

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

J. Parra-Arnau, A. Perego, E. Ferrari, J. Forné and D. Rebollo-Monedero

INTRODUCTION

In the scenario of social bookmarking, a user browsing the Web bookmarks web pages and assigns free-text labels (i.e., tags) to them according to their personal preferences. The benefits of social tagging are clear – tags enhance Web content browsing and search. However, since these tags may be publicly available to any Internet user, a privacy attacker may collect this information and extract an accurate snapshot of users' interests or *user profiles*, containing sensitive information, such as health-related information, political preferences, salary or religion. In order to hinder attackers in their efforts to profile users, this report focuses on the practical aspects of capturing user interests from their tagging activity. More accurately, we study how to categorise a collection of tags posted by users in one of the most popular bookmarking services, Delicious (<http://delicious.com>).

METHODOLOGY

As frequently done in collaborative tagging sites, the profile of a user is modelled as a *tag cloud*, that is, a visual representation where tags are weighted according to their frequency of use. Note that tag clouds are in essence equivalent to normalized histograms of tags.

Nevertheless, the representation of a user profile as a cloud, or equivalently as a normalized histogram, of all the tags submitted by users is clearly an inappropriate approach; not only because of the intractability of the profile, but also because it makes it difficult to have a quick overview of the user interests. For example, for users posting the tags “welfare”, “Dubya” and “Katrina” it would be preferable to have a higher level of abstraction that enables us to conclude, directly from the inspection of the user profile, that these users are interested in politics. In this sense, the categorization of tags allows us represent user profiles in a tractable manner, on the basis a reduced set of meaningful categories of interest. Motivated by this, next we proceed to describe the methodology that we followed to categorise the Delicious data set retrieved by the Distributed Artificial Intelligence Laboratory (DAILabor), at Technische Universität Berlin [1]. This methodology, explained in what follows, is in line with other works in this field [2], [3].

The data set in question is organized in the form of triples (*username*, *bookmark*, *tag*), each one modelling the action of a user associating a bookmark with a tag. To accomplish

J. Parra Arnau, J. Forné and D. Rebollo-Monedero are with the Department of Telematics Engineering, Universitat Politècnica de Catalunya, C.\ Jordi Girona 1-3, E-08034 Barcelona, Catalonia. E-mail: {[javier.parra](mailto:javier.parra@entel.upc.edu),[jforne](mailto:jforne@entel.upc.edu),[david.rebollo](mailto:david.rebollo@entel.upc.edu)}@entel.upc.edu

A. Perego is with the Institute for Environment and Sustainability of the Joint Research Centre, European Commission, I-21027 Ispra, Italy. E-mail: andrea.perego@jrc.ec.europa.eu.

E. Ferrari is with the Department of Theoretical and Applied Science, University of Insubria, Via Mazzini 5, I-21100 Varese, Italy. E-mail: elena.ferrari@uninsubria.it.

this categorisation, first we carried out some preprocessing to filter out those tags considered as spam. For this purpose, we collected some statistics about the number of characters contained by tags. After observing that 98% of tags had less than 23 characters, we dropped those tags with a number of characters over 22. In addition, we eliminated those posts with more than 50 tags, as they are usually spam [2]. Additionally, posts with no tags were not considered. After this simple preprocessing, the number of triples reduced to 1,149,895, and, consequently, the number of users, bookmarks and tags to 9,207, 349,658 and 54,024, respectively.

In a second stage, we aimed at identifying clusters or groups of semantically similar tags. As normally done in the literature, we performed a clustering analysis based on the *co-occurrence* between tags, that is, the number of times each pair of tags simultaneously appears in a same bookmark. Specifically, we modelled the relationships among tags as a matrix of co-occurrence $\{c_{ij}\}$, where each element c_{ij} with $i \neq j$ corresponds to the co-occurrence between tags i and j , and each element c_{ij} with $i = j$ contains the self-occurrence, i.e., the absolute frequency of appearance of tag i . Note that, clearly, this is a symmetric matrix and that each row (column) describes one tag in terms of the semantic similarity to the other tags.

In an attempt to concentrate on the significant relationships among these tags, we eliminated those rows satisfying $\sum_j c_{ij} < t$, for a certain threshold t . Similarly, we dropped those columns fulfilling an equivalent condition. In this regard, observe that, the higher the threshold, the lower the number of resulting tags, and thus the lower the number of triples containing those tags. Since we aimed at preserving at least 85% of the triples, and at the same time, we required the resulting tags to have a strong co-occurrence, we chose $t = 95$. In doing so, we obtained a reduced co-occurrence matrix with 5,999 tags. In conclusion, after this filtering process the number of triples, users and bookmarks became 985,273, 8,882, and 310,923, respectively.

Once we filtered the co-occurrence matrix, we proceeded to use a well-known clustering algorithm to create a two-level hierarchy of categories. But before applying this algorithm, we first required to specify a measure of similarity among tags. Recall that we modelled tags as rows and columns of a matrix, that is, vectors. As often done in the literature, we employed the cosine metric [4], a simple and robust measure of similarity between vectors. Equipped with this measure, we applied Lloyd's algorithm². As a result, we grouped the 5,999 tags into 20 categories. Afterwards, for each of these categories, we turned to apply the same algorithm to get 10 subcategories. This process yielded a total of 200 subcategories, which provided us with a granularity level thin enough as to define precise filtering policies, and sufficiently aggregated as to avoid noisy behaviors. The resulting categories were classified in decreasing order of popularity of their tags. Lastly, the tags in each subcategory were sorted in decreasing order of proximity to the centroid. The results of this categorisation may be found as follows.

² Lloyd's algorithm [5], which is normally referred to as k -means in the computer science community, is a popular iterated algorithm for grouping data points into a set of k clusters.

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #1									
Subcategory #1	Subcategory #2	Subcategory #3	Subcategory #4	Subcategory #5	Subcategory #6	Subcategory #7	Subcategory #8	Subcategory #9	Subcategory #10
palmas	serials	wrt54g	diagnostic	software	opensource	security	unix	linux	archivo
palmas	crack	iphone	computer	freeware	oss	linux/security	shellscript	distro	mauro
pda	maj	telephony	free&open	utilities	opensource.forge	nmap	sysadmin	distros	files
palm	programas	voip	audio/video	ware	open_source_portal	exploits	commands	distribuzioni	programmiewindows
pocketpc	serial	linksys	player	programs	projecthost	comp.network.security	adminspotting	livecd	dacontrollare
ppc	hpcv	asterisk	workstation	apps	softwarelibre	vulnerability	cli	distributions	antiguos
smartphone	cracks	sip	computers	software/windows	open-source	seguridade	comp.os.linux	knoppix	hardw_softw_libre
competition	warez	wifi	annoyance	windows	mozilla-bookmark	malware	un*x	redhat	software_wireless
moriarty	hackerecracker	wireless	itsstuff	soft	source	tcp-ip	sysadm	distribuições	webs_hardware
tex	logiciels	skype	operating	applications	software/linux	encryption	terminal	live-cd	microcontroller
blbtex	téléchargement	wi-fi	vcd	capture	open	firewall	backups	rpm	zil
handheld	cracking	router	techie	win32	freesoftware	hacking	rsync	command_line	
latex	logiciel	collaboration_tools	tiny	killerapp	gnu	watch.security	shell	distribues	
treo	informatique	smtp	tweak	win	open_source	secure	sync	colinux	
patents	securite	mesh	it	downloadsites	gpl	virus	filesystems	suse	
systemanalysis	hacker	imap	mpeg	windoze	blag-linux-and-gnu	antivirus	nix	programy	
cad	réseaux	spamassassin	techsupport	shareware	gnu/gpl	infosec	screen	fedora	
patent	patches	videoconferencing	tweaks	softwares	foss	sniffer	esr	packages	
peessoal	scan	gmail	recorder	synchronization	fsf	protection	availability	portais&midia	
rbshpc		wlan	ripping	bootcd	communauté	seguridad	comp.os.linux.tricks	linux:páginas:tecnicas	
logistics		postfix	favoriten	custo	fud	segurança	fs	kernel	
trabajo		aol	winxp	app.win	edit	worms	platforms/nix	operating_system	
		messengers	codec	application	libre	tcpip	shells	debian	
		dansk	tweaking	recovery	blag-software	spyware	software/vim	os	
		messaging/imap	xvid	cross-platform	longsight	adware	network.tools	virtualization	
			pc	file	free-software	ids	disk	system	
			disc	computerrelated	open:source	network	vi	slackware	
			customization	pcsoftware	foundation	networking	cygwin	distribution	
			pctech	operatingsystems	perspective	sec	servers	mplayer	
			dvr	devtools	opensource.adopter	md5	backup	wm	
			hardware	linuxsoftware	floss	mswindows	doc/tutorial	rescue	
			computing	freedownloads	openoffice	pgp	howto	bootable	
			computerinfo	downloads	private	ipsec	commandline	floppy	
			mediacenter	crossplatform	computerrelatedsites	cryptography	bsd	iptables	
			computerresources	putty	m5	ssh	anleitung	gnu/linux	
			computersandinternet	boot	thunderbird	tcp	filesystem	apt	
			formats	freebie	progs	crypto	admin	firewalls	
			techstuff	small	using	ports	monitoring	os_unix-like	
			burning	encoder	office	reverse	server	platforms	
			importedapril	fritz_use	chatterbot	segurança	administration	gentoo	
			computerstuff	beta	firefox_upload	pf	doc/howto	cluster	
			support	pick	check-out	vpn	aix	kde	
			tech.computer.random	bootdisk	grid	gpg	freebsd	netzwerk	
			multimedia	fritz.pick	rikhard	techsecurity	tuning	os.linux	
			codecs	fixit	blender	tcp/ip	generic-linux	yum	
			lab_export	software_developers	erp	forensics	embed	howtos	
			installation	telnet	mpeg-4	ms.ie	solaris	samba	
			computación	fritz.toolpick	tosort	openbsd	netbsd	gnome	
			htpc	catálogos	jmsa	ssl	config	computers/internet	
			monitor	free	native	pki	djb	user:harlock	
			programm	download		msft	neteng	driver	
			infrastructure	netops		viruses	articulos	user:soul	
			registry	ms.windows		from_gentoo_ff	broi	embedded	
			drivers	besplatno		ca	migration	x11	
			tecnologia	computersoftware		projects vpn	tutoriais	install	
			automation	vnc		cisco	ldap	sco	
			ff	screensaver			elvstone	sendmail	
			mythtv	archived				user-mode-linux	
			level	remote				routing	
			hdd	desktop				grafika	
			ordenadores	gratis				modem	
			pvr	windowsxp				laptop	
			informatica	todownload				páginas:diversas	
			divx	outliner				raid	
			mini-itx	clipboard				mutt	
			motherboard	xp				usb	
			ht	comp.software				setup	
			x86	emulator				procmall	
			mini	iso				tof	
			itv	ftp				zaurus	
			how-to	outliners				eliudhr	
			cpu	mirror				hp	
			mod	requirementsmanagement				novell	
			homemediasystem	dos				home-marks	
			bios	softwaredevelopers				laptops	
			simulator	freestuff				packaging	
			repair	rmttools					
			bootdisks	anti					
			portable	updates					
			san	pop3					
			firefox-exported	todoist					
			safari_export	programme					
			devices	sw					
			tivo	requirements					
			technik	jabber-clients					
			dell	microsoft					
			printer	engineering					
			notebook	convert					
			forum	program					
			hw	graphviz					
			dsl	groupware					
			suporte	mm					
			processor	upload					
			pcutilities	instant					
			importedopera	compression					
				platform					
				comp.os.windows.tips					
				projectmanagement					
				dtp					
				client					
				manager					
				webstats					
				zip					
				virtual					
				acrobat					
				ms					
				software&co					
				messenger					
				swf					
				log					
				vmware					
				rssed					
				matemática					
				im					
				icq					
				gimp					
				livre					
				messaging					
				rup					
				companies					
				outlook					
				configmanagement					
				jabber					
				exchange					
				longhorn					

