

DREJT SEMANTIKËS WEB,NJË PAMJE E PËRGJITHSHME E XML-SË

BOJAXHIU A., XHINA E., DHËMBI A.

Universiteti i Tiranës, Fakulteti i Shkencave të Natyrës, Departamenti i Informatikës

e-mail: aretibojaxhiu@fshn.edu.al

Përmbledhje

Në këtë dokument përshkruhet një udhëtim hipotetik tek nocioni i semantikës web duke përshkruar synimet, problemet, rezultatet e mundshme dhe ndikimin mbi përdoruesit e internetit. Mungesa e logjikës midis gjuhëve në www si për shembull XML, skema XML, XPath dhe gjuhët semantike websi për shembull: RDF, RDFS, dhe DAML+OIL krijojnë një gjendje problematike për qëllimet e semantikës web. Kjo rezulton nga një ndryshim në modelimin midis XML dhe logjikës web. Në këtë punim propozohet për të eliminuar këtë ndryshim, duke krijuar një bazë të përbashkët semantike për www dhe semantikën web, duke marrë ide nga të dyja. Nuk rezulton në thelb asnjë ndryshim në XML, por vetëm ndryshime të vogla në RDF. Lejon semantikën web për të qënë afër qëllimit të përshkrimit të webit, si një semantikë duke u bazuar në gjuhët e webit (përfshirë RDFS dhe DAML+OIL) të cilat janë ndryshuar në mënyrë të konsiderueshme për shkak të këtij qëllimi të përbashkët.

Abstract

In this paper we describe a hypothetical journey to the notion of a semantic web, describing the intents, the problems, the possible outcomes and the impact on web users. The semantic discontinuity between World WideWeb languages, e.g., XML, XML Schema, and XPath, and SemanticWeb languages, e.g., RDF, RDFS, and DAML+OIL, forms a serious barrier for the stated goals of the Semantic Web. This discontinuity results from a difference in modeling foundations between XML and logics. We propose to eliminate that discontinuity by creating a common semantic foundation for both the World-Wide Web and the Semantic Web, taking ideas from both. The common foundation results in essentially no change to XML, and only minor changes to RDF. But it allows the Semantic Web to get closer to its goal of describing the semantics of the World Wide Web. Other Semantic Web languages (including RDFS and DAML+OIL) are considerably changed because of this common foundation.

Fjalëkyçe: Semantika, Semantika web, XML, Modeli XML, WWW.

Hyrje

Termi “Semantika Web” është cituar nga Drejtori i W3C, Tim Berners-Lee për të treguar një evolucion të mundshëm të botës së gjerë aktuale web. Kjo është një mundësi, një qëllim; diçka që nuk mund të arrihet me anë të teknologjive aktuale në dispozicion.

Përkufizimi i semantikës web

Semantika është definuar si studimi i kuptimit dhe ndryshimit të kuptimit. Semantikë web është një web ku fokusi është vendosur më shumë në kuptimin e fjalëve sesa në vetë fjalët ku informacioni bëhet njohuri pas kryerjes së analizës semantike. Për këtë arsye, një semantikë web është një rrjet i dijes në krahasim me atë që kemi sot, që mund të përkufizohet si një rrjet i informacionit.

