Journal of Web Engineering, Vol. 1, No.1 (2015) 181-195

ORGANIZING INFORMATION IN MEDICAL BLOGS USING A HYBRID TAXONOMY-FOLKSONOMY APPROACH

YAMEN BATCH

Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology
National University of Malaysia, Selangor, Malaysia
yamenbatch@gmail.com

MARYATI MOHD. YUSOF

Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology
National University of Malaysia, Selangor, Malaysia
mmy@ftsm.ukm.my

Received October 11, 2014
Revised January 5, 2015

The retrieval of health-related information from medical blogs is challenging, primarily because these blogs lack a systematic method to organize their posts. This paper investigated the application of a hybrid taxonomy-folksonomy approach in medical blogs by reviewing the existing approaches that apply the hybrid taxonomy-folksonomy structure in web-based systems. The review showed that the hybrid structure was promising for enhanced classification of resources in web-based systems; particularly in a specific type of medical blog known as a physician-written blog. However, further research is needed to truly identify the long-term impact of the hybrid structure and its benefit in achieving a better organization of other categories of medical blogs.

Key words: Web-Based Systems, Medical Blogs, Folksonomy, Taxonomy
Communicated by: M. Gaedke & L. Carr

1 Introduction

The Internet is leading an evolution in the manner in which health information is delivered to health professionals and consumers [1]. The term “e-health” is broadly used to describe the use of the internet or web technology in healthcare [2]. One of the major advances of the internet was the introduction of the Web 2.0 technology in 2004 [3], which represented a new generation of websites and services that supported online collaboration and sharing among users [4]. This provided benefits for an easy-to-use and free social software [5]. Patients participating in Web 2.0 communities could share information about their conditions, diagnoses and medications with others [6]. Physicians also use Web 2.0 applications (e.g., wikis, blogs, forums) to share advice and expertise with other physicians and keep up-to-date with the latest advances in their specialties [7]. Web 2.0 also strengthened the relationship

between patients and doctors [8]. Some Web 2.0 sites can assist patients in selecting the best doctor or health service [9], or even make appointments with physicians [10].

A growing interest was witnessed in the web-based collaboration ware (namely wikis, blogs and podcasts) which was adopted in the dissemination of health-related information among health professionals and health consumers [11]. The use of blogs in the healthcare context is growing [12, 13]. A medical blog primarily discuss healthcare topics; i.e., its posts usually focuses on medical topics such as diseases, medical treatments and medications [14]. Medical blogs provide health professionals with new channels to share health information with patients and members of the public [15, 16]. These blogs are categorized based on their authors: blogs that were (1) physician-written, (2) nurse-written and (3) patient-written [17]. Patients use blogs to share their own experiences on health and diseases [17]; a few examples include Diabetes Mine blog and My Breast Cancer blog. In contrast, health professionals use blogs to share their practical knowledge and skills [17]; such as in the Clinical Cases blog and Kevin MD blog. Health professionals and health consumers produce significant health-related content through blogs [18]. However, similar to other Web 2.0 sites, medical blogs have drawbacks that include scattered information of an uncertain quality [16]. Retrieving the content of medical blogs is currently challenging, particularly in terms of extracting relevant health-related information; given the lack of systematic methods in the organization of blog posts [19]. Therefore, these blogs still do not meet the expectations of their users in terms of retrieving relevant health information from their posts [20].

One of the most important functions of web information retrieval systems is to organize contents in such a way that users are able to easily retrieve relevant information [21]. Information organization (i.e., how the information is structured) is a crucial aspect to consider in order to achieve an enhanced information retrieval result [22]. Two types of information organization schemes are used in web-based systems: taxonomies (i.e., predefined classifications) and folksonomies (i.e., user-generated classifications). Taxonomies offered consistent classifications of web resources; however, they were not able to represent users' vocabularies that continue to emerge in online communities. Folksonomies offered flexible and adaptable classifications of web resources; however, they lacked precision when describing resources. Several approaches have hybridized taxonomies with folksonomies to better organize and classify web resources; unfortunately, how a hybrid taxonomy-folksonomy structure can contribute to organizing information in medical blogs is still not well understood.

The aim of this paper is to explore the benefits of applying a hybrid taxonomy-folksonomy in medical blogs. To achieve this aim, first, we discussed the characteristics and content of medical blogs. Second, we compared taxonomies to folksonomies and reviewed the approaches that integrated both classifications in web-based systems. Third, we presented an example of applying the hybrid approach in physician-written blogs highlighting its benefits and limitations. Finally, the paper was concluded with some implications for further research.

2 The Characteristics of Medical Blogs

Medical posts contain information about medical concepts such as diagnoses, procedures or medications [23]. To gain a better understanding of their characteristics and contents, some of the research accomplished in the field of medical blogs were reviewed and discussed. To date, only a few studies have focused on the content of medical blogs. Table 1 summarized these studies.

