



# URM - Unlimited Register Machine

## A basic C# implementation

2015-10-02

### 1 Abstract

This paper briefly describes the **design choices** of a simple URM processor as described in the Computability class. During the process some interesting design issues arised and we are discussing them here.

The design is very simple and the implementation is far from being robust and well tested (this is no production code). Please refrain from production code considerations, such as thread safety, performance, fine grained error handling and such comments. We already know.

Funny story is that the model was implemented just to have a mean to test the logic of a simple program P we were writing on paper as a study session. Since we wanted an automated check on our assumptions "is this program returning the expected results?", we decided to implement the model.

By the way P was incorrect, for that matters ... (^\_^);

### 2 The URM model

The URM is defined as a **machine** (the processor) with an **unlimited tape** (the storage memory) that operates on a **program** in order to solve a **problem**. The problem is an unary function :  $N \rightarrow N$ .

#### 2.1 The Tape

The tape has "unlimited" squares (of course, in the software model it is limited by the underlying implementation : we used a List<long> having max 4G squares). Each square contains a non negative integer value (implemented as a 64 bit long).

The tape ensures the **square value constraints** (for example values  $\geq 0$ ) and exposes two methods only : **GetValue** and **SetValue**. The tape seems "unlimited" to the client because it **expands as needed** when new values are added or retrieved (aka "touched").

For instance: `tape.Set(5,1000)` (set the value 5 in the square at position 1000) : the tape silently expands to the 1000<sup>th</sup> position so that the client can use that position.

This is because we want the client to be able to do things such as :

- `z(5)`
- `s(25)`
- ...

... even if the underlying 5<sup>th</sup> and 25<sup>th</sup> squares do not exist : there is no need for the client to declare the existence of a square.

#### 2.2 The Program

The program is an **ordered list of statements**. The URM statement set is limited to four statements but our implementation is meant to be **extendable** : we can easily implement new statements, for instance to define a more evolved machine.

The implemented **statements** are the following :

- `z(n)` : the **zero** statement
- `s(n)` : the **increment** statement
- `t(n,m)` : the **transfer** function
- `c(n,m,q)` : the **conditional** statement

We also implemented a non standard URM statement which sets an arbitrary value in a square. More on this on the Initial Conditions chapter.



## 2.3 The Machine

The machine **defines a tape** and **expects a program** as an argument of the execute method.

The machine is **immutable**, which means that the Program Counter, which is supposed to points to the next statement, is not defined as a machine state field. Instead, it is defined on the stack as a local variable of the execute method. This is a design choice and it allows the machine to be thread safe, at least with regards of the program counter so far. We don't care of thread safety here, but immutability is always a good goal.

The lack of the program counter state, results in a problem for the statements which need to alter it during the program execution (i.e. the conditional statement). We have to deal with whose responsibility is to update the program counter. More on this on the program counter chapter.

Since we need to **track the loop condition**, we defined a **MaxStep counter** and we state that a machine is not looping if the processing finishes in a number of steps which is limited to MaxSteps.

If the loop condition is reached, the machine throws an LoopDetectedException. This is of course a design choice in order to be able to determine if the machine is looping or not (for example, unit tests need to "solve" the halting problem in a pragmatic way).