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #2									
Subcategory #11	Subcategory #12	Subcategory #13	Subcategory #14	Subcategory #15	Subcategory #16	Subcategory #17	Subcategory #18	Subcategory #19	Subcategory #20
web.design	usability	docs&books	webdesign	code/html	rich	validierung	styling-objects	webwork	web_tools
web.css	interaction_design	tutoriel	blogs/web-designer	css/blogs	online_editor	testtools	stylingobjects	servicewebmasters	web
web-programming-css	ia&usability	zen	design	webdesignresources	richtexteditor	validators	webdesign-development	programmierung	devtech
css/position	interface_design	gen	inspirasjon	css&dhtml	ttw	validator	design/graphicdesign	siteperso	webapplications
script.css	design/patterns	meinlinks	web-resources	technologiesweb	widget	validation	userexperience	persosites	diseny
web_design	uidesign	netscape	blogs/design	hjemmesider	desarrollo_web	accessibility	materiale-per-il-sito	misc.resources-sortme!	favicon
web-design	interface	webmisc	blog.webdev	web_standards	word_processing	barrierefrei	miscellany	tutoriales	www
diseño	ui	apprendre	webdesignweblogs	x/html	programming/web	wai	styleguides	manuales	systemdevelopment
design-web	ux	ma-barre-personnelle	design/inspiration	mediamaking	wysiwyg	w3c	développement	macromedia	lorem
stylesheets	interaction	plans	diseño	web_design_standards	dhtml	advancedhtmlcourse	transparency	ria	button
html/css	web-dev.usability	ruralportal	inspiración	webdeezign	nifty+tools	useability	uploaded	dreamweaver	practice
web+design	uxd	temporários	sitedesign	webstandars	ajax	webtech	totag	web_reference	ipsum
csssites	ia		diseinua	designportale	html/css/javascript	webresources		foros	webstuff
comp.www.css.layout	infodesign		web-designers	website.design	browser	wd			buttons
webdev	information_design		comp.www.inspiration	e-zine	js	webdeveldienste			computar
tableless	interfacdesign		design_directory	emagazine	javascript	tidy			loremipsum
web_development	hci		swapfile	technique	favelets	afrotech			browsercheck
css/layout	productdevelopment		blogart		script.javascript	barrierefreiheit			eines_web
my-website	info-arch		look_and_feel		texto	tilgjengeleghet			images.graphics.code
comp.www.css.reference	comp.www.usability		gurus		favelet	qa			itlinks
webstandarts	experience		siteawardz		programming/javascript	disability			website
web/design/css	user		businesscss		webpub	access			cheatsheet
web.html	usabilidade		design_resources		ie7				webpage
www/css	design/interface		zines		dom				favicons
layout	webapps-research		ala-authors		browsers				webmastering
webdesign-css	userinterface		zengarden		bookmarklet				screenshots
mv2003	guidelines		redesign		brp				tgh
html	infoarch		inspiration.website		internetexplorer				urls
web-dev	human_factors		firms		explorer				site
bbag/css	prototyping		webdeveloper		ie				widgets
layouts	interactiondesign		magasin		form				wildwildweb
web/design	sitemanagement		awards		diverseysteme				tricks
design-web.html	pattern		portal		sort				techniques
xhtml	usabilidad		stdesign		moz				style
web_standards	navigasjon		uncategorized_yet		msie				good
development/webdesign	firma		candy		technomanies				screencapture
web_tech	roi		wrk		uiconcept				uteis
kpx			company		png				typesetting
web_authoring			websites		table				resources
language:css			minimal						webmaster
tips/css			webshops						web_architecture
koodaus			agencies						recursos
css/referencesites			folk						techresource
omnibus			weeklypick						a-analisar
design/web			portals						wizards
alistapart									code-to-mutate
diseños									screenshot
estándares									dev_misc
web_programming									unfiltered
css/tricks									educationtools
webdev/css									cosasporhacer
webauthoring									webby
css/nav									xblog
webdevelopment									resource
csstools									tips
css-template									webpublishing
webstandards									user:alsanan
scripts-javascript									hypertext
csslayout									websitehelp
useful-design-stuff									demo
dev.web.css									usage
float									textile
webdevel									adobe
authoring									wiklink
css2									lists
desarrollo-web									help
tables									formatting
epr									ressources
csszengarden									dlib
tooltips									tom
www.design									atk
markup-sprachen									miscellaneous
menus									notable
comp.www.css.tricks									boolou
web-programming									naming
web_development_coding									calendars
design/css									printing
cssresources									eksempel
quirks									aver
standards									
web_coding									
webpagestuff									
examples									
css_and_xml									
programming/html									
web-development									
information_technology									
instruktioner									
w3									
css									
accesibilidad									
web/css									
quickreference									
templates									
webdev/artigos									
w3schools									
compatibility									
web_dev									
devel									
cssbug									
crossbrowser									
programming_web									
programming/css									
forms									
web_ref									
example									
web.example									
applicationdevelopment									
web&wai									
cssundxml									
techs									
designstuff									
menu									
webdev/layouts									
webutvikling									
standar									
css-xhtml-standards									
boxes									
tutorials									
www-sivut									
markup									
linkfest									
template									
web_resources									
tutorial									
programming/.net									
navigation									
standard									
computer&internet									
boulot									
bestpractice									
positioning									
estudos									
basics									
importedyahoo									
lesson									
bugs									
folder									

