Semantics of programming languages

William Steingartner

william.steingartner@tuke.sk

Faculty of Electrical Engineering and Informatics
Department of Computers and Informatics

Semantics of programming languages

CEEPUS Spring 2018/2019

Coalgebraic structural operational semantics

- Coalgebra allows to model behavior of program systems.
- Coalgebra defines structural operational semantics of programs written in some programming language.
- Coalgebra is constructed over category of state space by polynomial endofunctor.
- Every application of functor models one step of computation so as structural operational semantics.

Language *Jane*

- We briefly show the structural operational semantics of *Jane*.
- The following syntactic domains are introduced:

$n \in \mathbf{Num}$	– strings of digits;
$x \in \mathbf{Var}$	– names of variables;
$e \in \mathbf{Expr}$	– arithmetic expressions;
$b \in \mathbf{Bexpr}$	– Boolean expressions;
$S \in \mathbf{Statm}$	– statements.
$D \in \mathbf{Decl}$	– declarations of variables.

Syntactic domains $\mathbf{Num}$ and $\mathbf{Var}$ do not have internal structure from semantic point of view.

Language *Jane*: Syntax

Syntactic domain $\mathbf{Expr}$ contains well-formed arithmetic expressions according the following syntactic rule:

$$e ::= n \mid x \mid e + e \mid e - e \mid e * e.$$

Well-formed Boolean expressions from $\mathbf{Bexpr}$ can be of the following forms:

$$b ::= \texttt{false} \mid \texttt{true} \mid e = e \mid e \leq e \mid \neg b \mid b \wedge b.$$

Statements $S \in \mathbf{Statm}$ are standard (basic) Dijkstra's constructs (also called D-diagrams) - variable assignment, empty statement, sequence of statements, conditional statement and loop statement:

$$S ::= x := e \mid \texttt{skip} \mid S; S \mid \texttt{if } b \texttt{ then } S \texttt{ else } S \mid \texttt{while } b \texttt{ do } S.$$

State

- The main semantic domain in structural operational semantics is $\mathbf{State}$ which contains particular memory states;
- state $s \in \mathbf{State}$ is an abstraction of computer memory;
- each state is considered as function:

$$s : \mathbf{Var} \times \mathbf{Level} \to \mathbf{Z}.$$

A change of memory content (cell) is represented as an actualization of a state:

$$s' = s[((x, l), v) \mapsto ((x, l), \mathbf{n})].$$

Semantics of expressions

Semantic functions for arithmetic and Boolean expressions, resp., are

$$[\![e]\!] : \mathbf{Expr} \to \mathbf{State} \to \mathbf{Value} \qquad [\![b]\!] : \mathbf{Bexpr} \to \mathbf{State} \to \mathbf{Bool}$$

$$[\![n]\!]s = \mathbf{n}$$

$$[\![x]\!]s = [\![get]\!](x, s)$$

$$[\![e_1 + e_2]\!]s = [\![e_1]\!]s \oplus [\![e_2]\!]s$$

$$[\![e_1 - e_2]\!]s = [\![e_1]\!]s \ominus [\![e_2]\!]s$$

$$[\![e_1 * e_2]\!]s = [\![e_1]\!]s \otimes [\![e_2]\!]s$$

$$[\![\texttt{true}]\!]s = \mathbf{true}$$

$$[\![\texttt{false}]\!]s = \mathbf{false}$$

$$[\![e_1 = e_2]\!]s = \begin{cases} \mathbf{true} & \text{if } [\![e_1]\!]s = [\![e_2]\!]s \\ \mathbf{false} & \text{otherwise} \end{cases}$$

$$[\![e_1 \leq e_2]\!]s = \begin{cases} \mathbf{true} & \text{if } [\![e_1]\!]s \leq [\![e_2]\!]s \\ \mathbf{false} & \text{otherwise} \end{cases}$$

$$[\![\neg b]\!]s = \begin{cases} \mathbf{true} & \text{if } [\![\neg b]\!]s = \mathbf{false} \\ \mathbf{false} & \text{otherwise} \end{cases}$$

$$[\![b_1 \wedge b_2]\!]s = \begin{cases} \mathbf{true} & \text{if } [\![b_1]\!]s \leq [\![b_2]\!]s = \mathbf{true} \\ \mathbf{false} & \text{otherwise} \end{cases}$$

