

Magic-Sets for Datalog with Existential Quantifiers

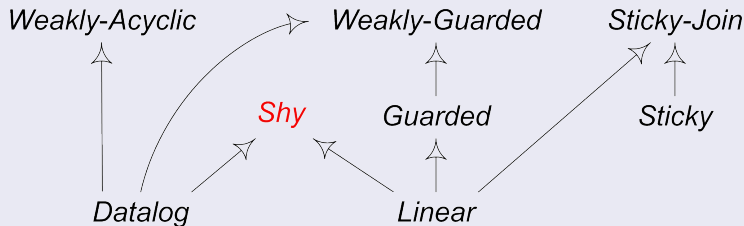
Mario Alviano, Nicola Leone, Marco Manna,
Giorgio Terracina, and Pierfrancesco Veltri

Department of Mathematics, University of Calabria

Datalog 2.0 2012
Vienna, 13 September 2012

- Datalog[∃]
 - Datalog + unrestricted existential quantification
 - Highly expressive
 - Undecidable reasoning

Several decidable fragments



- Datalog programs are usually evaluated bottom-up
 - Queries may be handled inefficiently
- Query optimization in Datalog \implies Magic-Sets
 - Identify relevant atoms for the query

Main contributions

Magic-Sets for Datalog[∃] programs

- Query-equivalent to the original program in general
- Closed on Shy
- Implementation in DLV
- Comparison with ontology reasoners

- Datalog programs are usually evaluated bottom-up
 - Queries may be handled inefficiently
- Query optimization in Datalog \implies Magic-Sets
 - Identify relevant atoms for the query

Main contributions

Magic-Sets for Datalog[∃] programs

- Query-equivalent to the original program in general
- Closed on Shy
- Implementation in DLV
- Comparison with ontology reasoners

Outline

- 1 Datalog[∃]
 - Syntax, semantics and queries
 - CHASE procedure and top-down evaluation
- 2 Magic-Sets
 - Example
 - Application to Shy and experiment

Datalog[∃] programs and queries

Rule structure

$$\forall \mathbf{X} \exists \mathbf{Y} \text{ atom}_{[\mathbf{X}' \cup \mathbf{Y}]} \leftarrow \text{conj}_{[\mathbf{X}]}$$

Query structure

$$\exists \mathbf{Y} \text{ conj}_{[\mathbf{X}]}$$

Example. Datalog[∃] program

$$r_1 : \exists Y \text{ father}(Y, X) \leftarrow \text{father}(X, Z)$$
$$r_2 : \text{grandFather}(X, Y) \leftarrow \text{father}(X, Z), \text{father}(Z, Y)$$

Database

$$\text{father}(a, b) \leftarrow$$
$$\text{father}(b, c) \leftarrow$$

Example. Queries

$$Q_1 : \text{grandFather}(a, Y)?$$
$$Q_2 : \exists Y \text{ grandFather}(Y, a)?$$

Datalog[∃] semantics

Let M be a set of ground atoms, and P be a program with rules

$$r : \forall \mathbf{X} \exists \mathbf{Y} \text{ atom}_{[\mathbf{X}' \cup \mathbf{Y}]} \leftarrow \text{conj}_{[\mathbf{X}]}$$

Atom entailment

$M \models a$ if there is a substitution σ s.t. $\sigma(a) \in M$.

Rule satisfaction

$M \models r$ if $M \models \sigma|_{\mathbf{X}}(\text{atom})$ whenever $\sigma(\text{conj}) \subseteq M$.

Definition (Model)

$M \models P$ if $M \models r$, for all $r \in P$.

Warning!

Unrestricted universe

How many models?

How many subset-minimal models?

$$r_1 : \exists Y \text{ father}(Y, X) \leftarrow \text{father}(X, Z)$$

$$\text{father}(a, b) \leftarrow$$

Datalog[∃] semantics

Let M be a set of ground atoms, and P be a program with rules

$$r : \forall \mathbf{X} \exists \mathbf{Y} \text{ atom}_{[\mathbf{X}' \cup \mathbf{Y}]} \leftarrow \text{conj}_{[\mathbf{X}]}$$

Atom entailment

$M \models a$ if there is a substitution σ s.t. $\sigma(a) \in M$.

Rule satisfaction

$M \models r$ if $M \models \sigma|_{\mathbf{X}}(\text{atom})$ whenever $\sigma(\text{conj}) \subseteq M$.

Definition (Model)

$M \models P$ if $M \models r$, for all $r \in P$.

Warning!

Unrestricted universe

How many models?

How many subset-minimal models?

$$r_1 : \exists Y \text{ father}(Y, X) \leftarrow \text{father}(X, Z)$$

$$\text{father}(a, b) \leftarrow$$

Query answering

Let P be a program, and Q be a query

$$\exists \mathbf{Y} \text{ conj}_{[\mathbf{X}]}$$

Query answers

$$\text{ans}_P(Q) = \{\sigma|_{\mathbf{X}} \mid M \models \sigma|_{\mathbf{X}}(\text{conj}) \text{ for all models } M \text{ of } P\}$$

Infinitely many models

Cannot count them in finite time!

