Assignment 4

Course Lecturer: Yizheng Zhao

May 29, 2022

⋆ **This assignment, due on 10th June, contributes to 10% of the total mark of the course.**

Question 1. OWA and CWA

Consider the database instance $\mathcal{D}_{\text{music}}$ given by:

StudioAlbum(Fantasy)   StudioAlbum(The_Eight_Dimensions)   StudioAlbum(Common_Jasmin_Orange)

DebutAlbum(Jay)   LiveAlbum(2004_Incomparable_Concert)

EP(Hidden_Track)   EP(Initial_D)   SoundtrackAlbum(Secret)   CompilationAlbum(Together)

Song(Herbalist_Manual)   Song(Elimination)

Singer(Jay_Chou)   Singer(Eason_Chan)

Composer(Jay_Chou)   Lyricist(Vincent_Fang)   Police(Black_Cat)

hasFriend(Jay_Chou, Vincent_Fang)   hasFriend(Jay_Chou, Will_Liu)

releases(Jay_Chou, Jay)   releases(Jay_Chou, Fantasy)   releases(Jolin_Tsai, Together)

sings(Eason_Chan, Elimination)   sings(Jolin_Tsai, Rewind)   sings(Jay_Chou, Herbalist_Manual)

writesMusicFor(Jay_Chou, Elimination)   writesMusicFor(Jay_Chou, Rewind)   writesMusicFor(Jay_Chou, Herbalist_Manual)   writesMusicFor(Ta-yu_Lo, Pearl_of_the_Orient)

writesLyricsFor(Jay_Chou, Elimination)   writesLyricsFor(Vincent_Fang, Rewind)

writesLyricsFor(Vincent_Fang, Herbalist_Manual)   DancesWith(Will_Liu, Herbalist_Manual)

Consider each of the following Boolean queries F (in DL notation).

(a) Album(Fantasy)

(b) StudioAlbum(The_Eight_Dimensions)

(c) LiveAlbum(Common_Jasmin_Orange)

(d) ¬LiveAlbum(Common_Jasmin_Orange)

(e) ¬EP(secret)

(f) ¬StudioAlbum ⊔ ¬LiveAlbum(2004_Incomparable_Concert)

(g) $\neg$StudioAlbum $\sqcup$ $\neg$LiveAlbum(Eason_Chan)

(h) $\exists$hasFriend.$\top$(Jay_Chou)

(i) $\exists$hasFriend.$\exists$dancesWith.Song(Jay_Chou)

(j) $\exists$hasFriend.Composer(Jay_Chou)

(k) $\exists$hasFriend.{Jay_Chou}(Vincent_Fang)

(l) DebutAlbum(2004_Incomparable_Concert)

(m) Song(Rewind)

(n) Singer(Jay_Chou)

(o) Singer(Jolin_Tsai)

(p) Lyricist(Jay_Chou)

(q) Composer(Jay_Chou)

(r) Composer(Ta-yu_Lo)

(s) Police(Jay_Chou)

(t) Police(Jolin_Tsai)

(u) $\neg$Singer-SongWriter $\sqcup$ $\neg$Police(Vincent_Fang)

(v) $\neg$Singer-SongWriter $\sqcup$ $\neg$Police(Ta-yu_Lo)

(w) Singer-SongWriter(Jay_Chou)

(x) Singer-SongWriter(Jolin_Tsai)

(y) $\neg$SongWriter(Vincent_Fang)

(z) $\neg$Dancer(Will_Liu)

- Write those Boolean queries (marked in red) in first-order logic (FOL) notation. (Note that for many queries there is no difference between DL notation and FOL notation).
- Query answering under closed world assumption: check for each Boolean F whether the answer to the query F given by $\mathcal{D}_{\text{music}}$ is "Yes" or "No".
- Query answering under open world assumption: check for each Boolean query F whether the certain answer to F given by $\mathcal{D}_{\text{music}}$ is "Yes", "No", or "Don't know".

