# 6
# Tag-Based Navigation and Visualization

Dimitar Dimitrov[1(✉)], Denis Helic[2], and Markus Strohmaier[3]

[1] GESIS – Leibniz Institute for the Social Sciences,
University of Koblenz-Landau, Cologne, Germany
`dimitar.dimitrov@gesis.org`
[2] Graz University of Technology, Graz, Austria
`dhelic@tugraz.at`
[3] GESIS – Leibniz Institute for the Social Sciences,
RWTH Aachen University, Cologne, Germany
`markus.strohmaier@gesis.org`

## 1 Introduction

Online information seeking has become an everyday task in lives of modern people. In principle, we distinguish between two strategies to explore and discover information spaces: search and navigation. Search implies a query formulation, whereas navigation is the process of finding a way to a given target by following hyperlinks. Navigation without a specific target is also referred to as browsing.

One of the main advantages of navigation as compared to search, which is tightly related to our cognitive abilities as humans—is that recognizing what we are looking for is much easier than formulating and describing our information need in a couple of keywords [34]. In literature, the formulation of an information need is also referred to as the *vocabulary problem* [29]. To overcome this problem in the early days of the Web, the information space has been structured by hand using predefined *controlled vocabulary* terms. As the Web continued to rapidly grow, the biggest disadvantage of this approach—the static structure—became more and more visible. Together with the rise of search engines, this led to the vanishing of even famous websites using controlled vocabularies such as for example DMOZ[1].

A new way of organizing a set of resources emerged with the introduction of social tagging systems. Prominent instances of social tagging systems on the Web are, e.g., BibSonomy[2], CiteULike[3], Delicious[4], and Flickr[5], where BibSonomy offers sharing of literature and bookmarks, CiteULike the sharing of citations, Delicious the sharing of bookmarks, and Flickr the sharing of photos. These

---

[1] DMOZ has been closed as of Mar 17, 2017, and it is no longer available under https://www.dmoz.org. The editors have set up a static mirror under http://dmoztools.net/.

[2] http://www.bibsonomy.org.

[3] http://www.citeulike.org.

[4] http://del.icio.us.

[5] https://www.flickr.com.

systems allow users to annotate a set of resources according to their needs with freely chosen words also called *tags*. This free-form annotation approached the vocabulary problem from a *social* angle and introduced new research directions, i.e., for structuring and visualizing the information space. Moreover, new models and theories for tag-based navigation have been developed and helped to establish it as a novel way of information access. *Tag-based navigation* is defined as the process of finding a way between two resources of a social tagging system following user assigned tags [38]. This way of exploring the information space is usually supported by a *tag cloud*. The tag cloud is a user interface that visualizes the tags describing a given set of resources. In that sense, a tag cloud is a textual representation of the topic or subject of the resource set and it captures its *aboutness* [24]. Navigation and browsing in a social tagging system are commonly initiated, e.g., by a system-wide tag cloud, by traversing tag hierarchy or by executing a search query typed into a traditional search box (see Sect. 3.1). Using tags as search terms (see "Tag-based social search" in this book [67]) or following recommended tags (see "Tag-based Recommendation" in this book [8]) are user activities very similar to tag-based navigation.

We organize our chapter in the following way. In Sect. 2, we describe the fundamental social tagging process for shaping the information space of a social tagging system. In Sect. 3, we discuss the tag cloud-based user interaction schema in a social tagging system, layouts, usefulness and evaluation of tag clouds, and visualization of trends in tagging data. We also give an overview of more complex interfaces that integrate tag clouds or expose tag hierarchies to users. We discuss clustering of tagging data in Sect. 4, as it is a way of dealing with one of the main problems with tagging data—the lack of structure. In this section, we give an overview over flat and hierarchical tag clustering. The flat tag clustering produces groups of similar tags, whereas the hierarchical tag clustering produces a tag hierarchy in which tags occupy a given hierarchy level based on, e.g., their generality. We show how tag-based navigation is modeled in Sect. 5. In general, models of user navigational behavior are used for providing navigational support such as recommending or highlighting links, adapting the navigational hierarchy, or even removing particular navigational links. In a particular case of tag-based navigation user models can be used to, for instance, adapt a given tag-cloud, include additional tags into the tag-cloud, or for ranking of resources whenever a given tag from a tag-cloud is selected. Finally, we discuss the navigability of social tagging systems from different theoretic perspectives in Sect. 6. Analyzing tag-based navigation with a plethora of network-theoretic tools allows us to evaluate and assess the quality, efficiency, and usefulness of the navigational structures imposed by various social tagging systems. We can use this information to further adapt and improve tag-based navigational constructs. For example, by measuring the average distance between resources in a social-tagging system we obtain a lower bound on the average number of clicks that a user needs to make to traverse between any two given resources. In the cases where the average distance exceeds a typical number of clicks that users make on the Web we have a strong indication for a poorly designed navigational interface that, consequently, we need to improve.

## 2  Social Tagging

In a social tagging system, users assign tags to resources. This process shapes the structure of the social tagging system and is called *social tagging*. The result of such a human-based annotation of resources is referred to as a *folksonomy*—a folk-generated taxonomy. A folksonomy is defined as a tuple $F := (U, T, R, Y)$ where $U, T$ and $R$ are finite sets, whose elements are called users, tags and resources, respectively, and $Y$ is a ternary relation between them, i.e., $Y \subseteq U \times T \times R$, called tag assignments. A folksonomy can also be seen as a tri-partite hypergraph where the node set is divided into three disjoint sets - $V = T \cup U \cup R$ with *hyperedges* expressed by one tag, one user and one resource - $t, u, r$. The presented definition follows the notion of Hotho et al. [44]. For further formal definitions of folksonomies the interested reader may consult "Tag-based social search" in this book [67].

As pointed out by Furnas et al., the social tagging process is the collective effort of solving the vocabulary problem [28]. In this sense, tags are beneficial for navigating the information space since they provide useful hints about a resource collection. Understanding how users create tags is important for: (i) designing user interfaces (see Sect. 3), (ii) designing clustering algorithms (see Sect. 4), (iii) modeling tag-based navigation (see Sect. 5), (iv) studying the theoretic navigability of social tagging systems (see Sect. 6). Steps towards gaining such understanding have been made by Golder and Huberman who studied the regularities in the users' activities and the tag frequencies in social tagging systems [31]. They also identified some of the problems that arise when users create tags such as synonymy (multiple tags that share the same meaning, e.g., little/small), polysemy (a tag that has many related meanings, e.g., wood (a piece of a tree)/wood (an area with many trees)), or homonymy (a tag that has different not related meanings, e.g., band (a musical group)/band (a ring)). Körner et al. presented a different view on the social tagging process by characterizing the users and their tagging motivations (see Table 1) [52]. The authors split the users into at least two main groups depending on their motivation:

1. Categorizers—users who try to divide the resources into categories by assigning tags sound with some personal or shared conceptualization.
2. Describers—users who try to assign tags that describe the resource best.

The categorizers assign tags to use them as a navigational aid and try to develop a consistent taxonomy. Resources are tagged according to a common characteristic important to the mental model of the user (e.g., "pictures", "projects", "drafts", or "archive"). The describers typically assign tags to support indexing of resources, and thus support search and retrieval tasks. The assigned tags are mainly descriptive and stemming from a dynamically changing open set of tags. Although, this user separation is very nice from a theoretic point of view, in the real world, social tagging system users probably belong to these two groups simultaneously. For example, a user can use a very small categorization schema while she is assigning a lot of descriptive tags. At the same time, a tag can be part
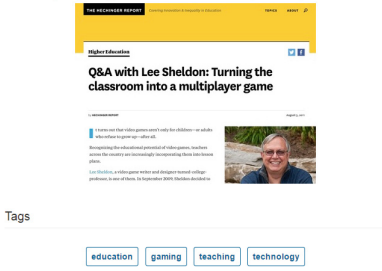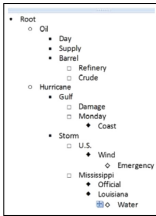
(a) Resource tag list in BibSonomy



(b) System-wide tag cloud in BibSonomy



(c) Tag hierarchy in tagFlake



(d) Tag hierarchy in ELSABer

**Fig. 1.** Tag-based user interfaces. BibSonomy (see footnote 2) provides a resource-specific tag list (a) to navigate between resources and a system-wide tag cloud (b) to initiate browsing. Browsing can also be initiated by a tag hierarchy as implemented in tagFlake (c) [21] and ELSABer (d) [56].

of a categorization schema and still be used as a descriptive tag. A more detailed description of the different tag types, their intended usage and classification is provided in "Tag-based social search" in this book [67].

## 3    User Interfaces and Visualization

Using tags to organize content introduced new research problems, i.e., exposing the content through user interfaces that leverage the advantages of the free-form annotation. In this section, we present interfaces developed to navigate the content of a social tagging system, i.e., the tag cloud and other interfaces that integrate tag clouds or facilitate tag-based browsing, e.g, through tag hierarchies.

**Table 1.** Tagging motivations as identified by Körner et al. in [52].

|                       | Categorizers   | Describers     |
|-----------------------|----------------|----------------|
| Goal of tagging       | Later browsing | Later retrieval |
| Change of vocabulary  | Costly         | Cheap          |
| Size of vocabulary    | Limited        | Open           |
| Tags                  | Subjective     | Objective      |

We also review research literature dealing with problems naturally arising with the introduction of these interfaces, e.g., tag selection, tag cloud layouts and usefulness, tag cloud evaluation, and trend visualization using tag clouds. Table 2 shows a brief summary of the contributions along the research lines discussed in this section.

## 3.1   Tag Clouds

Tag clouds are widely adopted across many social tagging systems because they visualize the information space in an intuitive way. A tag cloud is a textual representation of the topic or subject of a resource as collectively seen by the users and it captures the *aboutness* of the resource [24]. There are three possible visualizations of the relationship between users, tags and resources in a social tagging system. The first presents users and their connection to tags, the second shows users connections to resources and the last presents tags and how they connect to resources. From network theoretic perspective, these are all possible combinations related to the bipartite projections of the tripartite tagging hypergraph. When using tag clouds, however, it is up to the operators of a given social tagging system to decide which one of these combinations to offer.

Let us exemplify a tag cloud with the interaction schema of a user navigating a tag-resource bipartite network [39, 40]:

1. The system presents a tag cloud to the user for a given resource.
2. The user chooses a tag from the tag cloud.
3. The system delivers a list of resources tagged with the selected tag.
4. The user selects a resource from the list.
5. The resource is displayed and the process starts anew.

