Reasoning with Horn Description Logics Ontologies and Knowledge Graphs

David Carral, Ph.D.

TECHNISCHE UNIVERSITÄT DRESDEN

Slides available at
https://iccl.inf.tu-dresden.de/web/Reasoning-Horn-DLs-KGs/en

Motivation

Knowledge Graphs

What is a Knowledge Graph?

A Knowledge Graph is a data repository that is:

* **Normalised:** Data is decomposed into small units (“edges”)
* **Connected:** Knowledge is represented by relationships between these units

Extending KGs with **OWL terminological axioms**:

* Data integration
* Information extraction
* Automatic population
* Debugging

Wid **f OWL**
onto ith
Kno s!

Well, not really…

Scalability

KGs contain **large amounts of assertional information**:

WIKIDATA

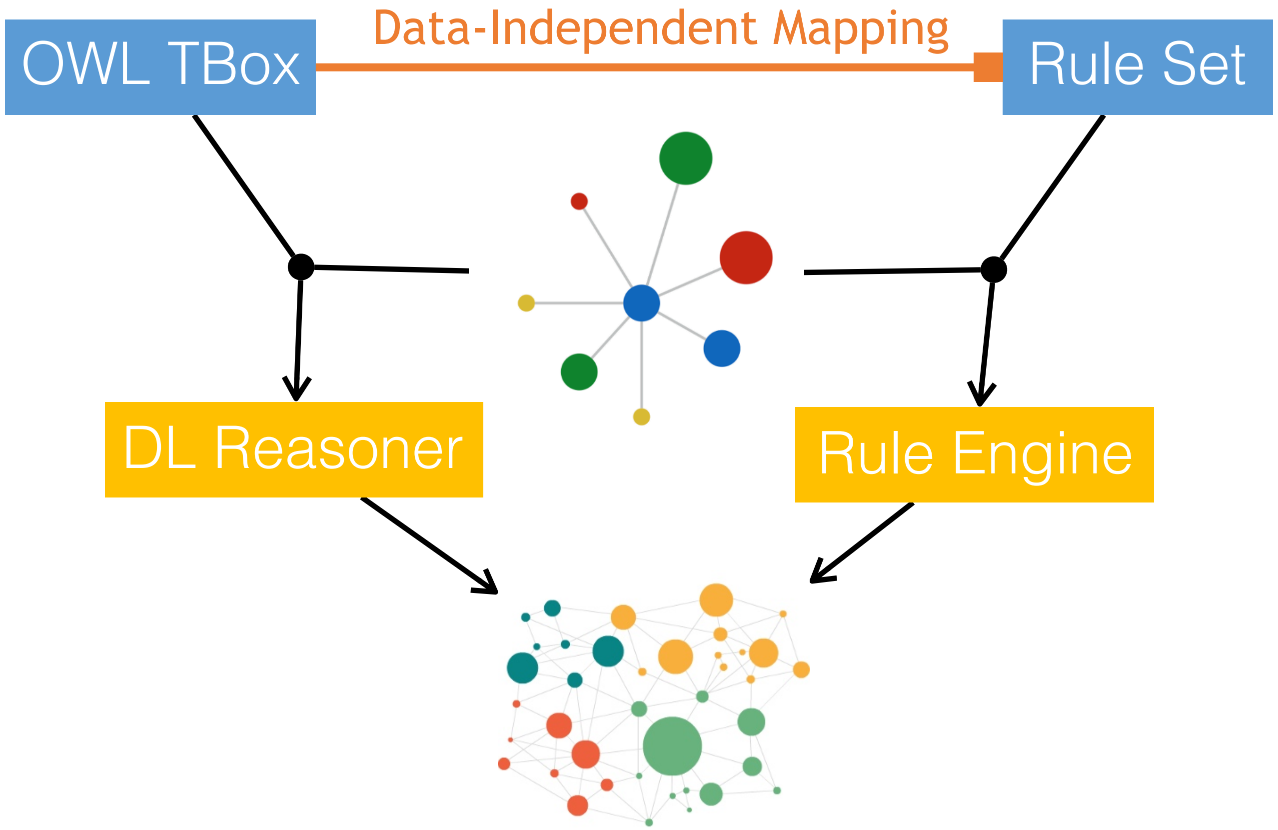
50×10^6 statements

70×10^9 facts, > 570×10^6 entities

Lack of tool support:

* **Traditional KR/DL View**: **Schema first** (class consistency, classification…)
* **Knowledge Graphs**: **Data first** (instance retrieval, CQ answering…)

From OWL to Rules



In **theory**:
* Correctness
* Complexity

In **practice**:
* Implement transformations
* Evaluate performance
* Further develop and optimise rule engines

From OWL to Rules

Acyclicity Notions

* A Practical Acyclicity Notion for Query Answering over Horn-SRIQ Ontologies [ISWC 2016]
* Restricted Chase (Non)Termination for Existential Rules with Disjunctions [IJCAI 2017]
* Tractable Query Answering for Expressive Ontologies and Rules [ISWC 2017]

The Combined Approach

* Pushing the Boundaries of Tractable Ontology Reasoning [ISWC 2014]
* The Combined Approach to Query Answering Beyond the OWL 2 Profiles [IJCAI 2015]
* The Combined Approach to Query Answering Horn-ALCHOIQ [KR 2018]

Reasoning with Nominal Schemas

* Towards an Efficient Reasoning Algorithm to Reason over DLs Extended with Nominal Schemas [RR 2013]

Acyclicity Notions for Horn DL Ontologies

The DL Horn-SROIQ

$$
\begin{aligned}
C_1 \sqcap \ldots \sqcap C_n &\sqsubseteq D \\
C &\sqsubseteq \forall S.D \\
C &\sqsubseteq \exists S.D \\
C &\sqsubseteq \leqslant 1\, S.D \\
C &\sqsubseteq \{d\} \\
R_1 \circ \ldots \circ R_n &\sqsubseteq S
\end{aligned}
$$

Figure 1: Where $C_{(i),D} \in \mathbf{C}$, $R, S \in \mathbf{R}$, and $d \in \mathbf{I}$.

