TaxoFolk: A hybrid taxonomy - folksonomy classification for enhanced knowledge navigation

Ching-Chieh Kiu and Eric Tsui

Knowledge Management Research Centre,
Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University,
Hung Hum, Kowloon, Hong Kong

Correspondence: Ching-Chieh Kiu c/o Professor Eric Tsui
Knowledge Management Research Centre, Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong
Email: kiuchingchieh@gmail.com

Abstract
Taxonomy is widely used in many of the website and directory navigation schemes for content/knowledge retrieval. However, information or content navigation support through taxonomy is often constrained due to its inability to take into account the full nomenclature and cultural nuances of knowledge seekers. The emergence and increasing adoption of collaborative tagging (social bookmarking) tools have provided lightweight and informal conceptual structures called folksonomies for knowledge retrieval. As for folksonomies, they reflect the vocabulary of the users. Hence, integrating folksonomies into a taxonomy combines the best of the two schemes as the resultant structure enhances taxonomy navigation with personsalisation for knowledge search and retrieval. This paper presents TaxoFolk, an algorithm for deriving hybrid taxonomy-folksonomy classification for enhanced knowledge navigation. The algorithm integrates folksonomy with a taxonomy through several unsupervised data mining techniques with augmented heuristics.

Keywords: Knowledge Navigation, Folksonomy, Taxonomy, Taxonomy-Folksonomy Integration, Collaborative Tagging

1 Introduction

Web resources and contents are usually indexed by a taxonomy to support knowledge navigation. Such a taxonomy is, typically, designed or created by owning authority, by a specialist (e.g. a taxonomist), or derived from the authors of the web resources (Spiteri, 2007). However, information or content navigation support through taxonomy is often constrained due to its inability to take into account the full nomenclature and cultural nuances of the knowledge seekers.

Several prominent collaborative tagging tools have emerged from the Web 2.0 movement. Most noticeably, these tools include Flickr (http://www.flickr.com/), bibsonomy (http://www.bibsonomy.org/), Furl (http://www.furl.net/), Ma.gnolia (http://ma.gnolia.com/), del.icio.us (http://del.icio.us/) and Digg (http://digg.com/). They provide users with the freedom to annotate their own (and those that are generated by others) resources with a set of tags (keywords) without relying on a top-down (highly regulated) controlled vocabulary (Specia & Motta, 2007). Such tagging tools allow users to store, organize, search, and manage resources through a lightweight conceptual structure, namely a “folksonomy”.

In social networking web sites, which typically are populated by user-generated content (UGC), *folksonomies* are often used and function as an alternative to formal taxonomies. While a user-tagged folksonomy provides much of the needed flexibility and freedom for users to label the stored information, a compromise on information accuracy and quality can also occur when disparate views and terms are being shared by many users operating in various contexts. Combining both folksonomy and taxonomy may help alleviate these problems (Barbosa, 2008; Owens, 2008; Ward, 2008; Hayman & Lothian, 2007). Such a hybrid model is also able to enhance the findability (AIIM, 2008) of content and produce a faster and less complicated method of content retrieval.

2 Problem Statements

Although taxonomies have been widely adopted for classifying and categorizing web resources, knowledge seekers continue to suffer from poor knowledge navigation and ineffective retrieval. A finding from Trant (2006) has revealed that more than 70% of the terms tagged by the users were not listed in a museum's documentation. This means that content classified by experts often have conceptual definitions that are different from those by other users. This is a common problem in formal taxonomies which, because of these differences, are unable to take into account the full nomenclature and cultural nuances of the users. The reason is that taxonomies are usually designed by a small group of experts who often have a different viewpoint and navigational behavior from the mass (general) users.

In addition, building and maintaining taxonomy is often an expensive and tedious task (Fichter, 2006; Kroski, 2006). Besides, taxonomies become outdated very quickly, whereas new concepts may emerge but they are not yet included in the taxonomy; in contrast, folksonomies accommodate easily such new concepts (Mitchell, 2005). Thus information or content navigation, searching and discovery support through a taxonomy can in fact be constrained and fail to facilitate collaborative opportunities (Barbosa, 2008).

There are some advantages and limitations of controlled vocabularies (taxonomies) and uncontrolled vocabularies (folksonomies). By blending both types of vacabularies, newly sprung characteristics of a hybrid taxonomy-folksonomy are depicted in Table 1. The benefits of a hybrid taxonomy-folksonomy model have been actively discussed (DowJones, 2008; Ward, 2008). They include:

1) Enhanced findability of content;
2) Improved knowledge searching and retrieval;
3) Enhanced taxonomy management process;
4) Existence of new navigational facets to better connect and display ; and
5) Classification of contents/web resources with minimal costs

To redress the above limitation and capitalizing on the benefits of taxonomies and folksonomies, this paper outlines an algorithm that integrates pre-processed folksonomy tags with a static taxonomy through unsupervised data mining techniques. These techniques include Formal Concept Analysis (FCA), K-Means, and Simple Matching Coefficients (SMC) algorithms.