Author	Objective	Method	Results
Denecke [14]	To introduce an algorithm to categorize posts according to their information type and to determine whether they contained medical content.	The algorithm used an existing information extraction system (SeReMeD) to extract entities on diagnoses, procedures and medications based on a fixed classification of related medical semantic types.	Distinguished informative posts from affective posts in medical blogs and identified medical-related information in posts.
Lagu, Kaufman [15]	To evaluate the content of blogs written by health professionals (i.e., health-related or otherwise) and the characteristics of medical bloggers.	Two hundred and seventy-one medical blogs were chosen and five entries per blog were reviewed to identify content.	Nearly half of the blogs discussed healthcare-related topics. Over half of the blogs had identifiable authors. More than half of the blogs described interactions with individual patients or promoted healthcare products.
Denecke [24]	To introduce a method to study diversity in medical blogs' topics.	The method was based on information extraction and domain knowledge and was applied to a set of medical posts.	The method helped improve post retrieval by detecting the medical category of the post.
Miller and Pole [12]	To analyze the content of health blogs.	Identified a sample of 951 health blogs in 2007 and 2008. All blogs were U.S.-focused and were updated regularly; their features and topics were analyzed.	Most blogs focused on health topics from professional and patient perspective.
Wagner, Paquin [25]	To assess the prevalence of various genetics-related topics and perceived credibility indicators in Genetics Blogs.	Conducted a content analysis of genetics blogs (n = 94) and archived blog contents published between June 15, 2007 and June 15, 2009.	In order to improve Genetics blogs access, bloggers should consider updating blog content frequently and providing information related to their expertise.
Greenberg, Yaari [26]	To examine the credibility of medical blogs and their published information as perceived by their readers.	Constructed six blogs on diabetes treatment. Each blog was viewed by approximately 60 participants. The participants were asked to fill in a short questionnaire that measured their perceived credibility of the blog and its information.	The participants have an attitude of scepticism and/or criticism of many aspects of the information in the blogs in spite of their desire and readiness to use this information.

Table 1 Summary of studies that discussed the content of medical blogs

The first study by Denecke [14] introduced an algorithm to classify medical posts according to their information type, and further distinguished informative posts from affective posts. A medical post was considered affective if it described the author's feelings on treatments, diseases or medications. A medical post was considered informative if it contained information on diseases, treatments or information related to healthcare. An affective post was regarded as unhelpful to the readers since it reflected authors' emotions and did not hold any medical concepts. The second study by Lagu, Kaufman [15] attempted to examine the content of medical blogs and discovered that nearly half (50.6%) of the sampled blogs included professional content, and that over half of the blogs had identifiable authors (i.e., authors that provided their real names, specialty and location). The third study by Denecke [24] aimed to determine diversity in medical blogs' content. He introduced a method that measured the diversity of a post's content and groups' search results, according to the dimensions of diversity, to improve the retrieval of other posts that were related to the same topic. The method was considered efficient since it presented various topical aspects that enabled users to see different outlooks of their queries. The fourth study by Miller and Pole [12] analyzed the content and characteristics of popular health blogs in the U.S. They noted that most blogs focused on bloggers' experiences with one disease or condition, or on personal experiences of health professionals. Half of the blogs were written from the perspective of professionals, and one-third of the blogs were written from the perspective of patients. The fifth study by Wagner, Paquin [25] assessed the prevalence of various genetics-related topics and perceived credibility indicators in Genetics Blogs. They conducted a content analysis on a population of 94 Genetics blogs. The included blogs were required to be written in English, to contain 2 or more posts and to have been updated in the past 6 months. Blog contents published between June 15, 2007 and June 15, 2009 were archived. The results showed that most blogs disclosed authors' full names (81%) and biographical information (67%). Many blog authors reported having genetics (67%) or life science expertise (59%). The authors concluded that in order to maintain or increase their blog traffic, genetics bloggers should consider to improve the blog credibility standards by updating its content frequently and providing information related to their expertise. The last study by Greenberg, Yaari [26] examined the credibility of medical blogs and the information published in them as perceived by their readers. They constructed six blogs, each with two posts; one post on conventional treatment and the other on an alternative diabetes treatment. Each blog was viewed by approximately 60 participants. The participants were asked to fill in a short questionnaire that measured their perceived credibility of the blog and its information. The results showed that the participants have an attitude of scepticism and/or criticism of many aspects of the information in the blogs in spite of their desire and readiness to use this information.

Although the credibility of medical blogs needs to be improved, the analysis of the above studies demonstrated that medical blogs provided a significant online medium that can be employed by physicians or patients to discuss and share medical topics. In addition, the extraction of medical topics of posts improved their information retrieval. Given that medical blogs will continue to increase in both number and content [27], unorganized blog entries could lead to ineffective content retrieval. A widely-used approach to organize blog posts was for the creator or viewer to add metadata. Users may add such metadata through two different methods: (1) by the use of tags (i.e., folksonomies) and (2) by the use of a set of predefined categories (i.e., taxonomies) [28]. The following section discussed folksonomies and taxonomies, and highlighted the differences between both types of classifications.

3 Information Organization Schemes: Taxonomies versus Folksonomies

A taxonomy is a controlled vocabulary, typically created by domain experts, that establishes hierarchal relationships between terms [21, 29] and represents coherent classification systems [30]. Taxonomies organize items of a particular domain by providing a one-thing-in-one-place model. Each item belongs to a specific category that, in turn, belongs to a more general one [31]. Examples of common taxonomies are the Linnaean system of classifying living things, and the Library of Congress Subject Headings (LCSH). The controlled vocabulary of taxonomies offered several characteristics, as follows [32]:

- It controlled the use of synonyms by establishing a single form of the term. This ensured that indexers applied the same terms to describe the same or similar concepts.
- It discriminated between homonyms, allowing the indexer to resolve clashes of meaning that arise when several terms have the same form with distinct meanings (e.g., "Java" the programming language, or "Java" the coffee, or "Java" the Indonesian island).
- It controlled lexical anomalies by minimizing any grammatical variations that could potentially create further noise during retrieval.
- It facilitated the use of codes or notations which can then be associated with terms.

