



Supporto alla navigazione nei sistemi di classificazione collaborativa: applicazione di tecniche semantiche alle folksonomie

Tesi di laurea di: David LANIADO Matr: 666797

Relatore: Marco COLOMBETTI

Correlatore: Davide EYNARD

Anno accademico 2005-2006
Sessione di laurea del 20 aprile 2007

Sistemi collaborativi

- **La cattedrale**

- “esperti”
- gerarchia
- isolamento



- **Il bazar**

- “utenti-sviluppatori”
- decentramento
- collaborazione
- condivisione





Folksonomie

- da *folk* + *taxonomy* (Vander Wals, 2004)
- gli utenti attribuiscono liberamente una o più parole chiave (*tag*) alle risorse
- due tipi di folksonomie:
 - *broad*: ogni risorsa può essere “taggata” da ogni utente
 - *narrow*: ogni risorsa può essere “taggata” solo da uno o pochi utenti
- modo per “ricordare in pubblico” (J. Schachter, 2004)
- vari contesti:
 - social bookmarking (del.icio.us)
 - fotografie (Flickr)
 - articoli scientifici (Bibsonomy)
 - articoli di blog (Technorati)

Tassonomie vs Folksonomie



- top-down
- esperti fanno il lavoro
- autorità centrale
- gerarchie di concetti
- vocabolario controllato
- *searching e browsing*
- bottom-up
- utenti fanno il lavoro
- “democrazia”
- spazio piatto
- parole chiave scelte liberamente
- *serendipity*



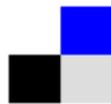
Folksonomie: vantaggi

- inclusive
- democratiche
- attuali
- *desire lines*
- rispecchiano il punto di vista degli utenti
- non binarie
- basso costo



Folksonomie: limiti

- non c'è controllo dei sinonimi
- scarsa precisione
- scarso *recall*
- mancanza di gerarchia
- *gaming*: possibilità di inquinare il sistema
- mancanza di standard



All items tagged **travel** – view **popular**

del.icio.us search

« earlier | later »

Airline tickets, Discount Airfare, Helping you find Cheap Flights At FareCompare save this

by arabella to travel airfare flights fare ... saved by 1330 other people ... 3 mins ago

Hotwire save this

by mikehinch to travel accounts ... saved by 1 other person ... 4 mins ago

BOOKING.COM online hotel reservations. Large selection of discounted hotels in all major destinations. Lowest rates, No payment in advance, Reservation is free, Cancellation is free save this

by atol557 to TRAVEL ... 4 mins ago

Cheap Airfare, Hotel Reservations, Car Rentals - Kayak.com save this

by arabella to travel airfare flights fare ... saved by 5642 other people ... 5 mins ago

South Africa hotels, lodges, guest houses, apartments, bed and breakfasts, holiday cottages, resorts save this

AA travel guides online

by pejrm to aa travel southafrica travelguides ... saved by 1 other person ... 6 mins ago

新宿、渋谷、池袋...、全国の駅周辺の情報を提供 / 駅街ガイド.jp save this

by mtso to travel japan ... saved by 4 other people ... 8 mins ago

related tags

- hotels japan pratumamo namo hotel pratu airlines flights thailand kong hong

active users

- sirpa wohler abbu underglasstable bradjg geekymama dreamaline asam.naseem chrisfoxinc

All items tagged **vacation** ([create tag description](#)) – view [popular](#)

 del.icio.us

« [earlier](#) | [later](#) »

[deep sea fishing](#) [save this](#)

Plan your next adventure travel vacation using our adventure travel company directory

by [gagarin777](#) to [deep sea fishing adventure travel vacation Andaman Nicobar islands island](#) ... 1 hour ago

<http://www.firstchoicevacationhomes.com/> [save this](#)

vacation home for rental,

by [darapan](#) to [vacation home](#) ... 2 hours ago

[Flickr: Photos from Casa Marcella](#) [save this](#)