The rest of the paper is outlined as follows; In Section 3 presents brief description of the current related works, meanwhile in Section 4 the algorithm for folksonomy-taxonomy integration is explained. In Section 5, the result from the experiment conducted on a trial dataset is discussed. Section 6 is the conclusion and future work.

Table 1 Characteristics of a hybrid taxonomy-folkosonomy (DowJones, 2008; Sampson, 2008)

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

3 Related Works

Laniado et. al. (2007) used WordNet to turn a folksonomy into a hierarchy of concepts and discovered the possibility of integrating an ontology by adding some explicit semantics, provided by a static hierarchy of concepts in the navigation interface of a folksonomy, in order to enrich the possibilities of navigation in a folksonomy, to help users orient themselves among keywords.

Ohkura et. al (2006) proposed an automated folksonomy system to determine whether a particular tag should be attached to an item and also to create a candidate tag set, which is a list of tags that may be attached to items, from weblog category names. This system is an automated multi-tagging system for weblog articles. For each weblog article, the system attaches multiple tags. Tag names and their concepts are automatically extracted from collected weblog articles. The system consists of the following three parts: a weblog articles crawler, a multi-tagger, and a user interface.

Hayman & Lothian (2007) developed a user portal called “myEdna” with a taxonomy directed folksonomy. Portal users are allowed to mark-up resources with keywords (tags) which prompted by a thesaurus. Users therefore can tag content with the most appropriate labels, but are limited to a list of terms which are controlled and directed by drop-down menus that allow the user to choose the most appropriate keywords.

Stock (2007b) mashes-up the benefits of the “old” science databases (professional indexing, citation indexing, full-text processing) and the benefits of folksonomies (authentic language use of the readers, multiple interpretations,

+ =

and new ranking options) to improve Science databases. Meanwhile, Trant (2006) employed social tagging and folksonomy to improve access to art information.

Although there are efforts on adoption of folksonomy to support and enhance contents browsing and indexing, and there has been ample research into taxonomy design, creation and maintenance, however, up to now, very few scientific publications have been found on work related to the co-leveraging of folksonomy and taxonomy for enhanced knowledge navigation. A recent workshop on the topic of taxonomy-folksonomy integration organised by the authors at a recent major international Knowledge Management conference provides yet further reinforcement of the above observation.

4 The TaxoFolk Algorithm

The algorithm for integrating folksonomy and taxonomy is depicted in Figure 1. The algorithmic algorithm comprises four major phases, which are (1) tag pre-processing phase, (2) domain contextualization phase, (3) contextual clustering phase, and (4) concept-tag consolidation phase. The detailed process of each phase is explained in the next sub-section and a general algorithm for taxonomy-folksonomy integration is shown in Table 2.

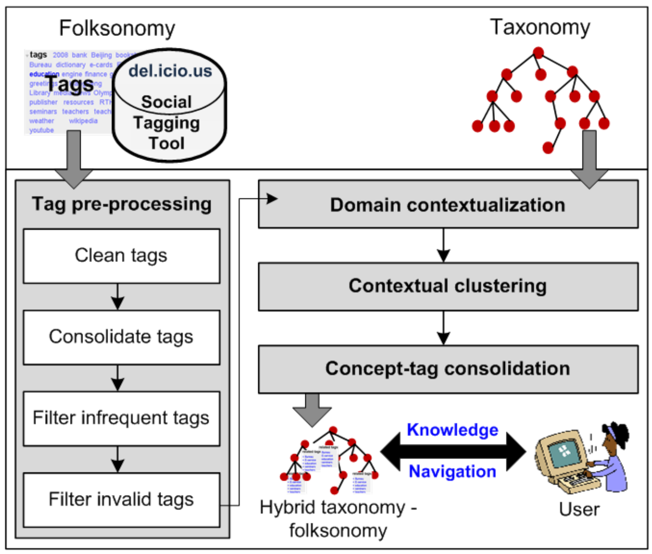

Figure 1 The taxonomy – folksonomy integration process

4.1 Phase1: Tag pre-processing phase

As a folksonomy is uncontrolled, tags are redundant as they can be freely and easily created. However, there are also redundancies, (instances of) incompleteness and inconsistencies. Terms can be mis-spelt and in words may appear in different forms, e.g. plural, singulars, various word tenses and pronouns. Therefore, the tag pre-processing process is primarily performed to reduce the noise in the tags obtained from social bookmarking. The following steps are performed in the listed sequence:

Table 2 A general algorithm for taxonomy – folksonomy integration

Input	A taxonomy and folksonomy tags (that describe the chosen taxonomy)
Step 1	*Tag pre-processing* Pre-processing of tags to identify candidate tags (phase 1)
Step 2	*Domain contextualization* Reasoning the hierarchical relationships in the taxonomy and the relationships between and among the folksonomy tags that are used to describe resources (phase 2)
Step 3	*Contextual clustering* Grouping candidate tags for the taxonomic concepts (phase 3)
Step 4	*Concept-tag consolidation* Integrating the candidate tags into the taxonomy (phase 4)
Output	A hybrid taxonomy-folksonomy (TaxoFolk)

1) *Clean tags:* A valid tag may consist of letters, numbers and symbols like a dash, a hyphen, dots and quotation marks. A tag that consists of other symbols is considered as an unusual tag; it will be filtered out.

