Introducere

Web-ul semantic este o extindere a web-ului actual, care permite găsirea, folosirea şi combinarea informațiilor de pe internet într-un mod mult mai uşor şi rapid. Noua abordare se bazează pe colectarea, prelucrarea şi publicarea de informații interpretabile de maşini şi pe metadate exprimate în limbajul **RDF** (Resource Description Framework). În prezent, conținutul web-ului este proiectat pentru a putea fi citit de ființe umane, nu pentru a putea fi refolosit de aplicații informatice. Web-ul semantic va completa şi dezvolta web-ul actual, creând un mediu în care agenți software vor putea procesa diverse sarcini sofisticate, un mediu în care informațiile vor avea un înțeles bine-definit. Astfel, se speră ca, în viitorul apropiat, calculatoarele să poată afişa, dar şi "înțelege" date.

Caracteristicile de bază ale Web-ului Semantic, respectiv înțelesuri bine definite ale conceptelor şi metadate procesabile automat de către calculatoare, folosite de către agenți software corespunzători, stabilesc o abordare eficientă de satisfacere a cerințelor **eLearning**-ului. Materialele pot fi interpretate semantic şi, la cererea utilizatorilor, pot fi reorganizate pentru a crea un nou modul activității didactice. În funcție de cerințele şi preferințele utilizatorului, materialele şi informațiile considerate relevante pot fi combinate într-un mod foarte simplu şi intuitiv. Acest proces se bazează pe interogări semantice şi pe navigarea printre materialele de studiu şi este posibil datorită **ontologiilor**, care furnizează definiții precise ale conceptelor şi ale noțiunilor utilizate.

În cadrul acestei lucrări, autorul şi-a propus şi a realizat proiectarea şi implementarea unei platforme destinate eLearning-ului. Utilizatorii acestei

platforme pot introduce texte din diverse domenii ale cunoaşterii, urmând ca infrastructura să deducă (folosind un algoritm propriu de combinare a clasificatorilor de text) domeniul din care face parte textul respectiv. De asemenea, platforma va propune materiale de studiu care abordează, la diverse niveluri, domeniul stabilit.

Lucrarea este structurată în cinci capitole.

Primul capitol prezintă caracteristicile principale ale Web-ului Semantic, tehnologiile şi conceptele care stau la baza acestuia, precum şi avantajele pe care le poate oferi acesta în cadrul procesului de eLearning.

Al doilea capitol este o trecere în revistă a infrastructurii ce reprezintă scopul final al acestei lucrări. Sunt prezentate principalele componente ale platformei, precum şi principalii algoritmi propuşi şi implementați în cadrul infrastructurii.

În cadrul celui de al treilea capitol, se face o analiză a stadiului actual al cercetării în domeniul web-ului semantic, punându-se accentul pe modalități de a raționa cu ajutorul conceptelor şi a relațiilor definite în cadrul ontologiilor (un aspect crucial pentru buna funcționare a unei infrastructurii bazate pe web-ul semantic). În continuare, sunt prezentate un studiu comparativ al limbajelor de interogare a bazelor de date RDF, precum şi un studiu comparativ al tehnologiilor semantice emergente în cadrul eLearning-ului.

Capitolul al patrulea prezintă conceptele şi tehnologiile pe care se bazează infrastructura de eLearning propusă în cadrul lucrării. Apoi, este prezentată, în detaliu, fiecare componentă a platformei. Aceasta include

proiectarea şi implementarea unei componente destinate instruirii asistate, ce identifică domeniul de învățare pe bază de antrenare şi clasificare, precum şi o componentă ce furnizează materiale de studiu pentru domeniul identificat. Domeniul este stabilit cu ajutorul ontologiei de domenii, o ierarhie ce conține 289 domenii şi subdomenii. Fiecărui domeniu frunză din ierarhie îi este asociat un document de antrenare, necesar procesului de clasificare.

De asemenea, sunt prezentați algoritmii utilizați în pre-procesarea documentelor de antrenare a clasificatorilor şi a textelor introduse de utilizatori.

