

RECUPERAREA CONCEPTULUI DE ONTOLOGIE ÎN CONTEXTUL WEB-ULUI SEMANTIC

Alexandru Ceașu

aceausu@racai.ro

Institutul de Cercetări pentru Inteligență Artificială,
Academia Română

Rezumat: Perspectivile în continuă schimbare ale web-ului au dus la adoptarea conceptului filosofic de ontologie în cadrul web-ului semantic. Ontologia, una din favoritele gândirii filosofice occidentale, în contextul web-ului semantic este subiectul unor considerabile transformări. Articolul trasează câteva din coordonatele evoluției conceptului de ontologie în gândirea filosofică și transformarea ontologiei în cadrul web-ului semantic.

Cuvinte cheie: ontologie, web semantic, OWL, OIL, RDF

Abstract: The continuously changing web perspectives lead to the usage of the concept of ontology to explain a possible meaningful web: the semantic web. Ontology, one of the favorites of the Occidental philosophical toughs, suffered a series of transformations. The paper covers some of these transformations and highlights a few key events in the evolution of the philosophical concept of ontology.

Keywords: ontology, semantic web, OWL, OIL, RDF.

Perspectivile în continuă schimbare ale web-ului au dus la adoptarea conceptului filosofic de ontologie în cadrul web-ului semantic. Ontologia, una din favoritele gândirii filosofice occidentale, în contextul web-ului semantic este subiectul unor considerabile transformări. Articolul trasează câteva din coordonatele evoluției conceptului de ontologie în gândirea filosofică și transformarea ontologiei în cadrul web-ului semantic.

Tehnologiile promovate în cadrul web-ului semantic au ca scop „eliberarea” datelor de sub regimul unor formate neextensibile, opace. Limbajele de adnotare ale web-ului semantic sunt fundamente ale unei sintaxe cu ajutorul căreia pot fi rezolvate probleme complexe cum sunt reprezentarea înțelesurilor și a relațiilor dintre ele. Perspectivile web-ului sunt în continuă schimbare oferind noi posibilități de comunicare și interacțiune dispuse oportunității și fortuitului unei descentralizări a informației desfășurată la o scară planetară. În acest context, web-ul semantic consacră un concept mai vechi – de această dată într-un cu totul alt domeniu decât ce al filosofiei – conceptul de ontologie.

Cum putem explica reintroducerea pe coordonatele tehnologiei informației a conceptului de ontologie de către „inginerii cunoașterii” (*knowledge engineers*) și de către „experții de domeniu” (*domain experts*)? Acesta este întrebarea la care cele două secțiuni de mai jos încearcă să dea un răspuns, din perspectiva evoluției conceptului de ontologie în gândirea filosofică și din perspectiva dezvoltării ontologiei în cadrul web-ului semantic.

Ontologia – concept filosofic

Până în secolul al XIX-lea, ontologia ocupa un loc privilegiat în taxonomia științelor filosofice. În concepția modernă, exista ontologia generală (care studia ființa în genere) și alte trei ontologii speciale, responsabile de regiunile ființei: teologia, cosmologia și egologia (sau psihologia). Ultimele trei depindeau în mod direct de ontologia generală, știință considerată „exactă” în acea vreme și, mai ales, exhaustivă (reclamând acoperirea tuturor departamentelor existenței, fără rest). Putem identifica, astfel, ontologia „ideală” pl tonică, ontologiile „categoriale” (cu debutul istoric în doctrinele elementelor oferite de pre-socratici și terminând cu tabelele categoriilor logice oferite de Hegel și Kant), ontologia teologică (autorii scolastici).

În secolul al XIX-lea și în secolul al XX-lea apar în filosofie o serie de ontologii total diferite care renunță la pretenția de exhaustivitate, acceptând faptul că o construcție teoretică, fie ea cât de generală, nu poate cuprinde decât o anumită parte a existenței. În special Nietzsche deschide drumul acesta, prin atacul asupra „idolilor” și „răsturnarea tuturor valorilor”, fiind urmat, însă, de Husserl și Heidegger (ontologii fenomenologice), Gadamer (ontologie hermeneutică), Foucault (arheologie).