2) *Consolidate tags:* Tags can be in the plural, singular, an abbreviation, an acronym and in various word tenses. Such tags will be consolidated into a root word, e.g. "*travels* and *travelled* to *travel*", and "*hong_kong*, *hong-kong*, *hongkong* and *hk* to *hong kong*" using natural language processing, morphology. Then, Levenshtein's similarity metric[1] is used to measure the similarity between the tags of a resource. Similar tags will be grouped together if the similarity value is equal to or above a certain threshold value. WordNet[2] is incorporated into this process for resolving the misspelling of tags.

3) *Filter infrequent tags:* Usually, a resource is tagged with many tags by many taggers. A user assigns a tag to a resource using his/her own definition of the resource. Therefore, a tag that is chosen to markup a resource by a user might not reflect the prevailing context or the norms of other users. Such tags are found infrequently hence need to be filtered out. Infrequent tags are filtered out according to the frequency of tag occurrence in a resource, F_t as denoted in below equation with a frequency threshold value.

$$F_t = \frac{\textit{Total number of the tag t used to markup the resource}}{\textit{Total number of tag used to markup the resource}} = \frac{t}{n}$$

4) *Filter invalid tags:* Some tags, despite their high frequency of occurrences, might be invalid. In WordNet, relatively new and very specific words still do not exist, for example *folksonomy*, *iBeam* or *3G*. Thus, the usage of the WordNet in the algorithm is restricted to being a spelling checker but inappropriate for checking invalid tags. In this process, Wikipedia[3] is used to filter such tags. As can be seen, these words (*folksonomy*, *iBeam* and *3G*) have been in the Wikipedia since 2004. Furthermore, Wikipedia concept definitions denoted by a uniform resource identifier (URI) does not change in most cases (Hepp et. al., 2008) and thus Wikipedia can be regarded as a viable lexical resource for filtering invalid tags.

4.2 Phase 2: Domain contextualization phase

In general, a taxonomy consists of concepts that are associated with a set of resources (e.g. documents, webpages, etc.) (Wikipedia, 2008). The hierarchical nature of the concepts is defined by the *subconcept-superconcept* relationship, also called *parent-child* relationship or *is-a* relationship. Hence, a taxonomy is formalized as a tuple $T := (C, S_C, R_T)$, where C represents the concepts of the taxonomy, S_C corresponds to the hierarchy of concepts and R_T is the resources that are associated by concepts.

A folksonomy describes the users, tags and resources, and allows user to designate tags to resources (Hotho et. al, 2006). In the algorithm, folksonomy is formalized as a tuple $F := (W, R_F)$, where W is tags and R_F is resources that are associated by tags.

Formal concept analysis (FCA) is applied at this phase to contextualize, through a reasoning process, the hierarchical relationships in the taxonomy and the relationships between and among the folksonomy tags that are used to describe resources. The Formal Concept Analysis (FCA) is an unsupervised data mining method mainly used for data analysis, where it provides a conceptual basis for structuring associations among concepts and for modeling concepts and corresponding attributes (Ganter & Wille, 1999). In FCA, a *formal context* is derived from the given data (taxonomy and folksonomy associated with their resources) as shown in the Figure 2.

A *formal context* is a triple $k = (O, A, R)$ where O are objects, A are attributes and R is a binary relation between O and A, where $R \subseteq O \times A$, $(o, a) \in R$ is read as "*the object o has the attribute a*". For a set of objects $X \subseteq O$ and $Y \subseteq A$, we define $X' := \{a \in A \mid (o, a) \in R \text{ for } \forall o \in X\}$ and $Y' := \{o \in O \mid (o, a) \in R \text{ for } \forall a \in Y\}$ respectively. A more detailed explanation of the FCA can be found in Ganter & Wille (1999).

Given a taxonomy $T := (C, S_C, R_T)$, it maps to a formal context, K_T where $(C , S_C) \subseteq O$, $R_T \subseteq A$ and R denotes the binary relation R_T corresponding to the C and S_C, and given a folksonomy $F := (W, R_F)$, it maps to a formal context, K_F where $W \subseteq O$, $R_F \subseteq A$ and R denotes the binary relation R_F corresponding to W.