Taxonomies are widely used in many websites for content indexing and retrieval [33]. They are used to organize information in hierarchical menus that enables easy access to corresponding web pages. Typical examples of web directories are Google and Yahoo! Directory [34]. However, content navigation support through a taxonomy is often limited [33] since content is classified by experts whose viewpoints differ from those of users [35]. Therefore, taxonomy classification is unable to represent users' vocabularies. A further limitation of taxonomies is that they are unable to index new resources added by users. Thus, such resources may be poorly classified, decreasing their retrievability [33, 36]. In addition, building and maintaining taxonomies is a tedious and an expensive task [35].

With the emergence of Web 2.0, another classification system called folksonomy was introduced. The word "folksonomy" is a fusion of the words "folk" and "taxonomy" [37]. A folksonomy is a bottom-up classification that resulted from a process called collaborative tagging, by which different users add tags to shared resources [31, 38-41]. Tagging systems became popular in Web 2.0 sites [42] as new tools that help users to organize, share and retrieve information [32, 43, 44]. Some of the most used collaborative tagging services are: *Delicious* (a social bookmarking website), *Flickr* (a web-based photograph management application), *CiteULike* (a tool for sharing academic papers) and *Technorati* (a search engine that allows blog authors to tag and search their blog posts) [45, 46]. Folksonomy tags provide a source of terms to describe new resources that continue to emerge in online communities [35, 47-49]. Unlike the one-thing-in-one-place model of taxonomies, folksonomy tags are able to describe different aspects of a resource [50]. Folksonomies have many advantages, including flexibility [51], low-cost, ease of use and the ability to express users' vocabularies [52]. In addition, many people enjoy tagging and are willing to add tags [53, 54].

Taxonomies are more suitable than tagging when organizing specific domains that tend to have stable content [32, 55]. On the other hand, folksonomies allows a higher adaptability when organizing information, compared to taxonomies [55, 56]. However, folksonomies are not able to replace taxonomies due to the fact that they deal with inconsistencies in terms of content retrieval [57], which is attributed to tagging resources without any form of vocabulary control [35, 58]. Many researchers

acknowledged the ambiguity of tags' meanings as the most significant challenge to collaborative tagging systems [59-62]. Namely, tagging systems are unable to identify the meanings of tags or the relationships between tags and resources [31, 52, 60, 63]. These semantic problems are associated with the numerous ways of using tags that include [35, 64]:

- Polysemy (the same tag can refer to different concepts: the tag "field" refers to a piece of land and to a branch of knowledge);
- Synonymy (different tags refer to the same concept: "car" and "automobile" both refer to a vehicle);
- Different lexical forms (various tag forms can refer to the same concept: car/cars, energy/energetic, pc/personal computer); and
- Different levels of abstraction (jazz or music).

Noruzi [65] asserted that folksonomies need to have a type of taxonomy control to maintain a better and consistent classification of resources. At the same time, folksonomy tags provide an opportunity for taxonomy-based systems to enhance the access to resources [66]. Table 2 summarized the differences between taxonomies and folksonomies.

Taxonomy	**Folksonomy**
Hierarchical, top-down and rigid classification	Flat, bottom-up and flexible classification
Controlled vocabulary	Free vocabulary
Created by experts	Created by users
Precise	Lack of precision
Can become outdated quickly	Adapts quickly to changes in users' vocabulary
Building taxonomies is tedious and expensive	Low-cost and easy to build classification

Table 2 Comparing taxonomies and folksonomies

Most classification schemes in web-based systems are either taxonomy or folksonomy-based. Combining both taxonomies with folksonomies may be a solution for delivering a superior classification system that can leverage the benefits of both classification schemes [67]. The following section presented a review of the existing approaches that adopted the hybrid model of taxonomy and folksonomy in web-based systems.

4 Combining Taxonomies and Folksonomies in Web-Based Systems

The combination of a taxonomy and a folksonomy resulted in a hybrid structure that offered the following advantages [33, 68]:

- Improved content searching and retrieval: Tagging resources by large group of users helps to equip search engines with large amounts of content tagged with keywords describing the users' received meaning which may never appeared in the content before.
- Enhanced taxonomy management process: Tags used with high frequency, or that meet the specific domain requirements, become candidates for inclusion in the taxonomy. These tags can help update taxonomy terms providing the consistency and broad usability of taxonomy.
- Creation of new navigational facets of resources: Indexing resources using human language goes beyond a computer's ability to analyze text, delivering results that are relevant from the user perspective.

- Classification of resources with minimal costs: The hybrid classification consists of taxonomy and folksonomy. The taxonomy is created once and folksonomy tags, which users add free of charge, are used to update the taxonomy. Thus, the cost of the hybrid classification is lower than having taxonomy alone.