**Table 2.** User interfaces literature

| Research line | Research work |
| --- | --- |
| Tag selection | Venetis et al. in [84], Skoutas and Alrifai in [80], Helic et al. in [39, 40] |
| Tag cloud layouts | Gambette and Véronis in [30], Jafee et al. in [46], Bielenberg and Zacher in [7], Kaser and Lemire in [48], Seifert et al. in [77], Viegas et al. in [85], Eda et al. in [23] |
| Tag cloud usefulness | Rivadeneira et al. in [73], Sinclair and Cardew-Hall in [78], Lohmann et al. in [60], Bateman et al. in [4], Zubiaga in [94], Kuo et al. in [53], Halvey and Keane in [32], Millen and Feinberg in [65] |
| Tag clouds over time | Lee et al. in [55], Collins et al. in [17],  Dubinko et al. [22], Russell in [74], Wagner et al. in [89] |
| Tag cloud evaluation | Skoutas and Alrifai in [80], Venetis et al. in [84], Trattner et al. in [82], Helic et al. in [39], Aouiche et al. in [2] |
| Integrated interfaces | Kammerer et al. in [47], Lin et al. in [58], Helic and Strohmaier in [36], Li et al. in [56], Di Caro et al. in [21] Vig et al. in [86–88] |

With this interaction schema, a user navigates on a tag-resource bipartite network using resource-specific tag clouds (see Fig. 1(a)). In step three of the interaction schema, the system may also provide a tag cloud that captures the aboutness of the currently presented list of resources. To initiate tag-based navigation, a social tagging system may present a system-wide tag cloud capturing the aboutness of the whole social tagging system (see Fig. 1(b)) or a user-wide tag cloud covering the tags assigned by the currently logged-in user (see Fig. 2(c)). Tag-based browsing can be also initiated by a tag hierarchy. For example, tagFlake by Di Caro et al. [21] and ELSABer by Li et al. [56] are user interfaces that offer top down tag hierarchy browsing. Both interfaces work in a similar manner. First, a tag hierarchy is presented on the left side of the screen. After a tag is selected, the associated resources are displayed on the right side of the screen (see Figure 1(c) and (d)).

**Tag Selection.** One of the first questions arising when designing a tag cloud refers to the tag cloud size, i.e., which and how many tags should be displayed. For example, there are systems presenting only twenty tags, whereas others offer a much bigger number (see Fig. 2). Proposed by Venetis et al., the simple TopN tag selection algorithm is very widely adopted [84]. To create a tag cloud, this algorithm considers only tags assigned to a specific resource. The algorithm selects the top $n$ tags with the highest resource-specific frequency to present in the tag cloud. If there are less than $n$ tags available for a resource, the remaining tag positions in the cloud are left empty. In the same work, Venetis et al. proposed also algorithms for selecting top tags based on standard text features and maximum resource coverage [84]. The tag selection problem has also been studied by Skoutas and Alrifai who introduced a tag selection framework based on frequency, diversity and rank aggregation [80].

**Tag Could Layouts and Functionality.** Figure 2(a), (b) and (c) shows the basic tag cloud layout. However, more sophisticated approaches exist. For example, Gambette and Veronis arranged tags in a tree structure (TreeCloud) so that their semantic proximity is reflected (see Fig. 2(d)) [30]. Proposed by Jafee et al., TagMaps is a unique layout using real geographical space to create a tag cloud for large collections of geo-referenced photographs [46]. Bielenberg and Zacher introduced a circular tag cloud layout and compared it to the typical rectangular layout [7]. In the proposed circular layout, the distance to the center and the font size of the tag represent its importance. In this layout, the distance between tags in the cloud does not reflect their similarity. Different researchers concentrated on the aesthetic issues regarding the tag cloud layouts. Kaser and Lemire arranged tags in nested HTML tables in which tag relationships are considered. To tackle white spaces in tag clouds, emerging due to different font size usage, they proposed to use the *min-cut placement* Electronic Design Automation algorithm [48]. In another work, Seifert et al. concentrated on the visual issues in layouts and proposed a new algorithm utilizing arbitrary convex polygons to bound tags and reduce white spaces [77]. Viegas et al. introduced Wordle,
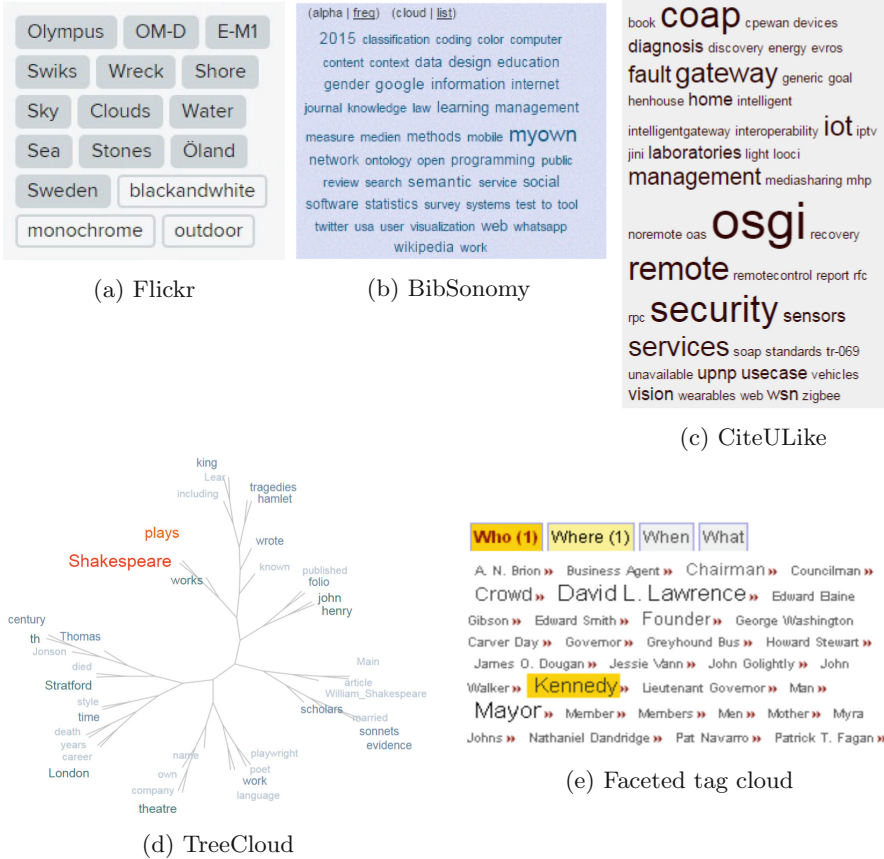
(a) Flickr

(b) BibSonomy

(c) CiteULike

(d) TreeCloud

(e) Faceted tag cloud

**Fig. 2.** Tag cloud layouts and functionality. Flickr (see footnote 5) (a) uses different colors in the tag cloud to distinguish between automatically and user assigned tags. BibSonomy (see footnote 2) (b) and CiteULike (see footnote 3) (c) use different font sizes to indicate tag importance and popularity. Changing the layout from a cloud to an alphabetically or frequency sorted list is offered by BibSonomy (b), whereas CiteULike (c) allows tag filtering. TreeCloud (d) uses a tree structure to reflect semantic proximity between tags [30]. In a faceted tag cloud (e), tags are classified into "Who", "Where", "When" and "What" facets [82].

a distinctive layout that concentrates on the balance of colors, typography and other visual features [85]. Eda et al. concentrated on experienced emotions when using tag clouds and proposed a layout in which the font size is determined by the tag's entropy and not by its content popularity [23].

In Sect. 3.2, we will show how tag clouds have been integrated into more complex interfaces. The tag cloud interface itself, however, can also provide additional functionality, e.g., tag sorting, filtering or faceting. Rearranging tags to present them as an alphabetically sorted list is helpful for finding the presence or absence of a given tag (see Fig. 2(b)). Filtering of tags is useful for tag clouds with large number of tags, i.e. system or user-wide tag clouds (see Fig. 2(c)). In a faceted tag cloud, tags are structured according to a classification schema which can be flat (see Fig. 2(e)) or hierarchical (see Fig. 1(d) and (c)). In Sect. 4, we will discuss algorithms for flat and hierarchical tag clustering used to populate these two types of interfaces with tags.

**Tag Cloud Usefulness.** Although tag clouds are very simple, they support users in multiple ways. For example, Rivadeneira et al. identified tag clouds to be useful for four different tasks [73]: (i) search: finding the presence or absence of a given target, (ii) browsing: exploring the cloud without a particular target in mind, (iii) gaining (visual) impression about a topic, (iv) recognition and matching: recognizing the tag cloud as data describing a specific topic. In an experiment on gaining impressions and recognition, Rivadeneira et al. studied the tag font size and the cloud layout. The results suggested that the font size has a strong effect on recognition. Although the different layouts changed the accuracy of impression, they had no significant effect on recognition.

Sinclair and Cardew-Hall studied the usefulness of tag clouds for different information retrieval tasks and found tag clouds especially useful for browsing scenarios [78]. In such scenarios, tag clouds support discovery of items (resources, users, or other tags) that a user might not have thought of or known about.

Halvey and Keane examined the usefulness of tag clouds for finding a specific target by comparing them to horizontal and vertical alphabetically sorted lists [32]. Their results showed that tag clouds are outperformed by both list types, suggesting that alphabetization aids users for orientation. They also experimented with the tag cloud typography, i.e., font sizes and found out that targets with larger font sizes are found more quickly. Regarding tag font size and position in the cloud, they came to similar conclusion as Rivadeneira et al. in [73], namely, that the font size strongly contributes to recall, whereas proximity to the largest tag has no effect.

Bateman et al. concentrated on the visual features of tag clouds and how they affect the visual search of a tag in the cloud [4]. They concluded that font size has a more significant impact on finding a tag than other visual features such as, e.g., color, tag string length and tag location.

Kuo et al. compared the usefulness of tag clouds and lists for summarizing search results from the biomedical domain [53]. They considered tag clouds superior to search result lists with respect to the presentation of descriptive information. However, tag clouds performed significantly worse when presenting relationships between concepts.

