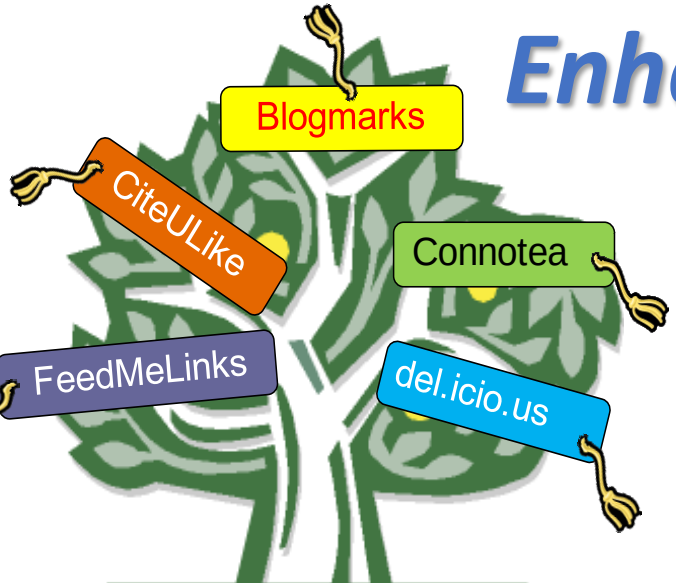
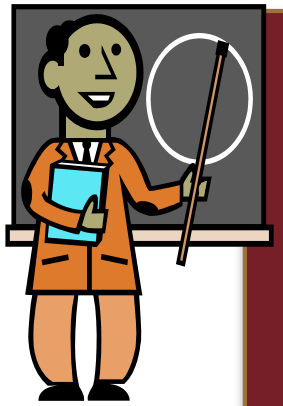
Taxonomy-Folksonomy Integration for Enhanced Knowledge Navigation



Eric Tsui & CC Kiu

Definition



"Taxonomies are composed of taxonomic units that are arranged frequently in a hierarchical structure. Typically they are related by parent-child relationships." –