Query answering

Let P be a program, and Q be a query

$$\exists \mathbf{Y} \text{ conj}_{[\mathbf{X}]}$$

Query answers

$$\text{ans}_P(Q) = \{\sigma|_{\mathbf{X}} \mid M \models \sigma|_{\mathbf{X}}(\text{conj}) \text{ for all models } M \text{ of } P\}$$

Infinitely many models

Cannot count them in finite time!

Outline

- 1 Datalog[∃]
 - Syntax, semantics and queries
 - CHASE procedure and top-down evaluation
- 2 Magic-Sets
 - Example
 - Application to Shy and experiment

Universal model

U is a universal model of P if

- $U \models P$ and
- for each model M of P there is a mapping h s.t. $h(U) \subseteq M$.

Theorem (Fagin et al. 2005)

$ans_P(Q) = \{\sigma|_{\mathbf{X}} \mid \sigma(\mathbf{X}) \text{ is a set of constants and } U \models \sigma|_{\mathbf{X}}(\text{conj})\}$

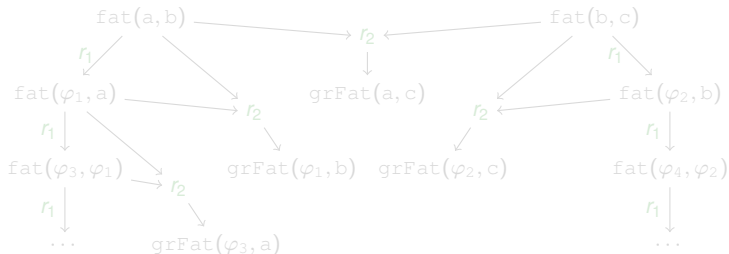
CHASE procedure

Constructs a universal model

CHASE procedure at work

 $D: \text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$ $r_1: \exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$ $r_2: \text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in
general! $\exists Y \text{grFat}(Y, a)?$

Yes!

Not query-driven!

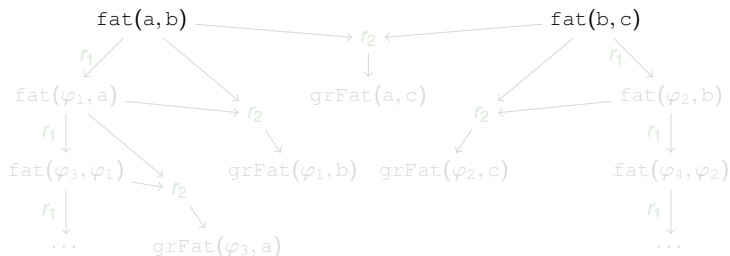
Also computes irrelevant atoms

CHASE procedure at work

D : $\text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$
 r_1 : $\exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$
 r_2 : $\text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in general!

 $\exists Y \text{grFat}(Y, a)?$

Yes!

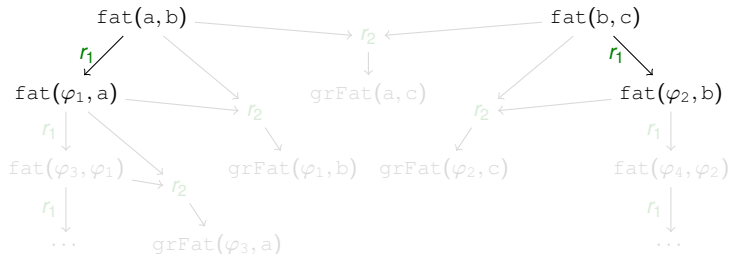
Not query-driven!

Also computes irrelevant atoms

CHASE procedure at work

 $D: \text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$ $r_1: \exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$ $r_2: \text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in
general! $\exists Y \text{grFat}(Y, a)?$

Yes!

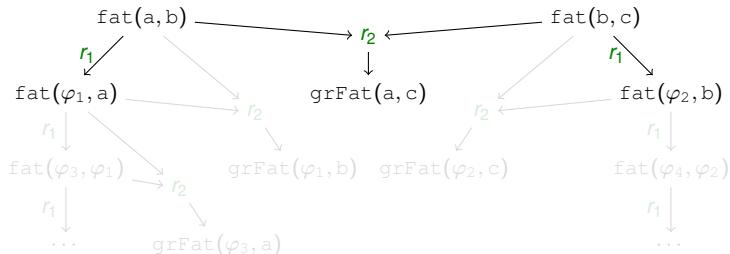
Not query-driven!

Also computes irrelevant atoms

CHASE procedure at work

 $D: \text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$ $r_1: \exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$ $r_2: \text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in
general! $\exists Y \text{grFat}(Y, a)?$

Yes!

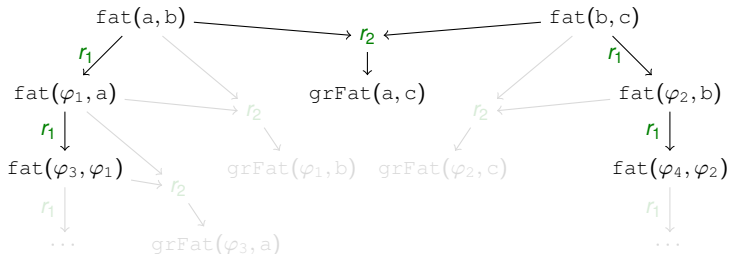
Not query-driven!

