

# Etiketa-lainoen Ikuskera Hobetzeko Multzokatzea

Arkaitz Zubiaga, Alberto P. García-Plaza, Víctor Fresno, Raquel Martínez

Universidad Nacional de Educación a Distancia  
{azubiaga, alpgarcia, vfresno, raquel}@lsi.uned.es  
<http://nlp.uned.es>

**Laburpena:** Erabiltzaileek aurrez anotatutako datuak berreskuratzeko baliabide interesgarria bilakatu dira markatzaile sozialak. Mota honetako webguneek etiketarik erabilienek osatutako etiketa-lainoa erakusten dute nabigazio aukera gisa. Etiketa-laino hauek, ordea, ez dute etiketen arteko antzekotasuna ez eta edukia kontuan izaten. Lan honetan SOM mapak erabiliz etiketen arteko erlazioak definitzeko metodoa aurkezten dugu. Erlazio hauek definitzeko, web dokumentuen edukietan oinarritu gara. Emaitza bezala lortutako mapa, ondorioz, etiketa-multzo ezberdinez osatzen da, multzo horietako bakoitzean termino esanguratsuenak erakutsiz nabigazioa eta ikuskera hobetzen direlarik. Azkenik, metodologia honek izan dezakeen aplikazio erreala aztertzen dugu.

**Hitz gakoak:** markatzaile sozialak, etiketak, etiketa-lainoa, multzokatzea

## 1. Sarrera

Etiketa sozialak erabiltzaileei web edukiak era irekian, elkarlanean eta egitura ez-hierarkikoan deskribatzeko aukera eskaintzen dien Web 2.0 fenomenoak dira. Markatzaile sozialetan aurretik beste erabiltzaile gehitutako orrialdeei etiketak erantsi diezazkiekete. Honi esker, zenbat eta erabiltzaile gehiago deskribatu orrialde bera, adierazgarriagoa izango da batutako etiketa-sorta. Adituek sortutako taxonomia klasikoetan ez bezala, etiketa sozialetan oinarritutako sistema folksonomikoek kontrolik gabeko berbategia osatu ohi dute, eta ondorio bezala sortzen den etiketa-egitura ez da ez finitua eta ez zehatza.

Sistema hauetan parte hartzen duen erabiltzaile kopuru handia dela-eta, anotatutako web dokumentu kopurua izugarri hazten ari da azken urteotan, etiketok mota ezberdinetako atazak hobetzeko interesa piztuz: web bilaketa eta nabigazioa, etiketa-lainoen ikuskera, eta jarioen iragaztea, besteak beste.

Markatzaile sozialei esker anotatutako datu-sorta handiak eskura daitezke, baina aurre-prozesamendua beharrezkoa da metadatu horiei etekina ateratzeko. Hori lortzera bidean, etiketen eta haien edukien arteko loturak aurkitzea da interes nagusienetakoa. Lehenago hainbat lan proposatu dira etiketen bidez informazio-atzitzea hobetzeko, baina guztiek ere etiketak batera agertzearen arabera egin dituzte multzokatzeak, batzuek erabiltzaileen datu gehigarriak kontuan izan dituzten arren.

Lan honetan, etiketei dagozkien web dokumentuen edukiak kontuan izanez etiketen arteko loturak deskubritzeko SOM mapetan oinarritzen den metodoa aurkezten dugu. Honi esker, mugakidetasunean oinarrituz markatzaile sozialetan etiketak gomendatzea ahalbidetzen du, eta baita jarioen zerbitzuak hobetzea ere.

Gainera, hizkuntza-ereduak baliatuz etiketa-multzo bakoitzari dagozkion termino esanguratsuenak eraztea ahalbidetzen du. Honenbestez, metodo honek etiketa-laino aberastuago bat sortzen laguntzen du, etiketa-multzo eta terminoen bidez nabigazioa eta ikuskera erraztuz.

Bestalde, metodo honek etiketak eta dagozkien dokumentuen arteko koherentziaren analisi kualitatiboa egiteko aukera eskaintzen du.

