Rule-based query answering method for a knowledge base of economic crimes

PhD thesis
@
RuleML 2011 Doctoral Consortium

Jaroslaw Bak
Institute of Control and Information Engineering
Poznan University of Technology

Outline

1. Problem statement
2. Overview of existing solutions
3. Rule-based query answering methods
   - with hybrid reasoning
   - with forward reasoning and extended rules
4. Rule-based mapping
5. Knowledge base of economic crimes – The Hydra case
6. Example queries and results
7. Conclusion and future work

Problem statement (1)

Research question:

How to efficiently query a relational database at the conceptual level defined in a rule-based system?

Problem statement (2)

Three main issues:

1. Rule-based query answering.
2. The combination of a rule-based system and a relational database.
3. The construction of the knowledge base (i.e. knowledge of economic crimes).

Overview of existing solutions (1)

- Backward chaining vs. forward chaining and OpenRuleBench initiative – Prolog wins [Bry et al]
- Magic Transformation [Ramakrishnan et al, Eiter et al) – a program *P* and a query *Q* are transformed into a new program, *magic(P∪Q)*
- Implementation alternatives for bottom-up evaluation – push and pull methods [Brass]

Overview of existing solutions (2)

- The combination of rules with relational databases [Motik et al., Calvanese et al, Lukácsy et al, Hustadt et al]
- The problem of applying rules and ontology in economic crimes is quite new – some works in information management exist [Biasotii et al, Casellas et al], not reasoning [Breuker]
- FFPoirot project – fraudulent Internet investment pages and the Nigerian letter fraud [Zhao et al]