The following four hybrid taxonomy-folksonomy approaches were: (1) coexistence of folksonomy and taxonomy, (2) folksonomy-directed taxonomy, (3) taxonomy-directed folksonomy and (4) folksonomy hierarchies [33, 69].

i. Co-existence: Taxonomy and folksonomy both exist in a given system. Each classification was independent and no relationship existed between the taxonomy terms and the folksonomy tags.
ii. Folksonomy-directed taxonomy: Both taxonomy and folksonomy co-exist; folksonomy served as a pool of candidate terms (tags) that may represent new terminology to enrich and update the taxonomy. This hybrid approach required a selection process that judge candidate tags based on their frequency and value within context.
iii. Taxonomy-directed folksonomy: This approach provided tag suggestions to users from a controlled set of terms in the form of drop-down menus, check boxes or type ahead. Such a hybrid structure was more consistent than traditional tagging systems and had better retrieval results than taxonomies.
iv. Folksonomy hierarchies: There were two kinds of folksonomy hierarchies: user-powered and automatic derivation. User-powered was a social classification created by a small population. In contrast, automatic derivation was created through clustering algorithms.

A number of studies explored the use of hybrid structures in web-based systems. Table 3 provided a summary of these studies and organized them according to the types of hybrid approaches mentioned above. One study by Alemneh and Rorissa [67] explored the potentials of user supplied tags or keywords in terms of complementing established controlled vocabularies in digital libraries. However, the study only discussed the theoretical aspects of combining folksonomies and taxonomies and did not provide an actual proof-of-concept.

Three of the existing approaches that combined taxonomies and folksonomies fell within the coexistence category. They were: *TaxoFolk, Tagsonomy* and *ICDTag*. TaxoFolk is an algorithm that integrates folksonomies into taxonomy to enhance classification and navigation of web resources [33]. The algorithm is comprised of four major phases: (1) tag pre-processing phase, (2) domain contextualization phase, (3) contextual clustering phase and (4) concept-tag consolidation phase. An experiment was conducted to evaluate the TaxoFolk approach. In this experiment, the chosen taxonomy was GovHK's portal (http://www.gov.hk/) and its folksonomy was obtained from the Delicious database. The taxonomy consisted of six levels that covered 752 web-sites. Each of these websites had a minimum of five users who had tagged it. The results of the experiment demonstrated that the techniques applied in the algorithm were promising in producing an efficient hybrid classification of resources.

Tagsonomy is a mechanism used to retrieve information from websites by applying a hybrid taxonomy–folksonomy approach, introduced by Sommaruga, Rota [34]. A web-based system called "Easy Access" was developed and tested with real users. The system was initially tested by simulating a number of different scenarios to verify that the system triggered the expected behavior. Each scenario was simulated to capture user interactions, and the results were compared with the retrieval of information; without using Tagsonomy.

Type of Hybrid Approach	Description	Advantages	Limitations
Co-existence *Empowering Digital Libraries Users through Combining Taxonomies with Folksonomies* [67]	Review of best practices and emerging trends in indexing resources in digital libraries.	The study explored the potentials of user supplied tags or keywords in terms of complementing established controlled vocabularies in digital libraries.	The study presented the theoretical aspects of combing folksonomies and taxonomies; without the proper proof-of-concept.
Co-existence *TaxoFolk* [33]	An algorithm that integrates folksonomies into taxonomy to enhance classification and navigation of web resources.	The techniques applied in the algorithm were promising to produce an efficient hybrid classification of resources.	TaxoFolk used data-mining algorithms that may sometimes give inaccurate results.
Co-existence *Tagsonomy* [34]	A web-based tool that combines a top-down classification defined by the website content manager, and a bottom-up classification defined by users to retrieve information from websites.	Tagsonomy helped users find relevant information by combining the users' search keywords and the predefined classification of information.	Tags were extracted from users' search keywords and did not originate from explicit tagging activities; therefore, tags did not reflect users' vocabularies.
Co-existence *ICDTag* [70]	A prototype for a web-based system that implements a combination of a taxonomy classification scheme and user-generated tags to organize physician-written blog posts.	The prototype improved the organization of posts in physician-written blogs by combining medical taxonomy (ICD-11 categories) with the user-added tags, which led to a better indexing and browsing of posts.	The results were limited to physician-written blog posts that discussed disease-related content only, and could not be generalized to other types of blogs.
Taxonomy-directed folksonomy *MyEdna* [71]	A taxonomy-directed folksonomy portal that allows users to label resources with tags prompted by taxonomies listed in drop-down menus.	MyEdna produced a more consistent categorization of resources by enabling its community users to make discussions and connections with the use of tags and resources.	The portal was designed on a conceptual level only, without proper development or evaluation.
Folksonomy hierarchies *FaceTag* [72]	A prototype for a semantic collaborative tagging tool that adds users' tags to a faceted classification scheme to improve the information organization in bookmarking systems.	Tag hierarchies and facets could possibly improve and disambiguate the meaning of tags, giving them a more coherent organization.	The prototype had not been tested in a real-world scenario.

Table 3 Summary of studies that combined taxonomy and folksonomy

For example, when several users selected various tags in the tag cloud during a search session, the number of tags that corresponded to the most visited pages increased. This preliminary test was useful to verify that Tagsonomy worked properly by helping users find relevant information with the combination of users' search keywords with predefined classification of information.

