On the Visualization of Hierarchical Relations and Tree Structures with TagSpheres

Stefan Jänicke(✉) and Gerik Scheuermann

Image and Signal Processing Group, Leipzig University, Leipzig, Germany
{stjaenicke,scheuermann}@informatik.uni-leipzig.de

Abstract. Tag clouds are widely applied, popular visualization techniques as they illustrate summaries of textual data in an intuitive, lucid manner. Many layout algorithms for tag clouds have been developed in the recent years, but none of these approaches is designed to reflect the notion of hierarchical distance. For that purpose, we introduce a novel tag cloud layout called *TagSpheres*. By arranging tags on various hierarchy levels and applying appropriate colors, the importance of individual tags to the observed topic gets assessable. To explore relationships among various hierarchy levels, we aim to place related tags closely. Various usage scenarios from the digital humanities, sports, aviation and natural disaster management point out the benefit of TagSpheres for different domains. In addition, we highlight that TagSpheres is also a novel layout approach for tree structures.

Keywords: Tag clouds · Text visualization · Hierarchical data · Tree layout

1 Introduction

The usage of tag clouds to visualize textual data is a relatively novel technique, which was rarely applied in the past century. In 1976, Stanley Milgram was one of the first scholars who generated a tag cloud to illustrate a mental map of Paris, for which he conducted a psychological study with inhabitants of Paris, aiming to analyze their mental representation of the city [1]. In 1992, a German edition of "Mille Plateaux", written by the French philosopher Gilles Deleuze, was published with a tag cloud printed on the cover to summarize the book's content [2]. This idea to present a visual summary of textual data can be seen as the primary purpose of tag clouds [3]. But the popularity of tag clouds nowadays is attributable to a frequent usage in the social web community in the 2000s as overviews of website contents. Although there are known theoretical problems concerning the design of tag clouds [4], they are generally seen as a popular social component perceived as being fun [5]. With the simple idea to encode the frequency of terms to a given topic, tag clouds are intuitive, comprehensible visualizations, which are widely used metaphors (1) to display summaries of textual

J. Braz et al. (Eds.): VISIGRAPP 2016, CCIS 693, pp. 199–219, 2017.
DOI: 10.1007/978-3-319-64870-5_10

data, (2) to support analytical tasks such as the examination of text collections, or even (3) to be used as interfaces for navigation purposes on databases.

In the recent years, various algorithms that compute effective tag cloud layouts in an informative and readable manner have been developed. One of the most popular techniques is Wordle [6], which computes compact, intuitive tag clouds and can be generated on the fly using a web-based interface.[1] Although the produced results are very aesthetic, the different used colors do not transfer information and the final arrangement of tags depends only on the scale, and not on the content of tags or potential relationships among them. Some approaches attend to the matter of visualizing more information than the frequency of terms with tag clouds – most often to compare textual summaries of different data facets.

In this paper, we present the tag cloud design *TagSpheres*, which endeavors to effectively visualize hierarchies in textual summaries. The motivation arose from research on philology. Humanities scholars wanted to analyze the clause functions of an ancient term's co-occurrences. Querying the large database, the scholars often face numerous results in the form of text passages. When only plain lists are provided to interact with the results, the discovery of significant text passages and the analysis of the contexts in which the chosen term was used becomes laborious. To support this task, we provide summaries of text passages in the form of interactive tag clouds that group terms in accordance to their distance to the search term. So, the humanities scholar gets an overview, and she is able to retrieve text passages of interest on demand.

We designed TagSpheres in a way that various types of text hierarchies can be visualized in an intuitive, comprehensible manner. To emphasize the wide applicability of TagSpheres, we list several examples from the digital humanities, sports, aviation and natural disaster management. That TagSpheres can be further used to generate layouts for tree structures is outlined in Sect. 5.

2 Related Work

Although tag clouds rather became popular in the social media, research in visualization attended to the matter of developing various layout techniques in the last years. A basic tag cloud layout is a simple list of words placed on multiple lines [7]. In such a list, tags are typically ordered by their importance to the observed issue, which is encoded by font size [8]. An alphabetical order is also often used, but a study revealed that this order is not obvious for the observer [5]. Later, more sophisticated tag cloud layout approaches that rather emphasize aesthetics than meaningful orderings were developed. A representative technique is Wordle [6,9], which produces compact aesthetic layouts with tags in different colors and orientations, but both features do not transfer any additional information. A Wordle showing the most important terms in Edgar Allan Poe's *The Raven* is given in Fig. 1(a).

Various approaches highlight relationships among tags by forming visual groups. In thematically clustered or semantic tag clouds, the detection of tags

[1] http://www.wordle.net/.

(a) Wordle of Edgar Allan Poe's *The Raven*.

(b) Comparing the co-occurrences of three Latin terms with TagPies.

Fig. 1. Tag cloud visualizations.

belonging to the same topic is supported by placing these tags closely [10]. Traditional, semantic word lists place clustered tags successively [11]. More sophisticated layout methods often use force directed approaches with semantically close terms attracting each other [12–14]. After force directed tag placement, tag cloud layouts can be compressed by removing occurring whitespaces [15].

Some methods generate individual tag clouds for each group of related tags, and combine the resultant multiple tag clouds to a single visual unity afterwards. An example is the Star Forest method [16], which applies a force directed method to pack multiple tag clouds. Other approaches use predefined tag cloud containers, e.g., user-defined polygonal spaces in the plane [17], polygonal shapes of countries [18], or Voronoi tesselations [19]. Newsmap uses a treemap layout [20] to group newspaper headlines of the same category in blocks [21]. Morphable Word Clouds morph the shapes of tag cloud containers in order to visualize temporal variance in text summaries [22]. For the comparison of the tags of various text documents, a ConcentriCloud divides an elliptical plane into sectors that list shared tags of several subsets of the underlying texts [23]. Due to the rather independent computation of individual tag clouds – which often leads to large whitespaces in the final composition step – the above mentioned methods can be seen as sophisticated small multiples. A rather traditional small multiples approach is Words Storms [24] that supports the visual comparison of textual summaries of documents.

Tag clouds also have been used to visualize trends. Parallel Tag Clouds generate alphabetically ordered tag lists as columns for a number of time slices and highlight the temporal evolution of a tag placed in various columns on mouse interaction [25]. In contrast, SparkClouds attach a graph showing the tag's evolution over time [26]. Other approaches overlay time graphs with tags characteristic for certain time ranges [27].

Only few approaches generate multifaceted tag cloud layouts in a single, continuous flow that includes the positioning of all tags belonging to various groups. RadCloud visualizes tags belonging to various groups within a shared elliptical area [28]. In Compare Clouds, tags of two media frames (MSM, Blogs) are comparatively visualized in a single cloud [29]. To support the comparative analysis of multiple tag groups, TagPies are arranged in a pie chart manner [30]. An example showing the comparative visualization of the co-occurrences of Latin terms is shown in Fig. 1(b).

Table 1. Characteristics of usage scenarios for TagSpheres.

