

Considerations Regarding the Use of Semantic Web Technologies in the Context of E-business Applications

Dr. Sabin-Corneliu BURAGA

Faculty of Computer Science, "A.I.Cuza" University of Iași, Romania

busaco@infoiasi.ro, <http://www.infoiasi.ro/~busaco/>

The paper will present main considerations regarding the use of semantic Web technologies (especially, Web ontologies) for designing and implementing e-business applications.

Keywords: *Semantic Web, Ontologies, E-business, Web Applications.*

1 Introducere

Spațiul World-Wide Web actualmente este compus din pagini (documente) marcate în limbaje precum *HTML – HyperText Markup Language*, *WML – Wireless Markup Language*, *SVG – Scalable Vector Graphics* – detalii în (Buraga, 2003; Buraga, 2004b; W3C, 2005) – conținând informații textuale, grafice și/sau multimedia destinate a fi citite și înțelese de către consumatorii umani. Principala activitate a calculatoarelor este cea de a reda aceste conținuturi și nu să le interpreteze într-o manieră automată și autonomă. Informațiile trebuie regăsite într-un mod inteligent și să poată fi procesate de către mașină. Conform creatorului spațiului WWW, Sir Tim Berners-Lee, *Web-ul semantic* reprezintă o pânză consistentă și logică a tuturor resurselor de pe Web, accentul punându-se pe interpretarea datelor de către mașină și nu pe reprezentarea lor (Berners-Lee *et al.*, 2001). Acest articol va pune în discuție o serie de probleme legate de utilizarea actualelor tehnologii ale Web-ului semantic în contextul aplicațiilor *e-business*. Unul dintre aspectele de interes va fi cel al utilizării ontologiilor pentru clasificarea resurselor și a componentelor software care intervin în cadrul unei aplicații Web.

2. Niveluri ale Web-ului semantic

Arhitectura Web-ului semantic este una funcțională, deoarece modul de constituire a acesteia se bazează pe specificarea incrementală a unor limbaje, pornind de la nivelul inferior (i.e. nivelul metadatelor) și ajungând la nivelurile superioare (e.g., nivelul logic). Limbajele disponibile pe fiecare nivel pot satisface

cerințe impuse de diferite tipuri (sau niveluri) de aplicații:

1. *nivelul metadatelor* pune la dispoziție cadrul general pentru exprimarea unor aserțiuni semantice simple. Modelul oferit conține concepte precum resursă și proprietate, utilizate să exprime meta-informații. Modelul se specifică via limbajul RDF (*Resource Description Framework*) și diverse vocabulare de metadata ca DCMI (*Dublin Core Metadata Initiative*), RSS (*Rich/RDF Site Summary*), FOAF (*Friend Of A Friend*) etc.;
2. *nivelul schemelor* oferă posibilitatea specificării de ontologii simple pentru a se putea defini o descriere ierarhică a conceptelor și proprietăților;
3. *nivelul logic* introduce limbaje ontologice mai complexe, capabile a modela ontologii sofisticate. Se dorește astfel constituirea unor servicii (*reasoning services*) pentru Web-ul semantic, de interes în ceea ce privește aplicațiile destinate *e-business*-ului (e.g., servicii de luare sau asistare în luarea deciziilor, servicii de monitorizare a vânzărilor *on-line* etc.).

3. Necesitatea exprimării ontologiilor

3.1 Preliminarii

Atunci când discutăm despre Internet și, în special, despre Web, ne gândim automat la *partajarea cunoștințelor* (aceasta implică activități de publicare, căutare, regăsire și interpretare a acestora). Evident, aceste activități nu pot fi efectuate doar procesând documente Web care folosesc vechile marcaje HTML (W3C, 2005), deoarece limbajul HTML este un limbaj pentru prezentarea datelor, nu destinat stocării sau atașării de se-

mantici (înțelesuri) acestora. Nici meta-lingajul XML – *Extensible Markup Language* – (Bray *et al.*, 2004) nu oferă, în mod direct, suport în definirea de semantici marcătorilor XML pe care putem să-i imaginăm și folosi. Rămâne în responsabilitatea utilizatorului asocierea de înțeles (*meaning*) fiecărui marcaj (*tag*) la care recurge (Buraga, 2004a; Buraga, 2005).