Although TaxoFolk and Tagsonomy provided feasible methods to integrate folksonomy with taxonomy, they had several limitations to consider. TaxoFolk produced the taxonomy–folksonomy structure using data-mining algorithms that may result in imprecise classification of resources. Tags within Tagsonomy were extracted from users' search keywords instead of explicit tagging activities; therefore, they did not reflect users' vocabulary.

A more user-driven realization of the taxonomy-folksonomy structure, called ICDTag, was introduced by Batch, Yusof [70]. It is a web-based system that implements a combination of a taxonomy classification scheme and user-generated tags to organize physician-written blog posts and extract information from these posts. In contrast to Tagsonomy, tags of ICDTag are explicitly provided by blog users. Therefore, the tags reflect the users' vocabulary. The ICDTag approach differs from TaxoFolk in producing the hybrid classification through the grouping of most frequently-used tags under medical categories, instead of applying data-mining algorithms.

MyEdna, a taxonomy-directed folksonomy portal was proposed by Hayman and Lothian [71], which allows users to label resources with tags prompted by a taxonomy listed in drop-down menus. The target website was the Education Network Australia (Edna), and the taxonomy used was established from the Australian education sector. MyEdna was expected to produce a more consistent categorization of resources and enable its community users to make discussions and connections using tags and resources. However, the portal was designed on a conceptual level only, without proper development or evaluation. As a result, the effectiveness of MyEdna in achieving its objectives was not assessed.

FaceTag was introduced by Quintarelli, Resmini [72] as a working prototype of a semantic collaborative tagging tool. It aimed to effectively add users' tags to a faceted classification scheme to improve the information organization capabilities of bookmarking systems. Tag hierarchies were semantically assigned to editorially established facets in order to enable flexible navigation of resources. Using semantically classifying user-added tags into facets, FaceTag may solve most of the semantic problems pertinent to polysemy and homonymy of tags. Preliminary user evaluations showed that the addition of tag hierarchies and facets improved and disambiguated the meaning of tags, giving them a more coherent organization. However, the FaceTag prototype has not been tested in a real-world scenario.

Four of the existing approaches (i.e., TaxoFolk, Tagsonomy, MyEdna and FaceTag) applied the hybrid approach in web-based systems. Motivated by their promising evaluations, the ICDTag approach [70] was implemented in medical blogs (i.e., physician-written blogs). The following section further elaborated on the ICDTag approach and how it was designed to improve the organization of medical posts.

5 ICDTag: An Application of the Hybrid Approach in Medical Blogs

One application of the hybrid taxonomy-folksonomy approach in medical blogs would be ICDTag [70], which is a prototype for a web-based system that allows physicians to organize posts using a hybrid taxonomy-folksonomy. ICDTag is introduced as an example because it is the only type of hybrid approach that is applied in medical blog context. The ICDTag was particularly meant for physician-written blogs written in English. Physician-written blogs were selected because they were better suited for generating and extracting medical information; since physicians were a major component of the medical blogging community [15] due to their use of blogs to discuss medical issues [17]. Following this approach, physicians who access blogs were presented with two classification schemes: (1) to tag medical posts using their own vocabulary or (2) to categorize posts in accordance to a predetermined classification scheme. The system also supported the extraction of information from medical posts. The hybrid taxonomy-folksonomy approach of ICDTag allowed users to assign the International Classification of Diseases (ICD-11) category to blog posts when creating their own posts. Afterwards, users could collaboratively tag posts using free-text words or phrases. Consequently, each blog post would have two attributes, a category (which belonged to a professional taxonomy) and a set of tags added by users (which represented a folksonomy); as shown in Figure 1.

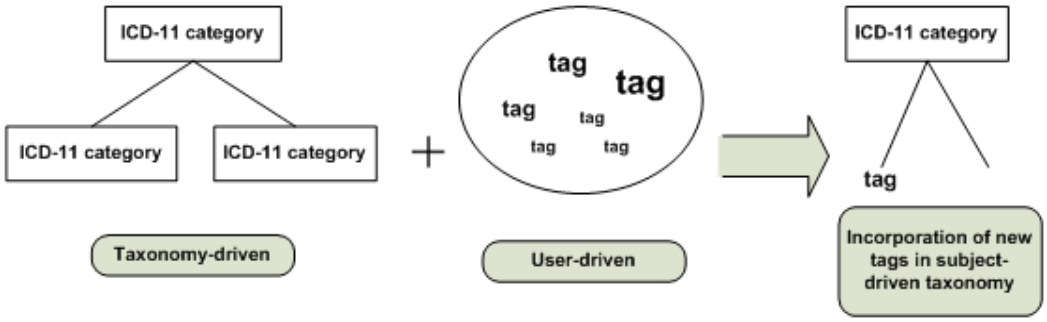

Figure 1 The integration of tags and ICD-11 categories [70]

The ICDTag prototype was evaluated using an experiment in which some physicians who were familiar with medical blogs were asked to use the prototype. The goal of the experiment was to analyze the dynamics and usage patterns of the prototype. Next, an online questionnaire was implemented to evaluate the main functions of ICDTag blogs from the perspective of end-users. The results of the questionnaire demonstrated that users had positively assessed the "browse" and "search" functionalities and the organization of the ICDTag blogs. By offering both types of classifications, the hybrid approach provided the following characteristics:

- Improved post-searching and retrieval: The existence of dual facets (i.e., tags and categories) led to better indexing, browsing and discovery of posts.
- Accuracy: The taxonomy classification of posts supported accurate retrieval of posts.
- Enhanced support and networking opportunities: Tags offered a way to share posts among physicians with common interests.