Ultimul capitol este rezervat descrierii detaliate a algoritmilor de clasificare de text implementați în cadrul infrastructurii. Aceştia sunt implementați în cadrul serviciului web, componentă care este destinată stabilirii automatizate a domeniului de lucru, pe baza textului introdus de utilizator.

Aceasta include o prezentare teoretică a tuturor algoritmilor, definirea formală a acestora, precum şi reprezentările acestora sub formă de scheme logice şi în pseudocod.

De asemenea, tot în ultimul capitol, este propus un algoritm propriu de combinare a rezultatelor clasificatorilor (ce produc clasamente de clase, nu valori numerice), care îmbunătățeşte performanța clasificării în comparație cu fiecare algoritm, luat individual.

Lucrarea se încheie printr-o scurtă secțiune de concluzii şi o bibliografie ce conține titlurile reprezentative pentru tematica abordată.

Capitolul 1. Caracteristicile web-ului semantic

1.1 Concepte şi tehnologii de bază

Web-ul semantic a fost inventat de Tim Berners-Lee, cel care a inventat şi URI (Uniform Resource Identifier), HTTP (HyperText Transfer Protocol) si HTML (HyperText Markup Language).

Tehnologiile care stau la baza web-ului semantic sunt incă într-un stadiu incipient şi, deşi viitorul proiectului pare foarte promițător, nu s-a ajuns la un consens în legătură cu direcția şi caracteristicile acestuia.

Informațiile ce pot fi găsite pe web, în fişierele html, pot fi utile în anumite cazuri, dar în altele, nu. O căutare Google returnează aproximativ 25% din totatlitatea rezultatelor relevante pentru utilizator, şi, foarte des se poate întâmpla ca anumite căutari să nu întoarcă nici un rezultat, chiar dacă există site-uri relevante pentru căutarea respectiva. De exemplu, se pot găsi uşor site-uri care oferă informații despre vreme, evenimente locale, programul tv, etc, dar toate sunt prezentate în HTML. Problema este că, în anumite contexte, este dificil să se preia datele respective pentru a fi folosite de fiecare utilizator în scopurile proprii.

Web-ul semantic poate fi văzut ca o soluție "inginerească". Acesta va fi construit cu sintaxe care folosesc URI. Un URI este, pur şi simplu, un identificator pentru Web (şiruri de caractare începând cu "http:" sau "ftp:", etc). Oricine poate crea un URI, iar orice resursă care are un URI poate fi considerată "pe Web".

Pentru Web-ul semantic, a fost nevoie de crearea unui limbaj care foloseşte un triplet de URI-uri (în limbaj natural, pot fi asemănate cu subiectul, predicatul şi complementul unei propoziții). Acest limbaj se numeste RDF (Resource Description Framework) şi are scopul de a standardiza trimiterea şi primirea de date. RDF este un limbaj XML ce prezintă informația folosind o serie de afirmații.

O altă parte foarte importantă a web-ului semantic o reprezintă ontologiile. Definiția cea mai des utilizată afirmă că o ontologie este definirea formală a unei conceptualizări acceptată de o anumită comunitate , dar nu există o definiție universal acceptată.

Ontologiile sunt utilizate pentru a reprezenta cunoştințele despre un anumit domeniu, furnizând vocabularul comun utilizat de comunitatea respectivă, definițiile conceptelor utilizate în vocabular, precum şi relațiile existente între aceste concepte.

Ambele tehnologii de bază, care formează coloana vertebrală a oricărei aplicații semantice, sunt prezentate pe larg în cele ce urmează.

1.1.1 Resource Description Framework (RDF)

Resource Description Framework (RDF) este un limbaj creat pentru descrierea informației legate de resursele World Wide Web. Inițial a fost dezvoltat pentru a reprezenta informațiile “metadata”, cum ar fi numele şi adresa de mail ale autorului unui document, legate de o anumită resursă web. Ulterior, conceptul a fost extins şi RDF a început să fie folosit pentru a descrie orice tip de informații în legătură cu orice obiecte ce pot fi

identificate pe Web, chiar dacă acestea nu pot fi preluate de pe Web. Exemple ar putea fi conținutul ultimelor ştiri, informații despre produsele existente pentru vânzare la magazine on-line (preț, specificații, disponibilitate),etc.