Consider the following non-Boolean queries F_i ($1 \leq i \leq 4$):

(a) $F_1(x) = \text{Singer}(x)$

(b) $F_2(x) = \neg\text{Singer}(x)$

(c) $F_3(x, y) = \text{hasFriend}(x, y)$

(d) $F_4(x) = (\text{Lyricist}(x) \vee \text{Composer}(x)) \wedge \neg\text{releases}(x, \text{Jay})$

For each query F_i, give

- for closed world assumption: answer(F_i,$\mathcal{D}_{\text{music}}$);
- for open world assumption: certanswer(F_i,$\mathcal{D}_{\text{music}}$).

Question 2. Querying with TBox

Following Question 1, consider now the TBox $\mathcal{T}$ given as:

$$\begin{aligned}
\text{StudioAlbum} &\sqsubseteq \text{Album}\\
\text{LiveAlbum} &\sqsubseteq \text{Album}\\
\text{StudioAlbum} \sqcap \text{LiveAlbum} &\sqsubseteq \bot\\
\text{EP} \sqcap \text{LiveAlbum} \sqcap \text{SoundTrackAlbum} &\sqsubseteq \bot\\
\text{DebutAlbum} &\sqsubseteq \text{StudioAlbum}\\
\text{Album} &\equiv \exists \text{hasTrack}.\text{Song}\\
\text{Singer} &\equiv \exists \text{releases}.\text{Album} \sqcap \exists \text{sings}.\text{Song}\\
\exists \text{releases}^{-}.\text{Police} \sqcap \text{Album} &\sqsubseteq \bot\\
\text{Lyricist} &\equiv \exists \text{writesLyricsFor}.\text{Song}\\
\text{Composer} &\sqsubseteq \exists \text{writesMusicFor}.\text{Song}\\
\text{SongWriter} &\sqsubseteq \text{Lyricist} \sqcap \text{Composer}\\
\text{Singer-SongWriter} &\sqsubseteq \text{Singer} \sqcap \text{SongWriter}\\
\exists \text{sings}^{-}.\top &\sqsubseteq \text{Song}\\
\exists \text{writesLyricsFor}^{-}.\top &\sqsubseteq \text{Song}
\end{aligned}$$

- Re-consider the Boolean queries F given in Question 1. Compute the certain answers in the context of $\mathcal{D}_{\text{music}}$, and in the context of $(\mathcal{T}, \mathcal{D}_{\text{music}})$.
- The addition of the TBox $\mathcal{T}$ to the database instance $\mathcal{D}_{\text{music}}$ allows one to draw new conclusions from $\mathcal{D}_{\text{music}}$, and may render some of the data (ABox assertion) α in $\mathcal{D}_{\text{music}}$ redundant, i.e., $(\mathcal{T}, \mathcal{D} \backslash \{\alpha\}) \models \mathcal{D}$. Can you identify all such assertions α?

Question 3. Computing $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ in $\mathcal{EL}$

Consider the $\mathcal{EL}$ TBox $\mathcal{T}$:

$$\begin{aligned}
\text{Guitarist} &\sqsubseteq \exists \text{plays_for}.\text{RockBand}\\
\text{Bassist} &\sqsubseteq \exists \text{plays_for}.\text{RockBand}\\
\text{Drummer} &\sqsubseteq \exists \text{plays_for}.\text{RockBand}\\
\text{RockBand} &\sqsubseteq \exists \text{managed_by}.\text{Manager}\\
\text{Manager} &\sqsubseteq \text{Employee}\\
\text{Manager} &\sqsubseteq \exists \text{managed_by}.\text{Manager}
\end{aligned}$$

and the ABox $\mathcal{A}$:

$$\begin{array}{rl}
\text{Guitarist(John_Lennon)} & \text{Bassist(Paul_McCartney)}\\
\text{Drummer(Ringo_Starr)} & \text{RockBand(Beatles)}\\
\text{managed_by(Beatles, Brian_Epstein)} &
\end{array}$$