The DL Horn-SROIQ

$$C_1 \sqcap \ldots \sqcap C_n \sqsubseteq D$$

$$C \sqsubseteq \forall S.D$$

$$\boxed{C \sqsubseteq \exists S.D}$$

$$C \sqsubseteq \leqslant 1\, S.D$$

$$C \sqsubseteq \{d\}$$

$$R_1 \circ \ldots \circ R_n \sqsubseteq S$$

Figure 1: Where $C_{(i)}, D \in \mathbf{C}$, $R, S \in \mathbf{R}$, and $d \in \mathbf{I}$.

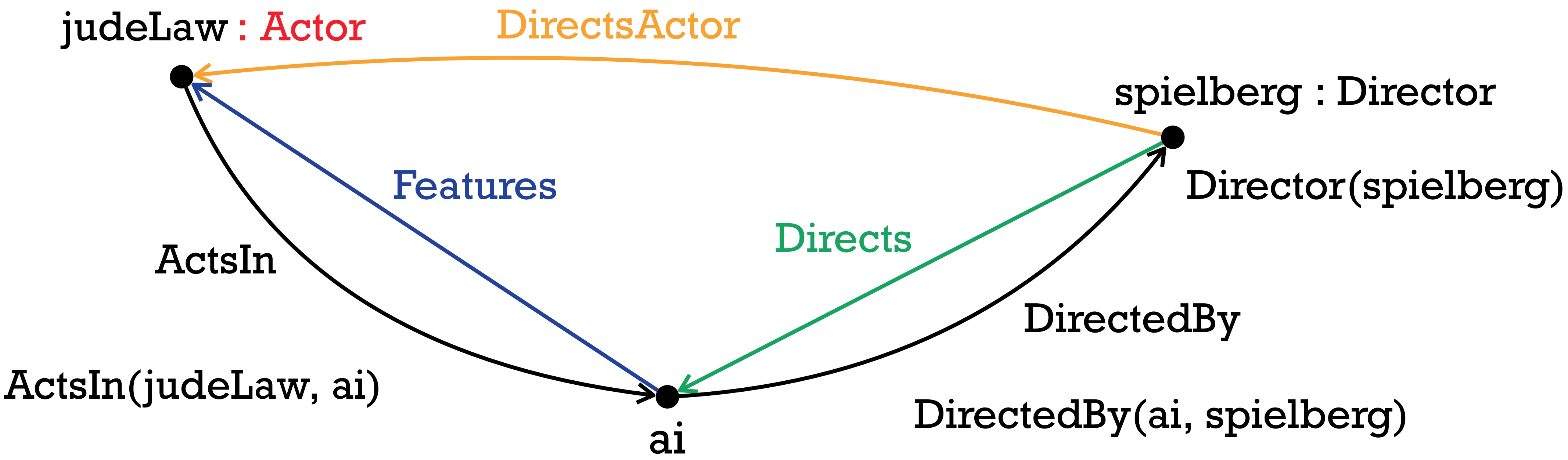
The Chase Algorithm

$\mathrm{Features}(x, y) \longrightarrow \mathrm{Actor}(y)$

$\mathrm{ActsIn}(x, y) \longrightarrow \mathrm{Features}(y, x)$

$\mathrm{DirectedBy}(x, y) \longrightarrow \mathrm{Directs}(y, x)$

$\mathrm{Directs}(x, y) \wedge \mathrm{Features}(y, z) \longrightarrow \mathrm{DirectsActor}(x, z)$



