Compatibilizarea Ontologiilor

-Referat-

Turlea Ana Cristina

Grupa 406 IS

Cuprins

[Introducere 3](#_Toc382851941)

[Ontologia 4](#_Toc382851942)

[Compatibilizarea Ontologiilor (Ontology Matching) 4](#_Toc382851943)

[Compatibilizare bazata pe coincidenta 6](#_Toc382851944)

[Extragerea corespondentelor folosind Algoritmi Genetici 8](#_Toc382851945)

[LOM 10](#_Toc382851946)

[Concluzie 10](#_Toc382851947)

[Bibliografie 11](#_Toc382851948)

Compatibilizarea Ontologiilor

# Introducere

Una din cele mai importante inventii din ultimele decenii o reprezinta Serviciile Web. Acestea au rolul de a impartasi informatiile intre aplicatii folosind internetul. Recent au aparut diverse puncte slabe, cum ar fi parcurgerea informatiilor fara a lua in considerare semnificatia lor. Aceastea determina nevoia de a aparea un nou Web care sa aiba mai multa relevanta pentru utilizator.

Semantic Web este o extensie a Web-ului actual deoarece reprezinta informatiile cu mai mult sens atat pentru oameni cat si pentru calculatoare. Permite descrierea continutului si serviciilor intr-o forma care poate fi inteleasa de masini si permite automatizarea proceselor precum adnotarea, descoperirea, publicitatea, publicarea si compunerea serviciilor. Web-ul Semantic permite descrierea formală a resurselor existente pe Internet (pagini Web, documente text şi multimedia, baze de date, servicii etc). Dintre avantajele acestuia se impune ca principală identificarea rapidă și precisă a resurselor relevante pentru utilizator precum şi exploatarea automată a resurselor de către agenţii inteligenţi. Ideea de Web Semantic a apărut în urmă cu aproximativ 15 ani şi a fost introdusa de către Tim Berners-Lee, inventatorul Web-ului. Aceasta forma a Web-ului permite calculatoarelor sa utilizare informatiile de pe web in acelas mod in care o fac oamenii. Cu alte cuvinte, semantica informatiilor este bine definita in web-ul semantic pentru a face posibila extragerea informatiilor. Totusi, web-ul contine informatii distibuite si eterogene. De exemplu sa presupunem ca doriti sa aflati mai multe informatii despre un profesor care preda informatica la o universitate din apropiere. L-ati vazut la o conferinta acum doua luni si va mai amintiti doar numele lui de familie. Cu structura actuala a web-ului este greu sa gasiti astfel de informatii, deoarece acestea sunt distribuite pe pagini web diferite. Totusi este usor sa aflati ce aveti nevoie cu ajutorul web-ului semantic.

A fost dezvoltat avand ca baza Ontologia, considerata ca fiind “coloana vertebrala” a Web-ului semantic. Cu alte cuvinte Web-ul curent se transforma din usor de citit de masini in usor de inteles de masini. Una din functiile Web-ului este de a construi o sursa de referinte pentru diferite subiecte, in timp ce Web-ul semantic este conceput pentru a construi un Web cu sens. Ontologia reprezinta fundamentul vocabularului si comunicarii in Web-ul semantic.

# Ontologia

Ontologia ofera o specificatie formala, explicita a unei conceptualizari. Prin urmare, faciliteaza schimbul de cunostinte asupra sistemelor distribuite. Cu alte cuvinte, aceasta permite sistemelor sau aplicatiilor sa coopereze desi nu au fost proiectate initial pentru a interactiona intre ele. Ontologia joaca un rol important in rezolvarea problemelor de interactiune intre aplicatiile unor diverse organizatii furnizand o semnificatie pentru domeniile comune.

Ontologia reprezinta conceptualizarea unui domeniu de cunoastere intr-un format destinat a fi procesat de catre calculator, modeland entitati, atribute, relatii si axiome. Ontologiile sunt o componenta cheie in rezolvarea problemei de eterogenitate semantica, permitand astfel aparitia corespondentelor semantice intre diferite aplicatii sau servicii web. Recent, ontologiile au devenit un subiect de cercetare popular in diferite domenii, inclusiv inginerie, comert electronic, management si procesarea limbajului natural. Scopul unei ontologii este de a realiza o semantica partajata, comuna, care poate fi transmisa intre oameni si intre aplicatii.