Lohmann et al. studied different tag cloud layouts, i.e., sequential layout (alphabetical sorting), clustered layout (thematic clusters), circular layout
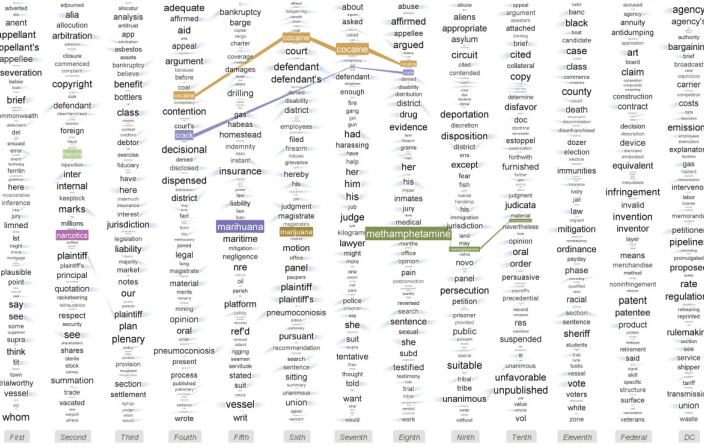
(a) PTC



(b) SparkClouds

**Fig. 3.** Tag clouds over time. PTC (a) compares the popularity of a given tag between multiple time periods [17]. In SparkClouds (b) a sparkline shows, e.g., if a tag is new or if it has experienced high popularity in the system over a time period [55].

(decreasing popularity), and their ability to support typical information seeking tasks [60]. For finding a specific tag, they suggest the sequential layout with alphabetical sorting. The thematically clustered layout performed best for finding tags that belong to a certain topic, whereas the circular layout with decreasing popularity is more appropriate for finding the most popular tags.

Overall, literature suggests that visualizing tags, i.e., their font size and layout of the cloud have significant effect on the tag cloud usefulness for tag-based navigation. Furthermore, the presented findings highlight the intrinsic connection of tag-based navigation and the way tagging data is visualized.

**Tag Clouds Over Time.** Tag clouds are also useful for visualizing trends and comparing resource collections over time. Research in this direction has been conducted by Lee et al. who introduced SparkClouds [55] and by Collins et al. who presented Parallel Tag Clouds (PTC) [17]. PTC is designed to con-

**Fig. 4.** Query expansion using tags as implemented in BibSonomy (see footnote 2). By clicking on the "+" sign of a related tag—tag assigned together with the search tag (top left)—the query can be expanded to refine the set of presented resources. Exploring the content using tags similar to the search tag is also possible (bottom left).

sider and understand changes across multiple resource collections by presenting their tag clouds simultaneously. It combines parallel coordinate plots visualization [45] with tag clouds and can highlight the underuse and overuse of a tag (see Fig. 3(a)). SparkClouds unifies sparklines [83] with typical tag cloud features to visualize evidence of change across multiple tag clouds (see Fig. 3(b)).

As discussed later on, the maturity of a social tagging system influences its navigability. Wagner et al. compared different methods for estimating the system's maturity, i.e., with respect to its semantic stability [89]. Taglines and Cloudalicious are two prominent tools for visualizing the usage and semantic stability of tags. Introduced by Dubinko et al., Taglines is a visualization working with the *river metaphor*—tags flow from left to right—and the *waterfall metaphor*—tags are presented in fixed slots through which they can "travel" over time [22]. With these metaphors, Taglines presents tags that possess a significantly high occurrence frequency inside a given time period, compared to outside this period. Proposed by Russel, Cloudalicious visualizes the evolution of tags over time [74]. The tool works with Delicious tagging data and produces
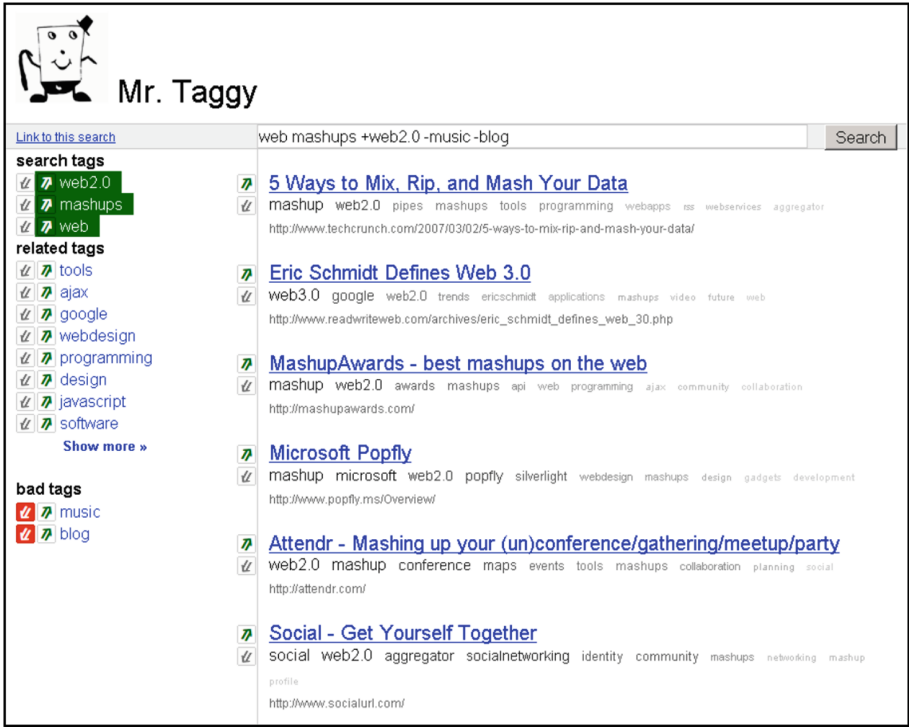
**Fig. 5.** MrTaggy as presented by Kammerer et al. in [47]. The MrTaggy browser allows a user to specify a query and then rate both presented resources (right) and related tags (left) by clicking on the arrows on the left. Clicking a tag arrow up or down refines the query by adding or excluding the tag from the query. Clicking a resource arrow up or down highlights similar results or excludes the resource from the result list.

a graph of the collective tagging activity for a given URL. It shows the relative weights of the most popular tags for the URL. Indications of stabilization are observed as the lines of the graph move from left to right. This pattern expresses the collective opinion of the users with respect to the URL. An even more interesting pattern in the graph are diagonal lines, as such lines suggest that users changed the URL describing tags.

**Tag Cloud Evaluation.** Skoutas and Alrifai, and Venetis et al. conducted research on tag cloud evaluation [80,84]. Both author groups propose very similar evaluation metrics for tag clouds with respect to coverage, overlap and selectivity. Apart from the evaluation metrics, Skoutas and Alrifai introduced a user navigation model that combined with the evaluation metrics allows tag cloud evaluation with respect to navigation. Aouiche et al. proposed an entropy-based metric for evaluating the informativeness of a tag cloud [2].
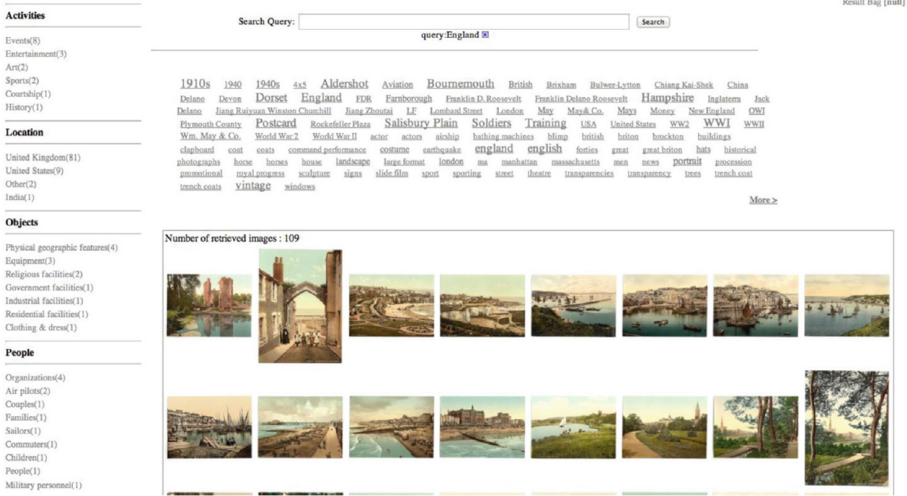
**Fig. 6.** The DPNF interface as proposed by Lin et al. in [58]. The DPNF interface integrates a search box to query the system and start navigation (top), controlled vocabularies as facets (left) and a tag cloud (middle) to explore an image collection (bottom).

Another method for evaluating tag clouds has been followed by Trattner et al. who performed a user study to evaluate tag-based information access in image collections [82]. In the study, they compared traditional and faceted tag clouds with a baseline (search-only) interface. Both tag cloud types performed better than the search-only interface with respect to a predefined search task. Additionally, the authors observed that the faceted tag cloud is more difficult to use initially, but it is considered as more powerful by the study participants in the long run.

Helic et al. showed that the navigability assumption—the widely adopted belief that tag clouds are useful for navigation—does not hold for every social tagging system [39]. Furthermore, they showed that the usefulness of tag clouds is sensitive to the adoption phase of the system, i.e., its maturity and that the navigability assumption may only hold for more mature systems. One very useful finding by Helic et al. is that the limitation of the tag cloud size to a practically more feasible size, e.g., five, ten or more tags does not influence the navigability. Depending on the maturity of the social tagging system and the tag cloud type (e.g., system-wide or resource-specific), however, a tag could covers hundreds or even thousands of resources. In such cases, the resources displayed after a tag selection are often sorted by their reverse chronological order and paginated which reduces the navigability. To tackle this problem, Helic et al. introduced a generalized pagination algorithm and experimented with different context preservation functions.
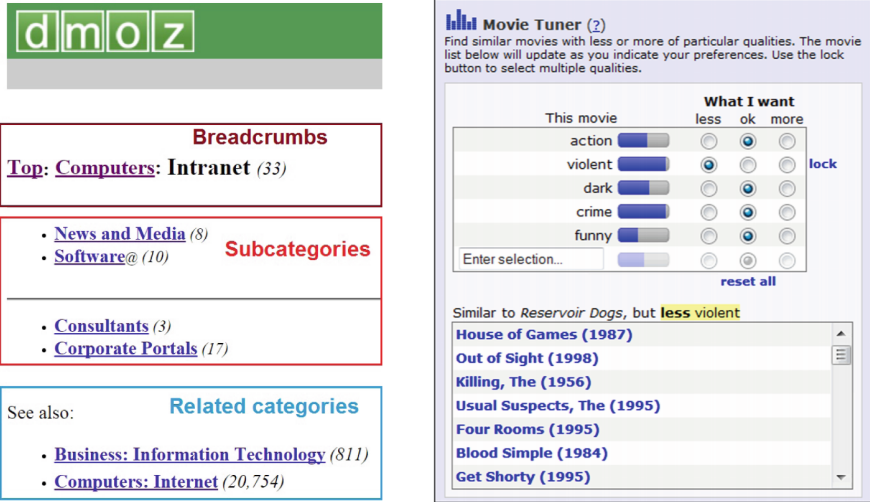
(a) Directory user interface

(b) Movie Tuner interface

**Fig. 7.** The directory user interface (a) as implemented in DMOZ (see footnote 1). The typical interface elements include breadcrumbs, subcategories and related categories. The Movie Tuner interface (b) as proposed by Vig et al. in [87]. The Movie Tuner interface offers users to critique a resource with respect to tags, e.g., "less violent" critique is applied to the movie *Reservoir Dogs*.

