# The chase procedure and its applications to data exchange

Adrian Onet

Concordia University
Montréal, Québec, Canada

DEIS 2010

# Background

Query:

$$q_1(W, X, Y, Z) \leftarrow R(W, X, Y', Z'), R(W', X, Y, Z)$$

Constraint $\bowtie [AB, BCD]$

$$\Sigma = \{R(W, X, Y, Z), R(W', X, Y', Z') \rightarrow R(W, X, Y', Z')\}$$

Tableau representation

$q_1$

| | | | |
|---|---|---|---|
| w | x | y' | z' |
| w' | x | y | z |
| w | x | y | z |

$\Sigma$

| | | | |
|---|---|---|---|
| w | x | y | z |
| w' | x | y' | z' |
| w | x | y' | z' |

Applying the constrains on query $q_1$, we obtain:

| q2 | | | |
|---|---|---|---|
| w | x | y' | z' |
| w' | x | y | z |
| w | x | y | z |
| w' | x | y' | z' |
| w | x | y | z |

$\equiv$

| q2 | | | |
|---|---|---|---|
| w | x | y | z |
| w | x | y | z |

$$q_2(W, X, Y, Z) \leftarrow R(W, X, Y, Z)$$

$$q_1 \equiv_\Sigma q_2$$

# Background

- Query Equivalence
- Query Optimization
- Logical implication

Nowadays:

- Data Exchange
- Data Repairs
- Peer Data Exchange

# Basic Notions - Dependencies

Embedded dependencies covers most of the practical constraints needed.

$$\forall \bar{x}\ \varphi(\bar{x}) \rightarrow \exists \bar{y}\ \psi(\bar{x}, \bar{y})$$

$\varphi$, $\psi$ represents conjunctions of atoms

- *tgd* $= \psi$ doesn't contain equality atoms*
- *egd* $= \psi$ contains only equality atoms
- *full tgd* $= \bar{y}$ is the empty vector
- *LAV* $= \varphi$ contains exactly one predicate

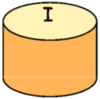
* - during this talk, if not mentioned otherwise, we consider only tgd's.

# Basic Notions - Instances

- $\mathbf{R} = \{R_1, R_2, \ldots, R_n\}$ - set of relational symbols
- Const - countable set of constants
- Null - countable set of labeled nulls
- $I$ instance over $\mathbf{R}$, $R_j^I \subset (\text{Const} \cup \text{Null})^{arity(R_j)}$
- $I$ ground instance over $\mathbf{R}$, $R_j^I \subset (\text{Const})^{arity(R_j)}$
- $h : dom(I) \to dom(J)$, such that $\forall c \in$ Const, $h(c) = c$ and $h(I) \subseteq J$ is called *homomorphism* from $I$ to $J$, denoted $I \to J$

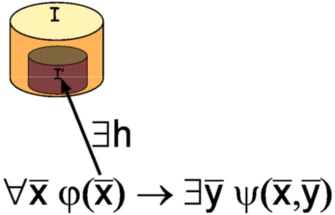
# Chasing tgd's

$$\sigma : \forall \bar{x} \; \varphi(\bar{x}) \rightarrow \exists \bar{y} \; \psi(\bar{x}, \bar{y})$$



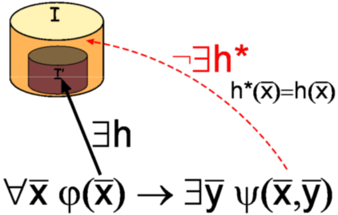

$$\forall \bar{\mathbf{x}} \; \varphi(\bar{\mathbf{x}}) \rightarrow \exists \bar{\mathbf{y}} \; \psi(\bar{\mathbf{x}}, \bar{\mathbf{y}})$$

$$\sigma : \forall \bar{x}\ \varphi(\bar{x}) \rightarrow \exists \bar{y}\ \psi(\bar{x}, \bar{y})$$



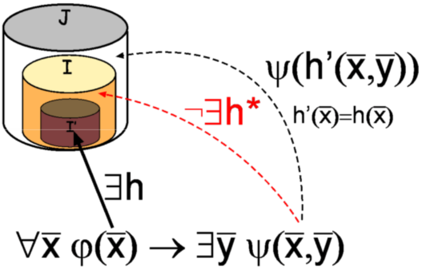
# Chasing tgd's

$$\sigma : \forall \bar{x}\ \varphi(\bar{x}) \rightarrow \exists \bar{y}\ \psi(\bar{x}, \bar{y})$$



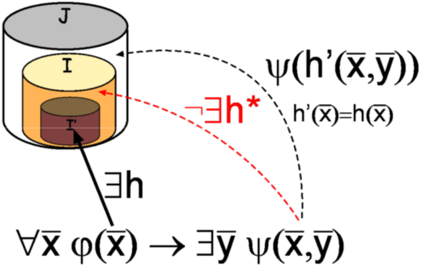
# Chasing tgd's

$$\sigma : \forall \bar{x}\ \varphi(\bar{x}) \rightarrow \exists \bar{y}\ \psi(\bar{x}, \bar{y})$$



# Chasing tgd's

$$\sigma : \forall \bar{x}\ \varphi(\bar{x}) \rightarrow \exists \bar{y}\ \psi(\bar{x}, \bar{y})$$




$$I \xrightarrow{\sigma, h} J$$

# Example Chasing tgd's

**Emp2**

| Name | Position |
|---|---|
| Ben | Analyst |
| John | Admin |

**Departments**

| DID | DName | MID |
|---|---|---|

$\sigma_2$ $\forall N,A,P,D\ \mathbf{Emp2(N,P)} \longrightarrow \exists E,I\ \mathbf{Employees(E,N,I)}$.