In domeniul web-ului semantic, ontologia este un vocabular care specifica sensul termenilor utilizati intr-un anumit domeniu. Cu alte cuvinte, ontologiile marcheaza informatiile intr-un mod in care sa le poata intelege calculatoarele mai bine semantica. Astfel, ontologiile joaca un rol important in realizarea corespondentelor in cadrul organizatiilor si in cazul Web-ului semantic, deoarece scopul lor este de a capta cunostintele dintr-un domeniu si a crea explicit semantica in mod generic. Este necesar să se stabilească mapari între conceptele de ontologii diferite pentru a capta corespondența semantică între ele. Cu toate acestea, stabilirea unei astfel de corespondență nu este o sarcina ușoara.

S-au construit diverse ontologii. Prin urmare ele trebuie sa fie accesate de alte aplicatii care fac uz sau schimb de informatii. Astfel de ontologii prezinta probleme de corespondenta, pentru care s-au dezvoltat mai multe solutii. Una dintre solutii este de a contrui o singura ontologie, dar aceasta este o solutie inadecvata, deoarece este mult prea inflexibila din punct de vedere al impartirii informatiilor. O alta solutie o reprezinta compatibilizarea ontologiilor, care joaca un rol important in rezolvarea corespondentelor in sisteme heterogene.

# Compatibilizarea Ontologiilor (Ontology Matching)

Compatibilizarea Ontologiilor este un factor cheie pentru interoperabilitatea web-ului semantic, precum si o tehnica utila in anumite sarcini de integrare a datelor clasice care intampina probleme de eterogenitate a semanticii. Preia Ontologiile ca input si determina ca output un set de corespondente intre entitatile corespunzatoare din acele ontologii. Aceste corespondente pot fi utilizate pentru diferite sarcini, cum ar fi fuzionarea ontologiilor, translatarea datelor sau navigarea pe web-ul de date (web data). Astfel compatibilizarea ontologiilor permite cunostintelor si datelor exprimate in ontologiile corespunzatoare sa interopereze.

In lucrarea (Arfaee, 2010) este preluata din “An integrative proximity measure for ontology alignment. *In Proceedings of 3rd International Semantic Web Conference*, 2004” de Euzenat J. si Valtchev P. o definitie a compatibilizarii ontologiilor: “date doua ontologii, fiecare descrizand un set finit de entitatati (care pot fi clase, proprietati, reguli, predicate etc.), sa se gaseasca relatiile (echivalenta) dintre aceste entitati”. Tot in aceasta lucrare s-a preluat din “A survey on ontology mapping. *Sigmod Record*, 2006” de Choi N., Song I. Y., si Han H., o divizare a abordarii compatibilizarii ontologiilor in trei grupuri. Primul grup contine corespondenta ontologiilor locale si globale. In acest fel, este usor de gasit relatiile dintre ontologiile locale si o ontologie globala datorita unui vocabular comun care leaga toate conceptele din ontologiile locale de aceleasi concepte dar din ontologia generala. Totusi, realizarea corespondentelor dintre ontologiile locale este un procedeu mai dificil. Pe de alta parte, al doilea grup prezinta crearea corespondentelor intre ontologiile locale. Aceasta mapeaza concepte similare din ontologia sursa cu concepte similare semantic din ontologia tinta. Ultimul grup se bazeaza pe fuzionarea ontologiilor intr-o ontologie coerenta. In continuare vom face referire doar la primul timp de compatibilizare a ontologiilor. In ontologiile de mari dimensiuni este imposibil de a examina toate corespondentele pentru a o selecta pe cea mai buna, deci este nevoie de anumite metode de a determina cea mai buna corespondenta. O astfel de metoda este extragerea corespondentelor tinand cont de anumite concepte.

Compatibilizarea ontologiilor vine ca solutie pentru integritatea datelor, nevoie aparuta din cauza existentei unei nepotriviri (mismatches). Printre motivele existentei acestor nepotriviri de date avem: reprezentarea subiectiva a unui domeniu (spre exemplu carte pentru copii despre animale vs compendiu zoologic), existenta diferitelor cerinte si worflow-uri privitoare la (tipurile de) aplicatii software (e-business vs. e-learning vs. e-entertainment, interactiune conventionala (desktop) vs. mediu mobil/TV), folosirea altor conventii de reprezentare a cunostintelor. Nepotrivirile apar la nivel de limbaj si la nivel ontologic.

La nivel de limbaj avem diferenta de sintaxa, expresivitate sau semantica a limbajelor ontologice actuale. La nivel ontologic avem diferente vizand structura semantica (exemple: termeni diferiti definind acelas concept, acelasi termen descriind mai multe concepte, conventii, paradigme de modelare eterogene, puncte de vedere diferite etc.).