Astfel, comunitatea utilizatorilor, chiar dacă nu vorbesc cu toții aceeași limbă, trebuie să partajeze aceeași *conceptualizare* a informației (mai mult, atât oamenii, cât și mașinile trebuie să fie capabile să utilizeze în același mod cunoștințele). Aceasta implică existența unui vocabular comun, a unui set comun de constrângeri și a unui cadru conceptual comun oferind relații între clase de informație. Vocabularul comun garantează că elementele sintactice (cuvintele, marcătorii etc.) au același înțeles atât pentru transmițător, cât și pentru destinatar (astfel de vocabular poate fi derivat dintr-un vocabular utilizat deja în cadrul unei comunități umane sau industrii). Cadrul conceptual comun ne va asigura că înțelegerea informațiilor nu va fi influențată de cunoștințele inițiale nespecificate (*unspecified background knowledge*) – vezi (Daum & Merten, 2003).

Apare astfel necesitatea de a exprima semantica într-un mod formal, procesabil de către calculator. Două concepte importante sunt:

- *relația (relation)* dintre conceptele vehiculate – de exemplu, pentru a defini termenul „*e-business*” ne bazăm pe „*business*”, iar când specificăm „*business*” apelăm la relația acestuia cu termenul „*economy*”;
- *constrângerea sau restricția (constraint)* – pentru a face distincție dintre „*e-business*” și „*business*”, vom defini o restricție privitoare la disponibilitatea pe Web în cazul *e-business*-ului.

Revenind la Web, acesta evoluează spre ceea ce va fi *Web-ul semantic* sau *The Web of Meaning*, conform celor declarate de Berners-Lee în (Berners *et al.*, 2001). Pentru a modela la nivel semantic informațiile, vom recurge la *ontologii*, subiectul principal al acestui articol.

3.2 Definiții și tipuri

Astfel, conform celor descrise în (Sowa, 2001), subiectul *ontologiei* este studiul *categoriilor* de lucruri (*things*) care există sau pot exista într-un anumit domeniu de interes. Rezultatul unui astfel de studiu, denumit *ontologie*, este un catalog de tipuri de lucruri care se presupune că există într-un domeniu de interes D (*domain*) din perspectiva unei persoane care utilizează un limbaj L cu scopul de a discuta despre D . Tipurile din ontologie reprezintă *predicatele*, *sensurile cuvintelor* sau *conceptele și tipurile de relații* ale limbajului L , atunci când e utilizat pentru a discuta despre domeniul D .

După Gruber, citat în (Buraga, 2004a), o definiție a conceptului de ontologie este următoarea: „O *ontologie* reprezintă conceptualizarea unui domeniu de cunoaștere într-un format destinat a fi procesat de calculator, format modelând entități, attribute, relații și axiome”. În inteligența artificială (iar acum și în domeniul Web-ului semantic), ontologia se referă la precizarea a ceea ce se consideră că „există”.

O ontologie informală poate fi specificată prin intermediul unui catalog de tipuri care sunt ori nedefinite ori definite numai prin propoziții în limbajul uman. Este și cazul documentelor XML, unii cercetători considerând că orice document XML poate fi considerat ca fiind o ontologie de nivel 0 (Trăușan-Matu, 2004). Ontologiile formale pot fi clasificate în funcție de maniera în care sub-tipurile sunt deosebite de super-tipurile lor:

- *ontologii axiomatice* – disting sub-tipurile prin axiome și definiții specificate într-un limbaj formal (logic);
- *ontologii bazate pe prototipuri* – disting sub-tipurile prin comparația cu un membru tipic ori specificând un *prototip* pentru fiecare sub-tip în parte.

