This is a preprint of an article accepted for publication in *Information Discovery and Delivery*. Huang, H., Yu H., Li W. (in press, 2025). Deep Learning-Enhanced Metadata and Dynamic Facet Representations for Temporal-Semantic Analysis of User-Generated Image Tags. *Information Discovery and Delivery*.

Deep Learning-Enhanced Metadata and Dynamic Facet Representations for Temporal-Semantic Analysis of User-Generated Image Tags

Hong Huang

Associate Professor at the School of Information
University of South Florida, Tampa, Florida, USA
ORCID ID 0000-0002-8957-4881
honghuang@usf.edu

Han Yu

Associate Professor at the Department of Applied Statistics and Research Methods
University of Northern Colorado, Greeley, Colorado, USA
han.yu@unco.edu

Wanwan Li

Associate Professor at the Computing and Mathematics Department
Oral Roberts University, Tulsa, Okaholma, USA
wanli@oru.edu

Abstract

Purpose

This study explores the temporal evolution of user-generated popular tags in Flickr, a collaborative image tagging system, through the lens of facet classification. It aims to uncover how tagging behaviors reflect shifts in technology, culture, and individual cognition, and how deep learning techniques can enhance our understanding of the semantic structures within folksonomies.

Design/methodology/approach

Popular tags from three benchmark years (2006, 2010, and 2015) were collected and manually categorized using a faceted classification scheme rooted in Ranganathan's model. To augment this analysis, we applied autoencoder-based deep learning models to extract latent semantic representations, and pretrained word embeddings to measure semantic similarity. This hybrid approach enabled both qualitative categorization and quantitative analysis of temporal-semantic patterns in tag usage.

Findings

The study found that over 80% of the most popular tags were consistently associated with spatial, personality, and material facets. Temporal analysis revealed a marked shift from time-based tags to more self-expressive, personality-oriented tags, reflecting users' growing inclination toward individualism and identity expression. Location-based tags remained persistent across all years, suggesting the global, place-based nature of user engagement. Deep learning–enhanced analysis revealed semantic groupings and confirmed that tagging behavior evolves in alignment with technological and cultural developments.

Originality/value

This study offers a novel integration of traditional facet analysis (Colon Classification) and deep learning to model the dynamics of user-generated metadata on social image platforms. It advances understanding of how folksonomies, when enriched by neural representation learning, can serve as both mirrors of cultural shifts and tools for more adaptive, user-centered metadata systems. The approach contributes methodologically to digital classification research and provides insights into the cognitive and social factors shaping online tagging behavior.

Keyword: Deep Learning, Colon Classification, Faceted Classification, Social Tagging, Autoencoder, Pretrained Word Embeddings

Introduction

Social tagging systems such as Flickr represent user-generated metadata frameworks in which individuals collaboratively assign descriptive tags to digital content, producing decentralized tag vocabularies known as folksonomies, which are typically less hierarchical and more fluid than formal classification schemes (Golder & Huberman, 2006; De Salve, Guidi, & Michienzi, 2021; Das et al., 2022). The social tagging systems provide some degree of navigability (Allam et al., 2020). Users in these communities tag resources, leading to the creation of a "folksonomy" or a collective classification system. There is an explicit social structure of Flickr, where users can declare their peers as "contacts" and join "groups" to share photos and comments (Capocci et al., 2010). As a crowd contributed social tagging system, people collaboratively tag or annotate their image using multi-facet tags to their contributed or their favorite image (Das et al., 2022). It will benefit catalogers or LIS professionals to compare the categorical patterns in Flickr and the ones developed in LIS practice (Nduka, 2022).

Flickr reflects collective user behaviors, interests, and evolving cultural trends (Rorrissa, 2010; Huang & Jorgensen, 2013; Sundararaj, & Rejeesh, 2021). Unlike traditional top-down classification systems, which are designed and maintained by experts like librarian, folksonomies emerge organically from user-generated metadata, offering insights into how individuals categorize and retrieve information in digital environments (Rahman, 2012; Zappavigna, 2018). In social tagging communities like Flickr, users label images with multiple tags that capture different facets of an image's content, context, and meaning (Stuart, 2019; Zhang et al., 2016). This collaborative tagging not only facilitates image organization and retrieval but also enables the study of evolving semantic patterns in social media metadata (Li et al., 2016; Qassimi, & Abdelwahed, 2019).

Moreover, empirical studies have shown that tags in social media environments exhibit long-tailed distributions, wherein a small subset of tags accrues high-frequency usage while a vast number of tags remain infrequently used (Kordumova, van Gemert, & Snoek, 2016). This characteristic suggests that while popular tags provide insights into dominant themes, rare tags may capture more context-specific user annotations (Desrochers et al., 2016). Image collection in Flickr holds social activities and display for people traveling location and time (Van Dijck, 2011), their self-expression and personal archival purposes (Cox, Clough, and Marlow, 2016).

Prior research has indicated that folksonomy structures exhibit inherent navigability and emergent categorization patterns, demonstrating similarities to conventional classification schemes employed in library and information science (Capocci et al., 2010; Allam et al., 2020). Despite these parallels, limited research has examined the extent to which traditional image classification schemes align with the hierarchical structures observed in folksonomy-based tagging systems.

This study aims to bridge this gap by investigating the classification schemes that underpin image tagging behaviors on Flickr. Specifically, it explores the consistency of representation in user-generated tags over time for fundamental classification facets, including color, place, time, matter, and personality. Additionally, it examines the alignment of social tagging with facet classification models like Colon Classification and evaluates the depth of folksonomy-based categorization within social media environments.

Furthermore, Research was reported to integrate spatial information (geo concepts), visual content (visual concepts), and textual metadata (semantics concepts) to compute tag relevance for Flickr photos (Gugulica and Burghardt, 2023). This study extends this line of inquiry by combining manual facet classification with embedding-space representation learning to analyze Flickr's popular tags over time. We first map tags into a dense semantic vector space using pre-trained word/tag embeddings and, where appropriate, apply encoder-style objectives (e.g., autoencoder variants) to obtain compact latent representations. These vectors support downstream analyses of semantic similarity, clustering, and structure discovery, allowing us to assess whether folksonomies exhibit increasing taxonomic regularities across time. Deep learning, centered on multilayer neural architectures capable of learning hierarchical representations, provides a flexible foundation for modeling such patterns (Taleb et al., 2021; Ye et al., 2022). In particular, word/tag embeddings and autoencoder-based representation learning have been shown to capture semantic relations and latent dimensions in user-generated metadata and tagging systems (Niebler, Hahn, & Hotho, 2017; Wang, Shi, & Yeung, 2015). Importantly, because our analyses operate on the embedding space, the same workflow is readily extensible to Transformer/LLM-derived contextual embeddings (e.g., sentence or tag-in-context encoders) in future work (Reimers & Gurevych, 2019).

Ultimately, this research enriches the broader conversation on knowledge organization by highlighting the dynamic, evolving nature of social classification. It provides empirical evidence on how traditional expert-driven schemes and user-driven folksonomies interact over time, offering practical insights for improving metadata structures in digital archives, content management systems, and information retrieval applications. The findings also inform best practices for designing more intuitive image indexing frameworks by demonstrating how embedding-based semantic representations, can be integrated with autoencoder-style compression and clustering to discover latent structure, while balancing the flexibility of folksonomies with the organizational rigor of expert classification systems.

Literature Review

Social tagging systems, like Flickr, facilitate user-generated metadata that influences information retrieval, content organization, and digital archiving (Jansson, 2017; Alemu, 2018). Unlike traditional classification systems structured by domain experts, folksonomies emerge organically through collective user participation, highlighting the interplay between cognitive categorization and social dynamics (Shirky, 2005).

Faceted Classification and Collaborative Tagging

Facet classification is a method used to organize knowledge by categorizing resources based on distinct, mutually exclusive facets, each representing a different perspective or angle such as topic or genre (Hackett, & Fisher, 2019; Centelles, & Ferrer, 2024). Each facet describes the resource from its unique viewpoint, helping to provide a comprehensive overview of the item's characteristics (Irwing, Hughes, Tokarev, & Booth, 2024). To classify a particular item, it is necessary to analyze the resource to determine the facets it contains (Hackett, & Fisher, 2019). The classification notation for the item is then created by synthesizing the class notations derived from these individual facets (Irwing, Hughes, Tokarev, & Booth, 2024; Sperberg-McQueen,2015). Ranganathan's Colon Classification introduces the use of facets as specific categories such as personality, matter, energy, space, and time (Ranganathan, 1962). These categories facilitate a flexible and granular indexing system, allowing for a more detailed organization of information (Cho et al., 2018). This method supports detailed categorization and retrieval by breaking down complex information into more manageable and comprehensible segments (Irwing, Hughes, Tokarev, & Booth, 2024). Studies have explored how faceted classification principles can be applied to folksonomy-based systems, revealing significant overlaps between expert-driven and user-generated metadata structures (Quintarelli, Resmini, & Rosati 2007; Denton, 2003; Irwing, Hughes, Tokarev, & Booth, 2024).