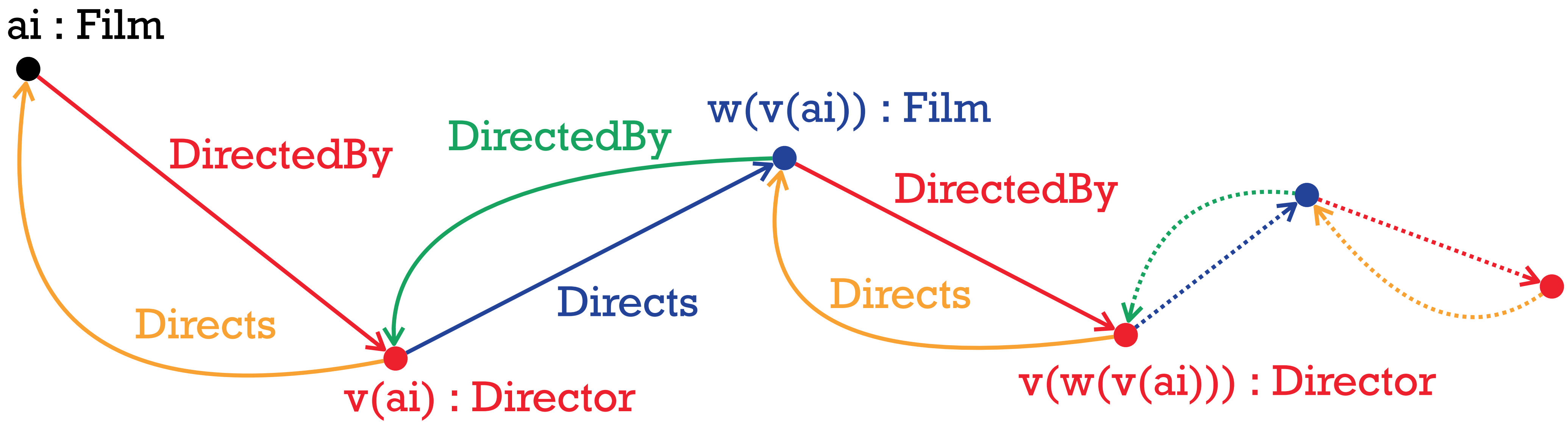
The Chase Algorithm

$\mathrm{Film}(x) \longrightarrow \exists v\, .\, \mathrm{DirectedBy}(x, v) \wedge \mathrm{Director}(v)$

$\mathrm{Director}(x) \longrightarrow \exists w\, .\, \mathrm{Directs}(x, w) \wedge \mathrm{Film}(w)$

$\mathrm{Directs}(x, y) \longrightarrow \mathrm{DirectedBy}(y, x)$

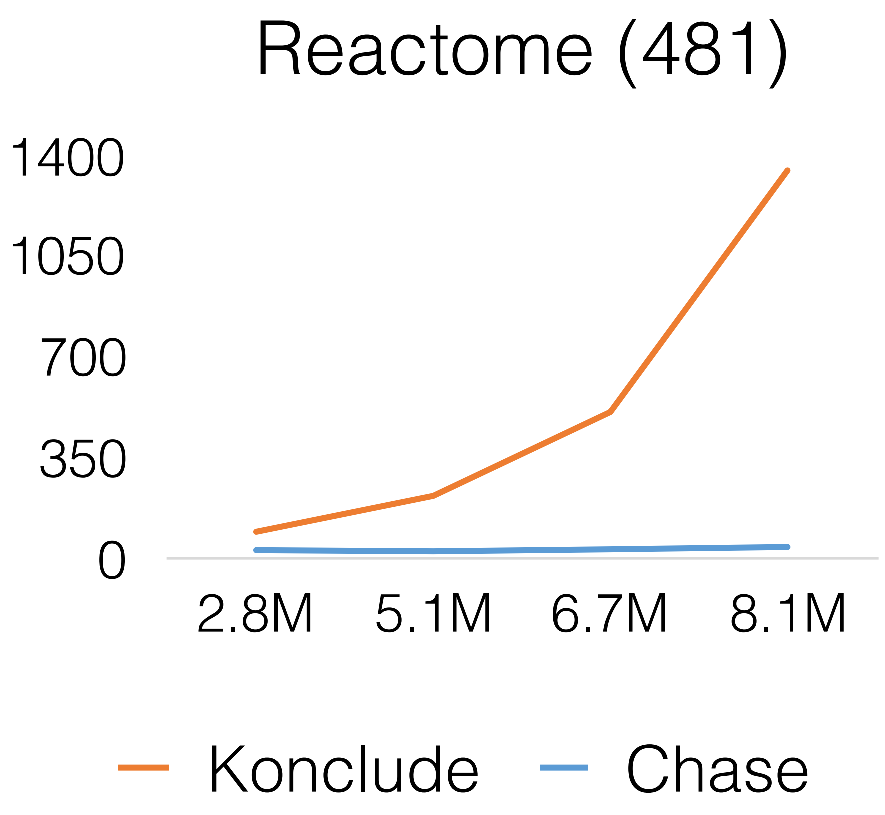
$\mathrm{DirectedBy}(x, y) \longrightarrow \mathrm{Directs}(y, x)$

