

Igor Ruiz Agúndez

Investigador S<sup>3</sup>Lab

Pablo García Bringas

Investigador Principal S<sup>3</sup>Lab



DeustoTech-Tecnológico  
Fundación Deusto

# Web Semantikoari sarrera

**Laburpena**—Web mundua asko aldatu da sortu zen momentutik, egunez egun irauliz. Gure artera heldu den azkenengo iraultza Web Semantikoa dugu. Artikulu honetan Web Semantikoa zer den eta zertarako balio duen aztertzeaz gain, zein elementuk osatzen duten ikertuko dugu. Halaber, Web 2.0-arekin duen erlazioa eta teknologia berri honen aurrean sorturiko ikuspuntu kritikoak ikusiko ditugu.

**Index Terms**—Informatika, Internet, Web Semantikoa.

## Sarrera

Denbora asko pasa da Tim Berners-Lee-k 1990ean *World-WideWeb* programa idatzi zuenetik. Hasiera batean dokumentuak elkarri lotzeko sortu zena, urteak igaro ahala handitzen eta handitzen joan da. *Hypertext link*-en bidez edo zer erreferentzia daiteke, edozein motatako objektuak elkarlotuz eta informazio espazio unibertsal bat sortuz.

Baina sarearen sortzaileek amets gehiago zituzten burutzeke. Sarean dagoen informazio guztia pertsonentzat ulergarria izateaz gain makinentzat ere ulergarria izatea nahi zuten. Honek informazioaren erabilera erraztuko luke, ordenatzeko, analizatzeko, esanahia emateko edota kategoriari ebazteko orduan. Makinak eta gizakia elkarlanean aritzeko bidea litzateke, alegia.

Behar hau asetzeko intentzioaz heldu da Web Semantikoa gure artera. Termino hau ulertzeko semantika zer den definitu beharra dago lehenengoz. Semantika, (semantikoues) edo garrantzitsua[1], grekotik datorren hitza da. Linguistikaren arabera,

pertsona eta gizartearen ikurren interpretazioan datza. Informatikaren zientzian ostera, esanahi hau zabaldu egiten da. Aztergai dugun kasuan, semantikak sarean dauden edukiak pertsona eta makinei ulertarazten dizkien elementua da. Web Semantikoa gaur egun erabiltzen dugun *World Wide Web*-aren luzapen bat da, non informazioaren semantika eta eskaintzen zaizkigun web zerbitzuak definiturik dauden [2]. Honi esker, bai pertsonak eta baita makinek ere sarean dagoen edukia uler dezakete. Laurogeita hamarreko hamarkadan, Tim Berners-Lee-ren hitzetan[3]: *Sarean dauden datu guztiak, edukia, estekak, eta pertsonen eta makinaren arteko eragiketak, analizatzeko gai izatea dut amets. Web Semantikoa-k hau egiztatuko luke. Zoritarrez oraindik sortzeke dago, baina gauzatzean, eguneroko merkatu mekanismoak, burokrazia eta gure eguneroko bizitza euren artean hitz egiten duten makinek kudeatuko dituzte. Batzuek agente inteligente-etan[4] luzaroan egindako lana azkenik gorpuztuko da. Bete nahi duen helburu garrantzitsuena datuak aplikazio, enpresa eta komunitateen artean konpartitzeko gaitasuna eskaintzea da. Elkarlan hau modu automatikoan edo eskuz prozesatuko litzakete, datuen arteko erlazio berriak azaleratuz.*

## Web semantikoaren elementuak

Lehenago aipatu den bezala, Web Semantikoaren oinarri garrantzitsuena interneten dauden datuen arteko erlazioen definizioan eta deskribapenean dago. Hori da gaur egun erabiltzen diren estekekin duten ezberdintasuna. Erlazioak edozein bi elementuren artean izan daitezke eta ez dago Web orrialde aktiboaren edo erabiltzearen dagoenaren kontzepturik. Gainera, erlazioak berak ere izendaturik daude. Aldiz, Web tradizionalen ez daude definituta eta

