

Ontológiai konzisztenciavizsgálat

Követelmény- és funkcionális specifikáció

Az ontológiák értékelését sokféle szempont szerint végezhetjük a technikai, logikai követelményektől kezdve a felhasználói élményekig bezárólag [Gómez-Pérez et al. 2004]. Az alábbi négy szempontból lehet (és szokták) megítélni az ontológiákat életciklusuk valamilyen szakaszában:

- ontológiai érvényességvizsgálat, technikai ontológiaértékelés (ontology evaluation)
az ontológiának milyen technikai, logikai, informatikai szabályoknak, kényszereknek kell megfelelnie, amely két nagyobb tevékenységterületre osztható tovább:
- ontológiaverifikálás (ontology verification)
az ontológia fogalmak meghatározásai korrekt módon, az ontológiai építkezés követelményeinek megfelelően történtek-e?
- ontológiaválidálás (ontology validation)
az ontológia fogalmai, állításai összhangban vannak-e a világnak azzal a szeletével, amelyet az ontológiával modellezni akartak?
 - használati ontológiaértékelés (ontology assessment)
adott használati célra, adott felhasználói tudás, motiváció, felkészültség mellett megfelelő módon használható-e az ontológia?

Az ontológiaépítés szakaszában az ontológiák belső, technikai, logikai minőségét kell elsősorban ellenőriznünk, a külvilágnak, a felhasználóknak való megfelelés ekkor még nem fontos. Az ontológiák technikai-módszertani ellenőrzésekor háromfajta szempontból vizsgálódhatunk:

- konzisztencia (consistency)
az ontológiai fogalmak meghatározásai logikailag érvényesek-e, és nincs-e ellentmondás a rendszeren belül a logikai mondatok (definíciók, illetve axiómák) között?
- teljesség (completeness)
a használati cél szempontjából tekintve minden fogalom definiálva van-e vagy levezethető-e más fogalmakból, és minden meghatározás teljes-e, érvényes-e?
- tömörség (conciseness)
az ontológia redundanciamentes-e, vagyis nincs-e szükségtelen, fölösleges fogalom benne (explicit redundancia), vagy nincs-e olyan fogalom benne, amely levezethető más fogalmakból (implicit redundancia)?

A fenti szempontrendszer sokféle ellenőrzésre ad lehetőséget, melyek közül első körben azt szedjük össze, hogy az ontológiák vázát alkotó taxonómiák ellenőrzésének milyen szempontjai lehetnek. A technikai szempontú érvényességvizsgálat szempontjait most meg kell fordítanunk, hiszen az ellenőrzés során a hibákat kell keresnünk (és megtalálnunk). A három vizsgálati szempont alapján az alábbi három fő *hibát* találhatjuk meg egy ontológiában:

- inkonzisztencia
- nem-teljesség
- redundancia

Ezeket természetesen a hibák típusa alapján tovább bonthatjuk, mielőtt azonban ezt megtennénk, rögzítenünk kell azt, hogy az ontológiáknak milyen elemtípusai lehetnek. Ez a szempont sokat segít majd a hibatípusok elkülönítésében. Egy komoly ontológiai elkötelezettséget rögzítve leszögezhetjük, hogy az ontológiákban az alábbi fogalomtípusok létezését feltételezzük:

- osztályfogalmak
példa: EMBER, KUTYA, HÁZ, FOLYÓ, KAVICS, TOLL
- relációfogalmak
példa: SZERETI, SZIGORÚ RENDEZÉS, APJA, ROKONA
- tulajdonságfogalmak
példa: PIROS, RIGID, TELJES, KÖNNYŰ, HOSSZÚ
- individuumfogalmak
példa: Pisti, Erzsébet-híd, JIX-936 rendszámú BMW 116i gépkocsi 2005-ben

A taxonómiaépítési hibák leírásához szükség van még egy másik fogalomcsoportra. Amikor a klasszifikációs munka során osztályokat hozunk létre, két dimenzió segítségével minősíthetjük ezt a tevékenységet. Mondhatjuk, hogy a létrehozott osztályok az értelmezési tartományra nézve:

