User oriented language for powerful data mining with Ferda

Michal Kováč

December 6, 2007

Table of contents part 1/2

Table of contents part 1/2

1. Introduction
    - Table of contents
    - What is Ferda Data Miner?
    - Ferda as programming language
    - What is missing? What should be done?
2. New functionality in Ferda
    - Network archive
    - Boxes for math
    - Lambda expression
    - Other new boxes

Table of contents part 2/2

3. Example – executing four fold task recursively
   - Motivation
   - Linear interpolation
   - Connection of boxes
   - Results

Table of contents part 2/2

3. Example – executing four fold task recursively
   - Motivation
   - Linear interpolation
   - Connection of boxes
   - Results
4. Next steps
   - Sequences and sets
   - Reuse of code
   - Better lambda
   - Summary

Ferda

What is Ferda?

- User oriented application

Ferda

What is Ferda?

- User oriented application
- For specification of tasks, execution and result browsing

Ferda

What is Ferda?

- User oriented application
- For specification of tasks, execution and result browsing
- Works with boxes

Ferda

What is Ferda?

- User oriented application
- For specification of tasks, execution and result browsing
- Works with boxes
- Ferda Data Miner = Ferda + boxes for data mining

Ferda

What is Ferda?

- User oriented application
- For specification of tasks, execution and result browsing
- Works with boxes
- Ferda Data Miner = Ferda + boxes for data mining

History

Ferda

What is Ferda?

- User oriented application
- For specification of tasks, execution and result browsing
- Works with boxes
- Ferda Data Miner = Ferda + boxes for data mining

History

- LISp-Miner not so user frienly

Ferda

What is Ferda?

- User oriented application
- For specification of tasks, execution and result browsing
- Works with boxes
- Ferda Data Miner = Ferda + boxes for data mining

History

- LISp-Miner not so user frienly
- Software project at MFF

Ferda

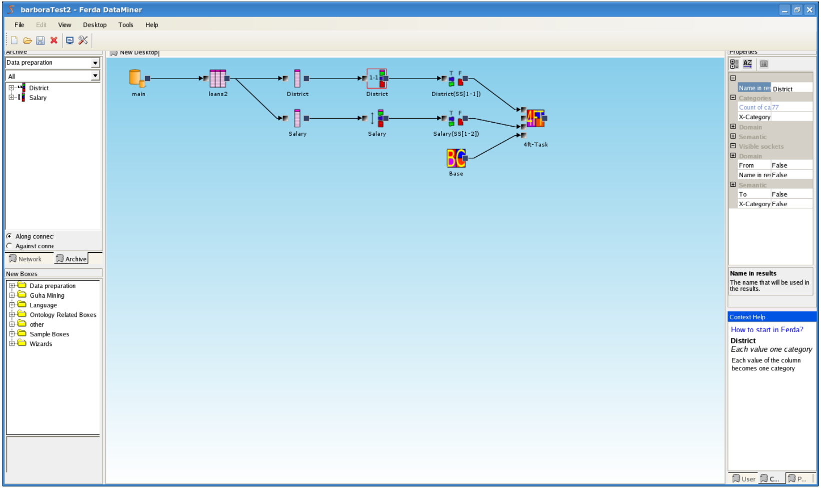
What is Ferda?

- User oriented application
- For specification of tasks, execution and result browsing
- Works with boxes
- Ferda Data Miner = Ferda + boxes for data mining

History

- LISp-Miner not so user frienly
- Software project at MFF
- Master theses

Screenshot



Name in res District
Categories
Count of ca 77
X-Category
Domain
Semantic
Visible sockets
Domain
From False
Name in res False
Semantic
To False
X-Category False

Name in results
The name that will be used in the results.

Context Help
How to start in Ferda?
District
Each value one category
Each value of the column becomes one category

Ferda as programming language

Box as function

Ferda as programming language

Box as function

- Box consists of functions

Ferda as programming language

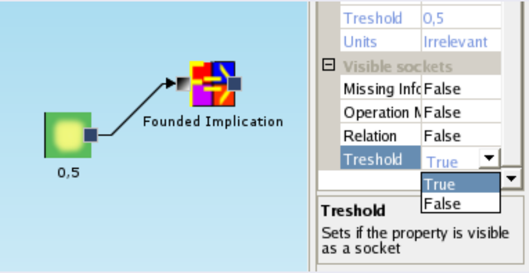
Box as function

- Box consists of functions
- Sockets are parameters of these functions

Ferda as programming language

Box as function

- Box consists of functions
- Sockets are parameters of these functions
- Property is also socket



Treshold	0,5
Units	Irrelevant
⊟ Visible sockets	
Missing Infc	False
Operation N	False
Relation	False
Treshold	True

True
False

Treshold

Sets if the property is visible as a socket

What is missing?

What is missing?

- Moving work from one project to another

What is missing?

What is missing?

- Moving work from one project to another
- Basic math boxes

What is missing?

What is missing?

- Moving work from one project to another
- Basic math boxes
- Recursion

What is missing?

What is missing?

- Moving work from one project to another
- Basic math boxes
- Recursion
- Other language boxes

What is missing?

What is missing?

- Moving work from one project to another
- Basic math boxes
- Recursion
- Other language boxes
- Ferda specific language boxes

What is missing?

What is missing?

- Moving work from one project to another
- Basic math boxes
- Recursion
- Other language boxes
- Ferda specific language boxes
- Data mining specific boxes for user programming

Network archive

What is network archive?

- New place where user can store connection

Network archive

What is network archive?

- New place where user can store connection
- Independent on project

Network archive

What is network archive?

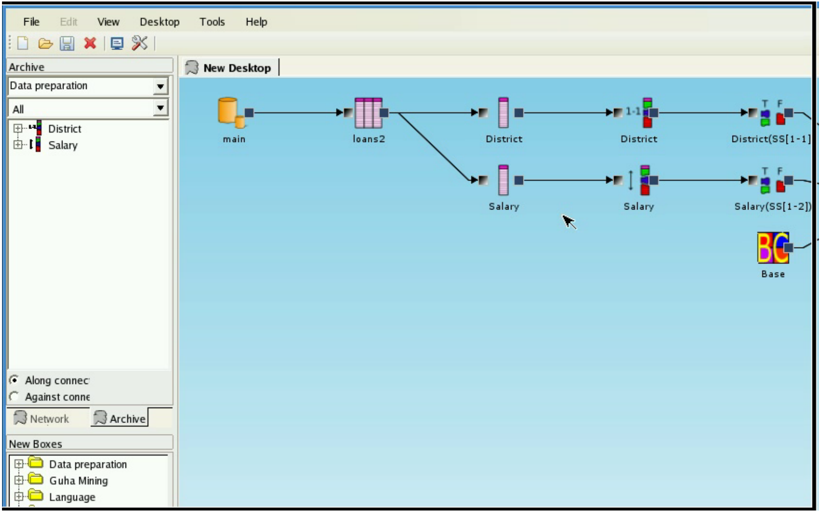
- New place where user can store connection
- Independent on project
- One network archive can be accessed from more computers

