FacetClouds: Exploring Tag Clouds for Multi-Dimensional Data

Manuela Waldner[1,2], Johann Schrammel[1], Michael Klein[1], Katrín Kristjánsdóttir[1], Dominik Unger[1], Manfred Tscheligi[1,3] *

[1]Cure - Center for Usability Research and Engineering, Vienna [2]Vienna University of Technology [3]ICT&S-Center, University of Salzburg

Abstract

Tag clouds are simple yet very widespread representations of how often certain words appear in a collection. In conventional tag clouds, only a single visual text variable is actively controlled: the tags' font size. Previous work has demonstrated that font size is indeed the most influential visual text variable. However, there are other variables, such as text color, font style and tag orientation, that could be manipulated to encode *additional* data dimensions.

FacetClouds manipulate intrinsic visual text variables to encode multiple data dimensions within a single tag cloud. We conducted a series of experiments to detect the most appropriate visual text variables for encoding nominal and ordinal values in a cloud with tags of varying font size. Results show that color is the most expressive variable for both data types, and that a combination of tag rotation and background color range leads to the best overall performance when showing multiple data dimensions in a single tag cloud.

Index Terms: H.5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces—Screen design (e.g., text, graphics, color)

1 Introduction

Tag clouds are well-known visual representations that encode how often certain words or *tags* appear in a text document, a data collection, or a socially maintained tag collection ("folksonomy"). In a conventional tag cloud, tags are displayed in alphabetical or random order, and the tag's font size encodes the frequency of the tag in the collection. Tag clouds are popular and effective tools for non-specific browsing of folksonomies [21], for grasping an impression of a community's interest [12], and for scanning a summary of descriptive information [14].

In addition to font size and the tags' locations within the cloud, there are additional intrinsic visual text variables that could be actively controlled to encode additional data dimensions (or *facets*). Examples of visual text variables include text color, font decoration, and orientation. Modifications of these visual text variables often serve purely aesthetic purposes [22]. While the expressiveness of visual variables has been thoroughly investigated for basic graphical marks, such as points or areas [2, 18], little is known about the effectiveness of visual *text* variables and the interferences of multiple visual text variables encoding multiple data dimensions.

In this paper, we present *FacetClouds* (Figure 1) – tag clouds that encode multiple ordinal and nominal data dimensions within a single, purely text-based representation. We start by exploring the design space of visual text variables in tag clouds and presenting the *FacetCloud* implementation. A series of three experiments systematically investigates the effectiveness of different visual text variables for visualizing ordinal and nominal data. We contribute the following experimental findings: (a) the effectiveness of **visual text variables encoding nominal or ordinal data** in a tag cloud containing tags of varying font size and (b) the effectiveness of **visual text variable combinations**, visualizing nominal and ordinal data dimensions in a single tag cloud. Our experimental findings can be utilized as guidelines how to effectively visualize more information in a single tag cloud than just the tags' frequencies.

*e-mail: waldner@cg.tuwien.ac.at, {*surname*}@cure.at

Figure 1: A FacetCloud showing 250 Graphics Interface authors (serif, horizontal tags) and sessions (vertical) from 2009 to 2011. Font size encodes the number of occurrences in the dataset. Text color encodes the publication year: the more red, the more recent.

2 Related Work

In many information retrieval tasks, users are confronted with multi-dimensional data. In such a dataset, each item can be assigned multiple, orthogonal categories and attributes, which are often referred to as *facets* [23, 6]. Traditionally, tag clouds have not been used for faceted browsing. Instead, they are usually employed as a loose formation of words (*tags*) of different size in a random or alphabetical sequence.

Kuo *et al.* [14] demonstrated that simple tag clouds slowed users down when looking for specific items in a collection, compared to conventional textual result lists, especially when answering relational questions. According to a study by Sinclair and Cardew-Hall [21], conventional search interfaces were assessed as more useful for accessing specific information. Tag clouds, on the other hand, were perceived as more suitable for browsing and non-specific information discovery [21]. Thus, tag clouds are rarely employed as the sole browsing and searching tool for online collections or database resources. Instead, they are often displayed as a complementary information retrieval interface, e.g. to support browsing of document collections [8].

The most common visual text variable manipulated by tag clouds is the tags' font size. Indeed, several studies have indicated that font size is the most effective visual variable in a tag cloud with respect to recall rate and for signaling tag importance [19, 11, 1]. The tags' spatial location within the cloud also influences their perceived importance. Tags in the upper left quadrant of the screen have been shown to receive more attention [20], have a higher recall rate [19] and can be spotted more quickly when asked to discover a specific tag [11]. Font weight and color are also strong visual features, while other effects are rather weak [1]. We complement these experimental findings by investigating the perceptual effects of visual text variables encoded *in addition* to font size, such as orientation and font color. Our focus lies on the suitability of visual text variables for classifying tags into multiple categories and for encoding additional numeric attributes – both for a single controlled visual variable and for combinations of multiple visual variables.

For more complex visual displays, tag clouds have also been used as augmentations for specific base representations with weighted text information. Tag clouds have been employed in combination with parallel coordinates [4], trend charts [5], as node-link diagrams [13], and on top of maps [9]. Lee *et al.* [16] render simplified line graphs of items' temporal developments into the tag's background. Similarly, Ganesan *et al.* [10] augment tags with emoticons to signal emotions behind community tags. Our work differs as we neither integrate the tag clouds into a specific base representation nor render additional information into the tags themselves. Instead, our goal is to encode additional data dimensions into tag clouds by controlling only the intrinsic visual variables of the text representations themselves without additional visual marks.