Semantics of statements

Semantic function is defined as follows:

$$[\![S]\!] : \mathbf{Statm} \to \mathbf{State} \to \mathbf{State}$$

$$\langle x := e, s\rangle \Rightarrow s[x \mapsto [\![e]\!]s] \quad (1_{os}) \qquad \langle \mathtt{skip}, s\rangle \Rightarrow s \quad (2_{os})$$

$$\frac{\langle S_1, s\rangle \Rightarrow \langle S_1', s'\rangle}{\langle S_1; S_2, s\rangle \Rightarrow \langle S_1'; S_2, s'\rangle}\ (3^1_{os}) \qquad \frac{\langle S_1, s\rangle \Rightarrow s'}{\langle S_1; S_2, s\rangle \Rightarrow \langle S_2, s'\rangle}\ (3^2_{os})$$

$$\frac{[\![b]\!]s = \mathbf{true}}{\langle \mathtt{if}\ b\ \mathtt{then}\ S_1\ \mathtt{else}\ S_2, s\rangle \Rightarrow \langle S_1, s\rangle}\ (4^{\mathbf{true}}_{os})$$

$$\frac{[\![b]\!]s = \mathbf{false}}{\langle \mathtt{if}\ b\ \mathtt{then}\ S_1\ \mathtt{else}\ S_2, s\rangle \Rightarrow \langle S_2, s\rangle}\ (4^{\mathbf{false}}_{os})$$

$$\langle \mathtt{while}\ b\ \mathtt{do}\ S, s\rangle \Rightarrow \langle \mathtt{if}\ b\ \mathtt{then}\ (S; \mathtt{while}\ b\ \mathtt{do}\ S)\ \mathtt{else}\ \mathtt{skip}, s\rangle \quad (5_{os})$$

Basic notions

- Coalgebra is mathematical structure, that is constructed over a base category by polynomial endofunctor;
- Category $\mathscr{C}$ is a structure, that consists of objects and morphisms between them;
- Functor is a kind of morphism between categories. Functor defined on one category is called endofunctor.

$$F : \mathscr{C} \to \mathscr{C}$$

The name *polynomial* is used for functor which has a form of polynomial:

$$FX = A_0 + A_1 \times X^{B_1} + A_2 \times X^{B_2} + \ldots + A_n \times X^{Bn},$$

where A_i, B_i are sets, and X stands for a state space, a class of objects in base category.

- Then, a coalgebra is a mapping:

$$c : X \to FX,$$

where c is n-tuple of morphisms of category that causes change of state.

Language *E-Jane*

We extend the base language with new constructs:

- Declarations of variables (without initializations):

$$D ::= \texttt{var}\ x \mid \varepsilon$$

  where ε is an empty declaration;

- block statement, user input and output statements:

$$S ::= \ldots \mid \texttt{begin}\ D; S\ \texttt{end} \mid \texttt{input}\ x \mid \texttt{print}\ e.$$

- A source program in *E-Jane* has a form

$$D_1; \ldots; D_m; S_1; \ldots; S_n,$$

  so it consists of a list of declarations and a list of statements.

Signatures for lists of declarations and statem/ents

First, we specify signatures for the lists of declarations and statements:

$$
\begin{array}{l}
\Sigma_{Decl\_List} = List_{fin}\,[Decl] \\
types: \quad Decl\_List, Decl \\
opns: \\
\quad init_d :\rightarrow Decl\_List \\
\quad head_d : Decl\_List \rightarrow Decl \\
\quad tail_d : Decl\_List \rightarrow Decl\_List
\end{array}
$$

$$
\begin{array}{l}
\Sigma_{Statm\_List} = List_{fin}\,[Statm] \\
types: \quad Statm\_List, Statm \\
opns: \\
\quad init :\rightarrow Statm\_List \\
\quad head : Statm\_List \rightarrow Statm \\
\quad tail : Statm\_List \rightarrow Statm\_List
\end{array}
$$

Signature for memory

An abstraction of a memory we specify as a signature $Memory$:

$$\Sigma_{Memory} = \begin{array}{ll} types: & Memory, Var, Value \\ opns: & init :\rightarrow Memory \\ & alloc : Var, Memory \rightarrow Memory \\ & get : Var, Memory \rightarrow Value \\ & del : Memory \rightarrow Memory \end{array}$$

- $init$ establishes an initial memory of a program;
- $alloc$ reserves a memory cell for declared variable;
- get returns (gets) an actual values of a variable;
- del deallocates (forgets) all memory cells defined on the highest nesting level.

