TECHNISCHE UNIVERSITÄT DRESDEN

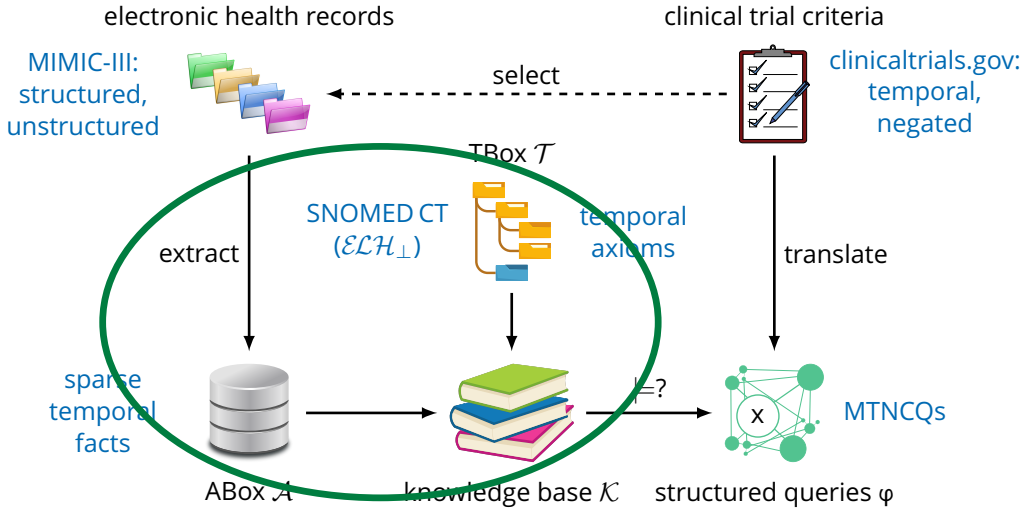
Stefan Borgwardt, Walter Forkel, Alisa Kovtunova

Institute of Theoretical Computer Science, Chair of Automata Theory

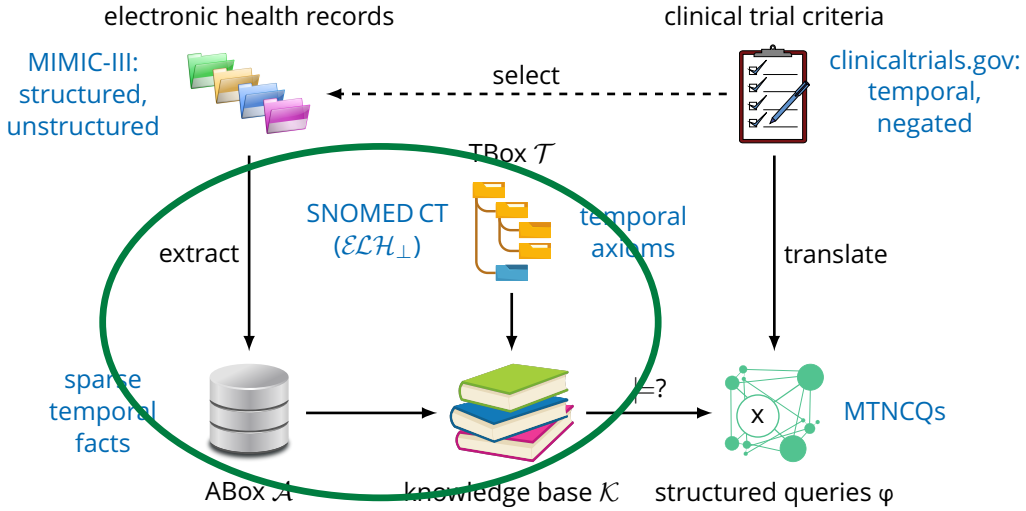
Finding New Diamonds: Temporal Minimal-World Query Answering over Sparse ABoxes

Bozen-Bolzano, 17th September 2019

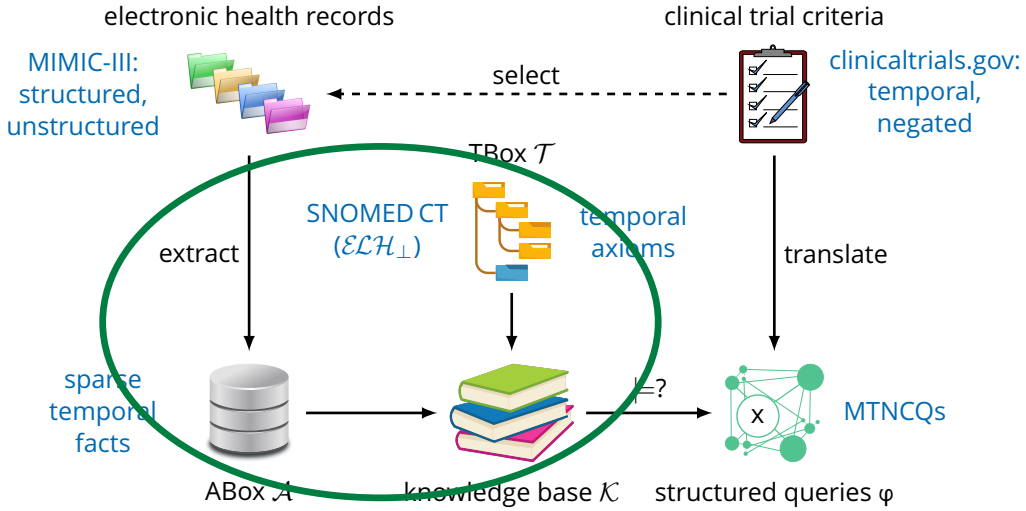
Motivation


electronic health records
clinical trial criteria
MIMIC-III: structured, unstructured
select
clinicaltrials.gov: temporal, negated
TBox $\mathcal{T}$
SNOMED CT ($\mathcal{ELH}_\perp$)
temporal axioms
extract
translate
sparse temporal facts
$\models$?
MTNCQs
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