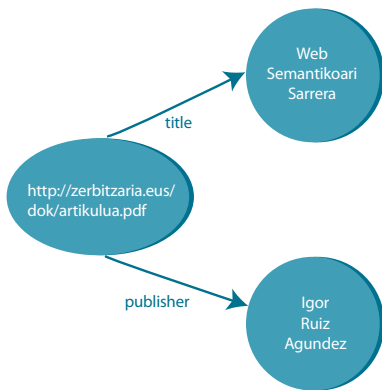
irakurleak esanahia testuinguruaren arabera igarri behar du. Erlazio hauen definizioak datuen elkartruke edo permuta automatiko hobea dakar. Elkartrukatze hau *Resource Description Framework (RDF)*-en bidez era formal batean definiturik dago. *RDF*-ak gaur eguneko esteken egitura zabaltzen du, estekari eta honen muturreko elementuei izena emanez. Hau *triples* edo hirukun izenaz ezagutzen da. Egitura honi esker datuak nahastu, ageri eta aplikazioen artean konpartitu daitezke. Modu matematiko batean ikusirik, egiturak grafo zuzendua eta etiketatua osatzen du. Grafoaren nodoek elementuak adierazten dituzte eta hauen arteko erlazioa zuzenduriko gezi izendatuak adierazten da. Eredu hau *RDF*-ak adierazteko modurik sinpleena da eta irudikapen grafiko bat ematen digu. Ikus dezagun *RDF* sintaxiaren adibide bat 1. zerrandan eta bere adierazpen grafikoa 1. irudian.

```
1 <rdf:RDF xml n s:
  rdf = "http://www.
  w3.org/1999/02/22_
  rdfsyntax_ns #"
2 <xmlns:dc="http://purl.org/
  dc/elements/1.1/">
3 <rdf:Description
  rdf:about="http://
  zerbitzaria.eus/dok/
  artikulua.pdf">
4 <dc:title>Web Semantikoari
  sarrera</dc:title>
5 <dc:publisher> Igor Ruiz
  Agundez </dc:publisher>
6 </rdf:Description>
7 </rdf:RDF>
```

Listing 1. RDF baten adibidea

Web Semantikoaren beharrak asetzeko hainbat tresna sortu dira, hona hemen hauek sailkatzeko proposamen bat:

- Erlazioek deskribaturiko informazioa eskuratzeko tresnak. (Adib., *SPARQL*)



**Irudia 1.** Adibideko RDF-aren adierazpen grafikoa

- Erlazio hauek eta lotzen dituzten elementuak modu zehatz batean klasifikatzeko eta karakterizatzeko tresnak, eraginkortasun automatikoa handituz. (Adib., *RDF Schemas*, *OWL*, *SKOS*)
- Elementuen arteko erlazio logikoak definitzeko tresnak erlazio berriak eta arrazonamenduetan mami eza aurkitzeko erabiltzen dira. (Adib., *OWL*, *Rules*)
- Datu iturriak ateratzeko eta beste datuekin elkartrukatzea bideratzeko tresnak. (Adib., *GRDDL*, *RDF*)

*RDF*-ek osatzen duten oinarriko zutabe honen inguruan beste elementu asko aurkitzen ditugu, Web Semantikoaren pila deritzona osatzen dutelarik. 2. irudian adierazita daude, ondoren xehetasun handiz aztertzen dira euren garapen mailaren arabera:

### A. Hypertext Web-aren teknologiak

- *IRI* (*Internationalized Resource Identifier*), *URI* (*Uniform Resource Identifier*)-ren orokortze bat da. Web Semantikoa osatzen duten elementuak instantzia bakarrean identifikatzeko erabiltzen da.
- *Unicode* testuak hainbat hizkuntzatan adierazteko erabiltzen da. Web Semantikoari hizkuntza ezberdinetan idatziriko dokumentuak lotzen laguntzen dio.
- *XML* (*Extensible Markup Language*) helburu orokorreko marka-hizkuntzak espezifikatzeko estandarra. Elementuak definitzeko askatasuna ematen du, datuen estrukturatzea bideratuz. Web Semantikoaren sintaxiaren arauak zehazten ditu.

### B. Web Semantikoaren estandarizaturiko teknologiak

- Lehenago aipaturiko *RDF* (*Resource Description Framework*)-a subjektu-predikatu-objektu erako espresioak sortzeko erabiltzen da, *triples* edo hirukunak sortuz. Elementuak eta hauen arteko estekak berak definitzen ditu.

Bere definiziorako *XML*-ren sintaxia erabiltzen du. Web Semantikoaren datuen arteko elkartruckerako erabiltzen da.

- *RDFS* (*RDF Schema*)-k *RDF*-an oinarrituriko elementuak definitzeko erabiltzen den hiztegi bat da. Propietateak eta klaseak hierarkia orokor batean emanez (Adib. *FOAF* pertsonak definitzeko eta *DOAP* proiektuak definitzeko) Web Semantikoan erabiltzen diren taxonomiak, hau da, klasifikazioak eratzen ditu.
- *OWL* (*Web Ontology Language*)-ek Logika deskriptiboan oinarrituz hiztegia gehiago definitzen du, klaseen arteko erlazioak, kardinalitatea, berdintasuna, propietateen ezaugarriak eta zerrendak, adibidez. Web Semantikoan erabiltzen diren ontologiak, hau da, domeinu baten barruko elementuen arteko erlazioak eratzen ditu.
- *SPARQL* (*Simple Protocol and RDF Query Language*) *RDF*-ei galderak egiteko hizkuntza da. Web Semantikoan datuei buruzko galderak egiteko erabiltzen da.