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #3									
Subcategory #21	Subcategory #22	Subcategory #23	Subcategory #24	Subcategory #25	Subcategory #26	Subcategory #27	Subcategory #28	Subcategory #29	Subcategory #30
dev_java	vcs	java	swing	actionsript	ai.rules	databases/oracle	computers/programming	programming	cplusplus
techinfomtl	sourcecodecontrol	j2ee	extremeprogramming	xaml	expertsystem	handytool	comp-sci	interpreters	c++
javanevs/info	version	java.jsp	agile	avalon	artificialintelligence	programming/ada	software_engineering	lanugages	c
development	version_control	java.servlet	design-patterns	matlab	bayes	quickreferencelists	programming.perl	programming.python	cpp
toolkits	scm	java.j2ee	software-engineering	nice	genetic	ref/web	software-development	functional	c#
framework	versioncontrol	hibernate	unittest	symbian	automata	referencecards	software-design	io	dotnet
ide	revision	imp.java	designpatterns	imageprocessing	ml	programming/delphi	code	prog-lang	sourcecentrals
debug	svn	container	softwaredevelopment	as	agents	reference_card	design-software	cookbook	c/c++
dokumentation	versioning	java.swing	patterns	excel	rule	development/uml	code/help	assembler	c-sharp
desenvolvimento	version-control	spring	testing	snmp	machine	refcards	coding	proglangs	csharp
api	subversion	josprojects	methodology	temporary	bayesian	referencecard	computerscience	fp	stl
se	source-control	javadoc	refactoring	from_local	inference	databases/mysql	developers	programming/python	gcc
software/opensource	cvs	prog_news	tdd	sets	alife	development/test	regular	computerlanguage	components
esi	bancodedados	graphicsandgames	ood	oracle	prolog	distributedcompilation	dev	languages	libs
bestpractices	sql	servlet	unit	sensors	fuzzy	tarjetas	softeng	plang	mono
core	db	java-webapp	essence	crm	machine_learning	gaelle	softdev	talk.programming	vb
frameworks	control	swing	objectoriented	corsos	machinelearning	quickref	algo	haskell	codeproject
mvc	database	jakarta	bulintesting		processing	programming/c++	develinsite	osdev	com
sucks	normalization	junit	oo		kr	xemacs	software/engineering	scheme	asp.net
build	sqlite	java-tools	process		bot	whitepaper	softwareengineering	objective-c	msdn
diagram	databases	ioc	oriented		simulation	documentacion	expressions	vm	
editors	configuration	struts	extreme		educationalresource	programming/c	computer_science	lambda	batch
lib	postgres	ant	heuristics		intro	programming/java	best-practice	ocaml	controls
programming	php_mysql	manageability	unittesting		r	mod_perl	softwaredev	py	winforms
media	postgresql	jdbc	canaryinthebitmine			pdf	best-practices	optimization	needsattentions
gui	sqlserver	jsp	objects			vim	software_development	talvmenni	
list	arch	swt	modeling			emacs	note_to_self	objc	
tbd		jakartaapache	comp.software-eng				sh	compiler	
develop		javadocs	process-improvement				algoritmos	threading	
programação		jms	coaching				optimisation	lisp	
scalability		freshmeat	quality				computer-science	programminglanguages	
link		persistence	wiki.c2				regexp	parsing	
plugin		mom	programacao				algorithms	programming/ruby	
guru		java.jdbc					swdev	assembly	
fritz.investigate		java.jms					programming	proglang	
dev-tools		snippet					programación	python	
sim		groovy					programming/perl	comp.lang.python	
idiom		lgpl					programmieren	compilers	
logging		documentacao&tutoriais					script	asm	
harvester		sourceforge					softwareengandinfo	debugging	
comparison		jvm					instruction	ruby	
diff		jboss					scripting	smalltalk	
sdk		java.applet					practices	gtk	
distributed		codegen					manual	continuations	
informática		concurrency					comp	parallel	
comparisons		developmenttools					doc	common	
corba		jython					docs	e-book	
codegeneration		toolkit					datastructures	twisted	
uml		aop					dbms	ada	
reports		eclipse					programacion	opengl	
plugins		jee					cs	qt	
development		macro					data-structures	squeak	
implementação		tomcat					regex	jscrip	
sun		concurrent					systemdesign	pl	
src		integration					algorithm	erlang	
documents		server-side					tao	benchmarks	
date		j2me					documentation	wxwindows	
imaging		jakartaapache.tomcat					oreilly	self	
e		orm					bash	tcl	
activex		poll					datastructure	threads	
homepage		mobiletec					dev/res	later	
unmarked		game-programming-java					graham		
ibm		benchmark					perl		
delphi		profiling					azubi		
eai		rssfeed					modules		
		reporting					technical		
		performance					cpan		
		lucene					eola		
		workflow					compsci		
		middleware					module		
		gc					prog		
		tip					acm		
		watch.sun					bloom		
		programming/gis					lore		
		informatik					imported_from_moz		
		bpm					syntax		
		certification					manuals		
							betc		
							oscon		
							tomonitor		
							faq		
							postscript		

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #4									
Subcategory #31	Subcategory #32	Subcategory #33	Subcategory #34	Subcategory #35	Subcategory #36	Subcategory #37	Subcategory #38	Subcategory #39	Subcategory #40
genius	geek/programming	mining	skills	ether	mind_map	careerinterview	deal	visualisation	emergence
geek	programmer	investigación	psychology	psychogeography	mindmaps	hiring	bpo	dataviz	self_organizing
gr8	readings	mis/web_infrastructure	pers	locative	conceptmap	interviewing	billing	designinformation	stigma
technology	readings/articles	data-mining	selfhelp	life_caching	pims	cv	ideas	newsmap	complexity
iservice	dogma	comp_sci	problem-solving	googleplex	mind-mapping	jobhunt	coop	informationdesign	autopoiesis
nerdy	essay	ceglowski	squishy	geo	map.mind	career	living	visualization	networks
tech	wisdom	datamining	weak-links	location	notetaking	resume	accounting	viz	behavior
tagme	ensayo	lsi	feedback-loops	mobile	brainstorming	employment	trucking	viz.web	dai
thought	toread	naturallanguage	meta	urban_coincidence	notespin	questions	fuel	visualize	groupthink
foo	paul_graham	informationtheory	conversation	location_awareness	information_processing	mgmt	to:read	information-design	socialnet
data_structures	essays	ir	terra-nova	sense_of_place	dev/effectivity	jobs	askmeif	information_display	social_theory
info	outsider	newsmastering	psych	geowanking	mind_mapping	job	property	cartographie	powerlaw
interesting	doctorow	artificial	argumentation	geolocation	brainstorm	business	rfid	visualizations	collaboration
relephant	team	analysis	sci.psychology	mob	pareto	management		infographics	reputation-management
nerds	toberead	data_mining	sociology	space/place	lionkimbro	freelance		graph	noosphere
nbt	article	ai.classification	life	palimpsest	undef	jobhunting		graphs	groups
wow	licencing	data	smut	data_culture	tasks	bizarticles		infovis	social_behaviour
blogin	managment	text	big-ideas	mobility	categories	interview		worldwide	socialcapital
structures	to_read	retrieval	perception	mobile_tech	think	careers		viz.data	everyone>anyone
developer	joel	lingüística	motivation	series60	organizer	interviews		visualization	emerging
cunya	rant	ai	critical	phones	knowledge_management	work		data_visualization	etch
information	working	decision	persuasion	ubicomp	informationscience	methods		zipcode	swarm
muse	a_lire	classification	lifestyle	gis	fops	leadership		diagrams	influence
anti-social	licence	structure	organizing	barcodes	lifelacks	consulting		mapping	singularity
making	articles	cybernetics	sleep	barcode	information-management	e-banking		graphing	chaos
miscwebsites	law/ip/copyright	tracking	self-improvement	new_york	topicmap	career_survival		info_vis	redes
userinterfaces	open_content	reasoning	timemanagement	context-aware	omgwtfbbq	eventmanagement		map	papers
guide	hackers	uno	education&research	space-place	human_behavior	mag_cio		infoviz	memetics
make	ninjas	statistics	passion	phone	kmap	hr		treemap	bigend
medialit	advice	stats	humanity	gsm	mindmap	iaandixd		viz_ref	urban_planning
oku	toprint	intertwingle	weapon	gps	obsessive	entrepreneur		actualités	facilitation
systems	skill	citeseer	man	mobil	notes	nonprofit		infopom	urban-planning
metric	fritz.readthis	market	weapon.system	cellphone	productivity	change		cartography	conferences
todo	o'reilly	bib		geodata	mindmapping			display	
technology-tosort	manifesto	statistiques		mobile_phones	thinking.tool			chart	
labs	readme	robot		context	knowledge-management			maps	
needs_investigation	opencontent	robotics		surveillance	mind			treemaps	
nist	read	experts		mobiles	knowledge management			ergonomy	
infolit	story-telling	stanford		storytelling	selfimprovement			knowledgemapping	
tech&society	chums	sia		cameraphones	thinking			design/infodesign	
tocheckout	comment	forskning		telecom	learn2learn			vis	
communication	memex			place	organization			zui	
geekdom	professional			cell	chandler			tree	
practical	to-read			cellphones	explore			demographics	
startup	rules			walking	gtd			innovative	
real	industry			sig	notebooks			processing.org	
interest	copyleft			instructionaltech	creativity			computery_things	
invention	commons			education/??	cognition			topic	
filters	intelligence				mapas			stat	
theory	manage				plm			human	
semiotics	cluetrain				km			tocheck	
idea	license				pim			charts	
unfiled	toblog				project_management			geography	
disruptive	licenses				ideje			kart	
craft	cyberspace				knowledge			fastadd	
etc	innovation				infoorg			crystalline	
id	kevinmarks				annotation			infolinks	
paper	decentralization				wacko				
timeline	stories				thoughts				
cyber	people				memory				
futurism	biz				planning				
informatics	talks								
macluhanisms	lecture								
computation	commercial								
machines	joito								
rad	organisations								
future	brownhen								
misc.creativity	ikc								
loc	pr								
import									
presentation									
presentations									
simplicity									
gadgets									
who									
huh									
time									
sustainable									
infonomics									
consume1									
value									
manufacturing									
wearable									
cyborg									
tobuy									
powerpoint									
nanotech									
current									
vr									
bbs									
idea.key									