We had a lot of maintenance issues with this Casa in Puerto Morelos. We recommend you don't rent it for your vacation unless you want to work on it for the owner. We paid them \$8000 to stay there for two weeks and had to perform our own maintenance becaus

by [pparks](#) to [Casa Marcella Puerto Morelos Vacation Rental](#) ... 3 hours ago

[Margarita Island forum: Night Life and Fun Activities - TripAdvisor](#) [save this](#)

by [oranges2oranges](#) to [vacation venezuela](#) ... 4 hours ago

[Felicità Japanese Garden](#) [save this](#)

by [domq](#) to [eastcoast vacation gardens](#) ... 4 hours ago

[Ireland Golf Vacation](#) [save this](#)

Golf Vacations in Ireland with the Irish golf vacation specialists. From all the famous golf links to the 'hidden gems' all over Ireland, Tailor-Made Golf ...

by [aileen695](#) to [Ireland Golf Vacation](#) ... 4 hours ago

[Harbor Gardens](#) ...

▼ **tags semantic tree**

- [-]location (10)
 - [-]region
 - [-]district
 - [+]country (8)
 - [+]state (1)
 - community (11)
 - state (6)
 - [-]borough
 - brooklyn (1)
 - bronx (1)
 - dc (1)
 - [-]geographical_area
 - [-]municipality
 - [+]city (11)
 - [+]town
 - [+]tract
 - caribbean (2)
 - northwest (1)
 - patagonia (1)
 - hotspot (1)
 - auvergne (1)
 - mideast (1)
 - riviera (1)
 - surroundings (1)
 - new_england (1)
 - wold (1)
 - wilderness (1)
 - new_world (1)
 - [+]area
 - cambria (1)
 - [-]point (1)

▼ **related tags**

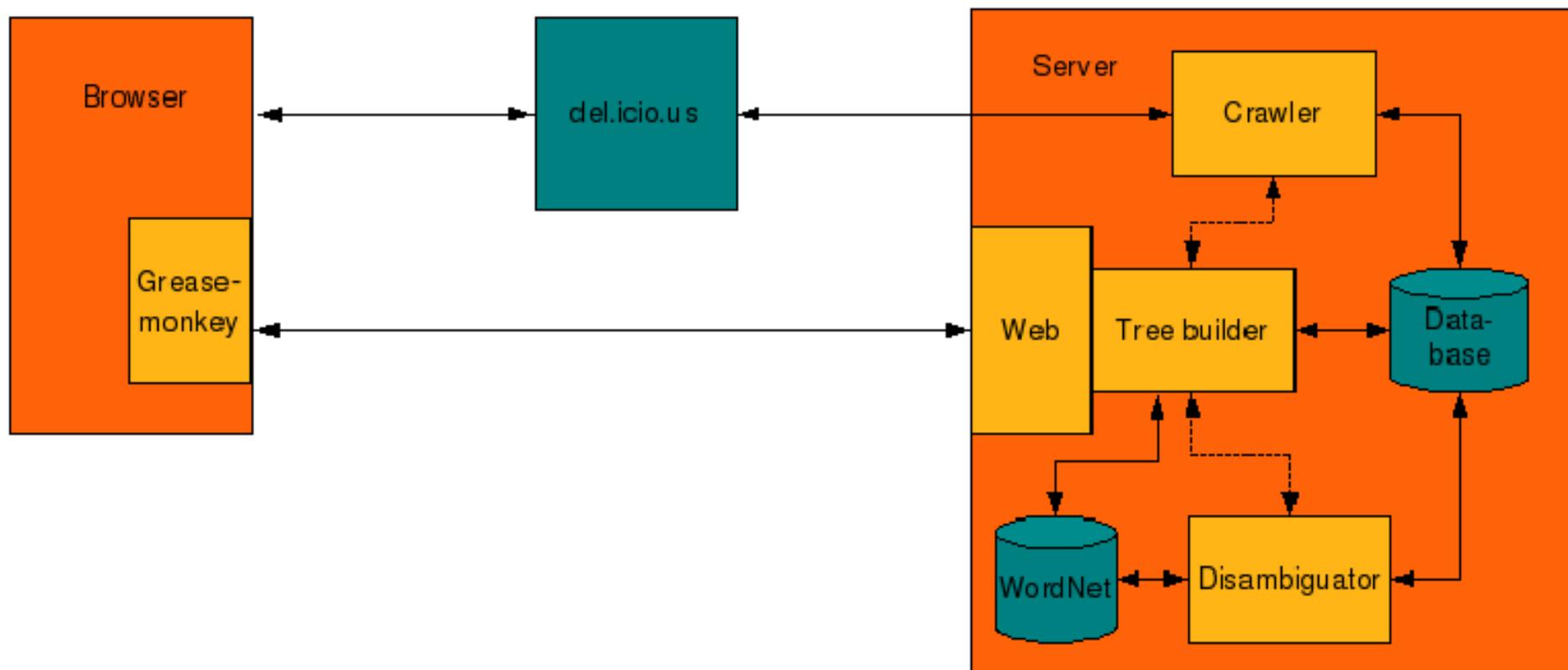
- travel
- tourism
- villa
- lodging
- lake
- marketing
- condo
- tahoe
- business
- villas
- and



