You will be able to attend in-depth table discussions on three of these cases on Day 2. The case outlines are provided to help you decide which case discussions you would like to attend.

9) ERIC TSUI

Lessons learnt from deploying a hybrid folksonomy-taxonomy approach to enhance navigation in a tourism industry KM system in Hong Kong

1. About the Case Organization
Showcasing a prototype of the TaxoFolk (Taxonomy-Folksonomy) system for enhanced knowledge navigation to the KM team of a large tourism company that already has strong KM governance and rooted in a culture of top down deployment of enterprise KM systems.

2. About the Challenge
- The company has already established taxonomies for classifying assets; however the taxonomy was set by a small group of staff and has not been updated regularly
- Taxonomy is getting out of date and users are increasingly frustrated as they waste time in locating knowledge assets
- Try to enlighten the team that taxonomy and search engines are not the only means to navigate knowledge assets; that user-labelled information can also be incorporated into the enterprise KM system
- Try to explain to the KM team that analyzing user labels/tags also help to better understand why certain words are used by users in qualifying information and this knowledge is valuable at taxonomy revamp stage as well as for search engine configuration. Additional vocabularies are often discovered from analyzing user tags/labels
- The KM team has approximately 12-15 members drawn from Corporate, HR, IT and other departments

3. What We Did
- Original research on combining taxonomy and folksonomy to enhance knowledge navigation in terms of a hybrid/extended site map
- Technologies involve user coding of clustering algorithms, (the TaxoFolk algorithm) codified IF-THEN rules to identify word stem and variants, checking of WORD-Net online dictionary and Wikipedia to verify the authenticity of words/terms and show relationships between them
- I explained the TaxoFolk algorithm, the rationale behind it, I showed the prototype and gave demonstrations

4. Challenges and Lessons Learned
- “Taxonomy” is very new to most Hong Kongers let alone the term Folksonomy. My biggest problem in promoting TaxoFolk is to survive the complexity of explaining several seemingly new and difficult-to-understand terms before the audience can ascertain what is the power of TaxoFolk
- Pick and approach companies/users that/who are familiar with information classification; companies that welcome bottom up contributions and especially those that are already using some kind of Web 2.0 tools
- Just deploy the prototype and let users try it for a pilot period. Let them experience the superiority in knowledge navigation and only then further explain how the hybrid navigation structure is put together
- Additional vocabularies revealed from the collected user tags/labels serve as an excellent source to further ascertain the gaps between use of naming and sometimes even the perceptions of corporate and perceptions of individual users, also useful for updating and maintaining the corporate taxonomy
- The system was not fully implemented, because the organization did not have the skills to re-develop the algorithms for TaxoFolk. However, when they finally upgraded their portal (SharePoint), they chose to adopt user tagging in the platform as a high priority.
5. Impact and Benefits

- Though the system was not implemented, the KM team finally realises the power and impact (offering an additional dimension to locate asset) of a tool like TaxoFolk.
- Common tags used by users turn out to be a valuable source of information to consider during taxonomy revamp especially on the setting of terms and phrases; further discussion with users on how certain tags/labels are used reveals the mental models of many users.
- At the time when I showed TaxoFolk to the organization, they were not piloting other projects hence I would say much of the change and understanding are attributed to TaxoFolk. The KM team appreciates that during taxonomy revamp, more consideration needs to be paid to a user’s perspective on what they perceive and their use of common terms (which may well be different from terms that the organization has adopted).
- Delivering explanations at various levels of abstraction, patience, and provide a prototype or demo system for users to try.

6. Next Steps

Tag clouds can also be generated from user contributed tags and are provided to users as an additional mean to navigate over knowledge assets. These tags have improved the findability of explicit assets in the organization. From the research perspective, we intend to apply for funds to support personalization of the TaxoFolk hybrid navigation system for individual users; at present it is one system for all users based on the contributed tags.

10) HALIZA JAILANI

Using knowledgebases using named entity recognition, query expansion and Linked Data, to support enhanced discovery

1. About the Case Organization

The National Library Board Singapore manages the National Library, 26 Public Libraries and the National Archives, with a mission to make knowledge come alive, spark imagination and create possibilities. NLB supports knowledge seeking activities by providing a trusted, accessible and globally-connected library and information service.

2. About the Challenge

With the merger of National Archives Singapore (NAS) and the National Library Board in November 2012, it became critical to harmonise library and archives metadata, name headings & controlled vocabularies so the public can discover resources from both agencies without having to go through two portals. Metadata and terminologies need to be aligned. Knowledge organisation processes need to be centralised for efficiency and cost savings, and common tools shared. Not many libraries and archives in the world have merged their collections, difficult to do as archives and libraries organise materials differently. The multi-level description of archives relates objects in a hierarchy and links the parts to a larger ensemble from the collection level perspective. Libraries organise at the item-level.