- egymástól elkülönülők, diszjunktak,
nem rendelkeznek közös példánnyal
- teljesek, kimerítőek,
az értelmezési tartomány mindegyik eleme beletartozik valamelyik osztályba
- diszjunktak és kimerítőek,
az osztályok nem rendelkeznek közös elemmel és együttesen teljesen lefedik az értelmezési tartomány minden elemét (ekkor azt mondjuk, hogy az osztályok partíciókat képeznek)

A fenti fogalmainkra támaszkodva az ontológiák taxonomikus vázának fő hibatípusait az alábbi módon határozhatjuk meg:

- inkonzisztencia hibák:
 - körkörösségi hiba (tranzitív kör létezése)
pl. az UTAS alá van rendelve az EMBER, az EMBER alá a NŐ, a NŐ alá az UTAS
 - szemantikus inkonzisztencia hiba
pl. a KUTYA a GŐZMOZDONY alá van rendelve
- particionálási hibák:
 - közös osztály diszjunkt taxonómiában,
diszjunkt felosztás (partíció) esetén egy részosztály több osztály alá van besorolva
példa: az AEROPLAN részosztály a VÍZI JÁRMŰ és a LÉGI JÁRMŰ osztályok alá van rendelve
 - közös individuum diszjunkt taxonómiában,
diszjunkt felosztás (vagy partíció) esetén egy individuum több osztály alá van besorolva
példa: Aeroplan típusú, CK-342 azonosítójú gép a VÍZI JÁRMŰ és a LÉGI JÁRMŰ osztályok alá van rendelve
 - besorolatlan individuum kimerítő taxonómiában,
kimerítő felosztás (vagy partíció) esetén egy individuum egyik osztály alá sincs

besorolva

példa: a NEMZETKÖZI JÁRAT és a BELFÖLDI JÁRAT osztályok mellett van egy olyan DBZ-02213 azonosítójú járat, amely egyik osztály alá sincs besorolva

- nem-teljességi hibák:
- nem teljesen lefedő fogalmi klasszifikáció,
nem szerepel az ontológiában a célterület minden fontos fogalma
példa: van VÍZI JÁRMŰ és LÉGI JÁRMŰ osztály, de nincs SZÁRAZFÖLDI JÁRMŰ osztály
- particionálási hibák,
a részosztályok létrehozásakor hiányoznak a particionáláshoz szükséges információk;
a hiba altípusai:
 - hiányzó elkülönülési kényszer
nem adják meg, hogy részosztályok diszjunktak
példa: megadják, hogy Buda és Pest részei Budapestnek, de nem rögzítik azt a tudást, hogy ezek teljesen elkülönülnek egymástól
 - hiányzó lefedési kényszer
nem adják meg, hogy a részosztályok teljesen lefedik az alaposztályt
példa: megadják, hogy Budapest 23 kerületét, de nem rögzítik azt a tényt, hogy ezek együttesen teljesen lefedik Budapestet
- redundancia hibák:
- részosztálya reláció többszörös megadása
két osztályfogalom között – direkt vagy indirekt módon – többször felveszik a részosztálya relációt
példa: a KUTYA [van fajtája] EMLŐS, az EMLŐS [van fajtája] ÁLLAT, a KUTYA [van fajtája] ÁLLAT (indirekt redundancia)
- példánya reláció többszörös megadása
egy individuum és több osztály esetében – direkt vagy indirekt módon – többször felveszik a példánya relációt
példa: BODRI [van példánya] KUTYA, a KUTYA [van fajtája] ÁLLAT, BODRI [van példánya] ÁLLAT (indirekt redundancia)
- azonos formális osztálydefiníció
eltérő nevek alatt teljesen azonos formális definíciója van két osztályfogalomnak
- azonos formális individuumdefiníció
eltérő nevek alatt teljesen azonos formális definíciója van két individuumfogalomnak

relációfogalmak nyelvi problémái

relációmegnevezés

Vannak relációk, melyeket a magyar nyelvben formális eszközökkel (pl. predikátumlogika segítségével) nem lehet pontosan, egyértelműen kifejezni. Példaként¹ vegyük az 'alárendeltje' relációt, melyet jelöljünk önmagával (tehát egyelőre még rövidítés nélkül). Ekkor a relációt kifejező predikátum a következő:

'X alárendeltje Y'

A fenti formális mondatot azonban nem tudjuk egyértelműen feloldani, mert az egyaránt jelentheti az alábbi két természetes nyelvi mondatot:

'X alárendeltje [van] Y' vagyis a relációban X a fölérendelt és Y az alárendelt

'X [van] alárendeltje Y-nak' vagyis a relációban Y az alárendelt és Y a fölérendelt

Természetes nyelven a toldalékok segítségével tehát pontosan kifejezhetjük a reláció pontos jelentését, ám a formális nyelvben ezt nem tudjuk megtenni, mert a magyar nyelv a viszony pontosítására alkalmas nyelvi jeleket a relátumokhoz, nem pedig a relációkifejezéshez köti. Mivel azonban a predikátumlogikai formalizmus alkalmazása megköveteli, hogy a természetes nyelv toldalékait levágjuk, ezért elveszítjük a reláció egyértelmű reprezentációjának lehetőségét. Az angol nyelv ebből a szempontból „szerencsésebb”, hiszen a konkrét példa relációja, a 'is subordinate of' formában egyértelműen kifejezhető.

relációs osztályfogalom vagy relációfogalom

konzisztenciaellenőrzés

relációkonzisztencia-vizsgálatok

dimenzió	tulajdonság	definíció
reflexivitás	reflexív	$\forall x(\mathbf{R}_{xx})$
	irreflexív	$\forall x(\neg \mathbf{R}_{xx})$
	nem-reflexív	$\exists x(\neg \mathbf{R}_{xx})$
szimmetricitás	szimmetrikus	$\forall x \forall y(\mathbf{R}_{xy} \supset \mathbf{R}_{yx})$
	aszimmetrikus	$\forall x \forall y(\mathbf{R}_{xy} \supset \neg \mathbf{R}_{yx})$
	antiszimmetrikus	$\forall x \forall y((\mathbf{R}_{xy} \wedge \mathbf{R}_{yx}) \supset x=y)$
	nem-szimmetrikus	$\exists x \exists y(\mathbf{R}_{xy} \supset \neg \mathbf{R}_{yx})$
transzitivitás	transzítív	$\forall x \forall y \forall z((\mathbf{R}_{xy} \wedge \mathbf{R}_{yz}) \supset \mathbf{R}_{xz})$

¹ Természetesen a 'fölérendeltje' relációval is ugyanez a probléma, de még sokáig sorolhatnánk a példákat, hiszen a 'példánya', az 'előzménye', 'utóda', 'változata' stb. relációk mindegyikével ugyanez a probléma.

	intranszitiv	$\forall x \forall y \forall z ((R_{xy} \wedge R_{yz}) \supset \neg R_{xz})$
	nem-tranzitiv	$\exists x \exists y \exists z ((R_{xy} \wedge R_{yz}) \supset \neg R_{xz})$
összefüggőség	összefüggő	$\forall x \forall y (x \neq y \supset (R_{xy} \vee R_{yx}))$
	erősen összefüggő	$\forall x \forall y (R_{xy} \vee R_{yx})$
	dichotóm	$\forall x \forall y (x \neq y \supset (R_{xy} \wedge \neg R_{yx}) \vee (\neg R_{xy} \wedge R_{yx}))$
	trichotóm	$\forall x \forall y (R_{xy} \wedge \neg R_{yx} \wedge x \neq y) \vee (\neg R_{xy} \wedge R_{yx} \wedge x \neq y) \vee x = y$

szabályok

total connectedness – teljes összefüggőség

Ha R az A halmazon értelmezett $R \subseteq A \times A$ Descartes-szorzat részhalmaza, akkor R reláció, amennyiben A minden elemére teljesül, hogy

$$\forall R \forall x \forall y (x \neq y \supset (R_{xy} \vee R_{yx}))$$

segmented connectedness – szegmentált összefüggőség

Ha R az A halmazon értelmezett $R \subseteq A \times A$ Descartes-szorzat részhalmaza, akkor R reláció, amennyiben A minden elemére teljesül, hogy .