Ontologiile pot fi de mai multe tipuri, în funcție de domeniul de interes al comunității cercetătorilor care definesc și utilizează o anumită ontologie. Un prim aspect important este cel al *nivelului de descriere* pe care îl poate oferi o ontologie (Buraga, 2004a; Daum & Merten, 2003):

- *vocabular* – reprezintă forma cea mai

simplă a unei ontologii; în acest sens, o schemă XML poate defini o ontologie (numită de nivel 0);

- *taxonomie* – deoarece una dintre cele mai naturale relații este cea de clasificare, se pot specifica relații între obiecte și clase, sub-clase etc.; în lumea obiectuală, relațiile se numesc *is-a*, taxonomiile oferind o vedere ierarhică a conceptelor modelate (un astfel de ontologie poate modela în mod corespunzător ierarhia categoriilor de personal dintr-o întreprindere);
- *sistem relațional* – desigur, pot exista relațiile non-ierarhice (un exemplu ar fi relația *este-ascultat-de* (*is-listen-by*) stabilită între un profesor și studentul lui, care nu implică nici o ierarhie între clasele „profesor” și „student”); aceste relații sunt tipice în cadrul sistemelor de baze de date relaționale (în fapt, fiecare schemă de bază de date relațională poate fi considerată a fi o ontologie);
- *teorie axiomatică* – constrângerile care pot surveni într-o ontologie sunt definite ca *axiome* (o axiomă reprezintă un enunț logic care nu poate fi dedus din alte enunțuri, dar poate fi folosit pentru demonstrarea altora); astfel, într-o ontologie privitoare la ierbivore, am putea defini o axiomă de genul „orice ierbivor consumă plante sau părți dintr-o plantă”; referindu-ne la sisteme de baze de date, constrângerile sunt definite prin intermediul regulilor de integritate (trebuie făcută observația că unele limbaje de programare oferă suport pentru specificarea constrângerilor – de exemplu, în C/C++ se pot defini aserțiuni; mai mult, există o ramură a informaticii care se ocupă doar de *constraint logic programming*).

Ontologiile pot diferi și în ceea ce privește scopul și contextul utilizării conținutului (setului de cunoștințe) pe care-l modelează. De exemplu, se pot crea ontologii specializate, folosite într-un domeniu de interes precum afacerile electronice, chirurgia sau aeronautica (e.g., ontologia dezvoltată de Boeing sau ontologia proceselor folosită în contextul serviciilor Web – a se consulta specificația OWL-S (OWL-S), dar și ontologii generale în care informațiile provin din orice domeniu – ca exemplificări, se pot menționa: Cyc – cu

variantele OpenCyc (OpenCyc) –, ontologia lui Sowa (Sowa, 1999) ori ontologia ABC (Brickley *et al.*, 1999).

Conform celor de mai sus, putem identifica patru tipuri de ontologii (Daum & Merten, 2003):

- de nivel cel mai înalt (*top-level* ori *upper-level*) – sunt cele mai generale, servind scopurilor unei comunități largi;
- specifice unui domeniu (*domain-related*) – considerate și verticale, se aplică unui domeniu specific de cunoaștere (de exemplu, industria de calculatoare sau domeniul academic);
- specifice unor sarcini (*task-related*) – se aplică unei sarcini sau grup de sarcini înrudite (de exemplu, analiza cerințelor software); se mai numesc și ontologii orizontale;
- specifice unei aplicații (*application-related*) – descriu conceptele care intervin într-o aplicație, referindu-se la specializarea unei ontologii specifice unui domeniu și/sau sarcini.

3.3 Specificarea ontologiilor

O ontologie va include (Sowa, 1999): categoriile (conceptele) fundamentale, proprietățile conceptelor și relațiile și distincțiile dintre concepte. Conform (Buraga, 2004a), se pot folosi diverse *limbaje de specificare*, ca exemple notabile putând fi date KIF sau succesorul său CL (*Common Logics*). Alte limbaje sunt bazate pe logici descriptive (*description logics*), pentru Web fiind propuse OIL (*Ontology Inference Layer*), RDFS (*RDF Schema*), DAML+OIL ori actuala recomandare a Consorțiului Web: OWL (*Web Ontology Language*) – detalii în (Buraga, 2004a; W3C, 2005). Cele din urmă, fiind bazate pe XML, pot fi folosite cu succes și pentru partajarea ontologiilor între aplicații, un aspect foarte important în contextul sistemelor bazate pe agenți software utilizate în domenii precum *e-business*-ul (de pildă, un sit de comerț electronic ar putea să-și clasifice produsele oferite pe baza unei ontologii specifice, acest aspect facilitând căutarea „inteligentă” a unui produs pe diverse criterii ori luând în considerație preferințele utilizatorului, pe baza categoriilor de cumpărături deja efectuate). După (Trăușan-Matu, 2004),

„ontologiile constituie liantul care integrează sisteme de baze de date, sisteme bazate pe cunoștințe, sisteme obiectuale în diverse aplicații integratoare și bazate pe colaborare”.

