Evolution of Knowledge Bases: DL-Lite Case

Evgeny Kharlamov
Free University of Bozen-Bolzano
KRDB Research Centre

Joint work with:
Dmitriy Zheleznyakov,
Diego Calvanese,
Werner Nutt

[KharlamovZheleznyakov’11]
[Calvanese&al’10]

Dagstuhl Seminar
October, 2011
On 12 April 1961, aboard the Vostok 3KA-3 (Vostok 1), Gagarin became both the first human to travel into space, and the first to orbit the earth. His call sign was Kedr (Siberian Pine, Russian: Кедр).[12]

• Two flavors of Web content:
  • Plain text
  • Knowledge = annotated data + ontologies
• Web content is ubiquitously dynamic
• Goal of our study: to understand how to incorporate new knowledge into the old one
Semantic CVS and Evolution

- **CVS** for managing ontologies
- Several people try to edit the same knowledge
- **Problem**: how to incorporate new knowledge into the existing one
Ontologies

- Ontologies are a prime mechanism to bring semantics using:
  - annotations (e.g., date, name, hasHusband)
  - meta annotations (e.g., class, property)
  - classifications of annotations (e.g., subclass-of)
  - properties of annotations (e.g., domain, range)
  - ...

- Semantic Web Technologies behind ontologies
  - Resource description Framework (RDF)
  - Ontology Web Language (OWL): e.g. $\text{OWL2 QL} = \text{DL-Lite}$
  - Rule Languages (e.g. OWL 2 RL)
### DL-Lite$_R$ Ontology Language

- **DL-Lite Knowledge Base (KB):** $K = (TBox, ABox)$

<table>
<thead>
<tr>
<th>TBox assertions:</th>
<th>DL formulas:</th>
<th>FO formulas:</th>
</tr>
</thead>
<tbody>
<tr>
<td>role inclusion</td>
<td>$R \sqsubseteq P$</td>
<td>$\forall x, y. (R(x,y) \rightarrow P(x,y))$</td>
</tr>
<tr>
<td>concept inclusions</td>
<td>$A \sqsubseteq B, A \sqsubseteq \exists R$</td>
<td>$\forall x . (A(x) \rightarrow \exists y. R(x,y))$</td>
</tr>
<tr>
<td>concept disjointness</td>
<td>$A \sqsubseteq \neg B, A \sqsubseteq \neg \exists R$</td>
<td>$\forall x . (A(x) \rightarrow \neg B(x))$</td>
</tr>
</tbody>
</table>

- **ABox assertions:** $A(c) \neg B(c), \exists R(c), R(a,c), R^-(a,c)$

- **Properties:**
  - extended **Krom** (2-clauses) **Horn Logic**: existentials in rules’ heads
  - No **real disjunction**: if $K \models A(c) \lor B(d)$ then $K \models A(c)$ and $K \models B(d)$
  - Every KB $K$ has a canonical model (embedible in every model of $K$)
Knowledge Base: DL-Lite Example

**Roles:** HasHb

**Concepts:** Wife, Husband, Single
Priest, Bishop, Cleric

**TBox:**
- Wife $\equiv \exists$HasHb
- Husband $\equiv \exists$HasHb$^-$
- Husband $\sqsubseteq \neg$Single
- Priest $\sqsubseteq$ Cleric
- Cleric $\sqsubseteq$ Single
- Husband $\sqsubseteq \neg$Priest

**ABox:**
- Wife(Mary), Husband(John)
- Priest(Adam), Priest(Bob)
- Bishop(Carl)

Diagram:
- Single
- Husband (John)
- Cleric
- Bishop (Carl)
- Priest (Adam, Bob)
- Wife (Mary)
- HasHb
- 1..n
Knowledge Evolution: DL-Lite case

Old Knowledge: $K$
New Knowledge: $N$
Evolved Knowledge: $K \bowtie N$

DL-Lite KB:

Evol. Operator: $\bowtie$

DL-Lite KB:

Evolved knowledge $K \bowtie N$ should

- be consistent - no logical contradictions
  \[\text{Cleric}(Bob) \land \neg \text{Cleric}(Bob)\]
- be coherent - no useless components
  \[\text{Cleric} \subseteq \text{Single} \]
  \[\text{Cleric} \subseteq \neg \text{Single}\]
- preserve important info, e.g. TBox (\sim ABox evol.) or ABox (TBox evol.)
- minimally different from old KB \sim minimality of change principle
Practical Considerations

- **Closure** under evolution:
  Evolution results should be *expressible* in DL-Lite

- **Efficiency**:
  Evolution results should be computable in *PTIME*
Knowledge Evolution: Old Story Again?