[1] http://sourceforge.net/projects/simmetrics/
[2] http://sourceforge.net/projects/jwordnet/
[3] http://www.ukp.tu-darmstadt.de/software/jwpl/

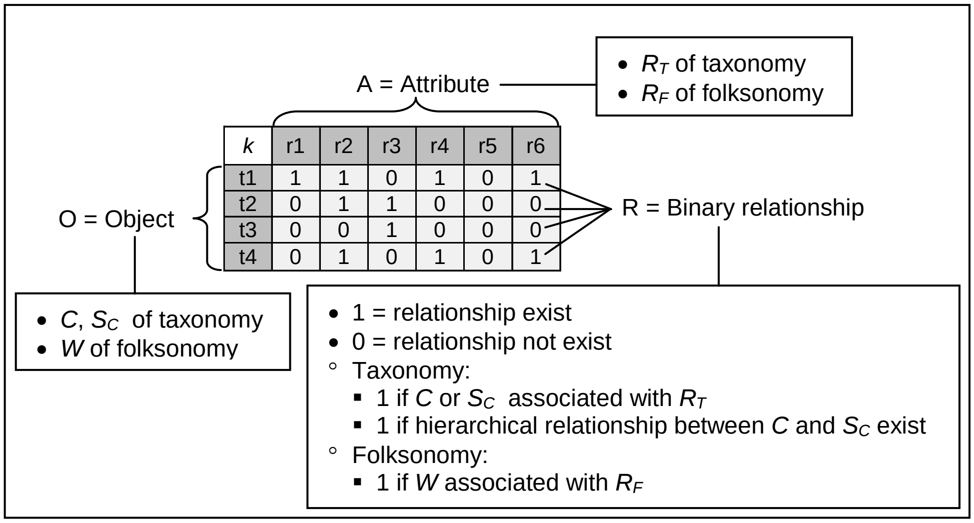

Figure 2 Formal context

4.3 Phase 3: Contextual clustering phase

In this phase, K-means is used to accelerate the taxonomy-folksonomy integration process. K-means clustering (Macqueen, 1967) is a simple unsupervised clustering algorithm that iteratively divides a data set into a number of clusters. For a more detailed explanation on the K-means please refer to (Macqueen, 1967).

In the contextual clustering phase, a formal context of a taxonomy is firstly grouped into a number of clusters to form an initial prediction model. Consequently, a formal context of a folksonomy is assigned to the "good" clusters defined in the model.

4.4 Phase 4: Concept - tag consolidation phase

A Simple Matching Coefficient (SMC) (Dunn & Everitt, 2004) incorporating a similarity threshold value, is used to measure the semantic distinction between concept and tag in each cluster. If the semantic distinction value is equal to or greater than a similarity threshold value, then the tag is regarded as related to the concept and it will be integrated with a taxonomy either as a *navigation* tag or a *label* tag (Figure 3) by applying the consolidated rules presented in Table 3.

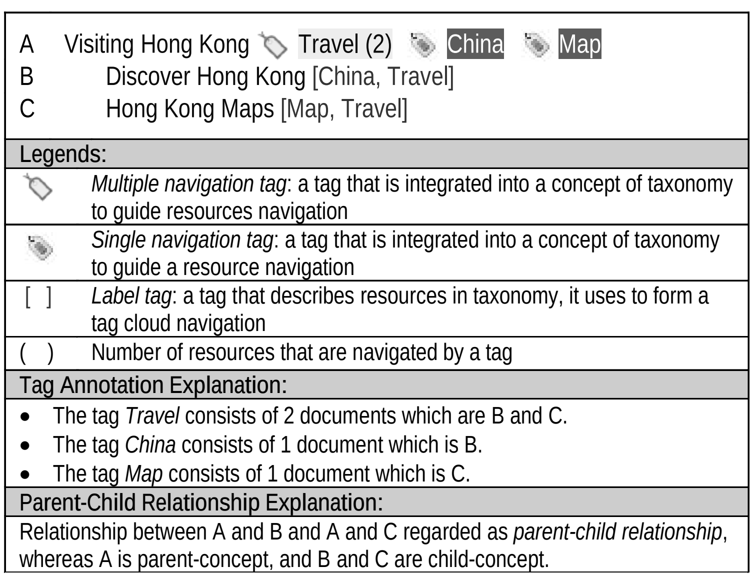

Figure 3 Sample of taxonomy and folksonomy integration

Table 3 Consolidated Rules for taxonomy – folksonomy integration

Given concept is denoted as C_O*, child concept is denoted as* C_C *and parent concept is denoted as* C_P*.*

Individual relationship
Given C_O is tagged by a set of tag, {*t*}
Rule 1: If SMC discovers $C_O \rightarrow \{t\}$,
then {*t*} is regarded as a ***label tag*** of C_O.

Parent-child relationship
Given C_C is tagged by a set of tag, {*t*}
Rule 2: If SMC discovers $C_P \rightarrow \{t\}$ AND $C_C \rightarrow \{t\}$,
then {*t*} is regarded as a ***label tag*** of C_C and as a ***navigation tag*** of C_P.
Rule 3: If SMC discovers $C_P \rightarrow \{t\}$,
then {*t*} is regarded as a ***navigation tag*** of C_P and C_C inherits the {*t*} as a ***label tag***.
Rule 4: If SMC discovers $C_C \rightarrow \{t\}$,
then {*t*} is regarded as a ***label tag*** of C_C.

5 Results and Analysis

The objective of this experiment is to evaluate and justify the techniques adopted in the algorithm to support a folksonomy's integration with a pre-defined taxonomy. In this experiment, the GovHK portal's (http://www.gov.hk/) (residents and non-residents) is the chosen taxonomy and its folksonomy is obtained from the del.icio.us (http://delicious.com/) database dated 21-06-2008.

