



IGOR LETURIA AZKARATE
Informatikaria eta ikertzailea

Apirilean izan zen Lyonen (Frantzia) WWW2012 Kongresua; W3C erakundeak, webaren gidaleroak zehazten dituen, urtero egiten duen kongresua. Kongresuko aurkezpenari eta workshopetako gaiak erreparatuta, webaren gaur egungo nondik norako eta erronken eta, beraz, etorkizuneko joeren ideia bat egin dezakegu.

WEBAREN etorkizunerako joerak

Azken urteotako gai izarretako bat *personalization* delakoa da, hau da, webeko erabiltzaile bakoitzari bere gustuen, ideien eta zaletasunen arabera edukia erakustea. Toki askotan aplikatzen da hori jada. Adibidez, hainbat bilatzailek norberari egokitutako emaitzak erakusteko aukera ematen dute, gure nabigazioko historialetan oinarrituta: aurrez egindako bilaketak, bilaketa horietako emaitzetatik zeinetan egin dugun klik, zein eduki aukeratu dugun sare sozialetan elkarbanatzeko... Musika entzuteko webgune eta zerbitzuek ere, guk edo gure lagunek aurrez entzun ditugun abestietan oinarrituta, gustuko izango dugula uste duten kanta berriak iradokitzen dizkigute. Berdintsu saiatzen dira on line dendak ere, aurreko gure erosketetan oinarrituta edo guk erositakoen berdinak erosi dituzten beste bezero batzuen erosketek kasu eginez. Komunikabide batzuk ere hasi dira azal ezberdina erakusten erabiltzaile bakoitzari, aurretik irakurri ditugun artikuluen informazioa erabiliz.

Printzipioz, pertsonalizazioak ideia ona badirudi ere, badago haren egokitasuna zalantzan jartzen duenik ere. TED hitzaldi benetan gomen-dagarri batean, Eli Pariserek ohartarazten du bakoitza bere burbuilan bertan goxo gelditzeko arriskua dagoela, munduan dagoen guztiaren informazio mugatu eta partziala soilik jasotzearen benetako arriskua. Lehen komunikabideek eta telebista-kateek iragazten zuten informazioa guretzat, eta mota bateko egunkariak eta telebista-kateak soilik irakurri eta ikusten bagenituen, alde bakarreko informazioa soilik jasotzen genuen. Internetek ikuspegia zabaldu zigun, baina, orain, pertsonalizazio-iragazkiek aniztasuna kaltetu dezakete berriz.

Bestalde, eduki pertsonalizatua erakutsi ahal izateko, enpresa erraldoi horiek guri buruzko informazioa biltzen dute, eta horrek pribatutasunaren kezka sortzen dio jende ugari. Oso interesgarria da Tim Berners-Lee webaren asmatzaile eta W3Cko sortzaile eta zuzendariak horren inguruan duen iritzia, WWW2012ko sarrera-hitzal-

dian adierazi zuena. Haren ustez, enpresek guri buruz duten informazioa beste batzuei saltzeko edo beste erabilera oker batzuetarako erabiltzea ez da zuzena eta ez litzateke egin beharko, baina ez du uste hori hain praktika ohikoa denik, azken finean ez baita hori izaten enpresa gehieneko negoziaren muina, eta hori egiten dutenak luzera merkatuak zigortzen ditu. Ez du partekatzen batzuek duten kezka, enpresek guri buruzko informazioa erabiltzeari buruzkoa, guri zerbitzu hobe bat emateko erabiltzera mugatzen bada. Arropa-denda baten adibidea jarri zuen: prakak erosten dituen dendako saltzaileak haren prakak neurria aurreko alditik gogoratzea ondo dago, beti probatzen ibiltzeko beharrik gabe (zerbitzu ona ematearen betiko maxima bat izan da "ezagutu zure bezeroa").

Edonola ere, enpresek beren informazioa ezertarako ere ez gorde eta erabiltzea nahi dutenek ere horretarako eskubidea dutela deritzo, eta enpresek hori errespetatu behar dutela; horretarako, W3Ck proposatu du HTTP protokoloan aukerako "do not track" goiburu bat sartzea: webgune batek hori duen eskaera bat jasotzen duenean, ez luke gorde beharko bezeroaren inongo daturik. Nabigatzaile ia guztiak inplementatu dute jada aukera hori, baina ikusi egin behar ea enpresek errespetatuko duten...