Domain	Digital humanities (see Sect. 4.1)	Sports (see Sect. 4.2)	Aviation (see Sect. 4.3)	Natural disaster management (see Sect. 4.4)
Task	Analyzing the clause functions of the co-occurrences of a search term T	Comparing the performances of teams in championships	Observing all direct flights from an airport or a city	Exploring the risks of natural disasters of countries (World Risk Index 2015)
H_1	Search term T	Best performing teams	Departure airport/city	Countries with very high disaster risks
$H_2, \ldots, H_n$	Co-occurrences in dependency on the word distance to T	Teams grouped by decreasing performance	Federal (H_2), continental (H_3) and worldwide flights (H_4)	Countries with decreasing disaster risks
n	4	5, 6, 8	2..4	5
$w(t)$	Number of (co-)occurrences of t	Number of a team's appearances	Inverse distance weighting between departure and arrival airports/cities	A country's disaster risk percentage
$p(t)$	Equally labeled tag of a higher hierarchy level	Same team if already placed on a higher hierarchy level	Previously placed tag of the same country/continent	Previously placed tag of the same continent
Strong tag relations	Equally labeled tags	Same teams if placed on multiple hierarchy levels	Departure/arrival airports/cities	N/A
Weak tag relations	Spelling variants	N/A	Airports/cities of the same country/continent	Countries of the same continent

Although techniques like TagPies or Parallel Tag Clouds are capable of visualizing sequences of tag groups, none of the mentioned approaches endeavors to visually encode generic hierarchical information intuitively in a single, compact, aesthetic tag cloud. TagSpheres – presented in this paper – are designed to fill this gap.

3 Designing TagSpheres

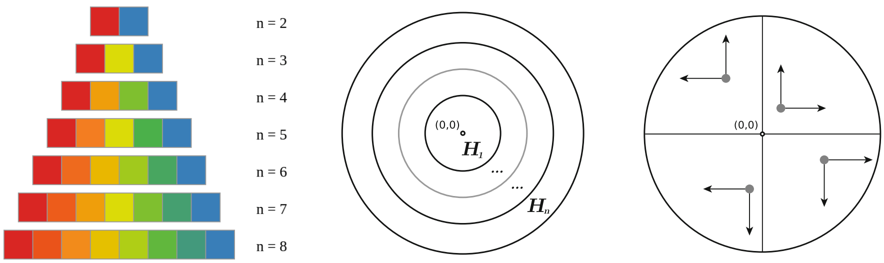
The central idea of TagSpheres is the visualization of textual summaries that comprise hierarchical information. This paper provides various usage scenarios that exemplify the existence of hierarchies in textual data (see Sect. 4). An overview of the characteristics of these examples is given in Table 1.

Given n hierarchy levels $H_1, \ldots, H_n$, the top hierarchy level H_1 contains tags representing the focus of interest of a usage scenario. All other tags are divided into $n-1$ groups in dependency on their hierarchical distance according to the observed topic, or to the tags on H_1. Each tag t in TagSpheres has a weight $w(t)$ reflecting its importance, and an optional predecessor tag $p(t)$ representing a relationship to another tag that was placed before t and usually belongs to a higher hierarchy level. In dependency on the observed topic, it might be necessary to place the same tag on several hierarchy levels to encode the change of a tag's importance among hierarchies. In such cases, predecessor tags help to visually link these tags.

3.1 Design Decisions

When designing TagSpheres, we use the following, well-established design features for tag clouds:

- **Font Size:** Evaluated as the most powerful property [31], font size encodes the weight $w(t)$ of a tag.
- **Orientation:** As rotated tags are perceived as "unstructured, unattractive, and hardly readable" [32], we do not rotate tags to keep the layout easily explorable.
- **Color:** Being the best choice to distinguish categories [32], various colors are assigned to tags belonging to different hierarchy levels. As TagSpheres encode the distance to a given topic, the usage of a categorial color map is inappropriate. Unfortunately, suitable sequential color maps as provided by the ColorBrewer [33] produce less distinctive colors even for a small number of hierarchy levels, so that adjacent tags belonging to different hierarchy levels are hard to classify. Following the suggestions given by Ware [34], we defined a divergent cold-hot color map using red for the first hierarchy level and blue for tags belonging to the last hierarchy level n. To avoid uneven visual attraction of tags, we only chose saturated colors that are in contrast to the white background. Example color maps for up to eight hierarchy levels are shown in Fig. 2(a).

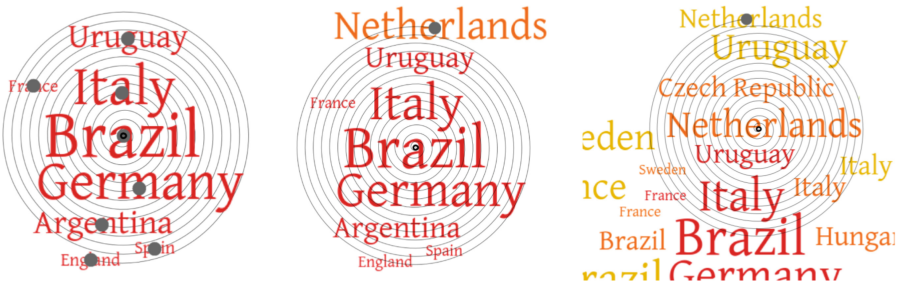


(a) Resultant color maps for $n = 2, \dots, 8$ hierarchy levels.

(b) Using spheres for the tags of different hierarchy levels.

(c) Vectors for occlusion check to guarantee hierarchical coherence.

Fig. 2. TagSpheres layout algorithm details.



(a) Placing all tags of H_1.

(b) Placing a tag without predecessor.

(c) Placing a tag with predecessor.

Fig. 3. Determining tag positions using an Archimedean spiral.

3.2 Layout Algorithm

In preparation, the tags are sorted by increasing hierarchy level, so that all tags within the same hierarchical distance to H_1 are placed successively. The tags of each hierarchy level are ordered by decreasing weight to ensure that important tags are circularly well distributed.

To avoid large whitespaces, a problem addressed by Seifert et al. [35], our method follows the idea of the Wordle algorithm [6] – permitting overlapping tag bounding boxes if the tags' letters do not occlude – to determine the positions of tags. So, we obtain compact, uniformly looking tag clouds for the underlying hierarchical, textual data. To ensure well readable tag clouds, we use a minimal padding between letters of different tags.

As shown in Fig. 2(b), we aim to visually compose tags of the same hierarchy level in the form of spheres around the tag cloud origin at $(0, 0)$. Initially, we iteratively determine positions for the tags of H_1 in the central sphere using an Archimedean spiral originating from $(0, 0)$. An example is given in Fig. 3(a). For each tag t of the remaining hierarchy levels $H_2, \dots, H_n$, we also use $(0, 0)$ as spiral origin, if $p(t)$ is not provided (see Fig. 3(b)). If $p(t)$ is defined, we use the predecessor's position as spiral origin (see Fig. 3(c)). As a consequence, hierarchically related tags are placed closely and visually compose in the form of rays originating from $(0, 0)$ as shown in Fig. 6(a). In contrast to other spiral based tag cloud algorithms, we avoid to cover whitespaces with tags of hierarchy level H_i within spheres of already processed hierarchy levels $H_1, \ldots, H_{i-1}$. Dependent on the quadrant in the plane, in which a tag shall be placed, we search for already placed tags intersecting two vectors originating from the dedicated position as illustrated in Fig. 2(c). If no intersections are found, we place the tag. This approach coheres all tags of a hierarchy level as a visual unity outside the inner bounds of the previously processed hierarchy levels' spheres.