Signature for configuration

In our approach, we will use *configuration* as a kind of state (snapshot of a memory) which we specify as follows:

$$\begin{aligned}\Sigma_{Config} = \quad & \Sigma_{Decl\_List} + \Sigma_{Statm\_List} + \Sigma_{Memory} + \\ & types : Config \\ & opns : next : Config \to Config \\ & \quad input : Config, Value \to Config \\ & \quad print : Config \to Value, Config\end{aligned}$$

- $next$ provides the next configuration;
- $input$ takes an input value and stores it into the given (concrete) memory cell;
- $print$ evaluates the value of an argument and returns it as an observable value.

Representation of types

We assign to particular specifications of types their representations as follows:

Type	**Representation**	
$Value$	$\mathbf{Value} = \mathbf{Z} \cup \{\perp\}$	
Var	$\mathbf{Var}$	
$Decl\_List$	$\mathbf{Decl\_List}$	$[\![D^*]\!] = [\![D_1; \ldots; D_n]\!]$
$Statm\_List$	$\mathbf{Statm\_List}$	$[\![S^*]\!] = [\![S_1; \ldots; S_n]\!]$
$Memory$	$\mathbf{Memory}$	
$Config$	$\mathbf{Config} = \mathbf{Program} \times \mathbf{Memory}$	

Representation of a program is :

$$\mathbf{Program} = [\![D^*; S^*]\!]$$

and an initial configuration of a program is:

$$config_0 = ([\![D^*; S^*]\!], m_0)$$

Representation of memory

We must to consider a level of nesting for a block statement:

$$l \in \mathbf{Level}, \quad \text{where } \mathbf{Level} \subseteq \mathbf{N}.$$

An actual memory $m \in \mathbf{Memory}$ we represent as a function:

$$m : \mathbf{Var} \times \mathbf{Level} \to \mathbf{Value}.$$

This function can be defined also by graph of a function:

$$graph(m) = \{((x, l), v) \mid m(x, l) = v\}.$$

We need an auxiliary function which returns the highest nesting level:

$$\begin{array}{ll} maxlevel : & \mathbf{Memory} \to \mathbf{Level}, \\ maxlevel(m) = & L, \text{where } \forall l_j . L \geq l_j, \end{array}$$

for

$$graph(m) = \{((x_1, 1), v_1), \ldots, ((x_i, l_j), v_k)\}.$$

Operations on a memory

An operation for initialization of memory we define as follows:

$$[\![init]\!] = m_0 = \{((\bot, 1), \bot)\}$$

variable	level	value
x_1	1	v_1
$\vdots$		
x_n	l	v_n

actual memory

variable	level	value
$\bot$	1	$\bot$

initial memory

Operations on a memory

The other operations in memory we define as follows:

$$
\begin{aligned}
[\![alloc]\!](x, m) = &\quad graph(m) \cup \{((x, maxlevel(m)), \bot)\} \\
[\![get]\!](x, m) = &\quad v, \quad \text{if } ((x, maxlevel(m)), v) \in m \\
[\![del]\!]m = &\quad graph(m) \setminus \{((x_i, maxlevel(m)), v_i) \mid i \in \mathbf{N}\}
\end{aligned}
$$

variable	level	value
$\vdots$	$\vdots$	$\vdots$
x	l	$\bot$

allocation of memory

variable	level	value
$\vdots$	$\vdots$	$\vdots$
x	l_{j-1}	v
x_i	l_j	v_k
$\vdots$	$\vdots$	$\vdots$
x_n	l_j	v_m

forgetting of local variables

Representation of configurations

The last step is to define the representation of configurations. Configurations $config$ are elements of the set $\mathbf{Config}$:

$$\mathbf{Config} = \mathbf{Program} \times \mathbf{Memory}.$$

An initial configuration before the execution of a program has a form:

$$config_0 = (\llbracket D_1; D_2; \ldots; D_m; S_1; S_2; \ldots; S_n \rrbracket, m_0).$$

After every step of program execution:

- a configuration is changed, an operation $\llbracket tail \rrbracket$ is applied on a list of declarations and statements;
- memory can be changed, this depends on an elaborated declaration or executed statement.