Rule-based query answering method

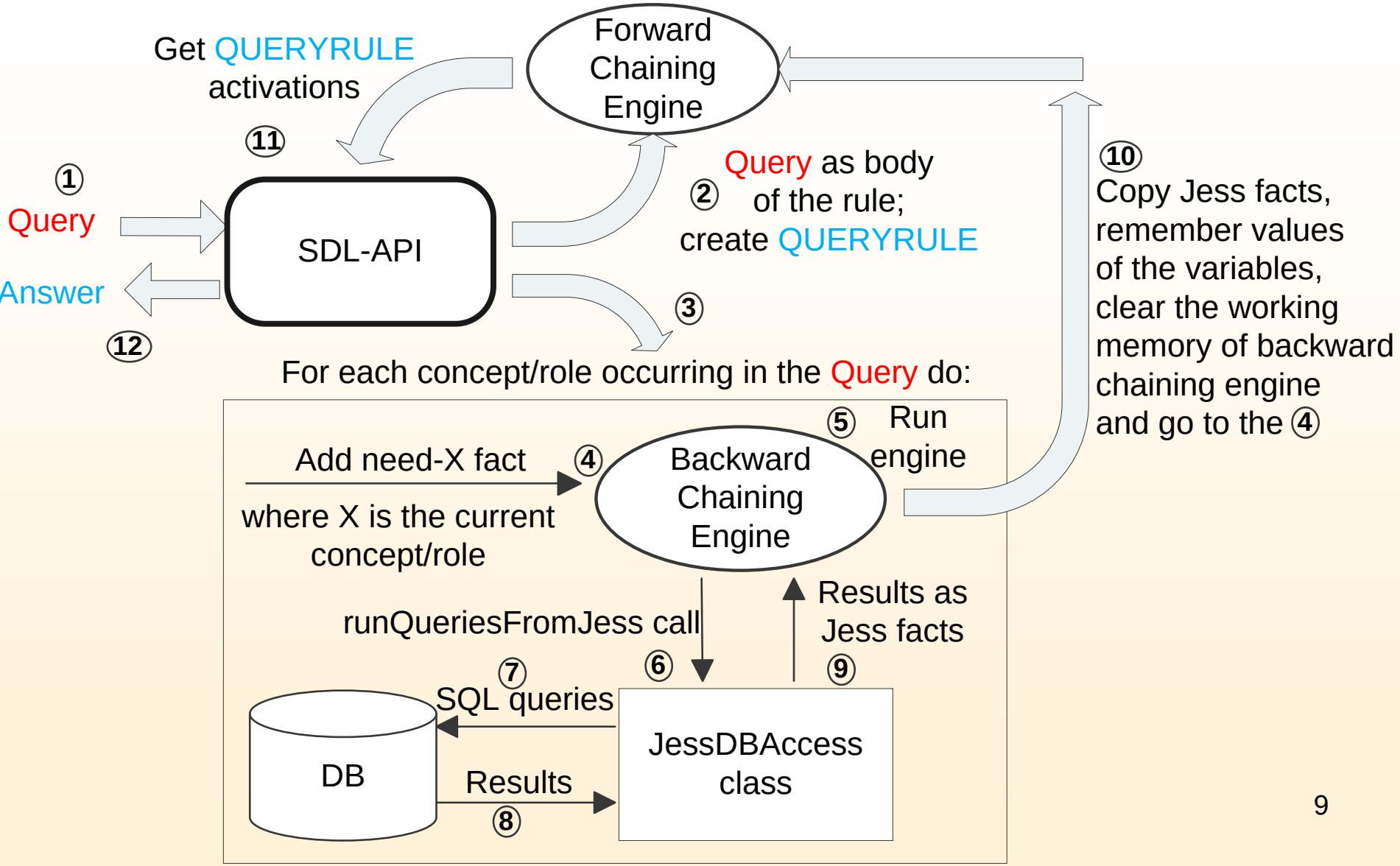
Assumptions

- We use a production rule system (the Jess engine) with forward and backward chaining
- We express the conceptual knowledge with the Horn-SHIQ ontology combined with SWRL (Horn-like) rules
- We use conjunctive queries in the form of directed graphs
- We use Datalog Safety and DL-safe rules to achieve the decidability of our system
- We apply our methods with the knowledge base of economic crimes: fraudulent disbursement and money laundering
- Our approaches are implemented in the Semantic Data Library (SDL)

Rule-based query answering method
Hybrid reasoning (1)

- Two Jess instances:
  - backward engine is used to gather data from a relational database
  - forward engine is used to answer a query
- Rules are expressed in the Jess language and obtained from the transformation of an OWL ontology with SWRL rules
- We developed an algorithm for grouping SQL queries
- This approach was presented at the RuleML2009 conference

Rule-based query answering method
Hybrid reasoning (2)



Rule-based query answering method
Forward reasoning with extended rules (1)

- Based on the modified magic transformation
- We use *extended* rules - rules that are generated automatically from the basic ones for the evaluation purposes
- Extended rules are query-independent – in contrast to the magic transformation
- Generation of the extended rules are based on the particular *sip* algorithm – goal- and dependency-oriented generation
- One instance of the Jess engine is used
- This method is more efficient than the hybrid one
- Accepted for the presentation at the RuleML2011@BRF conference

Rule-based query answering method

Forward reasoning with extended rules (2)

$$p_1(?x,?y),\ p_2(?y,?w),\ p_3(?w),?w \neq ?x \Rightarrow h_1(?x,?w)$$

$$p_1(?x,?y), p_2(?y,?w), p_3(?w),?w \neq ?x,\ h_1(?x,?w)^C \Rightarrow h_1(?x,?w)$$

$$h_1(nil,nil)^C \Rightarrow\ p_1(nil,nil)^C$$

$$h_1(nil,nil)^C \Rightarrow\ p_2(nil,nil)^C$$

Etc.

Applicable to: h_1(?x, ?w)

$$h_1(?x,?)^C,\ \ p_1(?x,?y) \Rightarrow\ p_2(?y,nil)^C$$

$$h_1(?x,?)^C,?x \neq nil \Rightarrow p_1(?x,nil)^C$$

Etc.

Applicable to: h_1(V, ?w) and h1(V, V)