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #5									
Subcategory #41	Subcategory #42	Subcategory #43	Subcategory #44	Subcategory #45	Subcategory #46	Subcategory #47	Subcategory #48	Subcategory #49	Subcategory #50
truthinmedia	environment	politics	online-blogs	news	government	boycottions	corporations	intellectual_property	tech_news
scum	green	gwb	firefox:toolbar<wbr>	usmedia	law	changerieschoses	world	ipr	planet
sympathy_for_the_devil	climate	bush	geek-stuff	worldnews	govt	bonnetrose	geopolitics	freeculture	news.tech
markets	republic	liberal	toolbar:/	newspaper	legal	activisme	economics	eff	tech-news
islamofascism	közársaság	anti-bush	uutisijasarakuvia	actualiteit	policy	by:tchort	cia	getinvolved	webmagazine
middleeast	sustainability	elections	ntic	generalnews	state	ohio	economy	freedom	t:news
cuba	eco	nausea	internet/news	e-media	united	signal_orange	globalization	newglobal	news&info
checks	disaster	election	tecnologia	nieuws	national		capitalism	neweconomy	news-technology
israel	ngo	election2004	geekstuff	nachrichten	politica		countries	newworld	technews
event	forecast	president	todos	nooze	egov		england	cory	readings/news
jewish	feminism	presidential	información	alternatemia	gov		corporate	commentary	daily
ruusia	oil	columns	olddaily	newsandviews	criminaljustice		soccer	littlebrother	onof
australia	gender	dubya	geeknews	newspapers	legislation		football	mediavc	tekkies
korea	worry	cheney	linuxnews	alternative-news	taxes		power	society	tech&science
boston	denver	kerry	metanews	periódicos/revistas	crime		econ	intellectual	actu
austria	michigan	progressive	dagliga	quotidien	military		politica	watch	daily-news
losangeles		people_bush	dailyites	opinions	california		suv	monstermedia	startlist
fr		left	kult	mainstream	florida		global	choice	e-zines
real_estate		activism	notclinux	cnn	poverty		spain	linkoftheday	it-magazines
islam		congress	errata	krant	police		grassroots	finalist	current_events
dc		libertarian	uutiset	anti-news	organizations		indymedia	people's	medios
investment		war	linux:noticias	newsandmagazines	switzerland		china	creative_commons	alternativenews
perso		foreignpolicy	linklist	giornali	texas		labor	fcc	imported101807
		fbush	nouvelles	mydailyreading	abuse		germany	monopoly	currentevents
		liberty	aa-popular	news.media	philadelphia		outsourcing	untagged	rss-feeds
		bushco	newssite	zeitungen			canada	mag	blogs/group
		conservative	dailyread	press			end_of_rome	organismos	newsandviews/daily
		constitution	importadas	mags			india	eu	computernews
		rumsfeld	dailies	periodicos				issues	newssource
		democrats	revista	periodicals					daily-morning
		voting	firefox_home_import	zeitung					tecnologia
		lies	veille	media					wired
		mesopotamia	infos	netmags					browse_me
		political	technologie	presse					noticias
		moore	noticias	mediaandvisualart					excellent
		watchdog	discussion	world_window					science/technology
		chomsky	popculture	guardian					ordinateur
		pol		onlineperiodicals					hi-tech
		michael-moore		newsmedia					reg
		america		diario					info-filter
		freespeech		noticies					readings/reviews
		usa		magazin					magazine
		iraq		nyheder					periodical
		democracy		keyword					newssites
		uspolitics		big					slashdot
		nh		mustread					cyberculture
		speeches		city					scitech
		michaelmoore		magazines					revues
		scienceandsources		center					gadget
		us		headlines					kulttuuri
		abu_ghraib		medien					to_sort
		opinion		diarios					ezine
		polls		charlotte					pdastuff
		georgewbush		nyt					work_bookmarks
		protest		columbus					natives
		op-ed		action					
		rights		journalism					
		homosexuality		show					
		christians		fox					
		conspiracies		aktuell					
		terrorism		alternative					
		corruption		actualite					
		idiocy		prensa					
		politik		international					
		humanism		gossip					
		polit		politique					
		propaganda		east					
		torture		internacional					
		socialism		zine					
		brian-bait		dallas					
		conspiracy		pages					
		fanatics		bbc					
		liberties		españa					
		abortion		nytimes					
		peace		chicago					
		anarchy		castellano					
		scary		windowsmobile					
		election_04		journal					
		anarchism		initial					
		numeracy		uk					
		fear		italiano					
		copyright		france					
		terror		yeahyeah					
		originales		onlinejournalism					
		cap							
		justice							
		fucked							
		morons							
		marriage							
		circlejerk							
		communism							
		vote							
		globalisation							
		globalnapolitika							
		dotgov							
		work							
		bright_moments							
		prison							
		editorial							
		united_states							
		gibt							
		belgium							
		fraud							
		afghanistan							
		report							
		hongkong							
		ethics							
		taiwan							
		infoshop-bookmarks							
		racism							
		un-tagged							
		queer							
		gay							
		sexuality							
		un							
		humanrights							
		iran							
		finnish							
		politically_correct							
		race							

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #6									
Subcategory #51	Subcategory #52	Subcategory #53	Subcategory #54	Subcategory #55	Subcategory #56	Subcategory #57	Subcategory #58	Subcategory #59	Subcategory #60
google/gmail	music/relations/sounds	searchengines	juarez	cool_tools_2004	bookmarking	bypass	services	webtool	web/services
email	netmaps	websearch	bittorrent/search	hobby.reading	delic	convenience	financial	redskaber	registration
google_mail	knowledge_bases	searchengine	file_sharing	web_solution	tag	miscinternet	banking	web/tools	geek-tools
web/google	nodes	search_engine	torrent	anonymity	tagging	pass	broadband	webtools	internet
mail	for-scott	search_engines	torrents	netsevice	linkblogging	tool-etc	isp	tool	systemadministration
webmails	culturilla	search	bit	websurftools	del.icio.us_api	web/registration	speed	from:furl	passwords
webmail	recommender	søk	trackers	mail.spam	folksonomy	webaccess	adsl	handy	traceroute
emails	visual-networks	searching	p2p	touse	web-application	webutility	provider	tools	anonymous
e-mail	search/similarity	internet.searchbots	bittorrent	greasemonkey	webservices/delicious	internet.websurfing	myclip	cooltools	networking/dns
anti-spam	graphical	opzoekprogramma's	bittorrent:multimedia	passwd	activism/open	listarapida	payment	service	whois
antisipam	coo	search_engines	bt	internet/ferramentas	tags	fakelogin	servicios	webapp	geekage
gmail	musika	search.engine	bittorrent	trolls	bookmark-logging	bugmenot	telecoms	useful	miscellaneous
spam	digg	buscas	filesharing	lefttotheirrowdevices	charting	subversive	wsh	utility	hostingproviders/free
hostingproviders/gmail	global_developments	engine	tracker	utiles	social_filtering	logins	servicosgratuitos	web2.0	lookup
emotemail	clustering	web/search	busquedas	coolinternetcrap	socialbookmarks	internetsearching	export	filter	address
filtering		searchers	downloading	account	furl	annoyances	microarray	ferramentas	password
messaging/bouncing		suchmaschinen_generell	file-sharing	web/service	info-management	must	webconferencing	by/k	web.service
messaging/spam		zoeken	broadcatching	filetransfer	ferramenta	internet_services	laszlo	sweetcode	url
correo		searchanddirectory	peer-to-peer	miscpersonal	del	web.tools		smart	anonymity
blogged		buscadores	descargas	file_storage	del.icio.us_hacks	web-tools		webservice	net
messaging/spf		suchmaschinen	peer	dropload	metadelicious	clever		util	domain
messaging/clients		cherche	special	revisar	id_genebooy	register		outils	internettools
hostingproviders		busca	privat	comp.security	visualseringer	netz		browsing	bandwidth
ops		referenceandsearch		adria	geekness	goodideas		magnolia	reference.alt
discussiongroups		search-engine		replace	full-text	kjekt		oppslag	proxy
nerdalert		arama		filestorage	collaborativefiltering	login		webapps	privacy
		toolsandreference		accounts	naviguer	dump		webbased	hosts
		moteur		goodies	popular	bestofweb		web.tool	zeczintelligence
		search-engines		paste	yoshi	faydali		utils	status
		suchmaschine		large	delicious	badassness		bestof	anon
		metasearch		personalization	navegación	web-sites		utilidades	sci_&_tech/it/tools
		suche		id_eyoumail	bookmark_manager	privacyandsecurity		app	authentication
		suchen-tools		mywebs	discovery	paranoia		interesantsoftware	web-based
		search/tools		transfer	links	temp		web_app	dns
		avoir		bigbrother	delfav			mssg	proxies
		engines		pratique	bookmarks			geeky	software/webbased
		recherche		exported	enlaces			webspace	host
		find			others			nytte	track
		mozillaproject			driven			a-classificar	comp.network
		searchtools			del.icio.us			incele	sécurité
		keywords			bookmark			online_tools	webhosting
		pagerank			irony			intriguing	ip
		seo			zeitgeist			web-content	web.browser
		directory			bookmarklets			quick	geographical
		google			digital_video			essential	webutilities
		si??			data_organization			reminders	domains
		locate			information_management			trick	wardriving
		everyday			hot			informatica/utilidades	urlinfo
		addresses			fred			usefull	hosting
		alt			mdntblu's			hacks	ietf
		posta						note	ms.ie_switch
		query						notepad	sso
		geral-legal						nerd	rfc
		sci_&_tech/it/search						hit=2	surf
		spider						referencia	
		search.google						sharing	
		indexing						collaborative	
		sem						personal_it	
		gthroughi						efficiency	
		fasetter						webnote	
		utgpkt						inventive	
		portali						en	
		plagiarism						korisno	
		promotion						skladistepodataka	
		definition						teknik	
		usetnet						postit	
		dmoz						nützlich	
								xuseful	
								ocr	
								sites	
								firefox_bookmarks	
								hack	
								linkcollections	
								index	
								geekery	
								online	
								archiving	
								webware	
								storage	
								dir	
								netsevice	
								kelly	
								hivemind	
								rel:start	
								webfeel	
								shared	
								extensions	
								busqueda	
								archive	
								recommendations	
								miscbookmarks	
								mike's	
								guia	
								spiders	
								converter	
								conversion	
								sci_&_tech/it	
								classifyme	
								jsled	
								top	
								varios	
								faves	
								athroughc	
								divers	
								dicas	
								chat	
								communications	
								webbrowser	
								wikis	
								how_to	
								post	
								knowledgebase	
								msn	
								extension	
								firefox	
								misc.chat	
								copy	
								telephone	
								contact	
								aim	
								bots	
								test	
								odn-candidates	
								sources	
								profile	