**Employees**

| EID | EName | DID |
|---|---|---|

**Emp1**

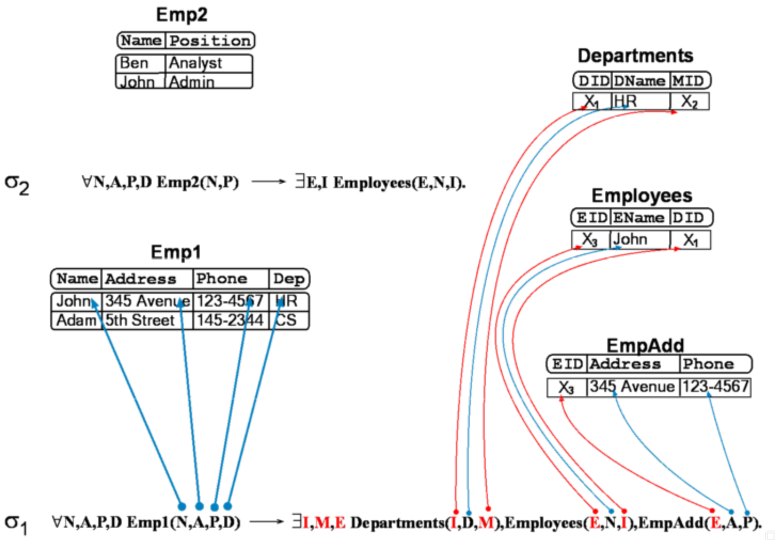
| Name | Address | Phone | Dep |
|---|---|---|---|
| John | 345 Avenue | 123-4567 | HR |
| Adam | 5th Street | 145-2344 | CS |

**EmpAdd**

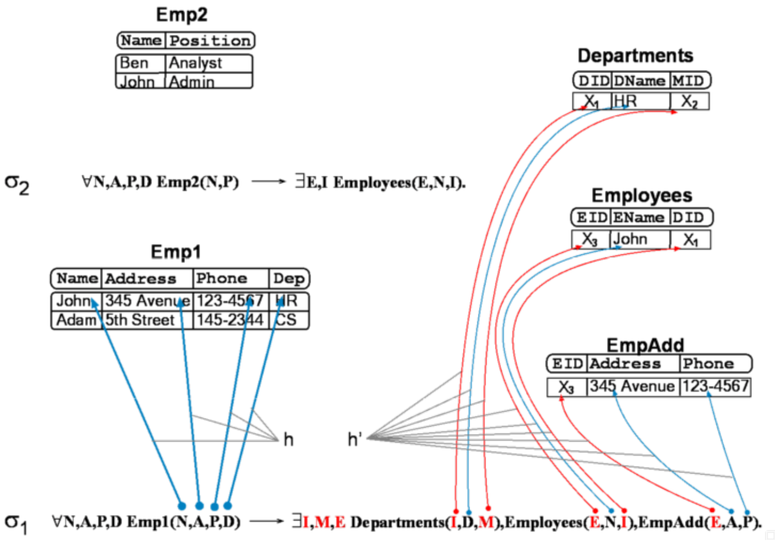
| EID | Address | Phone |
|---|---|---|

$\sigma_1$ $\forall N,A,P,D\ \mathbf{Emp1(N,A,P,D)} \longrightarrow \exists I,M,E\ \mathbf{Departments(I,D,M),Employees(E,N,I),EmpAdd(E,A,P)}$.

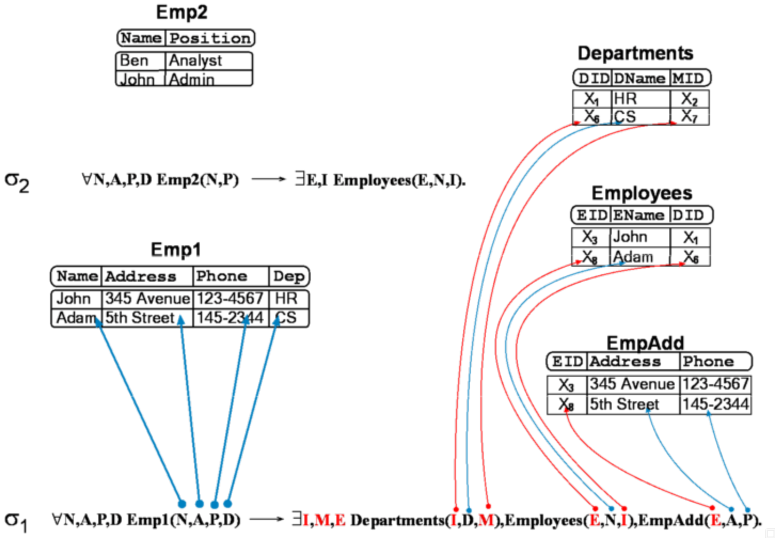
# Example Chasing tgd's



# Example Chasing tgd's



# Example Chasing tgd's



# Example Chasing tgd's

**Emp2**

| Name | Position |
|---|---|
| Ben | Analyst |
| John | Admin |

**Departments**

| DID | DName | MID |
|---|---|---|
| $X_1$ | HR | $X_2$ |
| $X_6$ | CS | $X_7$ |

$\sigma_2$ $\forall N,A,P,D\ \mathrm{Emp2}(N,P) \longrightarrow \exists E,I\ \mathrm{Employees}(E,N,I).$

**Employees**

| EID | EName | DID |
|---|---|---|
| $X_3$ | John | $X_1$ |
| $X_8$ | Adam | $X_6$ |
| $X_{11}$ | Ben | $X_{12}$ |

**Emp1**

| Name | Address | Phone | Dep |
|---|---|---|---|
| John | 345 Avenue | 123-4567 | HR |
| Adam | 5th Street | 145-2344 | CS |

**EmpAdd**

| EID | Address | Phone |
|---|---|---|
| $X_3$ | 345 Avenue | 123-4567 |
| $X_8$ | 5th Street | 145-2344 |

$\sigma_1$ $\forall N,A,P,D\ \mathrm{Emp1}(N,A,P,D) \longrightarrow \exists I,M,E\ \mathrm{Departments}(I,D,M),\mathrm{Employees}(E,N,I),\mathrm{EmpAdd}(E,A,P).$

# Example Chasing tgd's

**Emp2**

| Name | Position |
|---|---|
| Ben | Analyst |
| John | Admin |

**Departments**

| DID | DName | MID |
|---|---|---|
| $X_1$ | HR | $X_2$ |
| $X_6$ | CS | $X_7$ |

$\sigma_2$ $\quad \forall N,A,P,D\ \mathbf{Emp2}(N,P) \longrightarrow \exists E,I\ \mathbf{Employees}(E,N,I).$