Scelte di contesto

- folksonomia broad -> del.icio.us
- ontologia -> relazione di iperonimia sui sostantivi di WordNet: categorie generali contenute implicitamente nel linguaggio
- navigazione attiva (Firefox + Greasemonkey) -> estensione del browser per arricchire dinamicamente le pagine visualizzate con informazioni semantiche aggiuntive
- linguaggio di programmazione -> Perl

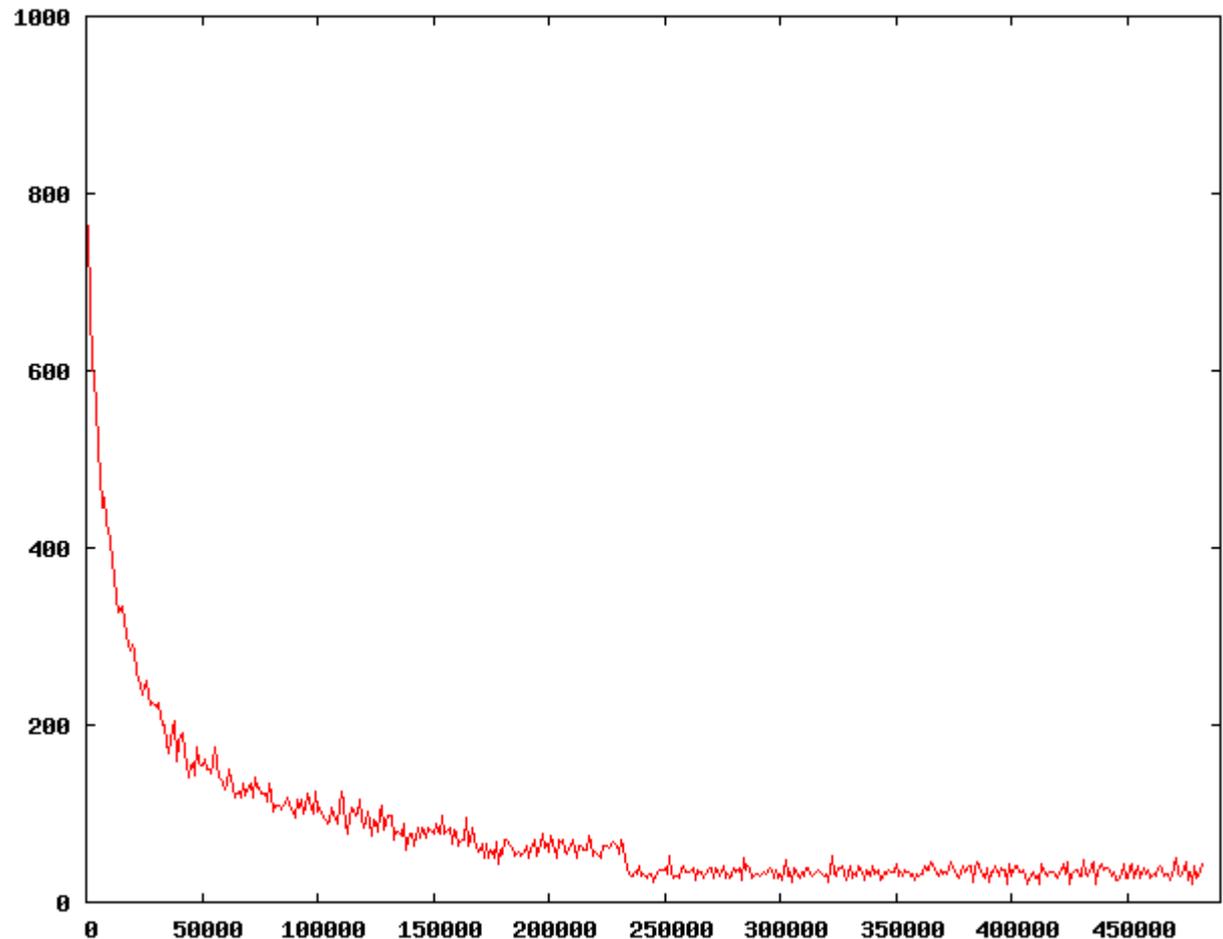
Architettura



Tag non presenti in WordNet



- Dati relativi a 30.000 utenti di del.icio.us:
 - 480.000 diverse tag
 - solo l'8% delle tag sono presenti in WordNet
 - oltre 20 milioni di azioni di tagging
 - il 68% delle volte la tag usata è presente in WordNet





Disambiguazione delle tag

- polisemia: Wordnet si basa su *synset*, gruppi di sinonimi che costituiscono le unità di significato fondamentali; ogni parola può appartenere a più *synset* -> occorre risolvere le ambiguità
- disambiguazione di ogni tag in relazione alla risorsa a cui si riferisce
 - contesto: le altre tag riferite alla stessa risorsa
 - metriche di correlazione: adapted lesk, Hirst & St. Onge
- algoritmo:
 - disambigua le prime N tag più usate per ogni risorsa confrontandole fra loro
 - disambigua ciascuna delle restanti tag confrontandola con le prime N
- vantaggi:
 - poco sensibile al rumore: si basa solo sulle tag più utilizzate
 - efficiente: il numero di confronti da effettuare non cresce esponenzialmente col numero delle tag attribuite a una stessa risorsa, ma resta limitato

L'albero semantico delle tag



- l'albero è un sottoinsieme della gerarchia dei concetti di WordNet, che comprende tutte le tag correlate a quella di partenza