Reducerea eterogenitatii poate fi realizata in doi pasi:

1. Potrivire (match)- determinarea alinierii la momentul proiectarii sau reutilizarii ontologiei
2. Procesarea alinierii - fuzionare (merging), transformare etc.

Alinierea (alignment) A intre si este o multime de corespondente peste si si prezinta o anumita cardinalitate: 1 la 1, 1 la N etc si pot fi atasate meta-date aditionale: metoda de aliniere, data alinierii, proprietati specifice etc.

Exista mai multe tehnici de compatibilizare. O posibila impartire pe categorii este urmatoarea:

* Tehnici de compatibilizare la nivel de termeni pe baza sirurilor de caractere:
  + Prefixul: net = network, pin = pinguin ??
  + Sufixul: ID = PID, word = sword??
  + Distanta de editare dintre doua cuvinte (numarul de operatii de inserare, stergere, substitutie a caracterelor pentru transformarea unui termen in altul): Distanta(NKN, Nikon) = 4.
* Tehnici de compatibilizare la nivel de termeni bazate pe limbaj: procesarea entitatilor (tokenization)
  + Hands-Free Kits --- <hands, free, kits>
  + Analiza morfologica pentru gasirea tuturor formelor de baza (lemmalization): Kits – Kit
  + Eliminarea cuvintelor nerelevante: ‘a’, ‘the’, ‘by’ etc.
* Tehnici de compatibilizare la nivel de termeni pe baza sensurilor: sinonimie, antonimie sau relatia de siblings (familia lexicala).
* Tehnici de compatibilizare la nivel de termeni pe baza glossei: numarul acelorasi cuvinte prezente in ambele glosse (texte) de intrare conduce la cresterea valorii de similiaritate a termenilor.
* Tehnici de compatibilizare la nivel de termeni pe baza taxonomiei: potriviri ale drumurilor marginite in grafurile ontologiilor (bounded path matching).
* Tehnici de compatibilizare la nivel de termeni bazate pe modele: verificarea satisfiabilitatii (SAT) pe baza axiomelor.
* Tehnici de compatibilizare la nivel structural bazate pe modele: utilizarea logicii descrierii.

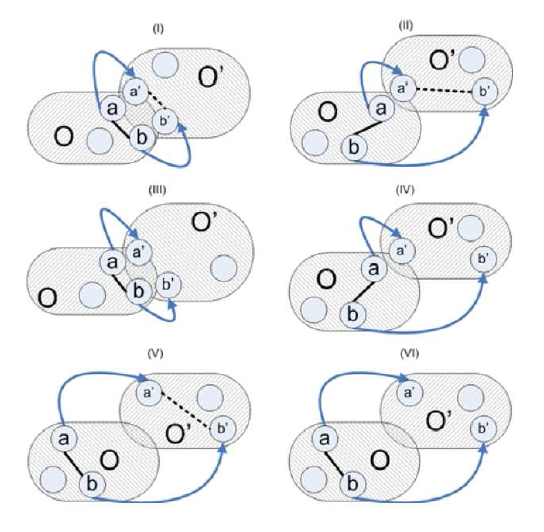
Aceste procese de compatibilizare pot avea loc secvential sau in paralel.

S-au dezvoltat si modele bazate pe asocierea unui graf fiecarei ontologii. Fiecare nod corespunde unui termen si fiecare arc corespunde legaturii dintre termeni. In continuare vom prezenta cateva metode de compatibilizare bazate pe grafuri.

# Compatibilizare bazata pe coincidenta

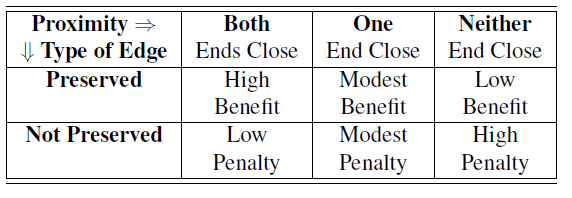
Modelul de compatibilizare bazat pe coincidenta defineste o functie ponderata ca masura de similitudine intre un set de concepte din doua ontologii. In lucrarea (Arfaee, 2010) este prezentata aceasta metoda prin idei extrase din lucrarea “Coincidence based scoring of mapping in an ontology alignment. *Journal of Advanced Computational Intelligence and Intelligent Informatics”* realizara de catre Haeri S. H., Abolhassani H., Qazvinian V., and Hariri B.