Dokumentu honen gainerakoa honela dago antolatuta. 2. atalak etiketen arteko loturak antzemateko lehenago proposatu diren hurbilketak aurkezten ditu. 3. atalean, gure lanean baliatu dugun datu-sorta aurkezten dugu lehenik eta behin, ondoren gure metodologia azalduz. 4. atalean, esperimentuen emaitzak erakutsi eta aztertu egiten dira, 5. atalean metodo hau aplikazio errealean ezartzeak izan dezakeen bideragarritasuna argituz. Azkenik, lanetik ateratako ondorioak eta etorkizunerako ildoak plazaratzen dira 6. atalean.

## 2. *Erlazionatutako Lana*

Lehenago hainbat metodo eta hurbilketa proposatu dira etiketen arteko erlazioak antzemateko. Horietako askok etiketak batera agertzeko probabilitatea hartzen dute kontuan etiketak multzokatzerakoan [5][2]. Beste lan batzuek, ordea, informazio gehigarria erantsi diote aurreko honi, hala nola erabiltzaile eta baliabideen arteko kointzidentziak [6][4].

Beste lan batzuetan lotura semantikoak bilatzen saiatu dira, ontologiak baliatuz. Horretarako, Wikipedia eta Google bezalako tresnek emandako informazioan oinarritu izan dira, besteak beste [1].

Aipatutako lan hauez gain, etiketa sozialetan oinarritzen diren hainbat sistema ere saiatu dira etiketen arteko erlazioak bilatzen. Tresna hauek kode itxia izan eta beraien funtzionamenduaren xehetasunik eskaintzen ez duten arren, badirudi etiketak batera agertzeko probabilitatearekin jokatzeko dutela. Esate baterako, Delicious berak etiketa bakoitzeko lotura izan lezaketen beste hainbat erakusten dizkigu, eta Flickr-ek etiketa bakoitzarekin lotuta daudenak multzo txikia-goetan banatzen ditu.

Gure lan honetan, ordea, etiketei dagozkien dokumentuen edukietan oinarritu nahi izan dugu, eta honek etiketen arteko loturak aurkitzeko duen egokitasuna aztertu, era berean erabiltzaileek definitzen dituzten etiketen eta edukien artean lotura zuzenik ba ote dagoen aztertuz.

## 3. *Gure Metodoa*

Lan honek etiketak eta beraien edukia aztertuz etiketa-lainoa berrantolatu eta beste era batean ikusteko metodoa aurkezten du. Lan honetarako hainbat pausu burutu behar izan dira: a) Datu-sorta eskuratu eta etiketa esanguratsuenak aukeratzea; b) Etiketak edukiaren arabera adieraztea; c) Etiketak antolatu eta ikusteko SOM bidezko multzokatzea; eta d) Multzo bakoitzari dagozkion termino adierazgarriak erauztea. Jarraian azaltzen ditugu pauso hauek guztiak.

### 3.1. *DeliciousT140 Datu-sorta*

Esperimentu hau burutu eta proposatutako metodoa probatu ahal izateko, Delicious gunean oinarritutako datu-sorta eskuratu genuen 2008ko ekainean. Datu-sorta honi DeliciousT140 izena eman diogu.

Datu-sorta hau sortzeko, webguneko 140 etiketarik erabilienetan oinarritu ginen, hau da, bertako etiketa-lainoan agertzen direnak. Etiketa-sorta honi T140 izena eman diogu. Honetan oinarrituz, T140 sortako etiketa bakoitzeko sarrera berriak monitorizatu genituen, 379.931 dokumentu bakar eskuratuz. Ondoren, dokumentu hauetako bakoitzari erabiltzaileek definitutako etiketak eskuratu genituen webgunetik, era berean dokumentuak dagozkien helbideetatik jaitsez, edukia eskuratzeko.