Also computes irrelevant atoms

CHASE procedure at work

 $D: \text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$ $r_1: \exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$ $r_2: \text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in
general! $\exists Y \text{grFat}(Y, a)?$

Yes!

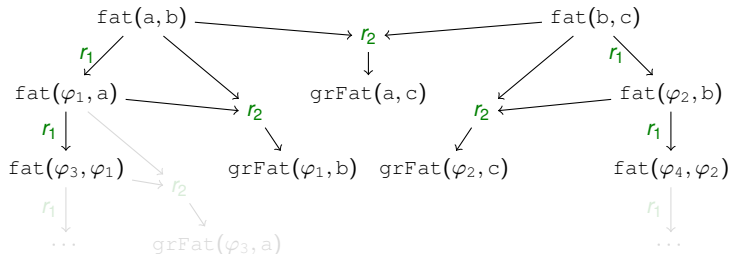
Not query-driven!

Also computes irrelevant atoms

CHASE procedure at work

 $D: \text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$ $r_1: \exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$ $r_2: \text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in
general! $\exists Y \text{grFat}(Y, a)?$

Yes!

Not query-driven!

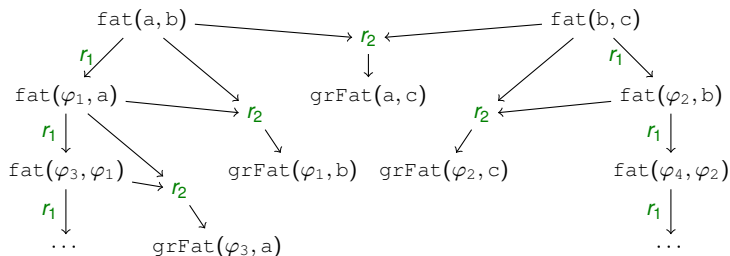
Also computes irrelevant atoms

CHASE procedure at work

D : $\text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$
 r_1 : $\exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$
 r_2 : $\text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in general!

 $\exists Y \text{grFat}(Y, a)?$

Yes!

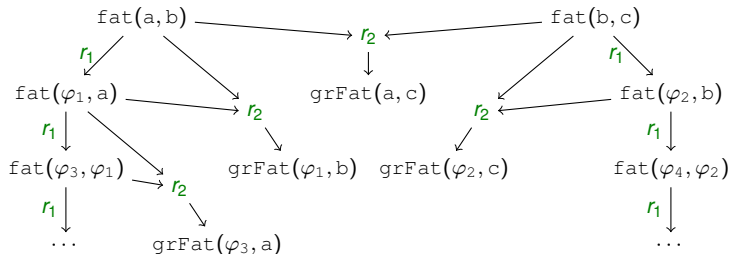
Not query-driven!

Also computes irrelevant atoms

CHASE procedure at work

 $D: \text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$ $r_1: \exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$ $r_2: \text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in
general! $\exists Y \text{grFat}(Y, a)?$

Yes!

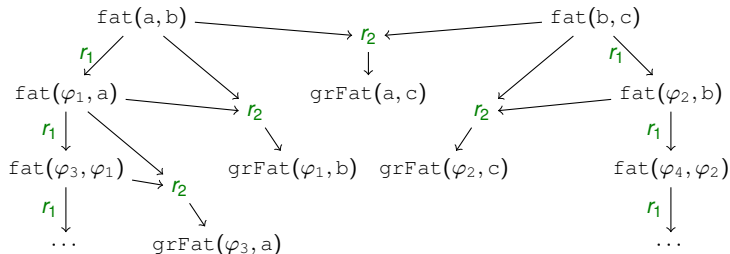
Not query-driven!

Also computes irrelevant atoms

CHASE procedure at work

 $D: \text{fat}(a,b) \leftarrow; \text{fat}(b,c) \leftarrow$ $r_1: \exists Y \text{fat}(Y,X) \leftarrow \text{fat}(X,Z)$ $r_2: \text{grFat}(X,Y) \leftarrow \text{fat}(X,Z), \text{fat}(Z,Y)$

Stop it?

Undecidable in
general! $\exists Y \text{grFat}(Y,a)?$

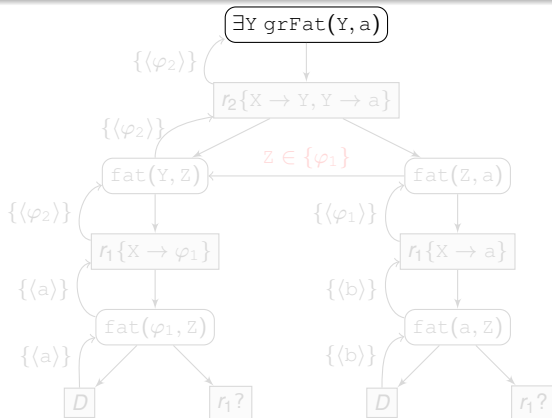
Yes!

Not query-driven!