3.3 Interactive Design

Implemented as an open source JavaScript library,[2] TagSpheres can be dynamically embedded into web-based applications. With mouse interaction, we enable the user to detect hierarchically related tags quickly. Thereby, we distinguish between strongly and weakly related tags, which are defined in dependency on the underlying usage scenario (see Table 1). Related tags are shown on mouseover (see Fig. 4). For strongly related tags we use a black font on transparent backgrounds having the hierarchy level's assigned color. In contrast, weakly related tags retain their saturated font color, but gray, transparent backgrounds indicate relationships.

TagSpheres provide a configurable tooltip displayed when hovering or clicking a tag to be used, e.g., to list all related tags and their weights. The mouse click function can be used for displaying additional information. e.g., to link to external sources, or to show text passages containing the chosen tag.

3.4 Limitations

The main objective of the presented layout algorithm is to combine a hierarchical information of textual data with the aesthetics of tag clouds. In contrast to the usual approach to always initialize an Archimedean spiral at the tag cloud origin $(0, 0)$ when determining the position of a tag, the usage of predecessor tags as spiral origins slightly affects the uniform appearance of the result in some cases (e.g., see Fig. 7). Occasionally, little holes occur, and – at the expense of visualizing the hierarchical structure of the underlying data – the tag cloud boundaries get distorted.

The proposed hot-cold color map used to visually convey hierarchical distance generates well distinguishable colors when the number of hierarchy levels is small. For a larger number of hierarchies as displayed in Fig. 6(c) or Fig. 10, closely positioned tags of different levels may become visually indistinct, especially when only few tags belong to a certain level.

The current TagSpheres design does not take the distribution of tags throughout different hierarchy levels into account. In use cases with a steadily increasing

[2] http://tagspheres.vizcovery.org.

or decreasing number of tags per hierarchy level it gets possible that a considerable proportion of the color maps' bandwidth is used for a comparatively small portion of tags. An assignment of colors taking the density distribution of the tags' weights into account could overcome this issue.

4 Use Cases

TagSpheres are applicable whenever statistics of unstructured text shall be visualized in the form of a tag cloud and a decent hierarchy among the tags exists. This section illustrates usage scenarios of TagSpheres for text-based data from four different domains: digital humanities, sports, aviation and natural disaster management.

4.1 Digital Humanities Scenario

Within the digital humanities project *eXChange*,[3] historians and classical philologists work with a database containing a large amount of digitized historical texts in Latin. Usually, humanities scholars pose keyword based search queries and often receive numerous results, which are hard to revise individually. As a consequence, the generation of valuable hypotheses is a laborious, time-consuming process. To facilitate the humanities scholars' workflows, we develop visual interfaces that attempt to steer the analysis of search results into promising directions.

TagPies – also developed within the *eXChange* project – are tag clouds arranged in a pie chart manner that support the comparison of multiple search query results [30]. Using a TagPie, humanities scholars analyze contextual similarities and differences of the observed terms – an example is given in Fig. 1(b). Whereas the tags of the same groups are placed in the same circular sectors in TagPies to support their comparative analysis, the intention of TagSpheres is the visualization of hierarchical information. This supports approaching a further research interest of the humanities scholars: the analysis and classification of a term's co-occurrences according to their clause functions. For this purpose, the scholars require four-level TagSpheres displaying the following tags:

H_1: search term T,
H_2: co-occurrences of T with word distance 1,
H_3: co-occurrences of T with word distance 2, and
H_4: co-occurrences of T with word distance 3 up to word distance m.

The font size of T on level H_1 encodes how frequent the search term occurs in the underlying text corpus; the font sizes of all other terms reflect their number of co-occurrences with T in dependency on the corresponding distance. On H_4, font sizes are normalized in relation to the distance range $m - 2$. A tag on hierarchy level H_i receives a predecessor tag if the corresponding term occurs on one of the previous layers $H_{i-1}, \ldots, H_1$.

[3] http://exchange-projekt.de/.

Two use cases provided by the humanities scholars involved in the *eXChange* project shall illustrate the utility of TagSpheres to support the classification of a term's co-occurrences by their clause functions.

The **first use case** (see Fig. 4) outlines the analysis of the co-occurrences of the Latin term ***morbo*** (disease). The humanities scholar discovered and classified terms in similar relationships to the given topic. In large distances, the humanities scholar found objects in form of affected parts of the body, e.g., head (*caput*), soul (*animo*) and limbs (*membrorum*), affected persons, e.g., son (*filius*), woman (*mulier*) and king (*rex*), and related places, e.g., Rome (*romam*), church (*ecclesia*) and *villa*. Closer to *morbo* (most often with distance 1 or 2), typical attributes and predicates can be found. Whereas attributes describe the type or intensity of the disease, e.g., pestilential (*pestifero*), heavy (*gravi*), deadly (*exitiali*) and acute (*acuto*), the occurring predicates illustrate the disease's progress, e.g., seize (*correptus*), dissappear (*periit*) and worsening (*ingravescente*). Adjacent to *morbo*, specific terms for "moral" diseases, e.g., greediness (*avaritiae*), arrogance (*superbiae*) and lust (*concupiscentiae*), and actual

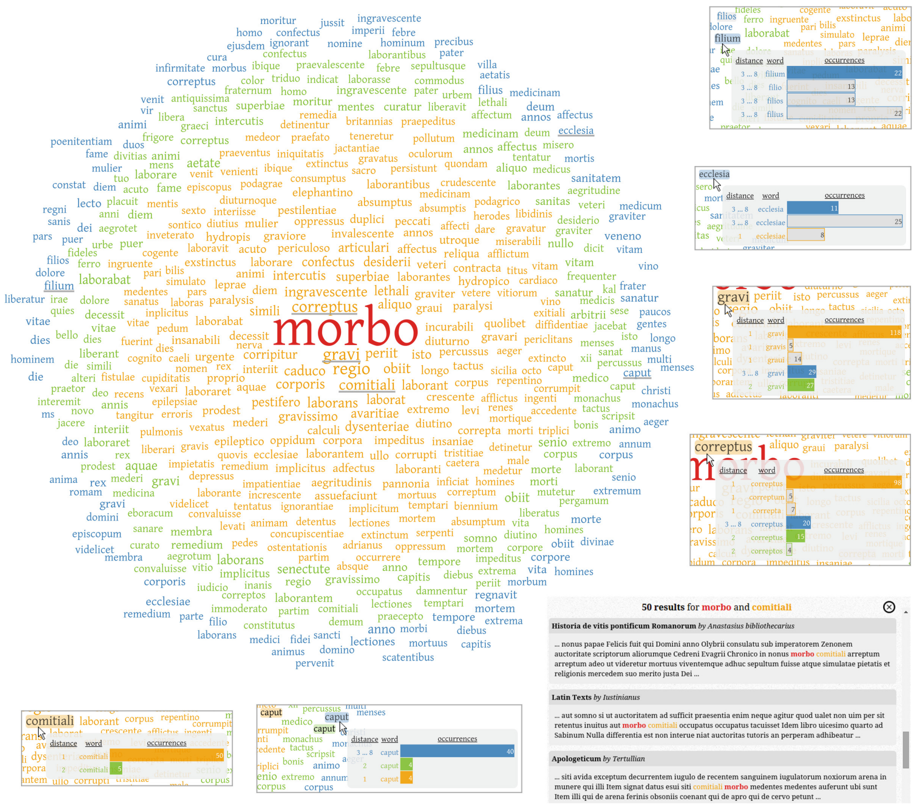


Fig. 4. The analysis of co-occurrences of the Latin term *morbo* (disease) on word distance.

diseases like jaundice ([morbo] *regio*), leprosy (*leprae*) and two common names for epilepsy (*[morbo] comitiali, [morbo] sacro*) occur.