For multi-dimensional information, dedicated information visualization techniques and user interfaces have been proposed. In the world-wide web, multi-dimensional metadata is often displayed in multiple lists, summarized by categories (*facets*) [23]. In contrast to tag clouds, the items' frequencies (the number of search results), are indicated as numerals next to each list item.

More complex browsing interfaces employ multiple coordinated views to display different dimensions in their most appropriate visual representation. *PaperLens* [15], *FacetLens* [17] and *GraphTrail* [8] integrate textual lists, temporal charts, and other visualizations in an interface for browsing different facets of large document collections. Others encode faceted information into a single representation. *FacetAtlas* [3] encodes keyword relationships via their relative locations and color-coded connection lines. Similarly, in *PivotPaths* [7], relationships between textual representations of people, resources, and concepts within document databases are visualized by connection lines. *FacetZoom* [6] combines hierarchical faceted browsing with zoomable interfaces. The main goal of our work is to determine whether a single tag cloud is a suitable alternative for supporting the visual presentation of faceted (or multi-dimensional) datasets, by controlling only the intrinsic visual text variables of the tags.

3 Facet Clouds

A dataset can contain multiple, orthogonal data dimensions, which can be either nominal or ordinal. Nominal data can be categorized based on some common characteristics. Consider the example of a document collection: typical categories are "keyword", "author" or "session". The latter two are discriminated by the tag rotation in Figure 1.

Ordinal data represents numeric or ranked attributes, assigned to each item in the dataset. Examples of ordinal values in a document collection are "publication year" and "citation count". In Figure 1, the ordinal value "publication year" is indicated by the tags' text color. The linear color ramp should help users comparing ordinal values between tags.

Nominal	
Text color	any user-defined color
Shape	normal, italic, bold, serif
Background color	any user-defined color
Orientation	0°, 90°, -90°, 180°
Ordinal	
Text color range	`[#999900, #990000]`
Transparency	`[0.2, 1.0]`
Background color range	`[#FFFF99, #FF9999]`
Font size*	`[14pt, 65pt]`

Table 1: Visual text variables of *FacetClouds*. *Font size is restricted to tag frequency in our experiments.

3.1 Visual Text Variables

To visualize multi-dimensional data in tag clouds, each data dimension needs to be mapped to a visual text variable. We rely on an adapted version of Bertin's [2] *retinal properties*, which are used to modify basic graphical units, like points or areas. Retinal properties are visual variables that the human eye is sensitive to, independent of the graphical unit's position [2, 18]: *color*, *shape*, *size*, *saturation*, *texture*, and *orientation*.

Mackinlay [18] classified the expressiveness of retinal techniques for nominal, ordinal, and quantitative data in classical information visualization techniques. According to his classification, size and saturation are perceived as being ordered and should therefore be avoided for encoding nominal measurements. The other retinal techniques (color, shape, texture, and orientation) are suitable for categorization. Appropriate retinal techniques for ordinal data are size, saturation, texture, and parts of the color spectrum.

Mackinlay's classification of retinal techniques only considers their expressiveness for manipulating basic graphical units like simple points, lines, or areas. In contrast, the basic graphical unit in tag clouds is text, which is a much more complex mark than the graphical primitives listed above. We therefore adjusted the list of retinal techniques for text-based graphical units, as summarized in Table 1.

The unmodified base tag in *FacetClouds* is rendered in fully opaque black with *Helvetica* font and variable font size. Helvetica was chosen as most tag clouds in the world-wide web use simple sans-serif fonts. Conventionally, within tag clouds the font size of the text represents the occurrence frequency of the tag in a data collection. We leave this mapping intact and seek to manipulate other visual text variables for encoding additional data dimensions:

The visual variable *color* represents the hue of the text color, both as a user-defined set of distinct colors for categorizing tags and as text color range from yellow to red with a low brightness to encode ordinal information. The text's *shape* is influenced by binary typographic settings, like font style (italic vs. normal), font weight (bold vs. normal), and font family (serif vs. sans-serif). Although these text shape parameters are orthogonal, we grouped them into one visual text variable to support the encoding of more than two tag categories with this variable. For controlling the *saturation* variable, we gradually adjusted the transparency value of the rendered tags on a white background (*i.e.*, tags with small values fade to white). Modification of a text's *texture* is rather uncommon – for instance by hatched fillings of letters to simulate a hand-drawn font. A more obvious approach would be to manipulate the texture of the text's background. To maintain text readability, we instead controlled the text's background color with pastel colors – lighter versions of the same color range as used for text color. Finally, we support four distinct tag rotations to distinguish tag categories based on their *orientation* in 90°steps.

We only consider retinal properties in Table 1 and therefore do not discuss the visual variable *position*. Since tags are usually tightly packed within the cloud representation, the individual tags cannot be precisely positioned. Instead, the position of each tag can only be marginally influenced, for instance by weighting the tags' centrality or their closeness depending on some data attributes.

3.2 FacetCloud Generator

We implemented a *FacetCloud* generator as a Tomcat Java web server with a JSP-front-end. The *FacetCloud* generator supports parsing of XML-databases. The operator specifies XML-tags representing nominal and ordinal fields to be parsed, and additionally defines a visual mapping for each field.