- Low-cost classification: Folksonomies enabled users to organize posts with minimal costs.

However, the taxonomy used in ICDTag (i.e., ICD-11 categories) only covered disease-related content, such as types of disease, causes, symptoms and treatments. Other contents that were unrelated to diseases, such as medical procedures and clinical trials, were not included in the ICD-11 categories. Hence, the ICDTag model was designed specifically for blog posts that only discussed disease-related information.

6 Conclusions and Future Work

To understand the implications of employing a combination of taxonomies and folksonomies in web-based systems, the existing hybrid approaches (i.e., TaxoFolk, Tagsonomy, MyEdna and FaceTag) were reviewed. Each of the approaches had the potential to enhance the classification of web resources. The hybrid approach provides high-level descriptions and representations of Web resources. Such a classification of Web resources could serve as a very powerful and flexible tool for increasing the user-friendliness and interactivity of online communities. In addition, combining the strengths of folksonomies and taxonomies offers powerful information organization and search capabilities. Using the hybrid approach, each Web resource has a semantic value (i.e., taxonomy term) and a social value (i.e., folksonomy tags). Thus, future applications for the hybrid approach will benefit from the semantic and social values of Web resources to extract more useful information. These applications might include mining users' opinions, recommender systems, personalized search and inferring semantic relations among Web resources.

Similar to web-based systems, integrating taxonomies with folksonomies in medical blogs had the potential to achieve better post classification. Folksonomy tags and taxonomy terms were used to describe different aspects of posts. Folksonomy tags represented users' opinions on posts, and each post was associated with a taxonomy term that reflected experts' viewpoints on the post. ICDTag was an example of the application of the hybrid approach for organizing posts in physician-written blogs, which improved organization by combining medical categories with user-added tags. However, the results of ICDTag could not be generalized to other types of medical blogs since it was limited to physician blog posts that discussed disease-related information. The hybrid approach improved both the structure and retrieval of posts from physician-written blogs. Thus, is worthwhile to apply the hybrid approach to other types of medical blogs to make them a more valuable and reliable source of health information for online medical communities. Further research and evaluations are needed to identify other long-term impacts of the hybrid approach in medical blogs, its benefit for medical bloggers and the kind of hybrid knowledge that it will generate. In future work, the hybrid approach can also be applied on different social media sites such as medical wikis and health forums. By using the hybrid approach, health professionals and health consumers will be able to better organize online medical resources by adding their own tags to these resources. Adding tags to medical resources have the potential to contribute to medical knowledge because tags might represent new medical terms.

References

1. Witteman, H. and L. O'Grady. E-health in the era of web 2.0. in Proceedings of the Virtually Informed: The Internet as (new) Health Information Sources, Final Conference of the Project Virtually Informed—The Internet in the Medical Field, The University of Vienna, Vienna, Austria. 2008.
2. Meier, C.A., M.C. Fitzgerald, and J.M. Smith, eHealth: extending, enhancing, and evolving healthcare. Annual Review of Biomedical Engineering, 2013. 15: p. 359-382.
3. Allen, M., Tim O'Reilly and Web 2.0: the economics of memetic liberty and control. Communication, Politics & Culture, 2009. 42(2): p. 6-23.
4. Cheung, K.-H., et al., HCLS 2.0/3.0: health care and life sciences data mashup using Web 2.0/3.0. Journal of Biomedical Informatics, 2008. 41(5): p. 694-705.
5. Giustini, D., How Web 2.0 is changing medicine. BMJ, 2006. 333(7582): p. 1283-1284.
6. Weitzel, M., et al., A Web 2.0 model for patient-centered health informatics applications. Computer, 2010. 43(7): p. 43-50.
7. Boulos, M.N.K. and S. Wheeler, The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Information & Libraries Journal, 2007. 24(1): p. 2-23.
8. Pittler, M., et al., Evidence-based medicine and Web 2.0: friend or foe? The British Journal of General Practice, 2011. 61(585): p. 302.
9. Bottles, K., Patients, doctors and health 2.0 tools. Physician Executive, 2009. 35(4): p. 22-5.
10. Afzal, M., et al. Social media canonicalization in healthcare: Smart cdss as an exemplary application. in e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on. 2012. Beijing, China: IEEE.
11. Murray, P., Web 2.0 and social technologies: what might they offer for the future of health informatics. Health Care and Informatics Review Online, 2008. 12(2): p. 5-16.
12. Miller, E.A. and A. Pole, Diagnosis blog: checking up on health blogs in the blogosphere. American Journal of Public Health, 2010. 100(8): p. 1514-1519.
13. Rozenblum, R. and D.W. Bates, Patient-centred healthcare, social media and the internet: the perfect storm? BMJ Quality & Safety, 2013. 22(3): p. 183-186.
14. Denecke, K. Accessing medical experiences and information. in European Conference on Artificial Intelligence, Workshop on Mining Social Data. 2008.
15. Lagu, T., et al., Content of weblogs written by health professionals. Journal of General Internal Medicine, 2008. 23(10): p. 1642-1646.
16. Moorhead, S.A., et al., A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. Journal of medical Internet research, 2013. 15(4).
17. Denecke, K. and A. Stewart, Learning from Medical Social Media Data: Current State and Future Challenges, in Social Media Tools and Platforms in Learning Environments, B. White, I. King, and P. Tsang, Editors. 2011, Springer Berlin Heidelberg. p. 353-372.
18. Fernandez-Luque, L., R. Karlsen, and J. Bonander, Review of extracting information from the social web for health personalization. Journal of Medical Internet Research, 2011. 13(1).
19. Kim, S., Content analysis of cancer blog posts. Journal of the Medical Library Association: JMLA, 2009. 97(4): p. 260-266.
20. Gavgani, V.Z. and V.V. Mohan, Application of Web 2.0 tools in medical librarianship to support medicine 2.0. Webology, 2008. 5(1).