$$\forall R (\forall x \forall y (x \neq y \supset (R_{xy} \vee R_{yx})))$$

no inverse redundancy – inverzredundancia-mentesség

logical statement of inverse relation

$$\forall R (\forall x \forall y (x \neq y \supset (R_{xy} \vee R_{yx})))$$

*Amikor relációra vonatkozó adatokat adatbázisban akarunk tárolni, akkor adatbázis-reprezentáció esetén ki kell zárni az inverzreláció-érték felvételét
inverzreláció létezése esetén csak az egyik irányú relációtípust szabad megengedni*

subordinate filtered – alárendelt szűrt - tranzitív

no transitive loop – tranzitív kör mentes

adott relációtípus esetén nem lehet kör az útvonalakban

OntoClean-módszertan

A fogalmakat – metatulajdonságaik mentén jellemezve – különböző típusokba sorolhatjuk. A metatulajdonságok alapján a fogalmak közti ontológiai kapcsolatokra szabályokat, kényszereket lehet felállítani, melyeket minden változtatás során/után ellenőrizni kell. Ezzel a kérdéssel külön tanulmány foglalkozik [OntoClean].

szófaji függés

Amikor veszek egy fogalomsort leíró szósort, mondjuk a következőt:

futás, repülés, úszás, mozgás

akkor a hagyományos módon úgy járhatunk el, hogy a fogalmakat reprezentáló szavak között felírunk relációkat, melyek voltaképp nem a szavak, hanem a fogalmak között érvényesek. Jelen példánkban a generikus alárendeltséget rögzíthetjük:

futás generikus fölérendeltje mozgás
repülés generikus fölérendeltje mozgás
úszás generikus fölérendeltje mozgás

Ha ezek után a vizsgált fogalmakhoz köthető más szófajú szavakat vesszük elő, mint például:

fut, repül, úszik, mozog
futni, repülni, úszni, mozogni
futó, repülő, úszó, mozgó
futandó, repülendő, úszandó, mozgandó

akkor nyilván ugyanazt a szemantikai viszonyrendszert állapíthatjuk meg a megfelelő szavak (pontosabban az általuk reprezentált fogalmak) között, például:

fut generikus fölérendeltje mozog
repül generikus fölérendeltje mozog
úszik generikus fölérendeltje mozog

Ebből következően az egy fogalomhoz tartozó, de különböző szófajú csoportokat úgy érdemes kezelni, hogy a fogalmi szinten megállapítjuk a szemantikai relációkat, míg a nyelvi szinten összekötjük az egy fogalomhoz tartozó szavakat, melyek mindegyik elemére származtathatjuk a fogalom ontológiai kapcsolatait.

{futás, fut, futni, futó, futandó} generikus fölérendeltje {mozgás, mozog, mozogni, mozgó, mozgandó}

Ennek megoldásához szükségünk van a különböző szófajú szavak fogalmakon keresztüli összekapcsolására, illetve ezek konzisztenciájának folyamatos ellenőrzésére.

a cselekvés maga	ige, főnévi igenév	cselekszik, fut, tanít, von, olvas
		cselekedni, futni, tanítani, olvasni, vonni
a cselekvés folyamata	főnév	cselekvés, cselekmény, tanítás, futás, futam, olvasás, olvasat,
a cselekvés ágense	folyamatos melléknévi igenév	cselekvő, futó, vonó, tanító
a cselekvés eredménye	befejezett melléknévi igenév	futott, tanított, olvasott, vonat, tapasztalat, cselekedet, tapasztalt
a cselekedet intenciója, szándéka, kötelezettsége, modalitása	beálló melléknévi igenév	futandó, tanítandó, olvasandó
a cselekedet lehetősége, megengedettsége, modalitása (?)		olvasható, futható, tanítható

tezauruszformátum-ellenőrzés

Az ontológiai konzisztencia vizsgálata során értelmes lehet olyan ellenőrző mechanizmus implementálása, amely egy tezaurusz-kimenetet feltételezve végez el ennek megfelelő

ellenőrzéseket. Ekkor az alábbi feladatokat kell megvalósítani (Ungváry Rudolf egy korábbi írása alapján):