Research on Flickr and other social tagging platforms has identified fundamental tag facets that align with traditional classification schemes, including location, objects, events, and personal expressions (Green, 2006; Stuart, 2019). While folksonomies inherently lack a rigid hierarchy, analysis of large-scale tag datasets suggests that users intuitively generate classifications resembling structured ontologies (Capocci et al., 2010; Staurt, 2019). The continued evolution of user-generated metadata highlights the potential for integrating folksonomies with existing cataloging systems to enhance information retrieval efficiency (Yu & Chen, 2020).

Semantic Insights in Social Tagging using Deep Learning Model

The semantic organization of tags in social media has been extensively studied through deep learning, large language models, and natural language processing (NLP) techniques (Gosal et al., 2019; Sit, Koylu, & Demir, 2020). Word embedding models like Global Vectors for Word Representation (GloVe) transform words into numerical vectors, capturing semantic meanings based on how words co-occur in large texts (Bhoir, Ghorpade, & Mane, 2017; Worth, 2023). This approach helps in image tagging by determining semantic similarities between terms, allowing the system to understand synonyms and manage words with multiple meanings depending on the context (Johnson, Murty, & Navakanth, 2024).

Recent developments in deep learning have revitalized the analysis of such user-generated metadata. At the heart of deep learning is representation learning, wherein neural architectures, especially deep networks, extract abstract, layered features from raw inputs (Bhoir, Ghorpade, & Mane, 2017; Worth, 2023). In the context of tagging, this enables systems to model semantic similarity and identify emergent themes. In the Transformer era, many recent systems implement this representation-learning workflow using pretrained Transformer/LLM encoders, mapping short texts (including tags) into a semantic vector space for similarity, clustering, and structure discovery (Devlin et al., 2019; Reimers & Gurevych, 2019). Because our downstream analyses operate on the embedding space, the same workflow is compatible with such contextual embeddings, even when instantiated with classic static embeddings. Word embeddings (e.g., GloVe, Word2Vec) represent a foundational deep learning concept that transforms terms into dense vectors in a high-dimensional semantic space (Johnson, Murty, & Navakanth, 2024). These vectors capture statistical co-occurrence and contextual usage, facilitating both synonym detection and conceptual grouping (Johnson, Murty, & Navakanth, 2024). The ability to compute similarity via distance metrics (e.g., cosine similarity) supports downstream tasks such as clustering and classification of tags based on semantic proximity (Johnson, Murty, & Navakanth, 2024). For example, if an image is tagged with "beach," GloVe helps infer that related tags like "sand" and "ocean" might also be appropriate due to the close relationships these words share in the embedding space. This not only enhances tagging accuracy but also supports zero-shot learning, where the system can tag images with terms it hasn't seen during training. Additionally, such embeddings improve search functionality, enabling more intuitive retrieval of images by broader concepts or themes, not just specific tags (Poonkodi, Arunnehru, & Anand 2021). Word embedding models, such as GloVe, have been employed to examine semantic relationships within folksonomies (Pennington, Socher, & Manning, 2014; Johnson, Murty, & Navakanth, 2024). These models capture co-occurrence patterns in large corpora, allowing researchers to quantify the similarity between tags and uncover latent thematic structures in tagging behaviors (Johnson, Murty, & Navakanth, 2024, Wang and Yu, 2022).

Beyond embeddings, *autoencoders* play a key role in unsupervised learning by compressing input data into lower-dimensional latent representations and reconstructing them, thereby revealing essential structure (Taleb et al., 2021). This study employs autoencoders to uncover latent semantic dimensions within tag distributions, capturing both local coherence and broader thematic trends.

The integration of semantic analysis with faceted classification further enhances our understanding of how users construct meaning within collaborative tagging environments (Gugulica & Burghardt, 2023). Recent deep learning work increasingly uses Flickr as a testbed for semantic analysis, leveraging images and tags to learn multimodal representations and support clustering and cross-modal retrieval (Feng et al., 2025). Techniques such as hypergraph embeddings, adaptive clustering, and ensemble prototype networks on Flickr datasets reveal semantic similarity patterns across images and tags (Comalada, Acuma, and Garcia 2025; Sattari and Yazici 2025; Liu et al., 2025).

While folksonomies differ from traditional taxonomies in their bottom-up structure, empirical studies indicate that they exhibit implicit hierarchies and semantic relationships that can be leveraged for improved information organization (Quintarelli, Resmini & Rosati, 2007; Kokla, & Guilbert, 2020). By integrating faceted classification principles and deep learning techniques, researchers can enhance our understanding of tagging behaviors and optimize metadata systems for digital content management.

Together, these methods illustrate how deep learning bridges cognitive folksonomic behavior with computational taxonomies, offering scalable, interpretable, and richly semantic analyses of user-generated content.

Research Objectives

This study explores the integration of folksonomy metadata into traditional cataloging practices by analyzing the semantic relationships and descriptive quality of user-generated tags. Our research investigates whether identifiable patterns exist in tag relationships based on subject matter and examines the connection between folksonomic classification categories and structured facet-based schemes like Colon Classification. We include mapping popular tags from Flickr to the Colon Classification scheme for the years 2006, 2010, and 2015 to study evolving patterns and assess how well folksonomy aligns with traditional cataloging methods. Additionally, we conduct an embedding-based semantic analysis within a broader representation-learning paradigm widely used in contemporary NLP, exploring the semantic relationships and clustering patterns of these popular tags using pretrained GloVe embeddings. This approach captures relationships among tags without relying on predefined semantic categories, allowing us to observe the evolving and dynamic nature of user-generated metadata in social classification systems and to consider implications for enhancing metadata practices in digital archives.

Methods

To examine the classification schemes that influence image tagging behaviors on Flickr over significant time intervals, this study analyzed the application of facet classification to popular tags during the years 2006, 2010, and 2015. These intervals were chosen to reflect the evolving nature of social tagging amid advancements in digital technology and shifts in user behavior. We employed autoencoders and pretrained word embeddings due to their ability to uncover latent semantic structures and subtle temporal changes in large, unstructured datasets.

Data Collection

We specifically selected datasets from Flickr for the years 2006, 2010, and 2015 due to their significance in technological and cultural evolution. The year 2006 represents the early proliferation of social tagging platforms; 2010 aligns with the surge of smartphone adoption and the emergence of influential photo-sharing platforms such as Instagram; and by 2015, Flickr had reached a mature user base, allowing us to examine stable yet evolving tagging practices. For each benchmark year, we identified a set of popular tags by querying Flickr's public API for the most frequently used tags associated with public photographs, resulting in 141 tags for 2006, 136 for 2010, and 139 for 2015. We defined "popular tags" as those exceeding a minimum frequency threshold within the year and excluded obvious duplicates, trivial variants, or purely technical strings. When near-duplicate tags (e.g., singular/plural forms or minor spelling variations) appeared among the popular set, we consolidated them into a single canonical form after manual inspection; tags that could not be unambiguously normalized were retained as separate entries and documented in our coding notes.

Because our deep learning analysis relies on pretrained Word2Vec embeddings, we further required that each tag appear in the embedding model's vocabulary. For the 2015 benchmark year, this filtering step yielded 133 tags for inclusion in the autoencoder and clustering stages, as reported in Table 2. Tags not covered by the embedding vocabulary were excluded from the semantic modeling pipeline but were still included in the Colon faceted classification and descriptive analyses when appropriate. This two-step process, frequency-based selection followed by vocabulary filtering, ensured that the tags used in the semantic modeling were both representative of popular usage on Flickr and technically compatible with the autoencoder-based analysis.

Facet Classification and Temporal Analysis

In this study, each tag from Flickr was meticulously categorized according to the primary facets of the Colon Classification system: Personality, Matter, Energy, Space, and Time. Two researchers independently analyzed the popular tags to maintain objectivity. They regularly discussed their findings to ensure a unified understanding of the tagging trends on Flickr. Following the recommendations by Miles and Huberman for robust qualitative reliability, which suggests a consistency rate of at least 80% for coder agreement (Miles & Huberman, 1994), our inter-rater reliability was assessed by independently classifying 30 randomly selected popular tags. This process resulted in a high agreement rate of 90%, indicating substantial reliability in our classification approach.