- ogni nodo corrisponde a una tag (disambiguata), o a un concetto che sia iperonimo di almeno una tag
- a ogni nodo sono associati due attributi:
 - il peso del nodo stesso, ovvero il numero di siti etichettati con quella parola in quella accezione
 - il peso del relativo sottoalbero, ovvero la somma di tutti i pesi dei nodi sottostanti
- passi necessari:
 - costruzione dell'albero
 - compressione
 - ordinamento dei rami in base al loro peso
 - stampa del risultato in formato HTML o XML



Costruzione dell'albero

- per ogni tag in una particolare accezione:
 - costruisci il cammino dalla tag alla radice dei concetti di WordNet
 - *merge* del nuovo cammino con l'albero già esistente, partendo dalla radice
- poichè l'albero ha dimensioni limitate, la complessità dell'algoritmo è lineare nel numero di tag correlate ricevute in ingresso e dipende da:
 - il fattore medio di ramificazione dell'albero b
 - la profondità dell'albero d



Compressione dell'albero

- problema: elevata granularità di WordNet
- esigenza: compattare l'albero senza renderlo troppo piatto e senza eliminare concetti essenziali
- algoritmo:
 - elimina i nodi che corrispondono a categorie di alto livello, fissate in una lista (come “ physical entity” o “causal agent”)
 - elimina i nodi interni dell'albero che:
 - non corrispondono a nessuna tag e
 - hanno un numero di figli minore di K oppure non hanno fratelli
- il valore predefinito di K è 2, ovvero solo i nodi che hanno un figlio solo vengono eliminati: la struttura della gerarchia viene preservata, ma i concetti più specifici possono salire nell'albero