The **second use case** (see Fig. 5) illustrates the exploration of the co-occurrences of the Latin term ***vino*** (vine). Like in the previous example, attributes of vine like precious (*pretioso*), sweet (*dulci*), new (*novo*), good (*bono*), white (*albo*) or "the best" (*optimo*) co-occur next to *vino*. Also closely positioned, usually with distance 1 or 2, are verbs describing (1) what people do with vine, e.g., drink (*postati, bibitur*), mix (*miscetur*) or swill (*lavabit*), and (2) what vine does to people, e.g., inebriate (*inebriatus, crapulatus*), rave (*furere*) or degenerate (*degenerantes*). In larger distances, subjects associated with vine can be found, e.g., people (*homines, populus*), saints (*sancti*), lord (*dominus*) or drunks (*ebrii*). Rather unexpected was the dominant usage of *vino* in Christian texts – visible through co-occurring terms like bread (*panem*), blood (*sanguis*), Body of Christ (*corpus, christi*) or sacrifice (*sacrificium*) – in contrast to a less frequent usage in classical texts. But, the humanities scholar stated that the visualization vividly reflects the classical tricolon "*vino* – *frumento* (grain) – *oleo* (oil)" as a list of important groceries in antiquity for soldiers to survive.

In this usage scenario, the interaction capabilities of TagSpheres are tailored according to the needs of the humanities scholars. Hovering a tag opens a tooltip showing the term's number of occurrences on all hierarchy levels as strongly related tags. Additionally, variant spellings or cases of the term are listed with their corresponding frequencies as weakly related tags to support the analysis process. An important requirement for the humanities scholars was the discovery of potentially interesting text passages, but they desired a straightforward access to the underlying texts in general. This so-called *close reading* is often reported as an important component when designing visualizations for humanities scholars [36]. TagSpheres support close reading by clicking a tag, which displays the corresponding text passages containing the search term and the clicked term with the chosen distance. For the first use case (Fig. 4, bottom right), text passages containing the terms *morbo* and *comitiali* are shown. In the second use case (Fig. 4, bottom right), we see text passages containing *vino* and *frumento*.

4.2 Championship Performances

This scenario illustrates how TagSpheres can be used to comparatively visualize performances in championships. Therefore, we processed a dataset containing the results of all national teams ever qualified for the FIFA World Cup. We receive the following six-level hierarchy:

H_1: FIFA World Champions,
H_2: second placed national teams,
H_3: national teams knocked out in the semifinal,
H_4: national teams knocked out in the quarterfinal,
H_5: national teams knocked out in the second round, and
H_6: national teams knocked out in the (first) group stage.

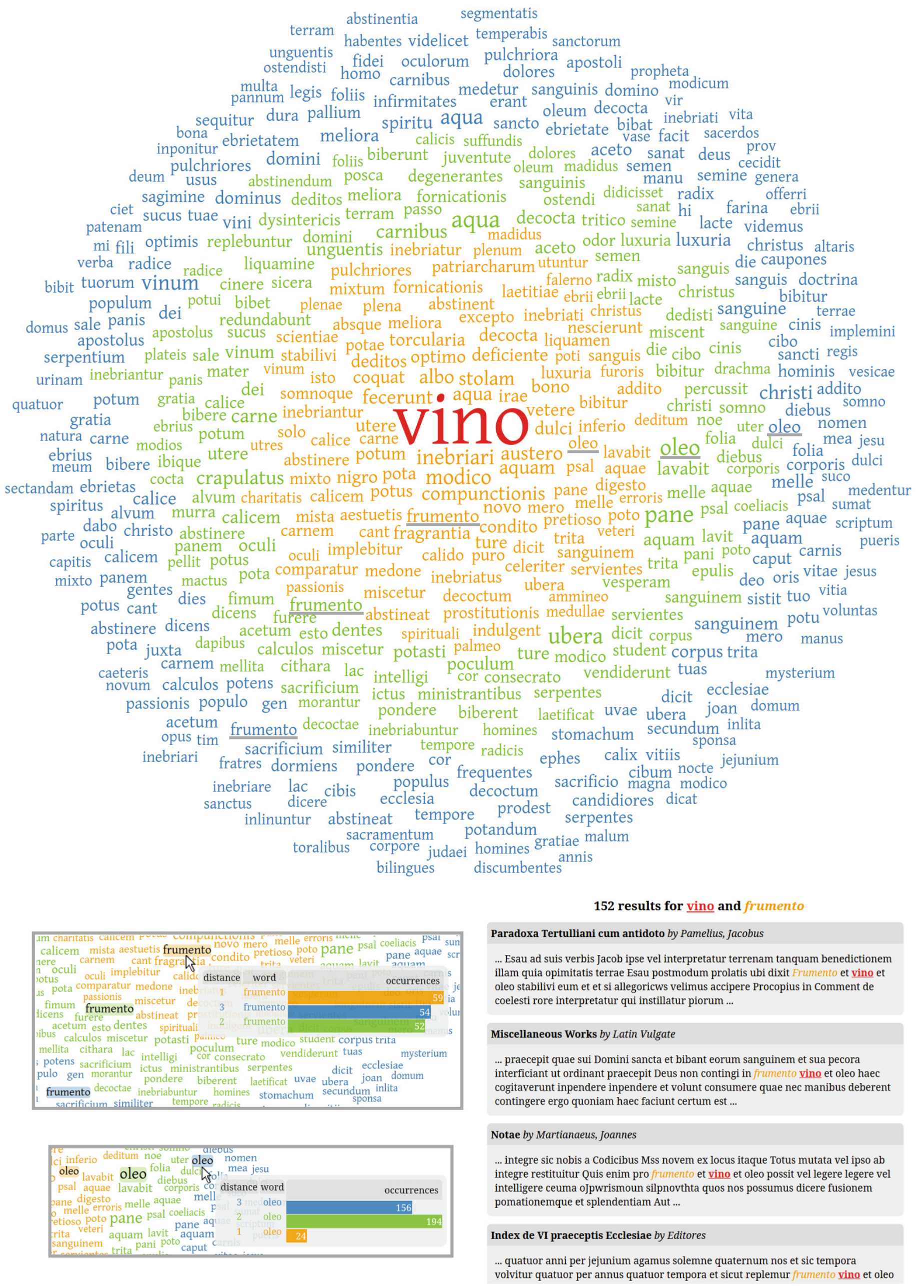


Fig. 5. The analysis of co-occurrences of the Latin term *vino* (vine) on word distance.