Network archive

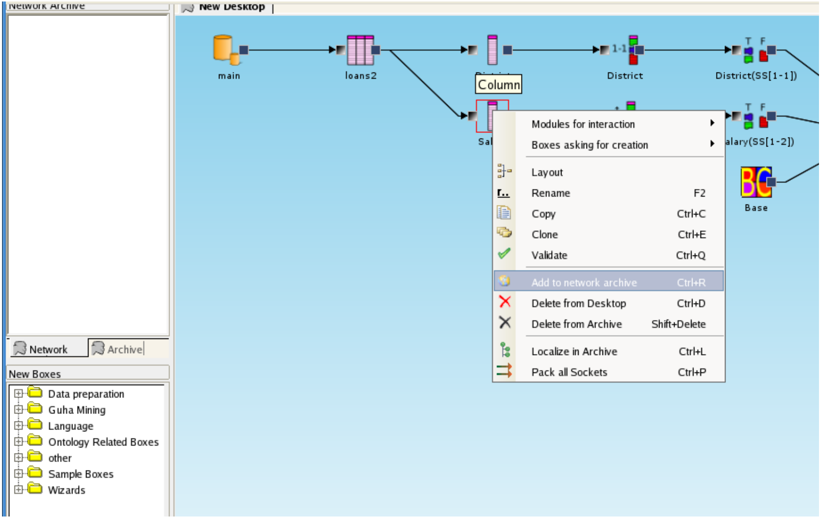
What is network archive?

- New place where user can store connection
- Independent on project
- One network archive can be accessed from more computers
- Way how to move connections from one project to another

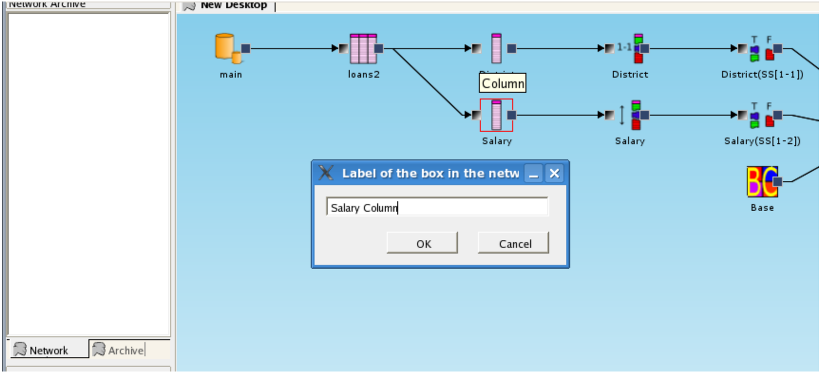
Movie – network archive



Screenshot – add a connection to the network archive



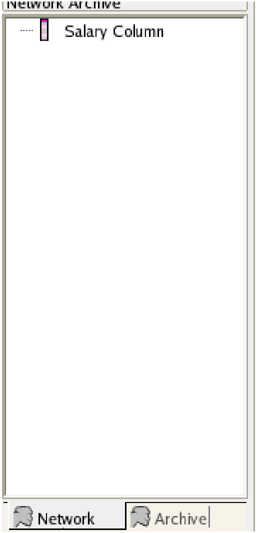
Screenshot – set a name of box in the network archive



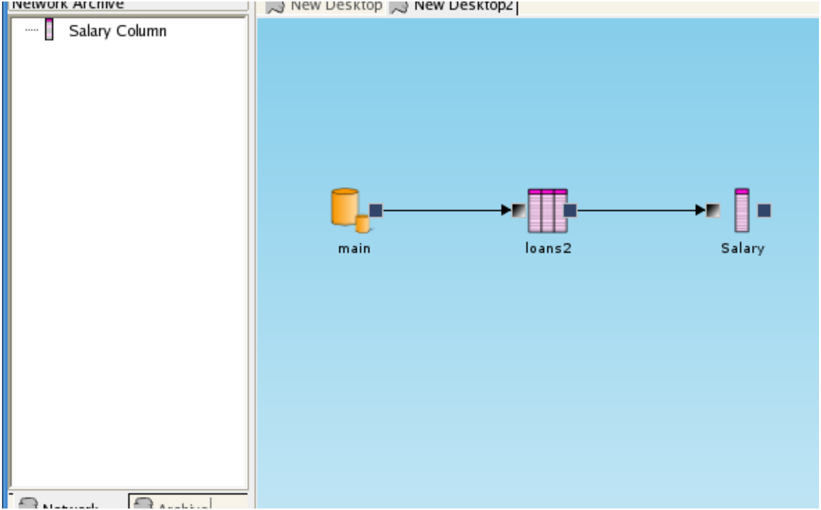
Label of the box in the netw

Salary Column

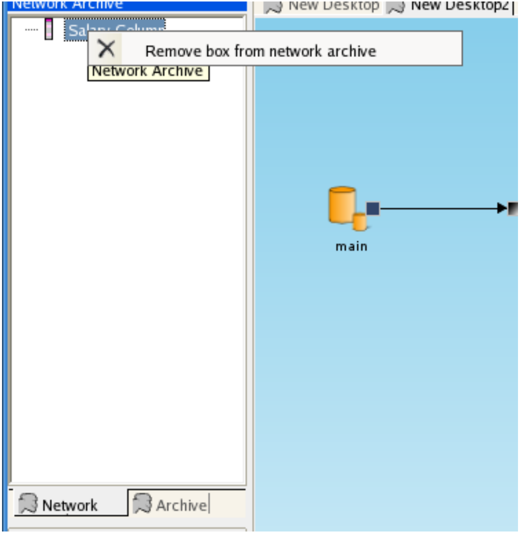
Screenshot – new box added to the network archive



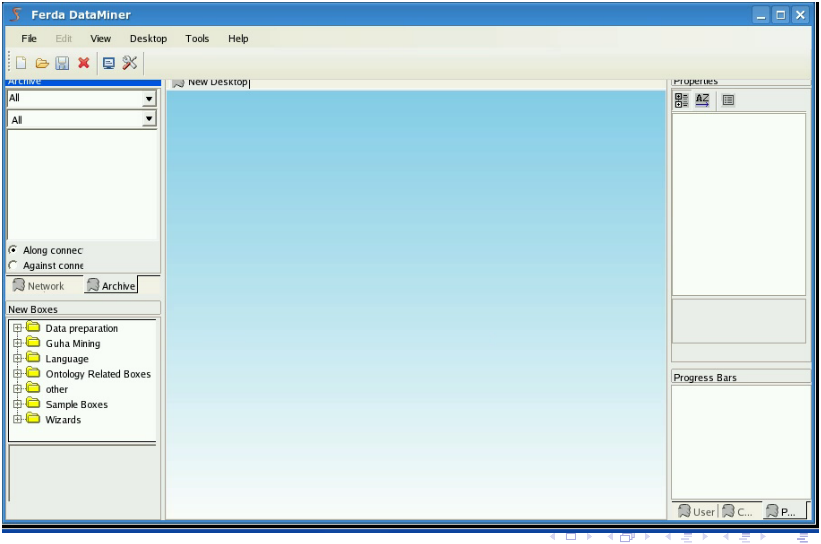
Screenshot – drop box to a desktop from the n. archive