21. Sommaruga, L., P. Rota, and N. Catenazzi, "Tagsonomy": Easy Access to Web Sites through a Combination of Taxonomy and Folksonomy, in Advances in Intelligent Web Mastering – 3, E. Mugellini, et al., Editors. 2011, Springer Berlin Heidelberg. p. 61-71.
22. Nielsen, J. and H. Loranger, Prioritizing Web Usability. 2006, Berkeley CA: New Riders Press.
23. Denecke, K. and W. Nejdl, How valuable is medical social media data? Content analysis of the medical web. Information Sciences, 2009. 179(12): p. 1870-1880.
24. Denecke, K. Assessing content diversity in medical weblogs. in First International Workshop on Living Web, collocated with the 8th International Semantic Web Conference (ISWC 2009). 2009. Washington, D.C, USA: CEUR-WS.
25. Wagner, L., R. Paquin, and S. Persky, Genetics blogs as a public health tool: assessing credibility and influence. Public health genomics, 2012. 15(3-4): p. 218-225.
26. Greenberg, S., E. Yaari, and J. Bar Ilan, Perceived credibility of blogs on the internet – the influence of age on the extent of criticism. Aslib Proceedings, 2013. 65(1): p. 4-18.
27. Lee, S.-F. and W.-J. Jih, Exploring the effects of blog visit experience on relationship quality: an empirical investigation with a cardiac surgery medical blog site. International Journal of E-Business Research (IJEBR), 2012. 8(2): p. 1-14.
28. Passant, A. Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs. in Proceedings of International Conference on Weblogs and Social Media. 2007. Boulder, CO, USA.
29. Pak, R., S. Pautz, and R. Iden, Information organization and retrieval: A comparison of taxonomical and tagging systems. Cognitive Technology, 2007. 12(1): p. 31-44.
30. Perry, P.B., et al. Leveraging a Technical Domain Taxonomy to Enhance Collaboration Knowledge Sharing and Operational Support. in SPE Digital Energy Conference and Exhibition. 2011. The Woodlands, Texas, USA: Society of Petroleum Engineers.
31. Golder, S.A. and B.A. Huberman, Usage patterns of collaborative tagging systems. Journal of Information Science, 2006. 32(2): p. 198-208.
32. Macgregor, G. and E. McCulloch, Collaborative tagging as a knowledge organisation and resource discovery tool. Library Review, 2006. 55(5): p. 291-300.
33. Kiu, C.C. and E. Tsui, TaxoFolk: a hybrid taxonomy–folksonomy structure for knowledge classification and navigation. Expert Systems with Applications, 2011. 38(5): p. 6049-6058.
34. Sommaruga, L., P. Rota, and N. Catenazzi, "Tagsonomy": Easy Access to Web Sites through a Combination of Taxonomy and Folksonomy. Advances in Intelligent Web Mastering–3, 2011: p. 61-71.
35. Spiteri, L.F., The structure and form of folksonomy tags: the road to the public library catalog. Information Technology and Libraries, 2013. 26(3): p. 13-25.
36. Choi, Y. Enhancing Access to the Web: Vocabulary Analysis on Users' Tags and Professionals' Index Terms. in Poster presented at iConference. 2010. Champaign, IL, USA.
37. Vander Wal, T. Folksonomy definition and Wikipedia. 2005 [cited 2009 29 December]; Available from: http://www.vanderwal.net/random/entrysel.php?blog=1750.
38. Zauder, K., J.L. Lazic, and M.B. Zorica. Collaborative Tagging Supported Knowledge Discovery. in Information Technology Interfaces, 2007. ITI 2007. 29th International Conference on. 2007.
39. McFadden, S. and J.V. Weidenbenner, Collaborative tagging: traditional cataloging meets the "wisdom of crowds". The Serials Librarian, 2010. 58(1-4): p. 55-60.
40. Zubiaga, A., Enhancing navigation on Wikipedia with social tags. arXiv preprint arXiv:1202.5469, 2012.
41. Apse, D. and T. Ley, Collaborative Tagging Applications and Capabilities in Social Technologies, in Open and Social Technologies for Networked Learning, T. Ley, et al., Editors. 2013, Springer Berlin Heidelberg. p. 185-188.