The nations' names are used as tags and font size encodes how often a national team partook a championship round *without* reaching the next level. Therefore, most nations occur on various hierarchy levels. If a tag t for a nation to be placed on H_i was already placed at a higher hierarchy level $H_{i-1}, \ldots, H_1$, we use the corresponding tag as predecessor $p(t)$.

Figure 6(a) shows the resultant TagSpheres. Especially this scenario illustrates the benefit of using the positions of predecessor tags as spiral origins for successor tags. In most cases, the various tags of a nation are closely positioned. Hovering a tag displays the all-time performance of the corresponding national team for all championship rounds in a tooltip. Expectedly, *Brazil* and *Germany* achieved very good results, especially in the last championship rounds. In contrast, *Italy* was often knocked out in the first round, but in case of reaching the semifinal (8x), *Italy* often became FIFA World Champion (4x). Few nations have a 100% success rate in the group stage. Qualified three times for the FIFA World Cup, *Senegal* always reached the quarterfinals. Most nations, e.g., *Sweden*, show the expected pattern "the higher the championship round, the smaller the number of appearances".

Analogously to the FIFA World Cup results, Fig. 6(b) illustrates the performances of all national teams ever participated the UEFA European Championship – pointing out Germany and Spain as most successful nations. Another example is given in Fig. 6(c) that illustrates the success of football clubs ever played in England's first league. Here, we use the average rank at the end of the seasons to cluster 68 clubs into eight hierarchy levels, and font size encodes the number of appearances.

4.3 Airport Connectivity

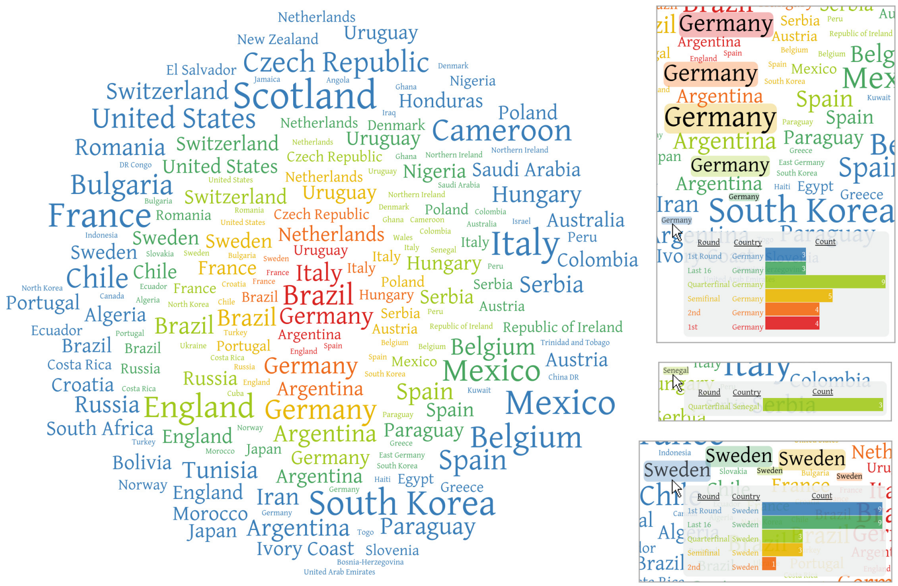
To analyze the federal, continental and worldwide connectivity of airports, we derived a dataset from the OpenFlights database,[4] which provides a list of direct flight connections between around 3,200 airports worldwide. With the selected departure airport d (or city) on H_1, all other airports (or cities) reachable with a non-stop flight cluster into three further hierarchy levels:

H_2: airports/cities in the same country as d,
H_3: airports/cities on the same continent as d, and
H_4: all other reachable worldwide airports/cities.

As tags we chose either airport names, the provided IATA codes,[5] or the corresponding city names. In this scenario, font size encodes the inverse geographical distance between the departure airport $d = \{lat_d, lon_d\}$ and an arrival airport $a = \{lat_a, lon_a\}$. To keep the deviation to the actual distance as small as possible, we apply the great circle distance G [37], defined as

$$G = 6378 \cdot \arccos\Big(\sin(lat_d) \cdot \sin(lat_a) + \cos(lat_d) \cdot \cos(lat_a) \cdot \cos(lon_d - lon_a)\Big).$$

[4] http://openflights.org/data.html.
[5] http://www.iata.org/services/pages/codes.aspx.



(a) Performances of all nations qualified for the FIFA World Cup.

(b) Performances of all nations qualified for the UEFA European Championship.

(c) Performances of English first league football clubs from 1888/89 – 2014/15.

Fig. 6. Visualizing championship performances with TagSpheres.

Predecessor tags are used to place airports or cities of the same country or continent closely. For a tag t to be placed on H_3, we choose the first placed tag with the same associated country as predecessor, if existent; for H_4, we choose the first placed tag with the same associated continent. Thus, a predecessor tag $p(t)$ in this scenario always belongs to the same hierarchy level as t.

Figure 7 shows TagSpheres for non-stop flights from various airports or cities. All examples show that airports/cities of the same countries/continents are placed closely in clusters. For Sydney, no tags are placed on H_3, and for Cagliari, no flight connections to airports outside Europe exist. When the user hovers a tag, the corresponding connection and the travel distance are shown in a tooltip. Clicking a tag redirects to Google Flights[6] listing possible flight connections.