Also computes irrelevant atoms

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

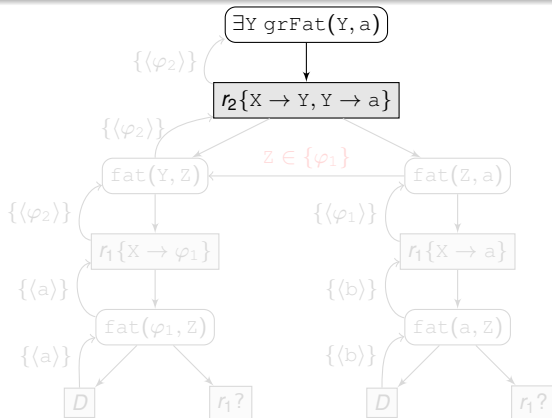
The query is true!



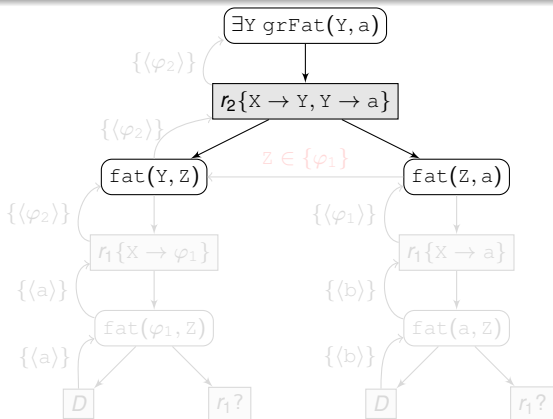
D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

The query is true!

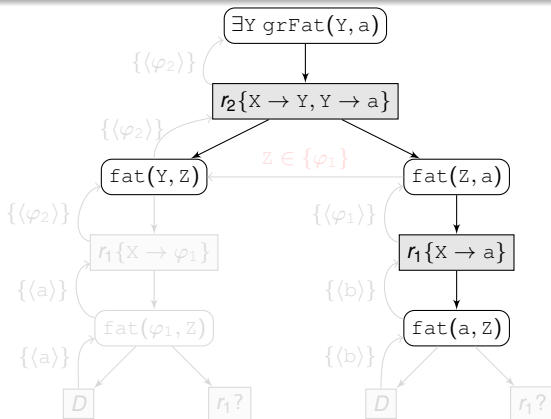


- D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$
- r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
- r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$ 

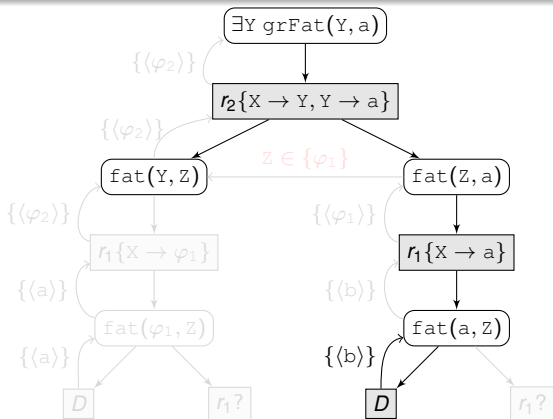
The query
is true!

D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$ 

The query
is true!

D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

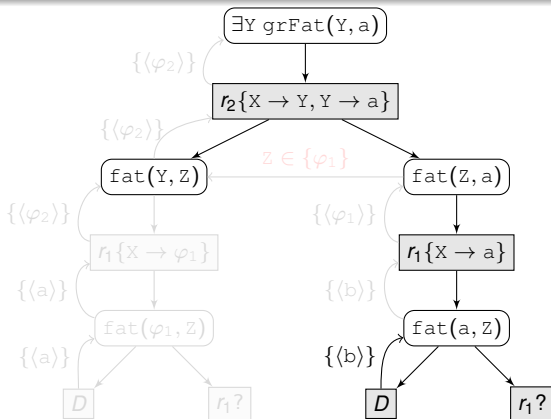
Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$ 

The query
is true!

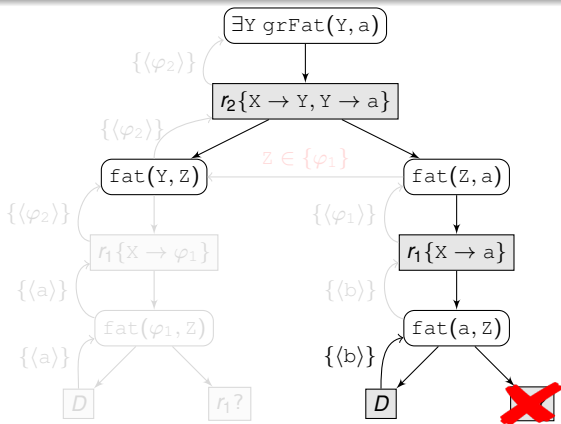
D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

The query is true!



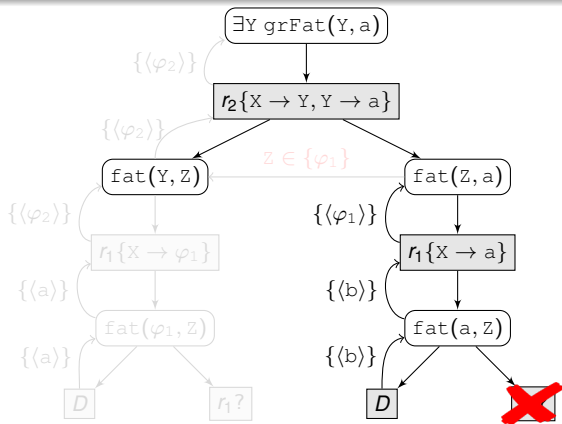
D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$ The query
is true!