$$h_1(?,?w)^C,?w \neq nil \Rightarrow p_2(nil,?w)^C$$

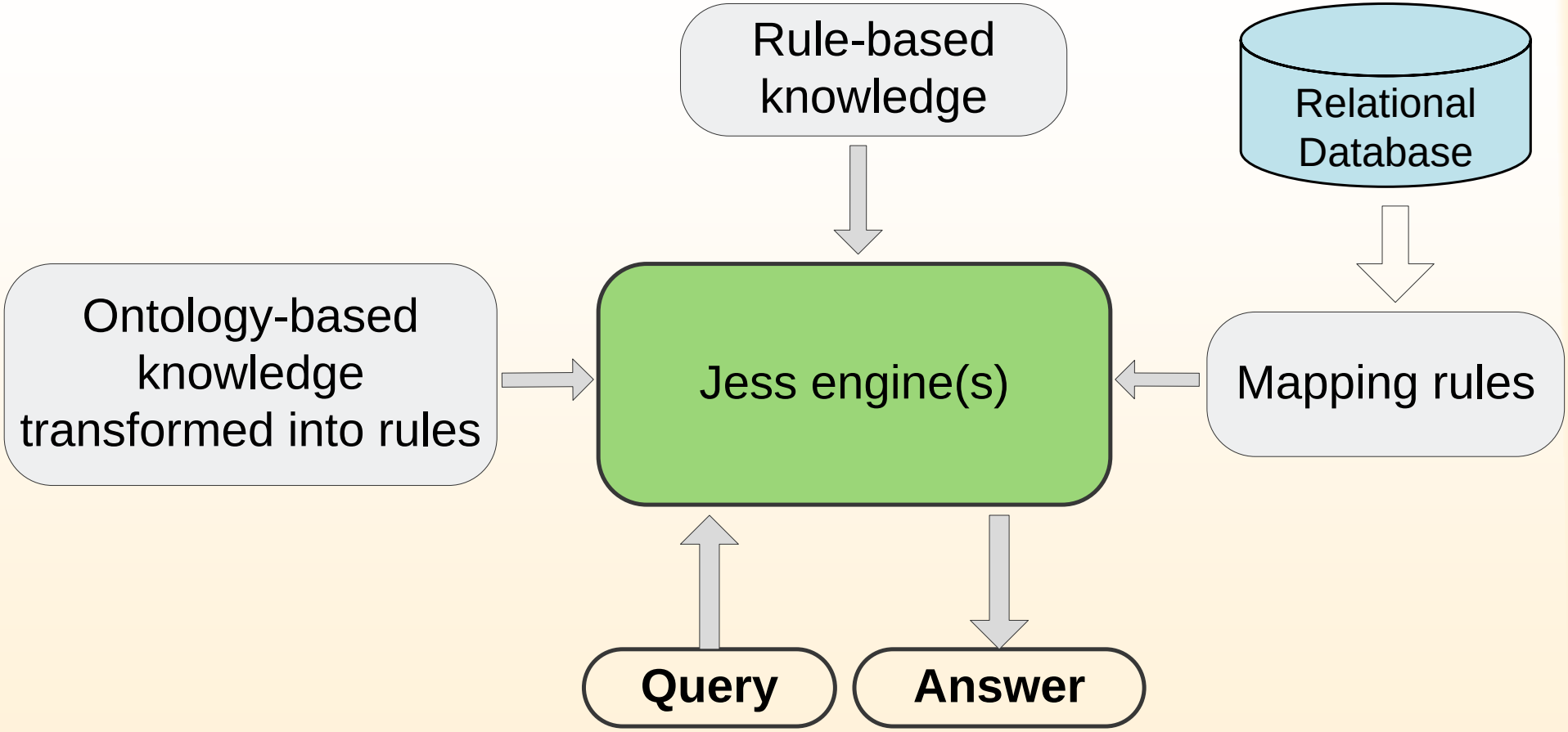
$$h_1(?,?w)^C,\ \ p_2(?y,?w) \Rightarrow\ p_1(nil,?y)^C$$

Applicable to: h_1(?x, V) and h1(V, V)

Etc.

Rule-based query answering method

The architecture of the SDL library



Rule-based mapping

- "Essential" concept/role:

  Buyer is-a Company is-a Institution

- For every essential concept/role a SQL query of the following form is created:

SELECT [R] FROM [T] <WHERE> <C, AND, OR>

where **R** – result columns, **T** – tables, **C** – constraints,
AND, **OR** – optional SQL commands

- Example: AdultMan

```
SELECT id FROM persons
WHERE age>18
AND gender='Male';
```