### C. Web Semantikoaren gauzatu gabeko teknologiak

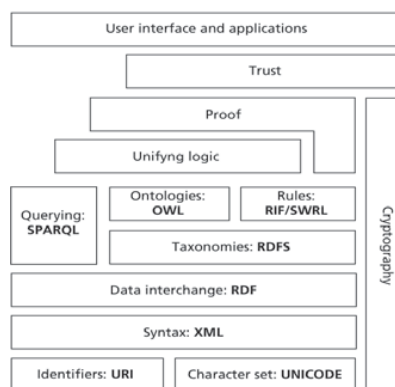
- *RIF* (*Rule Interchange Format*) edo *SWRL* (*Semantic Web Rule Language*), logikan oinarrituriko erregelak eta inferentzia motorren elkarlana bideratzeko sortua. Ontologiek objektuak makinenzako ulergarriak egiten badituzte, *RIF* erregelek informazio berria ontologia batetik nola inferitu behar den definitzen dute.
- *Kriptografiak* Web Semantikoan segurtasun geruza bat definitzen du. Edukiak kanal ez ziurretatik modu babestu batean transmititzeko gai izateko eta norik eginak diren egiaztatze erabiltzen da.
- Deribaturiko adierazpenenganako *konfiantza*, honela lortuko da: premisak egiaztaturiko iturrietatik datozela egiaztatuz eta informazio berria lortzean logika formala erabiliz.
- *User interface*-a edo makina eta gizakiaren arteko interfazea erabiltzaileek Web Semantikoa erabiltzeko azkeneko geruza da. Leiho moduan funtzionatuko du, internet ikusteko modu berri bat bideratuz.

### Web semantikoa eta web 2.0?

Web 2.0-ak erabiltzailearen parte hartzean datza, edukien sorreran lankidetzan interaktibo bat eraikiz. Sorturiko informazioa ez da modu erraz baten integratzen,  *mash-up*-en erabilera behartzen gaitu. Aldiz, Web Semantikoak erabiltzen dituen *RDF*-ekin ez da muga hau existitzen eta informazioa erantzun handiz itsasten da elkar.

Web Semantikoaren oinarriak erabiliz argitaraturiko informazioa makinek irakurri ahal izateaz gain, informazio honen berrerabilpenerako oztopoak ezabatzen ditu. *RDF*-ek ontologia ezberdinetatik datozen adierazpenak dokumentu bakarrean nahastu ditzakete.

Web 2.0-an erabiltzen diren etiketek (*tag*-ek) kutsu semantikoa ematen diote baina etiketa hauek *folksonomia*



**Irudia 2.** Web Semantikoaren pila

dira, hau da, hierarkia eta erlaziorik gabeko etiketatze soziala.

Web semantikoa aldiz, ontologietan oinarritzen da. Informatikarentzat ontologia bat domeinu definitu bat izanik, modu zehatz batean elementu guztiak antolatzean datza. Honen adibide famatu bat ingelesaren datu base lexikala edo *WordNet* dugu.

Ezberdintasun nabariena parte-hartze eremuan dago batik bat. Web 2.0-an eragile nagusia edukiak sortzen dituen gizakia da, blog baten idaztean adibidez. Eduki hau pertsonen artean *XML/RDF*-a erabiliz partekatuko da. Web Semantikoan, ordea, *OWL*-ekin elementu ezberdinen artean dagoen logika deskribatzen da, logika honen azkeneko erabiltzaileak makinak direlarik.

## Ikuspuntu kritikoak

Badira ere Web Semantikoaren aurrean kritika handiak egiten ari direnak, eta haien argudioak ere aipatu beharra dago.

Askoren ustez Web Semantikoak duen bideragarritasuna ez da oso handia gaur softwarearen ingeniariatzak dituen mugak direla eta. Gainera, iristear zegoen zerbaite bezala saldu izan da Web Semantikoak, eta momentuz ez da guztiz gauzatu.

Bada giza eskubideak arriskuan ikusten dituenik ere. Espero den Web Semantikoak heltzean oso erraza izango da sorturiko informazio guztia kontrolatzea. Makinek modu automatikoan edukiak blokeatu edo zentsuratu ahal izango lituzkete adierazpen askatasunaren kaltetan.

