A query formulation tool for semantic networks

Asaridis Christos

Advisors:

Researcher **Martin Doerr**

Assistant Prof. **Yannis Tzitzikas**

Institute of Computer Science
Foundation for Research and Technology - Hellas

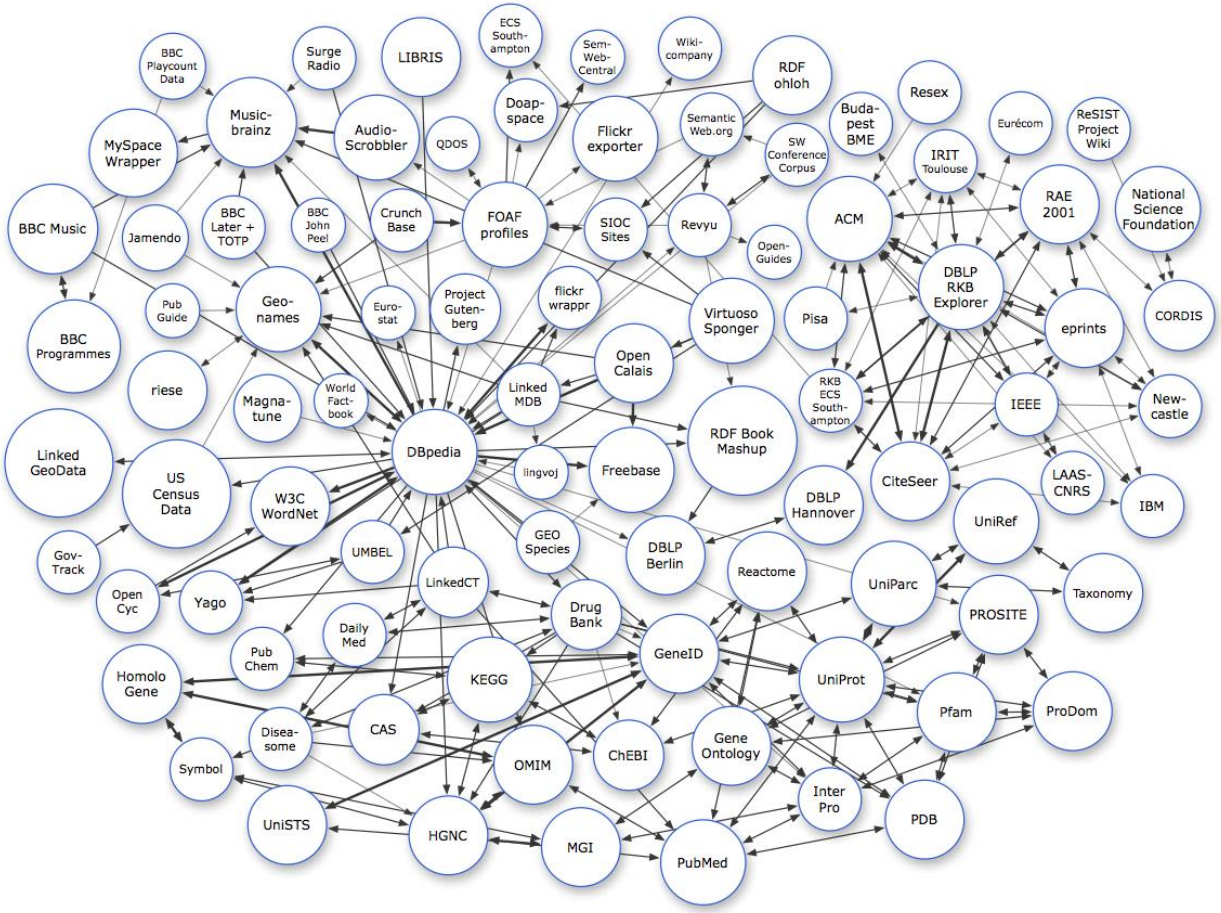
Computer Science Department
University of Crete

Outline

- ***Introduction***
- ***Problem***
- ***Implementation***
- ***Evaluation***
- ***Future Work***

INTRODUCTION

Semantic Web



- ***Web of data***
- ***RDF/S***
- ***RDF Triple Stores***

Cultural Heritage Semantic Networks

Triple Stores of:

- Heterogeneous data
- social and historical events

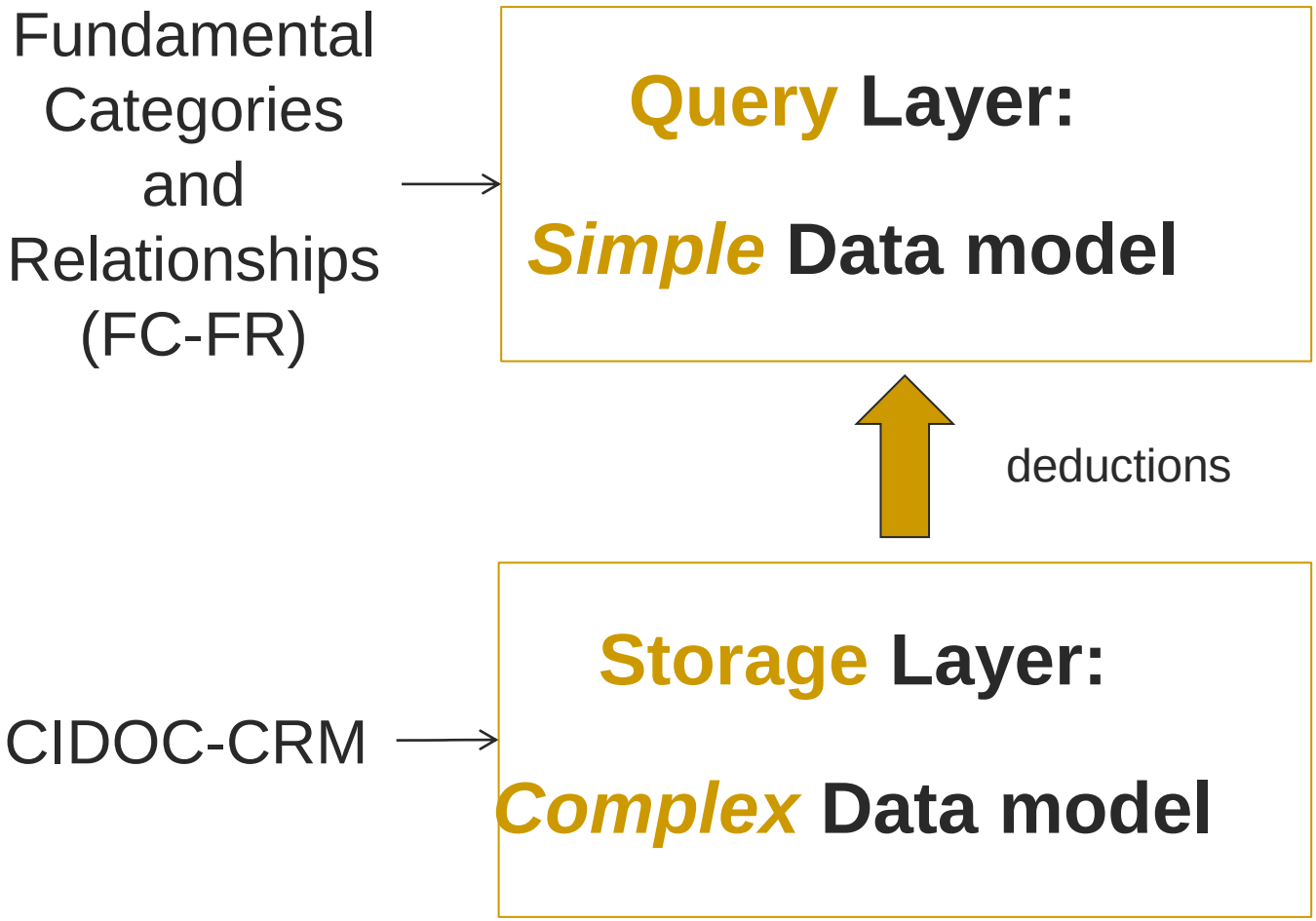
- composite structure
- diverse semantics
- multiple kinds of relationships

A Global Schema: The CIDOC CRM

Extensible **core ontology**

- More than 150 classes
- More than 250 properties
- Museum disciplines, archives and libraries

*Fundamental Categories and Fundamental Relationships**



*Katerina Tzomanaki, Martin Doerr, "Fundamental categories and relationships for intuitive querying CIDOC-CRM based repositories," FORTH technical report TR419, Tech. Rep., 2012.

Traditional way

```
select distinct  $Thing1  $Label
{
$Thing1 rdf:type crm:E70.Thing.
        optional{$Thing1 crmdig:L4F.has_preferred_label
        $Label.
        }{optional{
        $Thing1 crm:P130F.shows_features_of
        }}UNION
        {optional{
         $Thing1 rdf:type c
        $Thing1 crm:
        UNION
                                            $Information_Object.
                              rm:P67F.refers_to
             hing1 crm:P67F.refers_to     $Thing2.
        }
....
        UNION{
        $Thing1 crm:F1F.is_derivative_of  $tmpThing2.
    $tmpThing2 crmdig:L11B.was_output_of
        $DigMachEventX2.
         $DigMachEventX2 crm:P9B.forms_part_of      $Z1.
```

Very long SPARQL queries
Very difficult to write and comprehend

FC-FR Framework

E70.Thing -- (F3.is_same_as)[0,n] ->E70
E70.Thing--(F4F.is_comp > E70.Thing
E24.Physi --
_Thing --
->
rmation_Obje
F.refers_to ->
D1.Digital
_of)[0,n] ->
bject --
as_output_of ->
7.Digital_Machine_Event--
(P9B.forms_part_of)[0,n] ->
D2.Digitization_Process --
L1F.digitized -> E70.Thing --
(F4F.is_composed_of) [0,n] -> E70.Thing

Smaller paths - More readable and writable

Schema knowledge

Thing "is about or refers to"

What the end user sees

SPARQL vs FC-FR

```
select distinct $StartVar $Label {
$StartVar  rdf:type crm:E39.Actor.
optional { $StartVar crmdig:L4F.has_preferred_label $Label. }{
$StartVar  crm:P107B.is_current_or_former_member_of $var22.
{ $var22 crm:P62B.is_depicted_by $var23.}
UNION{ $var22 crm:P62B.is_depicted_by $tempvar24.$tempvar24 crm:P46B.forms_part_of $var23.}
{ $var23 crm:P108B.was_produced_by $var24.}
UNION{ $var23 crm:P108B.was_produced_by $tempvar25.$tempvar25 crm:P9B.forms_part_of $var24.}
{ $var24 crm:P7F.took_place_at $Endvar.}
UNION{ $var24 crm:P7F.took_place_at $var25 . $var25 crm:P89F.falls_within $Endvar.}
 }
 UNION {
{$StartVar  crm:P62B.is_depicted_by $var19.}
UNION{$StartVar  crm:P62B.is_depicted_by $tempvar20.$tempvar20 crm:P46B.forms_part_of $var19.}
```

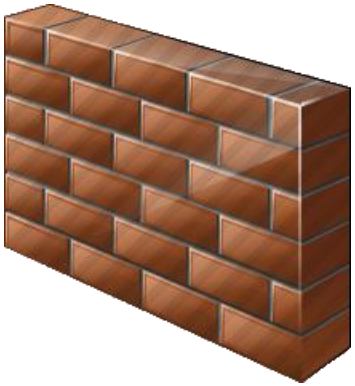
```
E39.Actor--(P107B.is_current_or_former_member_of)[0,n]-> E39.Actor:
{ E39.Actor -- P62B.is_depicted_by -> E24.Physical_Man-Made_Thing:
        { E24.Physical_Man-Made_Thing -- (P46B.forms_part_of) [0,n] -> E24.Physical_Man-Made_Thing:
            { E70.Thing-- {P53F.has_former_or_current_location  OR P54F.has_current_permanent_location} -> E53.Place:
                { E53.Place  --( P89F.falls_within)[0,n]-> E53.Place }
           OR E24.Physical_Man-Made_Thing -- P108B.was_produced_by -> E12.Production:
                { E12.Production --(P9B.forms_part_of)[0,n]-> E5.Event:
                    { E5.Event -- P7F.took_place_at ->E53.Place:
                        { E53.Place  --( P89F.falls_within)[0,n]-> E53.Place }}}}}
OR E39.Actor -- P67B.is_referred_to_by  -> E89.Propositional_Object:
        { E73.Information_Object -- P94B.was_created_by -> E65.Creation:
            { E65.Creation --( P9B.forms_part_of)[0,n] ->E5.Event:
                { E5.Event -- P7F.took_place_at ->E53.Place:
                    { E53.Place  --( P89F.falls_within)[0,n]-> E53.Place }}}
       OR E73.Information_Object -- P128B.is_carried_by -> E24.Physical_Man-Made_Thing:
            { E24.Physical_Man-Made_Thing -- (P46B.forms_part_of) [0,n] ->   E18.Physical_Thing:
                {E70.Thing -- {P53F.has_former_or_current_location  OR P54F.has_current_permanent_location}-> E53.Place :
                    {E53.Place  --( P89F.falls_within)[0,n]-> E53.Place }}}}
}
```

```
 UNION {
$StartVar  crm:P67B.is_referred_to_by $var5.
{ $var5 crm:P94B.was_created_by $var6.}
UNION{ $var5 crm:P94B.was_created_by $tempvar7.$tempvar7 crm:P9B.forms_part_of $var6.}
{ $var6 crm:P7F.took_place_at $Endvar.}
UNION{ $var6 crm:P7F.took_place_at $var7 . $var7 crm:P89F.falls_within $Endvar.}
 }
 UNION {
$StartVar  crm:P107B.is_current_or_former_member_of $var2.
{ $var2 crm:P62B.is_depicted_by $var3.}
UNION{ $var2 crm:P62B.is_depicted_by $tempvar4.$tempvar4 crm:P46B.forms_part_of $var3.}
{ $var3 crm:P53F.has_former_or_current_location $Endvar.}
UNION{ $var3 crm:P53F.has_former_or_current_location $var4 . $var4 crm:P89F.falls_within $Endvar.}
 }
 UNION {
{$StartVar  crm:P62B.is_depicted_by $var0.}
UNION{$StartVar  crm:P62B.is_depicted_by $tempvar1.$tempvar1 crm:P46B.forms_part_of $var0.}
{ $var0 crm:P53F.has_former_or_current_location $Endvar.}
UNION{ $var0 crm:P53F.has_former_or_current_location $var1 . $var1 crm:P89F.falls_within $Endvar.}
 }
}
```