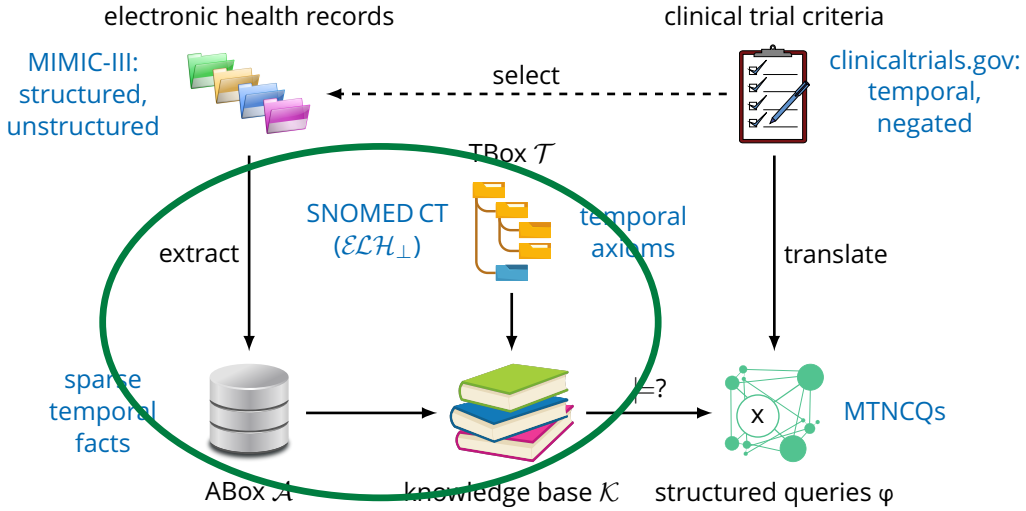
Motivation


electronic health records
clinical trial criteria
MIMIC-III: structured, unstructured
select
clinicaltrials.gov: temporal, negated
TBox $\mathcal{T}$
SNOMED CT ($\mathcal{ELH}_\perp$)
temporal axioms
extract
translate
sparse temporal facts
$\models$?
MTNCQs
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

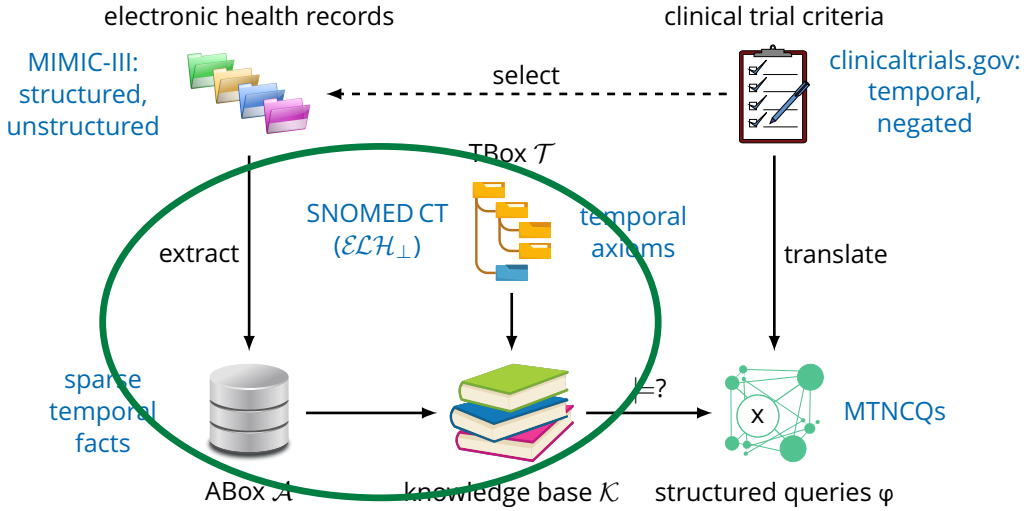
Motivation


electronic health records
clinical trial criteria
MIMIC-III: structured, unstructured
select
clinicaltrials.gov: temporal, negated
TBox $\mathcal{T}$
SNOMED CT ($\mathcal{ELH}_\perp$)
temporal axioms
extract
translate
sparse temporal facts
$\models$?
MTNCQs
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

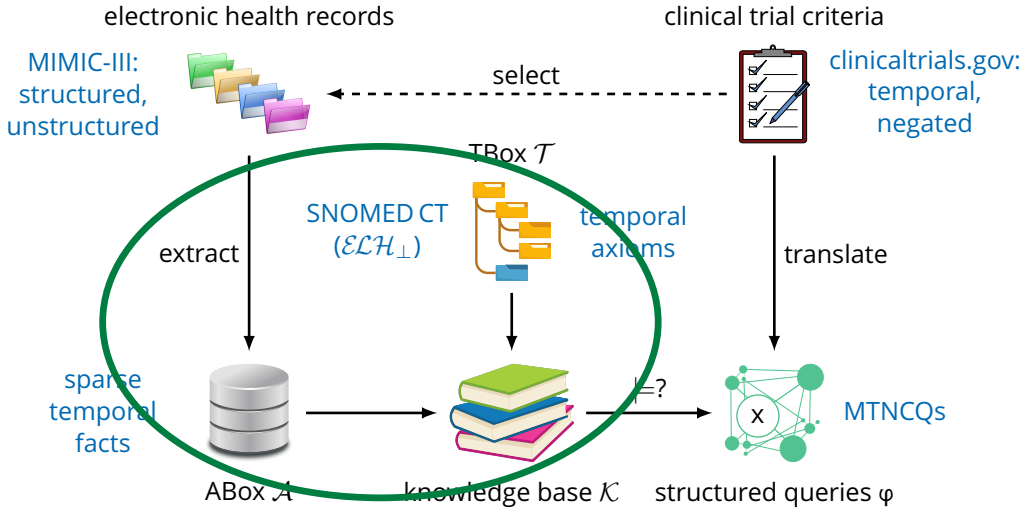
Motivation


electronic health records
clinical trial criteria
MIMIC-III: structured, unstructured
select
clinicaltrials.gov: temporal, negated
TBox $\mathcal{T}$
SNOMED CT ($\mathcal{ELH}_\perp$)
temporal axioms
extract
translate
sparse temporal facts
$\models$?
MTNCQs
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