3.3.1 Specificarea categoriilor

Primul aspect care trebuie luat în considerație este cel al specificării *categoriilor* ontologice. În teoria bazelor de date, categoriile sunt denumite, de obicei, *domenii*, iar în cadrul inteligenței artificiale se regăsesc sub numele de *tipuri* sau *sorturi*. Selectarea unor categorii determină tot ceea ce poate fi reprezentat într-o aplicație software sau o familie de aplicații. Sowa menționează în (Sowa, 2001) că „orice incompletitudine, distorsiuni sau restricții în eșafodajul de categorii vor limita inevitabil generalitatea fiecărui program și a bazei de date care utilizează acele categorii”. Criteriile prin care „populăm” ontologia dorită a fi creată/specificată sunt complexe, unul dintre aspectele importante fiind cel al experienței. În urma investigațiilor efectuate, se delimitează entitățile mentale (i.e., conceptele sau categoriile) ce pot fi diferențiate de alte categorii. Aceste entități pot intra în combinație cu altele, formând astfel noi concepte.

Sowa enumeră în (Sowa, 2001) diverse categorii ale ontologiei generale Cyc (scopul final al acestei ontologii este să cuprindă întreaga cunoaștere umană). Cea mai generală categorie se numește *Thing* și nu are proprietăți proprii. Actualmente, Cyc enumeră peste o sută de mii de tipuri de concepte având aproape un milion de fapte și axiome, varianta *open-source* fiind disponibilă la www.opencyc.org.

3.3.2 Specificarea proprietăților

Ceea ce diferențiază un concept de altul poate fi exprimat (și) prin intermediul *proprietăților* (numite și *attribute*). Astfel, fiecare concept (clasă ori *frame*) are asociate diverse proprietăți (*attribute*) care pot fi aplicate numai asupra claselor pentru care au fost definite. Se poate asocia același nume de atribut unor clase diferite, impunând anumite *restricții* ale valorilor atributelor considerate (aceste restricții sunt denumite *range restrictions*). După cum am văzut, o clasă

oferă un context pentru modelarea unui aspect specific unui domeniu.

O proprietate este o relație binară care poate specifica fapte privitoare la membrii unei clase ori la indivizi (Buraga, 2004a). Proprietățile se pot referi la tipurile de date (relații între instanțele de clase și literalii puși la dispoziție de limbajul de specificare sau diverse tipuri de date) ori la obiecte (relații între instanțele a două clase).

Definind o proprietate, în fapt se impune o restricție. Unele limbaje de specificare (precum OWL) pot oferi construcții prin care se exprimă faptul că o proprietate se poate defini ca fiind o sub-proprietate (specializare) a unei proprietăți deja specificate – în acest mod, proprietățile pot fi ierarhizate, similar claselor. Conform (Daum & Merten, 2003), restricțiile pot fi statice (structurale, semantice, de reprezentare, de proiectare) sau dinamice (tranziții și temporale).

3.3.3 Specificarea relațiilor

După cum se poate remarca din cele descrise mai sus, unul dintre aspectele cele mai importante îl reprezintă modul de specificare a relațiilor dintre concepte (în „jargonul” lui Sowa, modul de specificare a relațiilor dintre sub-tipuri și super-tipuri). Relațiile pot fi de hiperonimie („general-specific”), de meronimie („parte-întreg”) sau de antonimie. În mod normal, pentru organizarea și ilustrarea categoriilor ontologice în ierarhii se pot folosi arborii și grafurile. Fiecare ierarhie reprezintă o relație de ordine parțială (adică există cel puțin două elemente care nu pot fi comparate folosind relația de ordine considerată).