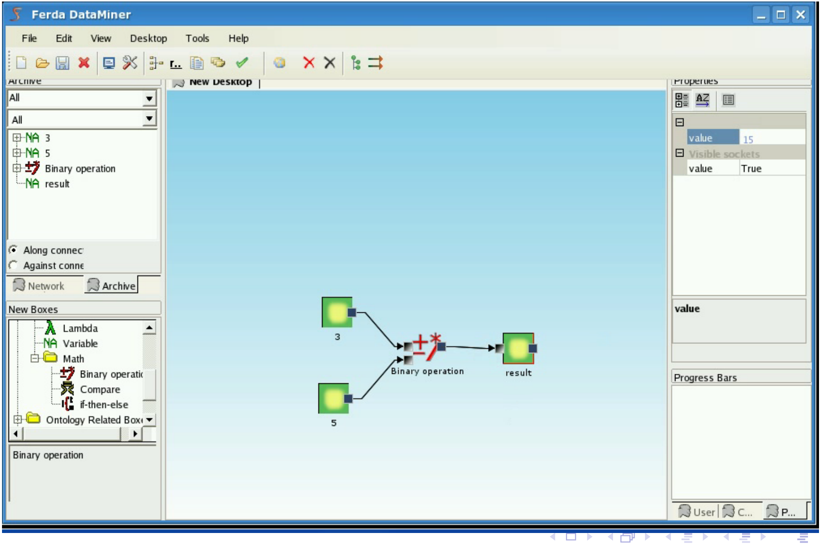
Screenshot – remove box from the network archive



Movie – Binary operation



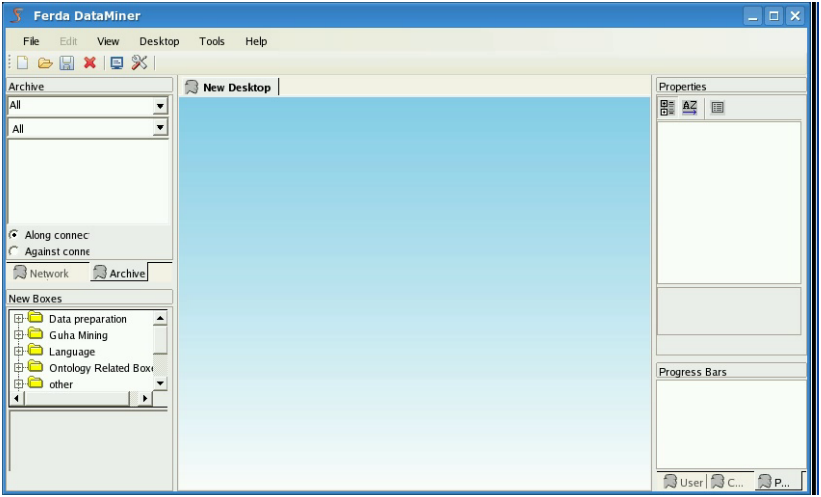
Binary operation



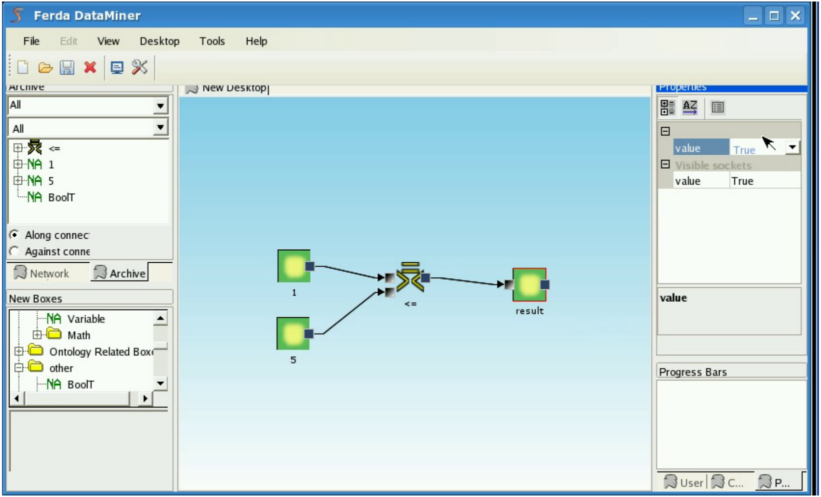
Properties:

value	15
Visible sockets	
value	True

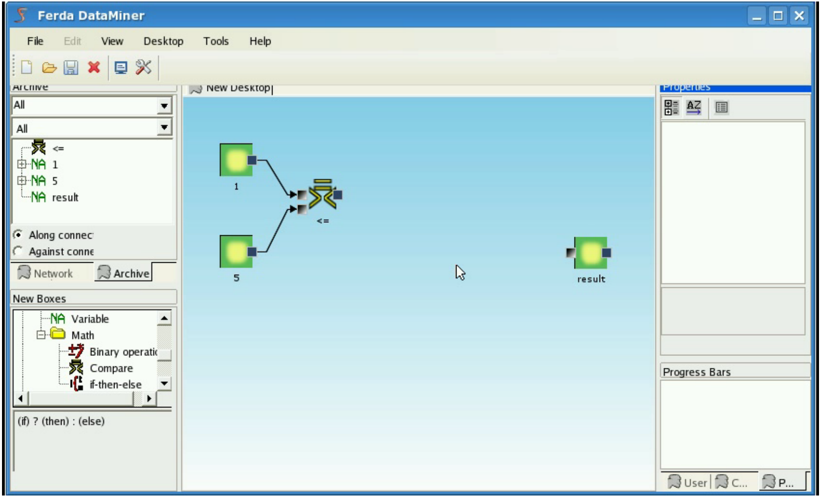
Movie – Comparision



Comparision

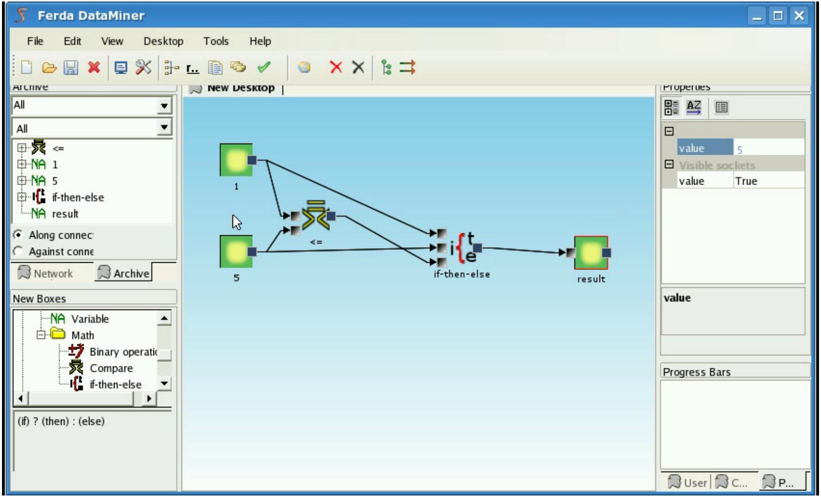


Movie – If expression



(if) ? (then) : (else)

If expressions



(if) ? (then) : (else)

value	5
Visible sockets	
value	True

Lambda expression

Basic facts

- From lambda calculus $(\lambda x.(1 + x))(9)$

Lambda expression

Basic facts

- From lambda calculus $(\lambda x.(1 + x))(9)$
- Basics of functional programming

Lambda expression

Basic facts

- From lambda calculus $(\lambda x.(1+x))(9)$
- Basics of functional programming

Lambda in C# 3

```
public delegate int function(int x);

public static void Main(string[] args)
{
  function plusOne = x => 1 + x;
  var a = plusOne(9);
  System.Console.WriteLine(a);
}
```