The taxonomy consists of 6 levels with linkages to 752 websites, out of which 100 websites were tagged in del.icio.us. To be included in this experiment, a minimum of 5 users tagging the website is needed. In the tag pre-processing phase, a threshold value 0.8 (Kiu & Lee, 2006) is used to discover the Levenshtein similarity of tags and a frequency threshold 0.1 is used to filter out the infrequent tags. The adopted thresholds serve to determine the candidate tags that are used to mark the resources. These statistics of the dataset and the candidate tags are shown in Table 4.

In this preliminary experiment, 3 to 6 clusters have been explored for grouping related candidate tags corresponding to concepts in the taxonomy, however it a 5 clusters configuration was eventually decided based on the *within cluster sum of squared errors values*. Applying the SMC with a similarity threshold of 0.9, three tags namely *hong kong*, *community* and *government* are discarded. These tags were used to mark-up the resources; they are not considered to be significant for enhanced knowledge navigation as they merely describe the nature of the resources.

Table 4 shows the candidate tags associated with the taxonomy concepts after the removal of these tags. After applying the consolidated rules, the tags were integrated with the taxonomy. The resulted taxonomy-folksonomy is depicted in Figure 4.

This preliminary experiment has demonstrated that the algorithm is both viable and feasible for integrating the folksonomy with the taxonomy using unsupervised data mining techniques. In this experiment, different frequency threshold values to filter out infrequent tags and similarity threshold values to integrate tags into the taxonomy have been tried with the algorithm in order to validate and to identify appropriate threshold values. Further justifications on the validity of the discovered threshold values used in this experiment and more will be subject to future research, possibly involving different taxonomies and sets of tags. The screenshot of the TaxoFolk navigation is shown in Figure 5.

As illustrated from Figure 4, the webpage entitled *E-Learning from RHTK* provides information on both e-learning and language learning which included resources on learning Japanese language has label tags, "*Japanese*", "*Language*" and "*Learning*". These tags appeared as navigational tags under the webpage *Interests & Hobbies*. Hence, for example, with "*Japanese*" surfaced as a navigational and label tag, user can instantly visualize and navigate to those very pages/folders that otherwise would not be easily noticed via the original (i.e. taxonomy only) structure.

Thus, such a hybrid taxonomy-folksonomy classification not only provides benefits but also enhances knowledge navigation and knowledge classification, More specifically, the hybrid classification provides

1) an enhanced and simplified way to navigate an existing repository based on commonly used user tags;
2) clues on users' indexing and information access preferences based on the inserted tags;
3) an user-oriented model for corporate knowledge navigation and discovery
4) valuable ongoing and timely input (on the use of user-preferred terms) for the taxonomy maintenance/revamp process
5) input to create/review meta-data and controlled vocabularies, two crucial elements of any formal taxonomy.