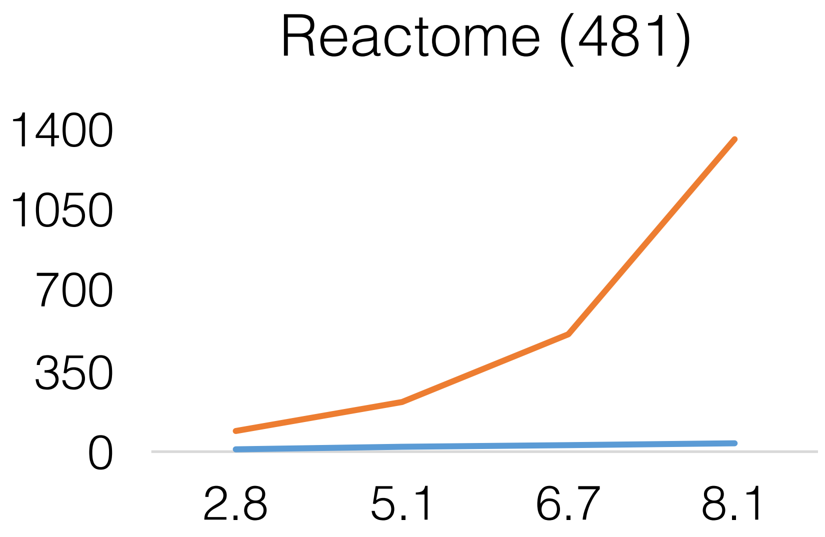


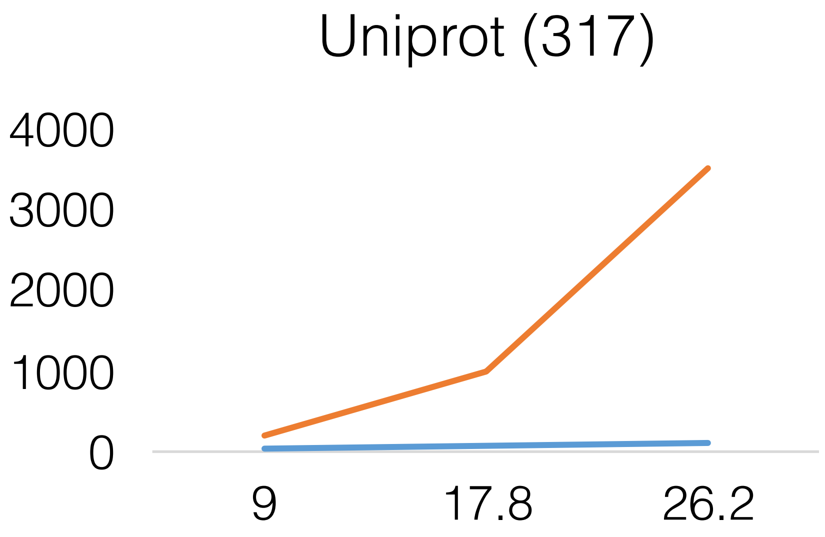
Motivation

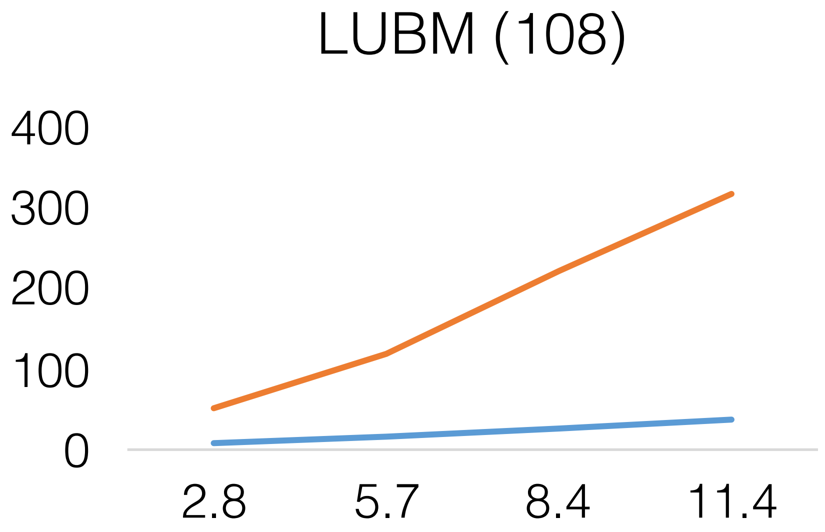
* Performance.
  * DL Reasoner: Konclude
  * Chase Implementation using the Datalog reasoner RDFox

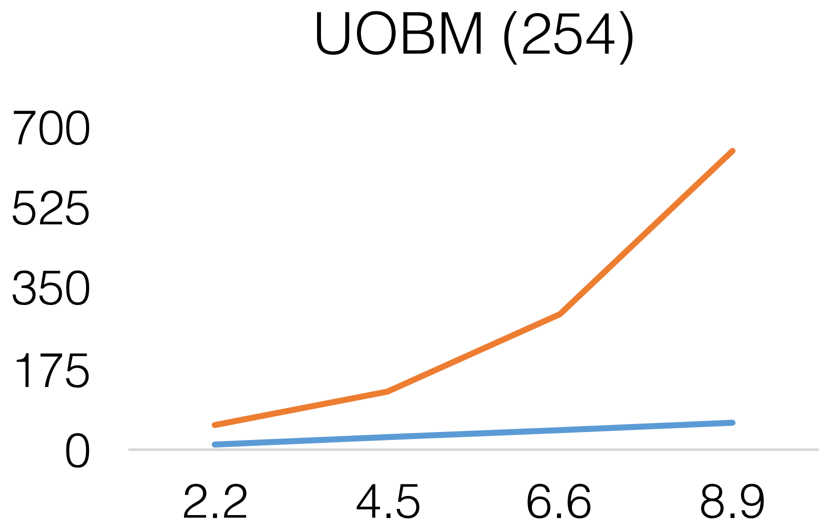
* Solves **conjunctive query answering**.




Reactome (481)
1400
1050
700
350
0
2.8
5.1
6.7
8.1


Uniprot (317)
4000
3000
2000
1000
0
9
17.8
26.2


LUBM (108)
400
300
200
100
0
2.8
5.7
8.4
11.4


UOBM (254)
700
525
350
175
0
2.2
4.5
6.6
8.9

Acyclicity Notions

* Weak Acyclicity (**WA**) [Theor. Comput. Sci. 2005]

* Joint Acyclicity (**JA**) [IJCAI 2011]

* Model-Summarising Acyclicity (**MSA**) and
  Model-Faithful Acyclicity (**MFA**) [J. Artif. Intell. Res. 2013]

* Restricted Joint Acyclicity (**RJA**),
  Restricted Model-Summarising Acyclicity (**RMSA**), and
  Restricted Model-Faithful Acyclicity (**RMFA**) [IJCAI 2017]