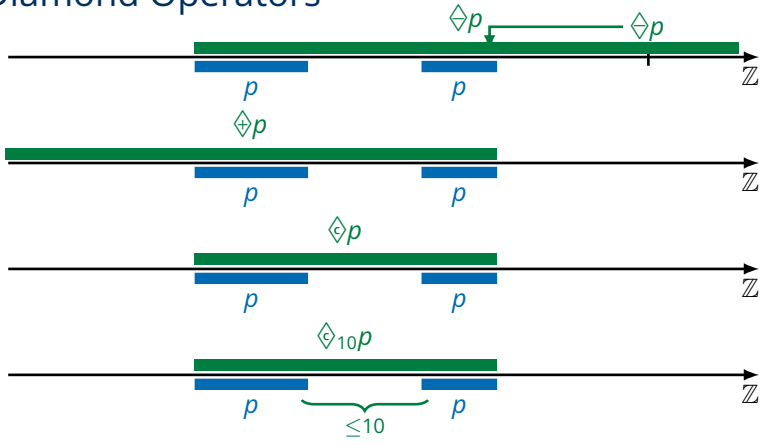
Motivation


electronic health records
clinical trial criteria
MIMIC-III: structured, unstructured
select
clinicaltrials.gov: temporal, negated
TBox $\mathcal{T}$
SNOMED CT ($\mathcal{ELH}_\perp$)
temporal axioms
extract
translate
sparse temporal facts
$\models$?
MTNCQs
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

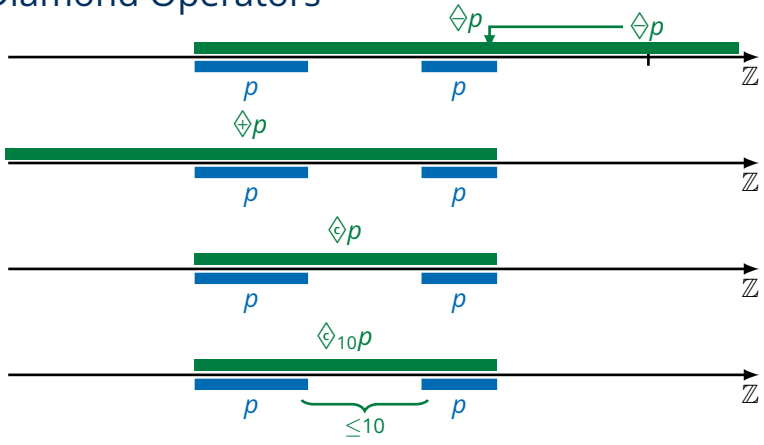
Motivation


electronic health records
clinical trial criteria
MIMIC-III: structured, unstructured
select
clinicaltrials.gov: temporal, negated
TBox $\mathcal{T}$
SNOMED CT ($\mathcal{ELH}_\perp$)
temporal axioms
extract
translate
sparse temporal facts
$\models$?
MTNCQs
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

