

web

IGOR LETURIA AZKARATE  
Informatikaria eta ikertzailea

## SEMANTIKOA

## BADAUDEN ETA BEHARKO DIREN TEKNOLOGIAK



123RF

**Maiatzeko artikuluan, Sir Tim Berners-Leek asmatutako World Wide Webak eta haren oinarrian dagoen HTML formatuak dituzten arazoak ikusi genituen, eta azaletik azaldu genuen zer den Berners-Leek berak proposatutako konponbidea, web semantikoa, alegia. Honetan, web semantikoaren azpian dauden teknologiak eta adibideak ikusiko ditugu, eta hura egia bihurtzeko dauden arazoak ere bai.**

Web semantikoan, objektuak, pertsonak... eta haien arteko erlazioak deskribatzen dira etiketa bidez. Etiketetan, orriaren itxura eta egitura azaldu beharrean, orriko elementuen esanahia jasotzen da. Horri esker, HTML sarearekin batera existituko litzatekeen sare paralelo bat sor daiteke, makinek ulertzeko moduko ezagutza-base bat, semantika adierazteko formatuetan kodetua. Behin makinek ulertuta, modu eraginkorren tratatu ahal izango lukete informazioa, eta milaka aplikaziotarako bidea ireki.

**TEKNOLOGIAK: RDF, OWL...**

Baina, horretarako, amaraunean dauden kontzeptu guztiak definitu ahal izateko, etiketatze semantikoko eskema eta formatuak behar dira. W3Ck formatu horiek definitu ditu hainbat estandarretan; RDF eta OWL dira garrantzitsuenak eta ezagunenak.

RDF (*Resource Description Framework*) baliabideak deskribatzeko balio duen formatu bat da, XMLn oinarritua. Hiru elementutan oinarritzen da: baliabideak, propietateak eta propietateen baliok. Baliabidea deskribatuko den hori da, eta URL batez (web identifikatzaile edo helbide batez) identifikatzen da. Propietatea baliabidearen ezaugarri bat da, deskribatu nahi dena. Eta baliok deskribatu nahi diren ezaugarrien balio konkretuak dira (ikus hurrengo orriko adibidea).

Era horretara, nahi duguna deskriba dezakegu. Baina gauza-mota bakoitza (pertsonak, musika taldeak, liburuak...) deskribatzeko erabili beharreko etiketetan ados jarri beharra dago; bestela, makinek ulertu gabe jarraituko lukete. Horretan laguntzen du OWL (*Web Ontology Language*) lengoaiak. OWLk ezagutza-arlo edo bizitzaren esparru jakin bateko objektu edo entitateak nola deskribatuko diren definitzea ahalbidetzen du.

**ADIBIDE ERREAL BAT: RSS FORMATUA**

Web semantikoaren ahalmenaren adibide txiki bat geure artean dugu aspalditik: blogek hasieratik, eta, gaur egun, Interneteko bestelako albiguneez ere erabiltzen duten RSS (*Really Simple Syndication*) formatua. Izatez, RDF mota bat da (jatorrizko izena *RDF Site Summary* da), albisteak deskribatzen espezializatutakoa. Bloggek berrikuntza handia ekarri zuten, informatika edo HTML ezagutza teknikorik izan gabe Interneten edukia sortzeko aukera eman baitzioten erabiltzaileari, eta jende berri asko hasi

## Elhuyar Zientzia eta Teknologia semantikoa



```
<RDF>
<Description about="http://www.zientzia.net/elhuyar.asp?Alea=253">
  <izenburua>Elhuyar Zientzia eta Teknologia</izenburua>
  <alea>253</alea>
  <egilea>Elhuyar Fundazioa</egilea>
  <data>2009ko maiatza</data>
  <artikuluak>
    <artikuluua resource="http://www.zientzia.net/artikuluua.asp?Artik_kod=14531" />
    . . .
  </artikuluak>
</Description>
<Description about="http://www.zientzia.net/artikuluua.asp?Artik_kod=14531">
  <izenburua>Web semantikoa: Interneten benetako ahala iristeko dago
  oraindik</izenburua>
  <egilea>Igor Leturia Azkarate</egilea>
</Description>
. . . .
</RDF>
```

zen Interneten testuak jartzen. Baina blogek ez zuketzen halako arrakasta izango RSS formatuagatik ez balitz.

Izan ere, blogak HTML formatuan soilik publikatu izan balira, hainbat blogen gaiak atsegin dituen irakurle batentzat ez litzateke erraza izango haiei jarraipena egitea. Aldian behin horietako guztietara sartu beharko luke, zerbait berririk dagoen ikusteko. Eta lan hori, gainera, askotan ezer berririk ez egoteko, edo ez gogoratzeko azkeneko aldian zein irakurri genituen... Azkenean, ezingo luke blog gutxi batzuen jarraipena baino egin.