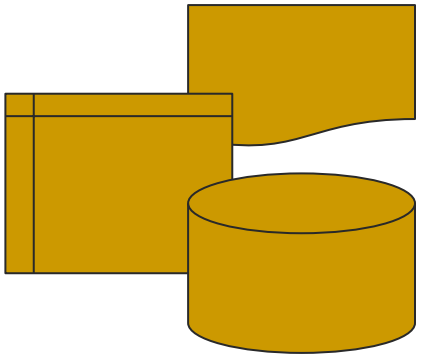
PROBLEM

Data Access Challenges

Information need



Semantic Web data



- Accessing data is a challenging task
  - Difficult to train users to write complex or comprehensive queries
  - Simple predefined sets of queries need specializations
- IT experts have to be involved

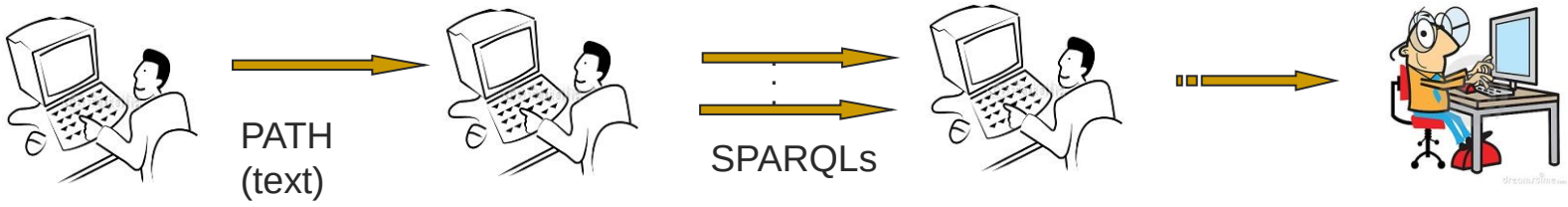
Existing Approach

Write Predefined Relationship

Validate and create SPARQL

Incorporate SPARQLs in the querying tool

Query



Paths' language

FR configuration tool

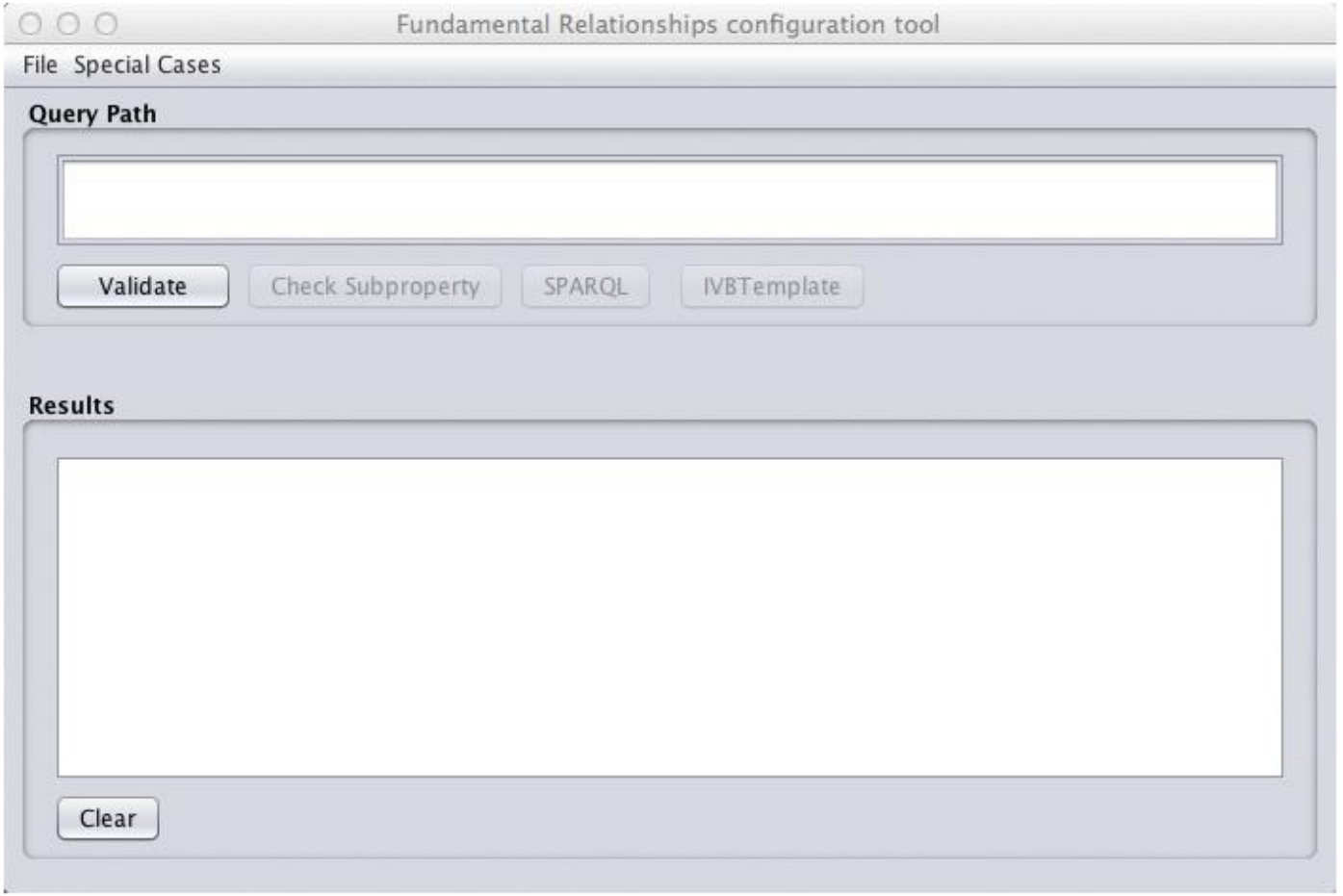
Querying component

Querying component

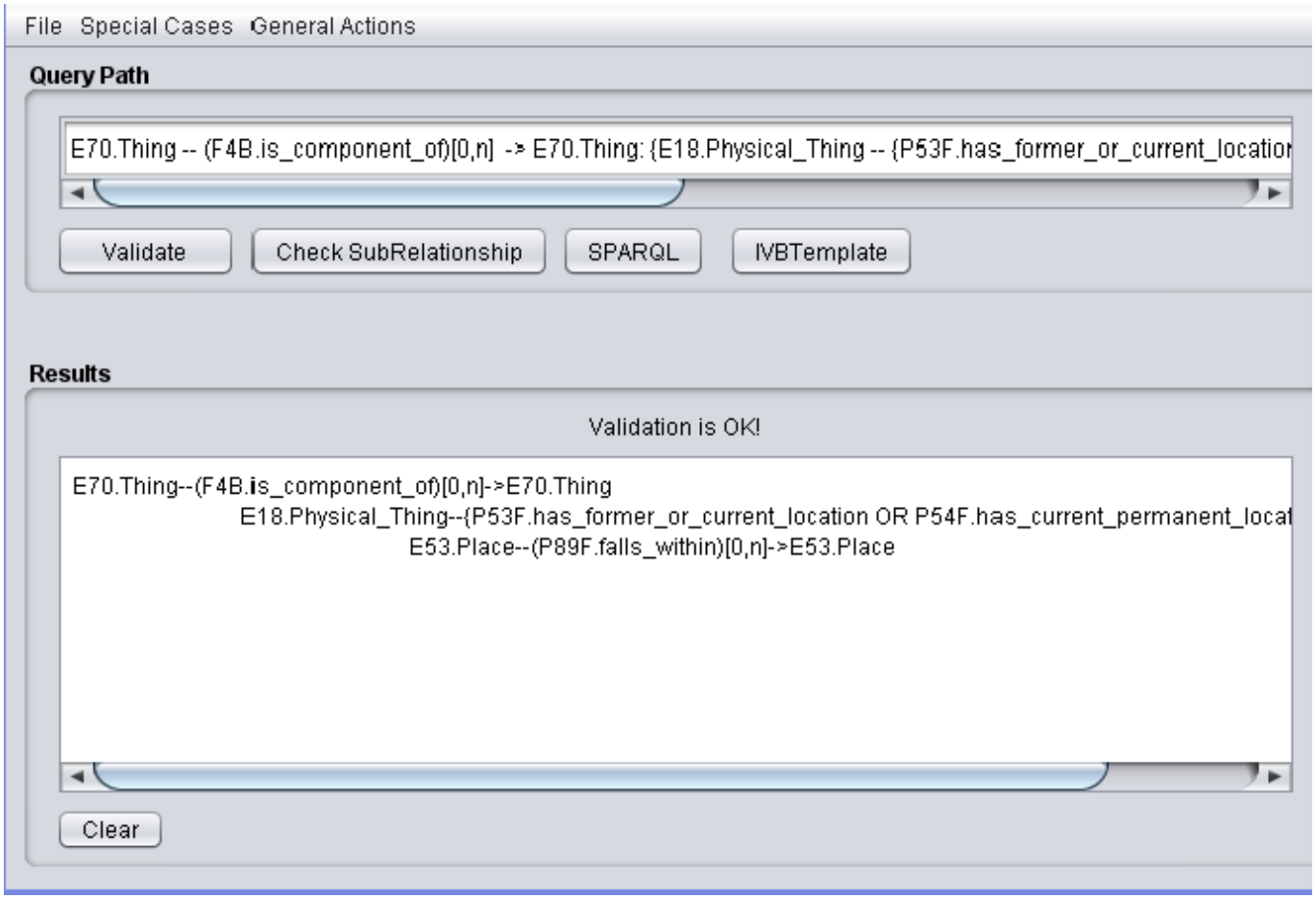
administrative user

end user

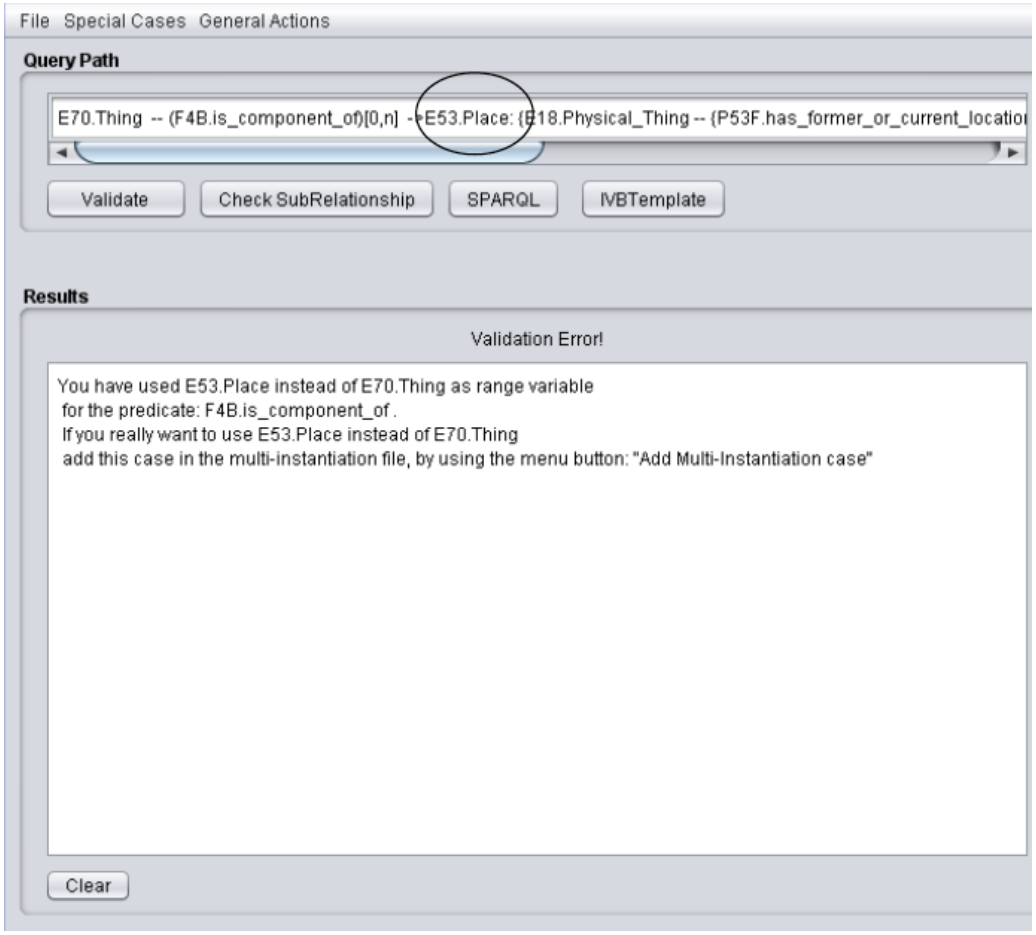
Text-based Editor



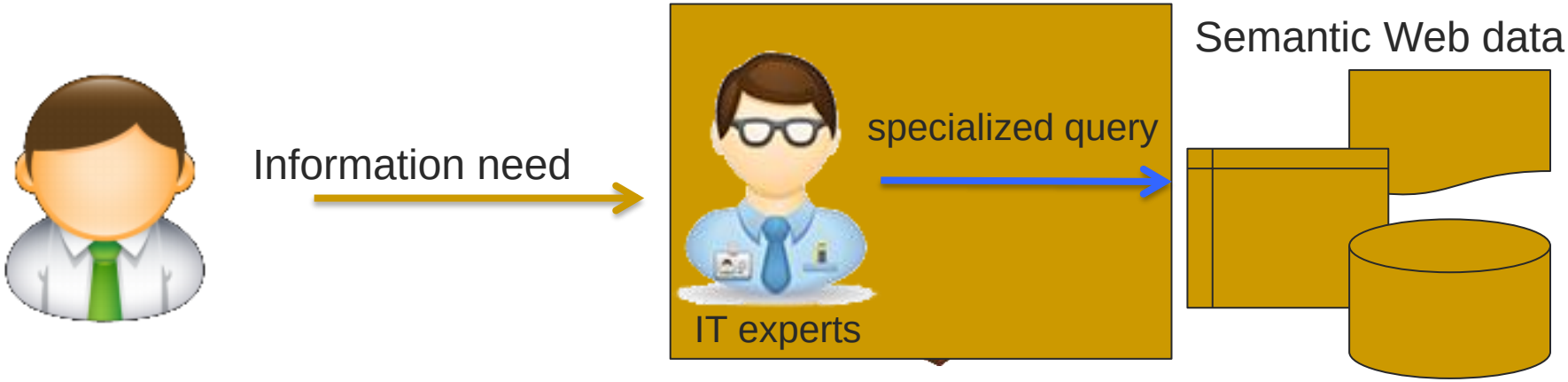
Check Validation



Report An Error



Need for New Querying Techniques



- Access to IT experts is a bottleneck
  - Understanding both domain and data takes years of training
  - Difficult even for IT experts to write complex queries
  - Wait days for response from overloaded IT department
  - Misunderstandings can lead to several iterations

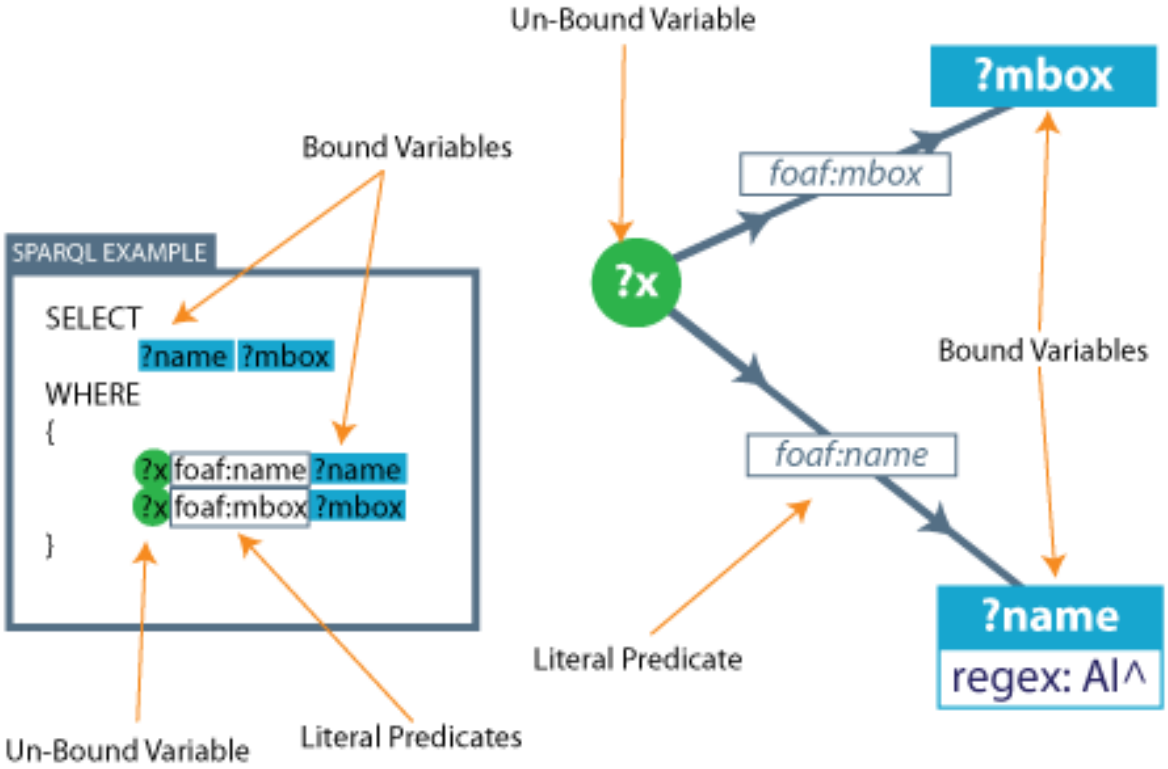
Visual Representation of Queries

- *User Interface query formulation techniques:*