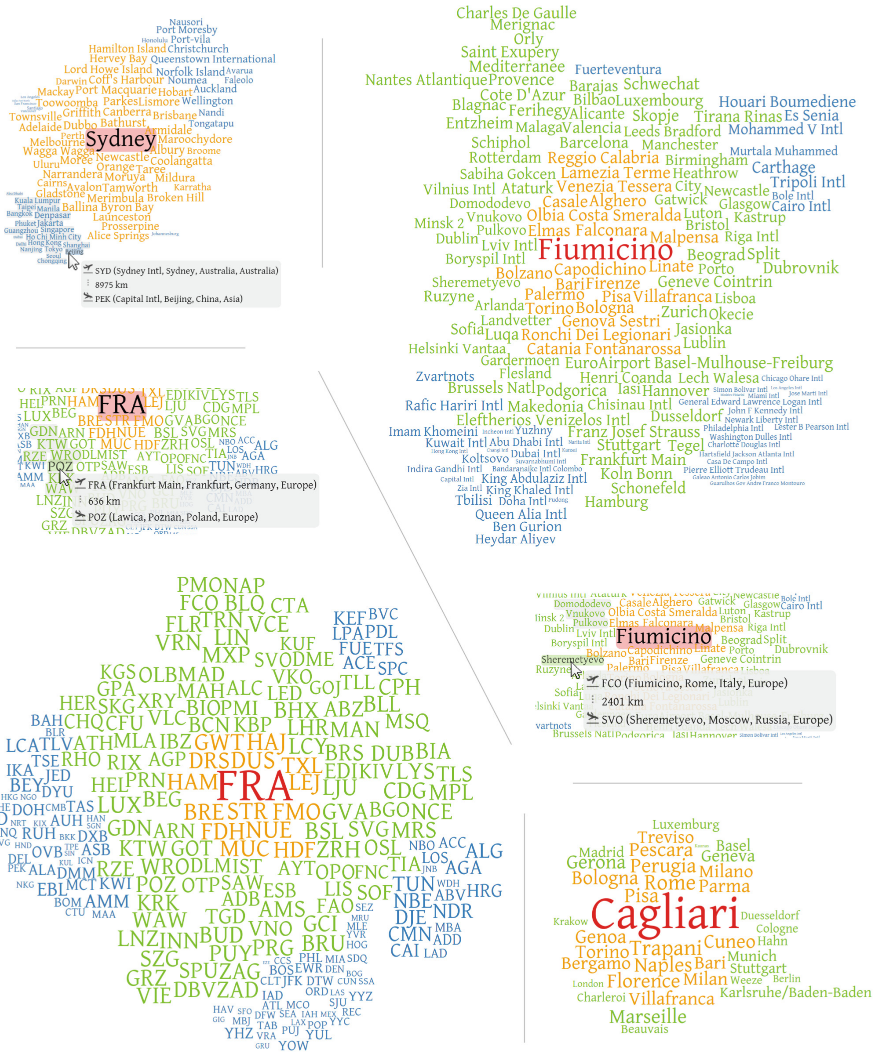


Fig. 7. Direct flight connections from airports in Sydney, Rome, Frankfurt and Cagliari.

6 https://www.google.com/flights/.

Fig. 8. World Risk Index 2015 visualized with TagSpheres.

4.4 World Risk Index 2015

The World Risk Report[7] analyzes disaster risks of countries. Thereby, the exposure of a country towards natural hazards (e.g., earthquakes, tsunamis, cyclones or floods) is compared to the country's vulnerability, which depends on living conditions and economic circumstances. Each country in the database receives a disaster risk percentage – Vanatu being the country with the highest risk (36.72%) and Qatar the country with the lowest risk (0.08%). All countries are clustered into five classes from *very high* to *very low* disaster risk, which are used to generate a thematic map[8] with countries colored according to these classes. The World Risk Index 2015 visualized with TagSpheres (see Fig. 8) uses the disaster risk classes as hierarchy levels:

H_1: countries with very high disaster risks (10.40%–36.72%),
H_2: countries with high disaster risks (7.31%–10.39%),
H_3: countries with medium disaster risks (5.47%–7.30%),
H_4: countries with low disaster risks (3.47%–5.46%), and
H_5: countries with very low disaster risks (0.08%–3.46%).

[7] http://www.worldriskreport.org/.
[8] http://tinyurl.com/htkw8h8.

In contrast to a thematic map, we highlight the actual, individual disaster risk percentage of each country with font size. To approximate geographical relations, we use predecessor tags to place country names belonging to the same continent closely.

5 Visualizing Tree Structures with TagSpheres

Numerous algorithms have been developed to visualize large tree structures [38]. Usually, explicit tree representations in the form of node-link diagrams focus on highlighting branching patterns, e.g., [39,40]; the visualization of values associated with individual nodes plays only a minor role. On the other hand, in implicit tree representations, e.g., tree maps [20], bubble charts [41] or pie chart variants [42], links are not drawn but hierarchical relationships between nodes are illustrated with nesting techniques. But, only few implicit tree layout algorithms communicate the actual values of nodes [43,44].

Applying the TagSpheres algorithm to tree structures yields an implicit node-link diagram that visualizes the values of nodes without explicitly displaying links. But, TagSpheres indicate structural relationships by using the parent of a node in the tree as predecessor tag, by applying variable font size to illustrate the number of a node's children, and by using the interaction functionality to highlight individual paths in the tree. This way, we gain a novel tree layout that rather favors the representation of nodes than links. Two examples presenting tree layouts generated with TagSpheres are outlined below.

5.1 Airport Connectivity

Using the OpenFlights database, we can construct a (minimum spanning) tree that reflects all possible flight connections from a selected departure airport d. As in Sect. 4.3, d is the only tag on hierarchy level H_1. All other hierarchy levels compose in dependency on the number of stops it takes to reach another airport. So, H_2 contains all airports reachable with a non-stop flight, H_3 contains all airports reachable with one stop, and so on. As the maximum number of stops is six, we get eight hierarchy levels. In case of multiple possible flight connections having the same number of stops when traveling between two airports, we keep the connection with the shortest geographical distance. Thus, each airport has a clearly defined predecessor. The resultant TagSpheres with Rome-Fiumicino (FCO) as departure airport is shown in Fig. 9. As the underlying tree is well balanced and the average number of children (outdegree) is relatively high (around 5.2 children per inner node), structural relationships are only faintly visible in the outer spheres. Paths are shown on mouse selection indicating the stops between d and the selected airport as well as available connecting flights to other airports. In contrast to other node-link diagrams, the values of all 3.228 nodes and their distances to the root node are easily recognizable with TagSpheres. Thereby, the font size of a tag reflects the number of connecting flights of the corresponding airport.

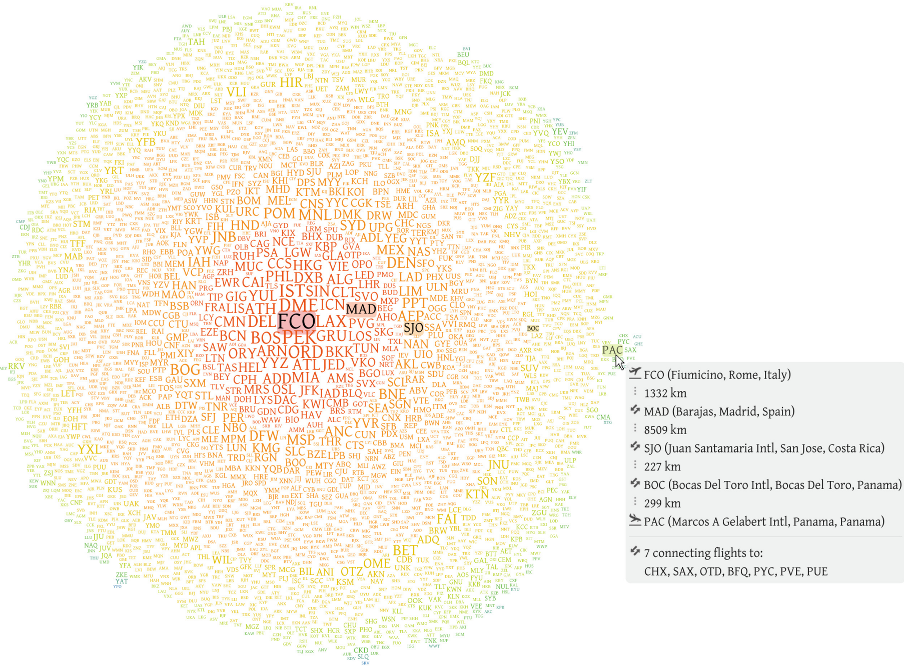


Fig. 9. Flight connections from Rome-Fiumicino (FCO).