**Employees**

| EID | EName | DID |
|---|---|---|
| $X_3$ | John | $X_1$ |
| $X_8$ | Adam | $X_6$ |
| $X_{11}$ | Ben | $X_{12}$ |

**Emp1**

| Name | Address | Phone | Dep |
|---|---|---|---|
| John | 345 Avenue | 123-4567 | HR |
| Adam | 5th Street | 145-2344 | CS |

**EmpAdd**

| EID | Address | Phone |
|---|---|---|
| $X_3$ | 345 Avenue | 123-4567 |
| $X_8$ | 5th Street | 145-2344 |

$\sigma_1$ $\quad \forall N,A,P,D\ \mathbf{Emp1}(N,A,P,D) \longrightarrow \exists I,M,E\ \mathbf{Departments}(I,D,M),\mathbf{Employees}(E,N,I),\mathbf{EmpAdd}(E,A,P).$

# Example Chasing tgd's

**Emp2**

| Name | Position |
|---|---|
| Ben | Analyst |
| John | Admin |

**Departments**

| DID | DName | MID |
|---|---|---|
| $X_1$ | HR | $X_2$ |
| $X_6$ | CS | $X_7$ |

$\sigma_2$ $\forall N,A,P,D\ \text{Emp2}(N,P) \longrightarrow \exists E,I\ \text{Employees}(E,N,I).$

**Employees**

| EID | EName | DID |
|---|---|---|
| $X_3$ | John | $X_1$ |
| $X_8$ | Adam | $X_6$ |
| $X_{11}$ | Ben | $X_{12}$ |

**Emp1**

| Name | Address | Phone | Dep |
|---|---|---|---|
| John | 345 Avenue | 123-4567 | HR |
| Adam | 5th Street | 145-2344 | CS |

**EmpAdd**

| EID | Address | Phone |
|---|---|---|
| $X_3$ | 345 Avenue | 123-4567 |
| $X_8$ | 5th Street | 145-2344 |

$\sigma_1$ $\forall N,A,P,D\ \text{Emp1}(N,A,P,D) \longrightarrow \exists I,M,E\ \text{Departments}(I,D,M),\text{Employees}(E,N,I),\text{EmpAdd}(E,A,P).$

# Chase Algorithm

$\textsc{Chase}(I, \Sigma)$

$I_0 := I$

$i := 0$

**repeat**

$\quad I_i \xrightarrow{\sigma,h} I_{i+1}$

$\quad i := i + 1$

**until** $I_{i-1} \neq I_i$

**return** $I_i$

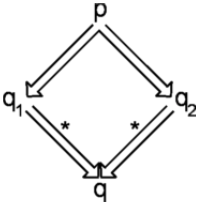
# Replacement System

- A pair $(A, \Rightarrow)$, is a replacement system if $A$ is a set of objects and $\Rightarrow$ is an antireflexive binary relation over $A$ called the transformation relation.
- by $\Rightarrow^*$ is denoted the reflexive transitive closure of "$\Rightarrow$".
- an element $p \in A$ is called irreducible if $p \Rightarrow^* q$ implies $p = q$.
- $(A, \Rightarrow)$ is finite if for all $p \in A$ there exists $n$ such that $p \Rightarrow^* q$ in at most $n$ steps and $q$ irreducible.
- $(A, \Rightarrow)$ is finite Church-Rosser if for all $p \in A$ if $p \Rightarrow^* q_1$ and $p \Rightarrow^* q_2$ and $q_1, q_2$ are irreducible, then $q_1 = q_2$.

# Church-Rosser Property (cont.)

### Theorem (Sethi)

*$(A, \Rightarrow)$ is finite Church-Rosser iff $(A, \Rightarrow)$ is finite and for any $p \in A$ if $p \Rightarrow q_1$ and $p \Rightarrow q_2$, then there exists $q \in A$ such that $q_1 \Rightarrow^* q$ and $q_2 \Rightarrow^* q$.*

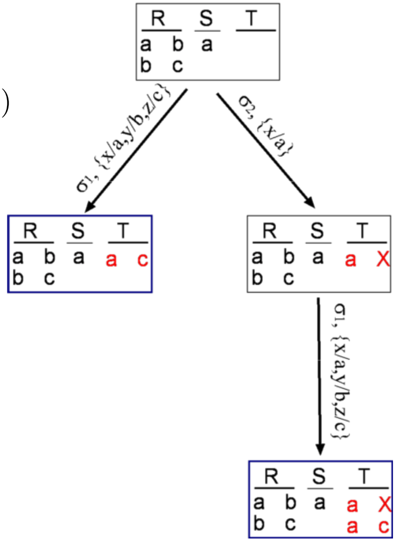



- let $\mathcal{I}$ be the set of all instances over schema $\mathbf{R}$ and $\Sigma$ a set of tgd's, then $(\mathcal{I}, \rightarrow_\Sigma)$ is a replacement system.

# Chase Properties: multiple results

$$\sigma_1 \quad : \quad R(x,y), R(y,z) \rightarrow T(x,z)$$

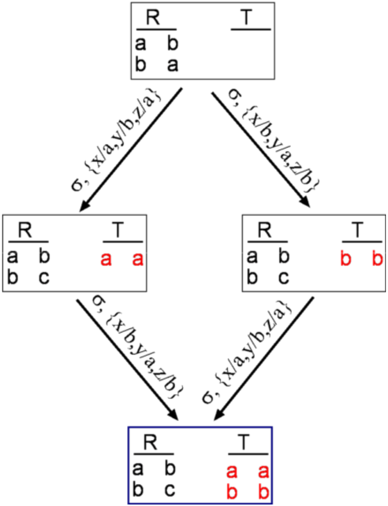
$$\sigma_2 \quad : \quad S(x) \rightarrow \exists Z \; T(x,Z)$$




- $Chase_\Sigma(I)$ denotes the set of all irreducible instances.

# Chase Properties: Church-Rosser for full tgd's

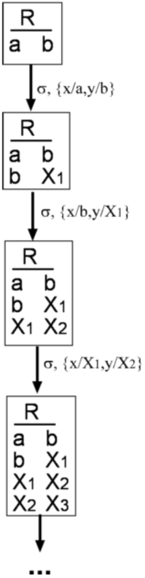
$$\sigma \quad : \quad R(x,y), R(y,z) \rightarrow T(x,z)$$




- If $\Sigma$ is a set of full tgd's, then the replacement system $(\mathcal{I}, \rightarrow_\Sigma)$ has the finite Church-Rosser property.