As an example, consider the *FacetCloud* in Figure 1: The XML-dataset originates from the DBLP computer science bibliography[1] (filtered by `venue:graphics_interface`). We reduced the dataset to conference publications in the years 2009 to 2011. As an additional data dimension, we added the `<session>` each publication was presented in. This information was obtained from the ACM Digital Library. The nominal field `<author>` was mapped to serif font, `<session>` to -90°rotation, while the ordinal field publication `<year>` is visualized by a text color range from dark yellow to dark red.

In the resulting tag cloud, each tag is associated with exactly one nominal field and 0 to N ordinal attributes. The ordinal attribute of a tag is defined by the average attribute values of all items associated with the tag. For instance, in the cloud in Figure 1, the text color of the tags is controlled by the average publication year of all papers associated with a certain author or session.

The tags are positioned according to a tag cloud layout algorithm described by Kim *et al.* [13]. Tags are mapped sequentially, starting from the center of the cloud's bounding box. After mapping a tag, the empty space around the tag is re-partitioned into empty space rectangles. This is repeated until all tags are placed into empty space rectangles or discarded, if no sufficiently large rectangle is available. The user defines the ordinal data attributes or a combination of attributes to determine the tags' centrality.

4 EMPIRICAL STUDY

To explore the potentials and limitations of the *FacetCloud* concept, we conducted a series of experiments with the aim of determining how many data dimensions can be successfully encoded into a single tag cloud, and which visual text variables are most suitable for encoding these dimensions. Three experiments were conducted to answer the following research questions:

1. **Tag Categories:** How many tag categories can be effectively distinguished in a tag cloud and which visual variables are best for discriminating the categories?
2. **Ordinal Data:** In addition to font size, which visual variables best aid users in recognizing the most relevant tags, and how do these additional variables affect users' abilities to recognize the largest tags?
3. **Combinations:** Are users able to distinguish visual variables encoding tag categories and ordinal data in a single tag cloud and which visual variable combinations support this discrimination?

For all three experiments, we created tag clouds consisting of 100 tags each, placed within a bounding box of 1000 x 700 pixels. To avoid any cognitive effects, we only showed blind text tags with equal length. For that purpose, we used five-letter words and longer words that were cropped to five letters out of the well-known blind text *lorem ipsum*. We chose to limit the tag length to five letters, in order to keep additional effects caused by word length and number of text pixels – which show a small, but reliable perceptual effect in tag clouds [1] – to a minimum. As we were interested in data dimensions encoded *in addition* to tag frequency, we produced tags with varying font size for all tag clouds. Tags' font sizes were assigned randomly based on a bell curve with a small standard deviation to produce few large (maximum 65pt font size) and a large number of small tags (minimum 14pt). The centrality of tags was redundantly coded with font size in experiment one, and with the combination of the numeric attributes mapped to the font size and other visual variables in experiments two and three.

In sum, 136 tag clouds were pre-computed for the three experiments. Each participant was confronted with all tag clouds, plus five warm-up clouds.

4.1 Participants

We recruited 24 paid participants (age 21-58, average age 39, 13 females) from a test subject database. All participants had normal or corrected vision and were not color blind according to self-report. Every participant was familiar with personal computers, and all except two participants reported daily internet use.

4.2 Procedure and Apparatus

Tag clouds were presented on a 22-inch monitor with 1680 x 1050 pixels. After filling out a demographic questionnaire, each participant started with a warm-up task. Once the participant felt familiar with the setup and the tasks, the first trial was initiated. The trials within each experiment were presented in randomized order to avoid learning effects. Before each trial, a short instruction text together with sample tags was displayed. After reading the instructions, the participant had to click a start button to start the trial timer and to make the tag cloud appear. In each experiment, the task was to find and click on three tags according to some criteria, specific to the individual experiments. Having consistently three targets across the entire study (*i.e.*, 408 clicks in total) represented a good balance between task complexity and effort. Users were instructed to completely solve the tasks as quickly as possible, and then to click the finish button outside the tag cloud. This would stop the timer again and lead to the next trial.

After each experiment, users were presented with paper printouts of tag clouds showing the visual encoding techniques used in the experiment (Figures 2, 4 and 6). Users had to rate the visual appeal and the understandability on a seven-point Likert scale for each printed tag cloud, and could also comment on the different encoding techniques. Overall, participation took 60 to 90 minutes to complete all experiments.

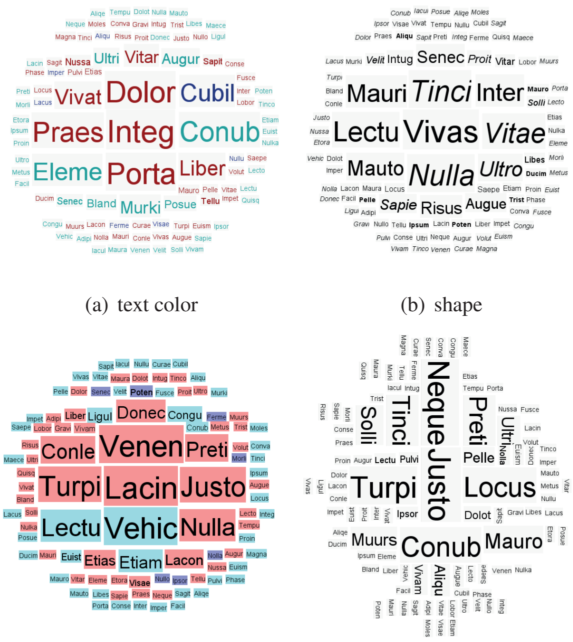
5 EXPERIMENT 1: TAG CATEGORIES

In the first experiment, different visual encoding techniques for categorizing tags were investigated. We controlled the following two independent variables: the *number of tag categories (N)* and the *visual encoding techniques (V)* to distinguish the tag categories. As visual encoding techniques, we used the visual text variables defined in Table 1, upper rows. Note that we did not mix or redundantly encode visual variables to encode different categories, such as the redundant encoding by font and orientation in Figure 1. Instead, we only controlled one visual variable per tag cloud, as shown in Figure 2. As text and background colors, we selected four basic RGB colors with distinct hue (red, cyan, blue, and yellow), which could be easily discriminated in a tag cloud by users of an informal pilot test. Based on the pilot users' feedback, we decreased the brightness for encoding the text color, and increased the lightness of the background color to maintain readability. The four levels of the shape and orientation encodings are listed in Table 1.