Figură marcantă pentru spargerea monopolului ontologiei generale este autorul german Friedrich Nietzsche. În primul rând, el consideră că omul nu poate înțelege decât ceea ce el însuși a creat [1]. Ca atare, sunt eliminate încă de la început ontologiile generale care impun omului o ordine „superioară”, exterioară lui – indiferent că este vorba despre o ordine divină (teologică), logică (sisteme categoriale) sau cosmologică (astrologia Evului Mediu). Nu avem instrumentele și nici capacitatea intelectuală de a surprinde ceva aflat în afara relației cu ordinea umană. În concluzie, ontologia – sau ordinea lumii – nu

este descoperită, ci creată. Mai mult, această creație nu este, ca la vechii greci, o „cunoaștere de dragul cunoașterii“, o „contemplație“, ci o activitate cu scop precis, o activitate utilă, programatică prin excelență – „creăm teorii pentru a pune stăpânire pe lucruri“ [1]. Ca atare, noi nu ne raportăm la lucruri, ci la lucruri așa cum apar ele în teoriile construite de noi: din nou, „realitatea“ nu este descoperită, ci „făcută“, „creată“. Noi creăm ceea ce numim „realitate“ prin intermediul unor „categorii“ – dar acestea nu sunt preluate din exterior ci inventate. Categoriile sunt, prin urmare, construcții fictive, imaginare, „falsificări“ care au un scop util, programatic, fără nici o urmă de intenții transcendente, așa cum spera Kant. Categoriile sunt, în mod evident, valori și nu „concepte științifice“, ele apar ca urmare a modului în care noi interpretăm ceea ce se află în jurul nostru. Iar interpretarea se sprijină pe intențiile, nevoile, idiosincraziile, nevrozele noastre, ea este conform fiziologiei noastre. O realitate independentă de ele nu există, în-sine. De la Nietzsche nu mai putem vorbi de sisteme categoriale – ci, în cel mai bun caz de taxonomii valorice: „calitățile reprezintă veritabila noastră idiosincrasie umană: a pretinde că aceste interpretări și valori umane ce ne revin să fie valori universale, și poate constitutive, aparține smintelii ereditare a vanității umane“ [1].

Întrucât avem de-a face cu taxonomii valorice, iar valorile nu pot fi universale, orice cunoaștere nu este decât o perspectivă: cunoaștem ceva privind dintr-un anumit punct, și vedem lucrurile așa cum se văd ele numai din acel punct. Punctul respectiv este dat de ceea ce ne constituie: valori, sentimente, simțuri. Desigur că există o infinitate de puncte și ca atare o infinitate de perspective. Sau, altfel spus, o infinitate de taxonomii, de ordini ale lumii, de ontologii. Nici una dintre ele nu este „adevărată“ în sensul modern al cuvântului, de corespondență cu realitatea. Ele sunt mai degrabă numite „utile“ decât „adevărate“. În orice caz, așa cum vor continua pragmatistii americani, orice idee, orice ordine a lumii, orice „ontologie“ este „adevărată“ în măsură în care ne este utilă. Astfel, dacă pentru un individ perspectiva teologică este foarte importantă și utilă pentru viața sa, atunci această perspectivă este categoric adevărată – la fel pentru perspectiva științifică, morală etc.

Variațiuni pe această temă întâlnim și la Husserl, Heidegger, Gadamer. Autorii postmoderni, în special Derrida, merg chiar mai departe: ei renunță la termenul ontologie pentru a prefera terminologia „jocurilor de limbaj“.