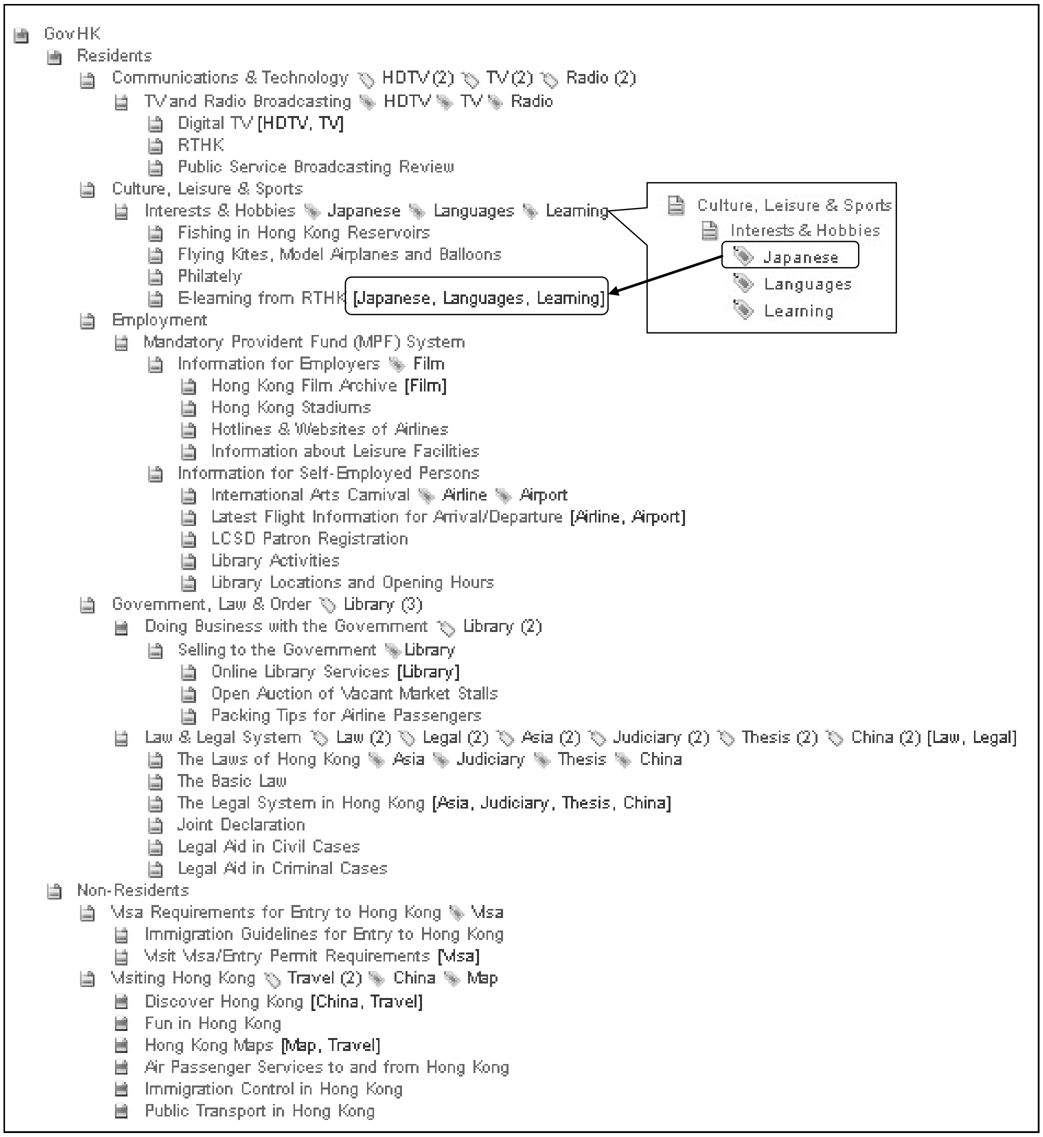

Figure 4 The resulting taxonomy with tags (please refer to Figure 3 for an explanation of tag annotation)

6 Conclusion and Future Work

In this paper, a novel approach for integrating folksonomy with a taxonomy using unsupervised data mining techniques, where prior knowledge is not required for the task is presented. The approach was implemented using the HKGov taxonomy with its folksonomy from the del.icio.us database. The result of the experiment has demonstrated that the techniques applied in the algorithm are promising and that it is feasible to use it to integrate folksonomy with taxonomy.

As for future work, the authors intend to experiment with other hybrid clustering techniques to combine the taxonomy and folksonomy and also to automate the folksonomy-taxonomy integration process. Additional experiments will be conducted to evaluate the TaxoFolk algorithm and the threshold values used in the algorithm to facilitate the automation of the taxonomy and folksonomy integration. Furthermore, additional trials have been planned and validations will be carried out in an industrial environment. In the longer term, the authors also aim to study the influence and impact brought about by TaxoFolk on user's navigation behaviour as well as the choice of terms in the corporate taxonomy revamp process.

In addition, the TaxoFolk framework can be extended to dynamic customize a personal hybrid taxonomy-folkosnomy based on user profile and other elements to support personal knowledge management (e.g., personalized digital library, learning repository etc.). Furthermore, the TaxoFolk also can be extended to support and build user-oriented travel web sites, consumer-centric online catalogs and user-centric file directories.

Acknowledgement

This research is funded by The Hong Kong Polytechnic University under Account code 1-45-37-0542. Its support is gratefully acknowledged.

Table 4 Number of tags associated the websites

ID	Websites	Tags from del.icio.us		Tags after tag pre-processing		Tags after applied clustering and SMC		Tags after applied consolidation rules	
		No. of taggers	No. of tags	Candidate tags	No. of tags	Candidate tags	No. of tags	Candidate tags	No. of tags
1	Discover Hong Kong	95	202	hong kong, travel, china	3	china	1	china, travel	2
2	GovHK - one-stop portal of the Hong Kong SAR Government	37	98	hong kong, government	2	-	0	-	0
3	GovHK: Residents	35	80	hong kong, government	2	-	0	-	0
4	E-learning from RTHK	34	68	learning, language, japanese	3	japanese, language, learning	3	japanese, language, learning	3
5	Visit Visa/Entry Permit Requirements	30	57	hong kong, visa	2	visa	1	visa	1
6	Online Library Services	27	52	hong kong, library	2	library	1	library	1
7	Hong Kong Maps	19	54	hong kong, map, travel	3	map	1	map, travel	2
8	The Laws of Hong Kong	20	38	law, legal	2	law, legal	2	law, legal	2
9	RTHK	16	39	hong kong, radio	2	radio	1	radio	1
10	The brand of Hong Kong	14	44	hong kong	1	-	0	-	0
11	Legislative Council	13	21	hong kong, government	2	-	0	-	0
12	Latest Flight Information for Arrival/Departure	13	29	hong kong, travel, airline, airport	4	airline, airport	2	airline, airport	2
13	Hong Kong Film Archive	11	38	hong kong, film	2	film	1	film	1
14	Hong Kong Wetland Park	9	24	hong kong	1	-	0	-	0
15	The Government of the Hong Kong Special Administrative Region - Immigration Department	6	12	hong kong	1	-	0	-	0
16	Digital TV	5	15	hong kong, tv, hdtv	3	hdtv, tv	2	hdtv, tv	2
17	The Legal System in Hong Kong	5	9	hong kong, communities, asia, legal, law, thesis, judiciary	7	asia, china, judiciary, thesis	4	asia, china, judiciary, thesis	4
18	Fun in Hong Kong's 18 Districts	5	8	hong kong	1	-	0	-	0
19	A Symphony of Lights	5	13	hong kong	1	-	0	-	0