We conducted a temporal analysis by examining tag clusters from the years 2006, 2010, and 2015 to trace the evolution of user tagging behavior on Flickr. This longitudinal comparison allowed us to identify persistent themes, detect emerging trends, and observe declining patterns. These insights provide a deeper understanding of the changing preferences and cultural shifts within the Flickr community over the selected time periods.

Semantic Embedding with Pretrained Model

In this study, we employed deep learning methods, specifically autoencoder neural networks, for capturing meaningful semantic representations of the tags collected from Flickr's dataset. An autoencoder is an unsupervised deep learning architecture designed to efficiently encode input data into a lower-dimensional latent space and then reconstruct the input from this latent representation. By doing so, it reveals essential underlying structures and semantic relationships among high-dimensional textual data (tags).

To model the latent semantic structure of tags, we first represented each tag using pretrained word embeddings trained on a large general-purpose corpus. Each tag was thus encoded as a fixed-length real-valued vector, which served as input to a feed-forward autoencoder. The autoencoder consisted of a symmetric encoder–decoder architecture with one hidden bottleneck layer that learned a low-dimensional representation of each tag vector. In our experiments, the input layer size matched the dimensionality of the pretrained embeddings, and the bottleneck layer reduced this dimensionality to a compact latent space designed to preserve the most salient semantic information.

To quantify semantic relationships among tags, we utilized pretrained Word2Vec skip-gram embeddings (Mikolov et al., 2013), which learn vector representations by optimizing the likelihood of context word prediction. As described in Bishop & Bishop (2024, Ch. 12.2), these embeddings map discrete words into continuous vector space where semantic similarity is preserved via spatial proximity.

Where the probability $p(w_{t+j}|w_i)$ is computed via the softmax function:

$$p(w_O|w_I) = \frac{\exp\,(V'_{WO} \cdot V_{WI})}{\Sigma_{\mathrm{i=1}}^{\mathrm{v}} \exp\,(V'_{WO} \cdot V_{WI})}$$

Where Vw and Vw' are the "input" and "output" vectors of word w and w', and V is the vocabulary size.

These embeddings serve as input to our autoencoder, which further encodes these vectors into a condensed latent representation. The autoencoder consists of multiple hidden layers designed to capture nonlinear relationships and semantic nuances within the tags, facilitating more meaningful clustering and analysis.

Representation Learning with Autoencoders

The embedded tag vectors were input into a deep autoencoder network to reduce dimensionality and capture latent semantic structure. An autoencoder consists of an encoder that maps input x to a latent representation z, and a decoder that reconstructs x from z. Training minimizes the reconstruction error:

$$\mathcal{L}_{\mathrm{AE}} = ||x - \hat{x}||^2$$

As emphasized in Mienye and Swart (2025), autoencoders are powerful for learning distributed, hierarchical representations from unlabeled data, ideal for unsupervised semantic analysis.

We trained the autoencoder to minimize reconstruction error between the input and output tag vectors using the Adam optimization algorithm with a mean-squared error loss function (Kingma & Ba, 2015; Vincent et al., 2008). The dataset of tag vectors for the focal year was randomly partitioned into training and validation subsets, with the validation set used to monitor model performance and guard against overfitting (Prechelt, 1997; Mahsereci et al., 2017). We employed early stopping based on validation loss, terminating training when no further improvement was observed for several consecutive epochs. To assess the stability of the learned representations, we repeated training with different random initializations and confirmed that the resulting latent-space clusters were qualitatively consistent across runs. The final latent vectors from the bottleneck layer were then used as input to the subsequent clustering and visualization steps described below.

Semantic Similarity and Clustering

Latent vectors generated from the autoencoder were evaluated using cosine similarity, a metric well-suited for high-dimensional semantic spaces. Semantic similarity among tags was quantified using the cosine similarity metric applied to the latent vectors generated by the autoencoder:

$$Cosine\ similarity\ (A, B) = \cos(\theta) = \frac{A \cdot B}{||A|| \times ||B||}$$

Where:

- A and B are vectors.
- A·B is the dot product of the vectors.
- ‖A‖ and ‖B‖ are the magnitudes (or norms) of the vectors.

This provided a pairwise similarity matrix for all tags, forming the basis for clustering. This measure evaluates the semantic relatedness of tags based on their encoded latent features, ensuring accurate grouping and retrieval in subsequent analyses.

Due to the high-dimensional nature of word embeddings, dimensionality reduction was essential. We utilized the t-distributed stochastic neighbor embedding (t-SNE) method to visually interpret and analyze tag relationships. The t-SNE algorithm preserves local structures and semantic relationships by projecting high-dimensional data onto a lower-dimensional space, enabling effective visualization.

To navigate the curse of dimensionality and to provide an intuitive visualization of the data structure, we employed the t-SNE algorithm (Wang, Huang, Rudin, & Shaposhnik, 2021). This non-linear dimensionality reduction technique is particularly well-suited for embedding spaces, preserving local structures and relationships between terms in a reduced-dimensional space.

Hierarchical clustering creates agglomerative clustering, is a "bottom-up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy (Sasirekha, & Baby, 2013). The linkage criterion determines the distance between sets of observations as a function of the pairwise distances between observations. One of the most popular methods is Ward's method (Ward, 1963), which minimizes the total within-cluster variance. At each step, the pair of clusters with the minimum between-cluster distance are merged. Following dimensionality reduction, hierarchical clustering (Ward’s method) was applied to systematically identify tag groups based on their semantic similarity. Ward’s method aims to minimize within-cluster variance.

The mathematical criterion for the Ward method is:

$$\Delta(\text{sum of squares}) = = \mathrm{SS}_{A\cup B} - (\mathrm{SS}_{A} + \mathrm{SS}_{B})$$

where $\mathrm{SS}_{\mathrm{AUB}}$ is the sum of squares within combined clusters *A* and *B,* and SS_{A}, and SS_{B} are sums of squares within individual clusters. This step revealed coherent thematic clusters reflecting the deep semantic structure of the data.

Together, these methods combine classical faceted classification with modern deep learning techniques to surface both interpretable and data-driven tag structures.

Result

Popular tags were categorized according to the primary facets of Colon Classification: Personality, Matter, Energy, Space, and Time. Users showed a marked preference for tags associated with Space and Personality, indicating an emphasis on location-based and personal expression content. Throughout the study period, there was a shift in tagging behavior, with a decline in the use of Time-related tags and an increase in tags linked to the personal and tangible aspects represented by the Personality and Matter facets. This evolution in tagging patterns suggests a trend toward more personalized and substantial content, moving away from ephemeral, time-specific details. Interpreted through an embedding-space lens, this pattern indicates that user-generated tags increasingly concentrate in semantic neighborhoods associated with place and self-expression rather than purely temporal markers.

Faceted Classification and Temporal Analysis

In Flickr, popular tags in 2006, 2010, 2015 were analyzed and well mapped into five fundamental facets in Colon Classification (Appendix 1, 2, 3). The distributions of the popular tags in the increasing order in these facets are: energy (6), time (18), matter (31), personality (31), space (56). The ~~exist~~ presence of high number of tags related to personality and space as popular tags reflects the nature of Flickr database that is a user center image management system.

To provide a comprehensive analysis of the evolving trends in the popular terms used in the Flickr database across the years 2006, 2010, and 2015, the result show stability in core facet themes and classifications. Many terms remain consistent across the years, indicating stable core themes in user behavior. For instance, terms like `animals`, `beach`, and `live` persist across all years (Appendix 1, 2, 3).

The comparison also found that new terms are becoming popular due to the emergence of technology & platforms. For example, the term `bw` (potentially indicating black & white photos) appears in 2010 and 2015 datasets, suggesting an increased interest or resurgence in black & white photography or possibly the usage of filters in platforms like Instagram (Appendix 1, 2, 3).

The absence of terms like `iphone` and `instagram` in the Flickr 2015 data may either indicate that they weren't top terms or that these datasets are a subset of larger datasets. There might have geographical shifts for people sharing images in Flickr database: While terms like `asia` and `australia` appear consistently, other terms such as `africa` (seen in 2006) are replaced by `amsterdam` in subsequent years. This may highlight evolving travel or photographic interests (Appendix 1, 2, 3).