Azkenik, gure datu-sortan ingelesezko dokumentuak bakarrik utzi genituen, gainontzekoak baztertzuz. Gainera, Delicousek gutxienez bi aldiz definitutako etiketen informazioa ematen duenez web dokumentu bakoitzarentzako, etiketa informaziorik ez zutenak kendu genituen. Azkenean, 144.574 dokumentu osatzen dute gure datu-sorta, 67.104 etiketa ezberdin dituztelarik definituta. Datu-sorta hau eskuragarri dago ikerketa lanetarako<sup>1</sup>.

### 3.2. *Etiketen Adierazpena*

Dokumentu bakoitzak bere pisudun etiketa-zerrenda du. Pisu hauek etiketa bakoitza esleitu duen erabiltzaile kopurua definitzen dute eta, honenbestez, dokumentuarekiko adostu duten garrantzia. Hori dela-eta, zerrenda horretarako etiketa garrantzitsuenetan oinarritzea erabaki genuen, batzuek apenas izan baitezakete garrantzia, erabiltzaile gutxi erabili dutela-eta. Etiketa garrantzitsuenak bakarrik kontuan izateko, atari bat definitu genuen minimo bezala; datu-sorta osoko batzuek bestekoa gainditzen duten etiketak baino ez genituen kontuan izan.

Zerrendetako lehen postuetan dauden etiketak sarriago erabiltzen dira gainontzekoak baino. Ondorioz, lehen postuetako etiketa hauetan oinarritzeak etiketen arteko erlazioak aurkitzerakoan emaitza hobek eman ditzakeela uste dugu.

Dokumentu bakoitzean, sorta osorako definitutako batzuek bestekoa gainditzen duten etiketak baino ez ditugu kontuan izan, beti ere T140 sortan barne daudela kontuan izanik. Beraz, dokumentua guztiz baztertzeko da etiketa esanguratsurik edo T140 sortako etiketarik ez badu.

<sup>1</sup><http://nlp.uned.es/social-tagging/delicioust140/>

Iragazte prozesu honen ondoren, T140 sortako etiketa bakoitzari esleitutako dokumentu kopurua aztertu genuen. Nabarmendu beharra dago banaketa ez dela uniforme. *design* da dokumentu gehien dituen etiketa, *11.856*rekin, eta *2008* gutxien dituen, *219*rekin.

Honekin iragazitako datu-sorta txikiago bat eskuratu genuen, gure beharretara egokituta. T140 etiketa sortako bakoitzari zegozkion dokumentuak esleitu, eta guztiak dokumentu bakarra bailiran batu genituen super-dokumentu batean, emaitza bezala 140 super-dokumentu sortuz. Dokumentu bakoitza hainbat super-dokumentutan egon daitekeenez (T140 sortako etiketa bat baino gehiago badu), hurbilketa honek etiketak batera agertzearen probabilitatea ere kontuan hartzen du inplizituki.

Super-dokumentu bakoitza termino-bektore bezala adierazteko, HTML markak kendu genituen lehenik eta behin, testu-laua eskuratuz. Horren ondoren, hitz hutsak ere kendu egin genituen, ingelesezko hitz hutsen zerrenda bat baliatuz, eta azkenik *stemming* prozesua burutu genion, Porter-en algoritmoa baliatuz.

*TF-IDF* funtzioaren bitartez terminoak pisatu ondoren, maiztasunaren (DF) arabera termino batzuk kendu genituen: maiztasun altuenekoak, eta baxuenekoak.

### 3.3. SOM Bidezko Etiketen Multzokatzea

Etiketen adierazpena burutu eta gero, SOM mapak [3] erabili genituen antzeko etiketak multzokatu ahal izateko. Honi esker, aurretik sortutako 140 termino-bektoreak sisteman sarrera bezala izanez, multzokako proposamena eskaintzen du, bektoreen gertutasunaren arabera.

SOM mapa konfiguratzerakoan, beharrezkoa da multzo kopuru bat definitzea. Esperimentu honetarako  $12 \times 12$  multzoko mapa sortu genuen, etiketa kopuruaren antzeko balio bat erabiltzearen. SOM mapak sortzeko software bezala SOMlib<sup>2</sup> erabili genuen.