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #7									
Subcategory #61	Subcategory #62	Subcategory #63	Subcategory #64	Subcategory #65	Subcategory #66	Subcategory #67	Subcategory #68	Subcategory #69	Subcategory #70
devhumour	toys	flash	cats	zombies	humor/webcomics	fun	pom	culture	funny
weirdwebpages	origami	animation	booze	freak	monday	novelties	sex	counterculture	humor
i'mbored!	geometry	avatar	alcohol	apocalypse	wednesday	web.toy	adult	popcult	timesink
linkwhore	crafts	avatars	cat	experiment	webcomic	diversions	pr0n	webzine	humour
hohohomm	papercraft	so	pets	emotion	friday	misc.fun	homosexual	ads	neato
webs	drawings	avatare	animals	creepy	mon-wed-fri	cool	erotica	advertising	wrong
funnies	plastic	curioses	dogs	tests	thursday	webtoys	nsfw	trucks	jokes
snark	folding	timewasters	dog	creatures	webcomics	curiosidades	girls	hollywood	satire
humorous	quirky	animations	addiction	lunacy	comic	generators	erotic	raelity	comedy
news.misc	knots	toy	pt-br	body	feedsite	justplainfun	adamhill	mobilewhack	badger
zoo!	kids	chill	buffy	death	dailyreads	amusing	brkchrnr	cultural	nta
lol	lego	homestarrunner	sweden	bees	saturday	whimsy	deusx	wack	timewaste
firefoxcrewpicks	bikes	web_art	sillygwailo	building	comix	coolstuff	women	commerce	parody
confessions		col-leccions	setre	robots	sticky	curiosos	male	japan	neato!funny!
serendipity		imatges	startrek	dreams	morning_dew	amusement		ad	joke
ludorije		monkeys	diversão	magic	stumble	brilliant		subculture	weirdstuff
memes		czaby	diversos	paradox	levity	freaky		detroit	weird
prive		mario	cartoniefumetti	survival	dilbert	bored		hah	silly
tidbits		south		remainder	comics	amusements		celebrities	matt
netculture				personality	cartoon	useless		disney	kitsch
procrastination				electricity	fumetti	underholding		danish	stupidity
absurd				violence	toons	neat		berkeley	whoa
coolrandomstuff				just	cartoons	aaaégeral		annarbor	schadenfreude
raw				drugs	th	random			harhar
stupid				house	weekly	hipster			surreal
div				love	geekculture	kedsomhed			ninja
ego				sonstiges	m	misc			dirtstyle
russianmusic_news				water	sinfest	ummm			fake
					bob	original			gross
					topten	goof			subvert
						coolsites			wtf
						ephemera			heyurl
						webtoy			insanity
						erie			funstuff
						excathedra			hmmm
						informacion			bizarre
						awesome			crazy
						wierd			dumb
						chris.h			cute
						linklog			fools
						geeks			evil
						manners			crap
						kornukopia			drinking
						nonsense			haha
						diverse			rants
						longread			silliness
						etiquette			odd
						neat_stuff			truth
						kate			ridiculous
						misc-amusements			strange
						inspirational			stupid
						found			gold
						ocio			animal
						cheat			initialimport
						relationships			sad
						story			relationship
						viral			videos
						hello			pirate
						diversion			dummy
						quizzes			simpsons
						nifty			bad
						consumerism			worldwideweird
						clock			ha
						tbt			disturbing
						soc			dating
						things			jesus
						sign			pirates
						junk			michel_v
						meme			cosmolinks
						boingboing			holidays
						funshit			column
						hoax			irc
						memories			
						numbers			
						crash			
						quote			
						pix			
						hoaxes			
						want			
						interestingsites			
						inventions			
						kultur			
						ikea			
						hmm			
						users			
						nelson.autotagged			
						oblink			
						quotes			
						massa			
						otros			

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #9									
Subcategory #81	Subcategory #82	Subcategory #83	Subcategory #84	Subcategory #85	Subcategory #86	Subcategory #87	Subcategory #88	Subcategory #89	Subcategory #90
dictionaries	primarysources	booklist	russian	librarians	arguing	language	gaiman	gutenberg	reference
dictionary	myths	unread	ru	libraries	arguments	english	fiction	freebooks	encyclopaedia
wörterbuch	history	nonfiction	fanfic	librarian	argument	etymology	prose	freeonlinebooks	encyclopedias
inglés	religion	christian	livres	library	rethoric	wordsmithy	sf	nice_reading	encyclopedia
dict	folklore	anglican	harrypotter	lawlibrary	acadaemia	toknow	lovecraft	books/online	books-and-reference
onlinedictionaries	mythology	catholic	fanfiction	opac	speaking	grammar	writers.writing	books	encyclopedia
thesaurus	myth	family	fandom	digitallibrary	discourse	words	lit	onlinebooks	search&research
translator	history/cs	judaistik/judaica	livres-lecture	legislative	languageologic	linguistics	writers	e-texts	language_english
translate	occult	pharmacy	expression	librarianship	debate	etym	horror	ebooks	look-it-up
slang	god	adoption	sanasto	digitallibraries	critical_thinking	lang	authors	etext	generalreference
hittegiak	studies		zh	librarystuff	rhetoric	lingua	author	angelsaksisch	reference
translation	historical		histoire	bibliography	sprog	misc.writing	scifi	literature	readyreference
acronym	skeptic		importedfavorites	census	relations	anagram	science_fiction	books/info	references
sprachen	spirit		e2	codes	logic	spelling	sci-fi	texts	internettecave
acronyms	lds		jdcupload2	british	misc.critical-thinking	communicate	sciencefiction	litteratur	referencesites
diccionario	victorian			healthcare	skepticism	write	cthulhu	ebook	fav:tools
translators	forteana				philosophy	lexicon	fantasy	freeful-textsites	start
scienze	atheism				universe	writing	fiction/meta	education-books	facts
lengua	medieval				cheating	corpus	novel	book	referencia
dictionnaire	past				academics	word	cyberpunk	e-books	urbanlegends
sprache	judaism				writings	terminology	writer	e-text	legends
glossary	steampunk					letters	mystery	literary	timezone
diccionarios	renaissance					latin	portuguese	rare	timezones
lexika	girardian					natural	tolkien	textbook	research&reference
deutsch	pagan					orwell		humanities	conversions
dizionario	humans					multilingual		libros	encyclopédie
german	consciousness					eng		classics	general
englisch	magick					editing		reading	portals/feeds
idiomas	tarot					anglophilia		books_and_reading	atlas
jargon	mormon					spoken		shakespeare	cultura
foreign	biography					dead		books-online	quotations
sanakirja	geschichte					wordplay		publicdomain	academic
lexikon	mysteries					ling		knjige	trivia
italian	celtic					metaphor		great	whitepapers
spanish	meditation					kieli		minimalist	knowledgebases
español						vocabulary		epublish	wissen
français						name		wrz	intouchwith
linguas						kanji		literatura	and
french						names		joyce	states
langue						nlp		oulipo	citation
shortcut						misc.writing.tips		publishing	calculator
swedish						greek		lessig	référence
anglais						wordnet		litteratur	the
						speech		criticism	quickhelp
						chinese		isbn	wikipedia
						japanese		reads	interests
						arabic		litterature	philo
						esperanto		poetry	archives
								mywebpagestuff	dream
								publishers	measurement
								litterature_children	to
								mpeg-7	all
									imported_home
									archive.unsorted
									imported2
									livros
									knowledge_data