### 3.2   Integrated Interfaces

Very often tag clouds are integrated into more complex user interfaces that allow a user to start, e.g., with a query formulation and then narrow down the presented results using the tag cloud. Some social tagging systems, e.g., BibSonomy also allow users to expand or narrow down the query using related or similar tags (see Fig. 4). Presented by Kammerer et al., MrTaggy is a very similar interface that allows searching and browsing of resources and tags by exploiting the relationships between them and the collected relevance feedback (see Fig. 5).

The Dual-Perspective Navigation Framework (DPNF) by Lin et al. introduced an interface seamlessly combining controlled vocabularies (metadata) and free vocabularies (social tags) [58]. By combining both vocabulary types, DPNF aims to provide better resource findability at each navigation step. In a user study, the authors compared the DPNF interface with a tag-only and metadata-only interface and concluded that the DPNF interface preforms best with respect to lookup and exploratory search tasks.

Another more complex interface than the tag cloud is the directory interface. It offers the following user interface elements: (i) breadcrumbs—provide a complete path to the root category, (ii) subcategories—deliver a list of links to more fine-grained categories, (iii) related categories—provide links to related categories. The directory interface is usually adopted in information systems with

hierarchical organization of resources, e.g, DMOZ (see footnote 1) (see Fig. 7a). In a social tagging system, the directory interface allows users to explore the content by navigating along tags organized in a hierarchy. For each tag in the tag hierarchy, the directory interface presents child tags as subcategories and tag siblings as related categories. Tag hierarchies are usually obtained by the state of the art hierarchical clustering algorithms discussed in Sect. 4.2. The navigability of social tagging systems using the directory interface has been studied by Helic et al. with simulations [36]. The authors discovered the limited ability of the user interface to present a tag hierarchy in its entirety and identified the breadth of hierarchy as the main problem reducing navigability.

Introduced by Vig et al., the Movie Tuner is fundamentally different from the interfaces presented so far as it is designed on the intersection of tag-based navigation and tag-based recommendation [87]. It allows a user to navigate a social tagging system by applying *critique* to resources (e.g., movies) with respect to tags (see Fig. 7b). This form of tag-based navigation is called tag-based critiquing. Unlike MrTaggy where a user can specify if a tag is relevant or not, Movie Tuner allows users to specify how relevant a tag is. For example, a user could explore the system for movies that are "less violent" than the movie *Reservoir Dogs*. Similar to the directory interface, Movie Tuner also operates on a special data structure called *tag genome* [86,88]. The tag genome captures the relevance of tags to resources and addresses three limitations of the social tagging process: (i) binary tag-resource relationships (the strength of the relationship is not reflected), (ii) tag sparsity (not all relevant tags may be assigned) and (iii) only positive tag-resource relationships (irrelevance of tags cannot be indicated as they are not assigned). The Movie tuner interface uses a multi-objective tag selection algorithm to choose the tags displayed for a given resource. The optimized tag selection objectives are: critique value, popularity and diversity. After applying a critique across one or multiple tags, the systems recommends resources that satisfy the user's critique. To respond to a user critique, the Movie Tuner interface resorts to an algorithm that selects resources based on the difference to the critiqued tags and similarity to the original resource.

The presented ways of integrating tag clouds into more complex interfaces, aim to achieve better overall user experience and to provide even better support for tag-based navigation.

## 4    Tag Clustering

As pointed out in Sect. 2, synonymy, polysemy and homonymy have been identified as problematic regrading the semantic of tags. Another crucial issue with social tagging data is the lack of structure. The efficiency of tag-based navigation and browsing, however, depends on the structure of the information space. To this end, creating groups of resources meaningful to users by exploiting, i.e., semantic relationships between tags is of special importance for tag-based navigation. Moreover, grouping tags that are semantically related to each other with additional taxonomy relations between tag groups allows us to come up with hierarchical tag-based interfaces. As various previous studies have shown users are

able to efficiently navigate hierarchical interfaces—hence providing such interfaces in a social tagging system is particularly important for supporting users in their explorations of the information space.

In this section, we present an overview of the state of the art algorithms for clustering tagging data. They tackle the above problems by organizing tags according to a classification schema. Depending on the classification schema, there are flat (see Table 3) and hierarchical (see Table 4) clustering algorithms. In general, the discussed tag clustering algorithms represent adaptations of existing state of the art clustering algorithms, e.g., K-Means, or Affinity Propagation. Unlike the algorithms for tag selection focusing on resource-specific tag clouds, the algorithms presented here create tag clouds for resource collections. For example, the flat clustering algorithms organize tags for the presentation through a faceted tag cloud or a system-wide tag cloud. The hierarchical clustering algorithms produce tag hierarchies suitable, e.g., for the directory interface or for interfaces such as ELSABer and tagFlake. We pay special attention to hierarchical clustering due to its importance for modeling tag-based navigation (see Sect. 5). The algorithms reviewed in this section are divided into three different classes: content-based, graph-based and machine learning.

### 4.1    Flat Tag Clustering

The content-based approach has been followed by Specia and Motta who proposed an algorithm for creating semantically related clusters of tags based on their co-occurrence [81]. The algorithm performs statistical analysis of the tag space and constructs a co-occurrence vector for each tag. Clusters are then created using cosine similarity between tags given their co-occurrence vectors. Zubiaga et al. introduced a content-based algorithm using unsupervised neural networks to obtain flat tag clusters [95]. Using language modeling techniques, the clusters are then labeled with the most discriminative tag in a cluster.

The graph-based approach has been adopted by Begelman et al. who proposed a recursive algorithm that uses spectral bisection to split a graph of connected tags into two clusters [5]. Similar to Begelman et al., Au Yeung et al. also introduced a graph-based clustering algorithm using a modularity function to evaluate the quality of division [3]. The authors evaluated their algorithm on three different networks based on users, co-occurrence of tags and context of tags. Hereby, the best results have been achieved using context tags networks.

The machine learning approach has been followed by Remage et al. and by Hassan-Montero and Herrero-Solana. Hassan-Montero and Herrero-Solana proposed an algorithm for tag clustering that considers the semantic relationships between tags [33]. Under a predefined number of clusters and a number of selected relevant tags, the proposed algorithm resorts to K-Means clustering on a tag similarity matrix estimated by means of the Jaccard coefficient. Ramage et al. studied the usage of K-Means clustering in an extended vector space model that contains not only tags but also texts from web pages [72]. They also proposed a novel generative algorithm based on latent Dirichlet Allocation (LDA) that uses tags and web pages texts. They found that the usage of tags in combination with web pages texts improves the cluster quality.

## 4.2  Hierarchical Tag Clustering

In this section, we cover the following three hierarchical clustering algorithms in a more detailed fashion: Hierarchical K-Means [20], Affinity Propagation [71] and Generality in Tag Similarity Graph [41]. Among many other algorithms presented here, these three algorithms are commonly used, simple to implement and exist in different variations.

**Hierarchical K-Means.** K-Means is probably the most prominent clustering algorithm [25,59]. The K-Means versions that we present here complement the flat clustering version from the previous section. For example, Zhong introduced a spherical online version of K-Means [93]. Dhillon et al. adapted the algorithm to work with textual data by replacing Euclidian distance with cosine similarity [20]. A combination of these two K-Means version creates a tag hierarchy in a top-down manner. The algorithm starts by splitting the whole input data into ten clusters. Clusters with more than ten samples are processed iteratively in the same manner, whereas clusters with less than ten samples are considered as leaf clusters. A special case is introduced to handle clusters with eleven samples which initially would have been also split into ten clusters. This special case gives freedom to the partitioning as it allows the division of clusters with eleven samples not into ten but into three clusters. Each node in the hierarchy is represented by the nearest tag to the centroid. This tag is removed from the actual tags contained in a cluster if the cluster is further partitioned.

**Affinity Propagation.** Affinity propagation has been originally proposed by Frey and Dueck [26]. The input of the algorithm is a set of similarities between data samples provided in a matrix. The diagonal of the matrix contains the self-similarity values representing the suitability of the data sample to serve as a cluster center. They are also called preferences. Specifying a number of desired clusters is not needed, however, there is a correlation between the preference values and the number of clusters (lower preference values imply a low number of clusters and vice versa). Affinity propagation characterizes each data sample according to its "responsibility" and its "availability" values. The responsibility expresses the ability of the sample to serve as an exemplar for other samples, whereas the availability shows the suitability of other data samples to be the exemplars for a specific data sample. Affinity propagation exchanges messages

**Table 3.** Flat clustering

| Algorithm type | Example approaches |
|---|---|
| Machine learning | Ramage et al. in [72], Hassan-Montero and Herrero-Solana in [33] |
| Graph-based | Begelman et al. in [5], Au Yeung et al. in [3] |
| Content-based | Specia and Motta in [81], Zubiaga et al. in [95] |

between data samples and iteratively updates the responsibility and availability values of each sample with a parameter $\lambda$ as an update factor. By adding structural constraints into the global objective function of affinity propagation, Plangprasopchok et al. adapted the algorithm to create a taxonomy [71]. Another approach to induce a hierarchy is to use the original version of affinity propagation recursively in a bottom-up manner. The algorithm starts with a matrix containing the top ten cosine similarities between the tags in a given dataset. The minimum of those similarities acts as preference for all data samples. The clusters are produced by selecting examples with associated data samples. Depending on an adjustable parameter specifying the ratio between the desired number of clusters and the data samples, the results are returned or another iteration starts. If the number of selected clusters in the previous run was too high, the preference values are lowered. Otherwise, they are increased. The sum of the connected data samples normalized to unit length represents the centroid of the cluster, while cosine similarity between the centroids of the clusters serve as input matrix for the next iteration. This process is repeated until the top-level is reached. As the output of the algorithm should be a hierarchy, each node in the hierarchy needs to represent a unique tag. To this end, the nearest tag to the centroid is selected as the tag representing the node. Additionally, the selected tag is removed from the actual tags contained in the leaf cluster and it cannot be used in lower hierarchy levels. The update factor $\lambda$ can be dynamically adjusted in each iteration. The algorithm terminates when a given number of iterations is reached or if the clusters are stable for at least ten iterations.