We used the following within-subjects factorial design:

	24	participants
x	3	**numbers of categories** N (2, 3, 4)
x	4	**visual encodings** V (Figure 2)
x	4	repetitions
	1152	total trials (48 per participant)

[1] `http://www.dblp.org/search/`



(a) text color (b) shape

(c) bg color (d) orientation

Figure 2: Visual encodings of FacetClouds in experiment one (all with three categories, seven targets and 93 distractors): (a) text color (targets in dark blue), (b) shape (bold targets), (c) background color (blue targets), and (d) text orientation (-90°rotated targets).

5.1 Task

Participants were asked to perform a visual search task with a varying number of targets and distractors, with 100 tags in total. Participants had to click three target tags in the cloud belonging to a certain category. The visual encoding for the category in question was described in the instruction screen before the trial, together with an example tag showing the target encoding. In the four repetitions, we varied the number of tags in the target category (3, 5, 7, and 9) to balance task complexity across the repetitions. The assignment of tags to the respective categories was randomized over the trials. In addition, we balanced the visual encoding of the target categories and the distractors.

The goal of the task was to demonstrate how effectively users could distinguish tags of different categories within a single tag cloud, irrespective of their relevance indicated by the font size.

5.2 Measurements

For each trial, we captured *task completion time*, monitored between the user clicking the start-button to make the tag cloud appear and the user clicking the finish-button after selecting three tags. In addition, we analyzed *false positives* (selected tags belonging to a distractor category) as an error measure. If users could not find three target tags after a subjectively long period, they could skip the trial. In this case, the trial was flagged as *error* and not included in the statistical analysis. Measurements of the four repetitions were accumulated for the analysis for each participant.

5.3 Hypotheses

We defined two hypotheses:

H 1.1: *(a) The greater the number of categories in the tag cloud, the longer it will take to find tags of a certain category, (b) independent of the visual encoding.*

H 1.2: According to a relevance ranking of visual variables for simple graphical marks [18], color is more relevant than shape, which is in turn more relevant than orientation for encoding nominal data. We expect that this ranking will also apply for tag clouds' visual text variables, so that *(a) text color and (b) background color will perform best, and (c) orientation worst*.

5.4 Results

We performed an ANOVA with *visual encoding (V)* and *number of facets (N)* as main factors, participant as random factor, and task completion time as dependent variable (*cf.*, Figure 3). Results show a main effect for V ($F_{3,69} = 129.007, p < .001$) and N ($F_{2,46} = 10.094, p < .001$), as well as an interaction between the two factors ($F_{6,138} = 14.114, p < .001$). Bonferroni-adjusted post-hoc comparisons revealed that *bg color* had a lower task completion time ($\bar{t} = 4.91$s) than all other visual encodings, while *shape* ($\bar{t} = 18.6$s) and *orientation* ($\bar{t} = 15.76$s) had the highest. Four categories ($\bar{t} = 12.83$s) took significantly longer to complete than three ($\bar{t} = 9.83$s) or two ($\bar{t} = 9.99$s). For *text color* and *bg color*, distinguishing two categories took significantly less time than distinguishing three. *Orientation* showed a significant increase in task completion time between three and four categories (from $\bar{t} = 14.29$s to $\bar{t} = 22.71$s). In contrast, *shape* could be solved significantly faster with three categories ($\bar{t} = 15.48$s) than with only two ($\bar{t} = 21.31$s).

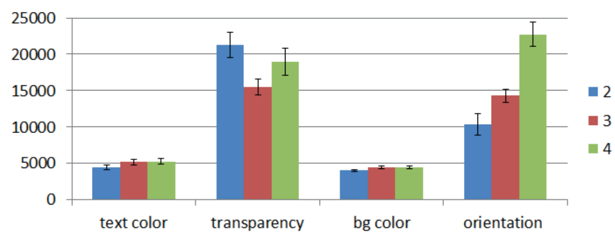


Figure 3: Interaction of visual encodings (horizontal axis) and number of categories (colored bars) for task completion time (in ms).

Overall, 29 trials (approximately 2.5% of all trials) had to be counted as error trials as users were unable to find three targets. 58.6% of the error trials are accounted to *shape*, 31.0% to *orientation*. On average, users tried to find targets for 35 seconds before canceling the task. 14 participants (of the total of 24) encountered at least one task they could not finish. A two-factorial ANOVA on false positives also revealed significant main effects (V: $F_{3,69} = 57,278, p < .001$ N: $F_{2,46} = 18,596, p < .001$) and interactions between the factors V and N ($F_{6,138} = 18,573, p < .001$). Post-hoc comparisons showed that *shape* caused a significantly higher number of false positives ($\overline{N_{fp}} = 0.17$) than any other visual encoding (*orientation*: $\overline{N_{fp}} = 0.01$, *text and bg color*: $\overline{N_{fp}} = 0.0$). For *shape*, the number of false positives increased significantly with each additional number of categories (from $\overline{N_{fp}} = 0.05$ for two categories to $\overline{N_{fp}} = 0.3$ for four).

