Ontológiák és adatbázisok – következtetés nyílt és zárt világokban

Szeredi Péter
szeredi@cs.bme.hu

BME VIK Számítástudományi és Információelméleti Tanszék

2008. március 26.

Bevezető példa: adatbázis

- Adott adatbázisok:

-nek publikációja	
Péter	publ1
Gergő	publ1
Miklós	publ2
Péter	publ2
István	publ3

Cikk
publ2

Könyv
publ1
publ3

- Mi a válasz a következő kérdésekre:
  - Hány cikke van Miklósnak?
  - Ki nem írt cikket?
  - Mik az egyszerzős publikációk?
  - Mik a nem-könyv publikációk?
  - Ki az, aki cikket írt, de könyvet nem?

Bevezető példa: Webes tartalom

- Adottak a következő, a világhálóról származó információk:

-nek publikációja	
Péter	publ1
Gergő	publ1
Miklós	publ2
Péter	publ2
István	publ3

Cikk
publ2

Könyv
publ1
publ3

- Mi a válasz a következő kérdésekre:
  - Hány cikke van Miklósnak?
  - Ki nem írt cikket?
  - Mik az egyszerzős publikációk?
  - Mik a nem-könyv publikációk?
  - Ki az, aki cikket írt, de könyvet nem?

Az adatbázis-lekérdezés szemantikája

- Relációs algebra – *zárt világ*
- Példa: „Ki az, aki cikket írt, de könyvet nem?”
  - Alaphalmaz: Péter, Gergő, Miklós, István, publ1, publ2, publ3
  - Cikkek: publ2
  - Cikket írók: Miklós, Péter
  - Könyvek: publ1, publ3
  - Könyvet írók: Péter, Gergő, István
  - Cikket írók, akik nem írtak könyvet: Miklós

Emlékeztető: az ALC nyelv szemantikája

- Példa: „Ki az, aki cikket írt, de könyvet nem?” $\longrightarrow$ Ki tartozik az alábbi fogalomba:

$$\exists \text{publikációja}.\text{Cikk} \sqcap \neg(\exists \text{publikációja}.\text{Könyv})$$

- Interpretáció: $\mathcal{I} =< \Delta, I >$
  - Δ az objektumok halmaza.
  - Az I függvény az atomi fogalmakhoz és szerepekhez halmazokat ill. relációkat rendel.
- Az összetett fogalomkifejezések szemantikája az $\mathcal{I}$ interpretacióban:

$$\begin{aligned}
\top^{\mathcal{I}} &= \Delta \\
\bot^{\mathcal{I}} &= \emptyset \\
(\neg C)^{\mathcal{I}} &= \Delta \setminus C^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(C \sqcup D)^{\mathcal{I}} &= C^{\mathcal{I}} \cup D^{\mathcal{I}} \\
(\exists R.C)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} | \exists b. \langle\, a, b \,\rangle \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\} \\
(\forall R.C)^{\mathcal{I}} &= \{a \in \Delta | \forall b. (\langle\, a, b \,\rangle \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}})\}
\end{aligned}$$

A Webes lekérdezés szemantikája

- Logikai következményfogalom: nyílt világ
- Példa: „Ki az, aki cikket írt, de könyvet nem?”
  - Számunkra érdekes egyed*nevek*: Péter, Gergő, Miklós, István, publ1, publ2, publ3
  - *Biztosan* cikk: publ2
  - Biztosan cikk-író: Miklós, Péter
  - Biztosan könyv: publ1, publ3
  - Biztosan könyv-író: Péter, Gergő, István
  - Biztosan könyvet nem író: *üres* (nincs ilyen)
  - Biztosan cikk-író, de könyvet nem író: *üres* (nincs ilyen)
- A web nyílt világában az információ nem teljes
- A lekérdezésre csak olyan választ adhatunk, amely a leírt világ *minden kiterjesztésében* igaz.

A Webes lekérdezés modellezése: leíró logikai adatdobozok

- A világban jelenlevő objektumok reprezentálására egy új névfajtát vezetünk be, az *egyedneveket*. jelölésük, a, b, c stb.
- Az adatdoboz (A-doboz) adatállításokat tartalmaz, ezek lehetnek:
  - fogalmi állítások: $C(a)$, pl. cikk(PUBL2),
  - szerepállítások: $R(a, b)$, pl. publikációja(PÉTER, PUBL2).
- Példa:

publikációja(PÉTER, PUBL1)	Cikk(PUBL2)
publikációja(GERGŐ, PUBL1)	
publikációja(MIKLÓS, PUBL2)	Könyv(PUBL1)
publikációja(PÉTER, PUBL2)	Könyv(PUBL3)
publikációja(ISTVÁN, PUBL3)	

- $\mathcal{I}$ interpretációs függvényt ki kell bővíteni: minden a egyednévhez $\mathcal{I}$ hozzárendel egy neki megfelelő $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ elemet
- $\mathcal{I}$ kielégíti a $C(a)$ fogalmi állítást ($\mathcal{I} \models C(a)$) csakkor, ha $a^{\mathcal{I}} \in C^{\mathcal{I}}$,
- $\mathcal{I}$ kielégíti a $R(a, b)$ szerepállítást ($\mathcal{I} \models R(a, b)$) csakkor, ha $\langle a^{\mathcal{I}}, b^{\mathcal{I}} \rangle \in R^{\mathcal{I}}$.