# Chase Properties: nonterminating chase

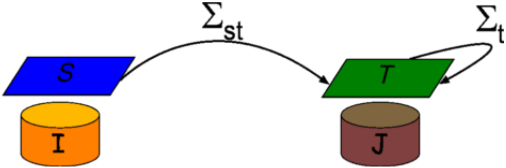
$$\sigma \quad : \quad R(x, y) \rightarrow \exists Z \; R(y, Z)$$



# Chase properties: Summary

- $I \xrightarrow{\sigma,h} J \Rightarrow I \subseteq J$.
- there may exist $J, J' \in Chase_{\Sigma}(I)$ such that $J \neq J'$.
- the Chase algorithm may not terminate.
- there exist a $\Sigma$ and instance $I$ such that it has both a terminating and a non terminating chase sequence.
- $\Sigma$ set of full tgd's $\Rightarrow$ $(\mathcal{I}, \rightarrow_{\Sigma})$ is finite.
- $\Sigma$ set of full tgd's $\Rightarrow$ the chase has the *Church-Rosser* property.

# Data Exchange, the problem




The data exchange setting $(\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$

- $\Sigma_{st}$ - specifies the relationship between $\mathbf{S}$ and $\mathbf{T}$
- $\Sigma_t$ - specifies the constraints that must be satisfied by $\mathbf{T}$

Instance $J$ is a solution for $(\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$ iff:

- $I \cup J \models \Sigma_{st} \cup \Sigma_t$
- $Sol(I)$ is the set of all solution for $I$

# Universal Solutions

Let $I$ be an instance and $(\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$ a data exchange settings.

$J$ is a universal solution (Fagin et al. ICDT03) for $I$ iff

- $J \in Sol(I)$
- $\forall J' \in Sol(I) \Rightarrow J \rightarrow J'$

**Theorem (Fagin et al. ICDT03)**

*If $J$ a finite instance from $Chase_{\Sigma_{st} \cup \Sigma_t}(I)$, then $J$ is a universal solution for $I$.*

# Certain Answers

If $Q$ is a query over $\mathbf{T}$ the certain answer on $\langle(\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t), I\rangle$ is defined as:

$$certain(Q, I) = \cap_{J \in Sol(I)} Q(J)$$

It turns out that universal solutions represents a good choice to get certain answers in data exchange:

## Theorem (Fagin et al. ICDT03)

*If $J$ is a universal solution for $Sol(I)$ and $Q \in UCQ$ then*

$$certain(Q, I) = Q(J)\downarrow^*$$

* - by $J\downarrow$ we mean the maximum subset of tuples from $J$ that contains only constants.

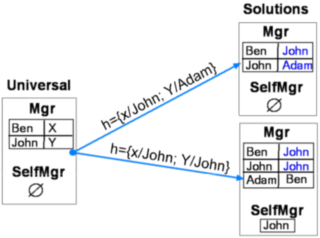
# Universal Solution in Data Exchange

Consider source instance:

| Emp |
|---|
| Ben |
| John |

And dependencies:

$$\Sigma_{st} : \quad Emp(x) \rightarrow \exists Y \; Mgr(x, Y).$$
$$\Sigma_{t} : \quad Mgr(x, x) \rightarrow SelfMgr(x).$$



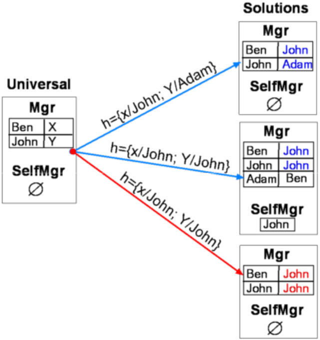
# Universal Solution in Data Exchange

Consider source instance:

| Emp |
|---|
| Ben |
| John |

And dependencies:

$$\Sigma_{st}: \quad Emp(x) \rightarrow \exists Y\ Mgr(x, Y).$$
$$\Sigma_t: \quad Mgr(x, x) \rightarrow SelfMgr(x).$$



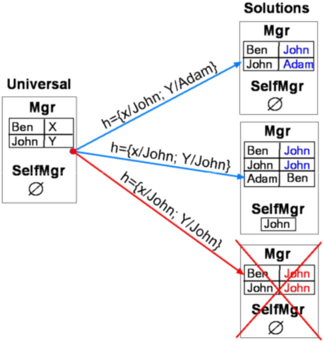
# Universal Solution in Data Exchange

Consider source instance:

| Emp |
|---|
| Ben |
| John |

And dependencies:

$$\Sigma_{st} : \quad Emp(x) \rightarrow \exists Y \; Mgr(x, Y).$$
$$\Sigma_t : \quad Mgr(x, x) \rightarrow SelfMgr(x).$$



# The Core

Consider source instance:

| R | |
|---|---|
| a | b |
| b | c |

| S |
|---|
| a |

And dependencies:

$$\Sigma_{st}: \quad R(x,y), R(y,z) \rightarrow T(x,z)$$
$$S(x) \rightarrow \exists Y\ T(x,Y)$$
$$\Sigma_t: \quad \emptyset$$

Universal Solution 1

| T | |
|---|---|
| a | X |
| a | c |

Universal Solution 2

| T | |
|---|---|
| a | c |

The core is the smallest universal solution.
The core is unique up to isomorphism.