*Blog eta albisteentzako etiketatze semantiko sinple batek hori egin badu, zer ez da gertatuko beste edonolako kontzeptuak ere semantikoki etiketatzen direnean...*

Baina blogek, pertsonentzako HTML bertsioaz gain, makinentzako RSS bertsioa ere bazuten. Bertsio horretan, azken sarrera edo artikulua agertzen ziren, horietako bakoitza etiketen bidez ondo ezberdinduta, eta bakoitzaren izenburua, egilea, data, laburpena, esteka eta abar ondo egituratuta, makinek ulertzeko moduan. Hala, RSS-irakurgailuak sortu ziren, norberak

Honela deskribatu ahal izango litzateke *Elhuyar* aldizkariaren ale bat eta haren artikulua RDF bidez (hau modu sinplifikatu eta ulerterrazean erakutsita dago, RDF formatuan beste etiketa-izen batzuk erabiltzen dira).

atsegin dituen blogen jarraipena egiteko. Irakurgailuak gure blog gogokoan RSSen jarraipen periodikoa egiten du, eta erabiltzaileari azkeneko sartu zenetik dauden berriak soilik erakutsi; horrela, hamarka edo ehunka blogen jarraipena egin daiteke. Eta blogetan espezializatutako bilatzaileak ere sortu ziren, eta RSSak batu eta filtratzeko zerbitzuak, egunkari eta aldizkarien webguneek ere jarri zuten RSSa, sare sozialek ere bai... Web 2.0ren iraultzaren benetako "errudunetako" bat RSSa izan zen.

Pentsa, blog eta albisteentzako etiketatze semantiko sinple batek hori egin badu, zer ez den

Blogekin, blogetan espezializatutako bilatzaileak sortu ziren, eta RSSak batu eta filtratzeko zerbitzuak; egunkari eta aldizkarien webguneek ere jarri zuten RSSa, sare sozialek ere bai... Web 2.0ren iraultzaren benetako "errudunetako" bat RSSa izan zen.

IRUDIA: FRANK PODGORAIK/123RF.



gertatuko beste edonolako kontzeptuak (pertsonak, salgaiak, ekitaldiak...) ere semantikoki etiketetzen direnean...

## WEB SEMANTIKOAREN AJEAK

Hala ere, ezin guztia ona izan. Badira jada urte batzuk web semantikoaren ideia sortu zela, eta gauzatzea asko kostatzen ari zaió. Izan ere, ez da lan erraza. Batetik, existitzen diren kontzeptu guztientzako ontologiak definitu eta adostu behar dira, eta, hainbat gauza jada egin diren arren, lan ikaragarria da.

Baina, bestetik, are garrantzitsuagoa dena, gero edukia formatu horietan sortu behar da, eta hori oso neketsua izan daiteke. Ezin dugu espero webguneak sortzen dituen jendeak RDF formatuan eskuz etiketatzea. Webguneak tresnak erabiliz sortzen dira aspalditik, eta tresna horiek izan beharko dira egokitu eta edukia formatu semantikoan ere sortuko dutenak, blog-plataformek zuzenean RSSa ere publikatzen duten bezala. Kasu jakin batzuetan, espero izatekoa da hori nahiko laster gertatzea, hala nola edukia berez nahiko egituratua den kasuetan (ekitaldien egutegiak, adibidez) edo enpresen interesekoak direnetan (on line dendetako produktuen deskribapen orriak, adibidez).

Zailagoa izango da gaur egun lengoia naturalen idatzitako testuetan agertzen den informazio guztia ere semantikoki etiketatzea. Testu batean pertsonak, liburuak, haien ezugarriak, haien arteko erlazioak eta abar deskribatzen direnean, hori semantikoki etiketatzea, tresna bisualen laguntzarekin ere, izugarritzko lana da. Eta ez dago, egutegiaren edo dendetako produktuen kasuetan bezala, automatikoki egiterik...

## MAKINAK TESTUA ULERTZEN?

Edo bai. Hainbat esperimintutan, Lengoia Naturalaren Prozesamenduko (NLP) teknikak erabiltzen ari dira testu arruntetatik automatikoki etiketatze semantikoak erazteko, batzuetan arrakastaz. Webgintzako tresnek agian horrelako LNPko teknikak integratu ditzakete, eta eduki sortzaileari lagundu diezaiokete etiketatze semantikoak sortzen, etorkizun ez hain urrun batean. Baina, bestalde, makinak hori benetan ongi egiteko gai badira, ez dago web semantikoaren beharrik, horrek esan nahi baitu makinak gai direla testua "ulertzeko", eta Interneteko bilatzaile eta bestelako agenteek zuzenean modu eraginkor batean tratatu ahal izango dituztela HTML formatuan dauden testuak.

Ez dakigu zein iritsiko den lehenago, semantikoki etiketatutako weba ala makinek testuaren semantika edo esanahia ulertzea. Eta, lehenengo kasua bada, ezin jakin zenbat eduki egongo den web semantikoan: web osoa etiketatuko den semantikoki, edo gauza jakin batzuk soilik (sinpleenak eta enpresen interesekoak), edo bien arteko zerbait... Edonola ere den, modu batera edo bestera, webean esanahiak gero eta ga-

*Edukia formatu horietan sortu behar da, eta hori oso neketsua izan daiteke. Ezin dugu espero webguneak sortzen dituen jendeak RDF formatuan eskuz etiketatzea.*

rrantzi handiagoa izango du, eta, semantikari esker, zerbitzu gero eta hobeak izango ditugu. Sir Tim Berners-Leek berak aurtengo martxoan esana da: "Weba ez dago amaituta. Gaur egunoko weba izebergaren punta besterik ez da. Teknologia berriak iritsiko dira, askoz ere botereitsuagoak, inoiz pentsatuko ez genituzkeen gauzak egitea ahalbidetuko digutenak. Onena iristear dago". Hala bedi! ●

Web semantikoari esker, *sting* hitza bilatu, eta musikariari buruzko sarrerak bakarrik jasotzeko aukera izango dugu. IRUDIA: ERIC MILLER/W3C.