Subjective ratings on visual appeal and understandability on a seven-point Likert scale for V were evaluated using a one-way ANOVA for each question. We found a significant effect for both questions (visual appeal: $F_{3,69} = 10.026, p < .001$, understandability: $F_{3,69} = 15.847, p < .001$). *Orientation* was rated significantly less visually appealing than *shape* and *text color*, and less understandable than all other visual encoding techniques.

5.5 Summary and Discussion

Our results show that distinguishing tags from four categories indeed takes longer than if fewer categories are present. However,

(a) text color range (b) transparency (c) bg color range

Figure 4: Visual encodings of FacetClouds in experiment two: (a) text color range, (b) transparency, and (c) background color range.

there is an interaction between number of categories and visual encoding. While the increase of task completion time for the color encoding techniques was rather small, categorization of rotated text was heavily affected by larger numbers of categories than two. We only evaluated 90°angles in this experiment. In the future, intermediate angles should be introduced, which will probably be distinguishable more easily. For the shape encoding, three categories were easier to distinguish than two. This difference potentially stems from the fact that we used italic and normal text as encoding when only two categories were present, and introduced bold as a third shape encoding for three categories, serif font as the fourth. Qualitative feedback from the users indicates that italic text was harder to distinguish from normal text than bold tags. During the experiment, we also informally observed an interaction between font size and font weight, causing users to select larger tags that appeared to be bold, when more than two categories were present. This could be a partial explanation for the significant increase in false positives for the shape encoding when bold tags were introduced with three categories. Investigating alternative fonts where shape encodings are visually more effective is an important future activity.

H 1.1: ***(a)*** *Performance indeed decreases with the number of categories, (b) but orientation and shape showed stronger effects than text and background color.*

As expected, the visual encoding is an important factor when it comes to visual discrimination of tag categories. Color encoding was the most effective visual encoding with respect to task completion times, especially when the color was encoded in the background of the tag. Shape and orientation yielded significantly lower performance results and user ratings. Orientation received the lowest subjective ratings, as some users found it "confusing" or "chaotic"[2] – even though only 90°steps were used, leading to a comparably tidy tag cloud appearance.

H 1.2: ***(a, b)*** *Color encoding is the most effective visual text variable for discriminating categories of tags, (c) but shape did not perform better than orientation.*

In summary, color seems to be the first choice when it comes to encoding of (up to four) tag categories. Although having the best performance values, six users described background color as "awkward", "unsettled", or "unattractive" in the interview. Text color – with slightly lower performance than background color – was praised for the choice of colors and its visual appeal.

6 Experiment 2: Ordinal Data

The goal of our second experiment was to investigate the best ways to visually encode ordinal data into a tag cloud. As the first independent variable, we controlled the *visual encoding* (V), based on the visual text variables for ordinal values summarized in Table 1, lower rows. We chose the color and transparency ranges based on subjective feedback of users in an informal pilot study (listed in Table 1). As in the first experiment, the tags' font size was varied according to a bell curve, but was not treated as separate independent variable. We rather see the font size as inherent property of every tag cloud and evaluate the effectiveness of the visual encoding V in combination with font size variations.

As the second independent variable, we asked users to perform two different *tasks* (T): either selecting the three largest tags or the three tags with the maximum values encoded by the visual variable V (*cf.*, Section 6.1).

As with font size, the ordinal values to be mapped to the visual encoding V were randomly assigned according to a bell curve, independently from the font size. The random assignment ensured that targets were evenly distributed across the tags, and with respect to font size. In contrast to the first experiment, the centrality of the tags was coupled to a combination of the value coded in font size and the controlled visual encoding V. As a result of this coupling, large tags were more evenly distributed within the cloud (*cf.*, Figure 4).

As for the previous experiment, we used a within-subjects factorial design:

	24	participants
x	3	**visual encodings** V (Figure 4)
x	2	**tasks** T (select largest, select maximum (V))
x	4	repetitions
	576	total trials (24 per participant)

6.1 Tasks

We asked users to perform two different tasks for each distinct visual encoding. In both tasks, users were asked to click on the three most relevant tags. In the *select largest*-task, relevance was defined by font size, *i.e.*, users had to select the three largest tags. In the *select maximum*-task, the goal was to select the three tags with the maximum ordinal value encoded by the independent variable V, namely: text color range, transparency, or background color range. The instruction text before each trial provided a textual description of the maximum value to look for, together with images of the tags with minimum and maximum visual encoding, respectively.

The aim of the tasks was to determine how quickly and accurately users could find the most relevant tags with respect to some ordinal data attribute. The combination of tasks was chosen to determine the interference of the visual variables V with the tags' font size.

6.2 Measurements

As in experiment one, we recorded *task completion times* for each trial. As *error* measure, we calculated the normalized deviation of the selected tags' assigned ordinal values to the correct maximum values. Since we always used the same bell curve to assign ordinal values to tags, the correct maximum values were constant across all trials, with the maximum tag encoding the value 100, and the two subsequent tags each encoding 97.

6.3 Hypotheses

The hypotheses for this experiment are as follows:

H 2.1: Previous research on tag clouds has shown that font size is the most prominent visual text variable in tag clouds [19, 11, 1]. We therefore hypothesize that *the largest tags will be detected (a) faster and (b) more accurately than maxima of the other visual variables.*.