Set $\mathbf{Config}$ is considered as a **state space** of coalgebra.

Semantics of declarations

Semantics of declarations is defined as a function on a memory:

$$[\![\texttt{var}\ x]\!] : \mathbf{Memory} \to \mathbf{Memory}$$

as

$$[\![\texttt{var}\ x]\!]m = [\![alloc]\!](x, m).$$

We extend set of variables $\mathbf{Var}$ with special (dummy) variables `begin` and `end` to ensure that we can identify the beginning and the end of a block statement. Such declaration increments level of nesting and we define its elaboration as follows:

$$\begin{aligned} [\![begin]\!]m = &\quad graph(m) \cup \{((\texttt{begin}, l+1), \bot)\}, \\ [\![end]\!]m = &\quad [\![del]\!]m. \end{aligned}$$

Extending the semantics of declarations for configurations

Because a state space of coalgebra is a set of configurations, we extend the definition for elaboration of declarations for configurations. We define a morphism:

$$[\![next]\!] : \mathbf{Config} \to \mathbf{Config},$$

and we define elaboration of declarations by the following functions:

$$\begin{array}{ll} [\![next]\!]([\![\texttt{var}\ x; D^*; S^*]\!], m) = & ([\![D^*; S^*]\!], [\![\texttt{var}\ x]\!]\ m), \\ [\![next]\!]([\![\texttt{begin}\ D^*; S'\ \texttt{end}; S^*]\!], m) = & ([\![D^*; S'\ \texttt{end}; S^*]\!], [\![begin]\!]m) \\ [\![next]\!]([\![\texttt{end}; S^*]\!], m) = & ([\![S^*]\!], [\![end]\!]m). \end{array}$$

This definition corresponds with the traditional definition for elaboration of declarations, where the declaration of a variable actualizes an environment of variables.

Semantics of statements

An execution of one step we define with the morphism $[\![next]\!]$:

$$[\![next]\!] : \mathbf{Config} \to \mathbf{Config}.$$

For an assignment statement and for the statement `skip`, this morphism is defined as follows:

$$[\![next]\!]([\![x := e, S^*]\!], m) = ([\![S^*]\!], m[((x, l), v) \mapsto ((x, l), [\![e]\!]m)])$$
$$[\![next]\!]([\![\texttt{skip}; S^*]\!], m) = ([\![S^*]\!], m)$$

In the sequence of statements, any statement can be executed in one or more steps. Hence we must to consider the following situations:

$$[\![next]\!]([\![S_i; S_{i+1}; \ldots; S_n]\!], m) = \begin{cases} ([\![S_{i+1}; \ldots; S_n]\!], m'), & \text{if } \langle S_i, m\rangle \Rightarrow m'; \\ ([\![S_i'; S_{i+1}; \ldots; S_n]\!], m'), & \text{if } \langle S_i, m\rangle \Rightarrow \langle S_i', m'\rangle. \end{cases}$$

Semantics of statements

An execution of conditional statement depends on a value of Boolean condition. Hence we define the first step of execution as follows:

$$[\![next]\!]([\![\texttt{if } b \texttt{ then } S'_i \texttt{ else } S''_i; S_{i+1}; \ldots; S_n]\!], m) = \begin{cases} ([\![S'_i; S_{i+1}; \ldots; S_n]\!], m), & \text{if } [\![b]\!]m = \mathbf{true}; \\ ([\![S''_i; S_{i+1}; \ldots; S_n]\!], m), & \text{if } [\![b]\!]m = \mathbf{false}. \end{cases}$$

The first step of execution of a loop is a transformation to the semantically equivalent conditional statement, and we define this step as follows:

$$[\![next]\!][\![\,(\texttt{while } b \texttt{ do } S'_i; S_{i+1}; \ldots; S_n, m)\,]\!] = [\![next]\!]\,([\![\texttt{if } b \texttt{ then } S'_i; \texttt{while } b \texttt{ do } S'_i \texttt{ else skip}; S_{i+1}; \ldots; S_n]\!], m)$$

Semantics of statements

For the statements of user input and output, we define appropriate representations of the operations in signature for configurations:

$$[\![input]\!]([\![\texttt{input}\ x; S_{i+1}; \ldots; S_n]\!], m) = \lambda v'.([\![S_{i+1}; \ldots; S_n]\!], m'),$$

where m' is an actualized memory

$$m' = m[((x, maxlevel(m)), v) \mapsto ((x, maxlevel(m)), v')],$$

where an input value v is stored into a memory cell allocated for the variable x on the highest nesting level.