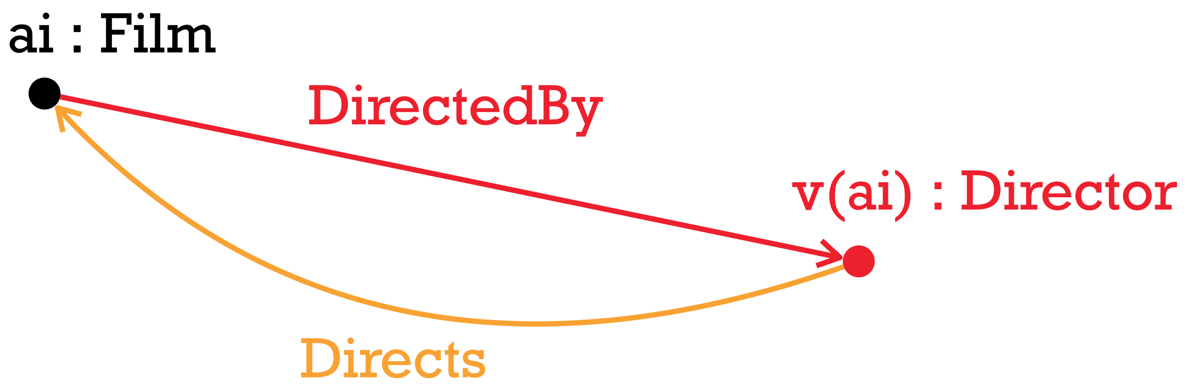
Datalog-first Restricted Chase

$\mathrm{Film}(x) \longrightarrow \exists v\,.\,\mathrm{DirectedBy}(x, v) \wedge \mathrm{Director}(v)$

$\mathrm{Directs}(x, y) \longrightarrow \mathrm{DirectedBy}(y, x)$

$\mathrm{Director}(x) \longrightarrow \exists w\,.\,\mathrm{Directs}(x, w) \wedge \mathrm{Film}(w)$

$\mathrm{DirectedBy}(x, y) \longrightarrow \mathrm{Directs}(y, x)$



The MFA Check

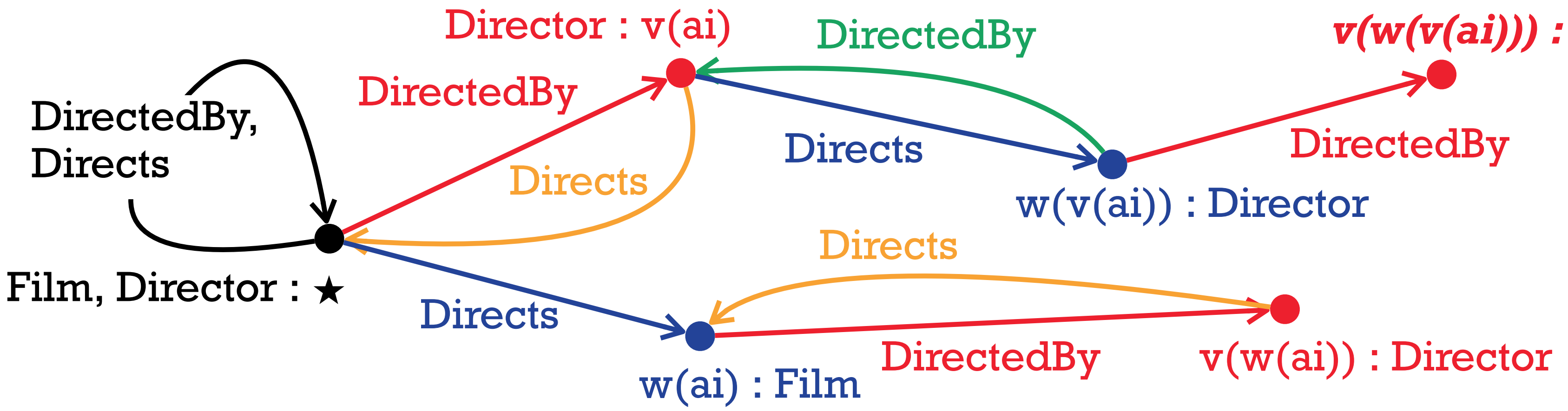
* Approach: perform chase, check if it stops; give up if a **cyclic skolem term** (with a repeated function symbol) appears.
* Termination may depend on given facts but, if the approach terminates on the **critical instance** (the set of all possible facts using a single constant "★"), then it terminates on all sets of facts.

$\mathrm{Film}(x) \longrightarrow \exists v\,.\,\mathrm{DirectedBy}(x, v) \wedge \mathrm{Director}(v)$

$\mathrm{Director}(x) \longrightarrow \exists w\,.\,\mathrm{Directs}(x, w) \wedge \mathrm{Film}(w)$

$\mathrm{Directs}(x, y) \longrightarrow \mathrm{DirectedBy}(y, x)$

$\mathrm{DirectedBy}(x, y) \longrightarrow \mathrm{Directs}(y, x)$



Not MFA!

The RMFA Check: Blocked Checks

* Problem: restricted chase termination is not monotone!
* In particular, it always terminates on the critical instance.
* **Idea:** for each fact that occurs in the chase sequence, we can re-trace a necessary fact set the must have been derived to derive this fact. By checking these facts we can in some cases determine that the application of the rule and substitution that generate this fact are **blocked**.