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #10									
Subcategory #91	Subcategory #92	Subcategory #93	Subcategory #94	Subcategory #95	Subcategory #96	Subcategory #97	Subcategory #98	Subcategory #99	Subcategory #100
blogs	messaging/loaf	feed	socialsoftware	enjoying	blogosfæren	blogsoftware	identity	blogresearch	blog
weblog	messaging/greenlisting	tools/rss	social-networks	quicklink	cyberstudier	blog_tools	s60	blogreading	blog-tw
blog/meta	ver	newsfeeds	social-networking	português	akademisketekster	blogging_software	cellular	blogconversations	br
meta_blog	loaf	feedreader	social	brazil	cyberkultur	tools/blog	etcon	blogwriting	unchecked
weblogs	bitácoras	aggregators	social-software	marketing	metafilter	blogware	slide	blogbusiness	mtblog
techblog	verification	realisimplesyndication	socialsw	itm	spaceplace	weblog-tools	conference	lj	cblog
www.blogging	braytek	rss_newsmaster_tools	social_network	vc	mefi	blog_software	consultants	livejournal	tw
daily_blog	bitacorras	feeds	yasns	bblog	diplom	blog-publishing-tools	collaboration_issues	pesquisa	facet
blogstuff	messaging/procmail	rss/readers	social-network	francophone	oai	blogtool	wap	librarians	webblog
blogindex	es	aggregation	social_networks	sharepoint	blogtalk	blogtech	mobilekomunikacije	lis	cn
blog-en		web-aggregator	social_software	nl		blogdevelopers	oip	friends	blogaggregate
linkfilter		aggregator	socialnetworking	baby		blogtools	teens	dl	posts
namac		rssreader	connectedness	researchers		outils-blog	yonderboy	eros	blogger
blogsearch		rss/xmlcitacivijesti	ssw			weblogging	fallo	ecomunities	fifthestate
topic_blog		blog-tools	socialnetwork			blogging	backchannel	ict	educacao
blog_design		rss	socialnetworks			blog/engine	newsworthy	call	edublogging
analytics/blog		rss/xml	softwaresocial			movable_type		montreal	person
linkblog		syndication	software/social			blog_tech			typepad
web_women		newsreader	sn			blog_tech_mt			textos
blogosphere		messaging/deflexion	socnet			blogsoft			friend
att		blogs_tools	sns			blog-docs			emergentdemocracy
blogar		webtoolservice	team.r			ping			cibercultura
navelgazing		sosofunctions	ss			blogging/tools			planet
wget		blogthings	shirky			blog/tool			emrgent
metaweb		blogs&newsfeeds	friendster			weblogtools			linkage
mediascape		bloglines	digital_identity			photoblogging			sc-blog
bloggers		blog??	xfn			movabletype			persian
weblog/blog		newnet	sna			coreblog			me
eng-blog		messaging/syndication	community			blog-tool			my
conceptdesign		opml	socialcomputing			comp.weblog_mt			
blogosfera		microcontent	socialstudies			personalserver			
blogroll		subscription	blog,			mt			
markpilgrim		messaging/convergence	trends			flog			
node		readers	group			carnets			
blog/or		bpomag	orkut			carnet			
blogservice			comunidades			kms			
interweb			reputation			mtswitch			
frequent			media_theorizing			photolog			
intercom			connection			moveabletype			
blog_biz			cyberkultura			feature			
critique			socioware			intranets			
reading_blogs			communities			sixapart			
recursion			trust			cmsundblogs			
famous			presence			wp			
linking			cop			trackback			
ongoing			social_technology			publish			
linkdump			smartmobs			blosxom			
technorati			foaf			textpattern			
blogosphere			digitalid			mblog			
often			topics			flickr			
drivel			cooperation			wordpress			
features			irc/bot			algemeen			
trend						moblogging			
bbcwmin						survey			
art_blogs						comments			
efficient						enterprise			
smartmob									
electronicpublishing									
digicult									
digitalnoizdavstvo									
metablogging									
lukas									
blogsites									
alliance									
licensing									
to_blog									
cc									
digitalculture									
ff_bookmarks									
electronic_culture									
advocacy									
newmedia									

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #11									
Subcategory #101	Subcategory #102	Subcategory #103	Subcategory #104	Subcategory #105	Subcategory #106	Subcategory #107	Subcategory #108	Subcategory #109	Subcategory #110
freeimages	inspiration	design/icons	desing	ruins	art	graphics	architecture	digitalcameras	font
stock-photo	pixelart	graphics/html/icons	music&art	subways	fine-arts	graphic	prefab	digitalcamera	typography
digit	nice_flash	incons	t:artwork	mapping-spaces	art_history	graphique	modernism	photography	soft/fonts
stockphotos	src:adamkell	práctico	montagens	urban_exploration	painting	icon	urbandesign	news,reviews	fonttypography
stock_photography	design_web	webpics	bookmarx	urban	poster	graphic_design	houses	photos_numériques	fonts
stockphoto	graphicdesign	multimedia/images	norml	urbanrelics	graffiti	design/graphics	firm	photo	fuentes
digitalphoto	studio	web.dev	visualart	abandoned	művészet	visual	urbanism	prasowka	tipografia
imagenes	agency	carts	gráficos	underground	street-art	visualdesign	archi	digitalnafotografija	type
royaltyfree	logos	bullets	customisation	cities	art_ref	pixels	wankers	digicam	typo
stockpix	awesomeness	image/icons	themas	cities.london	street	signs	postmodernism	canon	handwriting
licensefree	portfolio	webdev/imagens	digitalart	subway	illustration	designsites	landscape	nikon	tipos
design	coolsite	arrows	skins	tube	webart	fonts/css/icons	female	foto	tipografia
design/stock	concept	prints	skin	london	arts_humanities	kuvankásittely	architects	camera	kreativeinkauf
archivo	print	ikoner	wallpapers	metro	nature_science_art	isometric	possiblepbs	analyse	boatbuilding
resources/webdesign	coolshit	mails	dessin	buildings	curiosities	tiles	brunettesandravens	fotografia	
stockart	portfolios/companies	piccies	desktops	drift	generative	vector	trabalho	photoblog	
stockimages	individual	gifs	skinning	publicspace	collective	rasterbator	exemplos	album	
imgs	generator.x	iconos	themes	foxmarks	sculpture	design.resources	schools	photoblogs	
free/Images	portfolios	infographic	wallpaper	newyork	posters	gif,graphics,fonts		cameras	
picfodder	design/grid	gif	artwork	ny	galleries	symbol		photographie	
freecontent	goodexamplesofsites	clip_art	galleries	crafty	creative_technology	textures		panorama	
design/stockphotos	motion	web_material	imagens	tut	netart	grafix		digital	
photostock	tufte	collecting	koal	danger	museums	gfx		chernobyl	
galeria	interactive		desktop/mod	moving	arte	backgrounds		panoramas	
imagen	webagency		customize		cool_sites	wowl		photographer	
stock	interior		web&pcstuff		fractal	pixel		fotografia	
photography/various	gof		karen'sfavorites		kunstlich	icons		fotografie	
stockphotography	design_companies		customizing		analog	texture		photoblog	
picture	designlink-portale		skinner		artsyfartsy	design/art		exif	
photomesh	experimental		windows.shell		net.art	draw		photographers	
repository/graphics	imagery		eyecandy		drawing	logo		lighting	
stock/royaltyfree	firm designer				vintage	visuals		photoblogs	
image	showcase				kunst	symbols		kameras	
imagenes	modern				museum	creation		galeria	
hi-res	uml.classdiagram				collections	grafik		age	
images	institution				historie	escher		natura	
clipart	industrial				jp	creative		fatolog	
pics	apartment				artsci	graficos			
res	design,illustration				hist	profoto			
projekt	arts/culture				vj	mysoftware			
illustrations	motiongraphics				pop.culture	homepagedevelopment			
stock_photos	wicked				exhibitions	spiff			
photograph	furniture				electronic_art	use			
photos	designer				fractals	photoshop			
see	new_mods				coleccion	ascii			
pictures	flash/motion				light	mega			
fotos	product				rt	max			
photographs	revistas				ilustración	collection			
ent	vfx				vienna	privatization			
stock-photography	branding				media_production	mediaarchive			
repository	glass				wood	graphisme			
jpg	contest				historia	classix			
pd	flags				presents	computer_graphics			
pic	device				obsess	geneticalgorithms			
gallery	best				toronto	intellectual-property			
graffix	brand				tattoo	tiling			
share_photos	linear				med	models			
img	corp					illustrator			
jpeg	dainternetexplorer					fw			
bilder	thomazo					cg			
photosharing	amsterdam					grafiktipps			
viewer						maya			
public_domain						ga			
digital_commons									
pretty									
grafico									
xxx									
jackblack									