**Generality in Tag Similarity Graph.** Introduced by Heymann and Garcia-Molina, this algorithm receives a *tag similarity graph* as input [41]. The tag similarity graph is an unweighted graph in which each tag is represented by a node and two nodes have an edge between them if the similarity between

**Table 4.** Hierarchical clustering

| Algorithm type | Example approaches |
|---|---|
| Machine learning | Dhillon et al. in [20], Zhong in [93] (K-Means), Schmitz et al. in [75] (Association rules), Di Caro et al. in [21] (LSA), Candan et al. in [12] (LSA), Li et al. in [56] (Decision trees) |
| Graph-based | |
| - Graph clustering | Muchnik et al. in [66], Lancichinetti et al. in [54] |
| - Affinity propagation | Frey and Dueck in [26], Plangprasopchok et al. in [71] |
| Content-based | |
| - Generality in tag similarity | Heymann and Garcia-Molina in [41], Benz et al. in [6], Helic and Strohmaier in [36] |
| - Other | Schmitz in [76], Brooks and Montanez in [11] |

their respective tags is above some threshold. The algorithm starts by setting the most general node (central node in the similarity graph) as root of the hierarchy. All other nodes are added to the hierarchy in descending order of their centrality in the similarity graph. For each candidate node, the similarity between all currently present nodes in the hierarchy and the candidate node is calculated. The candidate node is added as a child of the most similar node in the hierarchy if their similarity is above a given threshold. Otherwise, the candidate node is added as a child of the root. The algorithm makes three main assumptions:

1. Hierarchy representation assumption—the edges representing a given hierarchy are also present in the similarity graph.
2. Noise assumption—there are noisy connections between unrelated tags (mainly due to spamming activities).
3. General-general assumption—the noisy connections between tags occur more often in the higher levels of a given hierarchy.

According to Heymann and Garcia-Molina, the hierarchy representation assumption is essential for detecting hierarchies based on similarity measures. Since tagging data exhibits a lot of noise [13], noisy tags would be of high degree in the tag similarity graph. Thus, they would occupy high hierarchy levels which would eventually reduce the ability of the produced hierarchy to guide navigation. This makes the second assumption also fundamental as it accounts for noisy tag connections. The general - general assumption is based on the intuition that higher level (more general) tags are likely to co-occur by chance. Inserting the more general (central) nodes in the similarity graph in the top of a hierarchy assures short hierarchy distances between the most general tags.

The authors mention the possibility to use different similarity measures as well as different centrality measures. Typical versions of the algorithm are degree centrality as centrality measure and co-occurrence as similarity measure (DegCen/Cooc) and closeness centrality and Cosine similarity (CloCen/Cos). As pointed out by Heymann and Garcia-Molina, more control over the properties of the hierarchy is possible by dynamically adjusting the similarity threshold.

**Other Algorithms.** Muchnic et al. discussed an algorithm for condensing a hierarchy based on metrics for estimating the hierarchy level of single nodes in a network [66]. Clauset et al. presented a general approach for extracting hierarchies from network data demonstrating that the existence of a hierarchy can simultaneously explain and quantitatively reproduce several commonly observed topological properties of networks, e.g., right-skewed degree distributions, high clustering coefficients and short path lengths [16]. Lancichinetti et al. proposed an approach for discovering hierarchies based on overlapping network community structures [54]. They introduced a fitness function for estimating the quality of cover and used it to find the most appropriate community for each network node.

Benz et al. computed the generality in tag similarity graph algorithm by Heymann and Garcia-Molina by using co-occurrence as similarity measure and degree centrality as centrality measure [6]. Additionally, they introduced an extensive preprocessing of the data to remove synonyms and resolve ambiguous tags. Helic and Strohmaier also adapted the generality in tag similarity graph algorithm to control for the breadth in the top levels of the created hierarchy as they identified it as a navigability reducing factor [36].

Li et al. presented the Effective Large Scale Annotation Browser (ELSABer) to browse social annotation data [56]. The algorithm creates a hierarchy using a decision tree and tag features containing, e.g., tag coverage, inverse coverage rate and intersection rate. Schmitz et al. applied association rule mining to extract hierarchies from tagging data concentrating on ontology learning and emergent semantics [75]. Another algorithm for creating tag hierarchies has been presented by Candan et al. who constructed a hierarchy by transforming the tag space into a tag graph and then minimizing its spanning tree [12]. The algorithm uses a similarity lower bound to prevent a context drifting of the tags in the hierarchy. Di Caro et al. described an algorithm that extracts the most significant tags from text documents (not from tagging data) and maps them to a hierarchy so that descendant tags are contextually dependant on their ancestors within a given document corpus [21]. Both algorithms by Di Caro et al. and by Candan et al. applied latent Semantic Analysis (LSA).

Brooks and Montanez presented an agglomerative clustering algorithm to induce a tag hierarchy using abstract tags and abstract tag clusters [11]. Schmitz introduced a subsumption-based algorithm for inducing tag hierarchies [76]. The algorithm uses co-occurrence statistics and builds a graph of possible parent-child relationships. For each node, the best path to a root is calculated under the consideration of reinforced possible parents. The paths are then composed into a tree.

## 5   Modeling Navigation in Social Tagging Systems

As shortly mentioned in Introduction models of user navigational behavior have been extensively used to improve information retrieval capabilities of the Web-based information systems—the most famous example being the Google's random surfer model used to improve rankings of search results. Similarly to the random surfer model various other navigational models have been applied for providing further navigational improvements such as adaptations of links and navigational interfaces by, for example, inserting, highlighting or removing links. For exactly these reasons navigational models have been also applied and adapted for social tagging systems.

In this section, we present two main frameworks for modeling tag-based navigation: Markov chains and decentralized search. Modeling navigation in tagging systems has been recognized as an important step towards better understanding of user navigation behavior [35], which in turn has major practical implications such as implementing more efficient user interfaces [36].

Markov chains have been regularly applied for modeling navigation on the Web, i.e., on information networks [92]. On the other hand, decentralized search approaches has been applied to study the navigational efficiency of broad and narrow folksonomies [35], to evaluate a folksonomy from a pragmatic point of view with respect to tag-based navigation [38] and to build directories for social tagging systems [36].

As shown in Sect. 3.1, tag-based navigation is facilitated either by traversing a tag hierarchy or by navigating between tags connecting resources on a tag-resource network. For pragmatic reasons, however, tag-based navigation is often modeled on a tag-tag network projected over a tag-resource network. Such network mappings reduce complexity and are shown to be effective, e.g., in the field of ontology learning [64].

## 5.1   Markov Chain Models

Navigation on the Web is the process of following links between web pages. Markov chains model navigation on the Web by assigning transition probabilities between web pages also called *states* [9,19,57,79]. Although Markov chain models can also be of higher order (the transition probability between two states depends on several previous states), first order Markov chains (the transition probability depends only on the current state) are more commonly used due to their simplicity. Navigation on the tag-tag network of a social tagging system is modeled with Markov chains by representing each tag as a state. Transition probabilities between states are then assigned according to the distance between the tags in a tag hierarchy induced, e.g., by the algorithms presented in the previous section.

## 5.2   Decentralized Search

Decentralized search is an algorithm designed by Kleinberg to explain the ability of humans to efficiently search other people in huge social networks [49,50]. The algorithm has been since its invention also used to model navigation in information networks. To model navigation, the *decentralized search* algorithm passes messages between network nodes. In decentralized search, the message holder forwards a message to one of its immediate neighbor nodes until the intended message recipient (the target node) is found. For selecting the next step in the navigation process, decentralized search resorts to a *background knowledge* to rank the neighbors of the current node (also called candidate nodes) and forward the message to one of them.

When modeling tag-based navigation, the background knowledge is a tag hierarchy and the message is passed to the neighbor $j$ with the shortest hierarchy distance $d(j,t)$ to the target node $t$.

In Fig. 8, we see an example of decentralized search in a tag-tag network using a tag hierarchy as hierarchical background knowledge. The goal in this example is to find a path between the start node 13 (marked yellow) and the target node 33 (marked red). To select the next node, the algorithm looks up
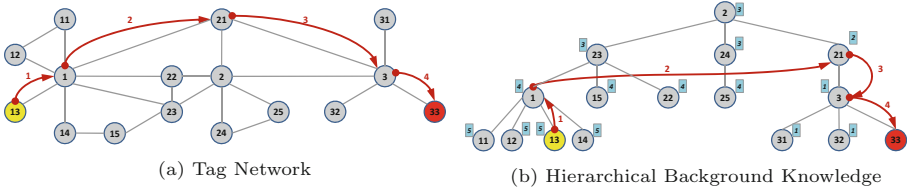
(a) Tag Network                          (b) Hierarchical Background Knowledge

**Fig. 8.** Decentralized search using hierarchy as background knowledge as shown by Helic et al. in [38]. (Color figure online)

the distance of the neighbors of the current node to the target node in the hierarchical background knowledge. The neighbor with the shortest hierarchy distance is then selected. For the first step, node 1 is selected since it is the only adjacent node to 13. At step two, the set of neighbors of node 1 contains 11, 12, 13, 14, 21, 22 and 23 and the node with the shortest distance to the target node is node 21 (number in boxes in (b) provides the distance of each node to the target node). The procedure is repeated until the target node is reached. The red arrows show the resulting path.

In the given example, the message is always passed to the node with the shortest hierarchy distance. Thus, a distance greedy action selection is used to model a confidently navigating user which is a plausible scenario in a social tagging system. The action selections presented next are also applicable although originally developed for modeling navigation on information networks [37]:

$e$**-greedy:** The e-greedy action selection chooses the candidate node $j$ with the shortest distance to the target node $t$ with a probability $1-e$. With a probability $e$, another candidate node is chosen uniformly at random.

**Softmax Rule:** The softmax rule [10,18] chooses a candidate node with shortest distance to the target node with a probability $p(j) \propto e^{cf(j)}$. Hereby, $f(j)$ represents the fitness function calculated from the distances $d(j,t)$, and $c$ is the user's confidence in her intuition. For high values of $c$, the softmax rule selects the candidate node with the shortest distance to the target nodes, thus, reduces to greedy selection. For small values of $c$, the softmax rule is tuned to select other candidate nodes based on $f(j)$, thus, it models a user with low confidence.

**Inverse Distance Rule:** The inverse distance rule [63] is very similar to the softmax rule as it selects the candidate node with a probability $p(j) \propto f(j)^{-c}$. The parameter $c$ expresses again the confidence. The main difference to the softmax rule is the different probability distribution.

**Dacaying e-greedy:** The decaying e-greedy rule [37] is based on the idea that humans do not possess sufficient intuition in the beginning of the navigation process, but their intuition becomes better and better during the process. The rule is based on a decay function that adapts $e$ at every step of the navigation. Different decay functions are possible, but normally $e(t) = e_0 \lambda^{-t}$ is used. Hereby, $e_0$ is the initial value of $e$, and $\lambda$ is a decaying factor at step $t$.