The output statement does not change a memory, but it changes a configuration (output statement is removed from the sequence of statements):

$$[\![print]\!](\texttt{print}\ e; S_{i+1}; \ldots; S_n]\!], m) = ([\![e]\!]m, ([\![S_{i+1}; \ldots; S_n]\!], m))$$

Semantics of statements

During the execution of a block statement, an elaboration of a special variable `begin` is elaborated:

$$[\![next]\!]([\![\texttt{begin}\ D^*; S'_i\ \texttt{end}; S_{i+1}; \ldots; S_n]\!], m) = \\ ([\![D^*; S'_i\ \texttt{end}; S_{i+1}; \ldots; S_n]\!], [\![begin]\!]m)$$

and an execution of a block statement continues by elaboration of real local variables (if present) and by execution of statements inside a block.

For an interruption of program execution we define a function

$$abort : \mathbf{Config} \to \mathbf{Config},$$

which immediately terminates running program which ends in an undefined configuration $config_\perp$:

$$abort(config) = (\varepsilon, m_\perp),$$

Base category for coalgebra

We construct a category $\mathscr{C}onfig$ where:

- configurations $config = (\llbracket D^*; S^* \rrbracket, m)$ are category objects;
- mappings $\llbracket next \rrbracket, \llbracket input \rrbracket, \llbracket print \rrbracket$ and $abort$ are category morphisms;
- we define an identity morphism for each configuration;
- composition of morphisms and associativity of composition holds.

An undefined configuration is also a terminal (final) object of a category:

$$config_\perp = (\varepsilon, m_\perp)$$

because there exists a possibility to abrupt the program execution from any configuration.

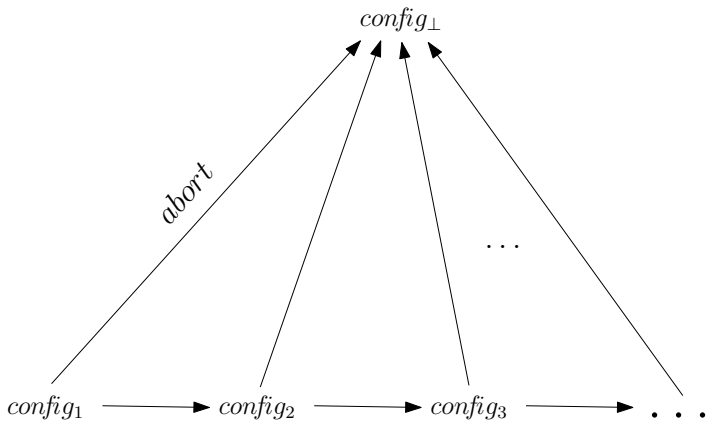
Colimit in base category

In case of infinite loop we must to guarantee that an execution of such a statement cannot be defined in category. Hence we require that each infinite path in category must to have a colimit:

- Let $config_1 \to config_2 \to config_3 \to \ldots$ be an infinite path (diagram $\mathcal{D}$) in category;
- object $config_\perp$ is a colimit of diagram $\mathcal{D}$

$$\text{colimit } \mathcal{D} = \bigcirc_{i \in N} [\![next]\!]([\![\texttt{while } b \texttt{ do } S; S^*]\!], m)$$

if from every object exists a morphism into this object.