Lambda expression

Other languages

Lambda in F#

```
let onePlus x = 1 + x
do printf "%s" (onePlus(9))
```

Lambda expression

Other languages

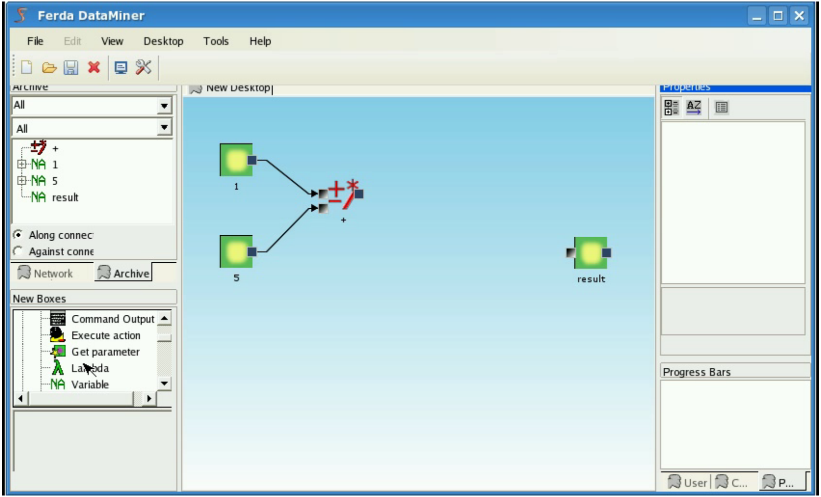
Lambda in F#

```
let onePlus x = 1 + x
do printf "%s" (onePlus(9))
```

Lambda in Python

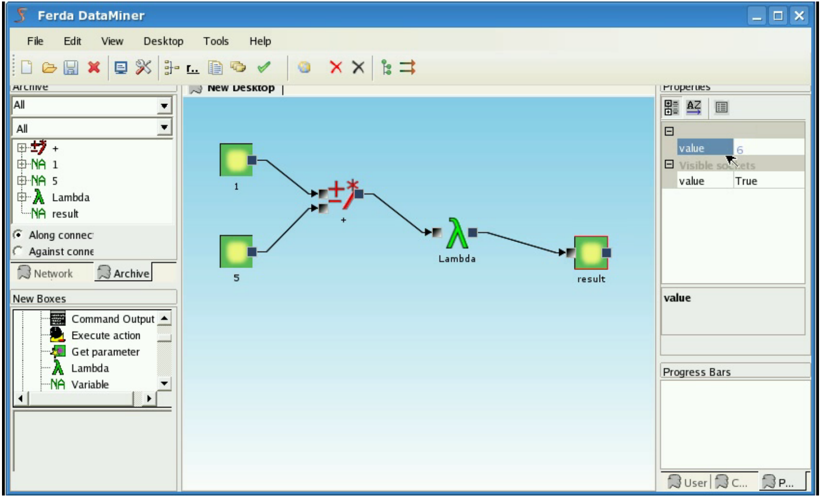
```
plusOne = lambda x: 1 + x
print plusOne(9)
```

Movie – Lambda basics in Ferda



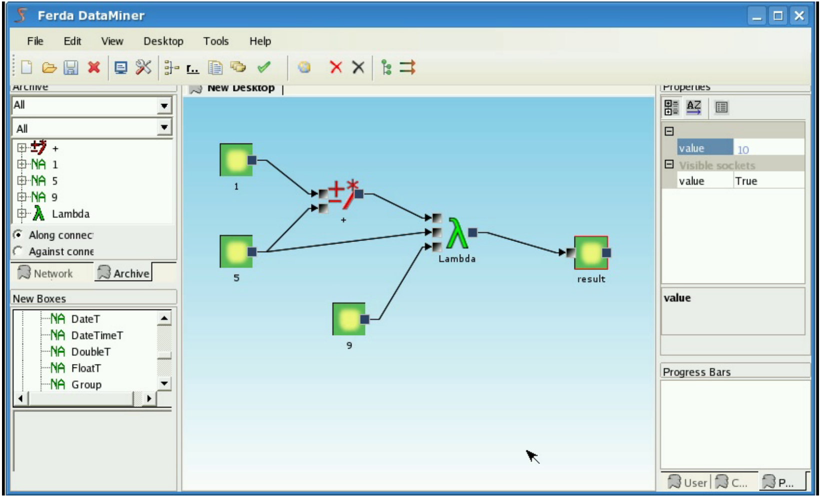
Basic use in Ferda

Lambda without parameters



Basic use in Ferda

One constant parameter specified



Implementation of lambda

How it really works

Algoritm

- Values of variables are cloned (whole subtree)

Implementation of lambda

How it really works

Algoritm

- Values of variables are cloned (whole subtree)
- Main function is cloned with substitution and returned

Factorial in C#

Structural version

First version of factorial

```
public static int Factorial(int x)
{
  if (x == 0)
  {
    return 1;
  }
  else
  {
    return x * Factorial(x - 1);
  }
}
```

Factorial in C#

Expresion version

Second version of factorial

```
public static int Factorial2(int x)
{
  return (x == 0) ? 1 : x * Factorial2(x - 1);
}
```

Factorial in other languages

Python

```
fac = lambda x: x == 0 and 1 or x * fac(x - 1)
```

Factorial in other languages

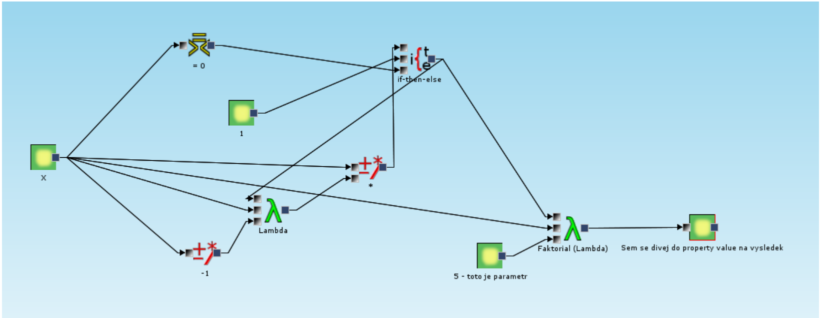
Python

```
fac = lambda x: x == 0 and 1 or x * fac(x - 1)
```

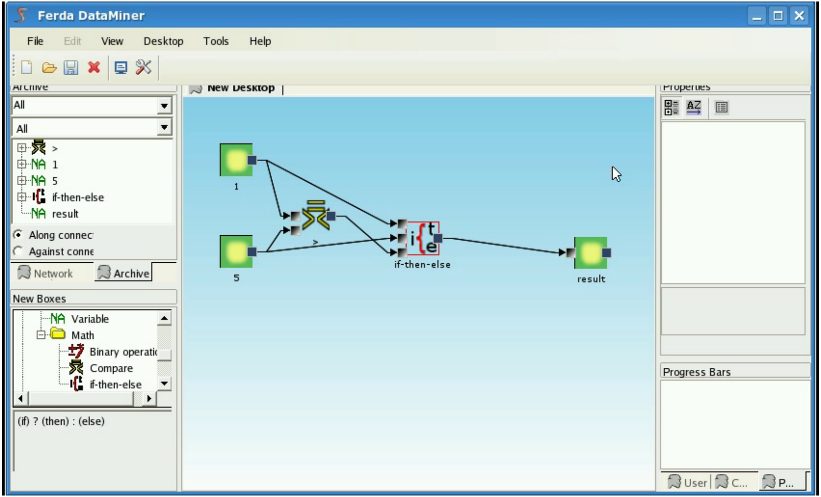
F#

```
let rec factorial n =
    if n=0 then 1 else n * factorial(n - 1)
```