Rrjetet e informacionit

Rrjetet e informacionit janë krijuar kur një numër i burimeve të informacionit janë të ndërlidhur ngushtë me njëri tjetrin. Ky rrjet është i shpërndarë, kur informacioni është i ruajtur në vende të ndryshme gjeografike dhe është i decentralizuar kur nuk ka kontroll të një nyje në një tjetër. Ideja origjinale e internetit ishte rritja në mënyrë të përshkallëzuar të rrjeteve të informacionit, duke lejuar pronarët e informacionit (autorët ose mirëmbajtësi i informacionit), të shërbejnë atë direkt dhe të krijohet një nyje e rrjetit me shumë pak përpjekje. Mungesa e përshkallëzimit të sistemeve të centralizuara të informacionit është për shkak të faktit se shpenzimet e menaxhimit rriten më shumë, linearisht me sasinë e informacionit të ruajtur dhe me sasinë e njerëzve të përfshirë. Kjo krijon një pikë ngopje, ku duke shtuar më shumë burime të sistemit (koha, njerëzit, kosto), më shumë rrit shpenzimet sesa rrit vetë ata. Ideja revolucionare e asaj që bëhet e njohur si world wide web, ishte vendosja e kostove të menaxhimit drejtpërdrejt mbi pronarët e informacionit në shkallë lokale shumë të vogël. Kjo u bazua në provat që kostot e menaxhimit të individëve të vetëm mbi informacionin e tyre, rritet në mënyrë lineare për aq kohë sa ata mbeten të vetmit për të menaxhuar këtë informacion.

Kur informacioni bëhet njohuri

Supozoni se ju dhuroni librin tuaj të preferuar. Sa informacion keni kaluar? (është një çështje e numërimit të karaktereve). Sa njohuri keni kaluar? (Nuk ka përgjigje). Njohuria varet kryesisht nga konteksti, i cila është përcaktuar si bashkimi i njohurive që tashmë kemi. Informacioni duhet të analizohet semantikisht para se të përkthehet në njohuri dhe analiza semantike është e pamundur të bëhet në qoftë se ju nuk keni instrumentet për të kryer atë, ose thjesht sepse konteksti, njohuritë e mëparshme nuk janë në gjendje të krijojnë lidhjet me informacionin e ri. Në qoftë se ajo nuk mund ta integroj atë, ajo mbetet pa lidhje dhe nuk mund të shtojë njohuritë tuaja të përgjithshme. Informacioni do të jetë i padobishëm (nganjëherë referohet si ndotja e të dhënave).

Web-i si semantikë

Kur ju shfletoni një faqe, browser-i juaj merr një numër bitesh nga rrjeti dhe i interpreton ato në faqet e internetit. Çfarë do të bëni me to? Ju i lexoni ato, ju përpiqeni ti kuptoni ato dhe të veproni si pasojë. Interneti është tashmë semantikë kur konteksti juaj është mjaft afër të qenit në gjendje për kryerjen e një analizë semantike në të, por, në qoftë se jo, informacioni që ju jep browseri nuk do të thotë asgjë për ju dhe nuk do gjeneroj asnjë veprim. Kështu që rezulton krejtësisht i padobishëm. Për shembull, duke shfletuar faqet anglisht në qoftë se ju nuk kuptoni anglisht ose duke shfletuar faqet kineze në qoftë se ju nuk kuptoni, ku ideogramat janë dy shembuj të shkallëve të ndryshme të dijes të përcjella nga informata web në varësi të kontekstit semantikë të përdoruesit.

Për anglisht, ju mund të jeni në gjendje për të marrë më shumë nga ajo, sepse të paktën alfabeti është pjesë e njohurive (nëse jetoni në një vend perëndimor), ndërsa për kinezët kjo është shumë e vështirë. Analiza semantike në përmbajtjen e web nuk kryhet vetëm nga njerëzit: Programe kompjuterike lëvizin në web duke marrë informacion për programe të tjera që kryejnë analiza semantike për ta indeksuar atë. Ky informacion është përpunuar pastaj drejtpërdrejt nga njerëzit ose nga programe të tjera kompjuterike në nevojë. Duhet të theksohet se për sasi të vogël të informatave, procesi i analizës semantike mund të kryhet nga njerëzit, por në shkallë shumë të madhe, detyra duhet të kryhet nga programet kompjuterike për të mbajtur shpenzimet të arsyeshme dhe të përputhshme me shkallën e rritjes së rrjetit apo frekuencës së përditësimit.