### 3.4. Hizkuntza Ereduen Bidezko Terminologia Erauzketa

Behin mapa sortuta, eta etiketen multzokatzeak ezagunak zirenean, etiketa zein multzo bakoitzarentzako termino esanguratsuenak erauzi nahi genituen. Termino bat etiketa edo multzo batean esanguratsua dela esateak, bertan sarritan agertzen dela esan nahi du, baina beti ere gai-nontzekoetan sarri agertzen ez delako. Beraz, dibergentzia kontuan izan beharra dago nolabait. Dibergentzia kontuan izateko, KLD funtzioan oinarritu ginen:

$$\text{KLD} = P_n(t) \cdot \text{Log} \frac{P_n(t)}{P_m(t)} \quad (1)$$

non  $P_n$  t terminoak multzoan agertzeko duen probabilitatea den, eta  $P_m(t)$  t terminoak sorta osoan agertzeko duena.

Honenbestez, KLD balio altu batek multzo zehatz batean agertzeko sorta guztian agertzeko baino probabilitate askoz altuago duela adierazten du. Honetan oinarrituz, etiketa zein multzo bakoitzarentzako termino esanguratsuak erauzi genituen.

## 4. Esperimentuak

Aurreko pausoak burutu eta gero (datu-sorta eskuratzeko, etiketen adierazpena, SOM bidezko etiketen multzokatzea eta terminologia erauztea), multzokatutako etiketak eta erauzitako terminoak mapa batean adierazi genituen.

### 4.1. Nola Irakurri Mapa

Emaitza modura lortutako mapak<sup>3</sup> etiketen banaketa ikustea ahalbidetzen du. Elkarrekin dauden multzoek gertutasuna adierazten dute; ondorioz, etiketen edukien arteko gertutasuna adierazten dute. Era berean, multzo batean hainbat etiketa agertzeak edukien arteko gertutasuna adierazten du. Garrantzi handia du, beraz, multzoen arteko gertutasuna aztertzeak.

Gainera, mapa sortzerakoan, etiketa-lainoen filosofia mantentzen saiatu gara, etiketak letra-tamaina ezberdinekin adieraziz. Laburbilduz, mapa hau Delicious-en jatorrizko etiketa-lainoaren berregituraketa bat da, antzeko etiketak multzokatu eta termino esanguratsuak erakutsiz.

<sup>2</sup><http://www.ifs.tuwien.ac.at/~andi/somlib/>

<sup>3</sup>Mapa osoa ikusteko: <http://nlp.uned.es/social-tagging/asonam2009/>

3d\_net advertising ajax apple architecture art article articles au audio **blog** blogging blogs book books business collaboration community computer cooking cool css culture database **design** development diy download economics education email english environment fashion fic finance firefox flash flex flickr food free freeware fun funny game games google graphics green hardware health history home howto humor illustration images inspiration interesting internet iphone java javascript jobs jquery language learning library lifehacks linux mac marketing math media mobile movies mp3 music news online opensource osx performance photo photography photos photoshop php politics portfolio productivity **programming** python rails recipe recipes reference research resources ruby science search security seo shop shopping social socialmedia socialnetworking **software** teaching tech technology tips tool **tools** toread travel tutorial tutorials tv twitter typography ubuntu **video** videos visualization web Web2.0 webdesign webdev wiki windows wordpress work writing youtube 2008

	0	1	2	3	4	5	6	7	8	9	10	11
0	apple osx mac	<b>software</b> freeware windows	howto ubuntu <b>linux</b>	java performance	tech computer hardware		email library	flex flash		web wordpress	css javascript jquery <b>webdesign</b>	portfolio inspiration illustration
1	<b>programming</b> python_net development	tips	security					<b>reference</b>	tutorials	ajax	typography	<b>design art</b>
2	3d graphics opensource			audio			rails ruby					fashion
3					videos <b>video</b> youtube					images	photoshop	photo shopping shop <b>photography</b>
4	games game mobile					tool		tv movies	cool			diy
5	science database			google search politics	news				interesting			funny
6	visualization math				seo jobs	work		economics finance	travel	green home		fic au humor
7	architecture					2008		articles		environment		language english
8	lifehacks					community		culture				health
9	iphone			research	online	internet						food cooking recipes recipe
10	collaboration socialnetworking php	<b>tools</b>		resources		twitter	advertising					
11	flickr productivity firefox wiki photos	<b>tutorial</b>		education teaching learning	technology blogging <b>web2.0</b>	social business socialmedia	marketing media		<b>music</b> mp3 download blogs <b>blog</b>	article	free toread writing	books book fun history webdev