All items tagged **pasta** — view [popular](#)

del.icio.us ▾

« [earlier](#) | [later](#) »

[Sweet Potato Gnocchi with Brown Butter and Sage](#) [save this](#)

by [questhoya](#) to [recipes](#) [pasta](#) [gnocchi](#) [sweet-potato](#) [ricotta](#) ... [saved by 6 other people](#) ... 1 hour ago

[Pasta with cauliflower - The Boston Globe](#) [save this](#)

by [dcrowley](#) to [recipes](#) [pasta](#) ... 1 hour ago

[Weight Watchers Spaghetti Carbonara Recipe-Kitchen Crafts 'n' More](#) [save this](#)

by [plm5087](#) to [pasta](#) ... 6 hours ago

[Aglio e olio - en utflykt i det italienska köket: PASTA MED MOROT, SPENAT OCH BALSAMVINÅGER eller PASTA CON CAROTE, SPINACI E ACETO BALSAMICO](#) [save this](#)

by [horrorhead](#) to [mat recept](#) [pasta](#) [huvudrätt](#) [lunch](#) ... 8 hours ago

[The Kitchen Pantry: Noodles, again and again and again...](#) [save this](#)

by [skippytp](#) to [recipes](#) [pasta](#) [asian](#) ... [saved by 2 other people](#) ... 14 hours ago

[Fish on Fridays: Linguine with Clam Sauce](#) [save this](#)

by [lofiEDDIE](#) to [recipes](#) [clam](#) [pasta](#) ... 15 hours ago

[Chicken or Turkey Tetrazzini - Allrecipes](#) [save this](#)

by [hulmeka](#) to [chicken](#) [turkey](#) [tetrazzini](#) [dinner](#) [recipes](#) [pasta](#) [comfort](#) ... [saved by 1 other person](#) ... 15 hours ago

[cooking: Spaghetti Sauce?](#) [save this](#)

▼ **tags semantic tree**

- [-] **food** (102)
 - [-] **pasta** (283)
 - [macaroni](#) (9)
 - [noodle](#) (9)
 - [spaghetti](#) (6)
 - [ravioli](#) (3)
 - [gnocchi](#) (2)
 - [penne](#) (2)
 - [vermicelli](#) (1)
 - [tortellini](#) (1)
 - [lasagna](#) (1)
 - [rigatoni](#) (1)
 - [lasagne](#) (1)
 - [fettuccine](#) (1)
 - [-] **produce**
 - [+] [vegetable](#) (12)
 - [+] [edible_fruit](#)
 - [veggie](#) (2)
 - [-] **meat** (5)
 - [+] [poultry](#)
 - [sausage](#) (9)
 - [+] [beef](#) (7)
 - [+] [cut](#)
 - [pork](#) (5)
 - [veal](#) (1)
 - [lamb](#) (1)
 - [+] [seafood](#) (16)
 - [+] [fish](#) (6)
 - [+] [baked_goods](#)
 - [chocolate](#) (1)
 - [takeout](#) (1)
 - [-] **communication**
 - [+] [message](#)
 - [+] [written_communication](#)

▼ **related tags**

- [cooking](#)
- [recipes](#)
- [vegetarian](#)
- [italian](#)
- [healthy](#)
- [main](#)
- [parmesan](#)
- [food](#)
- [sauce](#)
- [sausage](#)
- [gnocchi](#)

▼ tags semantic tree

- [-]food (102)
 - [-]pasta (283)
 - macaroni (9)
 - noodle (9)
 - spaghetti (6)
 - ravioli (3)
 - gnocchi (2)
 - penne (2)
 - vermicelli (1)
 - tortellini (1)
 - lasagna (1)
 - rismall ring-shaped stuffed pasta
 - lasagne (1)
 - fettuccine (1)
 - [-]produce
 - [+]vegetable (12)
 - [+]edible_fruit
 - veggie (2)
 - [-]meat (5)
 - [-]poultry
 - chicken (13)
 - turkey (1)