5.2 Bible Family Tree

More than 600 verses of the Bible describe familial relationships, e.g., between husbands and wives or between fathers and children. Tying all these information together results in the Bible family tree.[9] It contains 416 nodes (persons), the maximum depth of the tree is 74, and the average number of children of inner nodes is 1.7. Using a vertical dendrogram layout[10] supports the analysis of global structural relationships, but the values of nodes are only locally visible. With TagSpheres, the values of all nodes are readable in the global view. In contrast to the previous example, the sparseness of the tree and scaling the font size according to the outdegree of a node fairly indicate present relationships, which can be further explored with mouse interaction.

[9] Bible genealogy data taken from BibleFamilyTree.info, Copyright © 2013 by The Psalm 119 Foundation. Used by permission. (http://www.ThePsalm119Foundation.org).

[10] http://biblefamilytree.info/.

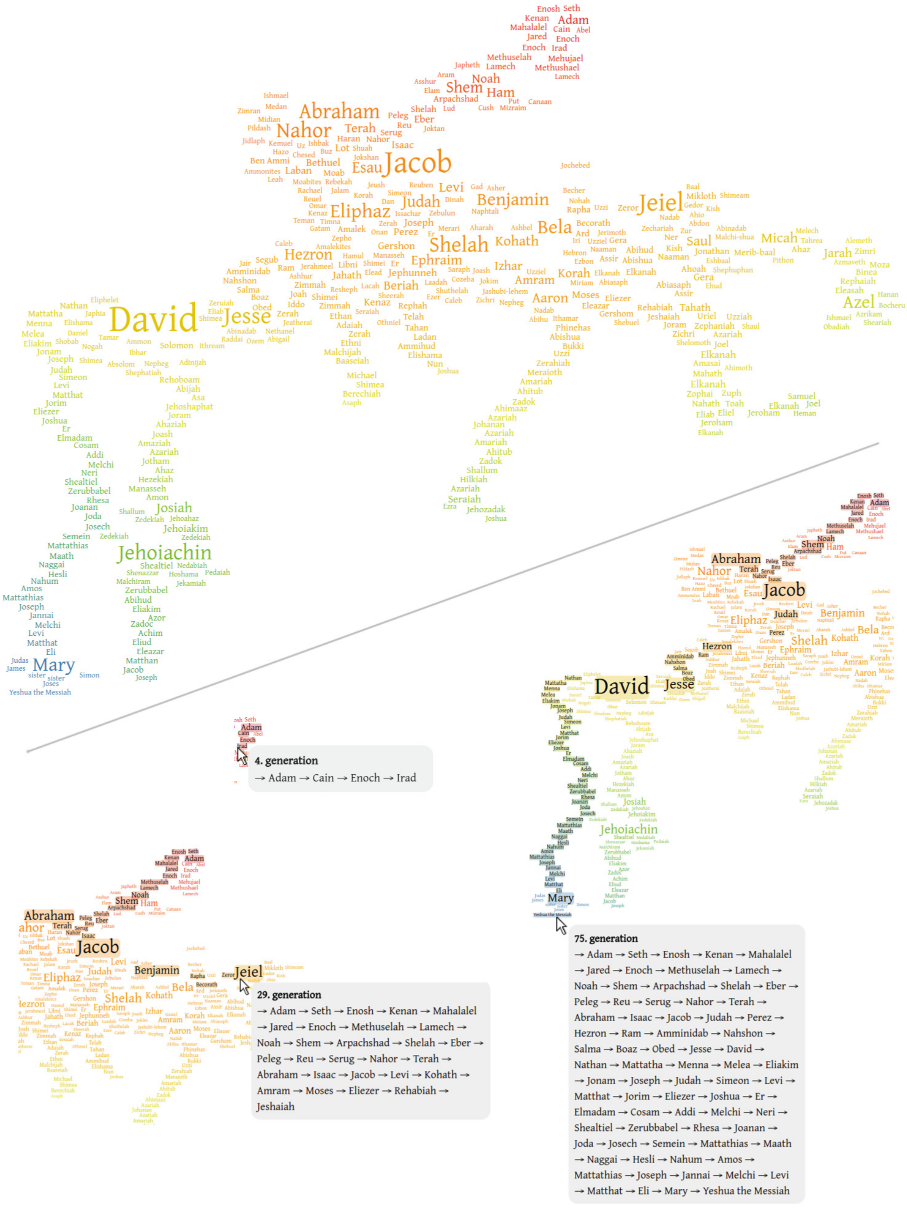


Fig. 10. Bible family tree visualized with TagSpheres.

6 Conclusion

We introduced TagSpheres that arrange tags on several hierarchy levels to transmit the notion of hierarchical distance in tag clouds. We accentuate relationships between different hierarchy levels by placing hierarchically related tags closely. The original motivation to design TagSpheres was to support humanities scholars in analyzing the clause functions of a search term's co-occurrences (see Sect. 4.1). Aspects of a positive evaluation of the TagSpheres design during the corresponding *eXChange* project are outlined in the previous version of this paper [45]. Further usage scenarios in sports, aviation and natural disaster management outline the inherence of hierarchical textual information in various domains and the usefulness of TagSpheres as they provide an interesting view on this type of data. In addition, we pointed out that the TagSpheres also serves as a novel tree layout algorithm. Although the value of this approach is yet to be evaluated, two use cases in aviation and theology indicate it's potential.

Despite few listed limitations, TagSpheres might be applicable to a multitude of further research questions from other areas. Also imaginable is the combination of TagSpheres and TagPies to support the comparative analysis of different textual summaries with hierarchical information.

Acknowledgements. The authors thank Judith Blumenstein for preparing the digital humanities use cases. This research was funded by the German Federal Ministry of Education and Research.

References

1. Milgram, S., Jodelet, D.: Psychological maps of Paris. In: Environmental Psychology, pp. 104–124 (1976)
2. Deleuze, G., Guattari, F.: Tausend Plateaus. Kapitalismus und Schizophrenie II. Merve Verlag, Berlin (1992)
3. Sinclair, J., Cardew-Hall, M.: The folksonomy tag cloud: when is it useful? J. Inf. Sci. **34**, 15–29 (2008)
4. Viégas, F.B., Wattenberg, M.: TIMELINES: tag clouds and the case for vernacular visualization. Interactions **15**, 49–52 (2008)
5. Hearst, M., Rosner, D.: Tag clouds: data analysis tool or social signaller? In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences, p. 160 (2008)
6. Viégas, F., Wattenberg, M., Feinberg, J.: Participatory visualization with wordle. IEEE Trans. Vis. Comput. Graph. **15**, 1137–1144 (2009)
7. Viégas, F., Wattenberg, M., van Ham, F., Kriss, J., McKeon, M.: ManyEyes: a site for visualization at internet scale. IEEE Trans. Vis. Comput. Graph. **13**, 1121–1128 (2007)
8. Murugesan, S.: Understanding Web 2.0. IT Prof. **9**, 34–41 (2007)
9. Jo, J., Lee, B., Seo, J.: WordlePlus: expanding wordle's use through natural interaction and animation. Comput. Graph. Appl. **35**(6), 20–28 (2015). IEEE
10. 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). doi:10.1007/978-3-642-03655-2_43