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #12									
Subcategory #111	Subcategory #112	Subcategory #113	Subcategory #114	Subcategory #115	Subcategory #116	Subcategory #117	Subcategory #118	Subcategory #119	Subcategory #120
fsicstuff	mp3	music	mp3s	circumvention	exactaudiocopy	vst	station	hi-fi	sydney
ispstuff	mp3blog	netlabel	electronic_music	drm	exact	fx	netradio	surround	events
computerreviews	blog/mp3	sheetmusic	electronicmusic	copyright	compact	vsti	music_online	audiophile	venues
abitoffun	musicfilters	playlist	glazba	dmca	ripper	monkey	radiostations	loudspeaker	party
networkstuff	realaudio	dj	discography	intellectualproperty	eac	videoclips	audio.radio	room	bar
badminton	digitalmusic	obituary	music-all	fairuse	wave	trading	streaming	stereo	bay_area
workrelated	blog_mp3	bands	ambient	copyrights	entertainment	official	musicandradio	end	festivals
science-fiction	music/new	tablature	electro	aac	extraction		radio	high	austin
voice	mp3_blog	indie	musick	piracy	eaudiovideo		internetradio	acoustics	club
classical	playlists	records	synth	watch.musicindustry	error		livestream	speaker	clubs
american	audio/music	instruments	recommend	creativecommons	song		music_internettv	hifi	dance
johnadams	audio	electronica	sounds	copyright	mode		stream	dsp	finland
stevereich	mp3blogs	labels	independent	evol_digital	wav		tolisten	treatment	sanfrancisco
ives	music_download	riaa	samples	recentscool	cds		streams	manufacturers	la
cage	audioscrobbler	hip-hop	music-databases		id3		webradio	av	sony
phillipglass	listen	lyrics	recording		cd		radios		virginia
	flac	chords	electronic				music&movies		
	itunes	label	sound				npr		
	mashup	funk	remix				frequency		
	share	hiphop	soundart				podcast		
	ogg	rap	synthesis				broadcast		
	winamp	vinyl	musik				webcast		
	mix	synthesizers	covers						
	itms	punk	effects						
	music/tech	jazz	production						
	podcasting	rock	pop						
	garageband	synthesizer	listening						
	rip	guitar	webzines						
	tunes	midi	composition						
	ipod	instrument	artists						
	audiobooks	concert	musica						
	personalinterest	concerts	musicbusiness						
	live	piano	musique						
		tabs	spurl.netbookmarks						
		idm	ringtones						
		musicians	music/artists						
		musicvideo	techno						
		music/meta	check						
		ezines	música						
		artist	einkaufen						
		hardcore	demoscene						
		gamemusic							
		corporation							
		band							
		gitarre							
		musician							
		songs							
		metal							
		blues							
		instrumentos							
		importados							
		record							
		lr_work							
		músicos							
		diyaudio							
		fan							
		goth							

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #13									
Subcategory #121	Subcategory #122	Subcategory #123	Subcategory #124	Subcategory #125	Subcategory #126	Subcategory #127	Subcategory #128	Subcategory #129	Subcategory #130
taxonomies	part	semanticweb	xml	which	crossroads	php	a_classer	content_management	webservices
topic_maps	response	semantic	characterencoding	www-106	advanced	codesource	linguistiquechinoise	coding/scripts	rpc
thesauri	neil	semweb	charsets	goes	htptd	programming.php	linguistique	cms	rest
architect	mass	n3	charset	two	mod_rewrite	pear	écriture	bug_tracking	mime-type:text/html
emnekart	capacity	rdf	special.characters	focuses	http	smarty	tac	content	ws
architekt	issue	owl	programming/xml	like	markdown	technologies	wiki.startingpoints	project.management	next
bureau	domainnames	semantic_web	dtd	look	htaccess	thephpresourceindex	wiki.imp	projecto	web-services
büro	all101domainresources	semantic-web	xsl	e-business	webprogramming	skript	wiki	contentmanagement	web_services
architektur	evaluation	ontology	encoding	provides	scraping	programming/php	kn	faceted	uri
categorization	flow	metadata	xslt	wide	scripts	lamp		sitedevelopment	soap
vocabularies	squid	dublincore	xpath	similar	cache	completescripts		content-management	generation
taxonomy	ns	ontologies	xsd	specifications	gecko	scripting/php		bugtracking	protocol
topicmaps		agent	specification	description	pfw	php5		infomanagement	web/tech
engineering-docs		semantics	internationalization	we	caching	cms-groupware		imp.php	identifiers
facets		intelligent	microformats	bridge	webserver	bulletin		good-ideas	protocols
		multi	referenz	out	desarrollo	programaç;o		implementation	soa
		multiagent.com	xmlhttprequest	wsdl	pep	mysql		document	z39.50
		primer	xml-rpc	developperworks	revisit	spip		mambo	simple
		biblio	sgml.xml.perl	through	forsideblog	class		reader	flex
		interoperability	joelonsoftware		editor	comp.lang.php.tricks		providers	model
			utf-8		apache	programmazione		solutions	trees
			specs		delicious/code	remoting		pm	jonudell
			kt		templating	newsletter		projects	
			unicode		howto.build	phpbb		intranet	
			postel		web_server			drupal	
			parser		comp.www.server.apache			project	
			schema		servidores			lamache	
			spec		format			archival	
			xmlrpc		proxyproject			matrix	
			i18n		including			cm	
			xforms		firebird			frominternetexplorer	
			rtfm		cgi			nntp	
			dctagged		page			thesis	
			atom		mozilla				
			i10n		blogmarks				
			cocoon		xul				
			character		outils_informatique				
			introduction		mime				
			localization		dynamic				
			docbook		asp				
			svg		calendar				
					bug				
					ical				
					webdav				
					plone				
					iis				
					tmp				
					zope				

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #14									
Subcategory #131	Subcategory #132	Subcategory #133	Subcategory #134	Subcategory #135	Subcategory #136	Subcategory #137	Subcategory #138	Subcategory #139	Subcategory #140
biology	phys	documentations	mathematics	science	research	explain	distance-learning	education	cognitive
pathogen	sujets	oms	math	space	centres	enginyeria	free,	chineselanguage	brain
evolution	edu	csl	sci.math	astronomie	information+society	learn	free-course	pedagogy	brains
bioinformatics	ondes	researcher	problem_solving	espace	e-politics	cyclopedia	learn.improve	learning	culturedanimals
genome	electro-magnetisme	crossmedia	maths	nasa	assets	sci/tech	online-publishing	curriculum	a-life
creationism	seti	cscw	wolfram	astronomy	wealth	educational	opencourse	rmit	neuroscience
genetics	astro		calculus	sciencenews	profession	sermons	readingmaterial	ed	neuro
ecology	venus		exercises	cosmology	horse	mechanics	spinningsilk.org	training	medicine
nodalpoint			contests	moon	trade	ciencia	self_education	elearning	cognitivescience
dna			applet	physics	britain	lectures	free_knowledge	classes	anatomy
read_later			probability	mars	pbbkmar07	academia	opencourseware	teaching	cogsci
tag_later			sdu	exploration	institutions	openaccess	open,	school	neural
			chem	planets	workshop	user:djok	stufftoread	misc.learning	nanotechnology
			uczelnia	bilim	science&technology	open_access	lifelonglearning	lms	transhumanism
				wissenschaft	funding	children	web/edu	strength	medical
				ciencia		pub	randomlit	conditioning	disease
				earth		journals	onlinelearning	homeschooling	nano
				nova		oldbusiness	lifehack	e-learning	bio
				astronomia		ff-gev	research&science	assessment	cancer
				sci			higher-education	gradschool	biotech
				periodic			etudes		energy
				chemistry			akademiskesites		
				satellite			courseware		
				cosmos			free-media		
				nature			colleges		
				sciences			myblogs		
				quantum			favorite		
							lessons		
							publication		
							course		
							of		
							phd		
							cult		
							material		
							courses		
							org		
							mit		
							firefoxbookmarks		
							study		
							college		
							varios_revisar		
							bb		
							university		
							uni		

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #15									
Subcategory #141	Subcategory #142	Subcategory #143	Subcategory #144	Subcategory #145	Subcategory #146	Subcategory #147	Subcategory #148	Subcategory #149	Subcategory #150
pricechecking	finance	books&reviews	decorating	travel	tourisme	object	deals/vendors	norge	shopping
bargains	money	used&rarebooksellers	bayarea	flights	voyage	accessory	computershopping	low	t-shirts
bargain	investing	book_sellers	tiendas	airfare	transport	diy	onlinestores	hitch	clothes
techstores	stocks	bookstores	compras	airline	voyages	vendor	eshop	bed	clothing
deals	currency	used	weapons	booking	autos	component	surplus	spisedted	housewares
compare	bills	booksellers	guns	airlines	gas	construction	store	northern	shoes
coupons&rebates	credit	merchant	atlanta	travelhelp	cars	gear	sff	ferry	t-shirt
prices	bank	bookstore	georgia	air	car	pcb	vendors	happening	fashion
coupons	xe	ecommerce	geocaching	airplane	transportation	electronics	personalenrichment	linker	decor
stores	plan	gift	voodoo	travelling	auto	lcd	magasinage	cost	apparel
onlineshopping		pdx	trips	directions	paris	materials	cables	ferie	accessories
price		bücher	golf	backpacking	local	fabrication	consume	travelguide	abroad
sales		portland	fly	tourism	housing	custom	catalog	oslo	bags
landkarten		recreation	watches	discount	seattle	fab	motherboards	eating	kitchen
auction			rc	hotel	solar	parts	retail	save	gifts
auctions			wedding	hotels	vancouver	paint	hometheatre-avs	hostel	supplies
shop				train	vakantie	manufacturer	appliances	accommodation	catalogs
classifieds				cheap		recycling	shops	poor	fabric
buy				holiday			cases	acomodation	yarn
consumer				vacation			feldhaus	norway	bdsm
consumption				flight			carl	north	home
ebay				travelandleisure			wishlist	budget	imported:narn
craigslist				colorado			woodworking	reise	beauty
buying				camping					household
				outdoors					babies
				viajes					burningman
				trains					jewelry
				thailand					pet
				student					decoration
				hiking					cosmetics
				bus					flowers
				outdoor					sewing
				tokyo					plantsources
				rail					
				transit					
				parks					
				association					
				hotspot					
				karte					
				places					
				dublin					
				ireland					
				scotland					