Következtetés A-dobozon

- Definíció: $\mathcal{A} \models \alpha$: Az $\mathcal{A}$ A-dobozból következik az α állítás: ha minden $\mathcal{A}$-t kielégítő interpretáció ($\mathcal{A}$ modellje), biztosan kielégíti α-t.

$$\mathcal{A} \models C(a) \iff \mathcal{A} \cup \{\neg C(a)\} \ \textit{inkonzisztens}$$

- Adatdoboz „lekérdezése”:
  - *Példányvizsgálat* (*instance check*): Igaz-e hogy egy α adatállítás következménye-e egy $\mathcal{A}$ adatdoboznak.

    Példa: igaz-e, hogy $(\neg\mathsf{Könyv})(\mathsf{PUBL2})$. Ha T-dobozunk szerint a könyvek és a cikkek diszjunktak ($\mathsf{Könyv} \sqcap \mathsf{Cikk} \equiv \bot$), akkor a fenti adatdoboznak következménye ez az állítás.
  - *Példánykikeresés* (*instance retrieval*): egy adott C fogalomkifejezéshez meg kell állapítani, hogy mely egyednevek tartoznak *biztosan* az adott fogalomba.

    Példa: mik a példányai az $\exists\mathsf{publikációja}.\mathsf{Cikk} \sqcap \neg(\exists\mathsf{publikációja}.\mathsf{Könyv})$ fogalomnak? A fenti adatdoboz alapján ennek a fogalomnak nincs ismert példánya. Ha az adatdobozhoz hozzávesszük az $(\leqslant 1\, \mathsf{publikációja})(\mathsf{Miklós})$ állítást (Miklósnak legfeljebb 1 publikációja van), és feltételezzük a könyvek és a cikkek diszjunktságát, akkor $\mathsf{Miklós}$-t kapjuk válaszként.

Nyílt és zárt világ

- Zárt világban minden állítás vagy igaz, vagy nem.
- Nyílt világban lehetnek eldöntetlen állítások. Például bizonyos publikációkról tudjuk, hogy cikkek, bizonyosakról tudjuk, hogy nem cikkek, és lehetnek olyanok, amelyekről nem tudjuk, hogy cikkek-e.
- Zárt világban nincs szükség negatív állításokra, például, minden egyed amelyről nincs kimondva, hogy cikk, a ’nem-cikk’ fogalomba tartozik.
- Zárt világban alkalmazható a „meghiúsulásos negáció” (NF, Negation by Failure), azaz egy állítás bizonyításának sikertelensége az állitás tagadását eredményezi
- Nyílt világban alkalmazható az esetszétválasztás, mint következtetési módszer.

Egy klasszikus esetszétválasztós példa: Az Oidipusz család

- Az Oidipusz adatdoboz:

  gyereke(IOKASZTÉ,OIDIPUSZ) gyereke(IOKASZTÉ,POLÜNEIKÉSZ)
  gyereke(OIDIPUSZ,POLÜNEIKÉSZ) gyereke(POLÜNEIKÉSZ,THERSZANDROSZ)
  Apagyilkos(OIDIPUSZ) ¬ Apagyilkos(THERSZANDROSZ)

- Erre az $\mathcal{A}_{OI}$ A-dobozra vonatkozóan az alábbi kérdést szeretnénk feltenni:

  Van-e Iokaszténak olyan gyermeke, aki egyrészt apagyilkos, és akinek másrészt van egy olyan gyermeke, aki nem apagyilkos?

  azaz:

$$\mathcal{A}_{OI} \models (\exists \text{gyereke}.(\text{Apagyilkos} \sqcap \exists \text{gyereke}.\neg \text{Apagyilkos}))(\text{IOKASZTÉ})?$$

- A válasz: igen, de a bizonyításhoz esetszétválasztás szükséges!

Egy saját esetszétválasztós példa: Az alkoholisták

- (T1): A szülői jó példa hatalma: ha valakinek van nem-alkoholista szülője, akkor ő sem alkoholista. ($\exists$szülője.$\neg$Alkoholista $\sqsubseteq \neg$Alkoholista)
- (T2): Az ijesztő rossz példa ereje: ha valakinek van alkoholista barátja, akkor ő nem-alkoholista. ($\exists$barátja.Alkoholista $\sqsubseteq \neg$Alkoholista)
- A fenti T-doboz mellett tekintsük az alábbi A-dobozt:
  - A-nak szülője B: szülője(A, B) (1)
  - A-nak barátja B: barátja(A, B) (2)
- Ezen tudásbázisból következik, hogy

  A nem alkoholista ($\neg$Alkoholista(A))

Saját tapasztalatok

- SINTAGMA Szemantikus Információ-integrációs rendszer:
  - A fogalmi modellezés megvalósítása leíró logikai eszközökkel, *zárt világ* feltételezés mellett
- DLog leíró logikai következtető rendszer
  - *nyílt világ* feltételezés mellett
- Mindkét rendszert Prolog környezetben valósítottuk meg

Irodalom

- The Description Logic Handbook; Theory, Implementation and Applications; Edited by Franz Baader et al.; Cambridge University Press; 2003; ISBN-13: 9780511060632
  - *Daniele Nardi, Ronald J. Brachman:* An Introduction to Description Logics
  - *Franz Baader, Werner Nutt:* Basic Description Logics
  - *Alex Borgida, Maurizio Lenzerini, Riccardo Rosati:* Description Logics for Data Bases
- A szemantikus világháló elmélete és gyakorlata; Szeredi Péter, Lukácsy Gergely, Benkő Tamás; Typotex, 2005;
- Lukácsy Gergely PhD dolgozat: Semantic Technologies using Logic Programming
- Kétnapos műhelymunka, Edinburgh (2006 október): The Closed World of Databases Meets the Open World of the Semantic Web. ESIWiki: `http://wiki.esi.ac.uk/The_Closed_World_of_Databases_Meets_the_Open_World_of_the_Semantic_Web`