Rule-based mapping - Example

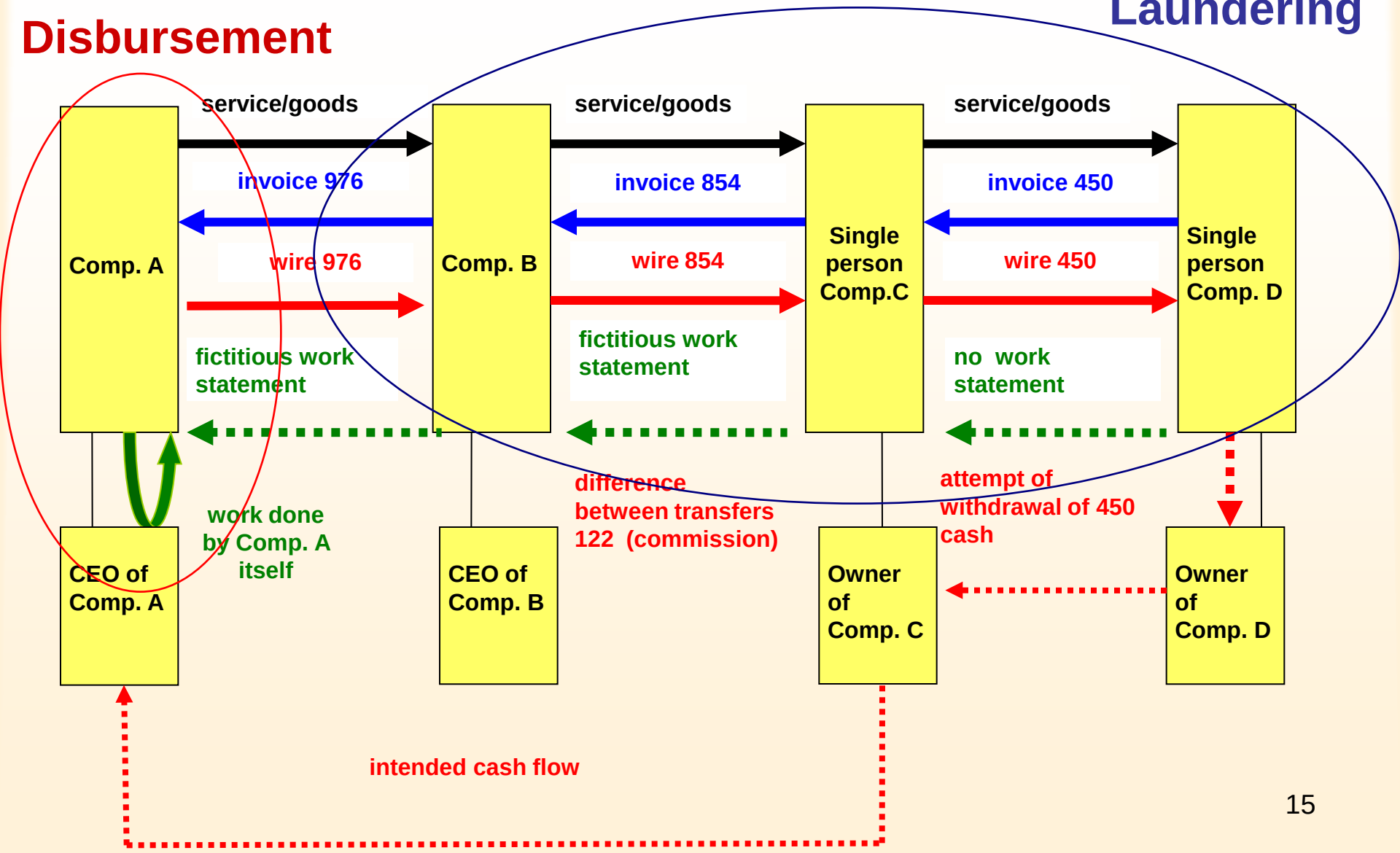
```
(defrule Def-MoneyTransferTo
  ?r<-(need-Triple
          (p "MoneyTransferTo")(s ?x)(o ?y))
 =>
 (bind ?query (str-cat
      "SELECT id, receiver FROM transfers;"))
 (?*access* runQueriesFromJess
     "Def-MoneyTransferTo"
      ?query
      "s;id;o;receiver;p;MoneyTransferTo;"
      (str-cat ?x ";" ?y ";")
      "triple"
      ?*conn*
      (engine))
      (retract ?r)
)
```