Problema compatibilizarii ontologiilor s-a redus la potrivirea, compatibilizarea a doua grafuri cu tip. Un graf cu tip este un graf ale carui muchii au asociate un tip. In acest mod, nodurile si, respectiv, grafurile cu tip indica conceptele ontologiei si relatiile dintre diferitele concepte. Deasemenea au introdus si conceptul de conservarea a muchiilor. Presupunem ca exista o muchie de tip t in graful ontologiei sursa. Muchia se numeste conservata daca si numai daca exista o muchie de tip t in graful ontologiei tinta unde m(a) si m(b) sunt corespondentele lui a si b in graful ontologiei tinta. Functia de pondere este definita prin sase categorii diferite cum sunt descrise in figura 1 preluata din lucrarea (Arfaee, 2010).

In aceasta figura, O si O’ sunt doua ontologii diferite si G si G’ sunt grafurile asociate. Mai mult, a si b sunt concepte din G iar a’, b’ sunt concepte din G’.

Figura

1. Categoria I: atat a’ si b’, cat si a si b sunt prea aproape dpdvd al distantei de editare a cuvintelor. Muchia dintre ele este muchie conservata. Astfel corespondentei ii este asociata o pondere mare. De retinut ca in acest model este mai importanta proprietatea muchiei de a fi conservata decat apropierea perechilor. In modelul bazat pe coincidenta perechea este reprezentata de doua varfuri ale unei ontologii intre care exista muchie.
2. Categoria II: in aceasta categorie, muchiile dintre perechi sunt conservate dar doar unul dintre elementele perechii sunt apropiate unul de celalalt. Ponderea corespondentei este moderata si mai mica decat ponderea din categoria precedenta.
3. Categoria III: in aceasta categorie, perechile sunt apropiate cu ambele capete dar muchiile nu se conserva, deci ponderea va scadea, dar nu foarte mult datorita nodurilor apropiate.
4. Categoria IV: in aceasta categorie, a si a’ sunt apropiate dar b si b’ sunt departate unul de celalalt, iar muchia dintre a si a’ nu este conservata. Aceasta corespondenta va avea ponderea mai mica decat cea a categoriei precedenta, deoarece doar un capat al perechiilor are distanta dintre varfuri mica.
5. Categoriile V si VI: in aceste doua categorii, varfurile sunt departate unul de celalat. Diferenta dintre cele doua ontologii este conservarea muchiilor, categoria VI va primi ponderea cea mai mica, din cauza neconservarii muchiilor.

Aceasta functie reprezinta o masura de similitudine. Ponderea unei compatibilizari este suma ponderilor corespondentelor dintre ontologiile sursa si tinta. Figura 2 reprezinta strategia de alegere a ponderilor in modelul bazat pe coincidenta.

# Extragerea corespondentelor folosind Algoritmi Genetici

Figura

Pentru a gasi cea mai buna corespondenta intre conceptele a doua ontologii s-au folosit si Algoritmii Genetici. Algoritmul presupune cateva iteratii pentru a dezvolta o populatie de indivizi. Fiecare iteratie a algoritmului este formata din doua faze consecutive de incrucisare si mutatie. Incrucisarea este procesul prin care (1) se selecteaza doi indivizi (parinti) din populatia curenta pe baza unei anumite functii si (2) se produc noi indivizi din acesti parinti. Mutatia este procesul prin care se schimba random anumite caracteristici ale fiecarui individ pentru a produce indivizi mai buni.

Algoritmul incepe cu o populatie random de corespondente si se incearca imbunatatirea lor prin folosirea unei functii de pondere prezentata in modelul de compatibilizare prin coincidente. Faza de incrucisare este un pic mai complicata. Ea consta in doua abordari diferite. Una bazata pe varfuri si una bazata pe perechi. In pasul de mutatie sunt alese aleator doua noduri din ontologia sursa si corespondentele lor din ontologia tinta sunt inlocuite cu mutatiile. Daca ponderea totala nu este imbunatatita, schimbarile sunt anulate.

S-au folosit doua functii de incrucisare diferite, una bazata pe perechi si una bazata pe noduri. In prima incrucisare, pentru toate perechile din graful ontologiei sursa se iau in considerare corespondentele parintilor si este aleasa corespondenta cu cea mai mare pondere. Aceasta functie este prezentata in figura 3. In cazul in care nodul target este deja asociat altui nod, se va face o corespondenta aleatoare.