D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

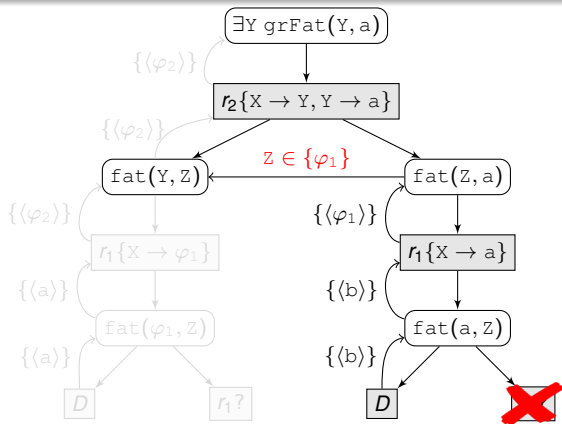
The query is true!



D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow; \text{fat}(\varphi_1, a) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

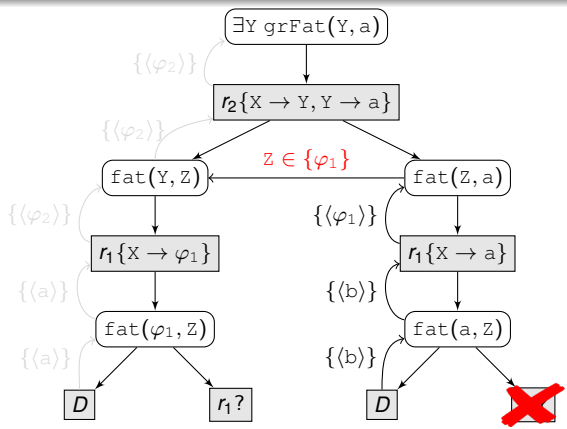
The query is true!



D : $\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow; \text{fat}(\varphi_1, a) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

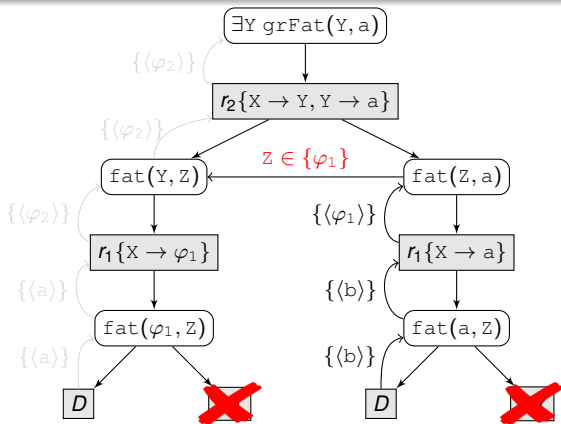
The query is true!



D : $\text{fat}(a, b) \leftarrow$; $\text{fat}(b, c) \leftarrow$; $\text{fat}(\varphi_1, a) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

The query is true!



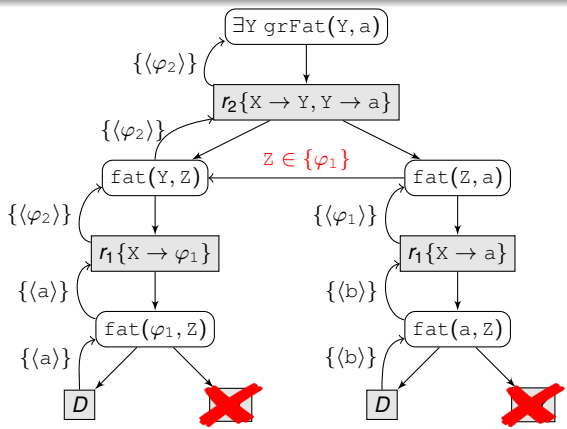
D : $\text{fat}(a, b) \leftarrow$; $\text{fat}(b, c) \leftarrow$; $\text{fat}(\varphi_1, a) \leftarrow$

r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

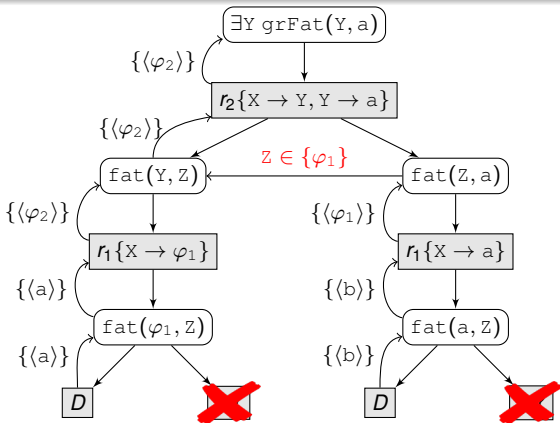
The query is true!