Temporal trends in Flickr tagging can be observed through the consistent use of terms related to seasons and months such as "April," "autumn," and "Christmas." These terms highlight the role of time-bound events and seasons in influencing how users tag their photos. The evolution of technology, particularly the rise of smartphones and platforms like Instagram which emerged in 2010 with user-friendly photo filters, has also likely shaped tagging behaviors, possibly encouraging trends such as black-and-white photography. Additionally, cultural and world events, such as a city hosting the Olympics, can lead to spikes in mentions, reflecting global cultural dynamics or significant happenings that capture public interest. Seasonal and temporal markers remain significant, emphasizing the impact of events, seasons, and festivals on the photographic content shared by users. Over time, while core themes persist, subtle shifts in terms suggest an evolution in interests or the emergence of new photography subcultures, indicating a dynamic interplay between technological advancements and cultural shifts in the Flickr community.

Table 1 Summary of comparative Analysis of Flickr popular tags across 2006, 2010, and 2015

	2006	2006 (%)	2010	2010 (%)	2015	2015 (%)
Energy	5	3.55	2	1.47	3	2.16
Matter	32	22.70	35	25.74	39	28.06
Personality	31	21.99	40	29.41	41	29.50
Space	56	39.72	49	36.03	47	33.81
Time	17	12.06	10	7.35	9	6.47
Total	141	100	136	100	139	100

In Table 1, it summarized the raw counts and respective percentiles for each category across the years 2006, 2010, and 2015 indicated that the “energy” category sees a slight increase from 2010 to 2015. The “matter” category has an increasing trend over the years. The “personality” category remains relatively stable between 2010 and 2015, with a slight increase. The “space” category shows a minor decrease from 2010 to 2015. The “time” category decreases in its percentage representation from 2006 to 2015. This table provides a clearer year-over-year comparison, highlighting changes in the popular terms used in the Flickr database across the years.

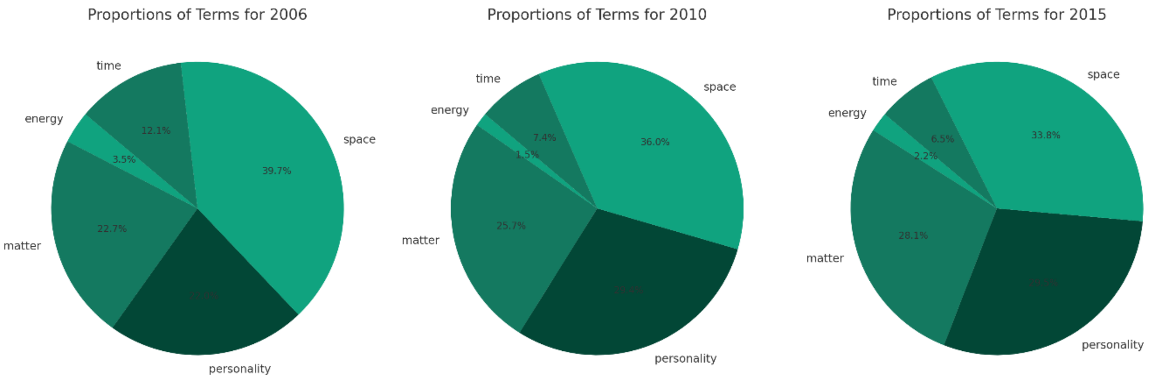

Figure 1. Facet classification proportions changes of popular tags in Flickr in 2006, 2010, and 2015

Figure 1 provides a comparative analysis of the proportions of terms for various categories across the years 2006, 2010, and 2015 on the Flickr platform. The pie charts further emphasize the balance shift in categories over the years. While the segments representing some categories remain relatively stable, others show a noticeable change in their proportions, underscoring the evolving nature of user tagging behavior. Figure 1 showed the tags classification in the popular tags in Flickr has fundamental facets: place, time, matter, and personality.

Autoencoder-Derived Semantic Structure

When comparing the facet classifications with the autoencoder model in Table 2, the following finding can be identified. Framed in embedding representation learning terms, the autoencoder serves as a nonlinear compressor over the tag embedding space, and the resulting clusters summarize dense semantic neighborhoods that can be compared to expert facets. Cluster 1 primarily overlaps with the `Personality` classification (61.11% or 22 terms). A quarter of the terms, 25% (9 terms), are associated with the `Time` classification. There's a minor overlap with `Matter` at 11.11% (4 terms). As for Cluster 2, it has a significant overlap with the `Matter` classification (55% or 22 terms). It also has a noticeable overlap with the `Space` classification (40% or 16 terms). There are minimal overlaps were found with `Energy` and `Personality` at 2.5% (1 term) for each. As for Cluster 3, it shows a dominant overlap with the `Personality` classification (61.9% or 13 terms). There's also a notable overlap with `Matter` at 28.57% (6 terms). It overlaps with `Energy` and `Space` are minimal at 4.76% (1 term) each. Regarding Cluster 4, it has a significant overlap with the `Space` classification (88.89% or 24 terms), with a minor overlap exists with `Personality` (11.11% or 3 terms). At last for Cluster 5, it demonstrates an equal overlap with `Matter` and `Space` classifications, both at 40% (4 terms each). `Energy` and `Personality` each have a 10% overlap (1 term each).

Table 2. Semantic Similarity Analyzed by Hierarchy Clustering for 133 Flickr popular tags in year 2015

Cluster	Classification	Percentage (%)	Raw Count
Cluster 1	Personality	61.11	22
Cluster 1	Time	25.00	9
Cluster 1	Matter	11.11	4
Cluster 2	Matter	55.00	22
Cluster 2	Space	40.00	16
Cluster 2	Energy	2.50	1
Cluster 2	Personality	2.50	1
Cluster 3	Personality	61.90	13
Cluster 3	Matter	28.57	6
Cluster 3	Energy	4.76	1
Cluster 3	Space	4.76	1
Cluster 4	Space	88.89	24
Cluster 4	Personality	11.11	3
Cluster 5	Matter	40.00	4
Cluster 5	Space	40.00	4
Cluster 5	Personality	10.00	1
Cluster 5	Energy	10.00	1
Total			133

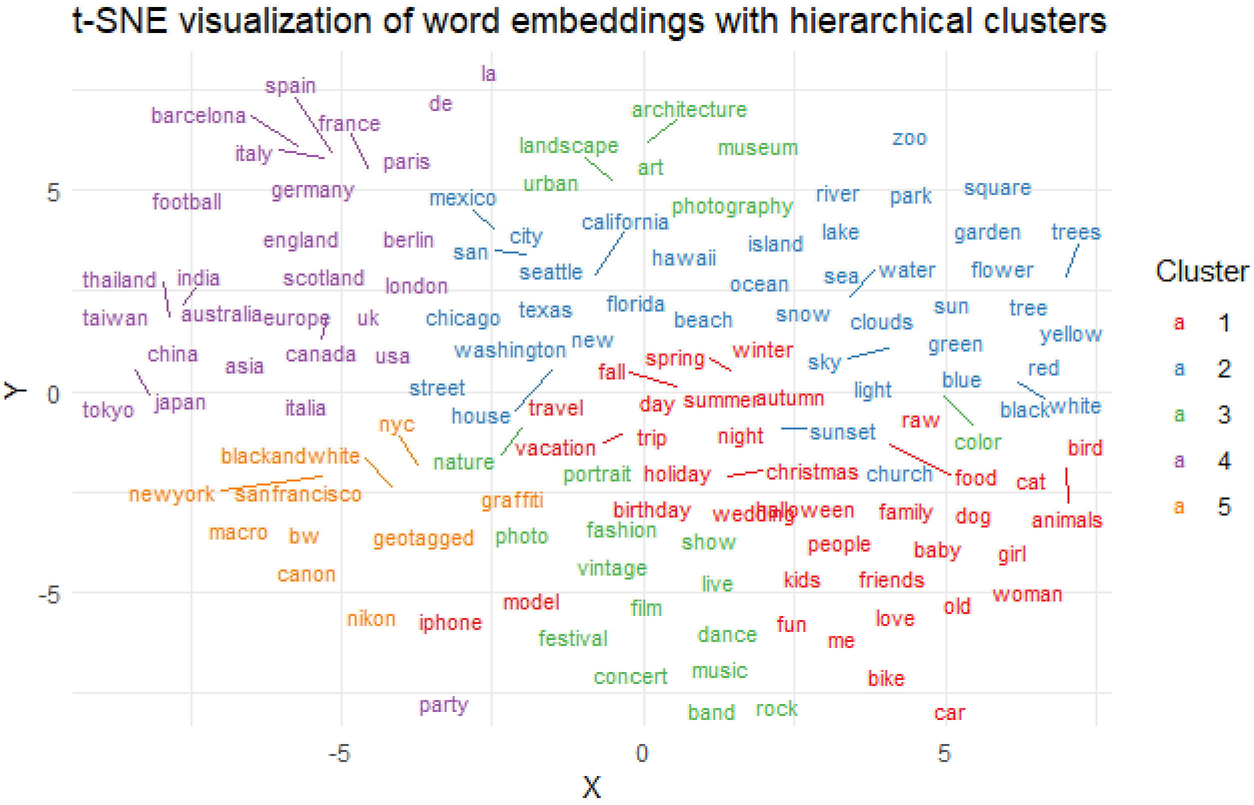

Figure 2. t-SNE visualization with hierarchical clustering of -tag embeddings for 2015 popular Flickr tags