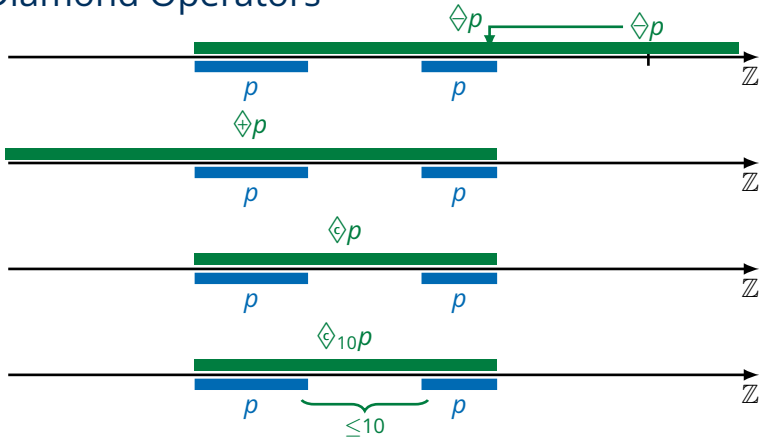
Metric Diamond Operators



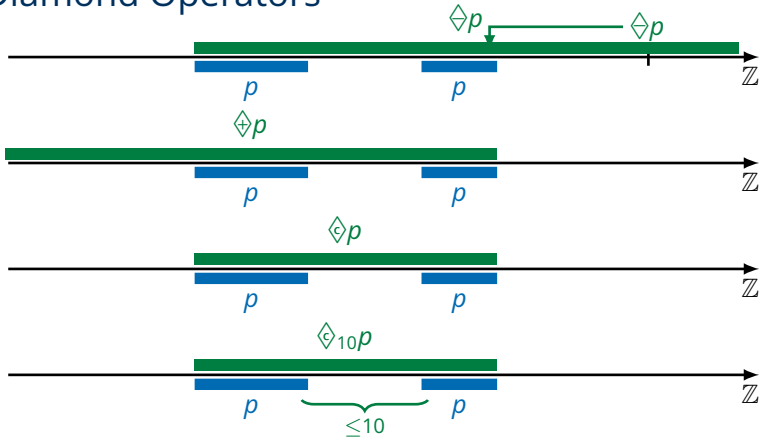
Metric Diamond Operators



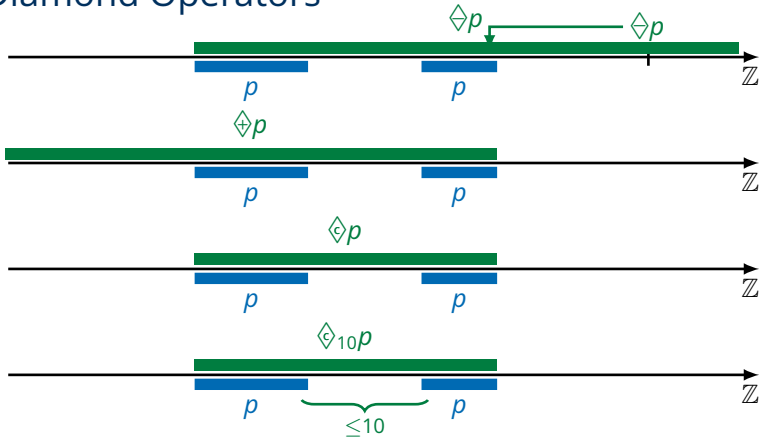
Metric Diamond Operators



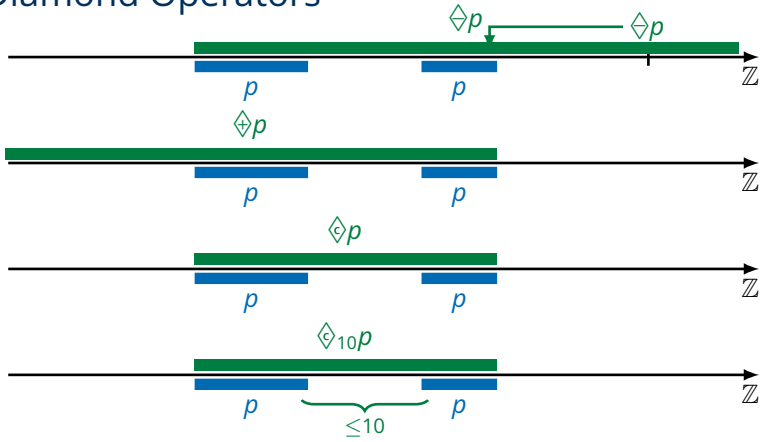
Metric Diamond Operators



Metric Diamond Operators



Metric Diamond Operators



- can be seen as closure operators $2^{\mathbb{Z}} \to 2^{\mathbb{Z}}$
- $\{\Diamond_1, \Diamond_2, \ldots, \Diamond, \Diamond, \Diamond, \Diamond\}$ is closed under pointwise $\cap, \cup, \circ$

A Temporal Description Logic

based on (Gutiérrez-Basulto, Jung, and Kontchakov 2016)

$$\Diamond\exists \text{diagnosis.Diabetes} \sqsubseteq \exists \text{diagnosis.Diabetes}$$

A Temporal Description Logic

based on (Gutiérrez-Basulto, Jung, and Kontchakov 2016)

$$\Diamond\exists \text{diagnosis.Diabetes} \sqsubseteq \exists \text{diagnosis.Diabetes}$$

$$\Diamond_8\exists \text{diagnosis.BrokenLeg} \sqsubseteq \exists \text{diagnosis.BrokenLeg}$$

A Temporal Description Logic

based on (Gutiérrez-Basulto, Jung, and Kontchakov 2016)

$$\Diamond\exists \text{diagnosis.Diabetes} \sqsubseteq \exists \text{diagnosis.Diabetes}$$
$$\Diamond_8\exists \text{diagnosis.BrokenLeg} \sqsubseteq \exists \text{diagnosis.BrokenLeg}$$

$\mathcal{TELH}_{\bot}^{\Diamond,\text{lhs}}$, a combination of $\mathcal{ELH}_{\bot}$ with diamond operators:

$$\Diamond A \sqsubseteq B,\ A_1 \sqcap A_2 \sqsubseteq B,\ \Diamond r \sqsubseteq s,\ \Diamond A \sqsubseteq \exists r.B,\ \exists r.A \sqsubseteq B, \quad A(a,i),\ r(a,b,i)$$

A Temporal Description Logic

based on (Gutiérrez-Basulto, Jung, and Kontchakov 2016)

$$\Diamond\exists\text{diagnosis.Diabetes} \sqsubseteq \exists\text{diagnosis.Diabetes}$$
$$\Diamond_8\exists\text{diagnosis.BrokenLeg} \sqsubseteq \exists\text{diagnosis.BrokenLeg}$$

$\mathcal{TELH}_{\bot}^{\Diamond,\text{lhs}}$, a combination of $\mathcal{ELH}_{\bot}$ with diamond operators:

$$\Diamond A \sqsubseteq B,\ A_1 \sqcap A_2 \sqsubseteq B,\ \Diamond r \sqsubseteq s,\ \Diamond A \sqsubseteq \exists r.B,\ \exists r.A \sqsubseteq B,\quad A(a,i),\ r(a,b,i)$$

$\Diamond = \Diamond$

Translation into first-order logic:

$$\forall x, t, t'.\, A(x,t') \wedge (t' \leq t) \rightarrow \exists y.\, r(x,y,t) \wedge B(y,t)$$

A Temporal Description Logic

based on (Gutiérrez-Basulto, Jung, and Kontchakov 2016)

$$\Diamond\exists\text{diagnosis.Diabetes} \sqsubseteq \exists\text{diagnosis.Diabetes}$$
$$\Diamond_8\exists\text{diagnosis.BrokenLeg} \sqsubseteq \exists\text{diagnosis.BrokenLeg}$$

$\mathcal{TELH}_\bot^{\Diamond,\text{lhs}}$, a combination of $\mathcal{ELH}_\bot$ with diamond operators:

$$\Diamond A \sqsubseteq B,\ A_1 \sqcap A_2 \sqsubseteq B,\ \Diamond r \sqsubseteq s,\ \Diamond A \sqsubseteq \exists r.B,\ \exists r.A \sqsubseteq B,\quad A(a,i),\ r(a,b,i)$$

$\Diamond = \Diamond$

Translation into first-order logic:

$$\forall x,t,t'.\, A(x,t') \wedge (t' \leq t) \rightarrow \exists y.\, r(x,y,t) \wedge B(y,t)$$

Theorem

Entailment in $\mathcal{TELH}_\bot^{\Diamond,\text{lhs}}$ is P-complete.