$\mathrm{Film}(x) \longrightarrow \exists v \,.\, \mathrm{DirectedBy}(x, v) \wedge \mathrm{Director}(v)$

$\mathrm{Director}(x) \longrightarrow \exists w \,.\, \mathrm{Directs}(x, w) \wedge \mathrm{Film}(w)$

$\mathrm{Directs}(x, y) \longrightarrow \mathrm{DirectedBy}(y, x)$

$\mathrm{DirectedBy}(x, y) \longrightarrow \mathrm{Directs}(y, x)$

Example: Suppose for a contradiction that the fact $\mathrm{Director}(v(w(c)))$ is derived during the computation of a chase sequence.

* Such a fact may only be derived via application of the **red rule** on the fact $\mathrm{Film}(w(c))$ which, in turn, may only occur in the chase via the application of the **blue rule**. Therefore, facts $\mathrm{Director}(c)$ and $\mathrm{Directs}(c, w(c))$ are also part of the chase when the fact $\mathrm{Director}(v(w(c)))$ is derived.
* Because the **green rule** is Datalog, $\mathrm{DirectedBy}(v(c), c)$ is also part of the chase (**Datalog-first!**).
* The **red rule** may not be applied to introduce a fact such as $\mathrm{Director}(v(w(c)))$.

The RMFA Check

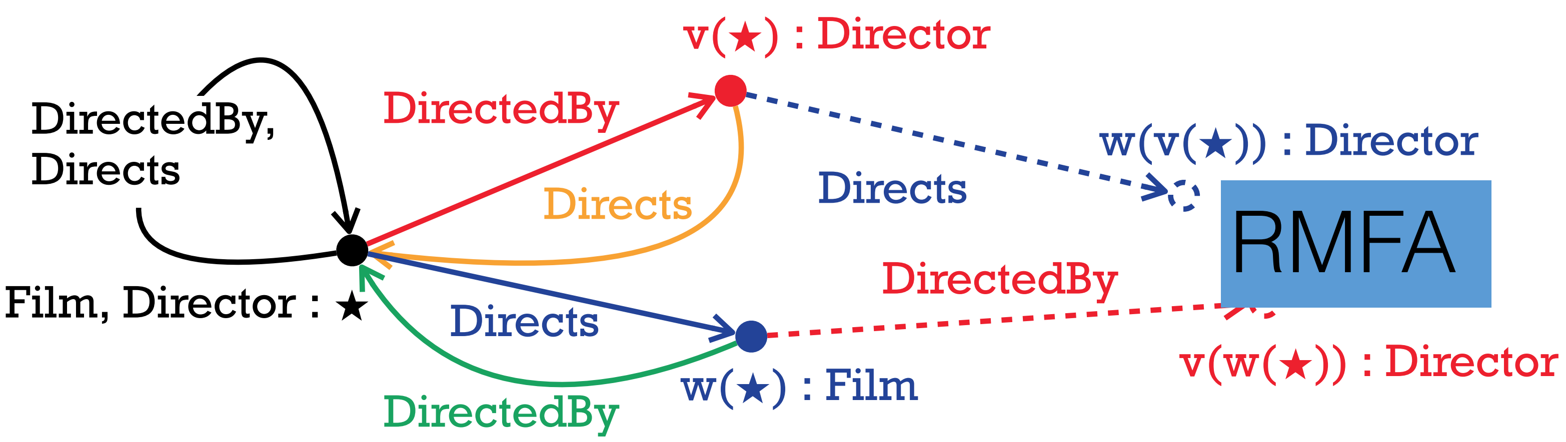
* Perform a chase like construction on the critical instance.
* Only apply an existential rule with respect to a substitution if this pair is not **blocked**.
* Give up if the procedure does not stop before the occurrence of a cyclic term.

$\mathrm{Film}(x) \longrightarrow \exists v\ .\ \mathrm{DirectedBy}(x, v) \wedge \mathrm{Director}(v)$

$\mathrm{Director}(x) \longrightarrow \exists w\ .\ \mathrm{Directs}(x, w) \wedge \mathrm{Film}(w)$