  - Graphical representations of query
  - User friendly representation of language
  - Menu-guided interfaces with look ahead mechanisms

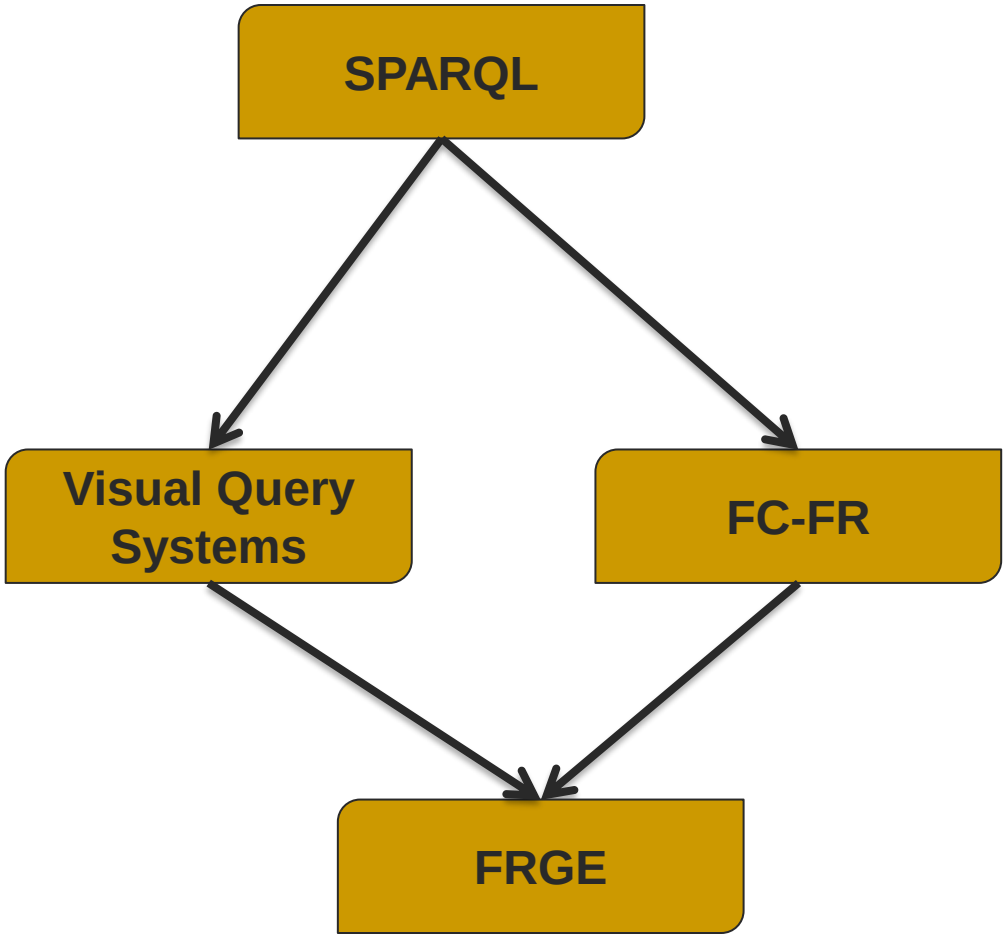
- *Existing Tools:*
  - Nitelight
  - VIQUEN
  - visKWQL

- ***Drawbacks:***
  - Dependency on SPARQL knowledge
  - Dependency on knowledge of schema structure

Nitelight


Un-Bound Variable
?mbox
Bound Variables
foaf:mbox
SPARQL EXAMPLE
?x
SELECT
?name ?mbox
WHERE
{
?x foaf:name ?name
?x foaf:mbox ?mbox
}
Bound Variables
foaf:name
?name
regex: Al^
Literal Predicate
Un-Bound Variable
Literal Predicates

Querying Semantic Networks



Visualizing Querying with FRGE

E70.Thing -- (F3.is_same_as)[0,n]->E70.Thing:

E70.Thing--(F4F.is_composed_of) [0,n] ->

E24.Physical_Man-Made

depicts->

E70.Thing

OR

E24.Ph

ies ->

refers_to ->

(F1F.is_derivative_of)[0,n] ->

al_Object -- L11B.was_output_of ->

.Digital_Machine_Event--

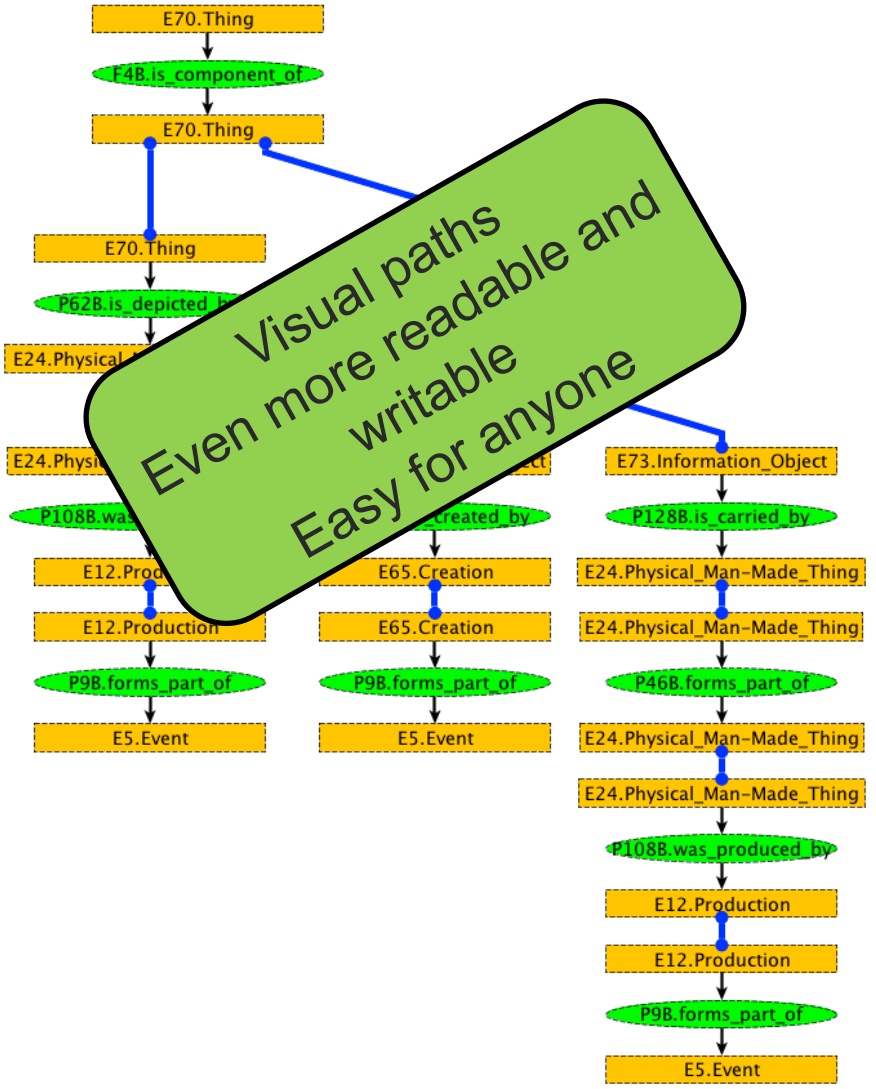
(P9B.forms_part_of)[0,n] ->

D2.Digitization_Process -- L1F.digitized ->

E70.Thing -- (F4F.is_composed_of) [0,n] ->

E70.Thing

IT expert is needed because of the requirement of scripting skills



Visual paths
Even more readable and writable
Easy for anyone

Contributions

Approach:

- *Simple syntax to write path expressions - machine and human readable*
- *Graphical editing and online validation*
- *Graphical visualization for better comprehension*

IMPLEMENTATION

Implementation Challenges

Three major challenges of implementation

- *Import Process*
  - Modeling the new language (Paths' Language)
  - Develop a parser able to recognize the new language
  - Convert from path language to graph
- *Editing Process*
  - Presentation capabilities
  - Editing capabilities
  - Set the constraints for valid path composition
  - Base platform ready to accept extra plugins
- *Export Process*
  - Modeling the decompose process of the graph
  - Develop an export mechanism

Implementation

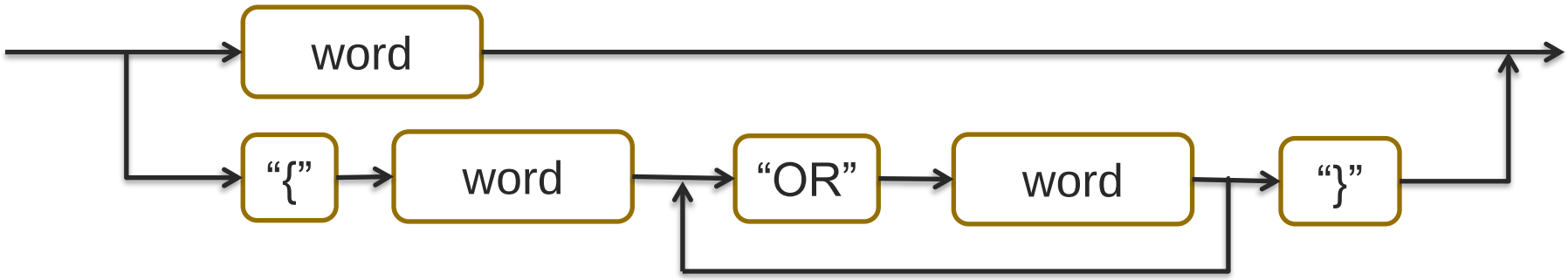
- ***Parser***
- *Fundamental Relationship Graphical Editor (FRGE)*
- *Export algorithm*

Paths' Language Grammar (1/2)

Triple

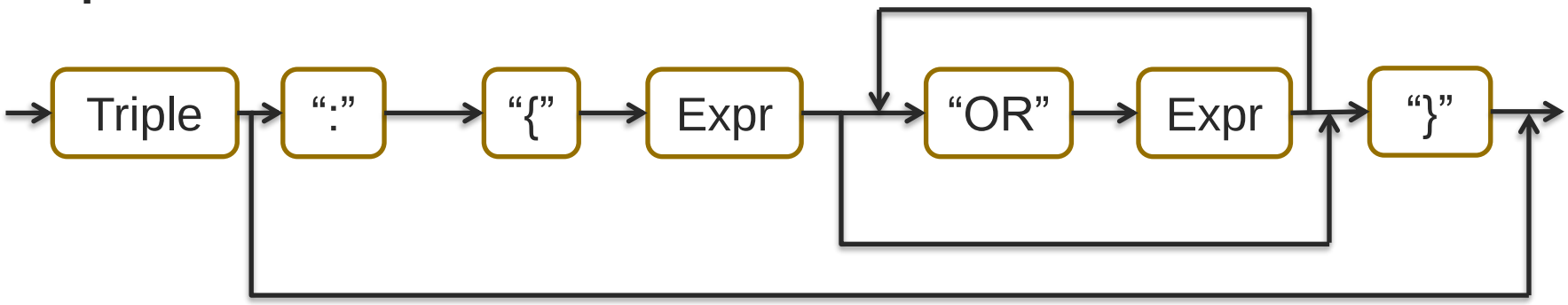
Subject / Object

Predicate



Paths' Language Grammar (2/2)

Expression



Implementation

- *Parser*
- ***Fundamental Relationship Graphical Editor (FRGE)***
- *Export algorithm*

Portability

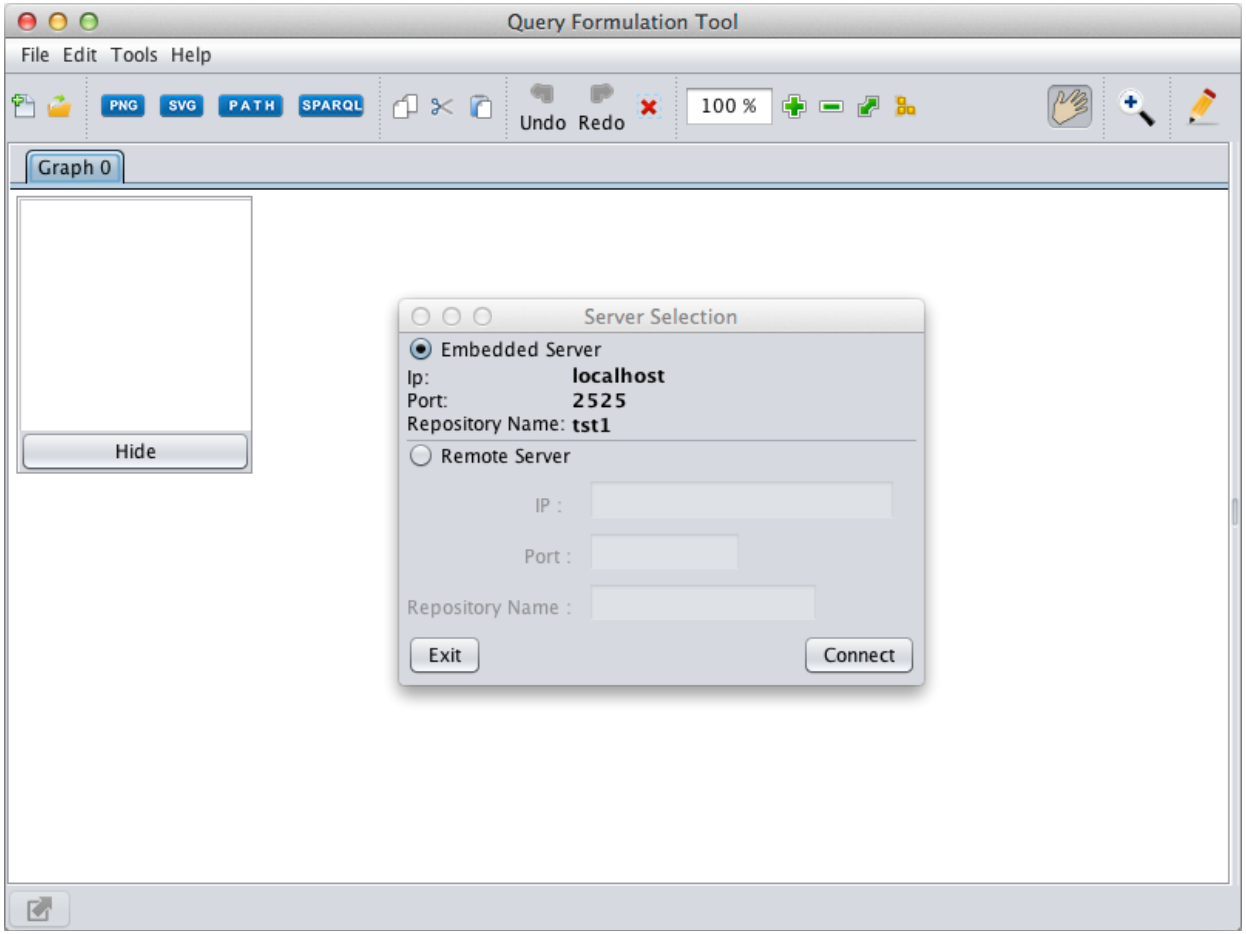
Easy to transfer

- *The FRGE is implemented in Java version 7*
  - Cross platform
  - Independent from hardware
- *Embedded server*
  - No need for internet connection
  - Apache Tomcat 7.0.39