- Knowledge evolution - old problem in AI
- Was primarily studied for Propositional Logics
- We study:
  - whether propositional approaches make sense for DLs
  - new, DL specific approaches

- Two main approaches to evolution of prop. theories:
  1. **Model-Based Approaches** (MBA): return a set of models
  2. **Formula-Based Approaches** (FBA): return set of formulas
Outline

- Knowledge evolution and DL-Lite
- Model-Based approaches and Issues for DL-Lite
- ABox Evolution in restricted DL-Lite
- ABox Evolution in full DL-Lite
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:  

New Knowledge $N$:  

Diagram showing relationships and entities.
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

\[ \text{Single} \rightarrow \text{Husband (John)} \]

\[ \text{Wife (Mary)} \]

\[ \text{HasHb} \rightarrow 1..n \]

New Knowledge $N$:

\[ \text{Cleric} \]

\[ \text{Minister (Carl)} \]

\[ \text{Priest (Adam, Bob)} \]

\[ \text{Mod}(K) \rightarrow \text{Mod}(N) \]
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

New Knowledge $N$:
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge \( K \):

New Knowledge \( N \):
Local Model-Based Approaches

• Important information for preservation: new knowledge

Old Knowledge $K$:

New Knowledge $N$:

\[ \text{Mod}(K) \]

\[ \text{Mininal distance} \]

\[ \text{Mod}(N) \]

\[ \checkmark \]
Local Model-Based Approaches

• Important information for preservation: new knowledge

Old Knowledge $K$:

\[ \text{Mod}(K) \]

New Knowledge $N$:

\[ \text{Mod}(N) \]
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

<table>
<thead>
<tr>
<th>Single</th>
<th>Husband</th>
<th>John</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HasHb</td>
<td>1..n</td>
</tr>
<tr>
<td>Wife</td>
<td>Mary</td>
<td></td>
</tr>
</tbody>
</table>

New Knowledge $N$:

Old Knowledge $K$:

$\text{Mod}(K)$

New Knowledge $N$:

$\text{Mod}(N)$

HasHb

Single

Husband

John

Wife

Mary

Minister

Carl

Priest

Adam, Bob
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

[Diagram showing relationships between Single, Husband (John), Wife (Mary), and HasHb (1..n)]

New Knowledge $N$:

[Diagram showing relationships between Cleric, Minister (Carl), Priest (Adam, Bob), and Mod($K$) and Mod($N$) with check marks]

- Minimal distance
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

New Knowledge $N$:

- Important information for preservation: new knowledge
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

- Single
- Husband John
- HasHb
- Wife Mary

New Knowledge $N$:

- Cleric
- Minister Carl
- Priest Adam, Bob

Mod($K$)

Mod($N$)

$\sqrt{\text{Local Model-Based Approaches}}$

$\sqrt{\text{New Knowledge}}$

$\sqrt{\text{Old Knowledge}}$
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

Single ------- x ------- Husband

John

HasHb

$1..n$

Wife

Mary

New Knowledge $N$:

Cleric

Minister

Carl

Priest

Adam, Bob

Local Model-Based Approaches

$\text{Mod}(K)$

$\text{Mod}(N)$

minimal distance

minimal distance

√

√

√
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

```
Single
X Husband John
HasHb
1..n Wife Mary
```

New Knowledge $N$:

```
Cleric
Minister Carl
Priest Adam, Bob
```

$\text{Mod}(K)$

$\text{Mod}(N)$

\[
\sqrt{\text{Mod}(K)}
\]

\[
\sqrt{\text{Mod}(N)}
\]

• Important information for preservation: new knowledge

Local Model-Based Approaches
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge \( K \):

New Knowledge \( N \):

The result of evolution:
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

- Single
- Husband: John

Representation?