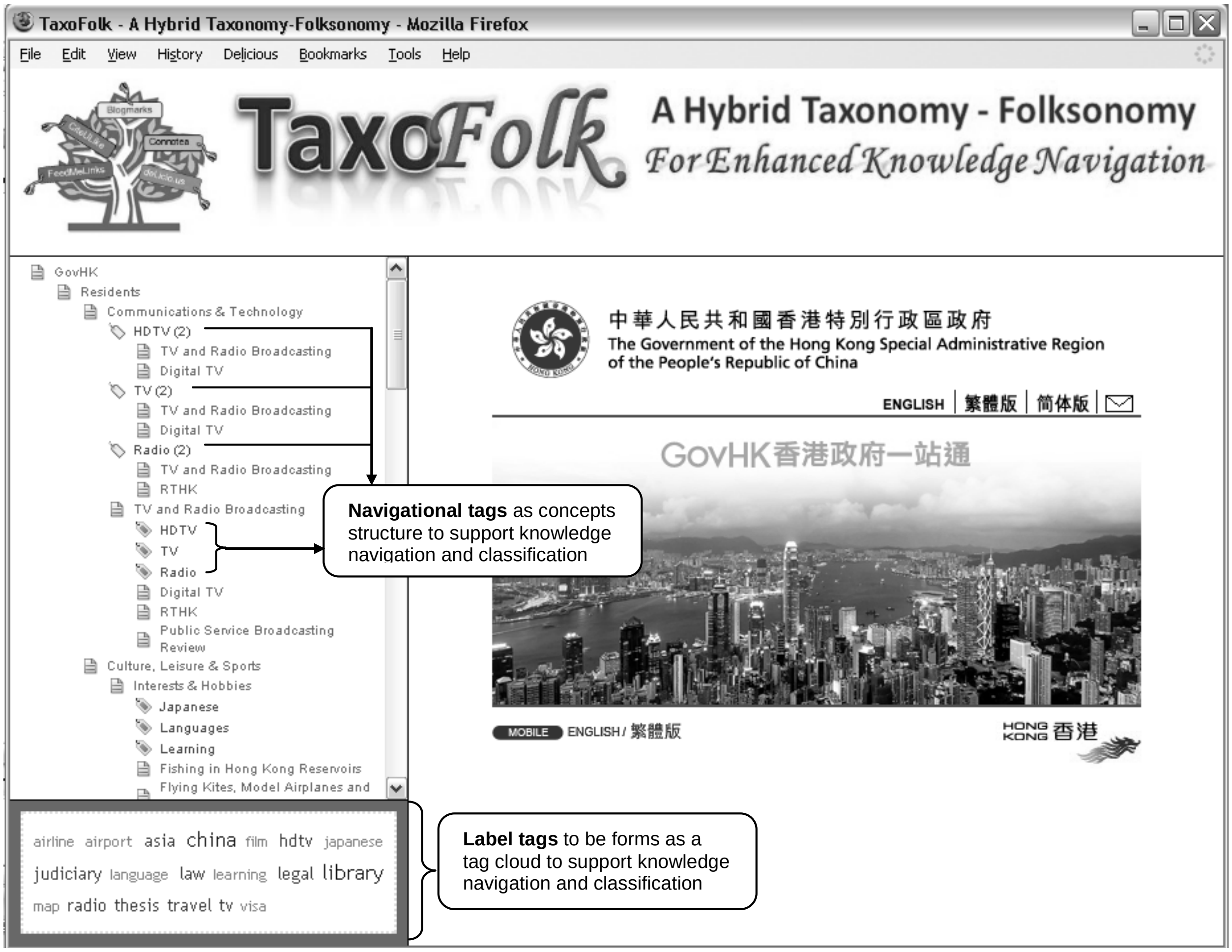

Figure 5 Screenshot of the TaxoFolk Navigation Interface

References

AIIM (2208) Findability: The Art and Science of Making Content Easy to Find. *Market IQ Intelligence Quarterly*.

BARBOSA D (2008) The Taxonomy Folksonomy Cookbook - Finding the Right Recipe for Organizing Enterprise Metadata. *Available: http://www.solutions. dowjones.com/cookbo ok/*

DOWJONES (2008) Folksonomies and Taxonomies in the Enterprise: Part Two. *Available: http://factiva.com/infopro/articles/Jun2008Feature.asp?node=menuElem1103*

DUNN G and EVERITT BS (2004) An Introduction to Mathematical Taxonomy. *Dover Publications*.

FICHTER D (2006) Intranet applications for tagging and folksonomies. *Online* **30(3)**, 43-45.

GANTER B and WILLE R (1999) Formal Concept Analysis – Mathematical Foundations. *Springer Verlag*.