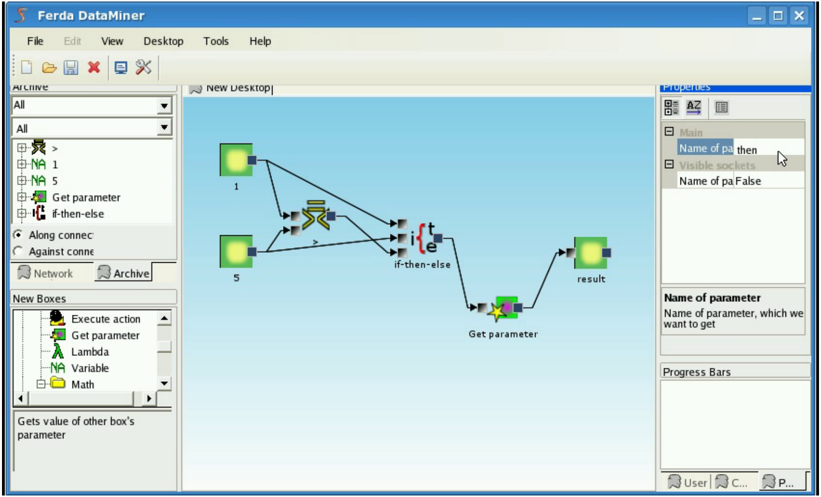
Factorial in Ferda



Movie – Get parameter



Get parameter



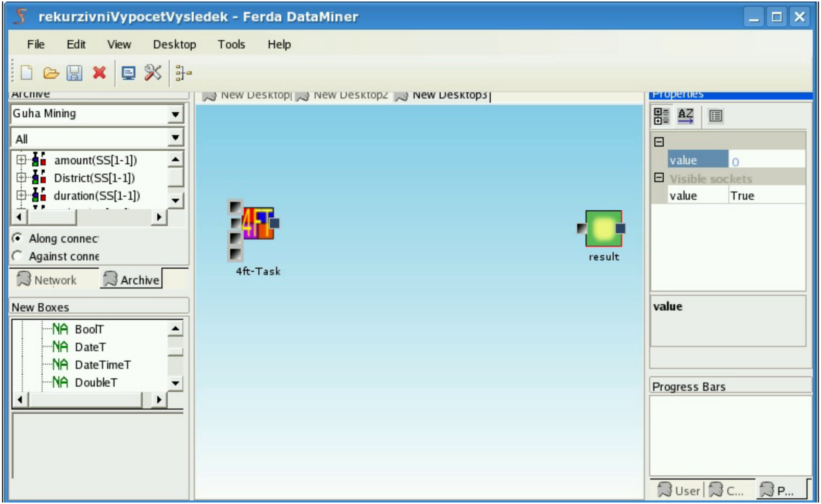
Name of pa: then

Name of pa: False

Name of parameter
Name of parameter, which we want to get

Gets value of other box's parameter

Movie – Execute action

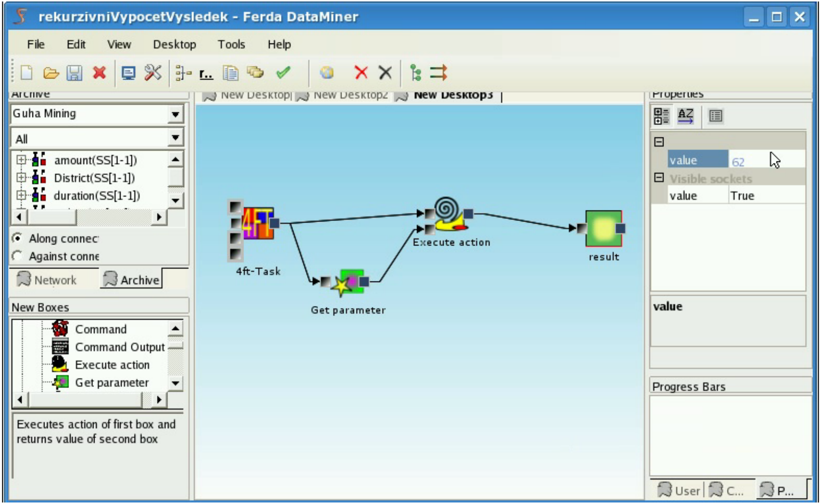


value 0

Visible sockets

value True

Execute action



rekurzivniVypocetVysledek - Ferda DataMiner

value	62

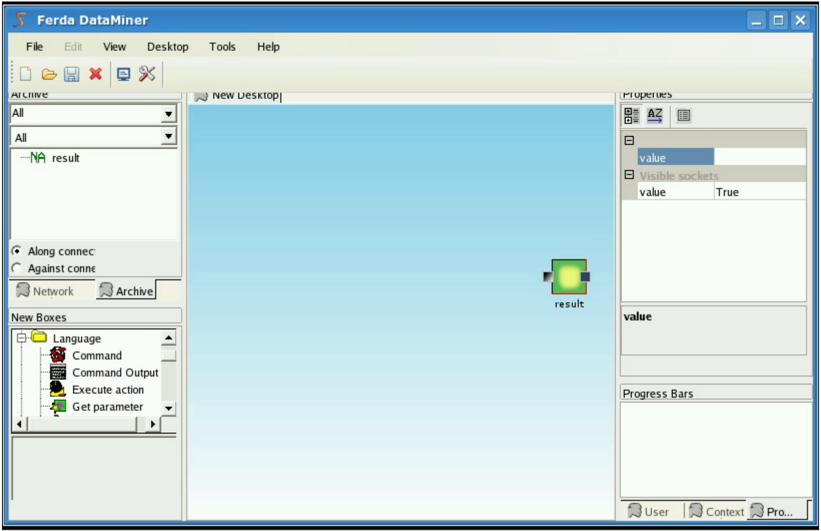
Visible sockets

value	True

value

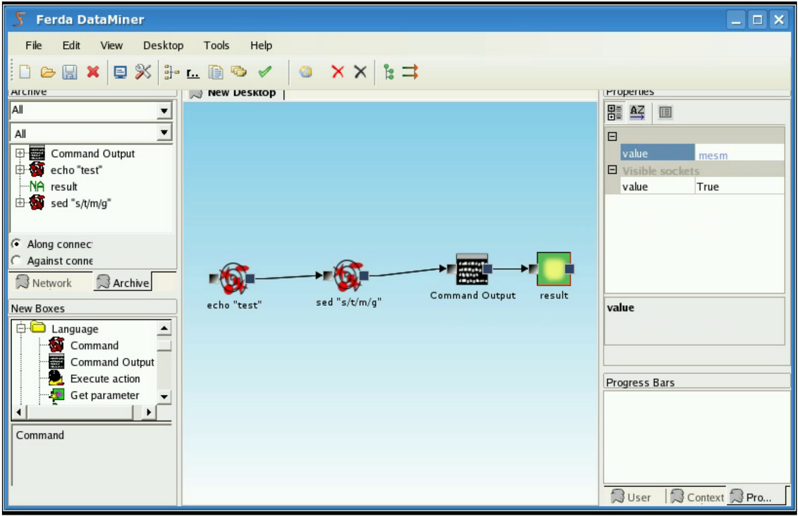
Executes action of first box and returns value of second box