42. Wang, J., et al., Personalization of tagging systems. Information Processing & Management, 2010. 46(1): p. 58-70.
43. Zeng, D. and H. Li, How useful are tags?—an empirical analysis of collaborative tagging for Web page recommendation. Intelligence and Security Informatics, 2008: p. 320-330.
44. Jäschke, R., et al., Challenges in Tag Recommendations for Collaborative Tagging Systems, in Recommender Systems for the Social Web. 2012, Springer Berlin Heidelberg. p. 65-87.
45. Nov, O. and C. Ye, Why do people tag?: motivations for photo tagging. Communications of the ACM, 2010. 53(7): p. 128-131.
46. Ding, Y., et al., Upper tag ontology for integrating social tagging data. Journal of the American Society for Information Science and Technology, 2010. 61(3): p. 505-521.
47. Avery, J.M., The democratization of metadata: collective tagging, folksonomies and Web 2.0. Library Student Journal, 2010. 5.
48. Mulatiningsih, B., Collaborative tagging in participatory webs: the death of information organisation? Asia Pacific Journal of Library and Information Science, 2012. 2(2): p. 108-116.
49. Peters, I., et al., Retrieval effectiveness of tagging systems. Proceedings of the American Society for Information Science and Technology, 2011. 48(1): p. 1-4.
50. Thomas, M., D.M. Caudle, and C.M. Schmitz, To tag or not to tag? Library Hi Tech, 2009. 27(3): p. 411-434.
51. Mathes, A., Folksonomies-cooperative classification and communication through shared metadata. Computer Mediated Communication, 2004. 47(10).
52. Kim, H.L., S. Decker, and J.G. Breslin, Representing and sharing folksonomies with semantics. Journal of Information Science, 2010. 36(1): p. 57-72.
53. Fichter, D., Intranet applications for tagging and folksonomies. Online, 2006. 30(3): p. 43-45.
54. Begelman, G., P. Keller, and F. Smadja. Automated tag clustering: Improving search and exploration in the tag space. in Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland. 2006.
55. Shirky, C. Ontology is overrated: categories, links, and tags. Clay Shirky's Internet Writings, 2005.
56. Derntl, M., et al., Inclusive social tagging and its support in Web 2.0 services. Computers in Human Behavior, 2011. 27(4): p. 1460-1466.
57. Cantador, I., I. Konstas, and J.M. Jose, Categorising social tags to improve folksonomy-based recommendations. Web Semantics: Science, Services and Agents on the World Wide Web, 2011. 9(1): p. 1-15.
58. Trant, J., Studying social tagging and folksonomy: a review and framework. Journal of Digital Information, 2009. 10(1).
59. Qingfeng, L. and S.C.Y. Lu, Collaborative Tagging Applications and Approaches. MultiMedia, IEEE, 2008. 15(3): p. 14-21.
60. Passant, A., J.G. Breslin, and S. Decker, a URI is worth a thousand tags. International Journal on Semantic Web and Information Systems, 2009. 5(3): p. 71-94.
61. Wetzker, R., et al., I tag, you tag: translating tags for advanced user models, in Proceedings of the third ACM international conference on Web search and data mining2010, ACM: New York, New York, USA. p. 71-80.
62. Dattolo, A., D. Eynard, and L. Mazzola, An integrated approach to discover tag semantics, in Proceedings of the 2011 ACM Symposium on Applied Computing2011, ACM: TaiChung, Taiwan. p. 814-820.
63. Beldjoudi, S., H. Seridi, and C. Faron-Zucker. Ambiguity in tagging and the community effect in researching relevant resources in folksonomies. in Proceedings of International Workshop User Profile Data on the Social Semantic Web. 2011. Heraklion, Crete, Greece.

64. Marchetti, A., et al. Semkey: a semantic collaborative tagging system. in Workshop on Tagging and Metadata for Social Information Organization at WWW 2007. 2007. Banff, Canada.
65. Noruzi, A., Folksonomies-Why do we need controlled vocabulary? Webology, 2007. 4(2).
66. Lu, C., J.-R. Park, and X. Hu, User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. Journal of Information Science, 2010. 36(6): p. 763-779.
67. Alemneh, D.G. and A. Rorissa, Empowering digital libraries users through combining taxonomies with folksonomies. Proceedings of the American Society for Information Science and Technology, 2012. 49(1): p. 1-3.
68. Spiteri, L., User-Generated Metadata: Boon or Bust for Indexing and Controlled Vocabularies?, in Indexing Society of Canada2013: Nova Scotia Trunk 3, Halifax, Canada.
69. Beatch, R. and P. Wlodarczyk. Hybrid approaches to taxonomy & folksonomy. in Semantic Technology Conference 2009. 2009. San Jose, CA, USA.
70. Batch, Y., M.M. Yusof, and S.A. Noah, ICDTag: A Prototype for a Web-Based System for Organizing Physician-Written Blog Posts Using a Hybrid Taxonomy-Folksonomy Approach. Journal of medical Internet research, 2013. 15(2).
71. Hayman, S. and N. Lothian. Taxonomy directed folksonomies. in New Developments in Social Bookmarking, Ark Group Conference: Developing and Improving Classification Schemes. 2007. Sydney, Australia: Citeseer.
72. Quintarelli, E., A. Resmini, and L. Rosati, Information architecture: Facetag: integrating bottom-up and top-down classification in a social tagging system. Bulletin of the American Society for Information Science and Technology, 2007. 33(5): p. 10-15.