11. Schrammel, J., Tscheligi, M.: Patterns in the clouds - the effects of clustered presentation on tag cloud interaction. In: Ebert, A., Veer, G.C., Domik, G., Gershon, N.D., Scheler, I. (eds.) HCIV -2011. LNCS, vol. 8345, pp. 124–132. Springer,
Heidelberg (2014). doi:10.1007/978-3-642-54894-9_9
12. Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M., Qu, H.: Context preserving dynamic word cloud visualization. In: Pacific Visualization Symposium (PacificVis), pp. 121–128. IEEE (2010)
13. Wang, J., Zhao, J., Guo, S., North, C., Ramakrishnan, N.: ReCloud: semantics-based word cloud visualization of user reviews. In: Proceedings of the 2014 Graphics Interface Conference, GI 2014, pp. 151–158. Canadian Information Processing Society (2014)
14. Liu, X., Shen, H.W., Hu, Y.: Supporting multifaceted viewing of word clouds with focus+context display. Inf. Vis. **14**(2), 168–180 (2014)
15. Wu, Y., Provan, T., Wei, F., Liu, S., Ma, K.L.: Semantic-preserving word clouds by seam carving. In: Computer Graphics Forum, vol. 30, pp. 741–750. Wiley Online Library (2011)
16. Barth, L., Kobourov, S.G., Pupyrev, S.: Experimental comparison of semantic word clouds. In: Gudmundsson, J., Katajainen, J. (eds.) SEA 2014. LNCS, vol. 8504, pp. 247–258. Springer, Cham (2014). doi:10.1007/978-3-319-07959-2_21
17. Paulovich, F.V., Toledo, F., Telles, G.P., Minghim, R., Nonato, L.G.: Semantic wordification of document collections. In: Computer Graphics Forum, vol. 31, pp. 1145–1153. Wiley Online Library (2012)
18. Nguyen, D.Q., Tominski, C., Schumann, H., Ta, T.A.: Visualizing tags with spatiotemporal references. In: 2011 15th International Conference on Information Visualisation (IV), pp. 32–39 (2011)
19. Seifert, C., Kienreich, W., Granitzer, M.: Visualizing text classification models with voronoi word clouds. In: Proceedings of the International Conference Information Visualisation (IV), London (2011)
20. Shneiderman, B., Plaisant, C.: Treemaps for Space-Constrained Visualization of Hierarchies (1998)
21. Weskamp, M.: Newsmap (2015). http://newsmap.jp/. Accessed 15 Nov 2015
22. Chi, M., Lin, S., Chen, S., Lin, C., Lee, T.: Morphable word clouds for time-varying text data visualization. IEEE Trans. Vis. Comput. Graph. **21**, 1415–1426 (2015)
23. Lohmann, S., Heimerl, F., Bopp, F., Burch, M., Ertl, T.: ConcentriCloud: word cloud visualization for multiple text documents. In: 19th International Conference on Information Visualisation (2015)
24. Castellà, Q., Sutton, C.: Word storms: multiples of word clouds for visual comparison of documents. In: Proceedings of the 23rd International Conference on World Wide Web, WWW 2014, pp. 665–676. ACM (2014)
25. Collins, C., Viégas, F., 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 (2009)
26. Lee, B., Riche, N., Karlson, A., Carpendale, S.: SparkClouds: visualizing trends in tag clouds. IEEE Trans. Vis. Comput. Graph. **16**, 1182–1189 (2010)
27. Shi, L., Wei, F., Liu, S., Tan, L., Lian, X., Zhou, M.: Understanding text corpora with multiple facets. In: 2010 IEEE Symposium on Visual Analytics Science and Technology (VAST), pp. 99–106 (2010)
28. Burch, M., Lohmann, S., Beck, F., Rodriguez, N., Di Silvestro, L., Weiskopf, D.: RadCloud: visualizing multiple texts with merged word clouds. In: 2014 18th International Conference on Information Visualisation (IV), pp. 108–113 (2014)

29. Diakopoulos, N., Elgesem, D., Salway, A., Zhang, A., Hofland, K.: Compare clouds: visualizing text corpora to compare media frames. In: Proceedings of IUI Workshop on Visual Text Analytics (2015)
30. Jänicke, S., Blumenstein, J., Rücker, M., Zeckzer, D., Scheuermann, G.: Visualizing the results of search queries on ancient text corpora with tag pies. In: Digital Humanities Quarterly (2016)
31. 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 (2008)
32. Waldner, M., Schrammel, J., Klein, M., Kristjánsdóttir, K., Unger, D., Tscheligi, M.: FacetClouds: exploring tag clouds for multi-dimensional data. In: Proceedings of Graphics Interface 2013, GI 2013, pp. 17–24. Canadian Information Processing Society (2013)
33. Harrower, M., Brewer, C.A.: ColorBrewer.org: an online tool for selecting colour schemes for maps. Cartogr. J. **40**, 27–37 (2003)
34. Ware, C.: Information Visualization: Perception for Design. Elsevier (2013)
35. Seifert, C., Kump, B., Kienreich, W., Granitzer, G., Granitzer, M.: On the beauty and usability of tag clouds. In: 12th International Conference on Information Visualisation, IV 2008, pp. 17–25 (2008)
36. Jänicke, S., Franzini, G., Cheema, M.F., Scheuermann, G.: On close and distant reading in digital humanities: a survey and future challenges. In: Borgo, R., Ganovelli, F., Viola, I. (eds.) Eurographics Conference on Visualization (EuroVis) - STARs. The Eurographics Association (2015)
37. Head, K.: Gravity for Beginners. University of British Columbia 2053 (2003)
38. Schulz, H.J.: Treevis.net: a tree visualization reference. IEEE Comput. Graphics Appl. **31**, 11–15 (2011)
39. Nguyen, Q.V., Huang, M.L.: A space-optimized tree visualization. In: IEEE Symposium on Information Visualization, INFOVIS 2002, pp. 85–92 (2002)
40. Holten, D.: Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans. Vis. Comput. Graph. **12**, 741–748 (2006)
41. Zhao, H., Lu, L.: Variational circular treemaps for interactive visualization of hierarchical data. In: 2015 IEEE Pacific Visualization Symposium (PacificVis), pp. 81–85 (2015)
42. Vliegen, R., van Wijk, J.J., van Der Linden, E.J.: Visualizing business data with generalized treemaps. IEEE Trans. Vis. Comput. Graph. **12**, 789–796 (2006)
43. Collins, C., Carpendale, S., Penn, G.: DocuBurst: visualizing document content using language structure. Comput. Graph. Forum **28**, 1039–1046 (2009)
44. Ghoniem, M., Cornil, M., Broeksema, B., Stefas, M., Otjacques, B.: Weighted maps: treemap visualization of geolocated quantitative data. In: IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, p. 93970G (2015)
45. Jänicke, S., Scheuermann, G.: Tagspheres: Visualizing hierarchical relations in tag clouds. In: GRAPP/IVApp (2016)