(Some) Completion Rules

$$\frac{\Diamond_{\star} A_1 \sqsubseteq A_2 \qquad \Diamond_{\bullet} A_2 \sqsubseteq A_3}{(\Diamond_{\bullet} \circ \Diamond_{\star}) A_1 \sqsubseteq A_3} \qquad \frac{\Diamond_{\star} A \sqsubseteq A_1 \qquad \Diamond_{\bullet} A \sqsubseteq A_2 \qquad A_1 \sqcap A_2 \sqsubseteq B}{(\Diamond_{\star} \cap \Diamond_{\bullet}) A \sqsubseteq B}$$

(Some) Completion Rules

$$\frac{\Diamond_{\star} A_1 \sqsubseteq A_2 \qquad \Diamond_{\bullet} A_2 \sqsubseteq A_3}{(\Diamond_{\bullet} \circ \Diamond_{\star}) A_1 \sqsubseteq A_3}$$

$$\frac{\Diamond_{\star} A \sqsubseteq A_1 \qquad \Diamond_{\bullet} A \sqsubseteq A_2 \qquad A_1 \sqcap A_2 \sqsubseteq B}{(\Diamond_{\star} \cap \Diamond_{\bullet}) A \sqsubseteq B}$$

$$\frac{i \in \Diamond_{\star} A(a) \qquad \Diamond_{\star} A \sqsubseteq B}{B(a, i)}$$

(only for representative time points i)

(Some) Completion Rules

$$\frac{\Diamond_{\star} A_1 \sqsubseteq A_2 \qquad \Diamond_{\bullet} A_2 \sqsubseteq A_3}{(\Diamond_{\bullet} \circ \Diamond_{\star}) A_1 \sqsubseteq A_3}$$

$$\frac{\Diamond_{\star} A \sqsubseteq A_1 \qquad \Diamond_{\bullet} A \sqsubseteq A_2 \qquad A_1 \sqcap A_2 \sqsubseteq B}{(\Diamond_{\star} \cap \Diamond_{\bullet}) A \sqsubseteq B}$$

$$\frac{i \in \Diamond_{\star} A(a) \qquad \Diamond_{\star} A \sqsubseteq B}{B(a, i)}$$

(only for representative time points i)

(Some) Completion Rules

$$\frac{\Diamond_{\star} A_1 \sqsubseteq A_2 \qquad \Diamond_{\bullet} A_2 \sqsubseteq A_3}{(\Diamond_{\bullet} \circ \Diamond_{\star}) A_1 \sqsubseteq A_3}$$

$$\frac{\Diamond_{\star} A \sqsubseteq A_1 \qquad \Diamond_{\bullet} A \sqsubseteq A_2 \qquad A_1 \sqcap A_2 \sqsubseteq B}{(\Diamond_{\star} \cap \Diamond_{\bullet}) A \sqsubseteq B}$$

$$\frac{i \in \Diamond_{\star} A(a) \qquad \Diamond_{\star} A \sqsubseteq B}{B(a, i)}$$

(only for representative time points i)

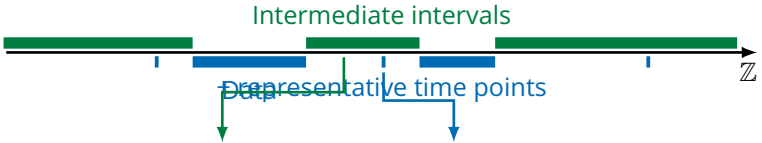
(Some) Completion Rules

$$\frac{\Diamond_{\star} A_1 \sqsubseteq A_2 \qquad \Diamond_{\bullet} A_2 \sqsubseteq A_3}{(\Diamond_{\bullet} \circ \Diamond_{\star}) A_1 \sqsubseteq A_3}$$

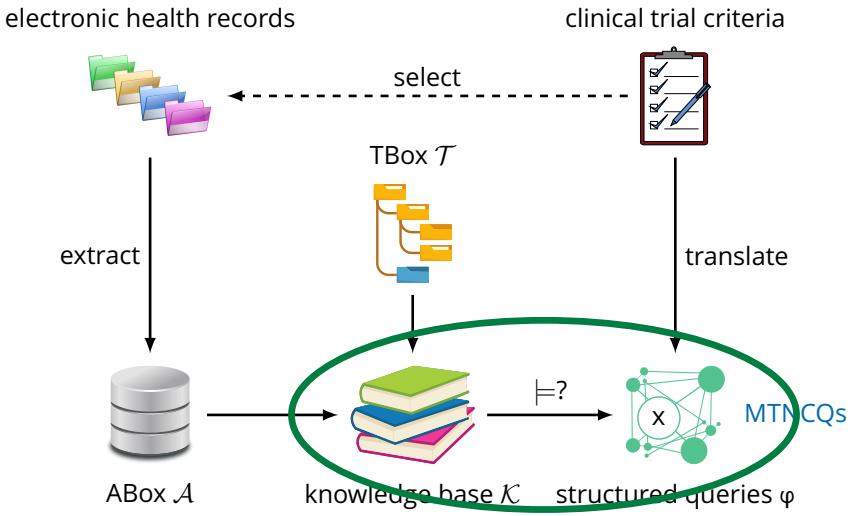
$$\frac{\Diamond_{\star} A \sqsubseteq A_1 \qquad \Diamond_{\bullet} A \sqsubseteq A_2 \qquad A_1 \sqcap A_2 \sqsubseteq B}{(\Diamond_{\star} \cap \Diamond_{\bullet}) A \sqsubseteq B}$$

$$\frac{i \in \Diamond_{\star} A(a) \qquad \Diamond_{\star} A \sqsubseteq B}{B(a, i)}$$

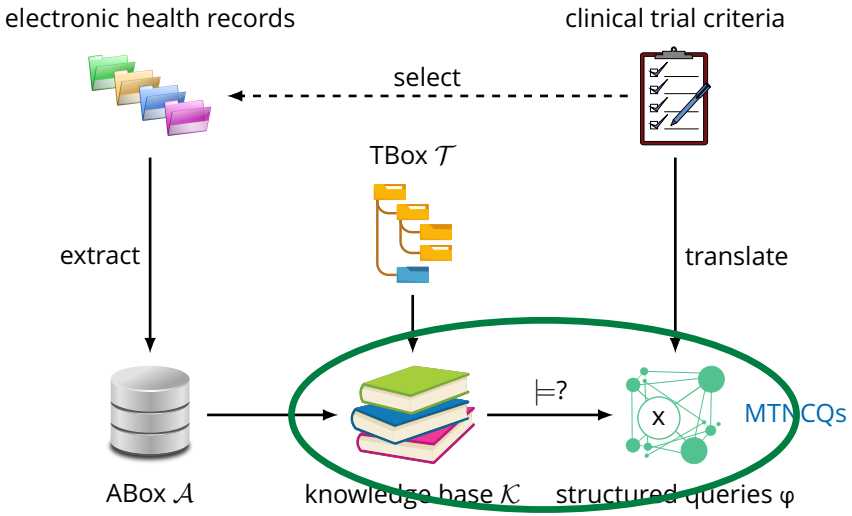
(only for representative time points i)



