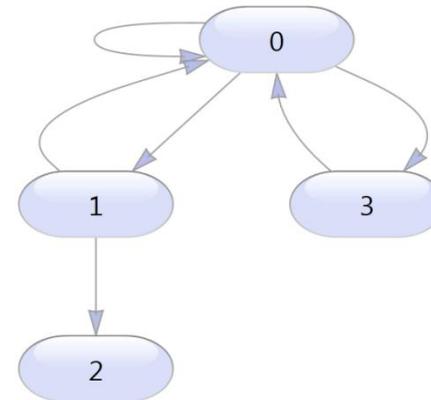
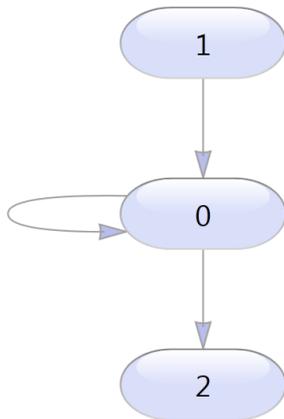
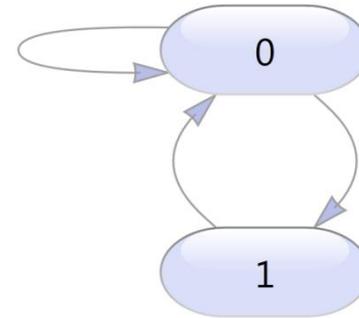
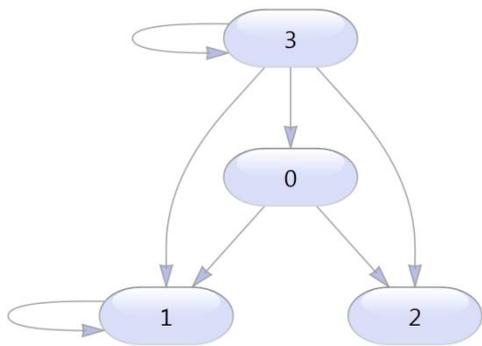
Without additional configuration, point exclusion is the default search mechanism.



# Solving – Search By Symmetry Exclusion

Symmetry exclusion prevents the solver from returning isomorphic solutions, and also externally randomizes the search for non-isomorphic solutions.

```
[Config(EXPLORATION_METHOD=SymmetryExclusion)]  
partial model P3 of Graph { ... }
```





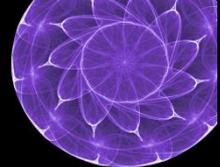
# Configuring Search - Introduce

There are several annotations for controlling the size and contents of the partial model.

Add  $n$  terms of the form  $f(\_, \dots, \_)$  to the partial model

*[Introduce( $f, n$ )]*

```
[Introduce(E, 10)]  
partial model P2 of Digraph { }
```



# Configuring Search - Cardinality

The solution must contain at least  $m$  distinct f-terms and at most  $n$  distinct f-terms

$[Cardinality(f, m, n)]$

---

The solution must contain at least  $m$  distinct f-terms

$[Cardinality(f, m)]$

---

`[Cardinality(E,5,10)]`

`partial model P2 of Digraph { }`



# Solving – Increasing The Size of Worlds

By default, the model finder does not automatically increase the size of the partial model after the first cardinality inference. However, this can easily be adjusted.

The model finder will now continue to increase the size of the partial model. At each step a requested size is reported and then cardinality inference and symbolic execution is repeated.

```
[Search]  
partial model P2 of Digraph { }
```

```
[Search(YourOwnStrategy)]  
partial model P2 of Digraph { }
```

Users can defined their own search preferences very easily.

Questions?

<http://research.microsoft.com/formula>

**FORMULA**

Modeling Foundations.