The t-SNE visualization of word embeddings with hierarchical clustering (Figure 2) reveals five distinct clusters, each representing a coherent thematic concept among the 2015 popular Flickr tags. One prominent cluster consists of geographical locations, including names of countries and cities such as Spain, France, Germany, Tokyo, Canada, USA, Scotland, and London, indicating a strong association with travel and place-oriented content. A second cluster captures words related to nature and urban landscapes, featuring terms like photography, museum, art, architecture, zoo, landscape, river, lake, ocean, and flower, which suggest themes of scenic beauty, cultural spaces, and outdoor photography. A third cluster focuses on seasons, travel, and events, containing words such as beach, snow, ocean, sea, winter, summer, fall, travel, vacation, portrait, and holiday, highlighting seasonal activities and tourism-related topics. Another cluster encompasses technology and photographic practice, where terms like blackandwhite, New York, San Francisco, macro, bw, Canon, Nikon, and iPhone point to discussions of photographic techniques, camera brands, and iconic shooting locations. Finally, a lifestyle and emotions cluster includes words such as model, birthday, wedding, fashion, music, concert, love, fun, old, woman, car, dog, food, cat, family, baby, and friends, reflecting themes of personal relationships, celebrations, and everyday life.

These clusters align in meaningful ways with the Colon facets summarized in Table 2. The geographic cluster is dominated by tags coded as Space, illustrating how location-oriented annotations form a dense semantic region in the embedding-derived latent space (corresponding to the high proportion of Space terms in Cluster 4). The lifestyle and emotions cluster is largely associated with the Personality facet, capturing self-expression and social roles (reflected in the strong Personality presence in Clusters 1 and 3). The technology/photography and nature/urban clusters blend Matter and Space, consistent with Clusters 2 and 5, where physical objects (e.g., cameras, animals, built structures) co-occur with place-related tags. Seasonal and event-related terms contribute Time facets within otherwise Personality- or Space-dominant clusters, indicating that temporality is often embedded within broader thematic contexts rather than forming an isolated group. Taken together, Figure 2 and Table 2 show that the autoencoder-derived semantic clusters and the Colon facet assignments converge on a small set of stable cognitive dimensions: Personality, Matter, Space, Time, and Energy, while also revealing how user-generated tags frequently mix multiple facets within a single semantic neighborhood.

The visualization effectively demonstrates how semantically related words are grouped, providing valuable insights for natural language processing (NLP), image tagging, and search optimization. The clustering method successfully organizes words into meaningful categories, making it useful for applications in digital content organization, social media analytics, and contextual search enhancements.

Discussion

The analysis of Flickr's popular tags over multiple years reveals consistent thematic patterns while also showcasing shifts in user behavior, influenced by technological advancements and cultural changes (Van Dijck, 2011; Khusro, Jabeen & Khan, 2021). The stability of certain core facets, such as Matter and Personality, suggests an inherent cognitive structure in how users assign tags, supporting the idea that folksonomies, despite their decentralized nature, exhibit underlying hierarchical tendencies (Holstrom & Tennis, 2020).

The classification of popular Flickr tags into Colon Classification facets—Personality, Matter, Energy, Space, and Time—revealed that users predominantly favored tags related to *Space* and *Personality*. These facets, which reflect geographical locations and individual expressions, accounted for the majority of high-frequency terms across all three time slices (2006, 2010, and 2015). Tags associated with *Time* and *Energy* were relatively sparse, suggesting a user preference toward more spatially and personally grounded annotations rather than action-based or temporal contexts.

This distribution aligns with the principle of data manifolds in deep learning, where high-dimensional data often lies on a lower-dimensional, semantically structured manifold. In this study, the manifold appears to reflect socially salient dimensions such as *place* and *identity*, suggesting that Flickr users organize content along stable cognitive axes. Prior work has shown that deep neural networks, especially autoencoders and embedding models, learn such low-dimensional manifolds to capture essential semantic structure (Bengio et al., 2013; Goodfellow et al., 2016). These learned manifolds support disentangled representations that are robust to variation and encode meaningful relationships within the data.

Temporal Evolution and Cultural Shifts

Our findings indicate a clear transition from general, descriptive tags toward highly personalized and identity-expressive annotations over the studied years. This shift is consistent with broader cultural movements towards individualism and self-expression, driven by widespread smartphone usage and the popularity of platforms emphasizing personal engagement. For example, increased occurrences of tags such as 'selfie', 'fun', and 'love' directly reflect evolving user behaviors shaped by technological advancements and emerging social norms. Most of the tags are perceptual properties (objects, people, color, visual elements, location, and description/number); and less in interpretive properties (e.g., people relationships) and viewer reaction property (e.g., conjecture).

Interestingly, it also found that the Shift in Platform Usage: The decrease in the "Place" category might be indicative of the emergence and popularity of other platforms like Instagram, where location-based tagging became a prominent feature. As users shifted some of their photo-sharing activities to these platforms, Flickr might have seen a slight decline in location-based tags. The research also found the Interplay with Global Events: Variations in certain categories might also reflect global events, trends, or popular culture influences during those years. For instance, significant global events might lead to a surge in "Time"-related tags as users capture and share momentous occasions.

The growth in “Personality” and “Time” tags might also reflect broader social media trends, where personal expression and capturing fleeting moments became paramount. The rise in "Personality" tags might also be indicative of a more connected global community. As users became more exposed to diverse cultures and lifestyles, their tagging behavior evolved to encapsulate more emotions, feelings, and personal experiences.

The stability in the "Matter" category might suggest that while tagging behaviors evolve, certain core categorizations based on objects, or the content of the photo remain consistent. This could reflect the inherent human behavior to classify based on tangible attributes. The variations in categories over time might also hint at the platform's evolving user base. As Flickr matured, it might have attracted a more diverse user base, leading to shifts in tagging patterns.

A key observation in this study is the decreasing reliance on Time-related tags and the increasing use of Personality-related terms. This shift may be attributed to the growing prevalence of personal expression in online platforms, where users focus more on self-representation rather than chronological documentation (Van Dijck, 2011). The rise of social media platforms, which emphasize ephemeral content and personal engagement, could have influenced this trend, leading to a reduction in time-based classifications on Flickr (Khusro, Jabeem, & Khan, 2021).

Autoencoder-Derived Semantic Structures in Social Tagging

The autoencoder model generated low-dimensional latent vectors that revealed semantically coherent clusters among tags. Clustering via Ward's method (Ward, 1963) identified tight groupings of conceptually related terms (e.g., *"beach," "ocean," "surf"*), supporting the model's ability to capture local semantic continuity. The autoencoder can be viewed as a downstream structure-inducing compressor over an embedding space. It concentrates semantic geometry into a compact latent representation that is more amenable to clustering and longitudinal comparison.

Users tend to tag images of travel and place with country and city names; they describe scenery and urban environments with landscape and architecture terms; they mark seasons, holidays, and vacations with temporal and event-related vocabulary; and they highlight equipment and technique with brand names and photographic jargon (Figure2). These patterns suggest that the latent space learned by the autoencoder is not an abstract mathematical artifact but a compact reflection of how people habitually organize their visual experiences in language. This finding is consistent with recent advances in autoencoder-based representation learning, where deep networks are used to uncover latent semantic structure in high-dimensional data without supervision. Studies have shown that such models, including variational and denoising autoencoders, are effective in capturing interpretable features and preserving semantic relationships for tasks such as clustering, retrieval, and visualization (Taleb et al., 2021; Ye et al., 2022).