$\mathrm{Directs}(x, y) \longrightarrow \mathrm{DirectedBy}(y, x)$

$\mathrm{DirectedBy}(x, y) \longrightarrow \mathrm{Directs}(y, x)$

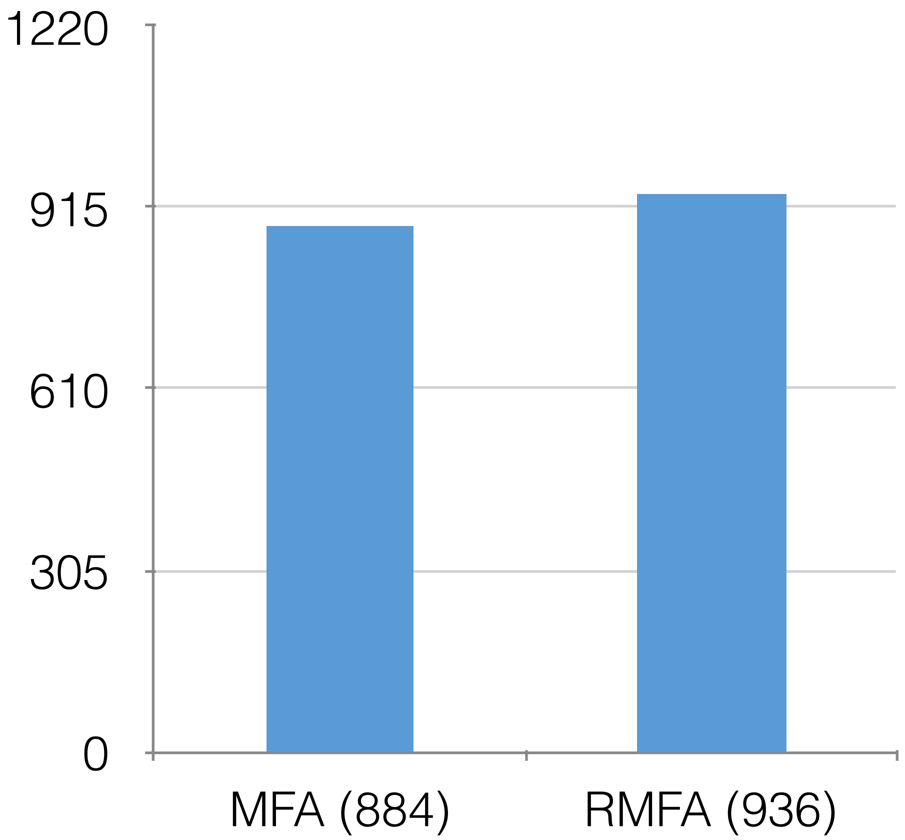


* Theorem [IJCAI 2017]: Deciding if a rule set is RMFA is 2ExpTime-complete.
* Slightly better bounds for DL Horn ontologies (ExpTime-complete).
* In practice: Did not encounter major performance issues even for a prototype implementation.

Real-world Coverage

* We studied 1220 ontologies obtained from two sources: MOWLCorp and Oxford Ontology Library.

* We developed a cyclicity notion, i.e., sufficient conditions for chase non-termination: Restricted Model-Faithful Cyclicity (**RMFC**)