Finitudinea individului și a discursului său va eșua întotdeauna în fața diversității și a „bogăției infinite“ pe care încearcă să o cuprindă. În al doilea rând – spune Derrida – „non-totalizarea mai poate fi interpretată și altfel: nu prin conceptul de finitudine, în sensul de condamnare la empiricitate, ci prin acela de joc. Dacă, atunci, totalizarea nu are sens nu e pentru că infinitatea unui câmp n-ar putea fi cuprinsă de o privire ori de un discurs finite, ci pentru că natura câmpului – altfel spus limbajul, și încă unul finit – exclude totalizarea: este, într-adevăr, câmpul unui joc, adică al unor substituiri nesfârșite în limitele unei mulțimi finite“ [2]. Nu mai avem de a face, ca în ipoteza clasică, de un câmp infinit: dimpotrivă câmpul este finit, dar îi lipsește tocmai ceea ce ar fi putut opri substituirile, și anume, centrul. Nu ne mai putem raporta la o „structură centrată“, cu un „joc întemeiat“, construit pe un fundament ce se sustrage jocului. Dimpotrivă, în lipsa centrului „absența unui semnificat transcendent extinde la infinit câmpul și jocul semnificației“.

Ontologia web-ului semantic

Pe coordonatele tehnologiei informației, o ontologie este definită ca o colecție de termeni folosiți pentru a descrie un domeniu specific, sau o reprezentare a unei conceptualizări comune a unui domeniu. Unele ontologii sunt largi, acoperind un mare număr de topici, altele sunt limitate la specificații foarte precise cu privire la un domeniu. O ontologie tipică conține o descriere ierarhică a conceptelor importante pentru un domeniu și descrie proprietățile esențiale ale unui concept printr-un mecanism de tipul atribut-valoare. Pentru precizarea relațiilor dintre concepte este folosită logica propozițională de ordinul întâi. De asemenea, o altă caracteristică a ontologiilor astfel concepute, o constituie posibilitatea ca indivizilor aparținând domeniului de interes să le fie atribuit un anumit concept din ierarhie pentru a le delimita tipul.

Ca atare, atomii unei ontologii sunt: (i) clasele; (ii) proprietățile și (iii) relațiile. De asemenea, o ontologie trebuie să aibă în vedere posibilitatea de a formula raționamente folosind primitivele menționate.

Actuala direcție de cercetare, preocupată de crearea ontologiilor, unifică cinci aspecte importante semnalate de comunități diferite: (i) o semantică formală și un suport de inferență eficient, oferit de logica descriptivă; (ii) puterice primitive de modelare epistemologică oferite de cercetarea orientată pe cadre; (iii) propuneri de standarde pentru notații sintactice oferite de comunitatea web; (iv) intuitivitatea reprezentării grafice a cunoașterii inspirată de cercetările privind rețelele semantice; (v) delimitarea modului în care sunt stocate datele de cel în care sunt vizualizate, distincție făcută, cu precădere, în studiile privind bazele de date relaționale.