Motorët e kërkimit kryejnë analiza semantike në faqet e internetit, dhe algoritmet që kryejnë këtë analizë semantike janë shumë komplekse dhe orientuesi i tyre është i bazuar në interpretimin HTML, meta-informacionin, topologjinë e lidhjeve dhe analiza gjuhësore natyrore. Një pjesë e problemit është se asnjë program e inteligjencës artificiale nuk është në gjendje që të përputhet me aftësinë e brendshme të njerëzve për të kryer analiza semantike, të nxjerrë kuptimin nga informacioni dhe konceptet e abstraguara nga shembujt. Por problemi i vërtetë është se shumica e përmbajtjes së gjetur në internet është bërë nga faqet HTML dhe HTML është projektuar për të përmbajë udhëzime për shfletuesit dhe se si të shfaqin faqet për përdoruesit. Ai përmban shumë pak informacion në lidhje me semantikën reale të faqes, e cila nuk ka (nëse faqet janë shkruar në HTML direkt), ose ka humbur kur transpozohet nga formate të tjera.

Kapacitetet dhe kufizimet semantike në HTML

Le të marrim një shembull për të paraqitur se sa e varfër është HTML për të shprehur semantikë jo-vizuale. Le të konsiderojnë tagun . Ajo nënkupton bold, i cila është një stil i veçantë fonti që është më i trashë dhe më i errët se

stilinormal. Ky stil është përdorur normalisht për të bërë fjalët të theksohet në kontekstin vizual të një fjalisë të caktuar. Pra, kur shkrimtarët HTML donin një pjesë të fjalisë së tyre në mënyrë të theksuar, ata përdorin tag sepse ata e dinë se si kjo do të jepet nga shfletuesit. Tani, supozojmë që ju dëshironi të lexoni këtë faqe për përdoruesit me aftësi jo-vizuale (ose vizuale të dëmtuara ose të kërkojnë pajisje audio), çfarë do të thotë tagu ? Gabimi është delikat, por i qartë: aspekti më i përdorur nga një koncept abstrakt është përdorur për të treguar vetë konceptin abstrakt.

Tagu duhet të ketë qënë përdorur pasi ai përfaqëson një koncept që është krejtësisht i pavarur nga konteksti i përdorimit dhe do të interpretohet si bold për tekstin, etj. HTML përmban një numër të etiketave që mund të përdoren për të përfaqësuar një përmbajtje abstrakte në të parë, por këto janë të dizenuara për të edituar faqet, kështu që ato përmbajnë semantikë për paragrafe, tabela, imazhe, por ata nuk mund të editojnë elemente të tilla si faturat, formula matematike, etj. HTML është e rëndësishme për të përshkruar faqet e internetit dhe një përdorim i mirë për editimin e saj, e shoqëruar me përdorimin e Cascading Style Sheets për të ndarë stilet e attributeve nga struktura e faqes. Problemi është se HTML trajton çdo informacion si faqet. Faqet janë të rëndësishme për shfletim, por nuk janë të mira për data mining; kur ju kërkoni një faqe ju nuk interesoheni për stilin apo strukturën e faqes, por vetëm për informacionin që ajo përmban.

Modeli XML

Shtimi i semantikës të HTML do të fragmentonte përdorimin e saj dhe do të rriste mospërputhjet mes implementimeve të klientit, duke shkaktuar më shumë dëme për web-in. Pra, W3C ka filluar në vitin 1997, me krijimin e infrastrukturës së re teknologjike për përmbajtjen e web, që do të ishte zgjeruar mjaftueshëm për të përcjellë informacion me semantikë specifike, përputh funksionalitetin për nevojën e web-it të sotëm si dhe për ta përmirësuar, por duke e mbetur aq e thjeshtë sa është e mundur. Pas shumë vitesh dhe dhjetra specifikimeve, modeli XML nuk është përfunduar ende, por është mjaft afër pikës ku vëmendja mund të silllet për problemet teknologjike që ky model paraqet web-in si një i tërë.