The presence of clusters representing geographical locations, nature, travel, photography, and lifestyle confirms that social tagging behaviors are not entirely chaotic but follow discernible semantic groupings. This suggests that folksonomies, while dynamic, maintain structured relationships that can be leveraged for improved information retrieval and metadata organization (Golder and Huberman, 2006; Das et al., 2022). When these clusters are read alongside the Colon facets in Table 2, a more refined picture emerges. The geographic cluster is dominated by tags coded as Space, while the lifestyle and emotion cluster is rich in Personality terms related to social roles, relationships, and affect. Technology- and nature-oriented clusters blend Matter and Space, reflecting the co-occurrence of physical objects, environments, and locations in everyday tagging practices. Seasonal and event-related tags introduce Time into these neighborhoods rather than forming a separate temporal cluster, indicating that temporality is typically embedded within broader narratives of travel, celebration, and daily life. In this sense, the autoencoder-derived structure recovers the same core cognitive dimensions as the Colon scheme: Personality, Matter, Space, Time, and Energy, while also revealing how users routinely mix facets within a single semantic neighborhood.

The analysis of Flickr's popular tags using semantic classification with word vectors and deep learning clusters revealed a complex landscape of tag categorization. The clusters exhibited diverse overlaps with multiple facets of classification, suggesting that while the clusters are distinct, they do not adhere strictly to a single facet. This was particularly evident in how some clusters showed dominant overlaps with specific classifications. For example, in Table 2, Cluster 4 was predominantly composed of terms related to Space, and Clusters 1 and 3 had a considerable number of terms associated with Personality.

Conversely, Clusters 2 and 5 presented a more mixed composition, showing significant overlaps with multiple classifications like Matter and Space. This mixed nature might indicate an inter-related theme within these clusters, highlighting the complexity of semantic relationships among tags. By aligning the clusters with established classifications, one can assess the relevance and coherence of each cluster to known themes, providing a foundation for more targeted investigations.

These observations have several implications for using deep learning to reveal latent structure in folksonomies. First, the close alignment between intuitive user clusters and Colon facets supports the idea that a faceted classification can serve as an interpretable “overlay” on top of data-driven representations. Deep models can thus uncover fine-grained neighborhoods and transitions within and between facets, while the faceted scheme offers a stable conceptual frame for interpreting those neighborhoods. Second, the presence of mixed-facet clusters underscores the importance of designing tagging, browsing, and recommendation interfaces that accommodate multi-faceted concepts rather than forcing tags into rigid single-category slots. In practice, this might mean supporting hybrid facet combinations such as “Personality and Time” (e.g., weddings in summer) or “Matter and Space” (e.g., architecture in specific cities) when organizing or retrieving social image content.

Furthermore, the comparison between deep learning-based clustering and manual facet classification reveals areas of alignment and divergence. While autoencoders and related representation-learning techniques successfully capture latent semantic structures and relationships, manual classifications provide insights based on human interpretation. The observed overlaps indicate that automated classification models in deep learning can be effectively integrated with traditional faceted classification frameworks to enhance metadata systems (Alijani, Tanha, & Mohammadkhanli, 2020; Asadi, 2021; Yu & Chen, 2020).

Finally, the convergence between autoencoder-based clusters and established patterns in social tagging research suggests that deep learning offers a powerful complement, not a replacement, to traditional knowledge organization approaches. Prior work has shown that folksonomies exhibit stable regularities such as power-law distributions and emergent semantic groupings; our findings illustrate that these regularities can be captured and refined in a latent space that is well suited to downstream tasks such as visualization, recommendation, and query expansion. Because this argument is grounded in latent-space geometry, it provides a direct path to future extensions using Transformer/LLM embeddings without changing the interpretive role of Colon facets. By aligning latent clusters with faceted categories, we provide a pathway for integrating user-generated tags, deep learning models, and expert-designed classification systems into a coherent framework for organizing and exploring large-scale social image collections.

These findings align with prior research indicating that social tagging systems are influenced by both cognitive and social factors (Iyer, Main, and Verlag 1995; Stuart, 2019). The evolving use of tags reflects broader shifts in technological adoption, cultural trends, and platform-specific affordances. The presence of stable classification facets suggests that folksonomies possess an inherent navigability that can be harnessed for structured information organization (Golder & Huberman, 2006; Irwing, Hufhes, Tokarev & Booth, 2024). As digital environments continue to evolve, understanding these tagging behaviors can provide valuable insights for improving content categorization in digital libraries, archives, and social media platforms (Van Dijck, 2011; Cox, Clough & Marlow, 2008).

Conclusion and Future Work

These results highlight the potential for representation-learning-enhanced folksonomies to approximate structured taxonomies. Although folksonomies are user-generated and non-hierarchical, our analysis revealed latent hierarchical and semantic patterns consistent with formal classification systems. This suggests that neural embedding models (including Transformer-era encoders) can bridge the gap between informal metadata and structured information retrieval frameworks.

Ultimately, this study demonstrates how neural representation learning in embedding spaces can enrich traditional metadata analysis by surfacing latent dimensions of meaning, enabling more flexible, adaptive, and cognitively aligned systems for digital archiving and content management.

The Flickr database's popular image tags provide a lens into the collective interests, activities, and priorities of its user base. As technology evolves, and as global events shape our lives, the way users tag and categorize their memories on platforms like Flickr will continue to evolve. Monitoring these trends offers valuable insights into broader shifts in society, culture, and technology. To provide more specific insights, a deeper dive into the complete dataset and potentially a more granular categorization of terms might be beneficial.

It is logical to analyze the classification scheme based on the tag distribution pattern. Faceted schemes easily accommodate new subjects by combining the concepts within facets in new ways, instead of having to fit new subjects into the existing organization of a list. If a faceted scheme is designed properly, it should be able to represent any possible subject within the sphere of human capabilities. Notation in faceted schemes tends to be relatively complex, due to their ability to synthesize facets; and their use of facet connecting codes, however it turns out to be more fit the dynamic changing in the social tagging system rather than those rigid non-facet schemes.

Facet classification provides a clue between the bottom-up tagging and the top-down controlled vocabularies(Quintarelli, Resmini & Rosati, 2007).

Tag clouds are visual interfaces for information retrieval that provide a global contextual view of tags assigned to resources in the system (Khusro, Jabeen, and Khan 2021). However, the tag clouds are flat and not enough to provide a semantic and multidimensional browsing experience (Vu, 2020). Researchers don't just look at the photos in isolation but considers both their visual elements and the surrounding information to categorize and understand the image with multifaceted relationships that exist in Flickr, encompassing various elements such as visual content, user interactions, tags, and time, all contributing to the holistic social media experience (Lin et al., 2012).

This study reveals that despite the inherently decentralized nature of folksonomies, user-generated tags on Flickr exhibit stable, implicit hierarchical structures that mirror traditional classification schemes. Over the observed period,from 2006 to 2015,core facets such as Matter and Personality consistently dominate, while a gradual decline in Time-related tags reflects shifting user behaviors influenced by evolving technologies and cultural trends. The integration of machine learning techniques, such as ~~word embeddings~~ pretrained embedding models and hierarchical clustering, has successfully unearthed latent semantic relationships among tags. These data-driven insights closely align with manual facet classifications, demonstrating that automated methods can effectively complement traditional approaches to metadata organization and enhance information retrieval in digital archives.

This study has several limitations that should be acknowledged. First, our analysis is restricted to popular tags from three benchmark years on a single social image platform, Flickr. As a result, the findings may not fully generalize to other platforms, domains, or to the long tail of less frequent tags that also contribute to the richness of folksonomies. Second, we focus exclusively on textual tags and do not incorporate visual features or other contextual signals from the underlying images; integrating multimodal information might reveal additional semantic patterns that are not captured by tag text alone. Third, we rely on a single pretrained embedding model and a specific autoencoder configuration, which, although sufficient for our purposes, represent only one of many possible deep learning setups. Future research could extend this work by examining other social media environments, by including image-based or multimodal representations, by exploring alternative architectures (such as transformer-based encoders and LLM embedding models), and by conducting multilingual or cross-cultural comparisons of faceted structures in social tagging systems.