Movie – Command and command output



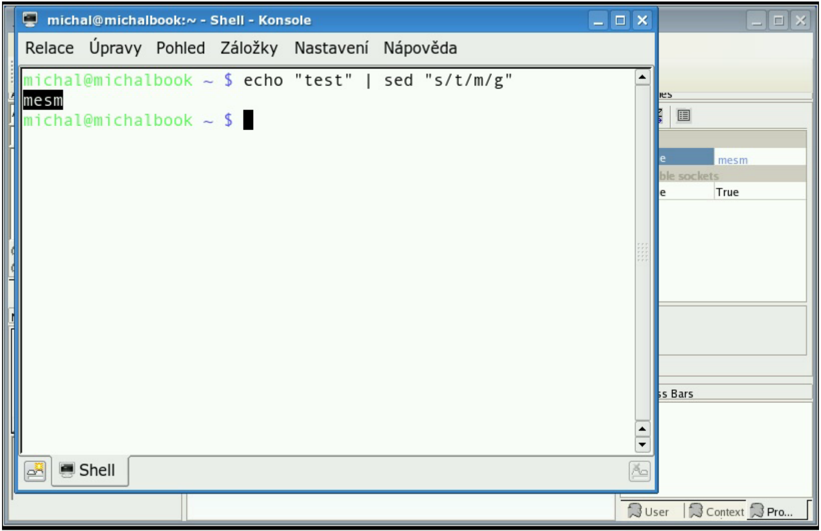
Command and command output

Example in ferda



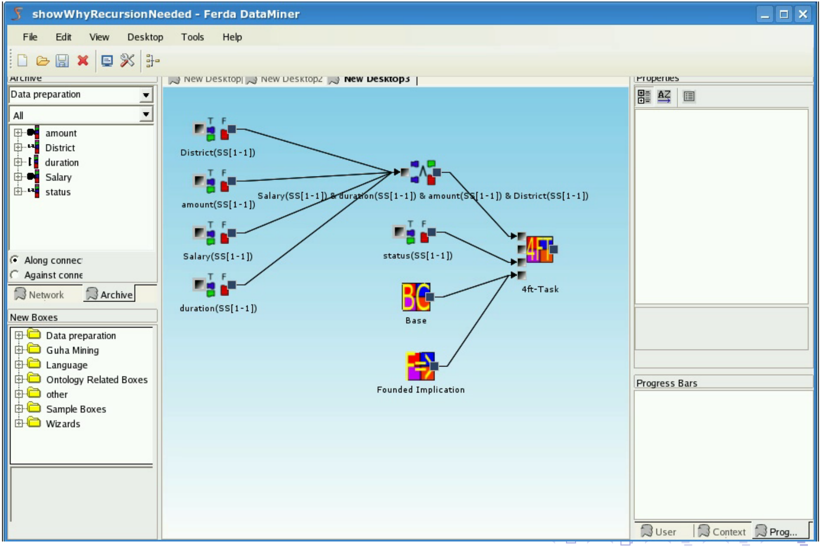
value	mesm
value	True

Command and command output



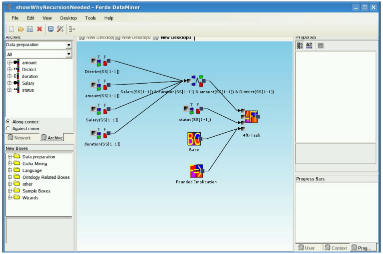
```
michal@michalbook ~ $ echo "test" | sed "s/t/m/g"
mesm
michal@michalbook ~ $ 
```

Movie – When lambda can be useful – 4FT task



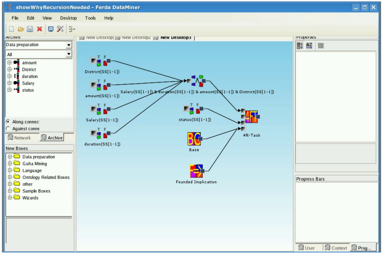
showWhyRecursionNeeded - Ferda DataMiner

When lambda can be useful – 4FT task



Problem

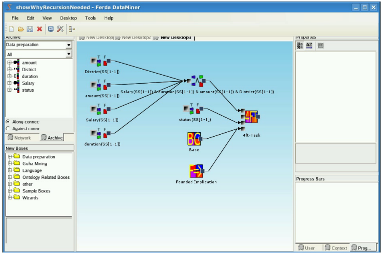
When lambda can be useful – 4FT task



Problem

- User tries some setting of quantifiers

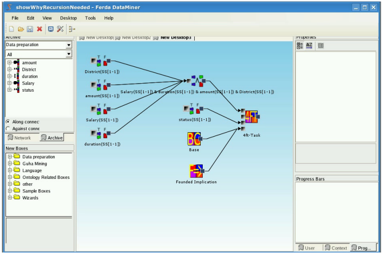
When lambda can be useful – 4FT task



Problem

- User tries some setting of quantifiers
- If he fails, he tries again with other settings?

When lambda can be useful – 4FT task



Problem

- User tries some setting of quantifiers
- If he fails, he tries again with other settings?
- It's manual and confusing

User specified automation of settings

User wants

- find the best settings for task

User specified automation of settings

User wants

- find the best settings for task
- suggest way how to find best settings

User specified automation of settings

User wants

- find the best settings for task
- suggest way how to find best settings
- have for different tasks different methods

User specified automation of settings

User wants

- find the best settings for task
- suggest way how to find best settings
- have for different tasks different methods

Programming finding best settings by user

User specified automation of settings

User wants

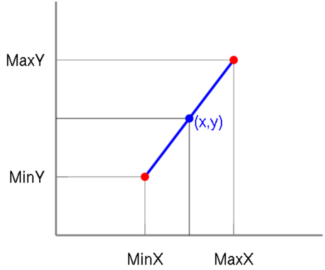
- find the best settings for task
- suggest way how to find best settings
- have for different tasks different methods