Anatomia e modelit XML

Modeli XML mund të kategorizohet si më poshtë:

- > Sintaksa (XML)
- > Kufizimet e validimeve (XML, XMLSchema)
- > Modulariteti (hapësira)
- > Aftësitë e hyperlinkut (XLink, XPointer, XInclude, XBase)
- > Metadata (RDF, RDFSchema)

Sintaksa

Sintaksa XML përcakton se si informacioni duhet të jetë i koduar për t'u konsideruar i mirë-formuar nga *parserat*. Sintaksa XML është bazuar në tekst dhe bazohet në të njëjtën sintaksë të përdorur nga HTML, por ajo e përmirëson atë duke treguar në mënyrë specifike elemente bosh për të lehtësuar kompleksitetin e parsimit dhe duke lejuar operacionin e rritjes.

Kufizimet e validimit

Dokumentet XML konsiderohen të mirë-formuar në qoftë se ata ndjekin sintaksë XML, por ato mund të jenë edhe më tej "të vërtetuar" në disa kufizime të validimit që mund të tregojnë strukturën e elementeve, llojet e vlerave, përmbajtjen e parazgjedhur, hapësira të lejuara, etj. Dokumentet XML konsiderohen të vlefshme në qoftë se ata janë të mirë-formuar dhe në qoftë se skemat e tyre janë ndjekur pa gabime.

Modulariteti

Modularitet në kontekstin XML është aftësia për të bashkuar përmbajtjen me kontekste të ndryshme semantike në të njëjtin dokument. Për shembull, përziejnë pamjen e faqes me formulat e matematikës dhe grafikë vektoriale për punime shkencore. Ky kapacitet është arritur duke lidhur çdo skemë me një URI të veçantë dhe unike, duke shmangur përplasjen e emrave dhe ruajtjen e pavarësisë semantike. Kjo është një tjetër risie modelit HTML, i cili nuk kishte aftësi të tilla më parë.

Aftësitë hyperlink

HTML përdor një etiketë të veçantë, modeli XML përdor nocionin e attributeve *namespaced* për të shtuar semantikë të veçanta të elementeve ekzistuese. Ky është një përmirësim i madhë pasi lejon shtimin e hyperlink në çdo skemë-dokument pa kërkuar krijimin e elementeve të veçanta lidhëse. Kjo është një qasje shumë më e saktë dhe e fuqishme sesa HTML.

Metadata

Metadata është e dhënë në lidhje me të dhënat. HTML shtoj tagun **meta** për të lejuar programet e kompjuterave që të ndihmojnë autorët e faqeve në indeksimin e faqeve, duke e vendosur fjalë kyçe që specifikojnë përmbajtjen e faqes. Mekanizmat e kërkimit me të mirë përdorin gjerësisht këto tague, kur janë në dispozicion, për të bërë kategorizimin e informacionit. Modeli XML shtrin këtë aftësi duke përcaktuar një framework për përcaktimin e burimeve (RDF), të cilat mund të konsiderohet si sintaksë për metadata, së bashku me skemat metadata (RDFS) që përcaktojnë strukturën e metadata dhe lidhjen e saj me skemat e tjera.