D : $\text{fat}(a, b) \leftarrow$; $\text{fat}(b, c) \leftarrow$; $\text{fat}(\varphi_1, a) \leftarrow$; $\text{fat}(\varphi_2, \varphi_1) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Top-down evaluation of $\exists Y \text{ grFat}(Y, a)$

The query is true!



D : $\text{fat}(a, b) \leftarrow ; \text{fat}(b, c) \leftarrow ; \text{fat}(\varphi_1, a) \leftarrow ; \text{fat}(\varphi_2, \varphi_1) \leftarrow$
 r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$
 r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

Outline

- 1 Datalog[∃]
 - Syntax, semantics and queries
 - CHASE procedure and top-down evaluation
- 2 Magic-Sets
 - Example
 - Application to Shy and experiment

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

Q : $\exists Y \text{ grFat}(Y, a)$?

r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

r_Q^* : $\text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

r_1^* : $\text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^*$: $\text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^*$: $\text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

⇒ $Q: \exists Y \text{ grFat}(Y, a)?$

$r_1: \exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

$r_2: \text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

$r_Q^*: \text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

$r_1^*: \text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^*: \text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^*: \text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Process the query:

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

⇒ Q : $\exists Y \text{ grFat}(Y, a)$?

r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

r_2 : $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

r_Q^* : $\text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

r_1^* : $\text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^*$: $\text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^*$: $\text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Process the query: produce r_Q^*

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

Q : $\exists Y \text{ grFat}(Y, a)$?

r_1 : $\exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

$\Rightarrow r_2$: $\text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$

r_Q^* : $\text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

r_1^* : $\text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^*$: $\text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^*$: $\text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Process $\text{mgc_grFat}^{\text{fb}}$ and r_2 :

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

$Q : \exists Y \text{ grFat}(Y, a)?$

$r_1 : \exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

$\Rightarrow r_2 : \text{grFat}(X, Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(X, Z), \text{fat}(Z, Y)$

$r_Q^* : \text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

$r_1^* : \text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^* : \text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^* : \text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Process $\text{mgc_grFat}^{\text{fb}}$ and r_2 : produce $r_{2,1}^*, r_{2,2}^*$ and modify r_2

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

$Q : \exists Y \text{ grFat}(Y, a)?$

$\Rightarrow r_1 : \exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

$r_2 : \text{grFat}(X, Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(X, Z), \text{fat}(Z, Y)$

$r_Q^* : \text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

$r_1^* : \text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^* : \text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^* : \text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Process $\text{mgc_fat}^{\text{fb}}$ and r_1 :

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

$Q: \exists Y \text{ grFat}(Y, a)?$

$\Rightarrow r_1: \exists Y \text{ fat}(Y, X) \leftarrow \text{mgc_fat}^{\text{fb}}(X), \text{fat}(X, Z)$

$r_2: \text{grFat}(X, Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(X, Z), \text{fat}(Z, Y)$

$r_Q^*: \text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

$r_1^*: \text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^*: \text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^*: \text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Process $\text{mgc_fat}^{\text{fb}}$ and r_1 : produce r_1^* and modify r_1

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

$Q: \exists Y \text{ grFat}(Y, a)?$

$\Rightarrow r_1: \exists Y \text{ fat}(Y, X) \leftarrow \text{mgc_fat}^{\text{fb}}(X), \text{fat}(X, Z)$

$r_2: \text{grFat}(X, Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(X, Z), \text{fat}(Z, Y)$

$r_Q^*: \text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

$r_1^*: \text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^*: \text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^*: \text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Process $\text{mgc_fat}^{\text{bf}}$ and r_1 ?

Magic-Sets: Example

- Aim at combining the benefit of the two approaches
- Only infer relevant atoms during the CHASE procedure

$Q: \exists Y \text{ grFat}(Y, a)?$

$\Rightarrow r_1: \exists Y \text{ fat}(Y, X) \leftarrow \text{mgc_fat}^{\text{fb}}(X), \text{fat}(X, Z)$

$r_2: \text{grFat}(X, Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(X, Z), \text{fat}(Z, Y)$

$r_Q^*: \text{mgc_grFat}^{\text{fb}}(a) \leftarrow$

$r_1^*: \text{mgc_fat}^{\text{bf}}(X) \leftarrow \text{mgc_fat}^{\text{fb}}(X)$

$r_{2,1}^*: \text{mgc_fat}^{\text{fb}}(Y) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y)$

$r_{2,2}^*: \text{mgc_fat}^{\text{fb}}(Z) \leftarrow \text{mgc_grFat}^{\text{fb}}(Y), \text{fat}(Z, Y)$

Process $\text{mgc_fat}^{\text{bf}}$ and r_1 ? **Do not pass binding to \exists -variables!**

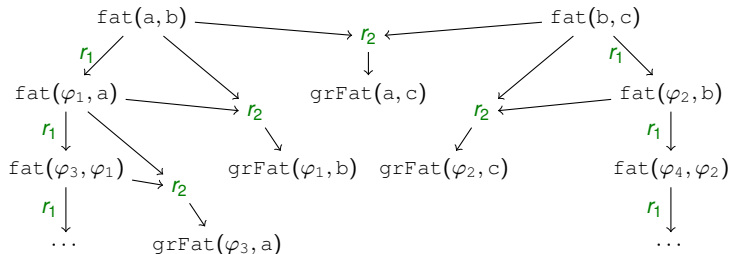
Magic-Sets: Impact on the CHASE procedure