RDF este creat pentru situații in care aceasta informație trebuie procesată de aplicații, nu doar afişată utilizatorilor. RDF furnizează o schemă comună pentru reprezentarea acestor informații pentru a putea fi interschimbată între aplicații, fără pierderea înțelesului semantic al informației.

RDF se bazează pe ideea identificării oricărui lucru folosind identificatori WEB (numiți Uniform Resource Identifiers – URI) şi descrierea oricărei resurse folosind proprietăți, respectiv valori ale proprietăților. De aceea, prin RDF se pot reprezenta afirmații simple despre resurse ca un graf în care nodurile şi arcele reprezintă resurse, respectiv proprietăți ale resurselor. Pentru concretizare, să considerăm afirmația “există o persoană a cărei identitate se află la adresa ” http://85.204.140.247/eu#eu, a cărei nume este Altăr Adam, şi a cărei adresa de email este adamalt@adamalt.ro. Aceasta afirmație poate fi reprezentată prin graful RDF din Figura 1:

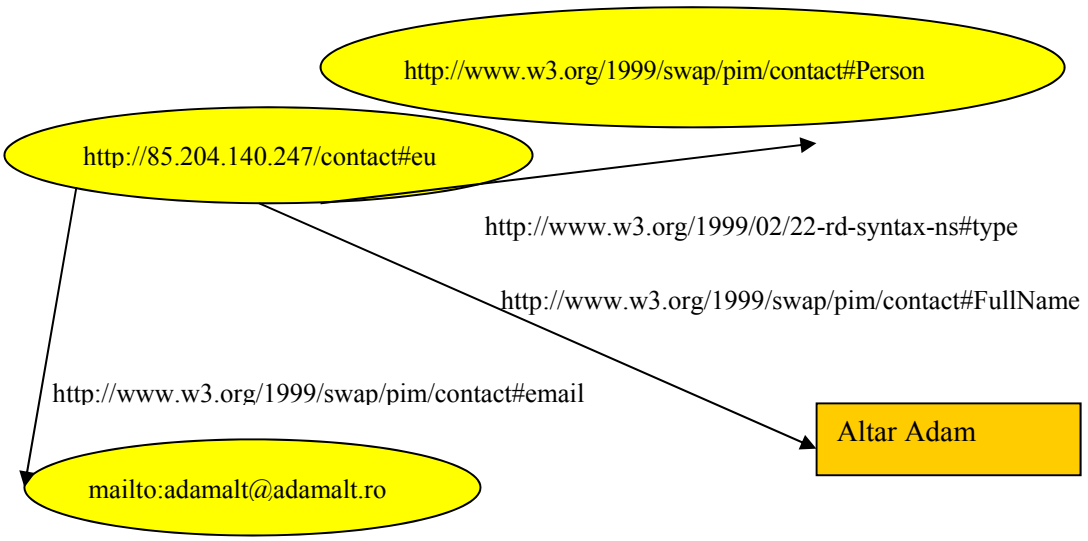

Figura 1 **Reprezentarea proprietăţilor unui obiect în limbajul RDF** ilustrează faptul că RDF-ul foloseşte URI-uri pentru a identifica persona (Altăr Adam, în exemplu, identificat la adresa http://85.204.140.247/contact#eu), tipuri de obiecte (Persoană, în exemplu, identificat la adresa http://www.w3.org/1999/swap/pim/ contact#Person), proprietăți ale acestor obiecte (adresa de email, în exemplu, identificată la adresa http://www.w3.org/1999/swap/pim/ contact#email) şi valori ale acestor proprietăți (mailto:adamalt@adamalt.ro, în exemplu, ca valoare a proprietății email). Această afirmație poate fi reprezentată şi ca sintaxă RDF/XML. Iată cum ar arăta afirmația de mai sus în sintaxă RDF/XML:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#"
```