- Compute the interpretation $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ as described in the slides.
- For $\mathcal{EL}$ concept queries, we know that $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ gives the answer "Yes" iff $(\mathcal{T},\mathcal{A})$ gives the certain answer "Yes". Check this for the following queries:
  - $\exists$plays_for.RockBand(John_Lennon);
  - $\exists$managed_by.Manager(Paul_McCartney);
  - $\exists$plays_for.$\exists$managed_by.Manager(Ringo_Starr).
- For more complex queries, $\mathcal{I}_{\mathcal{T},\mathcal{A}}$ can give the answer "Yes" even if $(\mathcal{T},\mathcal{A})$ does not give the certain answer "Yes". Check this for:
  - $F(x,y) = \exists z.(\text{plays\_for}(x,z) \wedge \text{plays\_for}(y,z))$.
  - $F = \exists x.\text{managed\_by}(x,x)$.

Question 4. Conjunctive queries over database and interpretation
Consider the following database $\mathcal{D}$ consisting of the following tables:

Person:	
ID	Name
2001	Jay_Chou
2002	Jolin_Tsai
2003	Stefanie_Sun
2004	Ta-yu_Lo

Enrollment:	
StudentID	Since
2002	2020
2003	2021
2004	2020

Attendance:	
StudentID	CourseID
2001	30000160
2002	30000160
2002	30000170
2003	30000180

Course:	
ID	Title
30000160	KR&P
30000180	PR&CV
30000170	NLP

- Define the finite first-order interpretation $\mathcal{I}_{\mathcal{D}}$ corresponding to $\mathcal{D}$.
- Reformulate each of the following SQL queries Q into first-order queries f_Q, and identify which of them are conjunctive queries.
- Answer Q in the context of $\mathcal{D}$ and f_Q in the context of $\mathcal{I}_{\mathcal{D}}$.

(a) SELECT * FROM Person

(b) SELECT Person.Name FROM Person, Attendance, Course
WHERE Person.ID = Attendance.PersonID
AND Course.ID = Attendance.CourseID
AND Course.Title = "KR&P"

(c) SELECT Person.Name FROM Person, Enrollment
WHERE Person.ID = Enrollment.PersonID
AND NOT EXISTS (
SELECT * FROM Attendance
WHERE Person.ID = Attendance.PersonID)

Question 5. Certain answers in different contexts
Consider the following $\mathcal{ALC}$ knowledge base $\mathcal{K} := (\mathcal{T},\mathcal{A})$ with:

$$\mathcal{T} := \{X \sqsubseteq Y, Y \sqsubseteq \exists r.X, X \sqsubseteq \forall r.Y, \forall r.X \sqsubseteq Y, W \equiv \neg V, \exists r.Y \sqsubseteq \neg V\}$$

$$\mathcal{A} := \{(\text{Jay\_Chou}, \text{Jolin\_Tsai}) : r, (\text{Jolin\_Tsai}, \text{Stefanie\_Sun}) : r, (\text{Stefanie\_Sun}, \text{Jay\_Chou}) : r,$$
$$(\text{Jolin\_Tsai}, \text{Jolin\_Tsai}) : r, (\text{Stefanie\_Sun}, \text{Stefanie\_Sun}) : r, \text{Stefanie\_Sun} : X\}$$

- Compute the certain answers to the following conjunctive queries in the context of $\mathcal{A}$.

- Compute the certain answers to the following conjunctive queries in the context of $\mathcal{K}$.

(a) $r(x, y) \land Y(y)$

(b) $\exists y(r(x, y) \land Y(y))$

(c) $\exists x, y(r(x, y) \land r(y, x))$

(d) $\exists z, w(r(x, y) \land r(y, z) \land r(z, x) \land r(z, w) \land W(w))$

Question 6 (with 1 bonus mark). Simpleness of ABox
Consider feeding arbitrary ABoxes rather than simple ABoxes as input to the problem of ontology-mediated querying. Does this affect the data complexity results?

Question 7 (with 1 bonus mark). k-colorability
Is it possible to show that the problem of conjunctive query entailment (CQ-entailment) in $\mathcal{ALC}$ is coNP-hard w.r.t. data complexity using a reduction from non-k-colorability in graphs? What if k is fixed?