From a practical standpoint, the alignment between autoencoder-derived clusters and Colon facets suggests several concrete applications for digital library and discovery systems. Facet–cluster correspondence can be used to refine faceted browsing interfaces by grounding facet labels and combinations in patterns that emerge from user tagging, thereby improving the intelligibility and usefulness of facet menus for end users. Deep learning derived clusters also offer a data-driven basis for tag recommendation and query expansion: systems can suggest additional tags or subject headings that are semantically proximate in the latent space, supporting richer description and more effective retrieval. In metadata workflows, librarians and system designers can use these clusters as "early warning signals" of emerging themes to monitor changes in cluster composition over time to identify candidate concepts for inclusion, subdivision, or revision in controlled vocabularies and subject heading lists.

Reference

Alemu, G. (2018). Metadata enrichment for digital heritage: users as co-creators. *International Information & Library Review*, *50*(2), 142-156.

Alijani, S., Tanha, J., & Mohammadkhanli, L. (2022). An ensemble of deep learning algorithms for popularity prediction of flickr images. *Multimedia Tools and Applications*, *81*(3), 3253-3274.

Allam, H., Bliemel, M., Al Amir, O., Toze, S., Shah, K., & Shoib, E. (2020, November). Collaborative ontologies in social tagging tools: a literature review of natural folksonomy. In *2020 Seventh International Conference on Information Technology Trends (ITT)* (pp. 126-130). IEEE.

Asadi, B. (2021). *Comparison of social tags and controlled vocabulary terms assigned to images: A feasibility study of computer-assisted image indexing*. McGill University (Canada).

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828. [DOI: 10.1109/TPAMI.2013.50]

Bhoir, S., Ghorpade, T., & Mane, V. (2017, December). Comparative analysis of different word embedding models. In *2017 International conference on advances in computing, communication and Control (ICAC3)* (pp. 1-4). IEEE.

Capocci, A., Baldassarri, A., Servedio, V. D., & Loreto, V. (2010, June). Friendship, collaboration and semantics in Flickr: from social interaction to semantic similarity. In *Proceedings of the International Workshop on Modeling Social Media* (pp. 1-4).

Centelles, M., & Ferrer, N. F. (2024). Taxonomies and ontologies in Wikipedia and Wikidata: an in-depth examination of knowledge organization systems. Hipertext. net, (28), 33-48.

Cho, H., Disher, T., Lee, W. C., Keating, S. A., & Lee, J. H. (2018). Facet analysis of anime genres: The challenges of defining genre information for popular cultural objects. KO Knowledge Organization, 45(6), 484-499.

Comalada, F., Acuña, V., & Garcia, X. (2025). Modelling cultural ecosystem services of river landscapes in the Iberian Peninsula with deep learning and social media images. *Journal of Environmental Management*, *394*, 127667.

Cox, A. M., Clough, P. D., & Marlow, J. (2008). Flickr: a first look at user behaviour in the context of photography as serious leisure. *Information* *research*, *13*(1), 13-1.

Das, P., Guda, B. P. R., Seelaboyina, S. B., Sarkar, S., & Mukherjee, A. (2022). Quality change: Norm or exception? Measurement, analysis and detection of quality change in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1-36.

De Salve, A., Guidi, B., & Michienzi, A. (2021). Exploiting homophily to characterize communities in online social networks. *Concurrency and Computation: Practice and Experience*, *33*(8), e5929.

Denton W. 2003. How to Make a Faceted Classification and Put It On the Web. retrieved from http://www.miskatonic.org/library/facet-web-howto.htm

Desrochers, N., Laplante, A., Martin, K., Quan-Haase, A., & Spiteri, L. (2016). Illusions of a "Bond": tagging cultural products across online platforms. *Journal of Documentation*, *72*(6), 1027-1051.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).

Elekes, Á., Schäler, M., & Böhm, K. (2017, June). On the various semantics of similarity in word embedding models. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 1-10). IEEE.

Fang, Z., Zou, Y., Lan, S., Du, S., Tan, Y., & Wang, S. (2025). Scalable multi-modal representation learning networks. *Artificial Intelligence Review*, *58*(7), 209.

Golder, S., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198–208.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Pres.

Gosal, A. S., Geijzendorffer, I. R., Václavík, T., Poulin, B., & Ziv, G. (2019). Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosystem Services*, *38*, 100958.

Green, R. 2006. Vocabulary Alignment via Basic Level Concepts. OCLC/ALISE research grant report published electronically by OCLC Research. Available online at: http://www.clc.org/research/grants/reports/green/rg2005.pdf

Gugulica, M., & Burghardt, D. (2023). Mapping indicators of cultural ecosystem services use in urban green spaces based on text classification of geosocial media data. *Ecosystem Services*, *60*, 101508.

Hackett, P. M., & Fisher, Y. (Eds.). (2019). *Advances in Facet Theory Research: Developments in Theory, Application and Related Approaches*. Frontiers Media SA.

Holstrom, C., & Tennis, J. T. (2020, November). Visibility, Identity, and Personal Expression: Qualitative Case Studies of Social Tagging on MetaFilter. In *Knowledge Organization at the Interface* (pp. 207-216). Ergon-Verlag.

Huang, H., & Jörgensen, C. (2013). Characterizing user tagging and Co-occurring metadata in general and specialized metadata collections. *Journal of the American Society for Information Science and Technology*, *64*(9), 1878-1889.

Irwing, P., Hughes, D. J., Tokarev, A., & Booth, T. (2024). Towards a taxonomy of personality facets. *European Journal of Personality*, *38*(3), 494-515.

Iyer H., Main F., and Verlag I., 1995. Classificatory Structures: Concepts, Relations, and Representations.

Jansson, I. M. (2017). Organization of user-generated information in image collections and impact of rhetorical mechanisms. *KO Knowledge Organization*, *44*(7), 515-528.

Johnson, S. J., Murty, M. R., & Navakanth, I. (2024). A detailed review on word embedding techniques with emphasis on word2vec. *Multimedia Tools and Applications*, *83*(13), 37979-38007.

Khusro, S., Jabeen, F., & Khan, A. (2021). Tag clouds: past, present and future. *Proceedings of the national academy of sciences, India section A: physical sciences*, *91*(2), 369-381.

Kingma, D. P. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kokla, M., & Guilbert, E. (2020). A review of geospatial semantic information modeling and elicitation approaches. *ISPRS International Journal of Geo-Information*, *9*(3), 146.

Kordumova, S., van Gemert, J., & Snoek, C. G. (2016). Exploring the long tail of social media tags. In *MultiMedia Modeling: 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016, Proceedings, Part I 22* (pp. 51-62). Springer International

Publishing.

Li, X., Uricchio, T., Ballan, L., Bertini, M., Snoek, C. G., & Bimbo, A. D. (2016). Socializing the semantic gap: A comparative survey on image tag assignment, refinement, and retrieval. *ACM Computing Surveys (CSUR)*, *49*(1), 1-39.

Liu, X., Zeng, H., Shi, Y., Zhu, J., Yang, K., & Yu, Z. (2025). Ensemble Prototype Networks for Unsupervised Cross-modal Hashing with Cross-Task Consistency. *IEEE Transactions on Multimedia*.

Luo, C., Zhan, J., Xue, X., Wang, L., Ren, R., & Yang, Q. (2018). Cosine normalization: Using cosine similarity instead of dot product in neural networks. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part I 27* (pp. 382-391). Springer International Publishing.

Mahsereci, M., Balles, L., Lassner, C., & Hennig, P. (2017). Early stopping without a validation set. *arXiv preprint arXiv:1703.09580*.

Mienye, I. D., & Swart, T. G. (2025). Deep autoencoder neural networks: A comprehensive review and new perspectives. *Archives of Computational Methods in Engineering*. Advance online publication. https://doi.org/10.1007/s11831-025-10260-5

Mikolov, T., Sutskever, I., & Chen, K. (2013). Greg S Corrado και Jeff Dean. *Distributed representations of words and phrases and their compositionality. Στο Advances in neural information processing systems*, 3111-3119.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. sage.

Nduka, S. C. (2022). *Awareness, accessibility, use of social media and social competence of library and information science postgraduate students in universities in Nigeria* (doctoral dissertation).

Niebler, T., Hahn, L., & Hotho, A. (2017). Learning Word Embeddings from Tagging Data: A methodological comparison. In *LWDA* (p. 229).

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Poonkodi, M., Arunnehru, J., & Anand, K. S. (2021). Transfer Learning-Based Image Tagging Using Word Embedding Technique for Image Retrieval Applications. In *Soft Computing and its Engineering Applications: Second International Conference, icSoftComp 2020, Changa, Anand, India, December 11–12, 2020, Proceedings 2* (pp. 157-168). Springer Singapore.