# 6  Theoretic Navigability of Social Tagging Systems

In this Chapter we discuss the navigability of social tagging systems from the network-theoretic perspective. Network-theoretic analysis of navigation allows us to theoretically evaluate and assess the quality, efficiency, and usefulness of the navigational structures imposed by various social tagging systems. Such theoretical analysis provides us with theoretical bounds on various aspects of social tagging systems and provides the first evaluation results and first indications for potential navigational bottlenecks and problems. Subsequently, we may remedy the problems even before performing expensive usability tests with real users. For example, network-theoretic tools allows us to study connectivity of various parts of our information systems—in this way we can identify completely disconnected or poorly connected groups of resources in our system.

So far, the theoretic navigability of social tagging systems has been studied from four different perspectives on which we want to shed light in this section: network theoretic, information theoretic, information foraging and tagging vs. library approach. Each perspective emphasizes that the navigability of a social tagging systems depends on the ability of the users to assign tags to resources, i.e., to solve the vocabulary problem.

## 6.1  Network Theoretic Perspective

Adamic et al. studied navigation in power-law degree distributed networks and showed that random walks naturally tend to select nodes with a high degree [1]. Based on this observation, they proposed a version of the decentralized search algorithm that exploits the degree distribution for finding the target node by passing the message to the candidate node with the highest degree. Such an algorithm makes each power-law degree distributed network theoretically searchable. Folksonomies—the data structures of social tagging systems—possess power-law degree distributions (see "Tag-based social search" in this book [67]), thus, they are easily navigable with Adamic's algorithm. In the case of tag-based navigation, however, the navigability of a folksonomy cannot be measured in this way as the algorithm does not exploit the semantic relationships between tags but only the network topology. Furthermore, tag hierarchies created by the algorithms presented in Sect. 4 are structures capturing not only semantic relationships between tags but also other useful properties of the social tagging process which cannot be neglected when looking at the network theoretic perspective of tag-based navigation. To this end, in this section we concentrate on two aspects: (i) the general navigability of a folksonomy as a graph and (ii) the ability of tag hierarchies to guide navigation in such a graph.

**Navigability of a Folksonomy as a Graph.** Cattuto et al. described the navigability of a folksonomy in terms of its "small world" network properties [13]. Small world networks are easy to navigate as all network nodes are reachable within few steps. In a pioneer work, Watts and Strogatz defined the class of

"small world" networks based on the characteristic path length and the clustering coefficient [91]. The characteristic path length is a global network topology measure specifying the average shortest path distance for all possible node pairs. The clustering coefficient is a local measure and specifies the extent to which the neighbors of a given node form a clique. Since folksonomies are tri-partite graphs, the above measures cannot be directly applied to study their the network properties. To this end, Cattuto et al. redefined the characteristic path length and the clustering coefficient for three mode data. After comparing observed folksonomies with two randomly generated folksonomies of equal size with respect to both measures, they found that the observed folksonomies have extremely high clustering coefficients and comparable to lower characteristic path lengths. Cattuto et al. also noticed the small characteristic path length values (about 3.5) which did not change significantly as observed folksonomies grew. This is a very important observation since it implies that on average every resource, user and tag is reachable from any other resource, user or tag within a couple of clicks. This high reachability explains also why folksonomies support serendipitous discovery [62].

**Navigation Supported by Tag Hierarchies.** The ability of tag hierarchies to guide tag-based navigation has been studied by Helic et al. using the small world network models [38]. In general, there are two types of small world network models—lattice-based (ring lattice model by Watts and Strogatz [91] and 2D-lattice model by Kleinberg [50]) and hierarchy-based (single hierarchy model by Kleinberg [51] and multiple hierarchies model by Watts et al. [90]). Essentially, those small world models generate networks in which the balance between the local network structure (short range links) and the global network structure (long range links) is used to guide navigation modeled, e.g., as decentralized search. In the hierarchy-based models, this balance is regulated through the distance distribution between the nodes of the hierarchies generating the network. Inspired by the above observations and models, Helic et al. proposed a theoretic evaluation of the suitability of tag hierarchies to support tag-based navigation in a tag-tag network. For all connected node pairs in the network, they suggested to measure the distance between the pair nodes in a given tag hierarchy and to create a distance distribution. The theoretic suitability of the tag hierarchy to support navigation is then estimated using this distance distribution. More precisely, the authors introduced an indirect comparison of the distance distribution of a tag hierarchy and the class of theoretically searchable networks according to Watts' model. A direct comparison is not possible due to the following differences: (i) in Watts' model, the degree distribution is uniform, whereas the tag degree distribution has been shown to follow a power-law; (ii) in a tag hierarchy, tags could be potentially attached everywhere in a hierarchy, which is not the case in the model of Watts where they would be attached only to leaves. To tackle these differences, Helic et al. adapted Watts' model to tagging networks and discussed the distance distributions of two synthetic tag hierarchies—the random and the homophily-based hierarchy. The homophily distance distribution
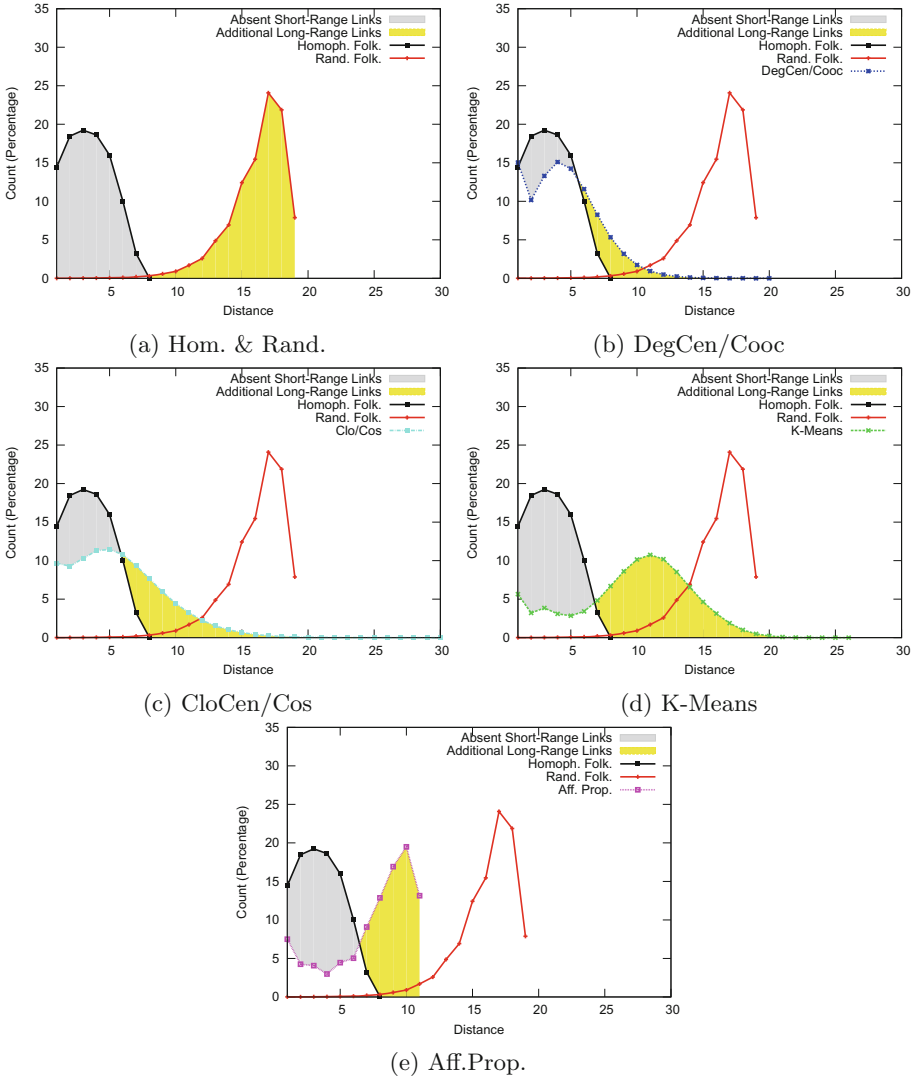
(a) Hom. & Rand.

(b) DegCen/Cooc

(c) CloCen/Cos

(d) K-Means

(e) Aff.Prop.

**Fig. 9.** Comparison of distance distributions for four hierarchical tag clustering algorithms as shown by Helic et al. in [38]. (Color figure online)

describes a hierarchy that only supports short range connections in the tag-tag network, whereas the random distance distribution mimics a tag hierarchy with both short and long range connections. None of the distributions is optimal (see Fig. 9(a)). The homophily distance distribution is dominated by short range links, whereas the random distance distribution is dominated by long range links. However, the homophilous tag hierarchy, which is lacking some long range links, is theoretically more suitable to guide tag-based navigation.

In Fig. 9, we also see a comparison between the distance distributions of tag hierarchies created by the algorithms in Sect. 4.2 on a BibSonomy dataset[6] and the two synthetic distance distributions—random and homophily-based distributions. The gray and the yellow areas represent the differences in the number of short range links and long range links, respectively. The gray area is called the Absent Short-Range Links area, whereas the yellow area is called the Additional Long-Range Links area. Theoretically, both areas need to be greater than zero but still rather small. Otherwise, the distance distribution will incorporate too many long range links and will become similar to the random distance distribution. The tag hierarchies created by DegCen/Cooc and CloCen/Cos versions of the generality in tag similarity graph algorithm should perform best from a theoretic point of view (see Fig. 9(b) and (c)) as they exhibit distance distributions with many short range links mixed with a few long range links. The distance distributions of the tag hierarchies induced by K-Means and Affinity Propagation seem theoretically less suitable since they possess too many long range links and too few short range links (see Fig. 9(d) and (e)).

## 6.2    Information Theoretic Perspective

The navigability of social tagging systems has also been studied by Chi and Mytkowicz from an information theoretic perspective [14,15]. In their work, Chi and Mytkowicz see social tagging as the collective effort of creating a mental map summarizing an information space. They suggest that users can benefit from this map as a navigational aid for efficiently exploring the information space. This idea is very similar to the idea of using hierarchies induced from folksonomies as background knowledge for modeling tag-based navigation with decentralized search or Markov models. With their work, the authors address the vocabulary problem—the ability of users to efficiently assign tags to resources, thus, to create mental maps of the information space. In their analysis, Chi and Mytkowicz calculated three information-theoretic measures: (i) entropy—a measure of uncertainty in a random variable, (ii) conditional entropy—a measure of the remaining entropy of a random variable given the that the value of the second random variable is known and (iii) mutual information—a symmetric measure of the independence of two random variables. Chi and Mytkowicz applied these measures to a Delicious folksonomy and found out that over time as the social tagging systems mature (i) the tagging efficiency is decreasing, thus, tags lose their descriptiveness, (ii) tags lose their ability to deliver conspicuous navigability and (iii) there is a decaying ability of the users to navigate between tags and resources.