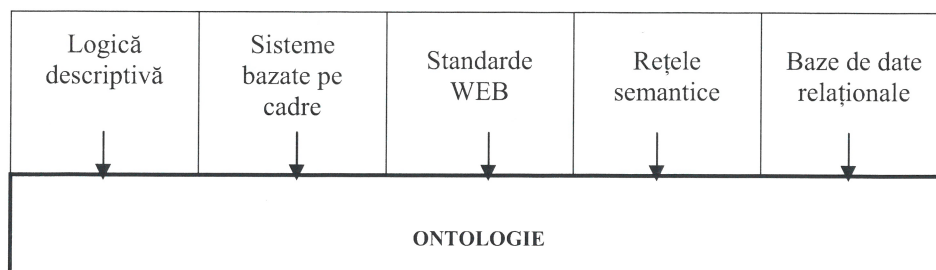


Figura 1. Filiația ideatică a noilor „ontologii“

Logica descriptivă descrie cunoașterea în termeni de concepte și de restricții a rolurilor, cele din urmă fiind folosite pentru a deriva automat taxonomii. Direcția principală de cercetare în reprezentarea cunoașterii este preocupată de a oferi teorii și sisteme pentru exprimarea unei cunoașteri structurate, pentru accesarea ei și pentru producerea de raționamente conduse de reguli. Logica descriptivă, cunoscută ca și logică terminologică, este rezultatul cercetărilor din domeniul rețelilor semantice și din preocuparea de a le conferi acestora o semantică formală și operațională. Logica descriptivă încearcă să identifice un fragment din calculul cu predicate de ordinul întâi. Acesta trebuie să posede o mare expresivitate, în același timp, păstrând eficiența și decidabilitatea unei eventuale proceduri de inferență. O caracteristică a logicii descriptive este definirea claselor (numite concepte) intensional în termenii descrierii proprietăților pe care obiectele trebuie să le satisfacă pentru a aparține conceptului. Aceste descrieri sunt exprimate folosind un limbaj ce permite construirea descrierilor compuse incluzând restricții peste relațiile binare între obiecte (numite roluri). Numeroase studii examinează extensia puterii expresive a unor asemenea limbaje și compromisul privind complexitatea computațională folosit pentru a deriva relațiile gen-specie (de asemenea problematică, dar mai puțin luată în considerare, este și derivarea relațiilor dintre indivizi și clase/concepte). În ciuda rezultatelor teoretice descurajatoare, există câteva implementări eficiente pentru logica descriptivă. Printre acestea se numără și limbajul OWL (Web Ontology Language).

Sisteme bazate pe cadre. Cea mai importantă primitivă a logicii cu predicate de ordinul întâi este predicatul. Sistemele bazate pe cadre, respectiv cele orientate pe obiecte, au o abordare diferită. Ele folosesc ca primitivă centrală clasa (respectiv, cadrul) caracterizată prin anumite proprietăți numite atribute. Aceste atribute nu sunt calificate global ci sunt aplicabile doar claselor pentru care au fost definite, iar același simbol poate fi definit ca atribut cu diferite tipuri de valori pentru clase diferite. Astfel, cu ajutorul cadrelor pot fi modelate diferite aspecte ale aceluiași domeniu. Multe alte rafinamente ale acestei paradigme de modelare au fost dezvoltate, contribuind la succesul de care se bucură acestea în ultimul timp. Astfel, au fost construite diferite sisteme bazate pe cadre și limbaje de programare orientate obiect, având o mare influență asupra comunității ingineriei software.

Rețele semantice. Structura unui limbaj de reprezentare a cunoașterii depinde în mod decisiv de scopul final pe care și-l propune. În cazul rețelilor semantice, scopul este formarea unui sistem logic care să poată exprima conținutul unor propoziții ale limbajului natural în cel mai simplu și direct mod posibil. Rețelele semantice sunt un sistem logic de reprezentare a semanticii limbajului natural. Spre deosebire de calculul cu predicate, conceput în principal pentru a servi studiilor de fundamentare a matematicii, rețelele semantice au fost concepute pentru a simplifica formalizarea unor aspecte ale limbajului natural. Ele sunt bazate pe o notație grafică anunțată, pentru prima oară, prin scrierile filosofului și logicianului Charles Sanders Peirce [3]. Acesta observă că relația specială pe care o împărtășește pisica cu rogojina este mai sugestivă dacă se folosește o expresie de genul pisică – pe – rogojină, decât dacă ar fi redată printr-o formulă a calculului cu predicate: $(\forall x)(\exists y)(pisică(x) \wedge rogojină(y) \wedge pe(x, y))$.

Baze de date relaționale. Pentru astfel de sisteme, datele au caracter empiric, fiind culese pe bază de observații și măsurători. Informația este rezultatul interpretării datelor de către un anumit subiect și conferă acestuia capacitatea de a lua decizii. În concepția și organizarea oricărei baze de date trebuie să se aibă în vedere independența datelor. Aceasta înseamnă că există o delimitare netă între reprezentarea fizică a datelor și imaginea pe care o are utilizatorul asupra acestor date. Modul concret în care este realizată memorarea și organizarea datelor este transparent pentru utilizator, deoarece acesta trebuie să fie preocupat doar de problema concretă pe care o are de rezolvat.