Flashback


electronic health records
clinical trial criteria
select
TBox $\mathcal{T}$
extract
translate
$\models$?
x
MTNCQs
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

Flashback


electronic health records
clinical trial criteria
select
TBox $\mathcal{T}$
extract
translate
$\models$?
x
MTNCQs
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

Metric Temporal Conjunctive Queries with Negation (MTNCQs)

"Diagnosis of Rheumatoid Arthritis (RA)
of more than 6 months and less than 15 years."

https://clinicaltrials.gov/ct2/show/NCT01198002

Metric Temporal Conjunctive Queries with Negation (MTNCQs)

"Diagnosis of Rheumatoid Arthritis (RA)
of more than 6 months and less than 15 years."

https://clinicaltrials.gov/ct2/show/NCT01198002

$$\Box_{[-6,0]}\big(\exists y.\mathrm{diagnosis}(x,y) \wedge \mathrm{RA}(y)\big) \wedge \neg\Box_{[-180,0]}\big(\exists y.\mathrm{diagnosis}(x,y) \wedge \mathrm{RA}(y)\big)$$

Metric Temporal Conjunctive Queries with Negation (MTNCQs)

"Diagnosis of Rheumatoid Arthritis (RA)
of more than 6 months and less than 15 years."

https://clinicaltrials.gov/ct2/show/NCT01198002

$$\Box_{[-6,0]}\big(\exists y.\mathrm{diagnosis}(x,y) \wedge \mathrm{RA}(y)\big) \wedge \neg\Box_{[-180,0]}\big(\exists y.\mathrm{diagnosis}(x,y) \wedge \mathrm{RA}(y)\big)$$

conjunctive queries with negation (NCQs)
+ arbitrary metric temporal operators

Metric Temporal Conjunctive Queries with Negation (MTNCQs)

"Diagnosis of Rheumatoid Arthritis (RA)
of more than 6 months and less than 15 years."

https://clinicaltrials.gov/ct2/show/NCT01198002

$$\Box_{[-6,0]}\big(\exists y.\mathrm{diagnosis}(x,y) \wedge \mathrm{RA}(y)\big) \wedge \neg\Box_{[-180,0]}\big(\exists y.\mathrm{diagnosis}(x,y) \wedge \mathrm{RA}(y)\big)$$

conjunctive queries with negation (NCQs)
+ arbitrary metric temporal operators

negative information is rarely entailed $\rightsquigarrow$ use "negation by failure"

Minimal-World Semantics

(Borgwardt and Forkel 2019)

Minimal-World Semantics

(Borgwardt and Forkel 2019)

Knowledge base $\mathcal{K}$:

$$\text{DiabetesPatient} \equiv \exists \text{diagnosis}.\text{Diabetes}$$

$$\text{DiabetesType1Patient} \equiv \exists \text{diagnosis}.\text{DiabetesType1}$$

$$\text{DiabetesType1} \sqsubseteq \text{Diabetes}$$

$$\text{DiabetesType1Patient}(p)$$

Minimal-World Semantics

(Borgwardt and Forkel 2019)

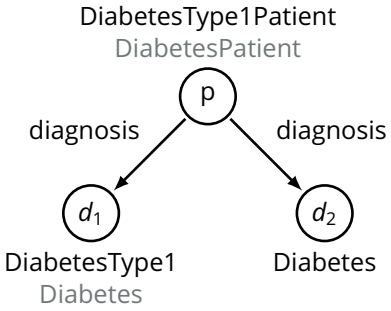
Knowledge base $\mathcal{K}$: $\text{DiabetesPatient} \equiv \exists \text{diagnosis}.\text{Diabetes}$

$\text{DiabetesType1Patient} \equiv \exists \text{diagnosis}.\text{DiabetesType1}$

$\text{DiabetesType1} \sqsubseteq \text{Diabetes}$

$\text{DiabetesType1Patient}(\text{p})$

canonical model $\mathcal{I}_{\mathcal{K}}$:



Minimal-World Semantics

(Borgwardt and Forkel 2019)

Knowledge base $\mathcal{K}$: $\text{DiabetesPatient} \equiv \exists \text{diagnosis}.\text{Diabetes}$

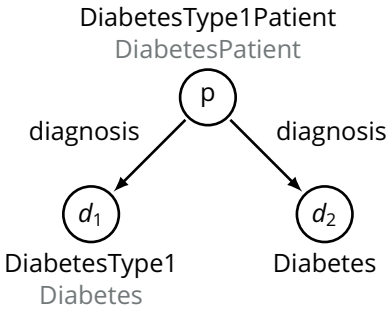
$\text{DiabetesType1Patient} \equiv \exists \text{diagnosis}.\text{DiabetesType1}$

$\text{DiabetesType1} \sqsubseteq \text{Diabetes}$

$\text{DiabetesType1Patient}(\text{p})$

canonical model $\mathcal{I}_\mathcal{K}$:

For positive queries φ: $\mathcal{K} \models \varphi$ iff $\mathcal{I}_\mathcal{K} \models \varphi$



Minimal-World Semantics

(Borgwardt and Forkel 2019)

Knowledge base $\mathcal{K}$: $\text{DiabetesPatient} \equiv \exists \text{diagnosis}.\text{Diabetes}$

$\text{DiabetesType1Patient} \equiv \exists \text{diagnosis}.\text{DiabetesType1}$

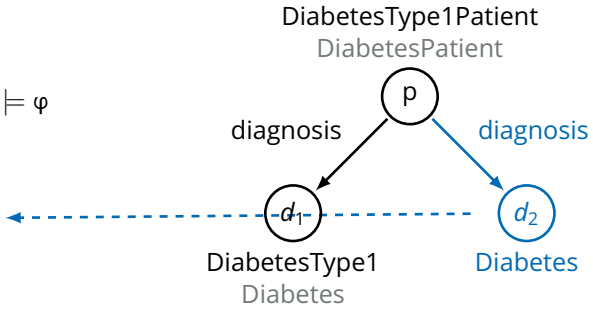
$\text{DiabetesType1} \sqsubseteq \text{Diabetes}$

DiabetesType1Patient(p)

canonical model $\mathcal{I}_\mathcal{K}$:

For positive queries φ: $\mathcal{K} \models \varphi$ iff $\mathcal{I}_\mathcal{K} \models \varphi$

Patient p has a diagnosis with Diabetes that is not DiabetesType1.