▼ ▼ related tags

- cooking
- recipes
- vegetarian
- italian
- healthy
- main
- parmesan
- food
- sauce
- sausage
- gnocchi

<ul style="list-style-type: none"> [-] vegetable (12) <ul style="list-style-type: none"> tomato (17) [-] squash (1) <ul style="list-style-type: none"> zucchini (5) [-] hot_pepper <ul style="list-style-type: none"> chili (2) chilli (1) potato (3) broccoli (3) radicchio (2) chou (2) mushroom (2) green (2) eggplant (1) cauliflower (1) artichoke (1) onion (1) [-] edible_fruit <ul style="list-style-type: none"> lemon (2) 	<ul style="list-style-type: none"> italian healthy main parmesan food sauce sausage gnocchi
--	---

sites tagged with pasta and broccoli on del.icio.us

« earlier | later »

Filmed by Bike [save this](#)

by nylon60mm to festival film Portland bicycle movie art community Cycling ... [saved by 17 other people](#) ... 1 hour ago

Classic Rendezvous Main [save this](#)

by nylon60mm to vintage history gallery bicycle ... [saved by 44 other people](#) ... 1 hour ago

Alta Bikes [save this](#)

Alta Bikes - 1 girls design bysykkel. Designet av Bleed, Norway Says, Frost. Adv. av Schnell&Melnychuck.

by inspirasjon to Design inspiration norway advertising bike bikes cycling bicycle cool ... [saved by 64 other people](#) ... 1 hour ago

BikeCult [save this](#)

by nylon60mm to nyc bicycle culture bikeshop ... [saved by 34 other people](#) ... 1 hour ago

Bicyclepaintings.com - Bicycle Paintings, by Taliah Lempert [save this](#)

by nylon60mm to art painting bicycle artist ... [saved by 25 other people](#) ... 2 hours ago

William F. Yurasko's Web site v.XII - Updated Daily [save this](#)

by johncabell to bicycle w&od ... 5 hours ago

Seattle International Randonneurs [save this](#)

by struan to seattle bicycle randonneur ... 6 hours ago

MRP Mountain Racing Products [save this](#)

by thinkalone to bicycles bicycle bike hiking cycling bicycling ... [saved by 13 other people](#) ... 7 hours ago

Mountain bike product reviews forums classifieds trails [save this](#)

by allenp109 to bicycle ... [saved by 33 other people](#) ... 7 hours ago

- Hutchinson Fusion Team Series Tour De France (Special Edition)

700x23 Tire [save this](#)

by allenp109 to bicycle ... [saved by 9 other people](#) ... 7 hours ago

RevoPower - Motorized Replacement Bike Wheel [save this](#)

by evangaffney to bicycle bicycles bike energy environment green revopower bikes engine

▼ tags semantic tree

- [+] instrumentality
- [+] activity (1)
- [+] location
- [+] living_thing
- [+] communication
- [+] facility (1)
- [+] social_group
- [+] action
- [+] mercantile_establishment
- [+] content
- [+] product (5)
- [+] group_action
- [+] measure
- [+] process
- [+] state
- [+] part (5)
- [.] quality
 - fitness (13)
 - resource (7)
 - power (2)
 - usability (1)
 - simplicity (1)
- [.] property
 - [+] chromatic_color
 - sustainability (6)
 - fashion (4)
 - extreme (2)
 - vintage (2)
 - lifestyle (2)
 - profile (1)
- [.] possession
 - [.] assets
 - [+] resource (6)
 - [+] share (2)
 - net (1)
 - [.] transferred_property
 - gift (4)
 - [.] cost
 - [+] payment
 - price (1)
 - rental (1)
- [+] decoration
- [+] covering
- [+] representation
- [.] way
 - [.] road (5)
 - trail (4)
 - street (2)
 - route (3)

▼ related tags

- [bike](#)
- [cycling](#)
- [lighting](#)
- [comparison](#)
- [lights](#)
- [north-carolina](#)
- [reviews](#)
- [bikes](#)
- [diy](#)
- [planning](#)
- [nc](#)