## 6.3    Information Foraging Perspective

Pirolli and Card proposed the information foraging theory to describe the human information seeking in a digital environment [69]. In subsequent work, the original theory has been adapted to model user navigation on the Web [27,70] and to an elementary social information foraging model (SIF) [68].

---

[6] http://www.kde.cs.uni-kassel.de/ws/dc09/.

In their work, the authors establish the SIF model as a useful mathematical tool for studying how social collaboration influences information foraging. SIF assumes that in the process of information foraging, *hints* (tags) are created and shared between individuals which makes navigation easier and improves the organization of the information space. The model presents a perspective that goes hand in hand and complements the perspectives presented earlier in this section. It also captures different aspects of the collective information foraging, i.e., time, diversity of the hints, cost of cooperation and social capital of the collaborators. SIF considers the effectiveness of the hints—tightly related to the entropy and mutual information analysis from the information theoretic perspective—in terms of their amount, validity, and interpretability by the forager in a certain step during navigation. According to the model, lowering the cost of effort associated with creating and sharing tags leads to higher productivity of the users of a social tagging systems. Click2Tag [43], a tagging technique that follows this idea, is shown to have lower tagging costs compared to the widely adopted "type-to-tag" approach.

## 6.4   Tagging vs. Library Approach

Macgregor and McCulloch discussed the advantages and disadvantages of the "tagging approach" (resources are annotated by users with freely chosen tags) and the "library approach" (resources are annotated by users with predefined controlled vocabulary) [61]. They proposed a definition of a controlled vocabulary and compared unrestricted free-form vocabularies emerged in social tagging systems to controlled vocabularies. Macgregor and McCulloch pointed out that controlled vocabularies have advantages in dealing with synonyms and homonyms, thus, provide good semantic clues. Compared to free-form vocabularies that exhibit a lot of noise introduced by the users, controlled vocabularies can handle lexical anomalies. Macgregor and McCulloch concluded that the precision and the recall of free-form vocabularies depend on the distribution of the tags. Consequently, general tags exhibit high recall and suffer precision, whereas specific tags suffer recall and enjoy precision. Heymann et al. made also a comparison of the navigational characteristics of the "tagging approach" and the "library approach" represented by tagging distributions and library terms distributions, respectively [42]. In their comparison, the authors focused on three major large scale organizational features of the tagging and library approaches: consistency, i.e., ability to deal with synonyms, quality, i.e., with respect to tag distributions and completeness, i.e., correspondence between tag and library terms. These organizational features give a different perspective to the "vocabulary problem" addressed from the information theoretic perspective by Chi and Mytkowicz. Their results suggest that tagging systems tend to be at least to some extent consistent, of high quality and complete. They found that: (i) synonyms are not problematic, (ii) moderately common user tags are perceived as even more helpful than library annotations assigned by an expert and (iii) top tags correspond to library terms.

# 7    Conclusion

In this chapter, we discussed the challenges researchers faced with the emergence of social tagging systems offering a free-form annotation of resources using tags. First, we presented the user interfaces that allow tag-based navigation, i.e., tag clouds and tag hierarchies. We paid special attention to the tag clouds as the most common interface that accounts for the unstructured nature of tagging data. We reviewed literature focusing on tag cloud usefulness, layouts and evaluation. Furthermore, we discussed trend visualizations with tag clouds. We also presented how tag clouds have been integrated into more complex user interfaces. We summarized the most popular state of the art algorithms for tag clustering used to populate, e.g., a tag cloud or to create a tag hierarchy. Lastly, we showed how tag-based navigation have been modeled and provided an overview of the different theoretic perspectives regarding the ability of folksonomies to support tag-based navigation, i.e., the network theoretic, the information foraging and entropy of information perspectives, and the "tagging approach" vs. "library approach".

# References

1. Adamic, L.A., Lukose, R.M., Puniyani, A.R., Huberman, B.A.: Search in power-law networks. Phys. Rev. E **64**(4), 046135 (2001)
2. Aouiche, K., Lemire, D., Godin, R.: Web 2.0 OLAP: from data cubes to tag clouds. CoRR abs/0905.2657 (2009). http://arxiv.org/abs/0905.2657
3. Au Yeung, C., Gibbins, N., Shadbolt, N.: Contextualising tags in collaborative tagging systems. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, HT 2009, pp. 251–260. ACM, New York (2009)
4. Bateman, S., Gutwin, C., Nacenta, M.: Seeing things in the clouds: the effect of visual features on tag cloud selections. In: Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia, HT 2008, pp. 193–202. ACM, New York (2008)
5. Begelman, G., Keller, P., Smadja, F.: Automated tag clustering: improving search and exploration in the tag space. In: Collaborative web tagging workshop at WWW 2006, vol. 50, Edinburgh, Scotland (2006)
6. Benz, D., Hotho, A., Stumme, G., Sttzer, S.: Semantics made by you and me: self-emerging ontologies can capture the diversity of shared knowledge. In: Proceedings of the 2nd Web Science Conference, WebSci 2010 (2010)
7. Bielenberg, K., Zacher, M.: Groups in social software: utilizing tagging to integrate individual contexts for social navigation (2006)
8. Bogers, T.: Tag-based recommendation. In: Brusilovsky, P., He, D. (eds.) Social Information Access. LNCS, vol. 10100, pp. 441–479. Springer, Cham (2018)
9. Borges, J., Levene, M.: Evaluating variable-length markov chain models for analysis of user web navigation sessions. IEEE Trans. Knowl. Data Eng. **19**(4), 441–452 (2007)
10. Bridle, J.S.: Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: Soulié, F.F., Hérault, J. (eds.) Neurocomputing, vol. 68, pp. 227–236. Springer, Heidelberg (1990). https://doi.org/10.1007/978-3-642-76153-9_28

11. Brooks, C.H., Montanez, N.: Improved annotation of the blogosphere via autotagging and hierarchical clustering. In: Proceedings of the 15th International Conference on World Wide Web, WWW 2006, pp. 625–632. ACM, New York (2006)
12. Candan, K.S., Di Caro, L., Sapino, M.L.: Creating tag hierarchies for effective navigation in social media. In: Proceedings of the 2008 ACM Workshop on Search in Social Media, SSM 2008, pp. 75–82. ACM, New York (2008)
13. Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V.D., Loreto, V., Hotho, A., Grahl, M., Stumme, G.: Network properties of folksonomies. AI Commun. **20**(4), 245–262 (2007)
14. Chi, E.H., Mytkowicz, T.: Understanding navigability of social tagging systems. In: Proceedings of CHI, vol. 7 (2007)
15. Chi, E.H., Mytkowicz, T.: Understanding the efficiency of social tagging systems using information theory. In: Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia, HT 2008, pp. 81–88. ACM, New York (2008)
16. Clauset, A., Moore, C., Newman, M.E.: Hierarchical structure and the prediction of missing links in networks. Nature **453**(7191), 98–101 (2008)
17. Collins, C., Viegas, F.B., Wattenberg, M.: Parallel tag clouds to explore and analyze faceted text corpora. In: IEEE Symposium on Visual Analytics Science and Technology, VAST 2009, pp. 91–98. IEEE (2009)
18. Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., Dolan, R.J.: Cortical substrates for exploratory decisions in humans. Nature **441**(7095), 876–879 (2006)
19. Deshpande, M., Karypis, G.: Selective markov models for predicting web page accesses. ACM Trans. Internet Technol. (TOIT) **4**(2), 163–184 (2004)
20. Dhillon, I.S., Fan, J., Guan, Y.: Efficient clustering of very large document collections. In: Grossman, R.L., Kamath, C., Kegelmeyer, P., Kumar, V., Namburu, R.R. (eds.) Data Mining for Scientific and Engineering Applications. MC, vol. 2, pp. 357–381. Springer, Boston, MA (2001). https://doi.org/10.1007/978-1-4615-1733-7_20
21. Di Caro, L., Candan, K.S., Sapino, M.L.: Using tagflake for condensing navigable tag hierarchies from tag clouds. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008, pp. 1069–1072. ACM, New York (2008)
22. Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., Tomkins, A.: Visualizing tags over time. ACM Trans. Web **1**(2), 7 (2007)
23. Eda, T., Uchiyama, T., Uchiyama, T., Yoshikawa, M.: Signaling emotion in tagclouds. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, pp. 1199–1200. ACM, New York (2009)
24. Fairthorne, R.A.: Content analysis, specification and control. Annu. Rev. Inf. Sci. Technol. **4**, 73–109 (1969)
25. Forgy, E.W.: Cluster analysis of multivariate data: efficiency versus interpretability of classifications. Biometrics **21**, 768–769 (1965)
26. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science **315**(5814), 972–976 (2007)
27. Fu, W.T., Pirolli, P.: SNIF-ACT: a cognitive model of user navigation on the world wide web. Hum.-Comput. Interact. **22**(4), 355–412 (2007)
28. Furnas, G.W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C., Naaman, M.: Why do tagging systems work? In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2006, pp. 36–39. ACM, New York (2006)
29. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. Commun. ACM **30**(11), 964–971 (1987)

30. Gambette, P., Véronis, J.: Visualising a text with a tree cloud. In: Locarek-Junge, H., Weihs, C. (eds.) International Federation of Classification Societies Conference, IFCS 2009, pp. 561–569. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10745-0_61

31. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. J. Inf. Sci. **32**(2), 198–208 (2006)

32. Halvey, M.J., Keane, M.T.: An assessment of tag presentation techniques. In: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, pp. 1313–1314. ACM, New York (2007)

33. Hassan-Montero, Y., Herrero-Solana, V.: Improving tag-clouds as visual information retrieval interfaces. In: International Conference on Multidisciplinary Information Sciences and Technologies, InSciT 2006 (2006)

34. Hearst, M.: Search User Interfaces. Cambridge University Press, Cambridge (2009)

35. Helic, D., Körner, C., Granitzer, M., Strohmaier, M., Trattner, C.: Navigational efficiency of broad vs. narrow folksonomies. In: Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT 2012, pp. 63–72. ACM, New York (2012)

36. Helic, D., Strohmaier, M.: Building directories for social tagging systems. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM 2011, pp. 525–534. ACM, New York (2011)

37. Helic, D., Strohmaier, M., Granitzer, M., Scherer, R.: Models of human navigation in information networks based on decentralized search. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media, HT 2013, pp. 89–98. ACM, New York (2013). https://doi.org/10.1145/2481492.2481502