+6% **+54%**

Restricted Chase: VLog Implementation

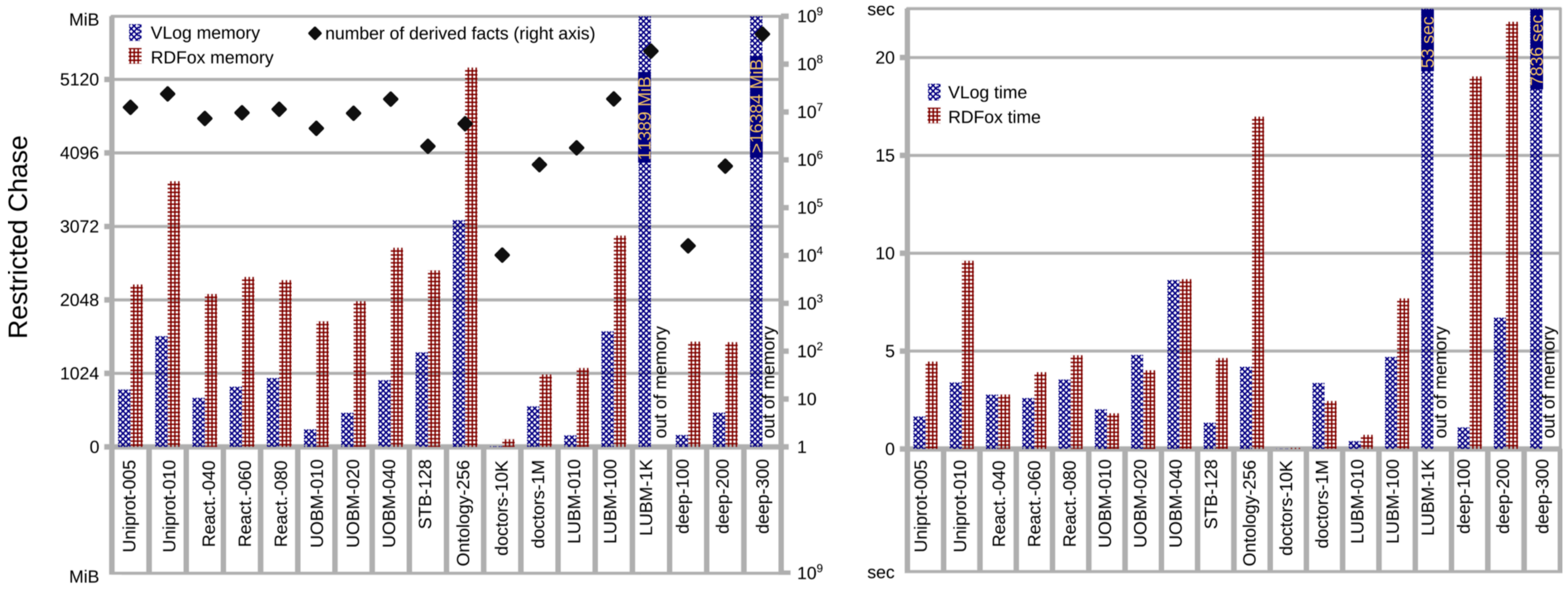


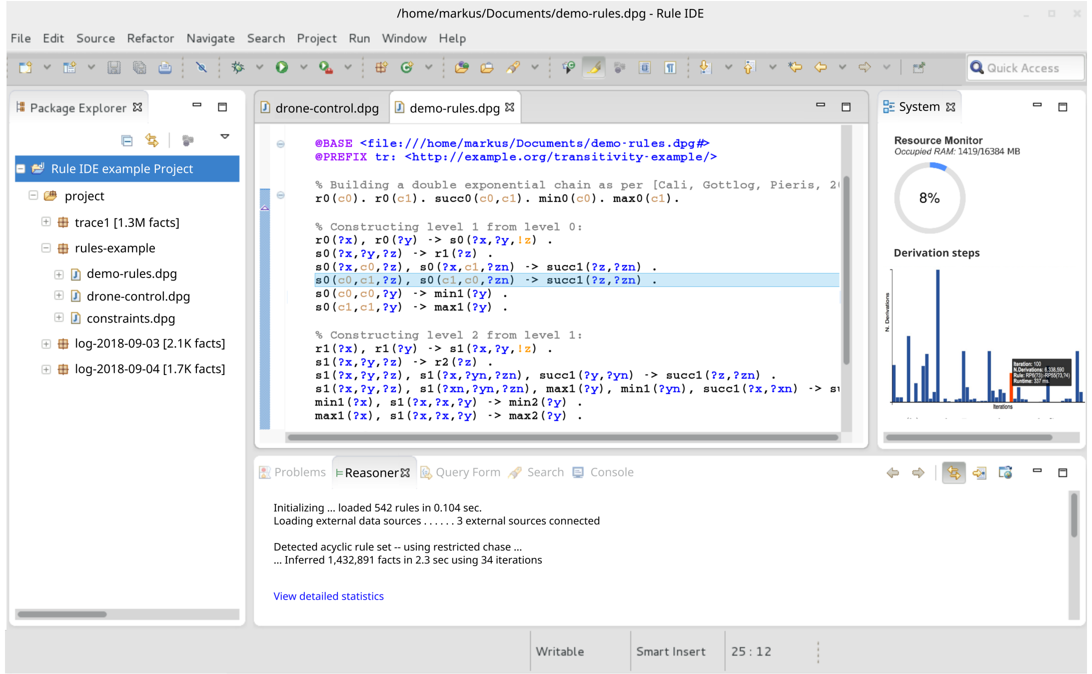
Fig. 1. Memory usage (left) and materialisation time (right) for VLog and RDFox

Conclusions

Problem Solved?

Once we develop efficient reasoners for KGs extended with OWL we are set!
Surely not, a lot of work still needs to be done:

* Expressive languages for KR on the web (**existential rules**)
* Support in KR languages (and implementations) with some form of **numerical reasoning** and can deal with metadata (**annotated logics**)
* Development of tools that can assist knowledges engineers on the creation of terminologies for KGs



Reasoning with Horn Description Logics Ontologies and Knowledge Graphs

David Carral, Ph.D.

TECHNISCHE UNIVERSITÄT DRESDEN

Slides available at
https://iccl.inf.tu-dresden.de/web/Reasoning-Horn-DLs-KGs/en

Bibliography

Bibliography: Rule Engines

* Efficient Model Construction for Horn Logic with VLog [IJCAR 2018]
  https://iccl.inf.tu-dresden.de/web/Article3046/en
* The Vadalog System: Datalog-based Reasoning for Knowledge Graphs [PVLDB 11]
  http://www.vldb.org/pvldb/vol11/p975-bellomarini.pdf
* Column-Oriented Datalog Materialization for Large Knowledge Graphs [AAAI 2016]
  http://korrekt.org/papers/Urbani-Jacobs-Kroetzsch_Vlog-datalog-materialization-AAAI2016.pdf
* Parallel Materialisation of Datalog Programs in Centralised, Main-Memory RDF Systems [AAAI 2014]
  http://www.cs.ox.ac.uk/ian.horrocks/Publications/download/2014/MNPHO14a.pdf

Bibliography: Data-preserving Transformations

* The Combined Approach to Query Answering Horn-ALCHOIQ [KR 2018] https://iccl.inf.tu-dresden.de/web/Inproceedings3098/en
* Restricted Chase (Non)Termination for Existential Rules with Disjunctions [IJCAI 2017] https://iccl.inf.tu-dresden.de/web/Inproceedings3140/en
* Tractable Query Answering for Expressive Ontologies and Rules [ISWC 2017] https://iccl.inf.tu-dresden.de/web/Inproceedings3163/en
* A Practical Acyclicity Notion for Query Answering over Horn-SRIQ Ontologies [ISWC 2016] https://iccl.inf.tu-dresden.de/web/Inproceedings3134/en
* The Combined Approach to Query Answering Beyond the OWL 2 Profiles [IJCAI 2015] https://iccl.inf.tu-dresden.de/web/Inproceedings3131/en
* Pushing the Boundaries of Tractable Ontology Reasoning [ISWC 2014] https://iccl.inf.tu-dresden.de/web/Inproceedings3122/en
* Towards an Efficient Reasoning Algorithm to Reason over DLs Extended with Nominal Schemas [RR 2013] https://iccl.inf.tu-dresden.de/web/Inproceedings3118/en

Bibliography: Acyclicity Notions

* Restricted Model-Faithful Acyclicity (**RMFA)**
  Restricted Chase (Non)Termination for Existential Rules with Disjunctions. [IJCAI 2017]
  https://iccl.inf.tu-dresden.de/web/Inproceedings3140/en
* Model-Faithful Acyclicity (**MFA**)
  Acyclicity Notions for Existential Rules and Their Application to QA in Ontologies [J. Artif. Intell. Res. 47]
  https://iccl.inf.tu-dresden.de/web/Article4005/en
* Joint Acyclicity (**JA**)
  Extending Decidable Existential Rules by Joining Acyclicity and Guardedness [IJACI 2011]
  https://iccl.inf.tu-dresden.de/web/Inproceedings3149/en
* Weak Acyclicity (**WA**).
  Data Exchange: Semantics and Query Answering [Theor. Comput. Sci. 336]
  https://www.sciencedirect.com/science/article/pii/S030439750400725X

SQID: Extending Wikidata with Rules