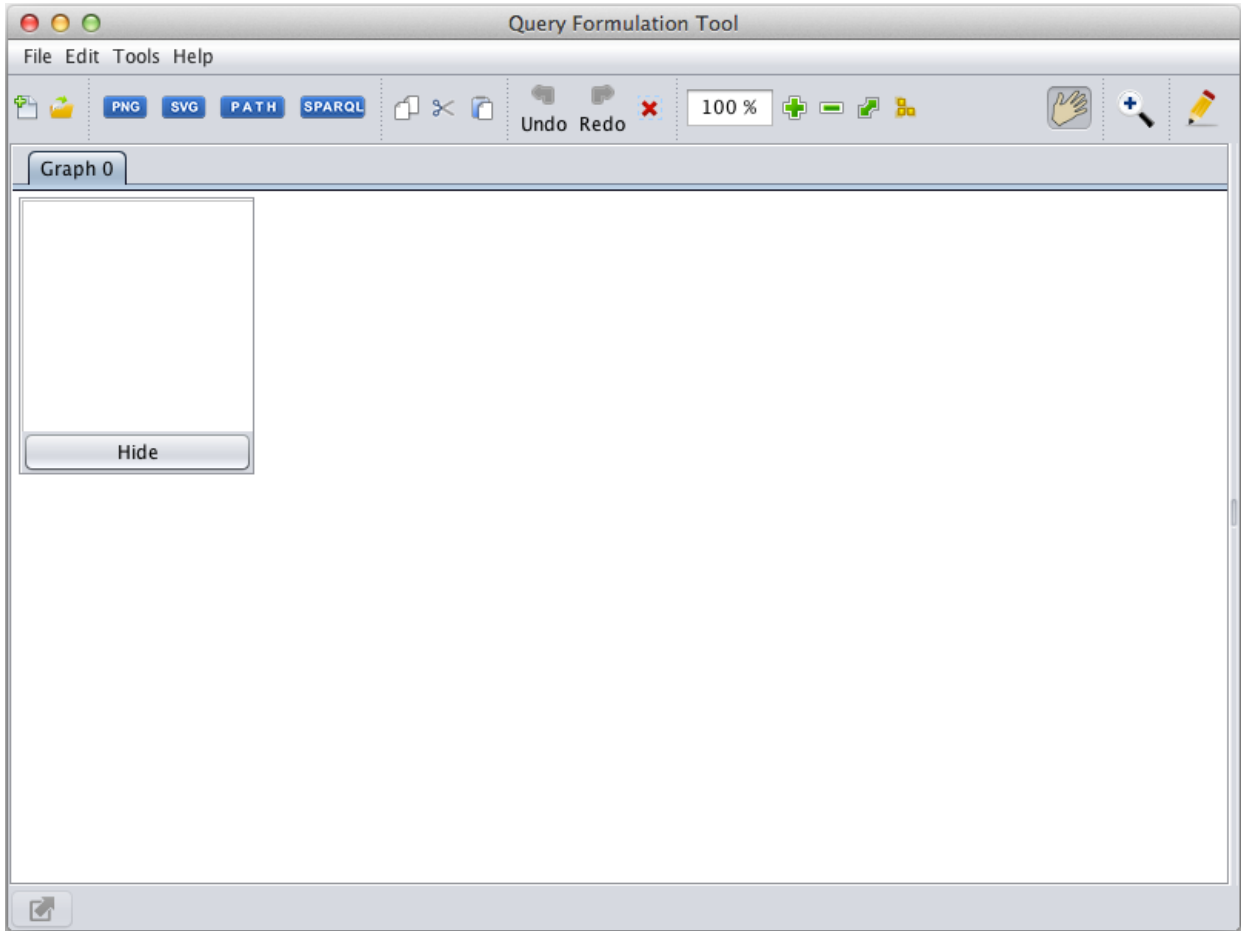
FRGE

FRGE
Fundamental Relationship Graphical Editor
Faculty
Data

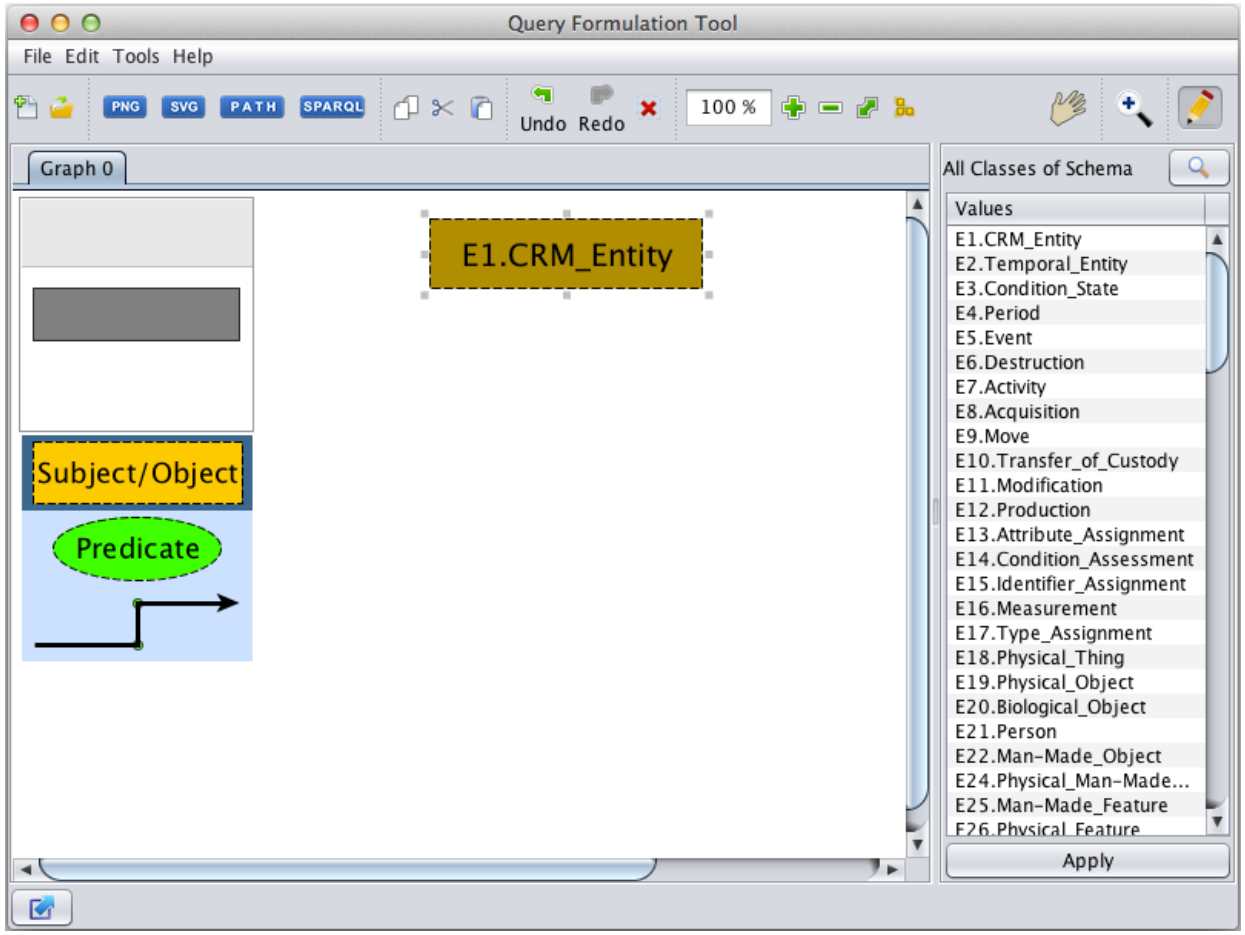
Select Server and Repository

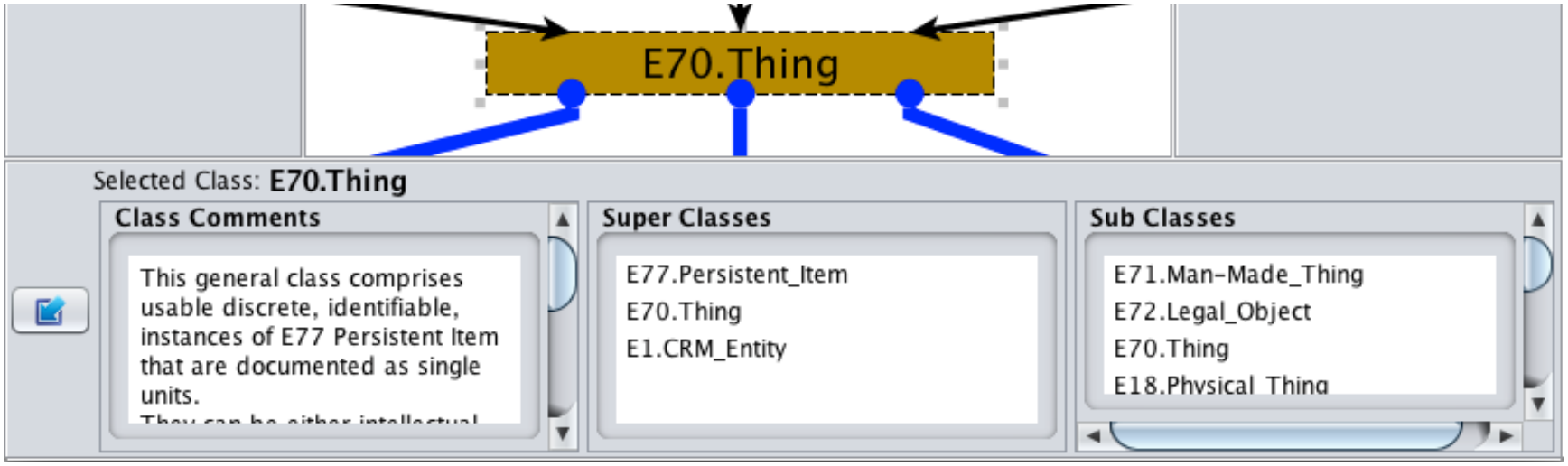


First View


Query Formulation Tool
File Edit Tools Help
PNG
SVG
PATH
SPARQL
Undo
Redo
100 %
Graph 0
Hide