H 2.2: Bateman *et al.*[1] showed that tag intensity had only a weak effect on perceived tag importance, in contrast to color, which showed a strong effect. However, they only compared two different colors, instead of a color range as in our experiment. Still, we believe that *text and background color will lead to (a) faster and (b) more accurate selection of maxima than transparency. Between text and background color, there will be no difference with respect to (c) task completion time and (d) accuracy.*

[2]All literal citations are translated to English.

6.4 Results

An ANOVA with *visual encoding (V)* and *task (T)* as main factors, participant as random factor, and task completion time as dependent variable showed main effects for both factors (V: $F_{2,46} = 9.302, p < .001$, T: $F_{1,23} = 8.221, p = .009$), as well as an interaction ($F_{2,46} = 4.197, p = .021$), as visualized in Figure 5. Post-hoc comparisons with Bonferroni corrections revealed that *transparency* yielded significantly higher task completion times ($\bar{t} = 7.01$s) than both, *text color* ($\bar{t} = 6.29$s) and *bg color range* ($\bar{t} = 5.97$s). *Bg color range* was significantly faster for solving the *select maximum*-task ($\bar{t} = 5.25$s) than the other two encoding techniques (*text color range*: $\bar{t} = 6.14$s, *transparency*: $\bar{t} = 6.46$s). Overall, the *select largest*-task was performed slower ($\bar{t} = 6.9$s) than the *select maximum*-task ($\bar{t} = 5.95$s).

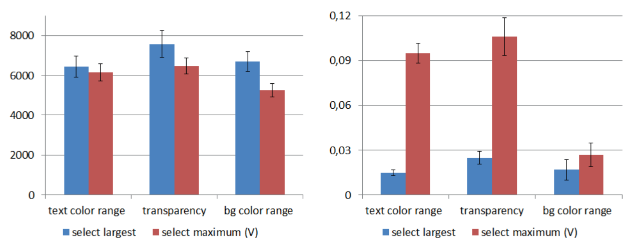


Figure 5: Interaction of visual encoding V (horizontal axis) and task T (colored bars) for task completion time in ms (left) and normalized error rate (right).

As a second dependent variable, we evaluated the error rate. An ANOVA again showed main effects for both factors (V: $F_{2,46} = 28.886, p < .001$, T: $F_{1,23} = 88.012, p < .001$), and an interaction between the two factors ($F_{2,46} = 22.151, p = .021$). The deviation of selected maximum values was significantly higher for *transparency* ($\bar{e} = 6.54\%$) than for the two other visual encodings (*text color range*: $\bar{e} = 5.49\%$, *bg color range*: $\bar{e} = 2.19\%$). For the *select largest* task, the error rate was lower ($\bar{e} = 1.88\%$) than for the *select maximum* task ($\bar{e} = 7.59\%$). For the *select maximum*-task, *bg color range* lead to a lower error rate ($\bar{e} = 2.7\%$) than both, *text color range* ($\bar{e} = 9.5\%$) and *transparency* ($\bar{e} = 10.6\%$). No individual users had a disproportionately big influence on the error averages.

We did not find any differences for the ratings of the subjective visual appeal ($F_{2,46} = 3.157, p = .056$) with respect to V, but we did find a significant difference for understandability ($F_{2,46} = 5.278, p = .009$). *Transparency* was rated as less understandable compared to the two color range encodings.

6.5 Summary and Discussion

Contrary to our expectations, it took participants longer to select the three tags with the largest font size than those with the maximum value with respect to the visual variable V. We interpret this as potential interference of visual variables. However, larger tags were selected with higher accuracy compared to the maximum value of V. These findings are consistent with an exploratory study by Bateman *et al.* [1]: They found that differences in font size could be detected very accurately, but their results also indicate that color and font weight are the more prominent visual text variables.

H 2.1: *Selecting the three largest tags is (a) slower, but* ***(b)*** *more accurate, compared to selection of the three maximum tags with respect to the visual encoding V.*

The least effective ordinal value encoding with respect to task completion time and error rate was transparency. It also received the lowest user ratings for understandability. In the interview, many users complained that faded tags were hard to read – even though reading the text was not required in our task. Background color range was the most effective visual encoding for finding the three tags with the maximum value. Looking at the tag cloud in Figure 4, this finding can be explained as follows: the large, directly adjacent boxes of similar shapes afforded for an effective side-by-side value comparison. Transparency and text color mainly affected the text itself, leading to areas of shapes that were smaller and more distant from each other, making direct comparison between tags more difficult. The study by Bateman *et al.* [1] showed that high intensity tags were likely to be selected as relevant tags. However, qualitative user feedback of our study suggests that subtle differences of transparency values were hard to determine when being asked to do an accurate ranking of tag relevance.

H 2.2: *As expected, transparency lead to the* ***(a)*** *highest task completion times and* ***(b)*** *lowest accuracy. (c, d) Background color range lead to better performance than text color range in the select-maximum-task.*

7 Exploratory Experiment 3: Combinations

In the final experiment, we explored combinations of nominal and ordinal data in one single tag cloud. All tag clouds encoded two categories and one ordinal attribute, in addition to a simulated tag frequency encoded in the font size (as in experiment two).

The first independent variable was the *combination of visual encodings* (*V1-V2*). We used all combinations of visual variables for nominal and ordinal data from Table 1. However, we removed combinations of color encodings for both, tag categories and numeric attributes, as they would lead to unreadable outcomes. The resulting eight combinations are visualized in Figure 6.