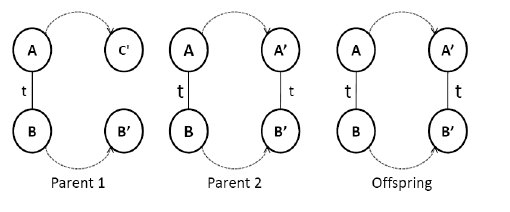
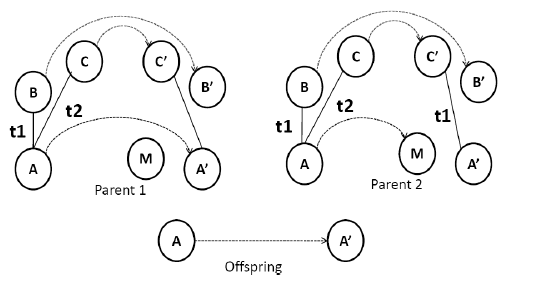
Ponderea unui nod este suma ponderilor tuturor perechilor in care nodul este inclus. Cea de-a doua functie de incrucisare se bazeaza pe aceasta caracteristica. Pentru fiecare nod al grafului ontologiei se va copia corespondenta cu cea mai mare pondere. Functia de incrucisare este prezentata in figura 4. In cazul in care nodul din ontologia target este asignat anterior unui alt nod, corespondenta aleatoare nu va fi realizata pana cand toate nodurile din ontologia sursa vor fi mapate, deoarece poate impiedica o asignare ulterioara mai buna a nodului target respectiv. Autorii acestui algoritm s-au concentrat pe cresterea ponderii in interatiile consecutive ale algoritmului genetic.

Figura 4

Figura 3

# LOM

Maparea ontologiilor bazata pe lexic (LOM) se bazeaza pe doua afirmatii: (1) oamenii ar trebui sa se implice in procesul de compatibilizare al ontologiilor, dar munca depusa poate fi redusa prin anumite instrumente inteligente. (2) cand limbajul natural al ontologiilor este similar, ar trebui folosita o metoda de a realiza corespondentele bazata pe lexic. Se folosesc patru tipuri diferite de potriviri pentru a construi sistemul:

1. Potrivirea totala a termenilor: LOM transforma toti termenii ai ambelor ontologii in litere mici, apoi se asigneaza 1 unei pereche daca a fost potrivita, si 0 altfel.
2. Potrivirea cuvintelor prin substitutie: in cazul in care intalnim o litera majuscula, o cratima sau underscore, LOM imparte termenul in doua cuvinte si apoi aplica o functie care asigneaza scoruri intre 0 si 1, proportional cu numarul de cuvinte potrivite. Deasemenea LOM inlatura cuvintele precum ‘a’, ‘the’, ‘in’ etc. inainte de atribuirea scorului.
3. Potrivirea sinonimelor: aceasta procedura extrage semantica cuvintelor folosind WordNet. Apoi determina un scor ca la (2).
4. Potrivirea tipurilor. LOM foloseste aceasta metoda aceasta daca nu se incadreaza in conditiile metodelor anterioare. In aceasta procedura se va mapa sinonimul unui termen cu termeni dintr-o ontologie “upper level”.

De retinut ca abordarea are anumite limitari in cazul in care ontologia are simboluri abstracte cum ar fi cele matematice.

# Concluzie

Compatibilizarea ontologiilor este un subiect important in web-ul semantic. Este un concept nou si exista diverse comunitati care incearca sa abordeze problema integrarii resurselor de pe Web. Toate aceste abordari de a rezolva problema integrarii resurselor pe Web au aparut in ultimul deceniu si sunt intr-un stadiu timpuriu de dezvoltare. Deci mai este munca multa de facut pana la obtinerea productivitatii dorite.

# Bibliografie

Arfaee, S. J. (2010, February 1). Ontology Matching Approaches in Semantic Web: A Survey. Edmonton, Canada. Retrieved from http://webdocs.cs.ualberta.ca/~jabbaria/Documents/Ontology%20Matching%20Approaches%20in%20semantic%20web.pdf

Miron, A. D. (n.d.). Semantic Web scurtă introducere. *Today Software Magazine*. Retrieved from http://www.todaysoftmag.ro/article/ro/4/Semantic\_Web\_scurta\_introducere\_82

Ontology Matching. (2013). Retrieved from http://ceur-ws.org/Vol-1111/om2013\_proceedings.pdf

Taye, M. M. (2010). Understanding Semantic Web and Ontologies: Theory and Applications. *Journal of Computing* *, 2* (6), 182-192. Retrieved from http://arxiv.org/ftp/arxiv/papers/1006/1006.4567.pdf