# No Universal Solution

Consider the instance:

| R | |
|---|---|
| a | b |

And dependencies:

$$\Sigma_{st}: \quad R(x,y) \rightarrow S(x,y).$$
$$\Sigma_t: \quad S(x,y) \rightarrow \exists Z\ S(x,Z).$$

This gives the following infinite chase sequence:

| S | |
|---|---|
| a | b |
| b | $Z_1$ |
| $Z_1$ | $Z_2$ |
| ... | |

Still there exists solutions:

| S | |
|---|---|
| a | b |
| b | a |

| S | |
|---|---|
| a | b |
| b | b |

# Chase termination

## Theorem (Deutsch et al. 2008)

*Consider an instance $I$ and a set $\Sigma$ of tgd's:*

- *it is undecidable whether some chase sequences of $I$ with $\Sigma$ terminates;*
- *it is undecidable whether all chase sequences of $I$ with $\Sigma$ terminates.*

## Theorem (Kolaitis et al. 2006)

*There exists a data exchange setting $(\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$, with the following properties:*

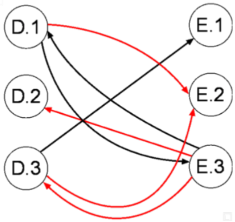
- *$\Sigma_{st}$ consists of one full tgd;*
- *$\Sigma_t$ consists of one egd, one full tgd and one tgd;*
- *the existence of solution is undecidable for this setting.*

# Chase termination: Weakly Acyclic Dependencies

Dependency Graph:

$\Sigma_t$:

$$Departments(did, dname, mgrid) \rightarrow \exists N\ Employees(mgrid, N, did).$$
$$Employees(eid, ename, did) \rightarrow \exists D, M\ Departments(did, D, M).$$




$\Sigma_t$ is weakly acyclic iff there is no cycle trough an existential edge.

- if a set of tgd's is weakly acyclic all chase sequences terminate.

# Chase Termination: Safe Conditions (Meier et al. 2009)

Let $\Sigma$ be a set of tgd's. The set aff($\Sigma$) defined as: $(R, i) \in$ aff($\Sigma$) iff

- $(R, i)$ contains an existential or
- $(R, i)$ is any position in the head of a dependency with a universal $x$ that appears only in aff($\Sigma$).

The propagation graph for $\Sigma$ is a directed graph (aff($\Sigma$)$, E)$, with $E$ as in the dependency graph with both regular and special edges.

$\Sigma$ is said to be safe if (aff($\Sigma$)$, E)$ doesn't contain any cycles going trough special edges.

# Chase Termination: Stratification

$\sigma_1, \sigma_2 \in \Sigma$; $\sigma_1 \prec \sigma_2$ iff

- $\exists\, I$ instance, and
- $I \models \sigma_2$, and
- $I \xrightarrow{\sigma_1, h} J$, and
- $J \not\models \sigma_2$.

Example:

$\sigma_1 : \; R(x, y) \rightarrow S(x)$

$\sigma_2 : \; S(x) \rightarrow R(x, x)$

Instance $I$:

| R | |
|---|---|
| a | b |

## Definition

- The chase graph for $\Sigma$ is a directed graph $G(\Sigma) = (\Sigma, E)$, where $(\sigma_1, \sigma_2) \in E$ iff $\sigma_1 \prec \sigma_2$.
- $\Sigma$ is stratified iff all cycles of $G(\Sigma)$ are weakly acyclic.

# Chase Termination: Stratification

$\sigma_1, \sigma_2 \in \Sigma$; $\sigma_1 \prec \sigma_2$ iff

- $\exists\ I$ instance, and
- $I \models \sigma_2$, and
- $I \xrightarrow{\sigma_1, h} J$, and
- $J \not\models \sigma_2$.

Example:

$\sigma_1 :\ R(x, y) \rightarrow S(x)$
$\sigma_2 :\ S(x) \rightarrow R(x, x)$

Instance $I$:

| R | |
|---|---|
| a | b |

$\sigma_1 \prec \sigma_2$

## Definition

- The chase graph for $\Sigma$ is a directed graph $G(\Sigma) = (\Sigma, E)$, where $(\sigma_1, \sigma_2) \in E$ iff $\sigma_1 \prec \sigma_2$.
- $\Sigma$ is stratified iff all cycles of $G(\Sigma)$ are weakly acyclic.

# Chase Termination: Stratification

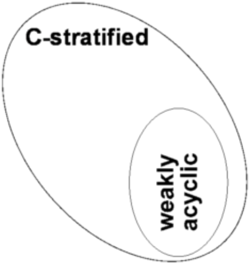
Theorem (Deutsch et al. 08)

*For every stratified set of tgd's and for all instances $I$ there exists a terminating chase sequence.*

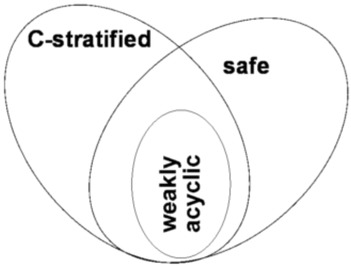
- the decision problem "is $\Sigma$ stratified?" is in coNP.
- the lower bound is open.