All items tagged **recipe** – view **popular**

del.icio.us search

« earlier | later »

Gothamist: Making Chocolate Babka For the Holidays save this

by nina to recipe ... 37 mins ago

eG Forums -> chocolate babka recipe save this in search of chocolate babka

by nina to food recipe ... 38 mins ago

レシピ検索No.1/料理レシピ載せるなら クックバツド save this

by hiroking to レシピ food cooking recipe 料理 ... saved by 326 other people ... 45 mins ago

Exclusively Food: Salmon Bites Recipe save this

by estelleihazs to food recipe savoury sweet ... 1 hour ago

VENISON DAUBE WITH CUMIN AND CORIANDER Recipe at Epicurious.com save this

by stevenecrane to recipe ... 1 hour ago

CUMIN HERB RICE PILAF Recipe at Epicurious.com save this

by stevenecrane to recipe ... saved by 1 other person ... 1 hour ago

biga deal save this

by randlejl to recipe bread ... saved by 2 other people ... 1 hour ago

日清製粉グループ: イタリア料理レシピ save this

by Exhumed to pasta italian recipe cuisine ... saved by 4 other people ... 2 hours ago

tags semantic tree

- [+]communication (2)
- [-]food (180)
 - [-]baked_goods
 - [+]cake (15)
 - [-]bread (18)
 - muffin (4)
 - [+]bun (1)
 - toast (2)
 - scone (1)
 - garlic_bread (1)
 - roll (1)
 - saltine (1)
 - [+]pastry (1)
 - [+]produce (1)
 - [+]meat (8)
 - [+]pasta (13)
 - chocolate (21)
 - [+]seafood (6)
 - [+]fish (3)
 - [-]breakfast_food
 - [+]cereal
 - muesli (1)
 - coconut (2)
- [-]food (7)
 - [-]nutriment
 - [+]course
 - [+]dish
 - [+]meal (10)
 - [+]sweet
 - [+]vitamin
 - mince (1)
 - delicacy (1)
 - [+]foodstuff
 - [+]beverage (5)

related tags

- food
- cooking
- dessert
- pasta
- salad
- beef
- レシピ
- sausage
- recipes
- baking
- lamb

« [earlier](#) | [later](#) »

Holistic Learning

[save this](#)
by [fournier48](#) to [blog ...](#)
just posted

[トレンドーズ社長 経
沢香保子の「人生を
味わい尽くす」プロ
グ](#) [save this](#)

by [tomojp](#) to [blog ...](#)
saved by 1 other
person ... just posted

Seth Godin article - Be a Better Liar.

[save this](#)
by [fflmike](#) to [blog
business article ...](#)
[saved by 20 other
people](#) ... just posted

[My RSS 管理人 プロ
グ \(工事予定\) : RSS
記事のクリック率
7% のほとんどは
「ポット」](#) [save this](#)

by [kenkity](#) to [rss Blog ...](#)
[saved by 12 other
people](#) ... just posted

L8dybug's Journal

[save this](#)

▼ tags semantic tree

- [+]written_communication
- [+]instrumentality
- [+]person (9)
- [+]activity
- [+]message (1)
- [-]content (6)
 - [-]knowledge_domain
 - [-]discipline
 - [-]science (113)
 - politics (184)
 - [-]economics (47)
 - finance (6)
 - psychology (34)
 - [-]physics
 - electronics (22)
 - optics (1)
 - astronomy (1)
 - [+]biology (10)
 - government (21)
 - math (15)
 - [+]geography (13)
 - [+]mathematics (10)
 - [+]anthropology (7)
 - maths (6)
 - ip (6)
 - medicine (5)
 - [+]linguistics (2)
 - sociology (3)
 - taxonomy (3)
 - [+]chemistry (1)
 - geology (1)
 - nlp (1)
 - technology (414)
 - [+]engineering

▼ related tags

- design
- web2.0
- news
- blogs
- art
- technology
- sex
- webdesign
- inspiration
- music
- web