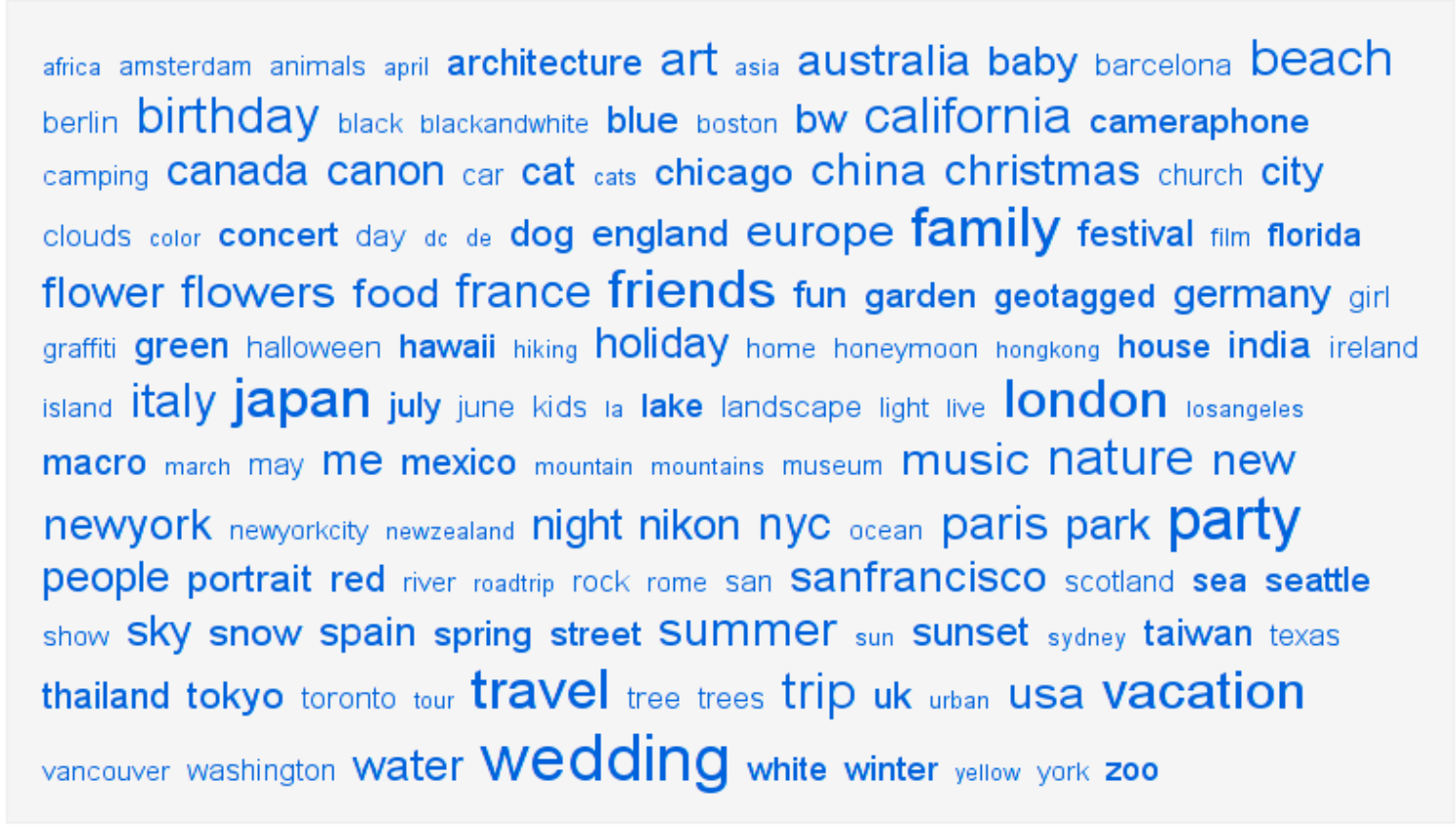
http://en.wikipedia.org/wiki/Taxonomy

"Folksonomy is a practice and method of collaboratively creating and managing tags to annotate and categorize content" –

http://en.wikipedia.org/wiki/Folksonomy

Tag Cloud: Most popular tags in Flickr

All time most popular tags



Characteristics

Taxonomy

- Central control / Created by corporate/taxonomist(s)
- Top down
- Navigation
- Rigid
- Corporate-defined vocabulary

Folksonomy

- Democratic creation / Created by the users
- Bottom up
- Association/Discovery
- Dynamic and extensible
- Personal vocabulary

Taxonomy + Folksonomy

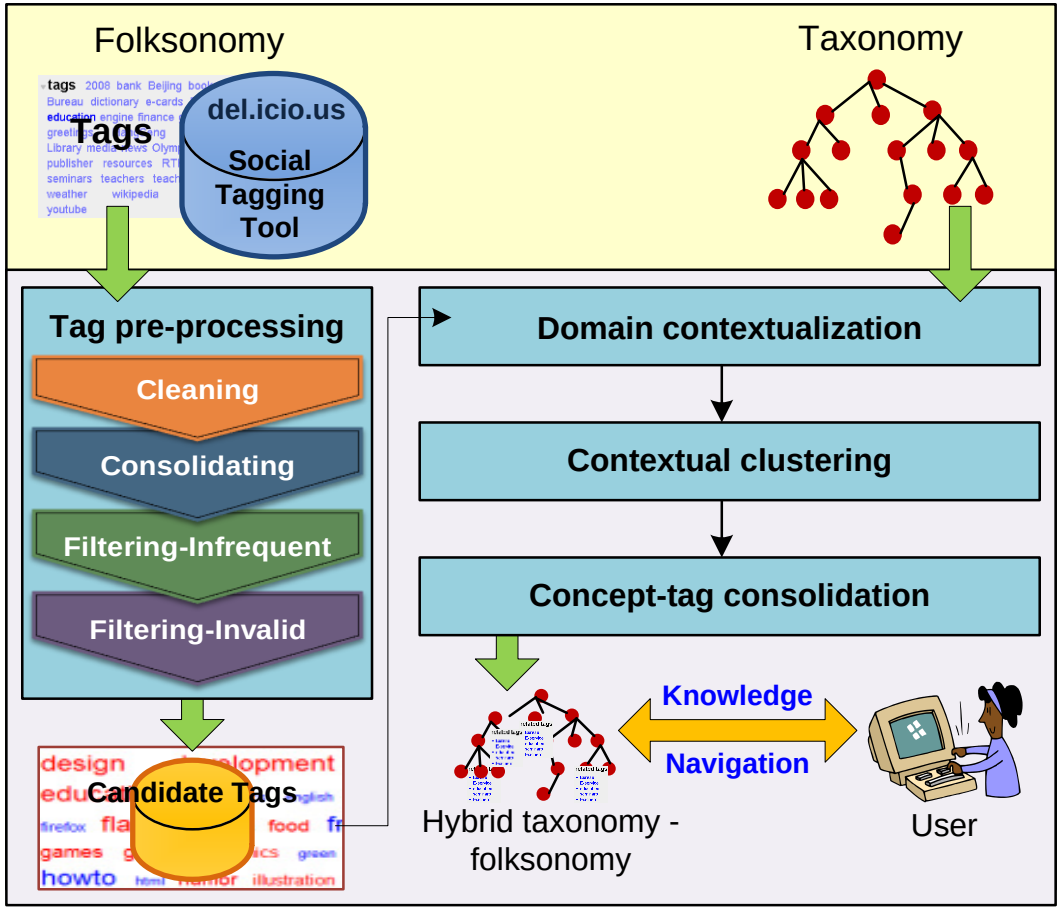
Taxonomy		Folksonomy	=	Taxonomy Hybrid Models
Central Control		Democratic Creation		Central control with continuous user input
Top-down		Bottom-up		Meet in the middle
Meaning to the author		Meaning to the reader		Cater for the growing community
Tedious process for making changes	+	Just do it	=	Suggestions, additions, deletions with governance models
Accurate		Good enough		Community validated and tested
Navigation		Discovery		From navigation to discovery by leveraging mass input
Restrictive		Expansive		Flexible and evolving
Defined Vocabulary		Personal Vocabulary		Community Vocabulary