Knowledge base of economic crimes

The Hydra Case

Fraudulent Disbursement

Money Laundering



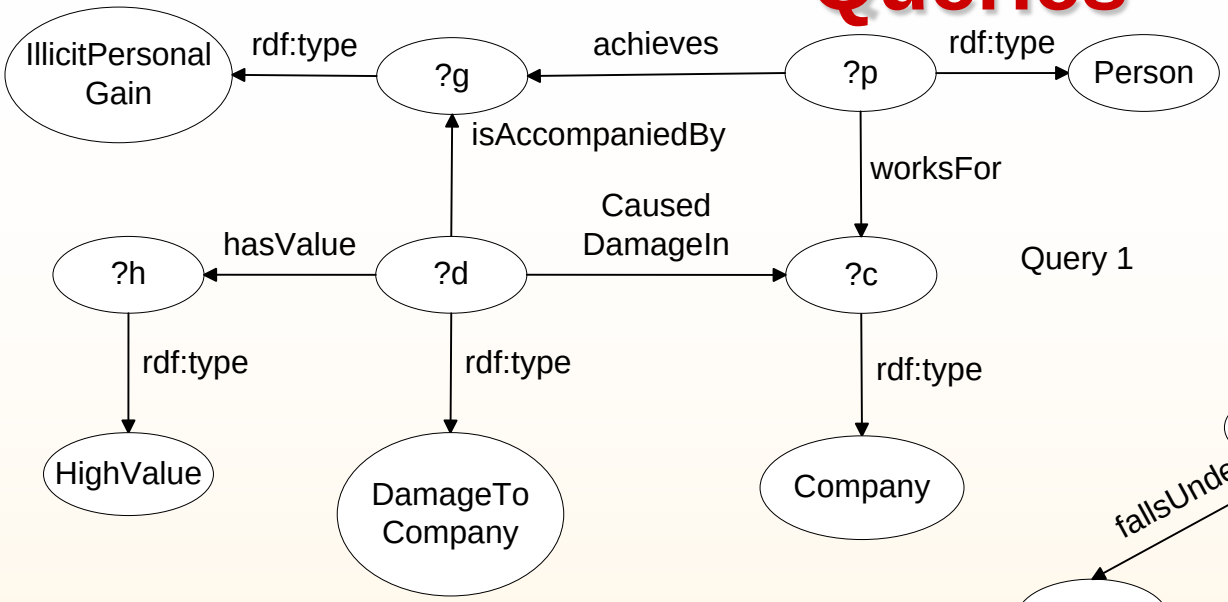
Example queries and results

Generation of the simulated input data

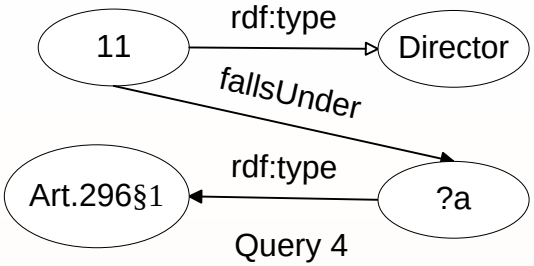
- Information about employees and their position in a company
- Invoices with all obligatory elements (payer, seller, product, etc.)
- Work approval documents (or the lack of them)
- Signatures on documents
- Goods and services
- Companies and their legal form
- Money turnovers: money transfers, payments and withdrawals
- Legal articles (name, ID and content)
- Information about illicit personal gains and damages to companies (with values)
- Other facts, like who knows about what (*Person knowsAbout document*) – these data come from testimonies.