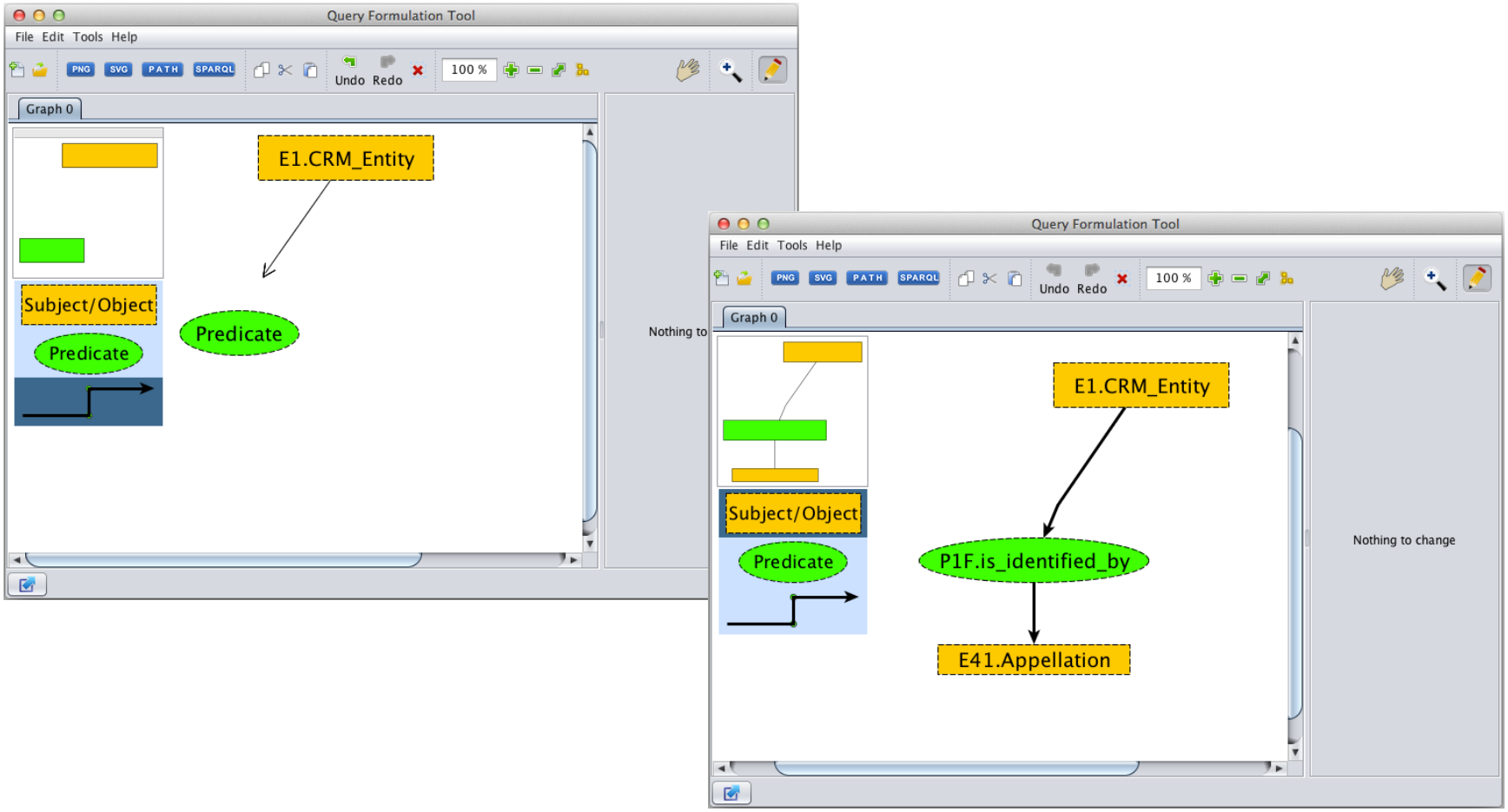
Nothing to Type



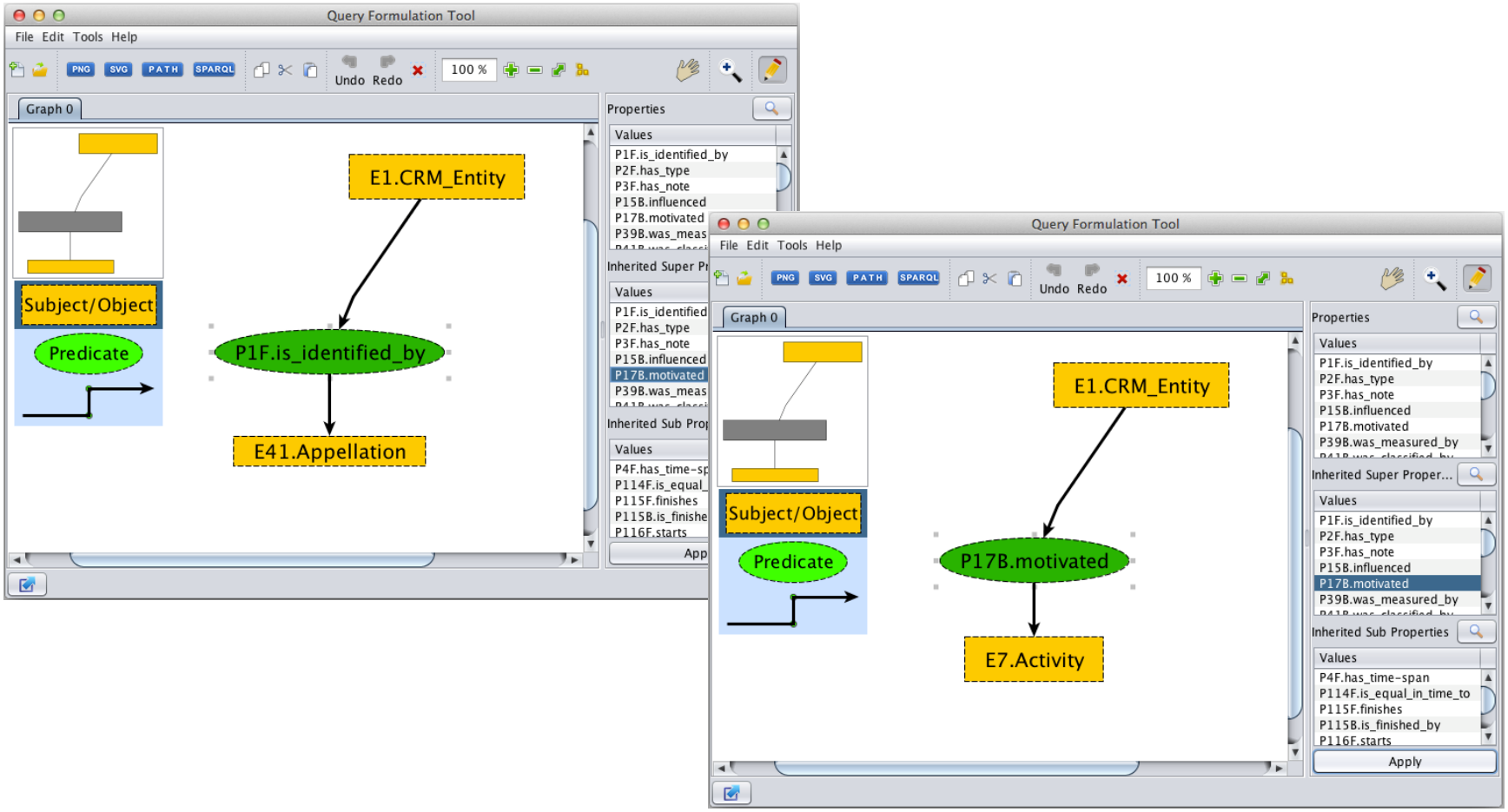


Information About Graph Elements

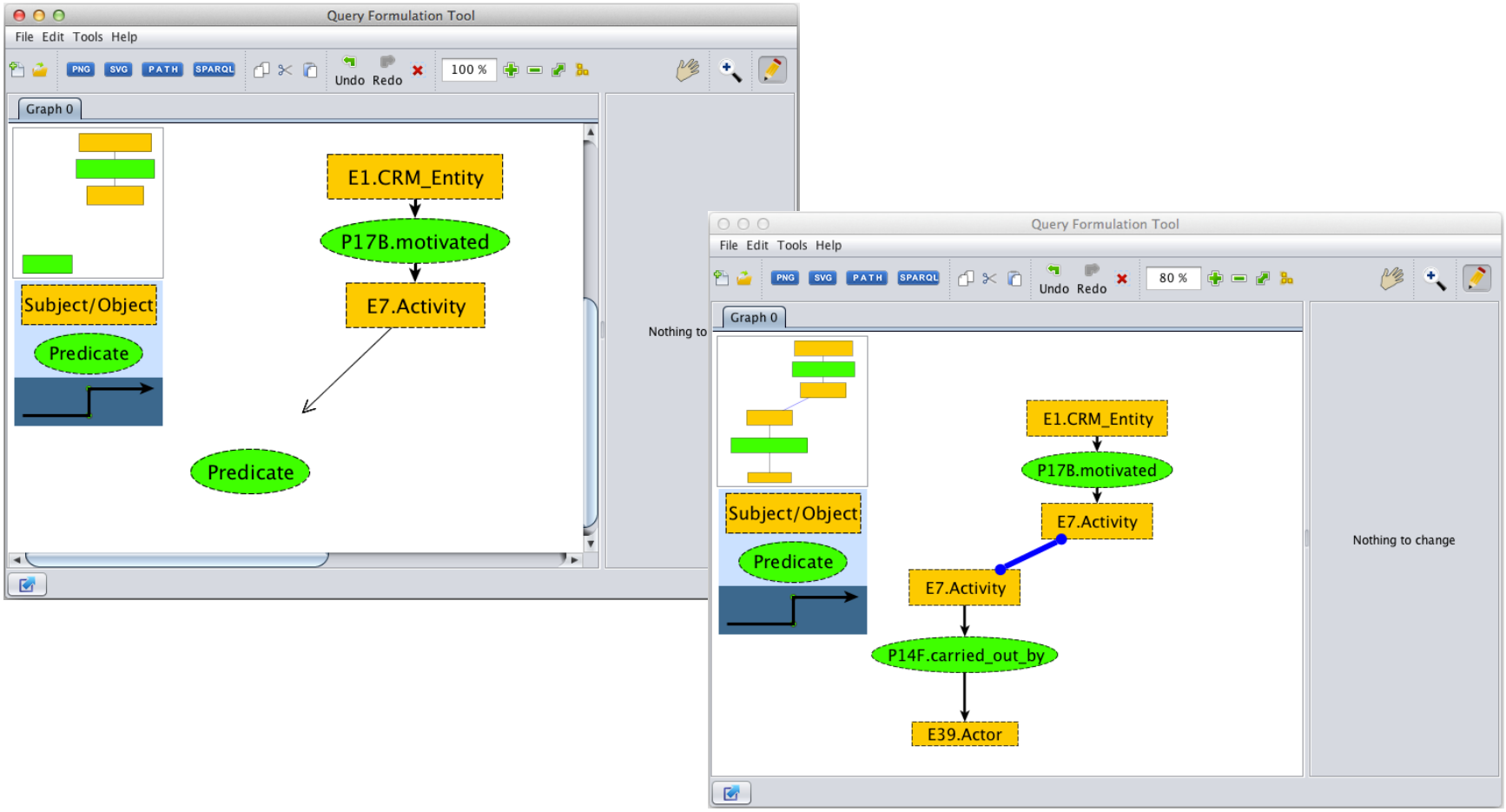
Auto Triple Complete



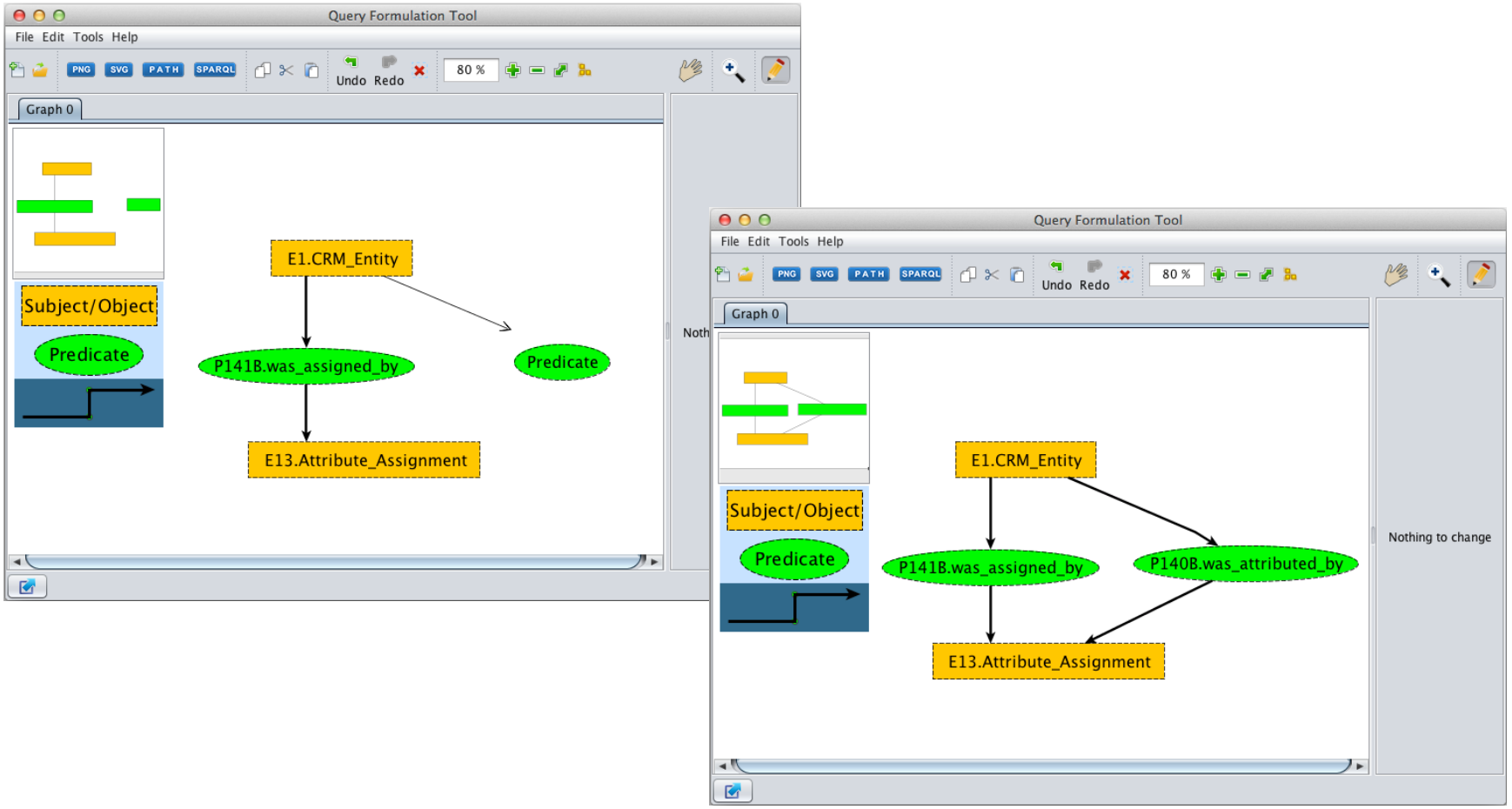
Auto Change Object



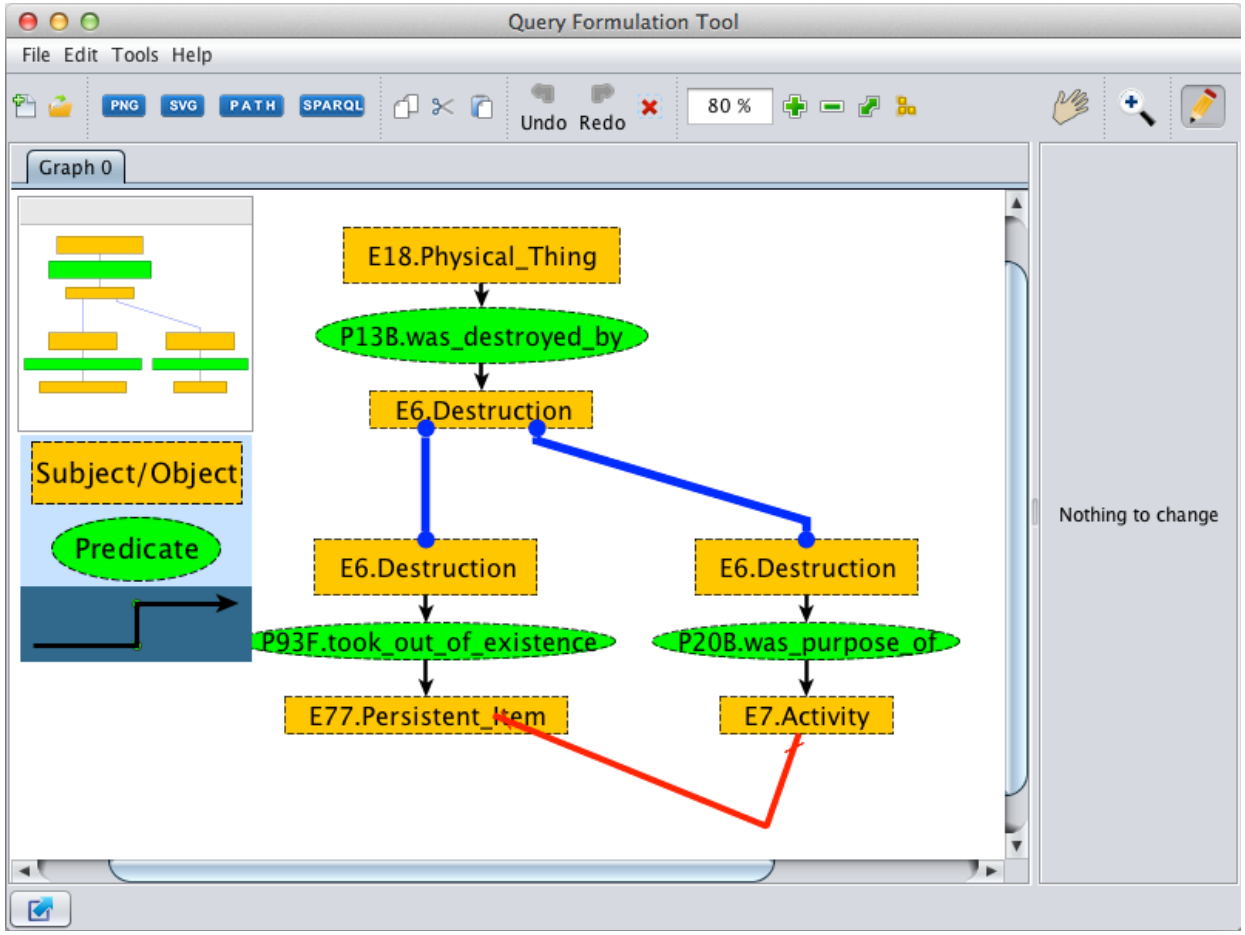
Auto Path Complete



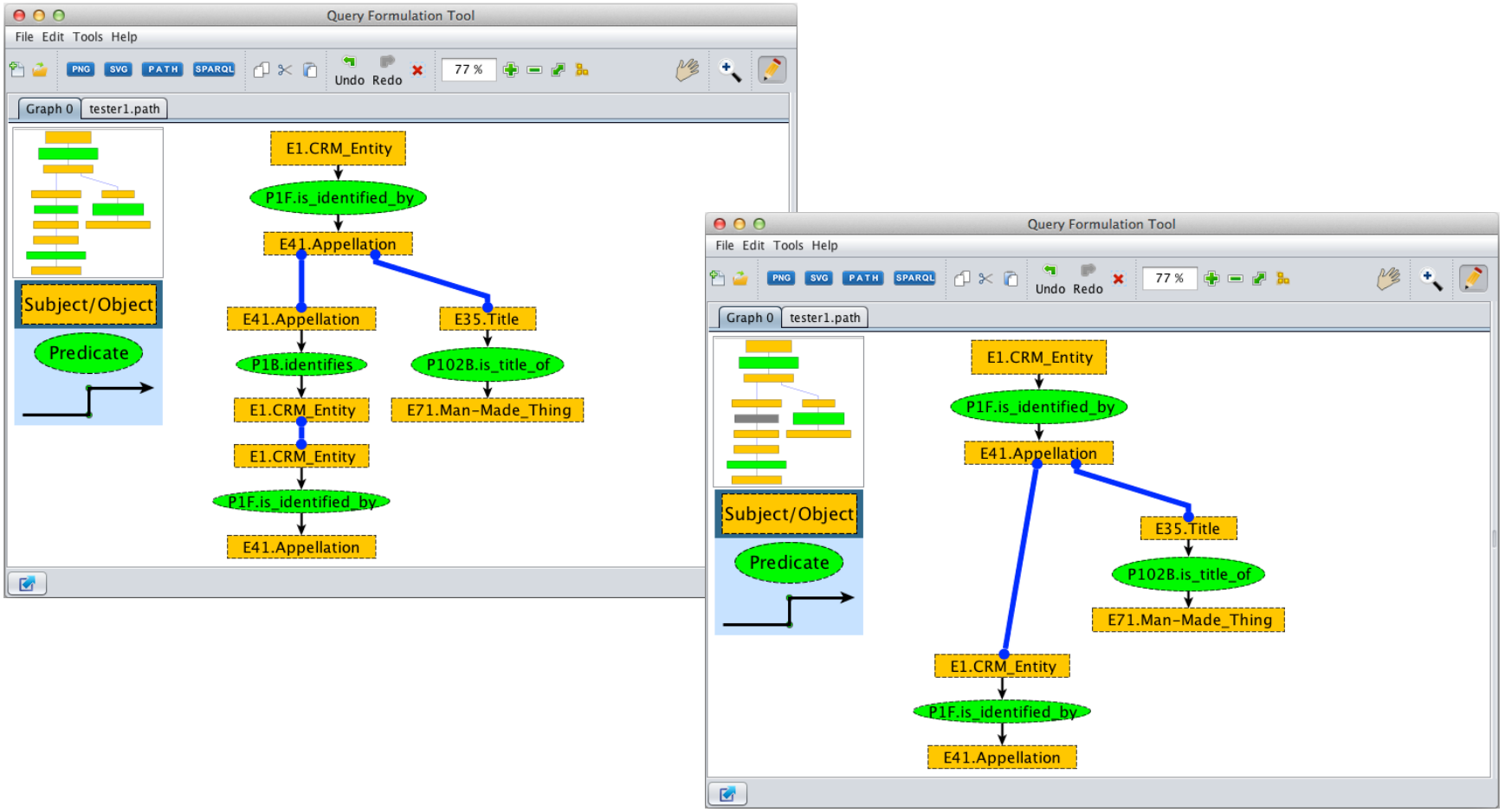
Logic in Auto Complete



Online Validation



Delete Action



User Individuals

File Edit Tools Help

Manage Prefixes

List of P

crm
crmdig