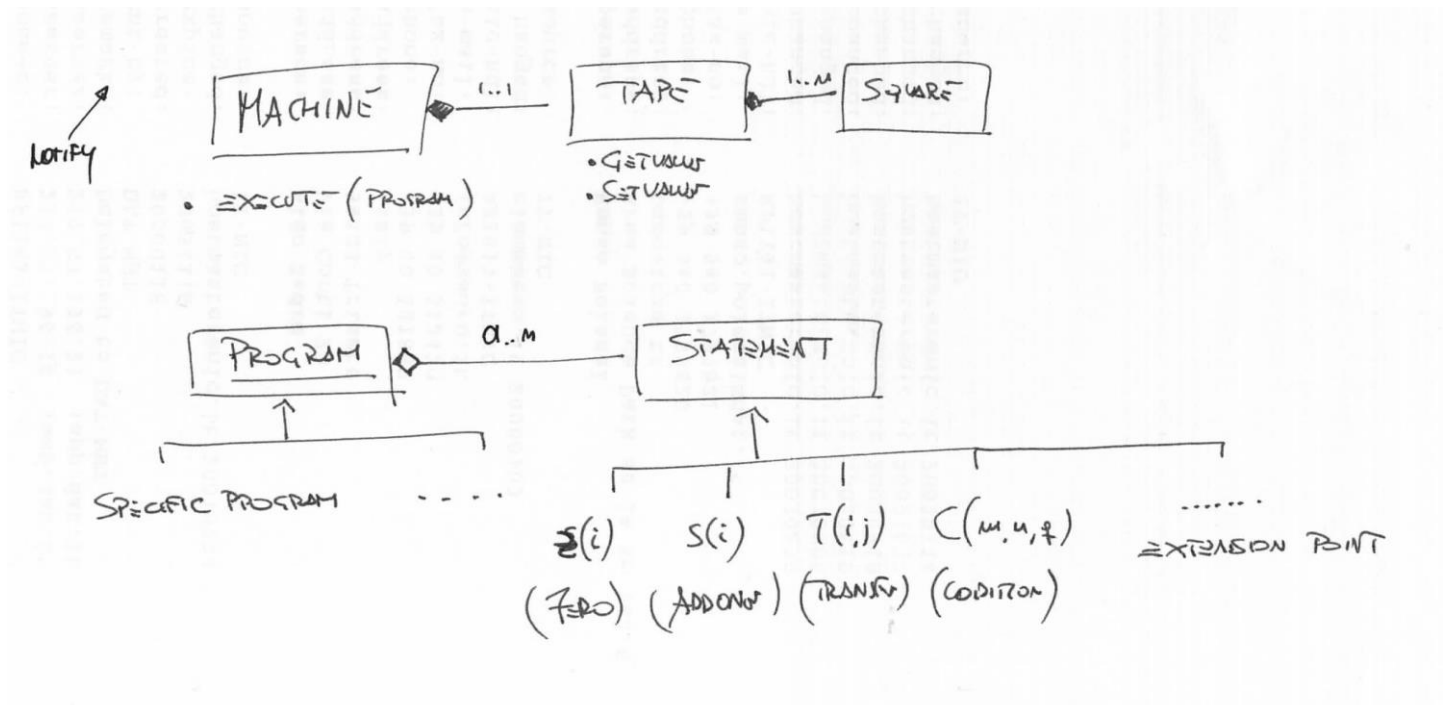
## 3 The implementation

The URM is implemented in OOP fashion with some functional coding style, such as a tendency to immutability of types.

Here we don't deal with a detailed code implementation, because it is easier to inspect the source code : we just point out some basic information and use cases.

### 3.1 The Class Diagram

The simplified class diagram :



### 3.2 The statements

We implemented the following basic standard URM statements :

- $z(n)$  : CStatementZero



- s(n) : CStatementAddOne
- t(n,m) : CStatementTransfer
- c(n,m,q) : CStatementCondition

We also implemented the following non-standard URM statement :

- set(n, v) : CStatementSet

### 3.3 The program

A concrete program can be coded by inheriting from an abstract program and by defining its statement list, as in the following example :

```
public sealed class CProgramXMinus1 : CAbstractProgram
{
    // =====
    // F(x) =
    //      : x - 1      when x > 0
    //      : 0          when x == 0
    // =====
    public CProgramXMinus1() : base (
        new IStatement[]
        {
            /* 01 */ new CStatementCondition(1,4,99), // x == 0 --> jpm end
            /* 02 */ new CStatementAddOne(3),         // k++
            /* 03 */ new CStatementCondition(1,3,7),   // x == k+1 --> jpm 7
            /* 04 */ new CStatementAddOne(2),         // k++
            /* 05 */ new CStatementAddOne(3),         // (k+1)++
            /* 06 */ new CStatementCondition(3),       // jpm 3
            /* 07 */ new CStatementTransfer(2,1)       // mov 2 -> 1
        }
    )
}
```

### 3.4 The client code

The client can code a program and invoke its execution on a machine with the following code :

```
try
{
    IProgram program = new CProgramXMinus1();
    long x = 1;
    long lResult = new CMachine(new long[]{x}).Execute(program);
}
catch (CMachineInLoopException ex)
{
}
```

The code :

- creates a specific program
- creates a machine passing the initial condition (an IEnumerable of values)
- invokes the execute method on the machine, passing the program
- retrieve the result of the execute method, which is the result of computation
- if the program enters a loop (which is a simulated loop condition) the machine throws a LoopDetectedException.

## 4 Points of interest

### 4.1 Initial Condition

The URM defines as "**initial condition**" the input values only. This is somehow counter-intuitive to us, since the "condition" term might refer to the initial "state" (i.e. the state of the entire tape : both the input values and the initial state). In URM, the tape defaults to zero in each and every cell, and before the processing, the input values are set in the leftmost squares.

In order to implement the definition of such initial condition, we can follow two strategies :



- we can define a new non-URM standard statement – say SetValue(n, v) - which allows us to explicitly set a value in a tape square
- we can pass an array of values to the machine before executing the program. The machine asks the tape to load such values.

#### 4.1.1 Initial condition defined by the Set Statement

We used this strategy mainly to try the extensibility feature of the statements as we are aware this is not the best way to set the initial condition.