Trebuie făcută observația că, în specificarea relațiilor și proprietăților, pot interveni metadatele (date privitoare la date) care adaugă informații suplimentare fiecărui element al ariei de cunoaștere modelată.

3.3.4 Exemplificări

Drept exemplu, figurăm relațiile dintre metalimbajele SGML și XML și limbajele de marcate derivate din acestea: HTML, respectiv XHTML – detalii și în (Bray *et al.*, 2004; Geroimenko, 2004; W3C, 2005); a se vedea figura 1.

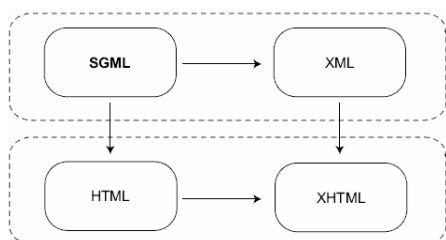


Fig. 1. Relațiile dintre meta-limbaje și limbajele derivate

Există disponibile diverse ontologii, specificate în mod uzual via construcții OWL, surprinzând diferite domenii de cunoaștere:

- ontologia camerelor digitale: www.xfront.com/owl/ontologies/camera
- ontologia alimentelor (*food ontology*): www.mindswap.org/~aditkal/food.owl
- ontologia proceselor (a serviciilor Web) OWL-S: www.daml.org/services
- ontologia standard de nivel superior IEEE SUO (*Standard Upper Ontology*): <http://suo.ieee.org>.

Pentru mai multe detalii, recomandăm cititorului să consulte și www.schemaweb.info.

3.4 Proiectarea ontologiilor

O metodologie referitoare la realizarea unei ontologii implică desfășurarea următoarelor etape:

1. *stabilirea domeniului de cunoaștere* – această fază presupune asamblarea resurselor de informație și expertiza a ceea ce vor deveni concepte (*things*) ale domeniului de interes; tot aici, se vor asigura consensul dintre posibilele alternative și consistența acestora;
2. *organizarea ontologiei* – se referă la proiectarea pe ansamblu a structurii conceptuale: identificarea principalelor concepte concrete și a proprietăților acestora, a relațiilor dintre concepte, crearea conceptelor abstracte, identificarea instanțelor, referențierea/includerea altor ontologii și așa mai departe;
3. *„umplerea” ontologiei* – această etapă are în vedere adăugarea conceptelor, relațiilor, indivizilor la nivelul de detaliere necesar pentru realizarea scopurilor ontologiei;
4. *verificarea ontologiei* – implică reconcilierea inconsistențelor dintre elemente, la nivel sintactic, logic și semantic; verificarea consistenței poate implica o clasificare automată, rezultând noi concepte bazate pe proprietăți individuale sau relații între clase;
5. *publicarea ontologiei* – faza presupune

verificarea finală de către experți ai domeniului și publicarea în cadrul mediului de dezvoltare (grup de interes, întreprindere, organizație, comunitate Internet etc.).

Se poate recurge la diverse instrumente de editare a ontologiilor, dintre care menționăm *Ontolingua*, *Protégé*, *SWOOP*, *Visio for Enterprise Architects* sau *WebODE*.

De asemenea, trebuie să avem în vedere ca aplicațiile folosite să ofere facilități pentru verificarea consistenței și coerenței ontologiei specificate. De notat faptul că declarațiile privi-toare la indivizi nu trebuie obligatoriu să fie stocate în cadrul aceluiași document.

Pot apărea complicații datorate mai multor factori, dintre care enumerăm doar necunoașterea *a-priori* a dimensiunii ontologiei, numărul persoanelor implicate în modelarea, verificarea și exploatarea ontologiei, utilizarea ontologiei de către utilizatori care nu sunt experți în domeniul considerat și – nu în ultimul rând – așa-numitul „fenomen” de *natural laziness*.