Momentuz web aplikazioek ez dituzte *RDF*-ak modu automatikoan sortzen. Beraz, eduki bat kaleratzeko orduan bi bertsio kaleratu behar dira: gizakiek ulertzeko bertsioa eta makinek ulertzeko bertsioa; azken honek denbora eta errekurtsio kontsumo handiagoa suposatzen du. Gainera eduki hau sortzeko aurretik prestaturiko erabiltzaile espezializatuak behar dira.

Azkenik, Web Semantikoak eraikitzekeo estandar berrien beharra zalantzan

jartzen dutenak daude. Euren ustetan, semantika gaur egungo baliabideekin eraiki daitekeela uste dute.

Hala ere, gaur egun ez da *HTML* hutsean eraikitako web semantikorik eraiki.

## Ondorioak eta etorkizuna

Ikusi dugunez, Web Semantikoak ekarpen asko dauzka gizarteratzeko. Makinak interneten dagoen informazioa ulertzeko gai diren momentua heltzean, egoera estarian dauden onura asko helduko dira gure eguneroko bizitzara. Web Semantikoak existitzen denaren hobekuntza bat dela gogoratu behar dugu, eta inoiz sor daitezkeen arazo guztiak konponduko ez dituen, internetek eskaintako baliabideak handituko ditu.

Oraindik lan asko dago egiteko, erabiliko diren estandarrak definitzen amaitu beharra dago eta hauen garapena eta ezarpena gauzatzeko dago. Hori bai, Tim Berners-Lee-ren ametsa benetakoa bihurtuko den eguna pixka bat hurbilago dago gaur.

Egun hori heltzean, gaur egun ezagutzen dugun *Internet*-a guztiz eraldatuko dela ziurta dezakegu.

## Eranskina A

### Terminoen hiztegia

*A. Hypertext*: Ordenagailu batean dagoen testua, irakurleak eskatzekotan erlazionaturiko informazioa eramateko gai dena. Honetarako hyperlink-ak erabiltzen dira.

*B. Hyperlink*: Dokumentu informatiko batean dagoen erreferentzia. Hau jarraituz gero, dokumentuaren beste atal batera eramango gaitu edota beste dokumentu ezberdin batera.

*C. HTML*: HyperText Markup Language, Web orrialdeak egiteko hizkuntza nagusia. Testuaren egituraketa bideratzen du.

*D. Mash-up*: Iturburu ezberdinak dituzten datuak erreminta bakar batean batzen dituen elementua.

*E. Etiketak (tag-ak)*: Hierarkiarik gabeko terminoa. Informazio bat

atxikitzen zaio hau definitzeko. Web 2.0-ari esker zabalduko hitza.

*F. Taxonomia*: Klasifikazioaren zientzia. Gauzak hierarkia jakin batean antolatzen ditu mota eta azpimotetan.

*G. Folksonomia*: Hierarkia eta erlaziorik gabeko etiketatze edo sailkatze sozial edo kolaboratiboa.

*H. Ontologia*: Informatikarentzat, domeinu definitu bat izanik, modu zehatz batean elementu guztiak antolatzean datza.

## Eranskina B

### Copyright oharrak

2. irudia ondoko[5] erreferentziaren Webgunetik lortu da egilearen baimenarekin. Gainerako irudi eta materialak egileak berak eginak dira.

## Erreferentziak

- [1] Diccionario gratuito en línea Griego-Español y Español-Griego. [Online]. Available: <http://es.freelang.net/enlinea/griego.php?lg=es>
- [2] Semantic Web - Wikipedia, the free encyclopedia. [Online]. Available: [http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web)
- [3] T. Berners-Lee and M. Fischetti, *Weaving the Web*, 1st ed. San Francisco: HarperOne, 1999.
- [4] Intelligent agent - Wikipedia, the free encyclopedia. [Online]. Available: [http://en.wikipedia.org/wiki/Intelligent\\_agents](http://en.wikipedia.org/wiki/Intelligent_agents)
- [5] Semantic Web Stack - Wikipedia, the free encyclopedia. [Online]. Available: [http://en.wikipedia.org/wiki/Semantic\\_Web\\_Stack](http://en.wikipedia.org/wiki/Semantic_Web_Stack)
- [6] The World Wide Web: A very short personal history. [Online]. Available: <http://www.w3.org/People/Berners-Lee/ShortHistory.html>
- [7] Página personal de Dr. Diego López de Ipiña. [Online]. Available: <http://paginaspersonales.deusto.es/dipina/>
- [8] Web 2.0 - Wikipedia, entziklopedia askea. [Online]. Available: [http://eu.wikipedia.org/wiki/Web\\_2.0](http://eu.wikipedia.org/wiki/Web_2.0)
- [9] W3C Semantic Web Activity. [Online]. Available: <http://www.w3.org/2001/sw/>
- [10] W3C Semantic Web FAQ. [Online]. Available: <http://www.w3.org/RDF/FAQ>