Minimal-World Semantics

(Borgwardt and Forkel 2019)

Knowledge base $\mathcal{K}$: $\text{DiabetesPatient} \equiv \exists \text{diagnosis}.\text{Diabetes}$

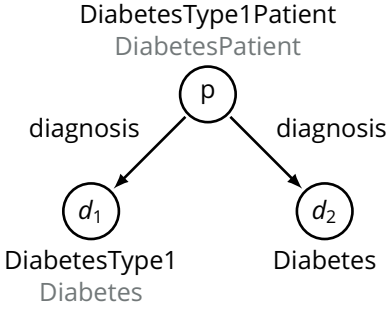
$\text{DiabetesType1Patient} \equiv \exists \text{diagnosis}.\text{DiabetesType1}$

$\text{DiabetesType1} \sqsubseteq \text{Diabetes}$

$\text{DiabetesType1Patient}(\text{p})$

Minimal canonical model $\mathcal{I}_\mathcal{K}$:

For positive queries φ: $\mathcal{K} \models \varphi$ iff $\mathcal{I}_\mathcal{K} \models \varphi$



Minimal-World Semantics

(Borgwardt and Forkel 2019)

Knowledge base $\mathcal{K}$: $\text{DiabetesPatient} \equiv \exists \text{diagnosis}.\text{Diabetes}$

$\text{DiabetesType1Patient} \equiv \exists \text{diagnosis}.\text{DiabetesType1}$

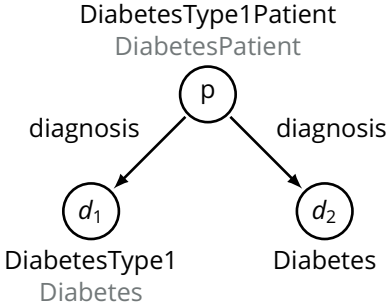
$\text{DiabetesType1} \sqsubseteq \text{Diabetes}$

$\text{DiabetesType1Patient}(\text{p})$

Minimal canonical model $\mathcal{I}_\mathcal{K}$:

For positive queries φ: $\mathcal{K} \models \varphi$ iff $\mathcal{I}_\mathcal{K} \models \varphi$

Minimal-world semantics:
use $\mathcal{I}_\mathcal{K} \models \varphi$ instead of $\mathcal{K} \models \varphi$



Temporal Minimal-World Semantics

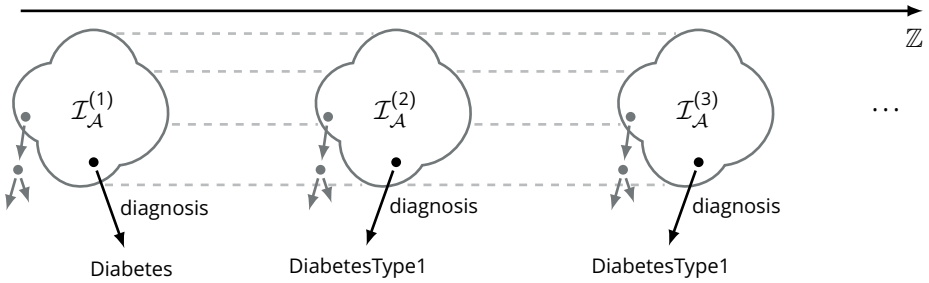
$$\Diamond\exists \text{diagnosis.Diabetes} \sqsubseteq \exists \text{diagnosis.Diabetes}$$

$$\Diamond\exists \text{diagnosis.DiabetesType1} \sqsubseteq \exists \text{diagnosis.DiabetesType1}$$

Temporal Minimal-World Semantics

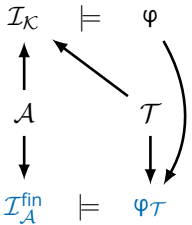
$\Diamond\exists$diagnosis.Diabetes $\sqsubseteq \exists$diagnosis.Diabetes

$\Diamond\exists$diagnosis.DiabetesType1 $\sqsubseteq \exists$diagnosis.DiabetesType1



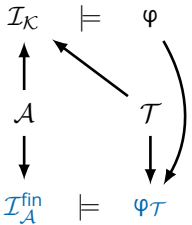
Combined Rewriting

$\mathcal{I}_\mathcal{K}$ can be infinite!



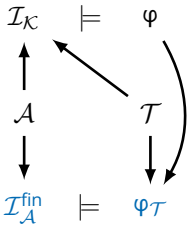
Combined Rewriting

$\mathcal{I}_\mathcal{K}$ can be infinite!



Combined Rewriting

$\mathcal{I}_\mathcal{K}$ can be infinite!

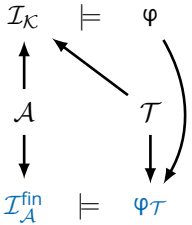


Combined Rewriting

$\mathcal{I}_\mathcal{K}$ can be infinite!

- NCQ rewriting from (Borgwardt and Forkel 2019)
- rewriting temporal operators into FOL:

$$\left(\Box_{[-6,0]}\psi\right)_\mathcal{T}(t) = \forall t'.\,(t' - t \in [-6, 0]) \rightarrow \psi_\mathcal{T}(t')$$

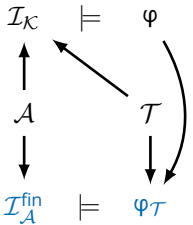


Combined Rewriting

$\mathcal{I}_\mathcal{K}$ can be infinite!

- NCQ rewriting from (Borgwardt and Forkel 2019)
- rewriting temporal operators into FOL:

$$\left(\Box_{[-6,0]}\psi\right)_\mathcal{T}(t) = \forall t'.\,(t' - t \in [-6,0]) \rightarrow \psi_\mathcal{T}(t')$$



Theorem

In $\mathcal{TELH}_\bot^{\Diamond,\text{lhs}}$, entailment of rooted MTNCQs is ExpSpace-complete, and P-complete in data complexity.