HAYMAN S and LOTHIAN N (2007) Taxonomy Directed Folksonomies. *World Library and Information Congress: 73rd Ifla General Conference and Council*, Durban, South Africa.

HAYMAN S and LOTHIAN N (2007) Taxonomy Directed Folksonomies: Integrating user tagging and controlled vocabularies for Australian education networks. *World Library And Information Congress: 73rd Ifla General Conference And Council*, 19-23, Durban, South Africa

HEPP M, BACHLECHNER D and SIORPAES K (2006) Harvesting Wiki Consensus - Using Wikipedia Entries as Ontology Elements. *Proceedings of the Workshop on Semantic Wikis at the ESWC2006 (ESWC2006)*, Budva, Montenegro.

HOTHO A, J¨ASCHKE R, SCHMITZ C and STUMME G (2006) Folkrank: A ranking algorithm for folksonomies. In K.-D. Althoff and M. Schaaf, editors, *Hildesheimer Informatik-Berichte,* 1/2006, 111–114.

KIU CC Kiu and LEE CS (2006) Ontology mapping and merging through OntoDNA for learning object reusability. *Educational Technology & Society*, **9(3)**, 27 – 42.

Kroski E (2006) The hive mind: Folksonomies and user-based tagging. *Available: http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-us er-based-tagging/*

Laniado D, Eynard D and Colombetti M (2007) Using wordnet to turn a folksonomy into a hierarchy of concepts. *In Semantic Web Application and Perspectives - Fourth Italian Semantic Web Workshop*. Bari, Italy. 192-201.

MacQueen JB (1967) Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1:281-297.

Mitchell RL (2005) Tag teams wrestle with web content *Computerworld* **38(16)**, 31.

Noruzi A (2006) Folksonomies: Uncontrolled Vocabulary. *Knowledge Organization*, **33(4)**.

Noruzi A (2007) Editorial. *Webology*, **4(2)**, editorial 12. *Available at: http://www.webology.ir/2007/v4n2/editorial12.html*

Ohkura T, Kiyota T and Nakagawa H (2006) Browsing system for weblog articles based on automated folksonomy. *In Proceedings of the WWW 2006 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*.

Owens L (2008) Why Social Tagging Won't Replace Formal Classification. Available: http://www.forrester.com/Research/Document/Excerpt/0,7211,44466,00.html

Sampson M (2006) What does Web 2.0 mean for Enterprise Search. *Available: http://www.slideshare.net/msampsonMNET/what-does-web-20-mean-foren terprise-search/*

Specia L and Motta E (2007) Integrating Folksonomies with the Semantic Web. *In Proceedings of the European Semantic Web Conference (ESWC 2007)*, Innsbruck, Austria: Springer.

Spiteri L F (2007) Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*, **4(2)**, Article 41. Available at: http://www. webology.ir/2007/v4n2/a41.html

Stock WG (2007b) Folksonomies and science communication: A mash-up of professional science databases and web 2.0 services. *Information Services & Use*, 27(2007)3, 97-103.

Trant J (2006) Social classification and folksonomy in art museums: Early data from the steve museum tagger prototype, *17th ASIS&T SIG/CR Classification Research Workshop*.

Trant J (2006) Social Classification and Folksonomy in Art Museums: early data from the steve.museum tagger prototype. *A paper for the ASIST-CR Social Classification Workshop.*

Ward T (2008) Taxonomy driven folksonomy. *Available: http://intranetblog.blogware.com/blog/_archives/2008/5/22/3707044.html*

Ward T. (2008) Taxonomy driven Folksonomy. *Available: http://intranetblog.blogware.com/blog/_ar chives/2008/5/22/3707044.html*

Wikipedia (2008) Taxonomy. *Available. http://en.wikip edia.org/wiki/Taxonomy*

About the authors

Ching-Chieh Kiu received her B.BA (Business Administration) in 1998 and M. IT (Information Technology) in 2000 from National University of Malaysia, Malaysia. She obtained her Ph.D (IT) in 2007 from Multimedia University, Malaysia. Currently, she is a research associate in Knowledge Management Research Center, The Hong Kong Polytechnic University. She is actively involved in applying data mining to pharmacy and corporate taxonomy / folksonomy projects. Prior to joining the KMRC, she was a lecturer at Multimedia University, Malaysia. Her joined research interests include Semantic Web technology, ontology interoperability, Knowledge Management, E-Learning, social network, taxonomy and folksonomy.

Eric Tsui was *Chief Research Officer*, Asia Pacific at Computer Sciences Corporation (Australia) from 2000 to 2005. Prior to that, he was a Principal Knowledge Engineer on expert systems projects at The Continuum Company. Currently, he is the *Professor of knowledge management* and Deputy Director of the Knowledge Management Research Centre at The Hong Kong Polytechnic University. His research areas are Knowledge Management Technologies, E-Learning and Taxonomies. Professor Tsui is a member of the Australian Computer Society, the Hong Kong Computer Society and a Vice President of the Hong Kong Knowledge Management Society. He has B.Sc. (Hons) (Sydney), PhD (Deakin) and MBA (Univ. of Southern Queensland) degrees.