The result of evolution:

- ✔ ✔ ✔
- ✗
Local Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $\mathbf{K}$:

Evolved KB $\mathbf{K'}$:
Global Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

New Knowledge $N$:
Global Model-Based Approaches

• Important information for preservation: new knowledge

Old Knowledge K:

New Knowledge N:
Global Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

$$\text{Single} \quad \times \quad \text{Husband} \quad \text{John}$$

$\text{Wife} \quad \text{Mary}$

$\text{HasHb} \quad 1..n$

$$\text{Mod}(K)$$

New Knowledge $N$:

$$\text{Cleric}$$

$$\text{Minister} \quad \text{Carl}$$

$$\text{Priest} \quad \text{Adam, Bob}$$

$$\text{Mod}(N)$$
Global Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

\[
\begin{array}{c}
\text{Single} \\
\times \\
\text{Husband} \\
\text{John} \\
\triangledown \\
\text{HasHb} \\
\downarrow \text{1..n} \\
\text{Wife} \\
\text{Mary}
\end{array}
\]

New Knowledge $N$:

\[
\begin{array}{c}
\text{Cleric} \\
\text{Minister} \\
\text{Carl} \\
\text{Priest} \\
\text{Adam, Bob}
\end{array}
\]
Global Model-Based Approaches

- Important information for preservation: new knowledge

**Old Knowledge K:**

- Single
- Husband: John
- Wife: Mary
- HasHb
- 1..n

**New Knowledge N:**

- Cleric
- Minister: Carl
- Priest: Adam, Bob

**The result of evolution:**

- ✓ ✓
- ✗ ✗
Global Model-Based Approaches

- Important information for preservation: new knowledge

Old Knowledge $K$:

```
Single  Married
      ×
  Husband  John
```

```
Wife  Mary
```

```
HasHb
1..n
```

Mod$(K)$

```
Representational ?
```

Evolved KB $K'$:

```
Single  Married
      ×
  Husband  John
```

```
Cleric
Carl  Adam, Bob
```

```
Wife  Mary
```

```
HasHb
1..n
```

Mod$(K')$

The result of evolution:

- ✓
- ✓
- ×
- ×
Distances Between Models

- All MBAs are based on
  - distances between interpretations
  - order on the distances to find minimums
- Distance for Propositional Logic
  - distance as a set \( \text{dist}_\cup(\mathcal{I}, \mathcal{J}) = \mathcal{I} \cup \mathcal{J} \)
  - distance as a number \( \text{dist}_{|\cup|}(\mathcal{I}, \mathcal{J}) = |\mathcal{I} \cup \mathcal{J}| \)
- Example:
  \[
  \mathcal{I} = \{p, q, r\} \quad \text{dist}_\cup(\mathcal{I}, \mathcal{J}) = \{q, r, s\}
  \]
  \[
  \mathcal{J} = \{p, s\} \quad \text{dist}_{|\cup|}(\mathcal{I}, \mathcal{J}) = 3
  \]
Distances Between Models: First Order Case

\[ \mathcal{I} = \{A(a), A(b), C(a)\} \quad \text{atom-based} \]

\[ \mathcal{J} = \{A(a), A(c), A(d), C(a)\} \quad \text{(relational) symbol-based} \]

\[ \text{dist}^a_\emptyset (\mathcal{I}, \mathcal{J}) = \{A(b), A(c), A(d)\} \]

\[ \text{dist}^a_{\emptyset \overline{\emptyset}} (\mathcal{I}, \mathcal{J}) = 3 \]

\[ \text{dist}^s_\emptyset (\mathcal{I}, \mathcal{J}) = \{A\} \]

\[ \text{dist}^s_{\emptyset \overline{\emptyset}} (\mathcal{I}, \mathcal{J}) = 1 \]
Space of Model-Based Approaches

Approach

global: $G$
local: $L$

set: $\emptyset$
number: $|\emptyset|$
symbols: $s$
atoms: $a$

Examples of Semantics:

$G^a_{\emptyset}$  $G^a_{\mid\emptyset\mid}$  $L^s_{\emptyset}$

Distance is built upon

Theorem: DL-Lite is not closed under every of the 8 MBAs: For every MBA there are $K, N$ s.t. $K \diamond N$ is not DL-Lite axiomatizable

- E.g. of issues: capturing evolution requires at least disjunction
**Inexpressibility: ABox evolution (TBox is fixed)**

**New Knowledge:** Single(John)

What happens with Mary?

**Intuition:** two cases are most likely
1. Mary is also single
2. Mary is married to another guy

MBAs gives another case:
3. Mary is married to either Bob or Adam but not to both

**Drawback 1:** counterintuitive

**Drawback 2:** inexpressible in DL-Lite

$$K \diamond N \not\models Priest(Adam) \lor Priest(Bob)$$

$$K \diamond N \not\models Priest(Adam)$$

$$K \diamond N \not\models Priest(Bob)$$
Inexpressibility: TBox evolution (ABox is fixed)

New Knowledge: Bishop $\subseteq$ Cleric

How does it affect the old KB?