38. Helic, D., Strohmaier, M., Trattner, C., Muhr, M., Lerman, K.: Pragmatic evaluation of folksonomies. In: Proceedings of the 20th International Conference on World Wide Web, WWW 2011, pp. 417–426. ACM, New York (2011)

39. Helic, D., Trattner, C., Strohmaier, M., Andrews, K.: On the navigability of social tagging systems. In: 2010 IEEE Second International Conference on Social Computing (SocialCom), pp. 161–168. IEEE (2010)

40. Helic, D., Trattner, C., Strohmaier, M., Andrews, K.: Are tag clouds useful for navigation? A network-theoretic analysis. Int. J. Soc. Comput. Cyber-Phys. Syst. **1**(1), 33–55 (2011)

41. Heymann, P., Garcia-Molina, H.: Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical report 2006–2010, Stanford University, April 2006

42. Heymann, P., Paepcke, A., Garcia-Molina, H.: Tagging human knowledge. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM 2010, pp. 51–60. ACM, New York (2010)

43. Hong, L., Chi, E.H., Budiu, R., Pirolli, P., Nelson, L.: SparTag.us: a low cost tagging system for foraging of web content. In: Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2008, pp. 65–72. ACM, New York (2008)

44. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: search and ranking. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 411–426. Springer, Heidelberg (2006). https://doi.org/10.1007/11762256_31

45. Inselberg, A.: The plane with parallel coordinates. Vis. Comput. **1**(2), 69–91 (1985)

46. Jaffe, A., Naaman, M., Tassa, T., Davis, M.: Generating summaries and visualization for large collections of geo-referenced photographs. In: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, MIR 2006, pp. 89–98. ACM, New York (2006)

47. Kammerer, Y., Nairn, R., Pirolli, P., Chi, E.H.: Signpost from the masses: learning effects in an exploratory social tag search browser. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, pp. 625–634. ACM, New York (2009)

48. Kaser, O., Lemire, D.: Tag-cloud drawing: algorithms for cloud visualization. arXiv preprint arXiv:0703109 (2007)

49. Kleinberg, J.: Navigation in a small world. Nature **406**(6798), 845 (2000)

50. Kleinberg, J.: The small-world phenomenon: an algorithmic perspective. In: Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing, pp. 163–170. ACM (2000)

51. Kleinberg, J.: Small-world phenomena and the dynamics of information. Adv. Neural Inf. Process. syst. **1**, 431–438 (2002)

52. Körner, C., Benz, D., Hotho, A., Strohmaier, M., Stumme, G.: Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 521–530. ACM, New York (2010)

53. Kuo, B.Y.L., Hentrich, T., Good, B.M., Wilkinson, M.D.: Tag clouds for summarizing web search results. In: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, pp. 1203–1204. ACM, New York (2007)

54. Lancichinetti, A., Fortunato, S., Kertész, J.: Detecting the overlapping and hierarchical community structure in complex networks. New J. Phys. **11**(3), 033015 (2009)

55. Lee, B., Riche, N.H., Karlson, A.K., Carpendale, S.: Sparkclouds: visualizing trends in tag clouds. IEEE Trans. Vis. Comput. Graph. **16**(6), 1182–1189 (2010)

56. Li, R., Bao, S., Yu, Y., Fei, B., Su, Z.: Towards effective browsing of large scale social annotations. In: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, pp. 943–952. ACM, New York (2007)

57. Li, Z., Tian, J.: Testing the suitability of Markov chains as web usage models. In: Proceedings 27th Annual International Computer Software and Applications Conference, COMPSAC 2003, pp. 356–361. IEEE (2003)

58. Lin, Y.L., Brusilovsky, P., He, D.: Finding cultural heritage images through a dual-perspective navigation framework. Inf. Process. Manag. **52**(5), 820–839 (2016)

59. Lloyd, S.: Least squares quantization in PCM. IEEE Trans. Inf. Theory **28**(2), 129–137 (1982)

60. Lohmann, S., Ziegler, J., Tetzlaff, L.: Comparison of tag cloud layouts: task-related performance and visual exploration. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 392–404. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03655-2_43

61. Macgregor, G., McCulloch, E.: Collaborative tagging as a knowledge organisation and resource discovery tool. Libr. Rev. **55**(5), 291–300 (2006)

62. Mathes, A.: Folksonomies-cooperative classification and communication through shared metadata (2004). http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

63. Miao, G., Tao, S., Cheng, W., Moulic, R., Moser, L.E., Lo, D., Yan, X.: Understanding task-driven information flow in collaborative networks. In: Proceedings of the 21st International Conference on World Wide Web, WWW 2012, pp. 849–858. ACM, New York (2012)

64. Mika, P.: Ontologies are us: a unified model of social networks and semantics. Web Semant. Sci. Serv. Agents World Wide Web **5**(1), 5–15 (2007)

65. Millen, D.R., Feinberg, J.: Using social tagging to improve social navigation. In: Workshop on the Social Navigation and Community based Adaptation Technologies (2006)
66. Muchnik, L., Itzhack, R., Solomon, S., Louzoun, Y.: Self-emergence of knowledge trees: extraction of the Wikipedia hierarchies. Phys. Rev. E **76**(1), 016106 (2007)
67. Navarro Bullock, B., Hotho, A., Stumme, G.: Accessing information with tags: search and ranking. In: Brusilovsky, P., He, D. (eds.) Social Information Access. LNCS, vol. 10100, pp. 310–343. Springer, Cham (2018)
68. Pirolli, P.: An elementary social information foraging model. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, pp. 605–614. ACM, New York (2009)
69. Pirolli, P., Card, S.: Information foraging. Psychol. Rev. **106**(4), 643 (1999)
70. Pirolli, P., Fu, W.-T.: SNIF-ACT: a model of information foraging on the world wide web. In: Brusilovsky, P., Corbett, A., de Rosis, F. (eds.) UM 2003. LNCS (LNAI), vol. 2702, pp. 45–54. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-44963-9_8
71. Plangprasopchok, A., Lerman, K., Getoor, L.: From saplings to a tree: integrating structured metadata via relational affinity propagation. In: Proceedings of the AAAI Workshop on Statistical Relational AI, July 2010
72. Ramage, D., Heymann, P., Manning, C.D., Garcia-Molina, H.: Clustering the tagged web. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, WSDM 2009, pp. 54–63. ACM, New York (2009)
73. Rivadeneira, A.W., Gruen, D.M., Muller, M.J., Millen, D.R.: Getting our head in the clouds: toward evaluation studies of tagclouds. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 995–998. ACM, New York (2007)
74. Russell, T.: Cloudalicious: Folksonomy over time. In: Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2006, p. 364. ACM, New York (2006)
75. Schmitz, C., Hotho, A., Jäschke, R., Stumme, G.: Mining association rules in folksonomies. In: Batagelj, V., Bock, H.H., Ferligoj, A., Žiberna, A. (eds.) Data Science and Classification, pp. 261–270. Springer, Heidelberg (2006). https://doi.org/10.1007/3-540-34416-0_28
76. Schmitz, P.: Inducing ontology from flickr tags. In: Collaborative Web Tagging Workshop at WWW 2006, vol. 50, Edinburgh, Scotland (2006)
77. Seifert, C., Kump, B., Kienreich, W., Granitzer, G., Granitzer, M.: On the beauty and usability of tag clouds. In: 12th International Conference Information Visualisation, 2008, IV 2008, pp. 17–25. IEEE (2008)
78. Sinclair, J., Cardew-Hall, M.: The folksonomy tag cloud: when is it useful? J. Inf. Sci. **34**(1), 15–29 (2008)
79. Singer, P., Helic, D., Hotho, A., Strohmaier, M.: Hyptrails: a bayesian approach for comparing hypotheses about human trails on the web. In: Proceedings of the 24th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland WWW 2015, pp. 1003–1013 (2015)
80. Skoutas, D., Alrifai, M.: Tag clouds revisited. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM 2011, pp. 221–230. ACM, New York (2011)
81. Specia, L., Motta, E.: Integrating folksonomies with the semantic web. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 624–639. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72667-8_44

82. Trattner, C., Lin, Y., Parra, D., Yue, Z., Real, W., Brusilovsky, P.: Evaluating tag-based information access in image collections. In: Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT 2012, pp. 113–122. ACM, New York (2012)

83. Tufte, E.R.: Beautiful Evidence. Graphis Press, New York City (2006)

84. Venetis, P., Koutrika, G., Garcia-Molina, H.: On the selection of tags for tag clouds. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM 2011, pp. 835–844. ACM, New York (2011)

85. Viegas, F.B., Wattenberg, M., Feinberg, J.: Participatory visualization with wordle. IEEE Trans. Vis. Comput. Graph. **15**(6), 1137–1144 (2009)

86. Vig, J., Sen, S.: Computing the tag genome. Technical report, 10 September 2010

87. Vig, J., Sen, S., Riedl, J.: Navigating the tag genome. In: Proceedings of the 16th International Conference on Intelligent User Interfaces, IUI 2011, pp. 93–102. ACM, New York (2011)

88. Vig, J., Sen, S., Riedl, J.: The tag genome: encoding community knowledge to support novel interaction. ACM Trans. Interact. Intell. Syst. (TiiS) **2**(3), 13 (2012)

89. Wagner, C., Singer, P., Strohmaier, M., Huberman, B.A.: Semantic stability in social tagging streams. In: Proceedings of the 23rd International Conference on World Wide Web, WWW 2014, pp. 735–746. ACM, New York (2014). https://doi.org/10.1145/2566486.2567979

90. Watts, D.J., Dodds, P.S., Newman, M.E.: Identity and search in social networks. Science **296**(5571), 1302–1305 (2002)

91. Watts, D.J., Strogatz, S.H.: Collective dynamics of small-worldnetworks. Nature **393**(6684), 440–442 (1998)

92. West, R., Leskovec, J.: Human wayfinding in information networks. In: Proceedings of the 21st International Conference on World Wide Web, WWW 2012, pp. 619–628. ACM, New York, (2012)

93. Zhong, S.: Efficient online spherical k-means clustering. In: Proceedings, 2005 IEEE International Joint Conference on Neural Networks, IJCNN 2005, vol. 5, pp. 3180–3185. IEEE (2005)

94. Zubiaga, A.: Enhancing navigation on wikipedia with social tags. CoRR abs/1202.5469 (2012). http://arxiv.org/abs/1202.5469

95. Zubiaga, A., García-Plaza, A.P., Fresno, V., Martínez, R.: Content-based clustering for tag cloud visualization. In: Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2009, pp. 316–319. IEEE Computer Society, Washington (2009)