Names are largely not standardised in NAS, where 8 different databases are managed individually by different teams. As a result a single person may have more than one form of name. Search results for this person are not unified and resources are retrieved according to the name a user enters. Merging of NAS collection as it is with the NLB collection will cause search results to be more fragmented. A search for an entity whether a person, an organisation or a place will not pull content resources about the entity into a single list. Like other NLs, NLB uses names authorised by Library of Congress Name Authority Cooperative Program (NACO) which observe strict rules for capturing every part of a name. Where names cannot be established in NACO, NLB uses a separate list from a local file.

In addition, the library world is undergoing a shift in thinking on how libraries catalogue resources and enable them to be discovered. The long-time library cataloguing standard AACR2 (Anglo-American Cataloguing Rules 2nd Edition) was replaced with a less rigid standard that is more attuned to the needs of a digital world, called RDA or Resource, Description and Access. Increasing efforts are also being made to break the coded information out of the catalogue record so that they can be used as Linked Data and library materials can become searchable on the Internet.

3. What We Did

NLB used a three-pronged approach to organise the merged collections and improve resource discovery: (i) Data harmonisation & query expansion (ii) Linked Data implementation, (iii) Knowledge-base building.
**Definitions**

“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](http://en.wikipedia.org/wiki/Taxonomy)

Tag Cloud: Most popular tags in Flickr

All time most popular tags

africa amsterdam animals april architecture art asia australia baby barcelona beach berlin birthday black blackandwhite blue boston bw california cameraphone camping canada canon car cat cats chicago china christmas church city clouds color concert day dc de dog england europe family festival film florida flower flowers food france friends fun garden geotagged germany girl graffiti green halloween hawaii hiking holiday home honeymoon hongkong house india ireland island italy japan july june kids la lake landscape light live london losangeles macro march may me mexico mountain mountains museum music nature new newyork newyorkcity newzealand night nikon nyc ocean paris park party people portrait red river roadtrip rock rome san sanfrancisco scotland sea seattle show sky snow spain spring street summer sun sunset sydney taiwan texas thailand tokyo toronto tour travel tree trees trip uk urban usa vacation vancouver washington water wedding white winter yellow york zoo
WIKIPEDIA'S CATEGORY STRUCTURE:
Visualization of the quad-hierarchized network of category pages in Wikipedia (UDC color coded), Gephi software: Fruchterman-Reingold layout. Only nodes are displayed, there are too many edges.

CATEGORY DISTRIBUTION UDC & WIKIPEDIA:
Distribution of UDC entries (categories) into nine main classes (inner ring) and distribution of Wikipedia category pages into 43 top categories (outer ring). Wikipedia categories are further assigned to corresponding UDC classes, and colored accordingly.

UNIVERSAL DEIMAL CLASSIFICATION:
Visualization of the UDC tree network, Gephi software: Fruchterman-Reingold layout. Edges are visible.

OCCURRENCES OF 43 TOP WIKIPEDIA CATEGORIES:
Visualization of Wikipedia's top category terms' occurrences in the Master Reference File of UDC in a tree-map layout (Magnaview software). The length of a column corresponds to how many times a category term is found in a UDC main class. Long columns indicate strong connections between a Wikipedia category and a UDC class. Each column is marked with its UDC main class number to show the connection.
## Taxonomy + Folksonomy

### Taxonomy
- **Central Control**
- **Top-down**
- **Meaning to the author**
- **Tedious process for making changes**
- **Accurate**
- **Navigation**
- **Restrictive**
- **Defined Vocabulary**

### Folksonomy
- **Democratic Creation**
- **Bottom-up**
- **Meaning to the reader**
- **Just do it**
- **Good enough**
- **Discovery**
- **Expansive**
- **Personal Vocabulary**

### Taxonomy Hybrid Models
- **Central control with continuous user input**
- **Meet in the middle**
- **Cater for the growing community**
- **Suggestions, additions, deletions with governance models**
- **Community validated and tested**
- **From navigation to discovery by leveraging mass input**
- **Flexible and evolving**
- **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
Collaborative tagging tools

TaxoFolk
A Hybrid Taxonomy - Folksonomy
For Enhanced Knowledge Navigation

TaxoFolk algorithm

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

New Territories
- Sha Tin Racecourse
- Ching Chung Koon
- Ping Shan Heritage Trail

Can visit a BEACH at here

Temple Street got an OPERA