Programming finding best settings by user

- biggest variability

Linear interpolation

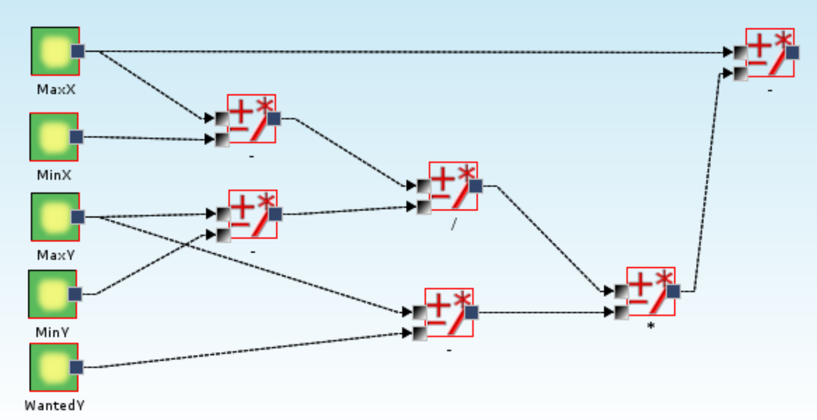
Basics



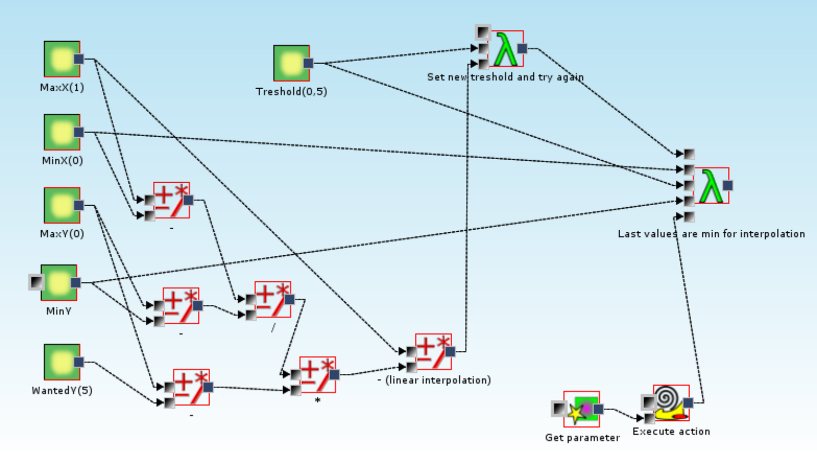
Formula

$$x = MaxX - (MaxY - WantedY)\frac{MaxX - MinX}{MaxY - MinY}$$

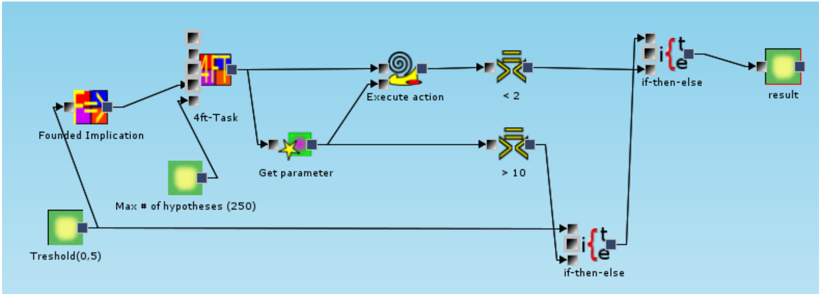
Interpolation as connection of boxes

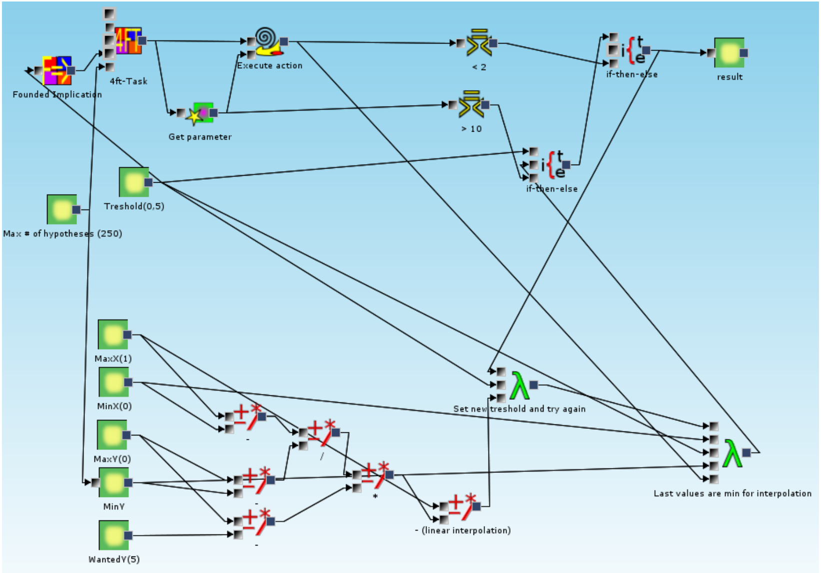


Result should be between

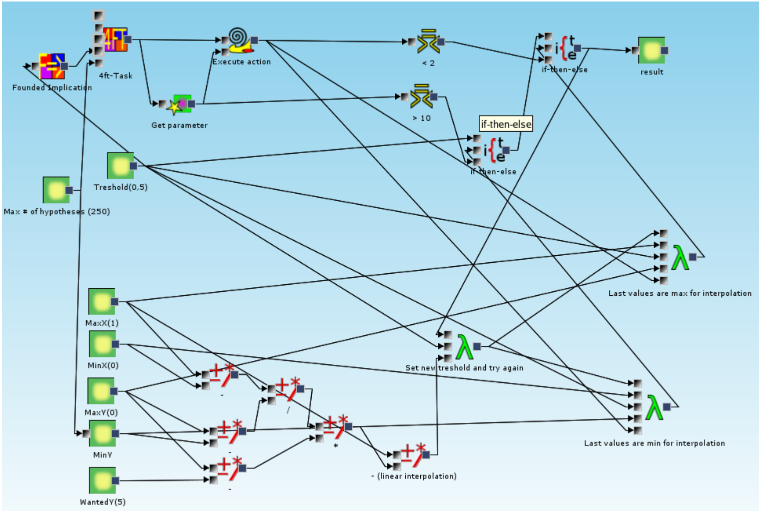


If not interpolate and set new max/min

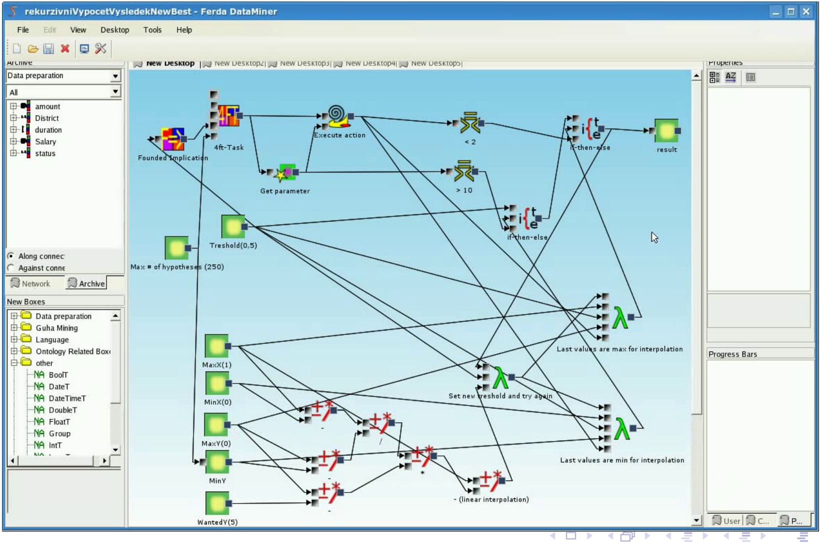



Founded Implication
4ft-Task
Execute action
< 2
if-then-else
result
Get parameter
> 10
if-then-else
Treshold(0,5)
Max # of hypotheses (250)
MaxX(1)
MinX(0)
MaxY(0)
MinY
WantedY(5)
-
/
-
*
-
- (linear interpolation)
Set new treshold and try again
Last values are min for interpolation

Connection


Founded Implication
4ft-Task
Execute action
Get parameter
< 2
> 10
if-then-else
result
if-then-else
if-then-else
Treshold(0,5)
Max # of hypotheses (250)
Last values are max for interpolation
MaxX(1)
MinX(0)
Set new treshold and try again
MaxY(0)
-
/
MinY
*
-
Last values are min for interpolation
- (linear interpolation)
-
WantedY(5)

Movie – Result



How to do it better

Is the recursive computing of 4FT needed

How to do it better

Is the recursive computing of 4FT needed

No – BestN algorithm

Is user programming needed

How to do it better

Is the recursive computing of 4FT needed

No – BestN algorithm

Is user programming needed

How to do it better

Is the recursive computing of 4FT needed

No – BestN algorithm

Is user programming needed

Yes – even with BestN for different projects different value function

Sequences and sets

Do we need sequences?