$Q: \exists Y \text{ grFat}(Y, a)?$

$\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$

$r_1: \exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

$r_2: \text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$



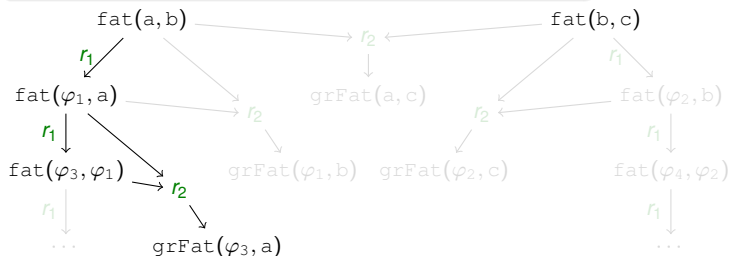
Magic-Sets: Impact on the CHASE procedure

$Q: \exists Y \text{ grFat}(Y, a)?$

$\text{fat}(a, b) \leftarrow; \text{fat}(b, c) \leftarrow$

$r_1: \exists Y \text{ fat}(Y, X) \leftarrow \text{fat}(X, Z)$

$r_2: \text{grFat}(X, Y) \leftarrow \text{fat}(X, Z), \text{fat}(Z, Y)$



Outline

- 1 Datalog[∃]
 - Syntax, semantics and queries
 - CHASE procedure and top-down evaluation
- 2 Magic-Sets
 - Example
 - Application to Shy and experiment

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, Y), s(Y), t(Y)$ $r_2 : \exists Y p(X, Y) \leftarrow$ $r_3 : s(Y) \leftarrow p(X, Y)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(Y) \leftarrow \text{mgc_q}^b(X), p(X, Y)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, Y) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(Y), p(X, Y)$

It's Shy!

It's not Shy!

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, Y), s(Y), t(Y)$ $r_2 : \exists Y p(X, \varphi) \leftarrow$ $r_3 : s(Y) \leftarrow p(X, Y)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(Y) \leftarrow \text{mgc_q}^b(X), p(X, Y)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, Y) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(Y), p(X, Y)$

It's Shy!

It's not Shy!

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, Y), s(Y), t(Y)$ $r_2 : \exists Y p(X, \varphi) \leftarrow$ $r_3 : s(\varphi) \leftarrow p(X, \varphi)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(Y) \leftarrow \text{mgc_q}^b(X), p(X, Y)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, Y) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(Y), p(X, Y)$

It's Shy!

It's not Shy!

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, \varphi), s(\varphi), t(Y)$ $r_2 : \exists Y p(X, \varphi) \leftarrow$ $r_3 : s(\varphi) \leftarrow p(X, \varphi)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(Y) \leftarrow \text{mgc_q}^b(X), p(X, Y)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, Y) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(Y), p(X, Y)$

It's Shy!

It's not Shy!

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, \varphi), s(\varphi), t(Y)$ $r_2 : \exists Y p(X, \varphi) \leftarrow$ $r_3 : s(\varphi) \leftarrow p(X, \varphi)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(Y) \leftarrow \text{mgc_q}^b(X), p(X, Y)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, Y) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(Y), p(X, Y)$

It's Shy!

It's not Shy!

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, \varphi), s(\varphi), t(Y)$ $r_2 : \exists Y p(X, \varphi) \leftarrow$ $r_3 : s(\varphi) \leftarrow p(X, \varphi)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(Y) \leftarrow \text{mgc_q}^b(X), p(X, Y)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, \varphi) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(Y), p(X, Y)$

It's Shy!

It's not Shy!

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, \varphi), s(\varphi), t(Y)$ $r_2 : \exists Y p(X, \varphi) \leftarrow$ $r_3 : s(\varphi) \leftarrow p(X, \varphi)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(\varphi) \leftarrow \text{mgc_q}^b(X), p(X, \varphi)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, \varphi) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(Y), p(X, \varphi)$

It's Shy!

It's not Shy!

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, \varphi), s(\varphi), t(Y)$ $r_2 : \exists Y p(X, \varphi) \leftarrow$ $r_3 : s(\varphi) \leftarrow p(X, \varphi)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(\varphi) \leftarrow \text{mgc_q}^b(X), p(X, \varphi)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, \varphi) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(\varphi), p(X, \varphi)$

It's Shy!

It's not Shy!

Preserving shyness

(1)

Informal definition

A program is Shy if nulls do not join in the CHASE.

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, \varphi), s(\varphi), t(Y)$ $r_2 : \exists Y p(X, \varphi) \leftarrow$ $r_3 : s(\varphi) \leftarrow p(X, \varphi)$ $r_Q^* : \text{mgc_q}^b(a) \leftarrow$ $r_{1,1}^* : \text{mgc_p}^{bf}(X) \leftarrow \text{mgc_q}^b(X)$ $r_{1,2}^* : \text{mgc_s}^b(\varphi) \leftarrow \text{mgc_q}^b(X), p(X, \varphi)$ $r'_1 : q(X) \leftarrow \text{mgc_q}^b(X), p(X, Y), s(Y), t(Y)$ $r'_2 : \exists Y p(X, \varphi) \leftarrow \text{mgc_p}^{bf}(X)$ $r'_3 : s(Y) \leftarrow \text{mgc_s}^b(\varphi), p(X, \varphi)$

It's Shy!