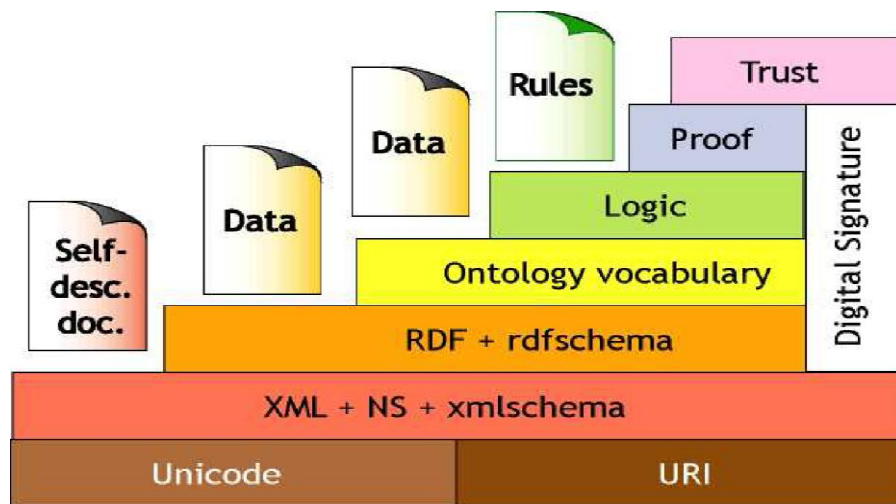


Figura 1. Anatomia e modelit XML

Arkitektura e makinës

Qëllimi përfundimtar i një semantike web është që të lejoj programet kompjuterike të ndihmojnë njerëzit duke bërë analiza semantike për ta, në grupe shumë të mëdha të informacionit dhe për të përmirësuar eksperiencën e përdoruesit në data mining.

Le të marrim një shembull për një faqe HTML:

```
<html>
<body>
<h1>My trip to the Java Island</h1>
<h3>by Stefano Mazzocchi (stefano@apache.org)</h3><p>The trip on the
island was great:
<ul>
<li>we tasted great Java coffee</li>
<li>we finished my latest program written in Java</li><li>we did a great trip on
the islands around Java</li></ul>
</p>
<p>The trip was arranged by Travels@Java.com Inc.
(info@javatravels.com).</p><p>Saluto tutti gli amici italiani che erano con
me.</p>
</body>
</html>
```

Analiza semantike e kësaj faqeje është shumë e vështirë për njerëzit dhe afërsisht e pamundur për programet kompjuterike (të paktën në ditët e sotme). Në vazhdim shkruajmë:

```
<?xml version="1.0"?>
```

```

<page xmlns="http://mysite.org/page"
xmlns:geo="http://geography.gov/terms"
xmlns:food="http://fao.org/metadata/food/en"
xmlns:man="http://onu.gov/metadata/mandkind/en"
xmlns:man="http://acm.gov/metadata/computing"
xmlns:email:http://ietf.org/schemas/email"
xmlns:com:http://nafta.org/terminology/en" xml:lang="en">
<title>My trip on the <geo:island>Java Island</geo:island></title><author>
<man:man man:age="25" email:address="stefano@apache.org"> Stefano
Mazzocchi
</man:man>
</author>
<content>
<p>My trip on the island was great: <ul>
<li>we tasted great <food:coffee>Java</food:coffee> coffee</li>
<li>we finished my latest program written in
<comp:lang>Java</comp:lang></li><li>we did a great trip on the islands around
<geo:island>Java</geo:island></li></ul>
</p>
<p>The trip was arranged by<com:companycom:type="incorporated"
email:address"info@javatravels.com"> Travels@Java.com Inc.
</com:company>
</p>
<p xml:lang="it">Saluto tutti gli amici italiani che erano con me.</p></content>
</page>

```

Kjo faqe përdor XML dhe hapësira për të shtuar informacione specifike semantike në lidhje me përmbajtjen e përfshirë, por heq të gjitha informacionet e stilit. Ndërsa për njerëzit është e vështirë që të kuptojnë këtë faqe se si është, për programet kompjuterike kjo është shumë më e lehtë pasi të gjitha informatat e nevojshme semantike janë vendosur nga autori i faqes dhe hapësirat identifikojnë lidhjet semantike *njëzëshëm*. Asnjë deduktim nuk ndodh kur një program kompjuterik kompilon këtë faqe për shkak se ajo nuk ka nevojë që ta kuptoj këtë faqe që të jetë në gjendje, të lidh përmbajtjen me zonat e saj semantike, të cilat janë identifikuar nga hapësirat.