It’s Not That Easy

Queries can distinguish more time points!

$\square_{[A,B]}$ sum of all numbers in φ (∞ does not count)

It’s Not That Easy

Queries can distinguish more time points!

$\square_{[A,B]}$ sum of all numbers in φ (∞ does not count)

It’s Not That Easy

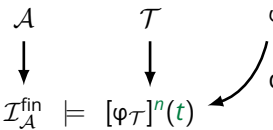
Queries can distinguish more time points!

$\square_{[A,B]}$ sum of all numbers in φ (∞ does not count)

It’s Not That Easy

Queries can distinguish more time points!

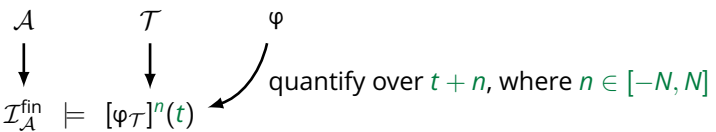
sum of all numbers in φ (∞ does not count)



quantify over $t + n$, where $n \in [-N, N]$

It's Not That Easy

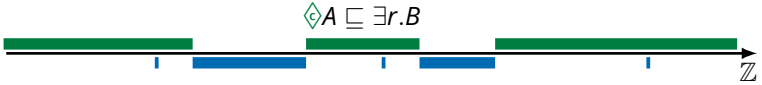
Queries can distinguish more time points!



N, and therefore $\varphi_{\mathcal{T}}$, is exponentially large. Model checking in FOL is in PSPACE.

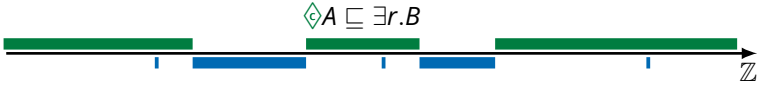
$\rightarrow$ EXPSPACE

Summary

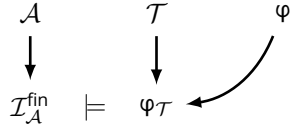


Entailment in $\mathcal{TELH}_{\perp}^{\Diamond,\mathsf{lhs}}$ over sparse data is P-complete.

Summary

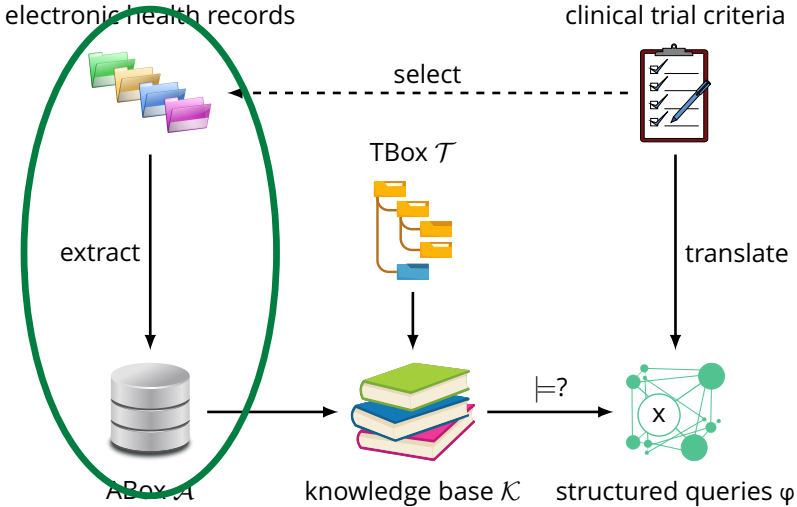


Entailment in $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs}}$ over sparse data is P-complete.

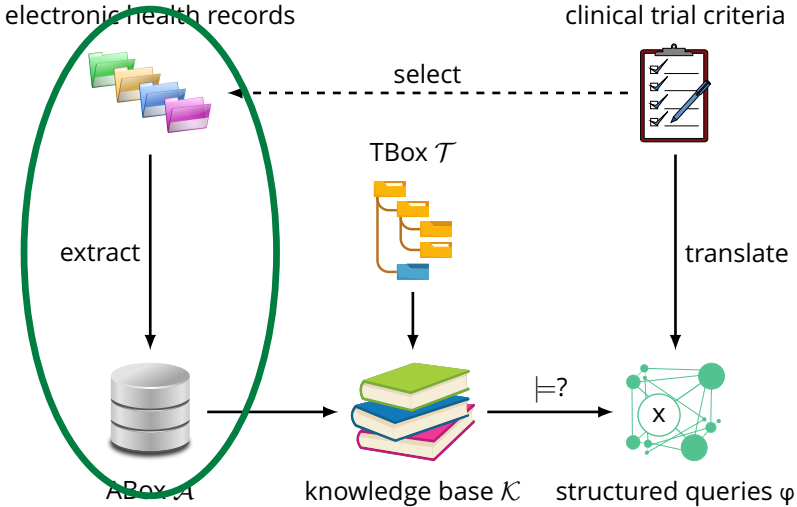


Answering MTNCQs in $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs}}$ is P-complete in data complexity.

Outlook


electronic health records
clinical trial criteria
select
TBox $\mathcal{T}$
extract
translate
$\models$?
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

Outlook


electronic health records
clinical trial criteria
select
TBox $\mathcal{T}$
extract
translate
$\models$?
ABox $\mathcal{A}$
knowledge base $\mathcal{K}$
structured queries φ

Thank you!

Borgwardt, Stefan and Walter Forkel (2019). “Closed-World Semantics for Conjunctive Queries with Negation over $\mathcal{ELH}_\bot$ Ontologies”. In: *Proc. JELIA'19*, pages 371–386.

Gutiérrez-Basulto, Víctor, Jean Christoph Jung, and Roman Kontchakov (2016). “Temporalized $\mathcal{EL}$ Ontologies for Accessing Temporal Data: Complexity of Atomic Queries”. In: *Proc. IJCAI'16*, pages 1102–1108.

Pictures from https://publicdomainvectors.org