# Chase termination: hierarchy

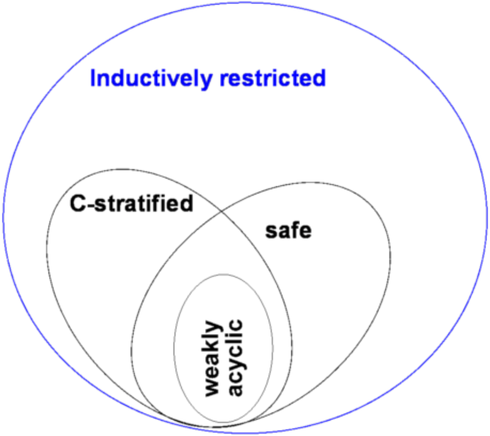
# Chase termination: hierarchy



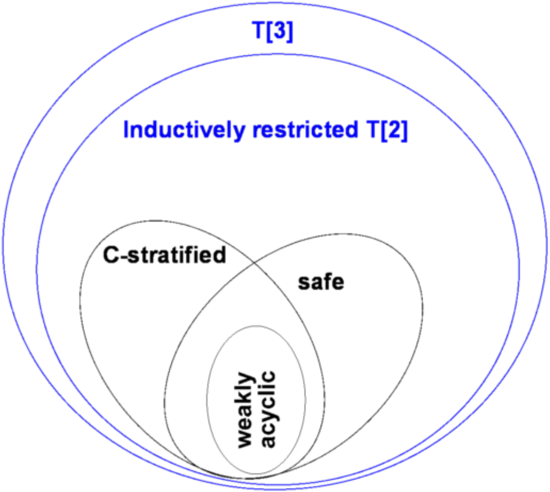
# Chase termination: hierarchy



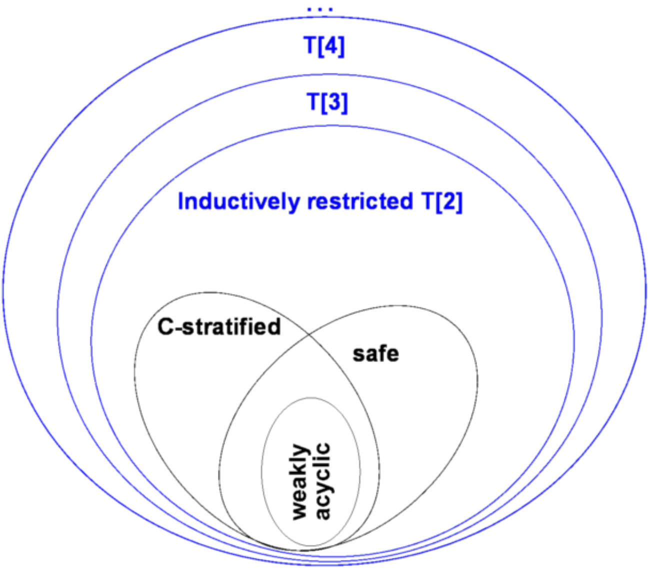
# Chase termination: hierarchy



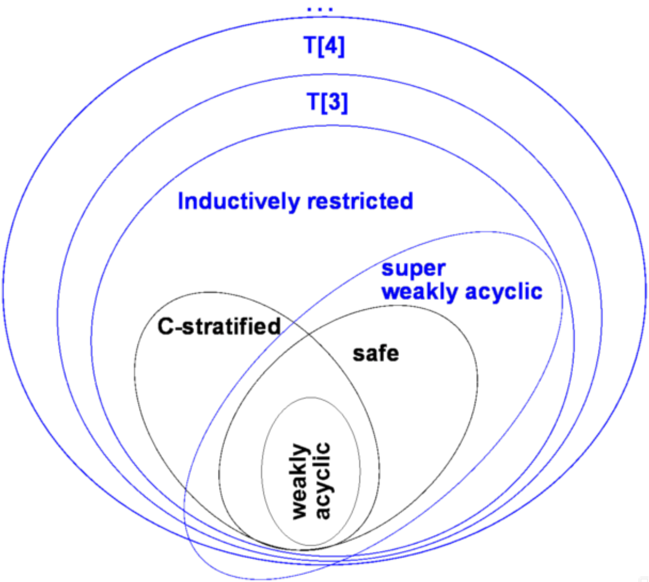
# Chase termination: hierarchy



# Chase termination: hierarchy



# Chase termination: hierarchy



# Chase termination: examples

- Stratified but not weakly acyclic:

$$\sigma : \; E(x,y), E(y,x) \rightarrow \exists Z, W \; E(x,Z), E(Z,W), E(W,x)$$

- Safe but not stratified:

$$\begin{aligned} \sigma_1 &: \; S(y,z), R(x,y,z) \rightarrow \exists W \; R(y,W,x) \\ \sigma_2 &: \; R(x,y,z) \rightarrow S(x,z) \end{aligned}$$

- Super-weak acyclic but not safe:

$$\begin{aligned} \sigma_1 &: \; N(x) \rightarrow \exists Y, Z \; E(x,Y,Z) \\ \sigma_2 &: \; E(x,y,y) \rightarrow N(y) \end{aligned}$$

# Chase termination: rewriting

Can we do better? YES

- Let $\mathbf{T}$ be one of the classes weakly-acyclic, stratified, C-stratified, safe condition or super weakly acyclic tgd's.
- Greco and Spezzano (VLDB 2010) introduced a new rewriting mapping $Adn$ such that for all $\Sigma$ set of tgd's over schema $\mathbf{R}$:
  - $\Sigma \equiv_{\mathbf{R}/\{\hat{\mathbf{R}}\}} Adn(\Sigma)$
  - let $Adn\mathbf{T}$ the set of tgd's such that $Adn(\Sigma)$ is in class $\mathbf{T}$.
  - $\mathbf{T} \subset Adn\mathbf{T}$.

# Chase termination: rewriting

$\Sigma_1$ :

$$\begin{aligned} \sigma_1 &: N(x) \to \exists y\ E(x,y) \\ \sigma_2 &: S(x), E(x,y) \to N(y) \end{aligned}$$

$\Sigma_2$ :

$$\begin{aligned} \sigma'_1 &: N(x) \to N^b(x) \\ \sigma'_2 &: S(x) \to S^b(x) \\ \sigma'_3 &: E(x,y) \to E^{bb}(x,y) \\ \sigma'_4 &: N^b(x) \to \exists y\ E^{bf}(x,y) \\ \sigma'_5 &: S^b(x), E^{bb}(x,y) \to N^b(y) \\ \sigma'_6 &: S^b(x), E^{bf}(x,y) \to N^f(y) \\ \sigma'_7 &: N^f(x) \to \exists y\ E^{ff}(x,y) \end{aligned}$$

$$\begin{aligned} \sigma'_8 &: N^b(x) \to \hat{N}(x) \\ \sigma'_9 &: N^f(x) \to \hat{N}(x) \\ \sigma'_{10} &: S^b(x) \to \hat{S}(x) \\ \sigma'_{11} &: E^{bb}(x,y) \to \hat{E}(x,y) \\ \sigma'_{12} &: E^{bf}(x,y) \to \hat{E}(x,y) \\ \sigma'_{13} &: E^{ff}(x,y) \to \hat{E}(x,y) \end{aligned}$$

# Chase flavors: Core Chase

CORE-CHASE$(I, \Sigma)$

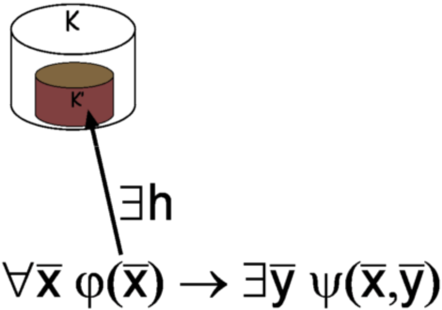
```
1  I_0 := I
2  i := 0
3  J = ⋃_{I_i --σ,h--> D} D
4  I_{i+1} = Core(J)
5  if I_i = I_{i+1}
6      then return I_i
7      else  i = i + 1; goto 3
```