The second independent variable was *task* (T): users either had to perform a visual search task as in the first experiment (*cf.*, Section 5.1) or the *select maximum*-task of the second experiment (*cf.*, Section 6.1). In the *select maximum*-task, the three maximum valued tags with respect to a normally distributed ordinal value had to be selected.

The following within-subject design was used in this experiment:

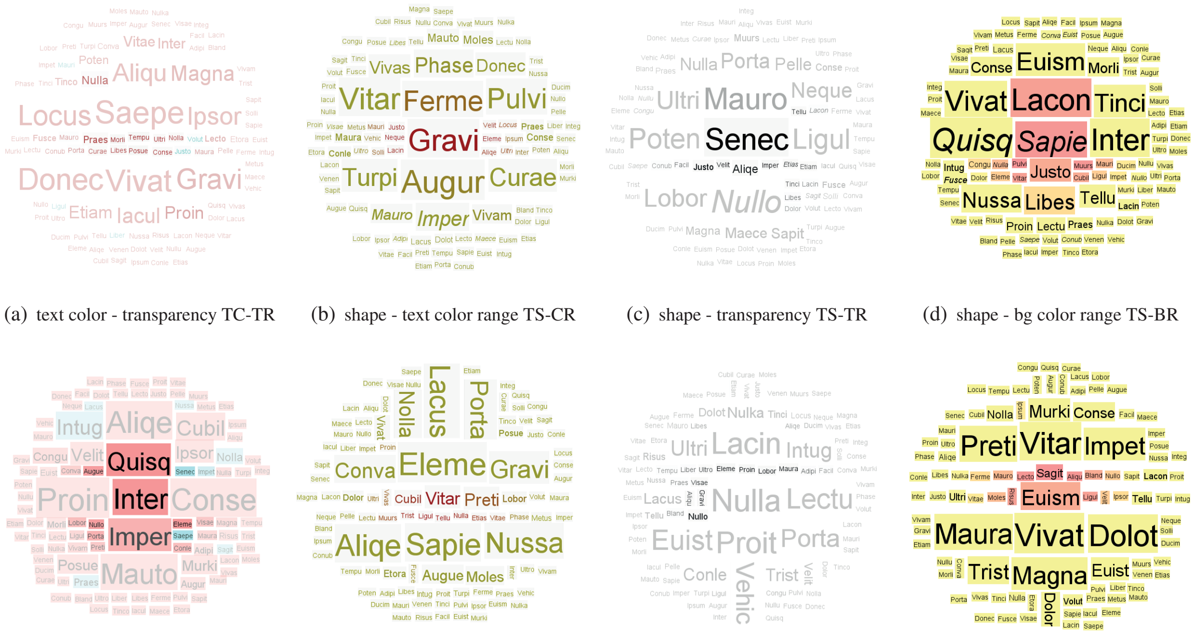
	24	participants
x	8	**encoding combinations** *V1-V2* (Figure 6)
x	2	**tasks** T (select 3 from category, select maximum (V2))
x	4	repetitions
	1536	total trials (64 per participant)

We captured *task completion time* for all trials, *false positives* and complete *error trials* for the *select 3 from category*-task (*cf.*, Section 5.2), and the *error rate* described as deviation from the correct maximum tags for the *select maximum*-task (*cf.*, Section 6.2).

This experiment was designed as exploratory study as we did not develop detailed hypotheses regarding the performance of the different combinations in advance. Instead, the aim of this experiment was to find out whether users could actually distinguish between two encoding techniques, visualizing orthogonal data dimensions. In addition, we wanted to find first evidence about which combinations of visual text variables leads to the best performance overall, when combining the two tasks of the first two experiments.

7.1 Results

Task completion time was analyzed using an ANOVA with *encoding combination (V1-V2)* and *task (T)* as main factors and participant as random factor. The goal of this analysis was to assess the effectiveness of a combination V1-V2 for both task categories, evaluated separately in the previous experiments. We found main effects for both factors (V1-V2: $F_{7,161} = 38.516, p < .001$, T: $F1,23 = 99.568, p < .001$) and an interaction between ($F_{7,161} = 48.973, p < .001$). Post-hoc comparisons revealed that users could solve the task significantly faster with the combination *orientation-bg color range* (OR-BR) ($\bar{t} = 6.03$s) than with all other combinations involving *orientation* (OR-CR ($\bar{t} = 7.79$s), OR-TR ($\bar{t} = 9.17$s)) and *shape*



(a) text color - transparency TC-TR (b) shape - text color range TS-CR (c) shape - transparency TS-TR (d) shape - bg color range TS-BR

(e) bg color - transparency BC-TR (f) orientation - text color range OR-CR (g) orientation - transparency OR-TR (h) orientation - bg color range OR-BR

Figure 6: Combinations of visual encodings in experiment three.

(TS-CR ($\bar{t} = 15.19$s), TS-BR ($\bar{t} = 16.83$s), and TS-TR ($\bar{t} = 18.24$s) to encode tag categories. The three slowest combinations were those using *shape* (TS-CR, TS-BR, and TS-TR), which is significant with respect to all others. Figure 7 illustrates the completion time results.