Standarde World Wide Web Consortium (W3C). Organizația menționată este formată din reprezentanți ai institutelor de cercetare și ai firmelor comerciale interesate de dezvoltarea/exploatarea rețelei globale de calculatoare. Standardele emise de o astfel de instituție nu au caracter imperativ, ele sunt recomandări, fiind publicate sub numele de specificații.

Iată, în viziunea consorțiului menționat, cum ar trebui să arate construcția unei ontologii folosind limbajele deja dezvoltate. De asemenea, în această figură sunt intuite și direcțiile ulterioare de cercetare, rămânând, încă, drept necunoscute procedurile de inferență ce vor trebui implicate în construirea unei ontologii.

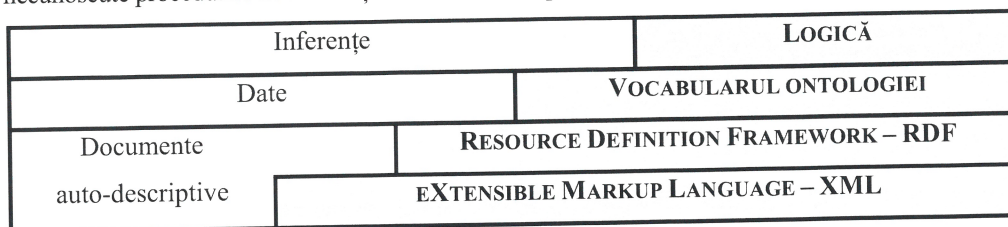


Figura 2. Construcția supraetajată limbajelor

Standardul XML (eXtensible Markup Language), apărut ca specificație în 1999, este un compromis între puterea de expresie a textului scris și portabilitatea / eficiența prelucrării documentelor codate în format binar. Fiind preocupat numai de aspectul sintactic al unui document, standardul XML cu greu poate fi denumit limbaj.

Limbajul „Resource Definition Framework” (RDF) [4,5] operează cu trei tipuri de entități: (i) resurse (care pot fi privite drept constante), (ii) proprietăți (aspecte specifice, caracteristice, atribute sau relații folosite pentru descrierea unei resurse) și (iii) enunțuri. Acestea din urmă sunt formate dintr-un predicat căruia îi sunt atribuite un subiect și un obiect. Enunțurile sunt folosite pentru a descrie proprietățile resurselor. Astfel, elementul de bază al limbajului RDF este tripletul format dintr-o resursă (subiectul) pus în legătură cu o altă resursă (obiectul) de o relație (predicat).

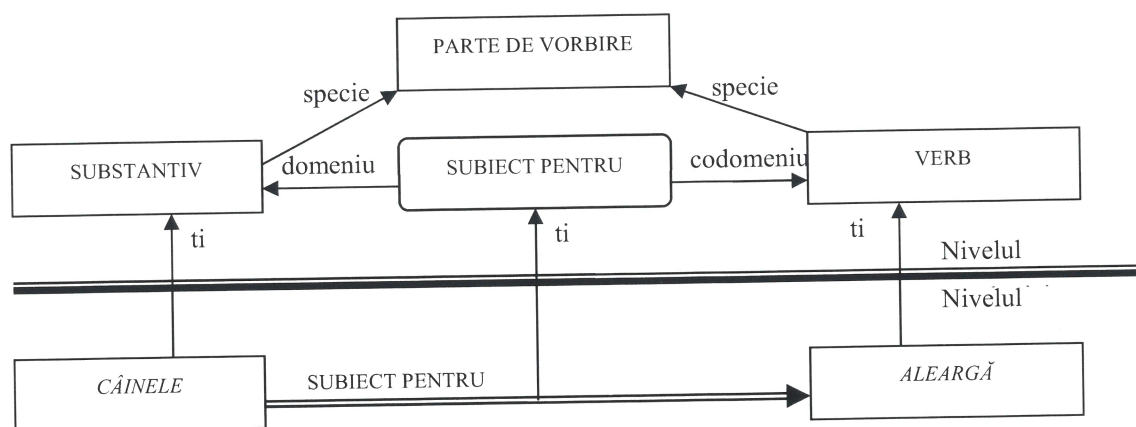


Figura 3. Exemplu de formalizare a propoziției „Câinele aleargă” folosind limbajul Resource Definition Framework