Coalgebra

- We have constructed base category $\mathscr{C}onfig$, which objects are configurations $config$ and morphisms are $[\![next]\!], [\![input]\!], [\![print]\!]$.
- We construct an appropriate form of polynomial endofunctor which will characterize this kind of systems:

$$Q : \mathbf{Config} \to \mathbf{Config},$$
$$Q(\mathbf{Config}) = 1 + \mathbf{Config} + O \times \mathbf{Config} + \mathbf{Config}^I.$$

- Coalgebra is defined as $c : X \to FX$ in general; we assign our state space and morphisms, then a coalgebra for programs in language *E-Jane* has a form:

$$\langle abort, [\![print]\!], [\![next]\!], [\![input]\!] \rangle : \mathbf{Config} \to Q(\mathbf{Config}).$$

Coalgebra models the behavior of a program such that in each step one of the following alternatives occurs:

- $Q(\mathbf{Config}) = 1$ – program aborts;
- $Q(\mathbf{Config}) = \mathbf{Config}$ elaborates a declaration/executes a statement;
- $Q(\mathbf{Config}) = O \times \mathbf{Config}$ provides an output value;
- $Q(\mathbf{Config}) = \mathbf{Config}^I$ takes an input value.

Example

We consider a program in language *E-Jane*:

```
var x; var y;
input x; input y;
if x <= y then begin
                  var z;
                  z:=x; x:=y; y:=z
                end
          else skip;
print x
```

For simplicity we introduce the following substitutions:

$$
\begin{aligned}
&D_1 = \texttt{var}\ x;\ D_2 = \texttt{var}\ y;\\
&S_1 = \texttt{input}\ x;\ S_2 = \texttt{input}\ y;\\
&S_3 = \texttt{if}\ x <= y\ \texttt{then begin var}\ z;\\
&\quad z := x; x := y; y := z\ \texttt{end else skip}\\
&S_4 = \texttt{print}\ x
\end{aligned}
$$

and we consider values $\mathbf{3}$ and $\mathbf{5}$ for variables x and y, resp.

Example

An initial configuration is $config_0 = (\llbracket D_1; D_2; S_1; S_2; S_3; S_4 \rrbracket, m_0)$.
Every application of functor Q represents one step of program execution:

$$Q(config_0) = \llbracket next \rrbracket(config_0) =$$
$$= (\llbracket D_2; S_1; S_2; S_3; S_4 \rrbracket, \llbracket var\ x \rrbracket(x, m_0)) =$$
$$= config_1$$

$$Q(config_1) = \llbracket next \rrbracket(config_1) =$$
$$= (\llbracket S_1; S_2; S_3; S_4 \rrbracket, \llbracket var\ x \rrbracket(y, m_0)) =$$
$$= config_2$$

$$Q(config_2) = \llbracket input \rrbracket(\llbracket S_1; S_2; S_3; S_4 \rrbracket, m_0) =$$
$$= (\llbracket S_2; S_3; S_4 \rrbracket, m_1) = config_3$$

$$Q(config_3) = \llbracket input \rrbracket(\llbracket S_2; S_3; S_4 \rrbracket, m_1) =$$
$$(\llbracket S_3; S_4 \rrbracket, m_2) = config_4$$

Example

$$Q(config_4) = [\![next]\!]([\![\textbf{begin var } z; z := x; x := y; y := z \textbf{ end}; S_4]\!], m_2) =$$
$$([\![\textbf{var } z; z := x; x := y; y := z \textbf{ end}; S_4]\!], [\![begin]\!]m_2) = config_5$$

$$Q(config_5) = [\![next]\!]([\![\textbf{var } z; z := x; x := y; y := z \textbf{ end}; S_4]\!], m_2) =$$
$$([\![z := x; x := y; y := z \textbf{ end}; S_4]\!], [\![\textbf{var } x]\!](z, m_3) = config_6$$

$$Q(config_6) = [\![next]\!](z := x; x := y; y := z \textbf{ end}; S_4]\!], m_3) =$$
$$([\![x := y; y := z \textbf{ end}; S_4]\!], m_4) = config_7$$