5. Utilizări ale ontologiilor în e-business

Unul dintre principalele domenii de utilizare a ontologiilor este, desigur, comerțul electronic. Siturile Web ale magazinelor virtuale pot beneficia de clasificările pe baze ontologice ale produselor oferite în vederea implementării unui motor intern de căutare inteligentă (Buraga, 2004a) și a unui sistem de recomandare pe baza preferințelor sau comenzilor anterioare ale vizitatorilor. Ontologiile pot fi folosite și în cadrul unor aplicații *e-business* oferind diverse servicii, precum cele de turism (Wright *et al.*, 2003).

Exploatând relațiile dintre diferite entități (utilizatori, resurse Web, tabele ale unor baze de date, module software etc.), se pot constitui servicii de raționament pentru luarea deciziilor sau pentru asistarea managerilor în diverse activități din aria *e-business* (Buraga, 2003; Buraga, 2004b; Buraga, 2005). De asemenea, tehnologiile Web-ului semantic pot fi folosite în cadrul întreprinderilor virtuale (Buraga & Cioca, 2005).

Din punct de vedere computațional, specificând și/sau utilizând o ontologie sau un set de ontologii se creează premisele realizării de

inferențe de către entitățile de calcul (e.g., agenții software, serviciile Web semantice etc.), cu importante implicații pe viitor în ceea ce privește modelarea și dezvoltarea aplicațiilor *e-business* inteligente disponibile în spațiul World-Wide Web.

Adăugându-se o dimensiune ontologică aplicațiilor Web se pun bazele unor noi tipuri de procese de afaceri, cu importante implicații asupra viitorului economiei digitale mondiale.

Referințe

- (Berners *et al.*, 2001) T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", *Scientific American*, 5, 2001
- (Bray *et al.*, 2004) T. Bray *et al.* (eds.), *Extensible Markup Language (XML) 1.0 (Third Edition)*, W3C Recommendation, 2004: www.w3.org/TR/REC-xml
- (Brickley *et al.*, 1999) D. Brickley *et al.*, *ABC: A Logical Model for Metadata Interoperability*, Harmony Project, Working Paper, 1999
- (Buraga, 2003) S. Buraga (coord.), *Aplicații Web la cheie*, Polirom, Iași, 2003
- (Buraga, 2004a) S. Buraga, *Semantic Web. Fundamente și aplicații*, Matrix Rom, București, 2004
- (Buraga, 2004b) S. Buraga (coord.), *Situri Web la cheie. Soluții profesionale de implementare*, Polirom, Iași, 2004
- (Buraga, 2005) S. Buraga, *Proiectarea siturilor Web* (ediția a doua), Polirom, Iași, 2005
- (Buraga & Cioca, 2005) S. Buraga, M. Cioca, „Semantic Web-based E-business Applications”, *The 7th International Symposium on Informatics in Economy (Information and Knowledge Age) – IE 2005 Proceedings*, Infocrec Press, Bucharest, 2005
- (Daum & Merten, 2003) B. Daum, U. Merten, *System Architecture with XML*, Elsevier Science, 2003
- (Geroimenko, 2004) V. Geroimenko, *Dictionary of XML Technologies and the Semantic Web*, Springer-Verlag, 2004
- (Sowa, 1999) J. Sowa, *Knowledge Representation: Logical, Philosophical and Computational Foundations*, Brooke Cole Publishing, 1999
- (Sowa, 2001) J. Sowa, „Ontologia și reprezentarea cunoștințelor”, în G. Constandache și Ș. Trăușan-Matu (coord.), *Hermeneutica și ontologia calculatoarelor*, Editura Tehnică, București, 2001
- (Trăușan-Matu, 2004) Ș. Trăușan-Matu, *Programare în Lisp. Inteligență artificială și Web semantic*, Polirom, Iași, 2004
- (Wright *et al.*, 2003) J. Wright *et al.*, „Using the ebXML Registry Repository to Manage Information in an Internet Travel Support System”, *Proceedings of the 6th International Conference on Business Information Systems – BIS 2003*, Colorado, 2003
- (OpenCyc) * * *, *OpenCyc*: www.opencyc.org
- (OWL-S) * * *, *OWL-S Specification*: www.daml.org/services
- (W3C) * * *, *World-Wide Web Consortium*: www.w3.org