If we set the initial conditions by executing such statements :

- those statements belong to the program, which is held responsible for defining its initial condition.
- as a consequence, we couple the initial condition to the program, which means that the program is not “parametric” any more (essentially we are killing the ability of the program to start in any initial condition)
- on the other hand, executing a program in any initial condition can lead to unexpected results, because the program itself “encodes” the squared being used and their semantics.

#### 4.1.2 Initial condition defined by the client before the program execution

This strategy is better because it reflects the URM model : since the model states that the initial condition is defined by setting values starting from the left side of the tape, we can simply pass the machine an ordered set of values during its construction. The machine will then ask the tape to load them one by one.

This way, the program does not need the SetValue statement any more and it is not held responsible for setting the initial condition. As a consequence, the program can run in any initial condition, which can also lead to unexpected results.

### 4.2 The program counter (PC)

The program execution is implemented by the machine by :

- setting a program counter to 1 (the PC is an integer “pointer” to the next execution waiting for execution)
- asking the program to fetch the next execution statement
- executing the next execution statement
- moving the program counter to the “next statement”

#### 4.2.1 Responsibility of the program counter updates

We need to decide who is **responsible** for the program counter update. It could reasonably be either the **machine** or the **statement**. This is a design task and we face more than one design choice.

##### 4.2.1.1 Statements update the PC

Since the program counter is incremented after some types of statements (e.g. arithmetic statements) and it is updated in a different fashion after other types of statements (e.g. conditional statements), we might be tempted to state that updating the PC is a responsibility of the statement.

The default statement behaviour might be “PC++” which might be overridden by statements in need of a different behaviour.

At first, this doesn’t sound so bad but we should notice that:

- the machine needs a mutable state (the PC), which we would like to avoid
- such a state must be observable and updatable by the statement and not by other types. As a consequence, since we don’t have “friend types” in c#, we should encapsulate the statement types as nested types of the machine type. There is nothing wrong about that, apart from unnecessary complexity.

In other words, this way we couple the statements to the machine executing them and we do it more than necessary. Of course if we state that a statement S can be executed by the machine M only, such a coupling is by design. However, for sake of genericity and simplicity we would like to avoid that.

##### 4.2.1.2 The Machine updates the PC

We think a better design would be to let the machine updating the PC using :

- a default behaviour (i.e. PC++)



- using the information provided by the statements after their execution, if they provide it (PC = statement result)

By doing so, we can let the machine be immutable (no PC state) in a more "functional" coding style.

Each statement returns a `StatementResult` which may or may not encapsulate the next PC value. If such a value is defined, we use it, otherwise we use the default behaviour (PC++) which is defined by the machine (as opposed to be defined by the statement)

### 4.3 The halting problem : faking the loop

In order to pragmatically "solve" the halting problem, we can state that "after N steps, the machine is meant to have entered a loop". The bigger the N, the less the difference from the "fake" and the "real" halting problem.

Of course by limiting the don't "solve the halting problem" (^ ^); ... we simply make sure that our machine computation, which will defined very simple actual computations, always ends.

In order do do so, we define a `Statement Instance` counter and we increment it after statement execution. A guard throws an exception when such a counter exceeds the maximum allowed steps for the machine.

### 4.4 Machine state notification

Since we can to **debug** our written-on-paper-program, we are very interested in inspecting the state of the machine after each statement execution.

To satisfy this requirement, the machine notifies its state :

- before the first statement execution (i.e. the initial condition)
- after each statement execution

The notification is implemented by the **observer pattern**, simplified by the DotNet event notion.

A client can subscribe such an event in order to be notified of the machine state after each execution step.

## 5 Addenda

### 5.1 Addendum 1 : the URM-program text parser

Right now, since we defined a mini framework, we can use it to **define a program by c# code**. Since the project aims to simulate a URM which can be fed with many programs, coding an URM-program in c# is not very friendly on a write-run-and-test basis.

We want to be able to write **URM programs as text files**, feed them into a command line application and let the application to parse them into a DotNet `IProgram` object, before executing them.

To do so, we define a simple parser which creates an `IProgram` by parsing a text containing URM statements. The parser is not relevant to this paper, therefore it is implemented in a quick-and-dirty fashion and as such it is not documented here.

#### 5.1.1 The statements

The parser is able to parse the URM statements as we write them in our `Computability` class, therefore :

- "z(n)" --> n being a non negative integer
- "s(n)"
- "t(m,n)"
- "c(m,n,q)"

By doing to, we define an invertible function that translates from a text to the equivalent `IStatement` and viceversa.

#### 5.1.2 The statement separators

The statements can be separated by ";" or by a line feed. While the line feed is useful in text files containing urm programs, the ";" is useful if we want to pass a small program on the command line.



### 5.1.3 Spaces and other characters

Spaces are ignored. Characters that are not used for defining the four statements make the parser raise a parsing exception.