Dow Jones Factiva & Michael Sampson

Taxonomy + Folksonomy [Challenges]

- Need to preserve the taxonomic structure
- Variation of languages
- No structure among tags
  - Tags are merely individual words
  - Space is not allowed in a tag
- Lots of misspellings
- Need to recover word stem from word tenses and parts of speech
- Abbreviations/Acronyms/Aliases exist
- Varying importance and frequency of tags
- Automate and customize the entire process

Our Approach [Methodology]



Techniques Used

- String normalization
- WordNet
- Wikipedia
- Data mining techniques
  - Formal concept analysis
  - K-means
  - Simple matching coefficient

TaxoFolk

A Hybrid Taxonomy - Folksonomy For Enhanced Knowledge Navigation

Attractions
- Most Popular
  - Hong Kong Island
    - The Peak
    - Madame Tussauds
    - Hollywood Road, Cat Street / Man Mo Temple
    - Western Market
    - Lan Kwai Fong & SoHo
    - Happy Valley Racecourse
    - Jumbo Kingdom
    - Ocean Park
    - Repulse Bay
    - Stanley Market & Murray House
    - Aberdeen
  - Kowloon
    - Ladies Market
    - Temple Street Night Market
    - Jade Market & Jade Street
    - Avenue of Stars
    - A Symphony of Lights
    - Clock Tower
    - Kowloon Walled City Park
    - Sik Sik Yuen Wong Tai Sin Temple
    - Lei Yue Mun Seafood Bazaar

Taxonomy

TaxoFolk algorithm

Attractions
- Most Popular
  - Hong Kong Island
    - The Peak — Sightseeing, Tram
    - Madame Tussauds — Museum, Waxworks
    - Hollywood Road, Cat Street / Man Mo Temple
    - Western Market — Shopping
    - Lan Kwai Fong & SoHo — Nightlight
    - Happy Valley Racecourse
    - Jumbo Kingdom
    - Ocean Park — Panda
    - Repulse Bay — Beach
    - Stanley Market & Murray House — Beach
    - Aberdeen — Sightseeing
  - Kowloon
    - Ladies Market — Shopping
    - Temple Street Night Market — Opera
    - Jade Market & Jade Street
    - Avenue of Stars
    - A Symphony of Lights
    - Clock Tower
    - Kowloon Walled City Park — City
    - Sik Sik Yuen Wong Tai Sin Temple
    - Lei Yue Mun Seafood Bazaar
  - New Territories
    - Sha Tin Racecourse — Horse
    - Ching Chung Koon
    - Ping Shan Heritage Trail

Temple Street got an OPERA

Can visit a BEACH at here

HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

uture • 啟迪思維 • 成就未來

Benefits of the hybrid navigation approach

- Enhanced and simplified way to navigate an existing repository based on popular user tags
- Added tags reveal users' indexing and information access preferences
- The hybrid structure delivers an user-oriented model for corporate knowledge navigation and discovery
- The hybrid structure provides valuable input to the taxonomy maintenance/revamp process
- Collected tags help to build/review meta-data and controlled vocabularies
- Tools developed to automate most of the steps in generating the hybrid structure