a probléma típusa		
hiányzó vezérszó	kapcsolódó szó vezérszóként nem szerepel	
inverzitás	Az összefüggések fordított (inverz) irányban fennállnak-e?	
ellentmondó kapcsolatok	Nemdeszkriptor nemcsak az L, vagy az L& vagy az LV relációban kapcsolódik deszkriptorhoz.	
reflexív kapcsolat	Egy szó önmagával van kapcsolatban	
ismétlődő kapcsolat	Ugyanazon két szó között ugyanaz vagy más reláció többször szerepel	
A megengedett számú kapcsolatoknál több szerepel	Ha egy kapcsolattípuson belül a megengedett relációelőfordulások számánál van több	Alsó = az adott relációban minimálisan kapcsolódó szavak száma Felső = az adott relációban maximálisan kapcsolódó szavak száma
	Ha más kapcsolattípusokban van a megengedett relációelőfordulások számánál több	Malsó = a más relációban minimálisan kapcsolódó szavak száma Mfelső = a más relációban maximálisan kapcsolódó szavak száma
Tranzitív hiba vagy kör van a tranzitív kapcsolatok láncában	Hiba, ha kifejezés tranzitív kapcsolata közvetlen kapcsolatként is előfordul. Például közvetlen kapcsolat áll fenn egy kifejezés fölérendeltjével és a fölérendelt fölérendeltjével. Ugyancsak hiba, ha a kifejezés tranzitív kapcsolatokon keresztül önmagával kerül újra kapcsolatban (azaz zárt kör keletkezik).	
Elütések ellenőrzése	A program bizonyos határig ellenőrizze a hibásan írt szavakat, az ún. elütéseket és a formálisan közel álló kifejezéseket (pl. Vásárlás - Vásárlás, Oktatási intézet - Oktatási intézmény).	
Hiányzó kapcsolatok, Hiányzó csúcsgalom (osztály)	A program mutassa ki, hogy mely deszkriptorok nem kapcsolódnak csúcsgalomhoz (osztályhoz)	
Hiányzó kapcsolatok, Kapcsolatok nélküli lexikai egység	A program mutassa ki azokat a lexikai egységeket, melyeknek semmiféle kapcsolata nincsen.	

A tezaurusz kezeléshez szükség van az ontológiai egységek tipizálására a deszkriptor-nem deszkriptor dimenzióban.

rendszer tulajdonságok

relata table

leftRelatums_relation_rightRelatums

hány relációtípus van a rendszerben

milyen tulajdonságai vannak a relációtípusoknak

hány relátum (elem) van a rendszerben

hány relátum van az egyes relációtípusok mentén

hány csúcsrelátum van relációtípusonként

a csúcsrelátumok listája relációtípusonként

a relátumok befoka / indegree

a relátumok kifoka / outdegree

a relátumok rangja / rank

hány hierarchia van a rendszerben

hány összefüggő részgráf van a rendszerben, typed Forrest

hány végrelátum van relációtípusonként

fokszám: az élek száma – hány csúcsra van közvetlen hatása

fokszámeloszlás: n-fokú csúcsok statisztikája

távolság: hány élen keresztül lehet legkevesebb lépésben eljutni egyik csúcsból a másikba

közelségi központosság: mennyire hatékonyan tud információt terjeszteni

közöttségi központosság: mennyire tudja az információ terjedését befolyásolni

egyéb

MEO nyulványok (szín, emberszerep, állat, növény, hely, idő, rokonsági)

relációk

fedés, közvetlenül megelőzi

hierarchikus struktúra

generikus reláció

szófüggvények

struktúra/reláció függvények

közvetlenül megelőzi (közvetlen megelőzője/immediate predecessor) reláció (pontosabban függvény)

apja, közvetlen felettese, szülője, fedője

inverze: közvetlen követője/immediate successor (gyereke, közvetlen leszármazottja, közvetlen beosztottja)

megelőzi reláció

relációstruktúra

közvetlen megelőzi lánc, vérvonal

szubreláció/subrelation

van átfedés a különböző nyelvű (nyelvfüggő) szeménék között:

magyarban nincs apai nagybácsi, más nyelvekben van ilyen fogalom

egyéb

autoritás

mikroelméletbe tartozás

nézet

workflow

metatípus: exoskeletális metaelem vs. alapelem

prototipikus bővíthetőség

Hivatkozások

[Gómez-Pérez et al. 2004]

Asunción Gómez-Pérez, Mariano Fernández-López, Oscar Corcho, *Ontological Engineering*, Springer, 2004

[OntoClean]

Szöts Miklós, *Az OntoClean metodológia ismertetése, problémái és továbbfejlesztési lehetőségei*

<http://ontologia.hu/Members/szots/OntoClean.pdf>