## Theorem (Deutsch et al. 08)

- Core-Chase$(I, \Sigma)$ *computes the core of the universal solution;*
- *if there exists a sequence such that* Chase$(I, \Sigma)$ *terminates, then* Core-Chase$(I, \Sigma)$ *terminates;*
- *if for* $(I, \Sigma)$ *there exists a universal solution, then* Core-Chase$(I, \Sigma)$ *terminates;*

# Chase flavors: Solution-aware chase

Let $\Sigma$ a set of tgd's, $K' \subseteq K$, $K \models \Sigma$

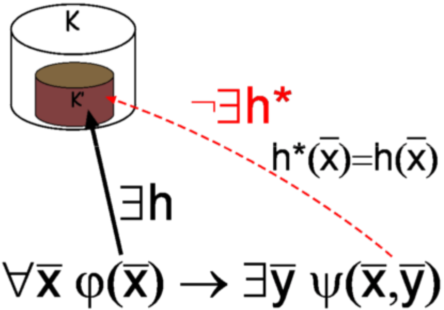



Theorem (Fuxman et al. 2006)

*The length of every solution-aware chase sequence of $K'$ with $\Sigma$ and $K$ is bounded by $p(|K'|)$.*

# Chase flavors: Solution-aware chase

Let $\Sigma$ a set of tgd's, $K' \subseteq K$, $K \models \Sigma$

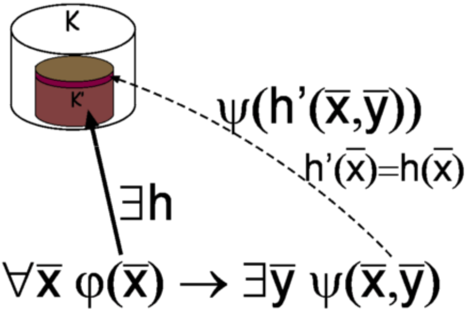



**Theorem (Fuxman et al. 2006)**

*The length of every solution-aware chase sequence of $K'$ with $\Sigma$ and $K$ is bounded by $p(|K'|)$.*

# Chase flavors: Solution-aware chase

Let $\Sigma$ a set of tgd's, $K' \subseteq K$, $K \models \Sigma$




Theorem (Fuxman et al. 2006)

*The length of every solution-aware chase sequence of $K'$ with $\Sigma$ and $K$ is bounded by $p(|K'|)$.*

# Chase flavors: Extended core-chase

Consider a disjunctive dependency:

$$\sigma : \forall \bar{x}\ \varphi(\bar{x}) \rightarrow \exists \bar{y} \bigvee_{1 \leq i \leq n} \psi_i(\bar{x}, \bar{y})$$

Extended Chase Step: $I \xrightarrow{\sigma,h} \{J_1, J_2, \ldots, J_p\}$

1. $\varphi(h(\bar{x})) \subseteq I$
2. $\neg\exists h', \neg\exists i$ such that $h'$ extends $h$ and $\psi_i(h'(\bar{x}, \bar{y})) \subseteq I$
3. $\forall i\ (1 \leq i \leq n)\ I \xrightarrow{\sigma_i,h} J_i,$ where $\sigma_i : \forall \bar{x}\ \varphi(\bar{x}) \rightarrow \exists \bar{y}\ \psi_i(\bar{x}, \bar{y})$

# Chase flavors: Extended core-chase

EXTENDED-CORE-CHASE($I, \Sigma$ set of DED's)

$L_0 := \{I\}$
$i := 0$
**for** $DED\ \sigma \in \Sigma$, h-applicable
  **do**
    $\forall I_j \in L_i$ run in parallel
    $I_j \xrightarrow{\sigma, h} K'_j$
    **for** each j
      **do**
        $K_j = \{\}$
        **for** $J \in K'_j$
          **do**
            $K_j = K_j \cup core(J)$

$L_{i+1} = K_j$
remove from $L_{i+1}$ all $M$ such that $\exists N \in L_{i+1}\ N \rightarrow M$
i:= i+1;
**if** $L_i = L_{i-1}$
  **then goto** 3
**return** $L_i$

# Chase and Date Exchange, beyond universal solutions

Data exchange settings $(\{S\}, \{R, T\}, \Sigma_{st}, \Sigma_t)$:

Source instance ($I$):

| S |
| --- |
| a |

dependencies:

$$\begin{aligned} \Sigma_{st} : &\quad S(x) \rightarrow \exists Y\ R(x, Y) \\ \Sigma_t : &\quad R(x, x) \rightarrow T(x) \end{aligned}$$

queries:

$$\begin{aligned} q_1(x) &\leftarrow \exists y\ R(x, y) \\ q_2(x) &\leftarrow \exists y\ (R(x, y) \wedge x \neq y) \vee T(x) \end{aligned}$$

- the universal model $U = \{S(a), R(a, X)\}$
- $cert_{q_1}(I) = dom(I) \cap q_1(U) = \{(a)\}$

# Chase and Date Exchange, beyond universal solutions

Data exchange settings $(\{S\}, \{R, T\}, \Sigma_{st}, \Sigma_t)$:

Source instance ($I$):

| S |
|---|
| a |

dependencies:

$$\Sigma_{st}: \quad S(x) \rightarrow \exists Y \; R(x, Y)$$
$$\Sigma_t: \quad R(x, x) \rightarrow T(x)$$

queries:

$$q_1(x) \leftarrow \exists y \; R(x, y)$$
$$q_2(x) \leftarrow \exists y \; (R(x, y) \wedge x \neq y) \vee T(x)$$

- the universal model $U = \{S(a), R(a, X)\}$
- $cert_{q_2}(I) = dom(I) \cap q_2(U) = \{\emptyset\}$

# Chase and Date Exchange, beyond universal solutions

Data exchange settings $(\{S\}, \{R\}, \Sigma_{st}, \Sigma_t)$:

Source instance (I):

| S |
|---|
| a |

dependencies:

$$\Sigma_{st}: \quad S(x) \rightarrow \exists Y\ R(x, Y)$$
$$\Sigma_t: \quad R(x, x) \rightarrow T(x)$$

query:

$$q_2(x) \quad \leftarrow \quad \exists y\ (R(x, y) \wedge x \neq y) \vee T(x)$$

- $\hat{\Sigma} = \Sigma_{st} \cup \Sigma_t \cup \{x = y \vee N(x, y); x = y, N(x, y) \rightarrow \perp\}$
- model set for $I$ and $\hat{\Sigma}$ is $U = \{\{S(a), R(a, X), N(a, X)\}; \{S(a), R(a, a), T(a)\}\}$ ;
- $cert_{q_2}(I) = dom(I) \cap \bigcap_{J \in U} q_2(J) = \{(a)\}$

# Homomorphisms (cont.)

- $h: dom(I) \to dom(J)$, such that $\forall c \in$ Const $h(c) = c$ and $h(I) \subseteq J$ is called *homomorphism* from $I$ to $J$, denoted $I \to J$. **(hom)**
- If $h$ is an injection then it is called *injective homomorphism*. **(ihom)**
- If $h(I) = J$ then $h$ is called *epimorphism* or *full homomorphism*. **(fhom)**
- If $h(I) = J$ and $h$ is also injective then $h$ is called *embedding*. **(emb)**

# Chase and Date Exchange beyond universal solutions

$F \in \{\mathbf{hom}, \mathbf{ihom}, \mathbf{fhom}, \mathbf{emb}\}$

## Definition (Deutsch et al. 2008)

A set $U$ of finite instances is an $F$-universal model set for s set of instances $K$ if it satisfies the following conditions:

1. $(\forall M \in K)(\exists T \in U) T \rightarrow_F M$;
2. $U \subseteq K$;
3. $U$ is finite;
4. $\neg \exists U' \subset U$ such that $U' \rightarrow_F U$.

## Theorem (Deutsch et al. 2008)

*Let $(\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$ be a data exchange setting with $\Sigma = \Sigma_{st} \cup \Sigma_t$ a set of NDED's. Let $U$ be a $F$-universal model set for $Sol_\Sigma(I)$ and $Q$ a query of arity $r$ over $\mathbf{T}$. If*

1. $F = \mathbf{hom}$ *and* $Q \in UCQ \cup Datalog$, *or*
2. $F = \mathbf{ihom}$ *and* $Q \in MonQ$, *or*
3. $F = \mathbf{fhom}$ *and* $Q \in UCQ^{\neg}$, *or*
4. $F = \mathbf{emb}$ *and* $Q \in UCQ^{\neg, \neq}$.

$$\text{then } cert_Q^\Sigma(I) = dom(I)^r \cap \bigcap_{J \in U} Q(J)$$

# Computing F-Universal model sets

Let $(\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$ be a data exchange setting with $\Sigma = \Sigma_{st} \cup \Sigma_t$ a set of NDED's.

- extend $\hat{\mathbf{S}} = \mathbf{S} \cup \{\hat{R} : \ R \in \mathbf{S}\} \cup \{N\}$;
- change $\Sigma$ to $\hat{\Sigma}$ by replacing each $\neg R(\bar{x})$ with $\hat{R}(\bar{x})$ and each $x \neq y$ with $N(x, y)$;
- if $F \in \{\mathbf{ihom}, \mathbf{emb}\}$ or $N$ appears in $\hat{\Sigma}$ extend $\hat{\Sigma}$ with:

$$x = y \vee N(x, y) \text{ and } x = y, N(x, y) \rightarrow \perp$$

- if $F \in \{\mathbf{fhom}, \mathbf{emb}\}$ extend $\hat{\Sigma}$ with:

$$R(\bar{x}) \vee \hat{R}(\bar{x}) \text{ and } R(\bar{x}), \hat{R}(\bar{x}) \rightarrow \perp$$

# Selected Bibliography


1. S.Abiteboul, R. Hull, and V.Vianu. *Foundations of Databases*. Addison Wesley, 1995.
2. A.V. Aho, C. Beeri, and J.D. Ullman. The theory of joins in relational databases. *ACM Trans. Database Syst., 4(3), 1979*.
3. C. Beeri, and M.Y. Vardi. A proof procedure for data dependencies. *J. ACM, 31(4), 1984*.
4. A. Calì, G. Gottlob, and M. Kifer. Taming the Infinite Chase: Query Answering under Expressive Relational Constraints. *Description Logics, 2008*.
5. A. Deutsch, A. Nash, and J. Remmel. The chase revisited. *PODS08*.
6. R. Fagin, P. G. Kolaitis, R. J. Miller, L. Popa. Data exchange: Semantics and query answering. *ICDT03*.
7. R. Fagin, P. G. Kolaitis, L. Popa. Data exchange: getting the core. *PODS03*.
8. A. Fuxman, P. G. Kolaitis, R. J. Miller, and W. C. Tan, Peer data exchange. *PODS05*.
9. G. Grahne, A. O. Mendelzon. Tableau Techniques for Querying Information Sources through Global Schemas. *ICDT99*.
10. G. Grahne, A. Onet. Data correspondence, exchange and repair. *ICDT10*.
11. S. Greco, F. Spezzano. Chase Termination: A Constraints Rewriting Approach. *VLDB10*.
12. P. G. Kolaitis, J. Panttaja, W. C. Tan. The complexity of data exchange. *PODS06*.
13. A. Hernich, N. Schweikardt. CWA-solutions for data exchange settings with target dependencies. *PODS07*.
14. L. Libkin. Data exchange and incomplete information. *PODS06*.
15. David Maier. The Theory of Relational Databases. *Computer Science Press 1983*.
16. D. Maier, A. O. Mendelzon, and Y. Sagiv. Testing Implications of Data Dependencies. *ACM Trans. Database Syst. 4(4), 1979*.
17. B. Marnette. Generalized schema-mappings: from termination to tractability. *PODS09*.
18. M. Meier, M. Schmidt, and G. Lausen. On Chase Termination Beyond Stratification. *PVLDB 2(1), 2009*.
19. R. Sethi: Testing for the Church-Rosser Property. *J. ACM 21(4): 671-679 (1974)*.