It's not Shy!

Preserving shyness

(2)

Observation

Shyness guarantees
absence of joins on nulls

 $Q : q(a)?$ $D : t(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, Y), s(Y), t(Y)$ $r_2 : \exists Y p(X, Y) \leftarrow$ $r_3 : s(Y) \leftarrow p(X, Y)$

Proposition

If the original program is shy, the modified program is shy as well and has the same models of the original program.

Theorem

Magic-Sets preserve shyness on the modified program.

Preserving shyness

(2)

Observation

Shyness guarantees
absence of joins on nulls

 $Q : q(a)?$ $D : t(a) \leftarrow ; \#dom(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, Y), \#dom(Y), s(Y), t(Y)$ $r_2 : \exists Y p(X, Y) \leftarrow$ $r_3 : s(Y) \leftarrow p(X, Y)$

Proposition

If the original program is shy, the modified program is shy as well and has the same models of the original program.

Theorem

Magic-Sets preserve shyness on the modified program.

Preserving shyness

(2)

Observation

Shyness guarantees
absence of joins on nulls

 $Q : q(a)?$ $D : t(a) \leftarrow ; \#dom(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, Y), \#dom(Y), s(Y), t(Y)$ $r_2 : \exists Y p(X, Y) \leftarrow$ $r_3 : s(Y) \leftarrow p(X, Y)$

Proposition

If the original program is shy, the modified program is shy as well and has the same models of the original program.

Theorem

Magic-Sets preserve shyness on the modified program.

Preserving shyness

(3)

 $Q : q(a)?$ $D : t(a) \leftarrow; \#dom(a) \leftarrow$ $r_1 : q(X) \leftarrow p(X, Y), \#dom(Y), s(Y), t(Y)$ $r_2 : \exists Y p(X, Y) \leftarrow$ $r_3 : s(Y) \leftarrow p(X, Y)$ $r_Q^* : mgc_q^b(a) \leftarrow$ $r_{1,1}^* : mgc_p^{bf}(X) \leftarrow mgc_q^b(X)$ $r_{1,2}^* : mgc_s^b(Y) \leftarrow mgc_q^b(X), p(X, Y)$ $r_1' : q(X) \leftarrow mgc_q^b(X), p(X, Y), \#dom(Y), s(Y), t(Y)$ $r_2' : \exists Y p(X, Y) \leftarrow mgc_p^{bf}(X)$ $r_3' : s(Y) \leftarrow mgc_s^b(Y), p(X, Y), \#dom(Y)$

Experiment

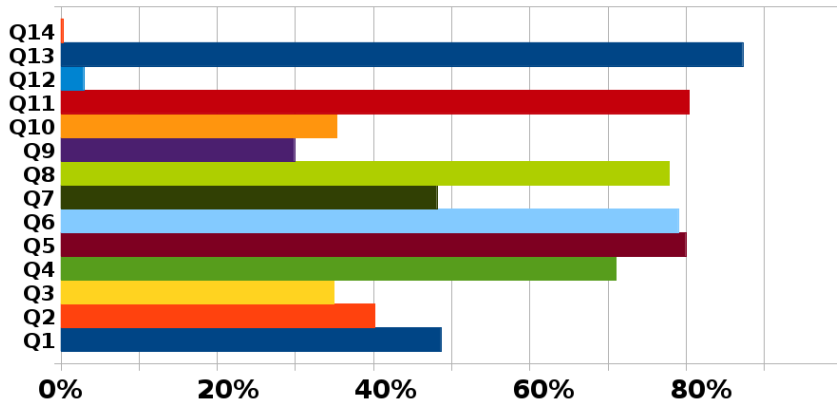
(1)

Dataset	System	# solved	Geom. Avg time
LUBM x10	DLV	14	2.87
	Pellet	14	84.48
	OWLIM-Lite	12	53.31
	OWLIM-SE	14	105.14
LUBM x30	DLV	14	9.70
	Pellet	0	–
	OWLIM-Lite	11	123.18
	OWLIM-SE	14	323.57
LUBM x50	DLV	14	16.67
	Pellet	0	–
	OWLIM-Lite	11	223.79
	OWLIM-SE	14	537.35

Experiment

(2)

Improvement on LUMBx50



- Novelty: handling of existential quantifiers
- Sound and complete for Datalog[∃] programs in general
- Preserving shyness is possible
- Implementation in DLV
 - Sensible performance improvements
 - Outperforms state-of-the-art ontology-based systems

Future work

- Explicit distinction between free and existential variables
- Application to other decidable classes
- Definition of new decidable classes, or linking existing fragments

- Novelty: handling of existential quantifiers
- Sound and complete for Datalog[∃] programs in general
- Preserving shyness is possible
- Implementation in DLV
 - Sensible performance improvements
 - Outperforms state-of-the-art ontology-based systems

Future work

- Explicit distinction between free and existential variables
- Application to other decidable classes
- Definition of new decidable classes, or linking existing fragments

Thank you!

Questions?