rdf
rdfs
xsd

Graph

URI of s

http://
RM_CI

Manage Disjoins

Search:

Disjoins

E2.Temp
E28.Conc
E41.App
E51.Cont
E73.Infor
E30.Righ

Manage Multi-Instantiations

Search:

Add

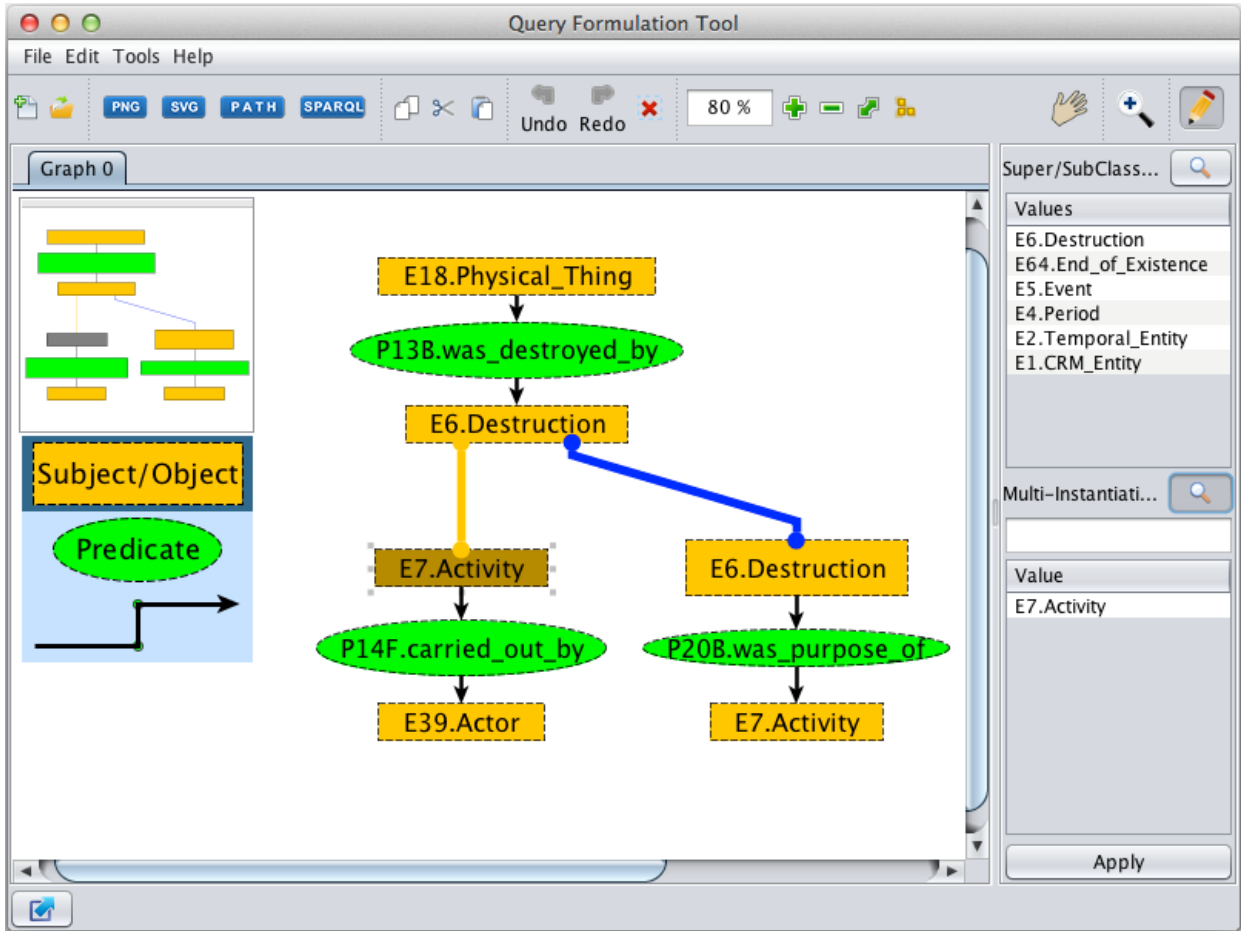
Remove

Close

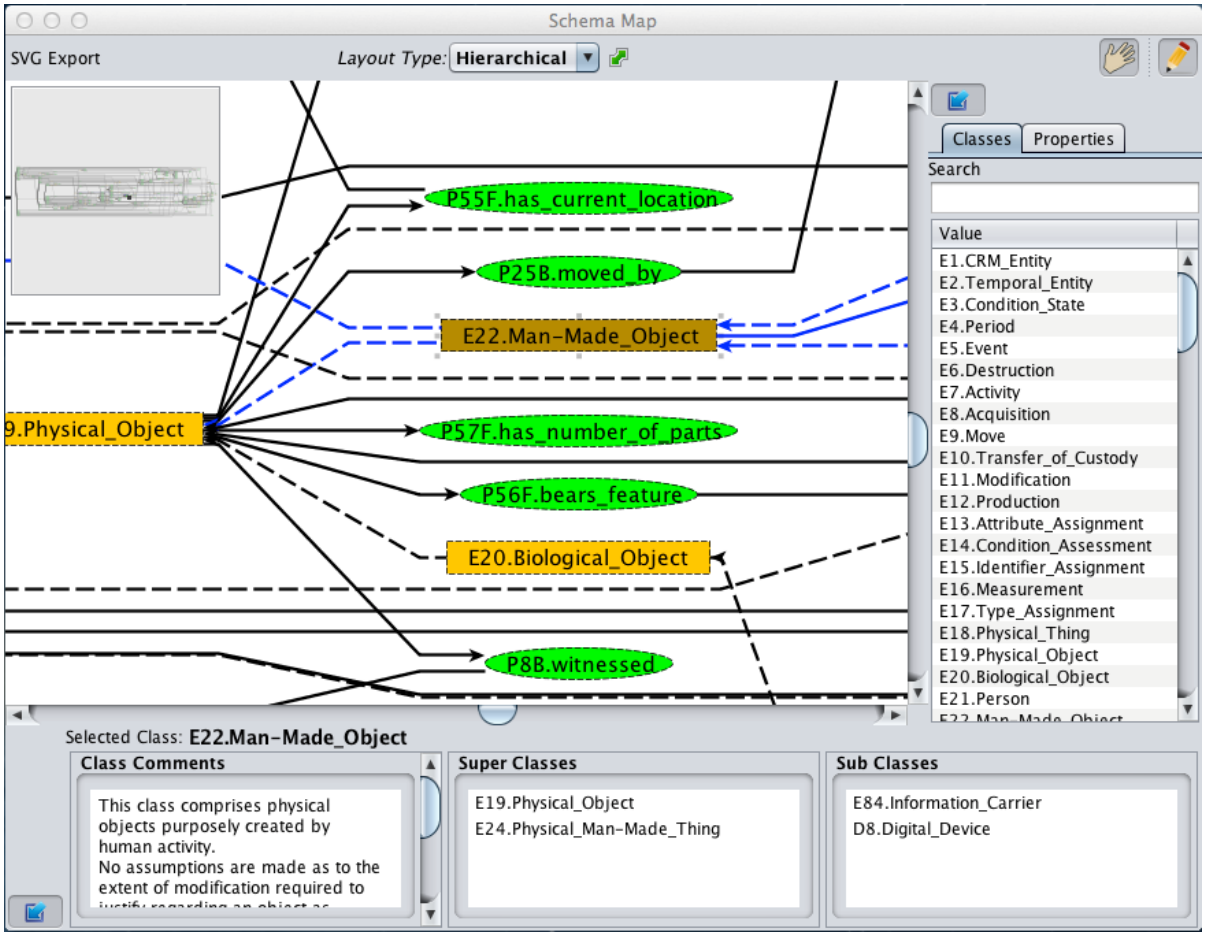
Multi-Instantiations

E26.Physical_Feature <-> E53.Place
E31.Document <-> E33.Linguistic_Object
E31.Document <-> E38.Image
E33.Linguistic_Object <-> E38.Image
E41.Appellation <-> E33.Linguistic_Object
E6.Destruction <-> E7.Activity