Rrjetet semantike si skema të metadata

Rrjetet semantike janë rrjetet e termave që lidhin semantikën bazuar në përvojën e jetës tonë. Në mënyrë që për një program kompjuterik të kuptoj se *ordine* do të thotë rregull, ajo duhet të ketë informacion që lidh keto dy fjalë bashkë, lloj mekanizmi i rrjetit që programi është në gjendje të përdori me rregulla të thjeshta për të dalë në konkluzion me kuptimin e duhur. Pra, në qoftë se motori i

kërkimit është i vetëdijshëm për faktin se *ordine* do të thotë rregull, metadata është shkëputur nga gjuha e përdorur që kur elementi *ordine* trashëgon të gjitha kuptimet e elementit rregull, të cilat supozohen të njohura nga përdoruesi. Specifikimi i skemas RDF përcakton një mënyrë për të treguar lidhje mekanike semantike mes skemave metadata dhe të lejoj motorët e kërkimit të jenë më fleksibile gjatë kërkimit.

XML si hapi i parë

Modeli XML është qartësisht një model shumë më mirë se ai i krijuar nga HTML dhe përmirëson të gjithë sistemet e informacionit. XML përballet me shumë probleme, i pari është kompleksiteti i vet dhe inkoherencat e dizajnit midis **specs**. Inkoherencat midis **specs** do të hiqet gjatë viteve, kompleksitetin i saj nuk mund ulet pa ndryshuar modelin.

Problematika

Problemi më i madh i semantikës web që përballet sot është mungesa XML në shfletues. Shfletuesi kryesor Microsoft IE performon shumë pak në aspektin e pajtueshmërisë së standardeve (edhe pse është versioni më i fundit) dhe rivali I tij kryesor Mozilla ende nuk është gati për këtë. Kjo i detyron ato site të përdorin XML për infrastrukturën e tyre web për ta fshehur atë pas transformimeve të dërguara për të përshtatur përmbajtjen HTML. Për më tepër, nuk ka asnjë mënyrë për motorët e kërkimit për të identifikuar veten e tyre në faqet e internetit, kështu që nuk ka asnjë mënyrë për faqet web për të ushqyer përmbajtjet XML në vend të HTML, të krijuara nga përmbajtja e tyre e fshehur XML. Shumë vende kanë filluar të përdorin teknologjitë XML në anën e serverit, sepse kjo i ndihmon ato për të reduktuar kostot e tyre të menaxhimit.

Shtimi i dimensioneve në web

Një zgjidhje e mundshme është bazuar në provën se të gjitha burimet e internetit janë në thelb një dimensionale në modelin HTML: kur të kërkoni për një URL të veçantë, ju jeni duke kërkuar URL për shfletim. Në kornizat botuese, edhe në URL-të statike nuk janë lidhur drejtpërdrejt me file të ruajtura në sistem, por është krijuar një output tjetër në varësi të përdoruesit që kërkon burimet.

Për shembull, agjentët WAP janë ushqyer me përmbajtje WML, shfletues HTML me HTML, klientët zanor me VoXML dhe kështu me radhë. Por edhe në qoftë se rezultati është i ndryshëm në varësi të parametrave të kërkesës, të gjithë klientët janë në gjendje të hyjnë në një dhe vetëm një pamje të burimit: pikëpamjen që u lejon atyre të shfletojnë.

Zgjidhja është zgjerimi i dritares URI për të lejuar kërkuesin të përcaktoj se për cilën pamje është i interesuar. Shfletuesat së shumti do të kërkojnë pamjet

shfletuese (që do të kthejnë përmbajtjen semantike vizuale, lidhëse dhe të asocuar), por ka shumë shikime të mundshme të një burimi:

- Pamje e parazgjedhur (default) - kthen pamjen kryesore të burimit (dhe kryesisht varet nga përdoruesit që kërkojnë faqen)
- Pamje lidhje - kthen listë të lidhjeve të përfshira në burim. Kjo është shumë e dobishme për *crawlers* që nuk kanë për të kryer analiza dhe mund të kthehen nga URI.
- Pamje semantike - kthen pamjen semantike të burimeve.
- Pamje view - kthen skemën e përdorur për të validuar burimin.
- Pamje metadata - kthehet informacion se si të lidhësh skemën e dhënë metadata me atë që gjendet në përmbajtjen e faqes.
- Pamje pipeline - kthen informacione për çfarë veprimesh bëhen për të krijuar burim të veçanta së bashku me loget, kohën dhe kompilimin e
- Lidhja-përktimi i pamjes - kthen pamjen e paracaktuar, por të gjitha lidhjet janë përktyer në përputhje me rregullat e dhëna në POST nga klienti.