Example queries and results

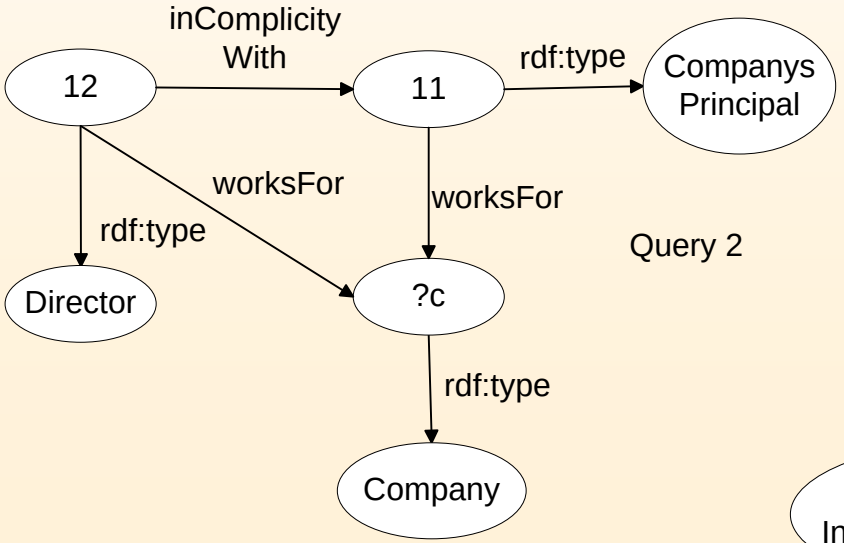
Queries



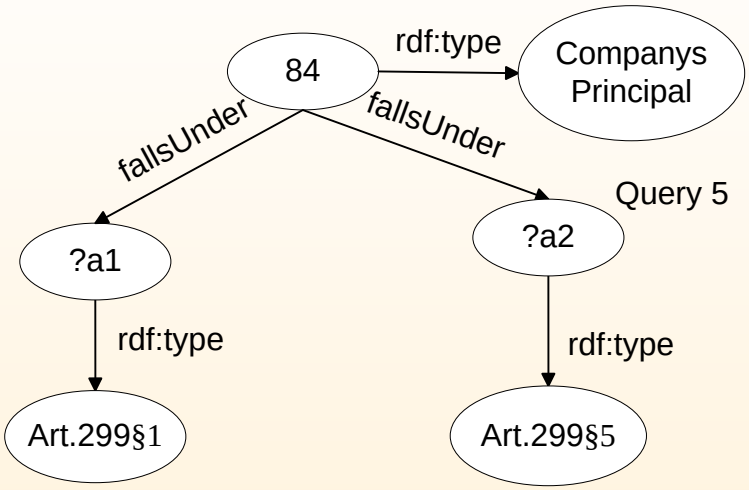
Query 1



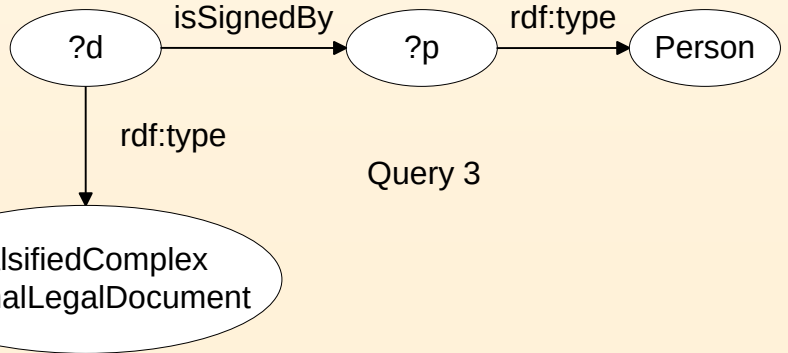
Query 4



Query 2



Query 5



Query 3

Example queries and results

Results

Query and info		Database 20		Database 100		Database 200	
		2010	2011	2010	2011	2010	2011
Query 1	[ms]	781	219	1328	891	1922	969
Results	[number]	54	54	474	474	1036	1036
Rules fired	[number]	74	251	441	1 630	796	3 001
Query 2	[ms]	2734	437	37141	4 125	163968	19 391
Results	[number]	1	1	1	1	1	1
Rules fired	[number]	1076	1 506	36260	13 179	225381	29 593
Query 3	[ms]	2875	359	36344	14 938	183047	116 593
Results	[number]	18	18	322	322	1004	1004
Rules fired	[number]	1367	2 005	38457	41 755	232583	359 681
Query 4	[ms]	5437	1 859	128719	35 656	Time exceeded 10 minutes	347 110
Results	[number]	1	1	1	1		1
Rules fired	[number]	2040	5 467	57091	58 520		597 711
Query 5	[ms]	9312	1 234	Time exceeded 10 minutes	34 500	Time exceeded 10 minutes	343 469
Results	[number]	1	1		1		1
Rules fired	[number]	2540	5 828		61 199		608 925

Conclusions

- We developed two approaches for rule-based query answering task: the hybrid reasoning and extended rules
- Approaches extend functionality of the Jess reasoning engine
- Approaches were tested with our knowledge base of economic crimes

- SDL library enables to query a relational database in terms of ontology concepts/roles
- Answer is always up-to-date
- Queries are in the form of directed graphs

Future work

- More optimizations in the query answering method
- Method for rule-dependent sips
- Comparison to other approaches (OpenRuleBench, OWLim, DLEJena etc.)
- Extension of rule-based knowledge base of economic crimes
- More formal description
- Graphical user interface for queries execution, rules creation and data (facts) analysis

- FINISH THE PhD ☺

THE END

Thanks for listening! ☺