- Every mature programming language has something like sequences

Sequences and sets

Do we need sequences?

- Every mature programming language has something like sequences
- Sequences can be simulated, but the price

Sequences and sets

Do we need sequences?

- Every mature programming language has something like sequences
- Sequences can be simulated, but the price

What we would like to do with sequences

Sequences and sets

Do we need sequences?

- Every mature programming language has something like sequences
- Sequences can be simulated, but the price

What we would like to do with sequences

- Make a sequence of functions

Sequences and sets

Do we need sequences?

- Every mature programming language has something like sequences
- Sequences can be simulated, but the price

What we would like to do with sequences

- Make a sequence of functions
- Do something for each item in the sequence

Sequences and sets

Do we need sequences?

- Every mature programming language has something like sequences
- Sequences can be simulated, but the price

What we would like to do with sequences

- Make a sequence of functions
- Do something for each item in the sequence
- Add an item

Sequences and sets

Do we need sequences?

- Every mature programming language has something like sequences
- Sequences can be simulated, but the price

What we would like to do with sequences

- Make a sequence of functions
- Do something for each item in the sequence
- Add an item
- Subsequence

Sequences and sets

Do we need sequences?

- Every mature programming language has something like sequences
- Sequences can be simulated, but the price

What we would like to do with sequences

- Make a sequence of functions
- Do something for each item in the sequence
- Add an item
- Subsequence
- Concatenation

What should be done for support of sequences?

New boxes

- Sequence as head/tail

What should be done for support of sequences?

New boxes

- Sequence as head/tail
- Sequence as array of items

What should be done for support of sequences?

New boxes

- Sequence as head/tail
- Sequence as array of items
- ForEach

What should be done for support of sequences?

New boxes

- Sequence as head/tail
- Sequence as array of items
- ForEach
- AddItem, Subsequence, Concatenate

What should be done for support of sequences?

New boxes

- Sequence as head/tail
- Sequence as array of items
- ForEach
- AddItem, Subsequence, Concatenate
- New group box

What should be done for support of sequences?

New boxes

- Sequence as head/tail
- Sequence as array of items
- ForEach
- AddItem, Subsequence, Concatenate
- New group box
- Box for conversion from sequence to group box

What should be done for support of sequences?

New boxes

- Sequence as head/tail
- Sequence as array of items
- ForEach
- AddItem, Subsequence, Concatenate
- New group box
- Box for conversion from sequence to group box

Changes to Ferda core

What should be done for support of sequences?

New boxes

- Sequence as head/tail
- Sequence as array of items
- ForEach
- AddItem, Subsequence, Concatenate
- New group box
- Box for conversion from sequence to group box

Changes to Ferda core

What should be done for support of sequences?

New boxes

- Sequence as head/tail
- Sequence as array of items
- ForEach
- AddItem, Subsequence, Concatenate
- New group box
- Box for conversion from sequence to group box

Changes to Ferda core

Socket should accept new group box the same way it accepts the old one

Sequence example on DM

Problem

You have a table and don't know anything about. You would like to know something about it.

Resolution

Sequence example on DM

Problem

You have a table and don't know anything about. You would like to know something about it.

Resolution

- Connect all columns you want to analyze to the sequence

Sequence example on DM

Problem

You have a table and don't know anything about. You would like to know something about it.

Resolution

- Connect all columns you want to analyze to the sequence
- Regarding types of colums and their data, create attributes for each column in the sequence

Sequence example on DM

Problem

You have a table and don't know anything about. You would like to know something about it.

Resolution

- Connect all columns you want to analyze to the sequence
- Regarding types of colums and their data, create attributes for each column in the sequence
- Create basic tasks

Sequence example on DM

Problem

You have a table and don't know anything about. You would like to know something about it.

Resolution

- Connect all columns you want to analyze to the sequence
- Regarding types of colums and their data, create attributes for each column in the sequence
- Create basic tasks
- See result of these tasks

Reuse of code

Problem description

Problem

The user wants to reuse boxes which he created.

We have

Reuse of code

Problem description

Problem

The user wants to reuse boxes which he created.

We have

- Project files

Reuse of code

Problem description

Problem

The user wants to reuse boxes which he created.

We have

- Project files
- Network archive

Reuse of code

Better network archive

What can be done

- Labels

Reuse of code

Better network archive

What can be done

- Labels
- More network archives

Reuse of code

Better network archive

What can be done

- Labels
- More network archives
- User access rights

Reuse of code

Project files

User could load boxes from other project

- Easy to implement

Reuse of code

Project files

User could load boxes from other project

- Easy to implement
- There is workaround with network archive

Reuse of code

Project files

User could load boxes from other project

- Easy to implement
- There is workaround with network archive

Including other project

Reuse of code

Project files

User could load boxes from other project

- Easy to implement
- There is workaround with network archive

Including other project

- Like other programming languages

Reuse of code

Project files

User could load boxes from other project

- Easy to implement
- There is workaround with network archive

Including other project

- Like other programming languages
- Strong

Reuse of code

Project files

User could load boxes from other project

- Easy to implement
- There is workaround with network archive

Including other project

- Like other programming languages
- Strong
- Harder to implement

Better lambda

- Parameter could be function with parameter, not only constant

Better lambda

- Parameter could be function with parameter, not only constant
- Lambda is slow

Better lambda

- Parameter could be function with parameter, not only constant
- Lambda is slow

How to make lambda quicker

Better lambda

- Parameter could be function with parameter, not only constant
- Lambda is slow

How to make lambda quicker

- Quicker creating of boxes

Better lambda

- Parameter could be function with parameter, not only constant
- Lambda is slow

How to make lambda quicker

- Quicker creating of boxes
- Functions could be called only once

Better lambda

- Parameter could be function with parameter, not only constant
- Lambda is slow

How to make lambda quicker

- Quicker creating of boxes
- Functions could be called only once
- Boxes could be clonned only in time it is really needed

Other things to do

- Most of modules for interacionts should work on top of functions not boxes

Other things to do

- Most of modules for interacionts should work on top of functions not boxes
- Boxes for GUHA – parts

Other things to do

- Most of modules for interacionts should work on top of functions not boxes
- Boxes for GUHA – parts
- User could choose on which computer a box is running

Other things to do

- Most of modules for interacionts should work on top of functions not boxes
- Boxes for GUHA – parts
- User could choose on which computer a box is running
- Box which can be programmed by user at runtime in some scripting language

Summary

Programming language in Ferda should

- offer more power for data mining

Summary

Programming language in Ferda should

- offer more power for data mining
- allow to do things which were not designed before

Summary

Programming language in Ferda should

- offer more power for data mining
- allow to do things which were not designed before

Next steps

Summary

Programming language in Ferda should

- offer more power for data mining
- allow to do things which were not designed before

Next steps

- there are many things to do