Terminologiari dagokionean, zerrenda bana dute bai multzoek eta bai etiketek. Zerrenda hauek garrantziaren araberako hurrenkera dute, goian esanguratsuenak ageri direlarik. Honenbestez, etiketa bati zein bere multzoari dagozkion terminoak aldi berean azter daitezke, alderatuz. Etiketako termino-zerrendan beltzez agertzen direnek neuronan ere ageri direla esan nahi dute.

## 4.2. Emaitzak

Gure esperimentuen emaitza gisa irudian erakusten den mapa eskuratu genuen. Etiketek izan beharreko multzokatze idealaren informaziorik ez dugunez, ezinezkoa zaigu analisi kuantitatiboa egitea, eta ondorioz analisi kualitatiboa egitea baino ez zaigu gelditzen. Lehen gainbegiratuan, hainbat ideia ikus daitezke.

Hainbat multzok erlazio logikoak erakusten dituzte. Esate baterako, (7/0) multzoa, *flex* eta *flash* etiketekin; (6/2) multzoa, *rails* eta *ruby* etiketekin; (4/3) multzoa, *videos*, *video* eta *youtube* etiketekin; (11/9) multzoa, *food*, *cooking*, *recipe* eta *recipes* etiketekin.

Beste multzo batzuetako etiketek duten lotura, ordea, gaian adituak diren erabiltzaileek ondorioztatu ahal izango lituzkete. Adibidez, (2/0) multzoa dauden etiketak: *howto*, *ubuntu* eta *linux*. Linux eremuan ohikoa izan daiteke hau, izan ere “howto” hitza oso erabilia baita.

Terminologiari dagokionean, etiketen arteko ezberdintasun nabarmenak ikus daitezke. Adibidez, *research* etiketak *student*, *journal*, *conference*, *researchers* eta *paper* terminoekin lotura handia dutela esan daiteke, baina *interesting* etiketarentzako ageri diren *offender*, *quote*, *statement*, *xxx* eta *jfrater* terminoak oso nahasiak dira, beraien arteko erlazio semantikoa aurkitzea oso zaila delarik. Honek, ordea, ezaugarri interesgarriak erakusten ditu, izan ere etiketa mota ezberdinen arteko diferentziak azaleratzen ditu: alde batetik, gai bati erreferentzia egiten dioten etiketak (*research*, *economics*, *recipes*), eta bestalde, subjektiboak, pertsonalak eta bestelakoak (*interesting*, *cool*, *toread*).

Etiketen beste multzokatze batzuk hasiera batean arraroak dira, baina terminologia azterteak argitzen du nolabait beraien arteko lotura. Adibide bezala, (0/4) multzoko *games*, *game* eta *mobile* etiketek erakusten dute mugikorrentzako jokuetan interesa duen komunitatea bada-goela. Era berean, (0/5) multzoko *science* eta *database* etiketek datu-base zientifikoen inguruko interesa erakusten dute. (11/3) multzoak ere, *photo*, *shopping*, *shop* eta *photography* etiketekin, argazkintzarako materialean interesa azaleratzen du.

Azkenik, (3/5) multzoan gertatzen den *google*, *search* eta *politics* etiketen multzokatze bitxia gure datu-sorta sortu genueneko joeraren ondorio dela dirudi, izan ere Ameriketako Estatu Batuetako hauteskundeek mugimendu handia izan baitzuten orduan sarean.