Multiple Instantiation



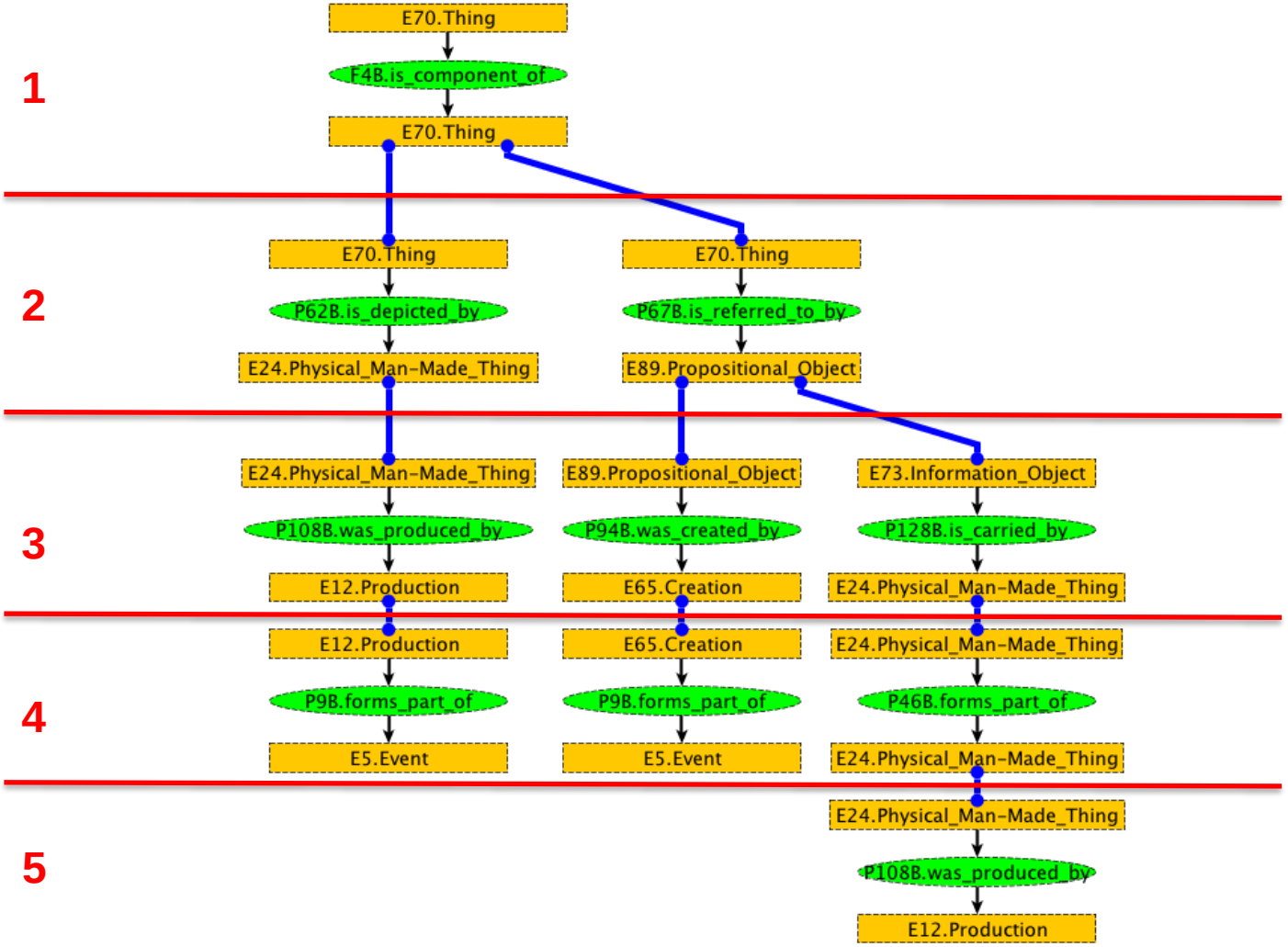
Schema Map


Schema Map
SVG Export
Layout Type: Hierarchical
P55F.has_current_location
P25B.moved_by
E22.Man-Made_Object
9.Physical_Object
P57F.has_number_of_parts
P56F.bears_feature
E20.Biological_Object
P8B.witnessed
Classes
Properties
Search
Value
E1.CRM_Entity
E2.Temporal_Entity
E3.Condition_State
E4.Period
E5.Event
E6.Destruction
E7.Activity
E8.Acquisition
E9.Move
E10.Transfer_of_Custody
E11.Modification
E12.Production
E13.Attribute_Assignment
E14.Condition_Assessment
E15.Identifier_Assignment
E16.Measurement
E17.Type_Assignment
E18.Physical_Thing
E19.Physical_Object
E20.Biological_Object
E21.Person
Selected Class: E22.Man-Made_Object
Class Comments
This class comprises physical objects purposely created by human activity.
No assumptions are made as to the extent of modification required to
Super Classes
E19.Physical_Object
E24.Physical_Man-Made_Thing
Sub Classes
E84.Information_Carrier
D8.Digital_Device

Implementation

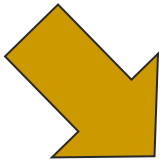
- *Parser*
- *Fundamental Relationship Graphical Editor (FRGE)*
- ***Export algorithm***

Path to Be Exported



Export Algorithm (1/2)

0	1	2	*NULL*	*NULL*
0	1	3	4	*NULL*
0	5	6	7	*NULL*
0	5	8	9	10



0	1	2	*NULL*	*NULL*
NULL	*NULL*	3	4	*NULL*
NULL	5	6	7	*NULL*
NULL	*NULL*	8	9	10

Export Algorithm (2/2)

- *One counter for opened brackets*
- *Two position pointers:*
  - First one points to last index of the current path (*n* row)
  - Second one points to first filled index of next row (*n+1 row*)

0 :{	1 :{	2		
	OR	3 :{	4 }}	
OR	5 :{	6 :{	7 }	
	OR	8 :{	9 :{	10 }}}

EVALUATION

Evaluation Categorization

Division of the evaluation process into the three main categories:

- Time to accomplish a task
- Correctness of queries
- Convenience of navigating and understanding the paths

Time to Accomplish a Task (1/2)

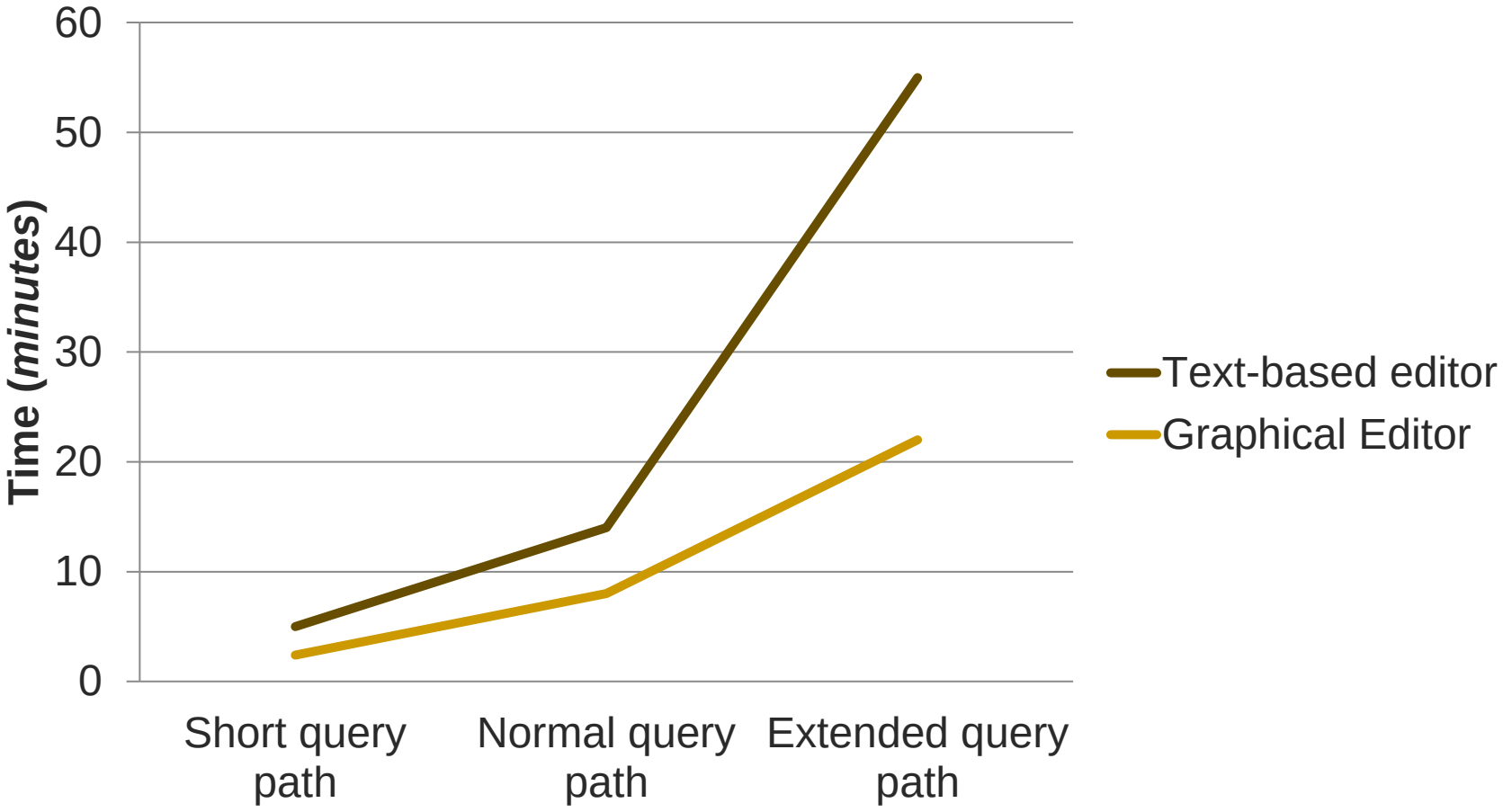
Test queries:

- *Short query path: 3 levels / 0 branches*
- *Normal query path: 5 levels / 2 branches*
- *Extended query path: 10 levels / 5 branches*

	Text-based Configuration Tool	Graphical Configuration Tool
Short query path	about 5 minutes	2 to 3 minutes
Normal query path	12 to 15 minutes	7 to 9 minutes
Extended query path	50 to 60 minutes	20 to 25 minutes

Time to construct the paths queries using different configuration tool

Time to Accomplish a Task (2/2)



Correctness of Queries

- *The criteria for this tests are:*
  - Accuracy and comprehension
  - Correct spelling and path structure
- *Text-based editor users ran into two kind of errors:*
  - Wrong spelling of classes and properties
  - Validity of the paths
- *Graphical editor users:*
  - No spelling errors
  - No validity errors

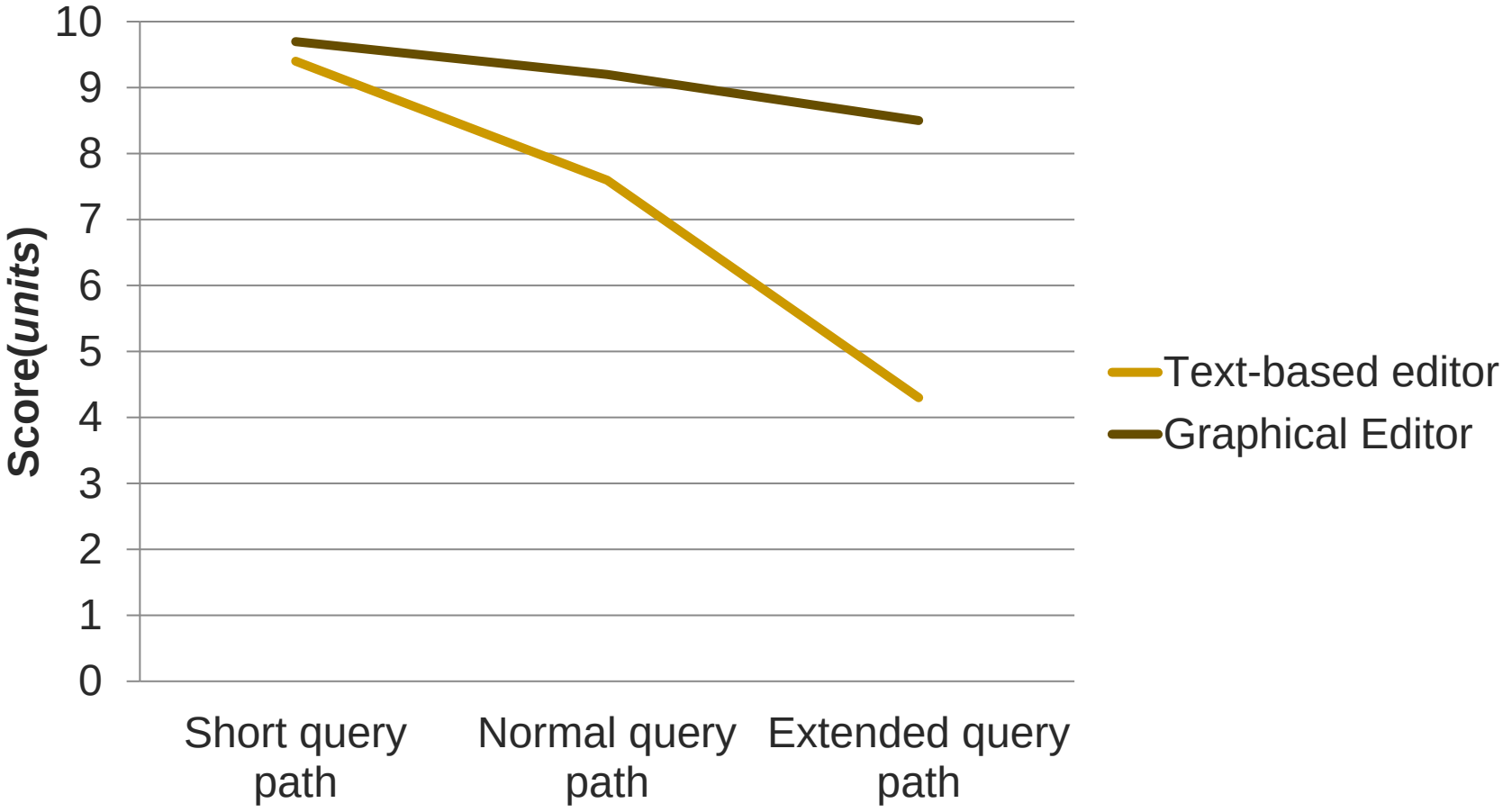
Convenience of Navigating and Understanding the Paths (1/2)

- *Tasks for this test are:*
  - Navigate through specific paths
  - Understand the paths flow and make small changes correspondingly

	Text-based Configuration Tool	Graphical Configuration Tool
Short query path	9.4	9.7
Normal query path	7.6	9.2
Extended query path	4.3	8.5

Score of these tests (*0-10*)

Convenience of Navigating and Understanding the Paths (2/2)



FUTURE WORK

Future Work

- *Extended validation tests*
- *Connection with instances and immediate stepwise execution*
- *Implement more querying features*
- *Implement extra plugins according to the needs*

Contributions

Approach:

- *Simple syntax to write path expressions - machine and human readable*
- *Graphical editing and online validation*
- *Graphical visualization for better comprehension*

- *SPARQL avoidance*
- *Time saving*
- *Simplified querying*

Video Demo

Questions?

Thank you for your attention!