



# A Semantic Tool to Support Navigation in a Folksonomy

David Laniado

August, 23 2007  
HP Research Labs, Palo Alto

# Collaborative Systems



- The **cathedral**

- “experts”
- central authority
- isolation



- The **bazaar**

- “users-developers”
- decentralization
- information sharing
- “organized chaos”





# Folksonomies

---

- *folk + taxonomy* (Vander Wal, 2004)
- users tag resources with freely chosen keywords
- two kinds of folksonomies:
  - *broad*: each resource can be tagged by any user
  - *narrow*: each resource can be tagged only by some users
- a way to “remember in public” (J. Schachter, 2004)
- different contexts:
  - social bookmarking (del.icio.us)
  - photographs (Flickr)
  - scientific papers (Bibsonomy)
  - blog posts (Technorati)

# Taxonomies vs Folksonomies

---



- top-down
- experts do the job
- central authority
- hierarchies of concepts
- controlled vocabularies
- *searching and browsing*
- bottom-up
- users do the job
- “democracy”
- flat space
- freely chosen keywords
- *finding and serendipity*

# Folksonomies: advantages

---



- inclusive
- democratic
- actual
- *desire lines*
- reflecting the users point of view
- overlapping
- low cost

# Folksonomies: limitations

---



- no synonym control
- lack of precision
- lack of recall
- lack of hierarchy
- possibility of *gaming*
- lack of standards



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 because

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

[Harbour 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)
        - midwest (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



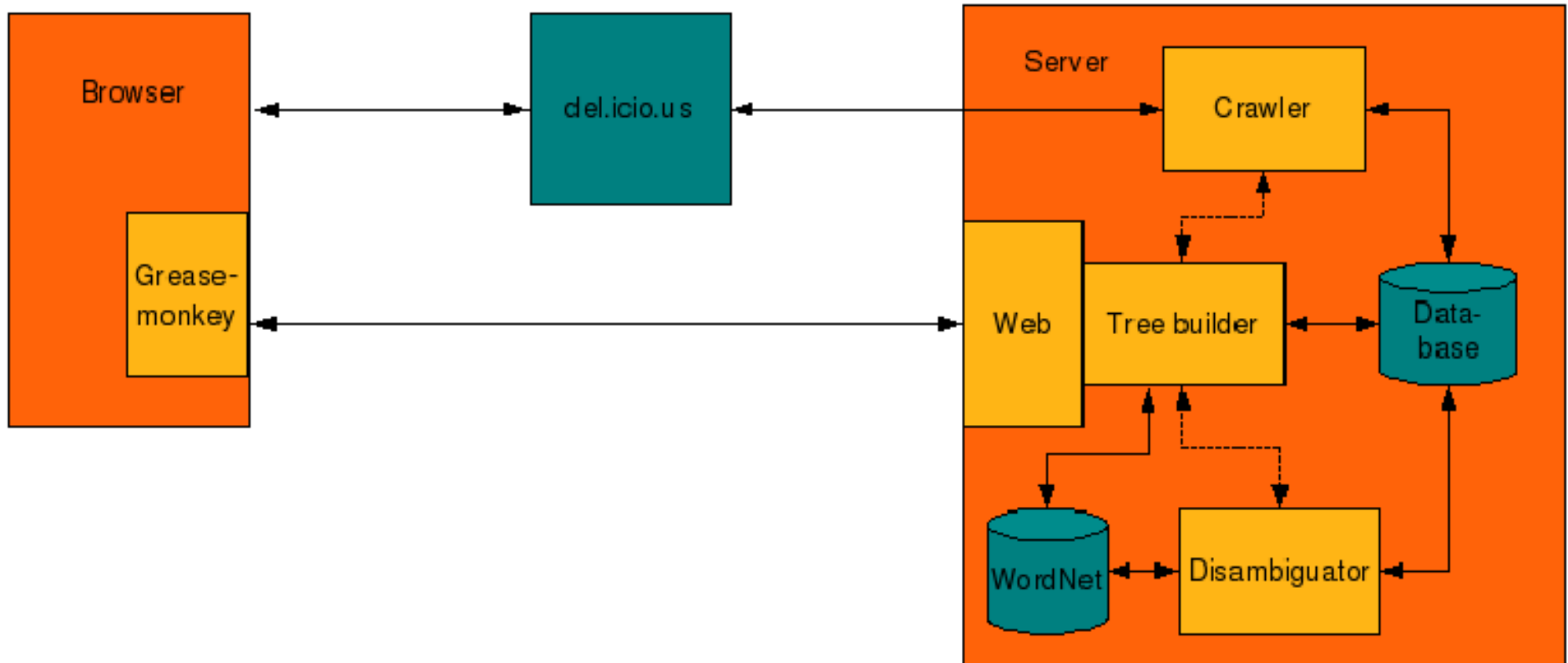


# Context choices

---

- a popular broad folksonomy -> del.icio.us
- ontology -> hypernymy relation on WordNet nouns: general categories implicitly contained in natural language
- *active navigation* -> Firefox + Greasemonkey: browser extension to dynamically enrich the pages with additional semantic information
- programming language -> Perl