### 5.1.4 Example

For example, we want to be able to compile the following :

```
string source = "z(1);s(2);t(1,3);c(1,2,3);";
IProgram program = new CProgramParser().Parse(source);
```

## 5.2 Addendum 2 : the URM-program text parser

Since we are now able – by Addendum 1 - to parse a text into a IProgram, we can write a console application which is fed a text program, executes it, shows the intermediate states and prints the processing result.

### 5.2.1 Command line arguments

We need to feed the application the following arguments :

- the program : can be a file path or an inline program
- initial value 1
- initial value 2
- .....
- initial value i

### 5.2.2 Example

Given the  $f(x)=x-1$  program, saved into the file "programs\XMinus1.urm" :

```
c(1,4,99)
s(3)
c(1,3,7)
s(2)
s(3)
c(1,1,3)
t(2,1)
```

We can execute it as follows : (we pass  $X = 0$ ) :  $0-1 = 0$  (by definition)

```
D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>com.scrigner.RAMModel.exe Programs\XMinus1.urm 0
#STEP STIDX      TAPE
00000 00001 c(1,4,99)  000
00001 00099      000 000 000 0[004.192]
Execution Completed. Result=[0]

D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>
```

The result is the shown table :

- the first column shows the step of the computation
- the second column shows the program counter (i.e. the next instruction address)
- the third column shows the instruction located at the program counter address
- the following columns show the touched tape values

In the previous example, we can see that, since  $x1 == x4$  ( $= 0$ ) we jumped to 99 : the result in X1 (i.e. 0).

We change the argument to 1 :  $1-1 = 0$





```
D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>com.scrigner.RAMModel.exe Programs\XMinus1.urm 1
#STEP STIDX      TAPE
00000 00001 c(1,4,99) 001
00001 00002 s(3)      001 000 000 000
00002 00003 c(1,3,7)  001 000 001 000
00003 00007 t(2,1)    001 000 001 000
00004 00008          000 000 001 000
Execution Completed. Result=[0]

D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>
```

We get a longer path through the program :

- we don't jump to 99, since  $1 \neq 0$
- we increment X3
- since  $X1 == X3$  we jump to I7
- we move X2 to X1
- we return  $X1 = 0$

Just another example : we change the argument to 3 :  $3 - 1 = 2$

```
D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>com.scrigner.RAMModel.exe Programs\XMinus1.urm 3
#STEP STIDX      TAPE
00000 00001 c(1,4,99) 003
00001 00002 s(3)      003 000 000 000
00002 00003 c(1,3,7)  003 000 001 000
00003 00004 s(2)      003 000 001 000
00004 00005 s(3)      003 001 001 000
00005 00006 c(1,1,3)  003 001 002 000
00006 00003 c(1,3,7)  003 001 002 000
00007 00004 s(2)      003 001 002 000
00008 00005 s(3)      003 002 002 000
00009 00006 c(1,1,3)  003 002 003 000
00010 00003 c(1,3,7)  003 002 003 000
00011 00007 t(2,1)    003 002 003 000
00012 00008          002 002 003 000
Execution Completed. Result=[2]

D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>
```

Here we can see iterations.

As a loop detection example, we can use the program loop.urm :



```
D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>com.scrigner.RAMModel.exe Programs\Loop.urm
#STEP STIDX TAPE
00000 00001 c(1,1,1)
00001 00001 c(1,1,1)
00002 00001 c(1,1,1)
00003 00001 c(1,1,1)
00004 00001 c(1,1,1)
00005 00001 c(1,1,1)
00006 00001 c(1,1,1)
00007 00001 c(1,1,1)
00008 00001 c(1,1,1)
00009 00001 c(1,1,1)
00010 00001 c(1,1,1)
Execution halted in a loop.

D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>
```

If we can execute 10 steps at most (value we set before executing the code), we enter a loop and the program ends in a loop condition.

As a last example, we execute the XMinusY program :

```
D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>com.scrigner.RAMModel.exe Programs\XMinusY.urm 5 2
#STEP STIDX TAPE
00000 00001 c(1,2,5) 005 002
00001 00002 s(2) 005 002
00002 00003 s(3) 005 003
00003 00004 c(1,1,1) 005 003 001
00004 00001 c(1,2,5) 005 003 001
00005 00002 s(2) 005 003 001
00006 00003 s(3) 005 004 001
00007 00004 c(1,1,1) 005 004 002
00008 00001 c(1,2,5) 005 004 002
00009 00002 s(2) 005 004 002
00010 00003 s(3) 005 005 002
00011 00004 c(1,1,1) 005 005 003
00012 00001 c(1,2,5) 005 005 003
00013 00005 t(3,1) 005 005 003
00014 00006 003 005 003
Execution Completed. Result=[3]

D:\Research\Univ\ComputabilitaELinguaggi\com.scrigner.RAMModel\com.scrigner.RAMModel\bin\Debug>
```