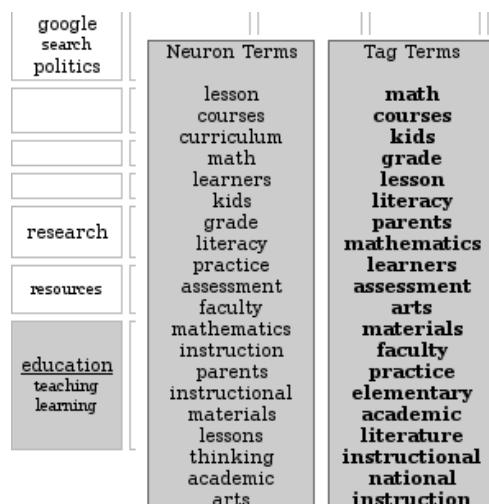
Gure datu-sortako T140 etiketentzako, gehienak multzo handiagoetan bana litezke, gai nagusiak sortuz. Guk aurkitutako kategoriak honakoak dira:

- **Informatika**, multzokatutako (0/0), (0/1), (0/2), (1/0), (1/1), (2/0), (2/1), (3/0) eta (4/0);
- **Web Diseinua eta Diseinu Grafikoa**, multzokatutako (9/0), (9/1), (10/0), (10/1), (11/0), (11/1), (11/2), (11/3), (10/3) eta (9/3);
- **Hezkuntza**, multzokatutako (3/9), (3/10) eta (3/11);
- **Sukaldaritza**, (11/8) eta (11/9) multzoak;
- **Denbora-pasa**, (9/11), (10/11), (11/11), (11/5) eta (11/6) multzoak; eta
- **Ekonomia**, (7/6) multzoa eta multzokatutako (5/11) eta (6/11).

Kategoria hauek ongi multzokatu izanak edukiak eta etiketen arteko koherentzia erakusten du. Kasu batzuetan, ordea, badirudi multzokatzeak ez direla guztiz zuzenak; (11/11) multzoan, denbora-pasa kategoriakoa izango litzatekeena, *webdev* etiketa *book*, *books*, *fun* eta *history* etiketekin batera ageri da. Era berean, (0/11) multzoan *productivity*, *firefox*, *wiki*, *photos* eta *flickr* etiketak multzokatu dira. (8/11) multzoan, bestalde, bi gai ezberdin nahastu direla dirudi: *music*, *mp3*, *download*, *blog* eta *blogs*. Hau gertatzearen arrazoietako bat etiketak eta beraien edukiaren arteko lotura batzuetan hain argia ez izatea izan liteke; adibidez, *blog* etiketa oso orokorra da, eta litekeena da gai asko nahastea bere baitan.

Beste etiketa batzuk ondo multzokatuta egon arren, badirudi aldameneko multzoarekin lotura handirik ez dutela. (4/3) multzoa, *video*, *videos*, eta *youtube* etiketak dituen, *tool* etiketa duen (5/4) multzoaren ondoan dago. Beste behin, *tool* etiketa oso orokorra delako gerta liteke hau.

Etiketa motak kontuan izanez, gai argi bati erreferentzia egiten ez dioten (pertsonalak, subjektiboak) etiketa gehienak bakarrik gelditu dira beraien multzoetan: (8/4) *cool*, (8/5) *interesting*, (6/10) *advertising*, (1/11) *tutorial*, (4/0) *online*, (5/7) *2008*, (5/4) *tool*.



Azkenik, badaude etiketa batzuk singularrean zein pluralean agertzen direnak. Kasu hauetako asko batera multzokatu diren arren, kasuren batean oso urruti dago bata bestearengandik, *tutorial* (1/11) eta *tutorials* etiketekin gertatzen den bezalaxe (8/1).