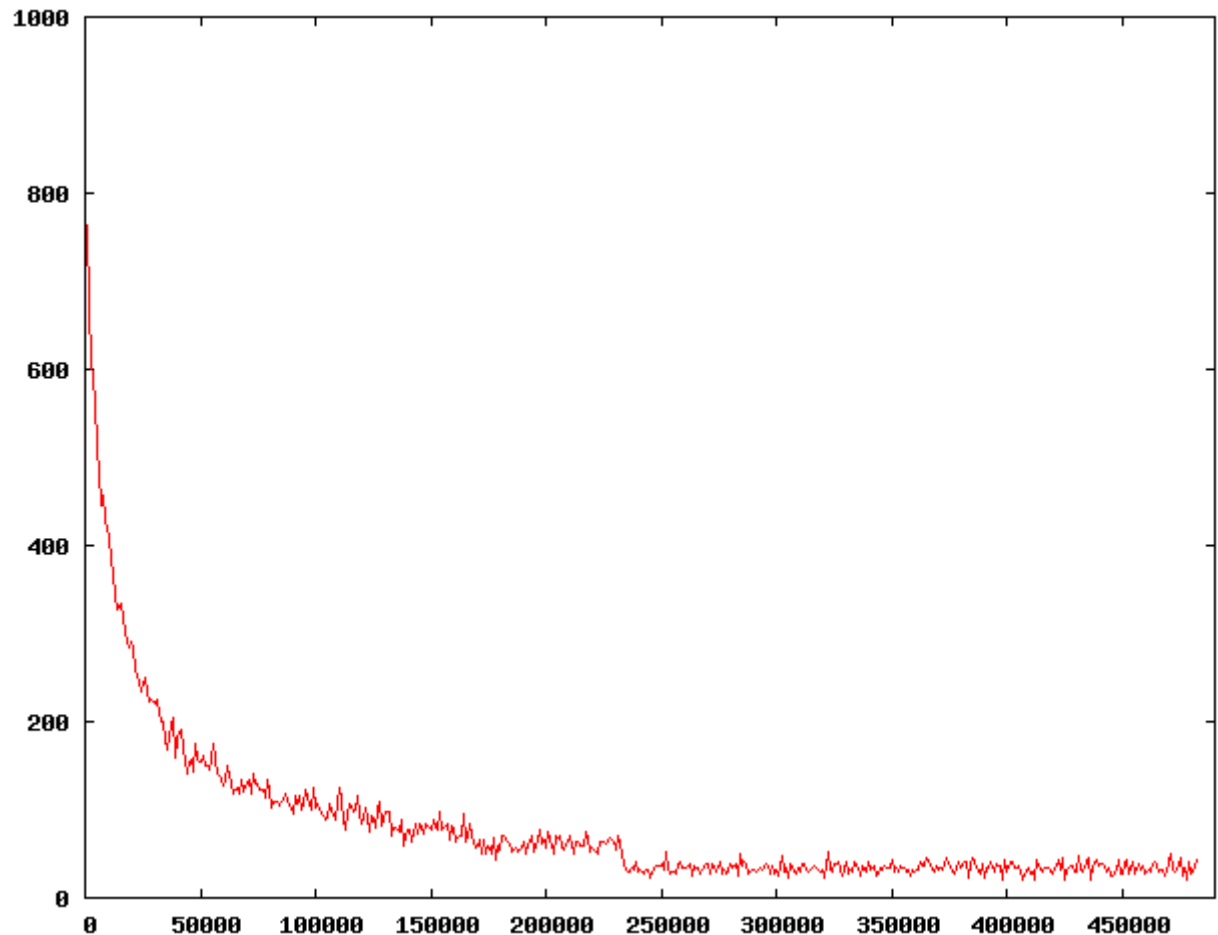
# System architecture



# Tags contained in WordNet



- Data relative to 30.000 del.icio.us users:
  - 480000 different tags
    - only 8% of tags are contained in WordNet
  - over 20 million tagging actions
    - 68% of the times the tag used is contained in WordNet





# Tag disambiguation

---

- polisemy: words can have more than one meaning -> need to solve ambiguities
- disambiguation of each tag in relation to the resource it is referred to
  - context: the other tags referring to the same resource
  - correlation metrics: adapted lesk, Hirst & St. Onge
- algorithm:
  - for each resource, take the first N most used tags, and compare them to each other to find the most likely meaning
  - disambiguate each of the remaining tags by comparing it to the first N
- advantages:
  - less noise sensitive: based only on the most used tags
  - efficient: the number of comparisons to be effected doesn't grow exponentially with the number of tags attributed to the same resource



# The tag semantic tree

---

- the tree is a subset of WordNet concept hierarchy, containing all the tags related to the chosen one
- each node represents a (disambiguated) tag or a concept, that is hypernym of at least one tag
- each node has two attributes:
  - the node's weight(number of resources tagged with that word in that acceptance)
  - the subtree weight (the sum of the weights of all nodes beneath)
- steps:
  - tree building
  - compression
  - branches ordering, according to their weight
  - result output, in HTML or XML format



# Tree building

---

- for each tag in a particular acceptation:
  - build the hypernym chain from the tag to WordNet roots of concepts
  - *merge* the new path with the existing tree, starting by the root
- as the tree has limited dimensions, the algorithm complexity is linear with the number of tags received in input and depends on:
  - the medium branching factor  $b$  of the tree
  - the tree depth  $d$

# Tree compression

---



- problem: WordNet's fine granularity
- need: compressing the tree without making it too flat or losing any interesting concept
- algorithm:
  - delete nodes corresponding to high level categories, fixed in a list (such as “physical entity“ or “causal agent”)
  - delete internal nodes that:
    - do not correspond to any tag and
    - have less than K children or have no siblings
- default value for K is 2, so the hierarchy structure is preserved, but more specific concepts can rise in the tree



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

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 [chimpsky](#) 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](#)





# Conclusions

---

- Filtering tags through a taxonomy of concepts, implicitly contained in a language lexicon, may improve the possibilities of exploring a folksonomy
- some of the main limitations of folksonomies can be partially solved:
  - lack of hierarchy
  - synonym control
  - gaming
- future work:
  - improving the tree compression algorithm
  - integrating WordNet with domain ontologies and multilingual wordnets
  - extracting data from different folksonomies
  - using tag disambiguation to filter the resources