Intuition:
Old knowledge is not affected, e.g., just add the assertion to the old KB

MBAs give strange model in evolution result:

$$M = \{ \text{Bishop(Carl), Cleric(Carl), } \neg \text{Single(Carl)}, ... \}$$

Hence, $\mathbf{K} \bowtie \mathbf{N} \not\models \text{Cleric} \subseteq \text{Single}$
Quest for Understanding

We want some understanding of:

• DL-Lite wrt evolution
  - What DL-Lite fragments are closed under MBAs?
  - What DL-Lite formulas are in charge of inexpressibility?

• Evolution wrt to DL-Lite
  - Can we capture DL-Lite evolution in richer logics?
  - What are these logics?

• Approximation
  - How to approximate DL-Lite evolution in DL-Lite?
Outline

• Knowledge evolution and DL-Lite
• Model-Based approaches and Issues for DL-Lite
• ABox Evolution in restricted DL-Lite
• ABox Evolution in full DL-Lite
DL-Lite\textsubscript{R}\textsuperscript{PR} fragment of DL-Lite\textsubscript{R}

- DL-Lite with restrictions on assertions involving roles
- Roles cannot be involved in disjointness
  - Forbidden role assertions:
  - Allowed only positive role assertions:
- Any concept inclusion assertions are allowed
- PR stands for Positive Role (interaction)
- Deciding
  \[K \nvDash \text{forbidden assertion}\]

\[
\exists R \sqsubseteq \neg B \quad \times
\]
\[
A \sqsubseteq \exists R \quad \checkmark
\]

can be (syntactically) checked PTIME in |K|
Example: Violating DL-Lite$_{PR}$ condition

New Knowledge: Single(John)

Forbidden assertions:
\[ \exists \text{HasHb} \sqsubseteq \neg \text{Priest} \]
\[ \exists \text{HasHb} \sqsubseteq \neg \text{Priest} \]

These assertions bring
- disjunction in $K \diamond N$ and
- inexpressibility of $K \diamond N$
Why is \( \text{DL-Lite}_R^{\text{PR}} \) Interesting?

- \( \text{DL-Lite}_R^{\text{PR}} \) is an extension of RDFS (its FO fragment) enriched with:
  - concept disjointness \( A \subseteq \neg B \)
  - mandatory participation \( A \subseteq \exists R \)
- Reasoning in \( \text{DL-Lite}_R^{\text{PR}} \) is as tractable as in \( \text{DL-Lite}_R \):
  - LogSpace query answering for CQs in data complexity
Evolution in DL-Lite\textsubscript{PR}^{\text{PR}}

\[ G^a_{\subseteq} \quad G^a_{\#} \quad G^s_{\subseteq} \equiv G^s_{\#} \]

\[ L^a_{\subseteq} \equiv L^a_{\#} \equiv G^a_{\subseteq} \equiv G^a_{\#} \quad L^s_{\subseteq} \text{ expressible} \quad \quad L^s_{\#} \text{ inexpressible} \]

Theorem:

• All atom based coincide: they give the same result

• Local and global symbol based semantics coincide

Theorem:

• Closed under local and global atom based and global symbol based.
• Not closed under local symbol based.

Algorithms:

We developed 3 polynomial time algorithms to compute

1. Atom based semantics
2. Global symbol based semantics
3. Minimal sound approximation of local symbol based
Evolution in DL-Lite\textsubscript{PR}\textsuperscript{R}

- DL-Lite\textsubscript{PR}\textsuperscript{R} is a maximal sublogy of DL-Lite\textsubscript{R} closed under the four atom based semantics

**Theorem:** \( T \) - DL-Lite\textsubscript{R} TBox s.t.:

\[
T \models \exists R \subseteq \neg B
\]

Then, for every of the 8 MBAs there are ABoxes \( A \) and \( N \) s.t. \((T, A) \diamond N\) is not axiomatizable in DL-Lite\textsubscript{R}
Example: Global Symbol-Based Evolution

- Recall: semantics traces changes in predicates only
- Example:

\[ A : C(a), C(b), C(c), B(e) \quad N : \neg C(a) \]

- We have to drop \( C(a) \) from \( A \)
- What to do with \( C(b), C(c) \) and \( B(e) \)?

\[ A' : \neg C(a), B(e) \]

- If there is an \textit{unavoidable change} in a concept than delete all the \textit{old information} about this concept
Example: Local Symbol-Based Evolution

- Example:

\[ \mathcal{A} : C(a), C(b), C(c), B(e) \quad \mathcal{N} : C(d) \]

- What to do with C(a), C(b), C(c) and B(e)?
  delete all about C

\[ \mathcal{A}' : C(d), B(e) \]

- If the new data makes any change in a concept, then delete all the old data about this concept

- Capturing this semantics requires disjunction

- The algorithm is a minimal sound approximation of the semantics
Concluding on Model-Based DL-Lite\textsuperscript{PR} Evol.