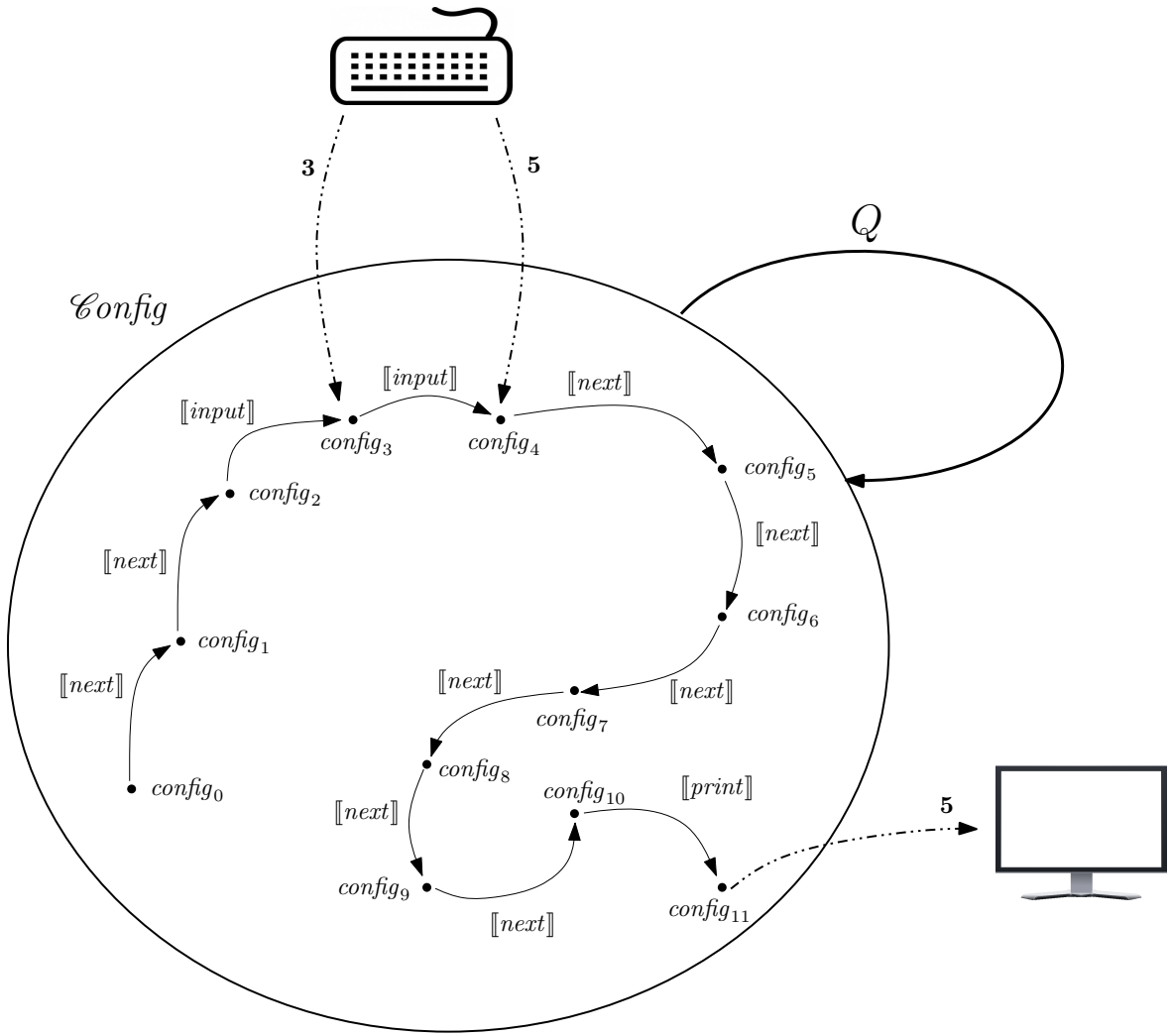
$$Q(config_7) = [\![next]\!]([\![x := y; y := z \textbf{ end}; S_4]\!], m_4) =$$
$$= ([\![y := z \textbf{ end}; S_4]\!], m_5) = config_8$$

$$Q(config_8) = [\![next]\!]([\![y := z \textbf{ end}]\!], m_5) = ([\![\textbf{end}; S_4]\!], m_6) = config_9$$

$$Q(config_9) = [\![next]\!]([\![\textbf{end}; S_4]\!], m_6) = ([\![S_4]\!], [\![end]\!]m_6) = config_{10}$$

$$Q(config_{10}) = [\![print]\!]([\![S_4]\!], m_6) = (\mathbf{5}, (\varepsilon, m_6)) = config_{11}$$

Example


3
5
Q
Config
⟦input⟧
⟦input⟧
⟦next⟧
config3
config4
config5
config2
⟦next⟧
⟦next⟧
config6
config1
⟦next⟧
⟦next⟧
⟦next⟧
config7
config8
config0
config10
⟦print⟧
⟦next⟧
5
config9
config11
⟦next⟧

Thank you for your attention.