După cum se observă din figură, resursele și proprietățile sunt construite separat, diferit față de sistemele bazate pe cadre, unde proprietățile aparțin claselor (respectiv, resurselor). De asemenea, foarte importantă este delimitarea clară între instanțele unor clase și relații, aparținând nivelului datelor, și nivelul ontologiei.

Pentru formalizarea vocabularului ontologiei este, în prezent, preferat limbajul Web Ontology Language (OWL) [6], [7], devenit recomandare a World Wide Web Consortium în 2003.

De ce nu a fost folosită logica cu predicate de ordinul întâi? Un posibil răspuns constă în imposibilitatea acestui tip de logică de a reține structura cunoașterii, respectiv a informației. De asemenea, logica cu predicate de ordinul întâi posedă o expresivitate prea mare pentru a putea fi viabilă computațional sau pentru a putea oferi proceduri eficiente de inferență.

Atomii limbajului OWL pot fi clasificați în (i) categorii fundamentale (respectiv categorii de bază, categoreme) și (ii) categorii derivate (sau sincategoreme).

Limbajul OWL se caracterizează prin categoria logică de bază numită clasă (sau concept dacă rămânem fideli logicii descriptive). În cadrul limbajului, construcția unor clase noi poate fi realizată fie prin definirea sa intensională (restriction) – prin enumerarea notelor sale, fie extensional (oneOf) – prin enumerarea indivizilor aparținând clasei, fie prin aplicarea unor operatori stricto sensu, cum sunt intersecția (intersectionOf), reuniunea (unionOf), diferența (complementOf), echivalența (equivalentClass) și contrarietatea (disjointWith).

Limbajul permite, de asemenea, calificarea instanțelor cu ajutorul operatorilor egalitate (sameIndividualAs) și incompatibilitate (differentFrom).

Tot o categorie derivată, foarte importantă pentru limbajul în discuție, este proprietatea. Aceasta are statutul de operator logic stricto sensu, respectiv de sincategoremă endogenă prin care, din una sau mai multe clase figurând ca argumente, se construiesc clase noi (sau definite clase noi).

O caracteristică a limbajului OWL constă în bogăția categoriilor logice derivate create de operatori stricto sensu aplicați sincategoremei proprietate. Faptul se datorează preferinței pentru construcția intensională a claselor. Acest tip de operatori pot fi grupați în trei grupe: (i) cei folosiți pentru descrierea naturii unei proprietăți, (ii) cei implicați în cuantificarea codomeniului proprietății și (iii) cei privind cardinalitatea acesteia.

Limbajul OWL folosește numai două tipuri de cuantificare: cea universală (allValuesFrom) și cea existențială (someValuesFrom). Tot în cadrul acestei grupe a sincategoremelor create de operatori stricto sensu implicați în cuantificarea codomeniului proprietății, introducem și un operator de tip special, prin care codomeniul este redus la o constantă (respectiv la o instanță a unei clase) – hasValue.

Cardinalitatea este un concept consacrat odată cu demersul pentru o logică simbolică a lui Chomsky. Prin cardinalitate (cardinality) poate fi specificat numărul de apariții a unei proprietăți. De asemenea, limbajul permite și menționarea numărului minim (minCardinality) și maxim (maxCardinality) de apariții al unei proprietăți.

În figura 4, sunt structurate cele trei posibilități de construcție a unei clase în limbajul OWL.