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #16									
Subcategory #151	Subcategory #152	Subcategory #153	Subcategory #154	Subcategory #155	Subcategory #156	Subcategory #157	Subcategory #158	Subcategory #159	Subcategory #160
television	importedfavorites	entertainment	cinema	documentación	video	ie7-add-ons	movies_and_meaning	filmrelatert	importedbookmarks
movies&tv	swinglinks	goingout	film	travail	dv	stumbleupon	meaning	history_in_film	furled
tv	springfavorites	stayingin	directors	radio/tv/movie/music	documentaries	tango_nuevo_geneve	spirituality	imovies	review
ontheair	altri	stage	movies	arts	vidéo	w-kandinsky	faith	socialrecommendation	rating
angel	hamburg	adelaide	filmsock	user:managir	quicktime	soundfx-midi		entertainme	anime
commercials		health&fitness	movie	digitalasset	subtitles	ylmusicgroups		inglese	fansub
media&more		theatre	director	martial	hdtv	searches		movies&music	video_games
personalinterests		theater	films	culture-literature	dvd	graphic-design		recensioni	uncategorized
cable		unsortedgrouped	trailers	asian	filmmaking	tickets		media/entertainment	nonclassés
firefly		melbourne	actors		média	sparda		movies&television	
saturdaynightlive		listings	festival		películas	bookmarksformarzzball		tv,films,entertainment	
btvs			new		equipment	dandapani		movies/information	
			cine			rajendran		academy-lib&archive	
			short			mídia		toplinks	
			documentary			toolbar		naslagwerk	
			michael			reviews		mov	
			times			trash		películas	
			a			old		filmsandthings	
			kino			bill		base	
			starwars			nyc		amg	
								cin?a	
								ontspanning	
								imdb	
								actor	
								cultuuragenda	
								cinéma	
								recommendation	
								buisness	
								datenbank	
								filme	
								entretenimento	
								celebrity	
								guides	
								de	
								ratings	
								filmes	

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #17									
Subcategory #161	Subcategory #162	Subcategory #163	Subcategory #164	Subcategory #165	Subcategory #166	Subcategory #167	Subcategory #168	Subcategory #169	Subcategory #170
food	fk427	delivery	products	osx	pests	apple	ergonomics	horticulture	nutrition
recipes	mediabase	helsinki	problem	mac	ipm	macintosh	rsi	plants	health
vegetarian	case	cuisine	airport	software/osx	pesticides	os_x	keyboard	agriculture	diet
recipe		ppi	bluetooth	mac_software	insects	macosx	input	botany	surveying
cooking		newton	amazing	macstuff	coffee	macresources	yoga	fruit	geomatics
chicken			beos	macos	caffeine	shortcuts			misc.health
vegan			multiple	osx/apps	organic	panther			wellness
sushi				dashboard		hints			
drink				iphoto		macs			
food/drink				rendezvous		mac_os_x			
cheese				tiger		platforms/osx			
pizza				mac-software		powerbook			
sweets				applescript		x			
culi				cocoa		troubleshooting			
drinks				macsoft		operating-systems			
restaurants				darwin		safari			
restaurant				outlscollaboratifs		ibook			
tea				instantmessaging		mac_stuff			
dining				zeroconf					
bread				finder					
foodblogs				tinderbox					
diariesblogs				software/meta					
preservation				laptop24may2006					
orixas									
religiãoeafins									
eminglês									

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #18									
Subcategory #171	Subcategory #172	Subcategory #173	Subcategory #174	Subcategory #175	Subcategory #176	Subcategory #177	Subcategory #178	Subcategory #179	Subcategory #180
retro	linuxgames	games	nude	retrogaming	jeux	consola	rpg	gaming	game_theory
commodore	mmog	webgames	blonde	nintendo	jeuxetscénarios	emulador	worldbuilding	game.board	vgames/psych
spectrum	gamers	game_design	topless	emulation	jocs	user:joel	roleplaying	game.rules	virtual_geographic
c64	mmorpg	virtual_worlds	brunette	videogames	poker	arcade	d&d	boardgamegeek	vgames/general
retrocomputing	gamer	gamedesign		roms	milkandcookies	darkside	narrative	game.rpg	virtual_communities
nostalgia	mmo	unreal		cheats	racing	mame	játék	game	play
atari	penguin	gametheory		emulators		streetfighter	publisher	game.computer	virtual_environments
publicidade	cyberanthropology	multiplayer		console		emuladores		game.design	conqwest
basic	spiel	nomic		emu		juegos		bookmarksforpbanders	futurelab
classic	strategy	abandonware		videogame		rom		porsche914	
tl	mud	interactivefiction		gamecube		emulacion		automobilesanddriving	
amiga	in	if		oldschool		modding		engineandconversions	
		text.only		jeuxetemulation				dreamcast	
		shockwave		videojuegos				abstract	
		web-games		nes				saturn	
		adventure		faqs					
		ludology		gba					
		beginner		xbox					
		rpgs		mods					
		boardgames		gamedev					
		machinima		gameboy					
		nethack							
		board							
		ps2							
		xd							
		go							
		games.design							
		card							
		edtech							
		giochi							
		puzzles							
		puzzle							
		rubik							
		quake							
		cards							
		jogos							
		chess							
		spiele							
		ww2							

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #19									
Subcategory #181	Subcategory #182	Subcategory #183	Subcategory #184	Subcategory #185	Subcategory #186	Subcategory #187	Subcategory #188	Subcategory #189	Subcategory #190
imported_home_opera	mif	themepark	ff1.5bmark	imported,	kitplane	blair/jim	argentina	tosqueira	compassion
sms.at	africa	rollercoaster	pardelib	test,	fursuits	ts	gobierno	badtrip	cruelty
projekte	mali	coaster	stratégie	designs	mascoats	sg1	organizaciones	bizarro	animalrights
firmen,shops&produkte	senegal	roller	affaires_electroniques	croyance	flying	fic	spanish-español	brasil	fabricantes
immigration	united-states	theme	mba	hå	makers	rec'd	educación	noticia	ciscocco
immigration-tosort	middle-east	park	norsk	pagangeneral	aviation	slash	colombia	portugal	plastimodelismo
	musea		emploi	krutch	knitting	sentinel	marine		cerebus
	morocco		kasvit	marriage	needlework	vasút	sailing		
	morgaine		h-imported	news-story	canadian	regulation	robotik		
	quiz		m7		caravan	poland	computervision		
	europa		soucástky		knit		chile		
	sudan				online-shopping		venezuela		
	asia						courts		
	netherlands								
	egypt								
	fetish								
	anthropology								

HIERARCHICAL CATEGORISATION OF WEB TAGS FOR DELICIOUS

Category #20									
Subcategory #191	Subcategory #192	Subcategory #193	Subcategory #194	Subcategory #195	Subcategory #196	Subcategory #197	Subcategory #198	Subcategory #199	Subcategory #200
calibration	zzz-tosort	www/design	colour	accessibilità	webcolors	cor	mek.untagged	colorselection	clr
	ass	dev.web	colors	chooser	graphics.css	design/color	mek.imported	design/colortools	vision
		comp.tools	color	color.scheme	colortheory	colortools		farge	
		cssdesign	rgb	colorforweb	generador	cores		webpages	
		onlinetools	sci.color	colourschemas	spiffy	design/colors		tools/webdesign	
		webdesigntool		colour_schemes	uwsu	couleurs		wheel	
		designing		site::design		design.color		design_tool	
		devel/web		usabilità		webdesign/color		colorschema	
		webdizajn		webpalette		colours		colorpickers	
		comp.www.tools		onlinetool		webdesign-color		javascript	
		website_design		palette		colorpicker		littlehelpers	
		web_general		web/design/color		computerarts		art/photography	
		design-web-resources		harmony		palette			
		aesthetics		colourblindness		colores			
		interfaces		sweet		colourtheory			
		web_service		docdes		colorchooser			
		generator				webdev/ferramentas			
						couleur			
						website_tools			

REFERENCES

- [1] [Online]. Available: http://www.dai-labor.de/en/competence_centers/irml/datasets/
- [2] M. Grahl, A. Hotho, and G. Stumme, "Conceptual clustering of social bookmarking sites," in Proc. Int. Conf. Knowl. Manage. (IKNOW), Graz, Austria, Sep. 2007, pp. 356–364.
- [3] L. Specia and E. Motta, "Integrating folksonomies with the semantic web," in Proc. Int. Semantic Web Conf., 2007, pp. 624–639.
- [4] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stum, "Evaluating similarity measures for emergent semantics of social tagging," in Proc. Int. WWW Conf. ACM, 2009, pp. 641–650.
- [5] S. P. Lloyd, "Least squares quantization in PCM," IEEE Trans. Inform. Theory, vol. IT-28, pp. 129–137, Mar. 1982.