Prechelt, L. (2002). Early stopping-but when?. In *Neural Networks: Tricks of the trade* (pp. 55-69). Berlin, Heidelberg: Springer Berlin Heidelberg.

Qassimi, S., & Abdelwahed, E. H. (2019). The role of collaborative tagging and ontologies in emerging semantic of web resources. *Computing*, *101*(10), 1489-1511.

Quintarelli, E., Resmini, A., & Rosati, L. (2007). 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*, *33*(5), 10-15.

Rahman, A. I. M. (2012). *Social tagging versus Expert created subject headings* (Doctoral dissertation, Oslo and Akershus University College of Applied Sciences, Oslo, Norway).

Ranganathan S. R. 1962. Elements of library classification. Asia Publishing House, Bom- bay, p 45-70.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Rorissa, A. (2010). A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology*, *61*(11), 2230-2242.

Sasirekha, K., & Baby, P. (2013). Agglomerative hierarchical clustering algorithm-a. *International Journal of Scientific and Research Publications*, *83*(3), 83.

Sattari, S., & Yazici, A. (2025). Semantic deep learning and adaptive clustering for handling multimodal multimedia information retrieval. *Multimedia Tools and Applications*, *84*(13), 11795-11831.

Shirky, C. 2005. Clay Shirky's Writings About the Internet. Ontology is Overrated:

Categories, Links, and Tags: http://shirky.com/writings/ontology_overrated.html Accessed 02/01/2025

Sit, M. A., Koylu, C., & Demir, I. (2020). Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. In *Social Sensing and Big Data Computing for Disaster Management* (pp. 8-32). Routledge.

Sperberg-McQueen, C. M. (2015). Classification and its Structures. *A new companion to digital humanities*, 377-393.

Stuart, E. (2019). Flickr: Organizing and tagging images online. *KO Knowledge Organization*, *46*(3), 223-235.

Sundararaj, V., & Rejeesh, M. R. (2021). A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites. *Journal of Retailing and Consumer Services*, *58*, 102190.

Taleb, A., Dutta, A., Montavon, G., & Samek, W. (2021). Multimodal explainable AI for brain age prediction using denoising autoencoders. *Neurocomputing*, 453, 145–154. https://doi.org/10.1016/j.neucom.2021.03.047

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*(11), 2579–2605.

Van Dijck, J. (2011). Flickr and the culture of connectivity: Sharing views, experiences, memories. *Memory Studies*, *4*(4), 401-415.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008, July). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning* (pp. 1096-1103).

Vu, B. (2020). *A Taxonomy Management System Supporting Crowd-based Taxonomy Generation, Evolution, and Management* (Doctoral dissertation, University of Hagen, Germany).

Wang, H., Shi, X., & Yeung, D. Y. (2015, February). Relational stacked denoising autoencoder for tag recommendation. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 29, No. 1).

Wang, J. and Yu, H. (2022). Measure-Theoretic Probability of Complex Co-occurrence and E-Integral. arXiv:2210.09913 [stat.ML] https://doi.org/10.48550/arXiv.2210.09913

Wang, Y., Huang, H., Rudin, C., & Shaposhnik, Y. (2021). Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMAP, and PaCMAP for data visualization. *Journal of Machine Learning Research*, *22*(201), 1-73.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244.

Worth, P. J. (2023). Word embeddings and semantic spaces in natural language processing. *International journal of intelligence science*, *13*(1), 1-21.

Ye, X., Qiu, Y., Zhang, J., & Song, G. (2022). A survey of deep learning methods for unsupervised and self-supervised representation learning. *Journal of Computer Science and Technology*, 37(4), 759–790. https://doi.org/10.1007/s11390-022-1835-5

Yu, W., & Chen, J. (2020). Enriching the library subject headings with folksonomy. *The Electronic Library*, *38*(2), 297-315.

Zappavigna, M. (2018). *Searchable talk: Hashtags and social media metadiscourse*. Bloomsbury Publishing.

Zhang, H., Shang, X., Luan, H., Wang, M., & Chua, T. S. (2016). Learning from collective intelligence: Feature learning using social images and tags. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, *13*(1), 1-23.

Appendix 1 Facet classification in 2006 Flickr popular tags

Personality	Matter	Energy	Space	Time
animals	beach	***camping***	***africa***	***april***
architecture	black	***hiking***	***amsterdam***	***august***
art	blackandwhite	live	australia	autumn
baby	blue	macro	barcelona	christmas
birthday	bw	***sunset***	berlin	day
cat	***cameraphone***		***boston***	fall
concert	canon		california	halloween
dog	car		canada	july
family	clouds		chicago	***june***
festival	color		china	***may***
friends	film		church	night
fun	flower		city	***O6***
girl	food		***dc***	***october***
graffiti	green		england	***september***
holiday	island		europe	spring
home	lake		florida	summer
honeymoon	light		france	winter
kids	mountain		garden	
landscape	nature		geotagged	
me	nikon		germany	
music	ocean		hawaii	
new	red		***hongkong***	
party	river		house	

people	rock		india
portrait	sea		ireland
roadtrip	sky		***italy***
show	snow		japan
travel	sun		london
trip	tree		mexico
vacation	water		museum
wedding	white		newyork
	yellow		newyorkcity
			newzealand
			nyc
			paris
			park
			rome
			san
			sanfrancisco
			scotland
			seattle
			spain
			street
			sydney
			taiwan
			texas
			thailand
			tokyo
			toronto
			uk
			urban
			usa
			vancouver
			washington

			york	
			zoo	

Note. Bold/Italics: Tags that dropped off (2006) to the most popular tags in Flickr.

Appendix 2. Facet classification in 2010 Flickr popular tags

Personality	**Matter**	**Energy**	**Space**	**Time**
animals	beach	live	***asia***	autumn
architecture	black	macro	australia	christmas
art	blackandwhite		barcelona	day
baby	blue		berlin	fall
band	bw		california	halloween
bike	canon		canada	**july**
bird	car		chicago	night
birthday	clouds		china	spring
cat	color		church	summer
concert	film		city	winter
dance	flower		de	
dog	food		england	
family	green		europe	
fashion	***iphone***		florida	
festival	island		france	
football	lake		garden	
friends	light		geotagged	
fun	**mountain**		germany	
girl	nature		hawaii	
graffiti	nikon		house	
holiday	ocean		india	
kids	***photo***		**ireland**	
landscape	***photography***		***italia***	
love	***raw***		japan	

me	red	london
model	river	mexico
music	rock	museum
new	sea	newyork
old	sky	newyorkcity
party	snow	nyc
people	sun	paris
portrait	***tree***	park
show	water	san
tour	white	sanfrancisco
travel	yellow	scotland
trip		seattle
urban		spain
vacation		street
vintage		taiwan
wedding		***taly***
		texas
		thailand
		tokyo
		toronto
		uk
		usa
		washington
		york
		zoo

Note. Bold/Italics: Tags that were added (2010) to the popular tags in Flickr. Bold/Underline: Tags that were dropped off (2015) to the popular tags in Flickr. Bold/Italics/Underline: Tags that were added (2015) to the popular tags.

Appendix 3. Facet classification in 2015 Flickr popular tags

Personality	**Matter**	**Energy**	**Space**	**Time**
animals	beach	live	asia	autumn
architecture	black	macro	australia	christmas
art	blackandwhite	***sunset***	barcelona	day
baby	blue		berlin	fall
band	bw		california	halloween
bike	canon		canada	night
bird	car		chicago	spring
birthday	clouds		china	summer
cat	color		church	winter
concert	film		city	
dance	flower		de	
dog	food		england	
family	green		europe	
fashion	***instagramapp***		florida	
festival	iphone		france	
football	***iphoneography***		garden	
friends	island		geotagged	
fun	lake		germany	
girl	light		hawaii	
graffiti	nature		house	
holiday	nikon		india	
kids	ocean		italia	
la	photo		***italy***	
landscape	photography		japan	
love	raw		london	
me	red		mexico	
model	river		museum	
music	rock		newyork	
new	sea		newyorkcity	
old	sky		nyc	
party	snow		paris	
people	***square***		park	

portrait	***squareformat***	san
show	sun	sanfrancisco
travel	tree	scotland
trip	trees	seattle
urban	water	spain
vacation	white	street
vintage	yellow	taiwan
wedding		texas
woman		thailand
		tokyo
		uk
		unitedstates
		usa
		washington
		zoo

Note. Bold/Italics: Tags that were added (2015) to the popular tags in Flickr.