## 5. *Eztabaida*

Aurkeztutako metodo honen aplikazio interesgarria izan liteke berau Delicious bezala sistema erreal baten ezartzea. Horri esker, erabiltzaile batek etiketa zehatz batean termino zehatz bat duten orrialdeetara harpidetzea ahalbidetuko luke, besteak beste. Askotan erabiltzaile batek ez ohi ditu etiketa zehatz bat duten dokumentu guztiak jaso nahi, gehiegi direlako. Adibidez, erabiltzaile batek “education” etiketa dutenak jaso nahi izan ditzake, baina bere baitan “literature” terminoa baldin badute bakarrik. Kasu honetan, sistemak “literature” terminoa duten “education” etiketadun dokumentuak eskainiko litzokie erabiltzaileari, baina nahi izanez gero baita multzo bereko etiketenak ere, hau da, “literature” terminoa bai, baina “teaching” edo “learning” etiketak izanez. Ohiko sistemaren bitartez erabiltzaileak ez lituzke azken dokumentu hauek jasoko, eta seguruenik bere interesekoak izango dira.

Arazo nagusi gisa, sistemaren eskalagarritasuna aztertu beharra dago. Posible al da sistema hau eguneratua mantentzea? Lehenik eta behin, maparen eguneraketa noizean behin burutu beharko litzateke (adib., astean behin) azken  $n$  dokumentuak kontuan izanez, webgune osoa kontuan hartzea gehiegi bailsateke. Mapa eguneratu bitarteko tarte horretan, dokumentu berri guztiak maparen azken bertsioan dagozkien multzoan eroriko lirateke.

Bestalde, honek denboran zehar etiketen bilakaera aztertzea ahalbidetuko luke, joera aldaketak antzemateko aukerarekin. “news” bezalako etiketa batek aldatzeko joera handia izan dezake, garai zehatz batzuetan berri konkretu batzuek izan dezaketen mugimendu handia dela-eta.

## 6. *Ondorioak eta Etorkizunerako Ildoak*

Lan honetan multzokatutako etiketa-lainoa eskuratzeko metodoa aurkeztu dugu, etiketa eta multzo bakoitzeko terminologia esanguratsua erauziz. Gainera, Delicious webguneko 140 etiketa erabilienetan oinarrituz datu-sorta bat eskuratu dugu, DeliciousT140.

Hurbilketa honek markatzaile sozialetan hurrenkera alfabetikoa izan ohi duen etiketa-lainoarekiko berrantolaketa bat egiten du. Berrantolaketa honek etiketen arteko loturak antzematen laguntzen du, terminologia esanguratsua nabarmenduz. Analisi kualitatibo batek emaitza positiboak erakusten ditu.

Etorkizuneko ildoan honakoak nabarmen ditzakegu: (1) etiketa bakoitzaren semantikan analisi sakonagoa burutzea; (2) gure metodo honen emaitzak etiketak batera agertzearen probabilitatearekin egindakoarekin alderatzea; (3) analisi hau beste hizkuntza batzuekin zein datu-sorta eleanizetan probatzea.

### *Eskerrak*

Lan hau MAVIR ikerketa sareak (S-0505/TIC-0267) eta QEAVis-Catiex proiektuak (TIN2007-67581-C02-01) finantzatu dute.

### *Bibliografia*

1. S. Angeletou, M. Sabou, L. Specia, and E. Motta. Bridging the gap between folksonomies and the semantic web. In Proceedings of ESWC 2007: Workshop on Bridging the Gap between Semantic Web and Web 2.0, 2007.
2. G. Begelman, P. Keller, and F. Smadja. Automated tag clustering: Improving search and exploration in the tag space. In WWW '06: Collaborative Web Tagging Workshop, 2006.
3. T. Kohonen. Self-Organizing Maps. Self-organizing maps. 3rd ed. Berlin: Springer, 2001, xx, 501 p. Springer series in information sciences, ISBN 3540679219, 2001.
4. P. Mika. Ontologies are us: A unified model of social networks and semantics. Web Semantics: Science, Services and Agents on the World Wide Web, 5(1):5-15, 2007.
5. P. Schmitz. Inducing ontology from flickr tags. In WWW '06: Collaborative Web Tagging Workshop, 2006.
6. X. Wu, L. Zhang, and Y. Yu. Exploring social annotations for the semantic web. In Proceedings of WWW '06, pages 417-426, New York, NY, USA, 2006. ACM.