RDF:CLASS		
subClassOf: <i>clasă</i>	OWL:CLASS	
	intersectionOf: <i>listă</i> unionOf: <i>listă</i> complementOf: <i>clasă</i> oneOf: <i>listă</i> equivalentClass: <i>clasă</i> disjointWith: <i>clasă</i>	OWL:RESTRICTION onProperty: <i>proprietate</i> allValuesFrom: <i>clasă</i> hasValue: <i>instanță</i> someValuesFrom: <i>clasă</i> cardinality: <i>nr. natural pozitiv</i> minCardinality: <i>nr. natural pozitiv</i> maxCardinality: <i>nr. natural pozitiv</i>

Figura 4. Modalități diferite de construcție a claselor descrise folosind o ierarhie gen-specie

După natura lor, proprietățile se caracterizează ca: (i) simetrice; (ii) tranzitive; (iii) funcționale; (iv) inverse; (v) invers-funcționale.

După compromisul pe care îl face între puterea expresivă și complexitatea computațională, limbajul OWL are trei niveluri (complet, aliniat logicii descriptive și redus). În varianta sa completă este foarte dificil de implementat o procedură eficientă de inferență, de aceea, operând câteva restricții (cea mai importantă fiind interzicerea folosirii metaclaselor) poate fi aliniat eficienței computaționale a logicii descriptive. Varianta redusă a limbajului, numită „OWL Lite”, a fost concepută având ca obiectiv optimizarea procesului de învățare mai degrabă decât complexitatea computațională sau puterea expresivă.

Perspective

Limbajele web-ului semantic sunt un compromis între comunități științifice care foloseau diferite limbaje de reprezentare a ontologiilor. De asemenea, sunt și un compromis între eleganța teoretică și posibilitatea de învățare a limbajelor web-ului semantic de către indivizi cu pregătire medie. Astfel, OWL (cel mai folosit limbaj al web-ului semantic) este departe de a constitui un limbaj logic „reusit” din punctul de vedere al unui logician. Încă nu are o capacitate inferențială (asupra acestui aspect nu s-a ajuns încă la un consens). De asemenea, nu distinge, încă, decât între două ipostaze ale unui discurs: denotatul și referința – propunerea pe care și-o face, aceea de a se putea ridica la un nivel intensional, nefiind realizată.

Necesitatea de a construi o rețea globală orientată, de data aceasta, înspre reprezentarea unor înțelesuri mai mult decât înspre afișarea unor date, impune dezvoltarea unui limbaj de reprezentare comun. Astfel, OWL are toate șansele de a deveni acel limbaj comun, iar după previziunile W3C – previziuni care s-au

dovedit de cele mai multe ori adevărate, în parte și datorită posibilităților sale de a le „face“ adevărate – vom asista la construirea unei rețele globale semantice în numai 3-4 ani.

Bibliografie

1. **NIETZSCHE, F.:** Voința de putere, traducere din limba germană de Claudiu Baci, Editura „Aion“, 1999.
2. **DERRIDA, J.:** Scriitura și diferența, Ed. Univers (colecția „Studii“), București, 1998.
3. **SANDERS, P. C.:** Manuscripts on existential graphs, în Collected Papers of Charles Sanders Peirce, A. W. Burks (ed.), vol. 4, Harvard University Press, Massachusetts, pp. 320-410.
4. **BRICKLEY, D., R.V. Guha** (eds.), RDF Vocabulary Description Language 1.0: RDF Schema, 2003, format electronic (<http://www.w3.org/TR/rdf-schema/>)
5. **KLYNE, G., J. J. CARROLL** (eds.) Resource Description Framework (RDF): Concepts and Abstract Syntax, 2003, format electronic (<http://www.w3.org/TR/rdf-concepts/>).
6. **DEAN, M., G. SCHREIBER** (eds.), OWL Web Ontology Language Reference, 2003, format electronic (<http://www.w3.org/TR/owl-ref/>).
7. **PATEL-SCHNEIDER, P. F., P. HAYES, I. HORROCKS:** OWL Web Ontology Language Semantics and Abstract Syntax, 2003, format electronic (<http://www.w3.org/TR/owl-semantics/>).