Avantazhet e XML në web semantikë

XML ofron një sintaksë thelbësore për strukturën e përmbajtjes në dokumente. XML nuk është aktualisht një komponent thelbësor i teknologjive semantike web në shumicën e rasteve, si sintaksë alternative, të tilla si turtle, por nga analiza që kemi bërë deri tani XML është pjesë thelbësore në aksesimin e shpejtë të informacionit në web dhe standartizim të përmbajtjes në web.

XML është duke u përdorur për një numër qëllimesh të ndryshme si:

- si një sintaksë bazë për gjuhë të tjera markup
- si semantikë markup e faqeve web
- si një format uniforme e të dhënave të këmbimit

XML sjell kriteret e mëposhtme për shërbimet web XML në lidhje me arkitekturën:

- Zgjatshmëri: Një sistem mund të funksionoj si duhet pa humbur funksionet kryesore gjatë një update-i.
- Neutraliteti: Kufizimet e kërkuara të një aplikimi janë të kufizuara.
- Struktura: XML paraqet edhe strukturën dhe përmbajtjen e dokumentit, ku ofron kontroll rritje të granularitetit të informacionit nëpërmjet transformimit dhe query.
- Ndërveprimi: Komunikimi dhe shkëmbimi i të dhënave ndërmjet sistemet heterogjene janë të mundshme.

Skenarët e mundshëm në të ardhmen

Ka disa skenarë të mundshëm në të ardhmen që mund të ndodhin:

XML në anën e klientit

Nëse të gjithë klientët do të mbështesnin XML, semantika web është vetëm një çështje ose dokument standardizues dhe skema metadata dhe i lidhin ato me hapësira të përshtatshme. Problemi është, siç është përshkruar tashmë, që nuk do të jetë e lehtë për të gjithë klientët e internetit për të mbështetur XML. Kjo ka shumë gjasa të krijoj një tërheje inerciale që do të ngadalësojnë miratimin e teknologjive të reja web.

XML në anën e serverit

XML do të miratohet në anën e serverit së shpejti, sepse ajo tashmë mundëson reduktimin e kostos në të dy aspekte: krijimit dhe mirëmbajtjen e informacionit të faqes. Motorët web janë në gjendje të mbështesin: agjentët e përdoruesit XML. Një web i tillë do të lejoj kërkime më të mira në nivel lokal në site, por nuk do të përmirësoj data mining në shkallë globale.

Burimet shumëdimensionale web

Kjo përmirëson në skenarin e mëparshëm duke futur mënyrë të standardizuar për faqet për të ndarë informacione semantike. Kjo do të largoj problemet në qoftë se aftësia e indeksimit është shtuar në informacionin semantikë. Kjo do të ketë miratimin e teknologjive semantike dhe kjo do të lejoj miratimin inkremenal.

Sfidat

XML mund të shërbejë si një metadata për të lehtësuar ndërveprimet dhe integrimin e ardhshëm të semantikës web për akses të të dhënave të pastruara.

Literatura

Bray T., Paoli J., Sperberg C.M. (1989): Extensible Markup Language (XML) 1.0
Erdmann M., Studer R.(1999): Ontologies as Conceptual Models for XML Documents. In: Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management
Shiyong Lu., Ming D., Farshad Fotouhi.(2002):The Semantic Web: opportunities and challenges for next-generation Web applications
Stefano M.(2006): Toward the semantic web a view of XML from outer space,
<http://www.w3.org/TR/REC-xml>