- All four atom-based semantics:
  - give reasonable / intuitive results of evolution
  - can be captured in DL-Lite\textsuperscript{PR} with a PTIME algorithm
- Symbol based semantics:
  - give unintuitive results of evolution
  - 2 out of 4 semantics cannot be captured DL-Lite\textsuperscript{PR} → does not look good for ABox evolution
- We have 6 model-based approaches to do evolution of RDFS ontologies
Outline

- Knowledge evolution and DL-Lite
- Model-Based approaches and Issues for DL-Lite
- ABox Evolution in restricted DL-Lite
- ABox Evolution in full DL-Lite
Winslett’s Semantics for ABox Evolution

- For the full $\text{DL-Lite}_R$ we focus on Winslett’s semantics (WS)
- WS corresponds to an atom-based semantics $L^a_\Theta$
- Atom-based semantics behave well for $\text{DL-Lite}_R^{PR}$
- WS proved its importance:
  - Updates of ABoxes in expressive description logic [Liu&al:2006]
  - DL-Lite ABox updates [De Giacomo&al:2006]

Recall: DL-Lite is not closed under WS
Prototypal Sets

• $S$ - a set of models, finite $\{ J_1, \ldots, J_n \} \subseteq S$ is a prototypal set for $S$ if for every $I \in S$, there $J \in \{ J_1, \ldots, J_n \}$ that can be homo. embedded in $I$

$$\forall I \in S, \exists J \in \{ J_1, \ldots, J_n \}: \quad J \hookrightarrow I$$

• We call elements of prototypal sets: prototypes

• Not all sets of models have prototypal sets

• DL-Lite KBs have canonical models $\Rightarrow$ prototypal sets are singletons

**Theorem:** $K$ - DL-Lite$_R$ KB, $N$ - ABox. Then, ABox evolution $K \diamond N$ under WS has a prototypal set
Prototypal Sets

Theorem: \( K - \text{DL-Lite}_R \text{ KB, N - ABox.} \)

Then,

ABox evolution \( K \bowtie N \) under WS has a prototypal set
Behavior of Prototypes for ABox Evolution

1. Take $I_0$ - slightly extended canonical model of $K$

2. Construct zero-prototype $I_0 \bowtie N \rightarrow J_0$

3. Construct other prototypes from $I_0 \bowtie N \rightarrow \{J_1, ..., J_n\}$

**Theorem:** $K$ - DL-LiteR KB, $N$ - ABox. Then,

- prot. set of ABox evol. $K \bowtie N$ is of exponential size in $|K \cup N|$  
- all prototypes can be constructed by ConstrProt
Approximating ABox Evolution

Theorem: \( K \)- DL-Lite\(_R\) KB, \( N \)- ABox. Then, a maximal sound approx. of ABox evolution \( K \diamond N \) exists and can be computed in PTIME

- Prototypal sets help in approximation of DL-Lite\(_R\) ABox evolution
- Maximal sound approximation:
  - \( K_{ap} \) is a **sound** approximation of \( K \diamond N \) if \( K \diamond N \subseteq \text{Mod}(K_{ap}) \)
  - \( K_{ap} \) is **maximal** if \( \forall K' \)- sound approx.:
    \( K \diamond N \subseteq \text{Mod}(K_{ap}) \subseteq \text{Mod}(K') \)
Conclusion

1. Inexpressibility

We described the space of 8 MBAs to evolution, and showed inexpressibility of all MBAs for DL-Lite

2. Reasons of inexpressibility

We found DL-Lite formulas responsible for inexpressibility

2. DL-Lite fragments

We found a maximal fragment of DL-Lite closed under 6 out of 8 evolution semantics

4. Approximation

We introduced prototypes and based on them showed how to do PTIME maximal sound approximation of Winslett’s semantics in DL-Ltie
Thank you!

**ACSI Project**
Artifact-Centric Service Interoperation
FP 7 grant, agreement n. 257593
http://www.acsi-project.eu/

**Webdam Project**
Foundations of Web Data Management
ERC FP7 grant, agreement n. 226513
http://webdam.inria.fr/
Presentation is Base On


Further Reading