▼ active users

- BazookaDance
- birdphone
- unfoldingrose
- jrhyen
- pensar_custa
- miguelyn
- jameskoole
- IzuXIII
- sacrifice94
- rightmindx
- luochao1987
- yark
- chengshan
- lokidesign
- futher
- eckartwalther
- joetomczak

All items tagged **health** – view **popular**

 del.icio.us

« [earlier](#) | [later](#) »

[Multiple Personality Disorder-DID](#) [save this](#)

by [ademajo](#) to [MentalHealth](#) [mental health](#) [medicine](#) [interesting](#) [dissociative](#) [mind&body](#) ... 19 mins ago

[RIDE Project - Semantic technology for eHealth](#) [save this](#)

by [sionbach](#) to [snomed](#) [health](#) [semanticweb](#) ... saved by 1 other person ... 19 mins ago

[Natural Medicine, Herbal Remedies–Treatment of ADD ADHD Anxiety UTI and more](#) [save this](#)

Great Natural Remedy site

by [dluntz](#) to [herbal](#) [natural](#) [treatment](#) [remedies](#) [health](#) ... saved by 6 other people ... 20 mins ago

[MediCorp Health System | Community Calendar](#) [save this](#)

Classes at Mary Washington Hospital

by [endswell](#) to [health](#) [scouts](#) ... 20 mins ago

[Navy SEALs.com - US Navy SEAL Workout](#) [save this](#)

by [dajamerson](#) to [Fitness](#) [workout](#) [health](#) ... saved by 392 other people ... 20 mins ago

['Cannabis' may help mentally ill](#) [save this](#)

by [deanjb](#) to [health](#) [bipolar](#) [cannabis](#) [medical](#) ... 22 mins ago

[www.whonamedit.com](#) [save this](#)

by [chimpky](#) to [reference](#) [science](#) [search](#) [tools](#) [health](#) [history](#) ... saved by 89 other people ... 22 mins ago

[blog.myspace.com/independz](#) [save this](#)

by [ralphjarmon](#) to [health](#) [blog](#) ... 25 mins ago

[Aging - State of Aging and Health Report](#) [save this](#)

by [afeeney](#) to [aging](#) [health](#) [demographics](#) [healthcare](#) ... saved by 1 other

▼ **tags semantic tree**

- [+]communication (2)
- [+]state
- [+]activity
- [+]living_thing
- [+]content (1)
- [+]instrumentality
- [+]food (16)
- [+]location
- [+]social_group
- [+]work (1)
- [+]agent
- [+]measure
- [+]process
- [+]quality (1)
- [+]structure
- [+]property
- [+]body_part
- [+]process
- [+]action
- [+]food (29)
- [+]group_action
- [+]possession
- [+]chemical_element
- [+]information (1)
- [+]trait
- [+]natural_object
- [+]shape
- [+]ability
- [+]facility
- [+]representation
- [+]magnitude_relation
- [+]material
- [+]happening

▼ **related tags**

- sex
- number
- discover
- penis
- source
- exercise
- enlargement
- size
- pills
- pill
- male

▼ **active users**

- [griffindorluv](#)
- [jmresearch](#)
- [gcubed](#)
- [dhanam](#)
- [kasthuri1981](#)
- [jude2004](#)
- [OutdoorCine](#)
- [tonysutherland](#)
- [marnroo](#)
- [mommychele](#)



Conclusioni

- il filtraggio delle tag attraverso una tassonomia di concetti contenuti implicitamente nel lessico di una lingua può migliorare le possibilità di esplorazione di una folksonomia
- si possono risolvere parzialmente alcuni problemi delle folksonomie:
 - mancanza di gerarchia
 - controllo dei sinonimi
 - gaming
- sviluppi futuri:
 - miglioramento dell'algoritmo di compressione dell'albero
 - ampliamento dell'ontologia, integrando WordNet con ontologie di dominio e wordnet locali multilingua
 - raccolta dei dati da diverse folksonomie
 - utilizzo della disambiguazione delle tag per filtrare le risorse