Pribatutasunaren defentsa gehiegizkoa, baina, kalterako ere izan daiteke. Europar Batasunean, adibidez, lege batek debekatu egiten du cookieak (erabiltzailearen preferentziak eta bestelakoak gordetzeko erabiltzen den sistema) gordetzea erabiltzailearen baimenik gabe. Cookieak gabe, ez da soilik webgune batek ezin duela pertsonalizazioa egin: logeatu ere ezin zaitezke egin! Lege horren arabera, legez kanpo daude webguneen % 90etik gora. Orain arte ez dute Estatuek legea betearazi, baina Britainia Handian milioi erdi libra arterainoko isunak jartzea ahalbidetu dute berriki, eta bertako webgune batzuk hasi dira cookieen baimen eskaera gogaikarria egiten.



WEB SEMANTIKOA

Azken urteotan asko entzuten den beste gai bat web semantikoarena da, eta hala izan da aurtengo kongresuan ere. 2009ko maiatzeko eta ekaineko zenbakietan idatzi genuen horri buruz. Testuz beharrean informazio egituratu esanahidunez osatutako web paralelo bat da, testuzkoa ez bezala, makinek ulertzeko eta egoki tratatzeko errazagoa dena. Haren gainean osatutako zerbitzu aurreratu *adimendunak*, testutik eduki egituratua erauzteko proiektuak... Halako aurkezpen ugari izan ziren.

Googlek berriki aurkeztu duen Knowledge Graph produktuak ere web semantikoan du oinarria. Pertsona, toki edo gauza konkretu bat bilatu badugu, ohiko web-orrien zerrendaz gain, informazio egituratu eta erlazionatua duen taula bat ere erakusten du.

Modako gaia da, eta zalantzarik ez datozen hila eta urteetan halako zerbitzu asko agertuko direla. Nahiz eta, egia zor, aspaldi ari diren esaten hurrengo izango dela web semantiko lehertuko den urtea, eta leherketa hori ez da iristen; baina egia da, halaber, poliki-poliki eta banan-banan ari direla gero eta halako zerbitzu gehiago agertzen.

LENGOAIA NATURALAREN PROZESAMENDUA

Aurtengo kongresuan berritasun bat izan da, Lengoaia Naturalaren Prozesamenduaren eta Hizkuntza-teknologiaren presentzia handia. Elhuyarko I+G Taldean duela hamarkada batetik lantzen ditugun teknologia horiek, zeinei buruz sarritan hitz egin dizuegun atal honetan (corpusak, itzulpen automatikoa, bilatzaile hobeak egiteko teknologiak, galderak erantzuteko sistemak, elkarrizketa-agentek...), tradizionalki toki gutxi izan dute webaren eta bilatzaileen

inguruko kongresuetan. Halako gaiak hizkuntzaren inguruko kongresuetan lantzen ziren edo hizkuntza-teknologiaren ingurukoetan, baina orain artean webaren eta bilatzaileen munduak ez du haien beharrik ikusi. Oso azaletik tratatu izan dute hizkuntza: *stemming* edo sasi-lematizaio sinple bat, hitzen bilaketa hutsa...

Teknologia horiekiko oraingo interesaren arrazoi nagusietako bat da aurreko metodo sinpleek goia jo dutela eta emaitzak hobetzeko hizkuntza sakonago aztertze beharra ikusi dutela. Bilatzaileen kasuan, adibidez, jabetu dira emaitza hobeak itzultzeko beharrezkoak direla hizkuntzaren arabarako bilaketa, bilaketa eleanitza, sinonimoen edo antzeko hitzen bilaketa, jendearen iritziairen analisi automatikoa haren arabarako rankinak egiteko, laburpen automatikoa, galderak erantzutea, eta halako teknika sakonagoak.

Beste arrazoi bat web semantikoarena da, paradoxikoa badirudi ere. Teorian, web semantikoarekin, makina gai da bertan dagoen informazio egituratua ulertzeko, hizkuntza ulertzeko beharrik gabe. Baina problema jakin batzuk ebazteko (erabiltzaileak lengoaia naturalean egindako galderak erantzuteko, adibidez), beharrezkoa da desanbiguazioa (galdera web semantikoko zein objektu konkreturi buruz den jakitea, edo zein propietate konkreturen ingurukoa). Edo testuetatik automatikoki web semantikoko eduki egituratua erauzi nahi bada, beharrezkoak dira hizkuntza-teknologiak.

Aipatutako hiru gai nagusi horiez gain, asko hitz egin zen sare sozialei buruz ere. Eta web mugikorra, HTML5a, bideoa, 3D... Eboluzionatzeko aukera eta bide ugari ditu oraindik webak, eta datozen urteotan izango gara eboluzio harri horren lekuko, ziur. ●

“Aurten berritasun bat izan da, Lengoaia Naturalaren Prozesamenduaren eta Hizkuntza-teknologiaren presentzia handia”