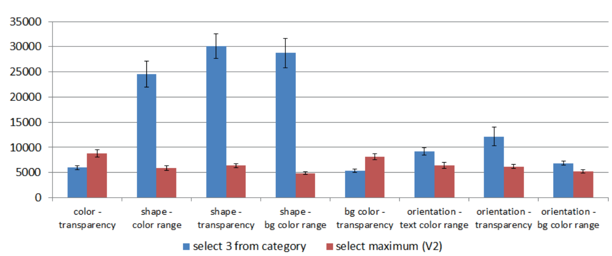


Figure 7: Task completion time (in ms) for each visual encoding combination (V1-V2) (horizontal axis) and task (colored bars).

An ANOVA of the error rates in the *select maximum*-task showed a significant difference of visual encodings ($F_{7,161} = 16,922, p < .001$). The combination *orientation-bg color range* (OR-BR) had a lower error rate ($\bar{e} = 2.29\%$) than all other encoding combinations, except for *shape-bg color range* (TS-BR) ($\bar{e} = 3.00\%$) .

We also found a significant effect of visual encodings on the number of false positives in the *select 3 from category*-task ($F_{7,161} = 7,845, p < .001$). *Shape-transparency* (TS-TR) led to significantly more false positive selections ($\overline{N_{fp}} = .094$) than all combinations not including *shape* (TC-TR, BC-TR, OR-CR, OR-TR, and OR-BR, all $\overline{N_{fp}} = 0$).

There is a significant effect for encoding combination V1-V2 on visual appeal ($F_{7,161} = 4,125, p < .001$) and on understandability ($F_{7,161} = 8,323, p < .001$). Post-hoc comparisons revealed that the visual appeal of *shape-text color range* (TS-CR) was rated higher than for *text color-transparency* (TC-TR) and *orientation-transparency* (OR-TR). Concerning understandability, *orientation-transparency* (OR-TR) was rated significantly lower than *orientation-bg color range* (OR-BR), *shape-bg color range* (TS-BR), and *shape-text color range* (TS-CR).

7.2 Summary and Discussion

The results showed that combinations including shape encoding for tag categories (TS-CR, TS-TR, TS-BR) lead to overall low performance. The combination of shape and transparency encoding (TS-TR) lead to the highest number of false positive selections for the *select 3 from category*-task. This indicates that the added transparency could have decreased the recognizability of italic versus non-italic tags observed in experiment one even more.

The encoding leading to the fastest overall completion times was the combination of orientation and background color range (OR-BR). The two combinations using background color range as ordinal data encoding (TS-BR, OR-BR) also had the lowest error rates for the *select maximum (V2)*-task, which is consistent with results from experiment two.

In contrast to the performance measures, the highest subjective score for visual appeal and understandability was achieved by the combination of shape and text color range TS-CR. Users described this combination as “balanced”, with “good readability”. It is important to note that readability was not captured by the performance measures in our experiments. However, it is, of course, a very important aspect for text-based visualizations. Assessing not only perceptual effects, but also text cognition for *FacetClouds* is the logical next step in the future.

Both our previous experiments have shown that all tasks could be solved faster and with higher accuracy if the data dimension in question was encoded by color. In this experiment, color could either be used to encode nominal or ordinal data. Results indicate that combinations that include background color ranges to encode ordinal values lead to higher overall performance, when both nominal and ordinal data dimensions need to be encoded. Background color range is the first choice if the ordinal facet to be encoded needs to be accurately detected. If the focus of the *FacetCloud* lies more on the visual aesthetic, text color range encoding for ordinal values is probably the better alternative. Thus, for encoding tag categories, other visual text variables than color have to be used. As neither shape nor orientation yielded convincing results for finding tags of a certain category, alternative color-independent encoding techniques for tag categorizations (such as tag bounding box border styles) or redundant encoding of shapes and orientation (such as those used in Figure 1) should be explored.

8 Conclusion and Future Work

FacetClouds encode multiple orthogonal data dimensions in a single tag cloud by varying different visual text variables. We explored the effectiveness of retinal properties on textual marks arranged in a tag cloud representation, as well as the combination of multiple visual text variables for encoding multiple data dimensions. Below, we summarize the results of our three experiments:

Apart from font size, **color** (both as text color or as the tag's background color) is the most effective visual text variable for encoding nominal as well as ordinal data.

Background color range supports more accurate estimation of the most relevant tags than text color and is therefore recommended for tasks that require fairly precise evaluation of ordinal tag values.

Tag **transparency** is disliked by the users and leads to inaccurate results when determining the most relevant tags.

Font size can be more accurately compared between tags than the other ordinal visual text variables in our experiments, but the comparison of tags is slower.

Text **orientation** in 90°steps and **shape** modifications performed significantly worse than color encoding for distinguishing tag categories. Many users perceived rotated tags as unstructured, unattractive, and hardly readable. Shape differences caused by serifs or font styles were hard to detect with the chosen Helvetica font.

When combining nominal and ordinal dimensions into a single tag cloud, **color** should be used to encode ordinal data. The most effective combination for our tasks was **orientation with background color range** (Figure 6h), while the combination of **shape and text color range** (Figure 6b) received highest subjective user ratings. As neither shape nor orientation encoding seem to be adequate alternatives to color for distinguishing tag categories, new nominal color-independent visual encoding techniques for tags or combinations of techniques should be explored in the future.

In our experiment, we used blind text with equal text length to concentrate on purely perceptual effects. Using real text and real data may lead to new, yet unexplored challenges. For instance, real text distorts the visual text variables, as long words make tags bigger and vice versa, which has been shown to have an effect on perceived tag importance [1]. Also, the users' concerns about reduced readability, especially for rotated and faded tags, could not be quantified by the performance measures in our experiments. It is therefore subject to future research to assess the perceived accuracy of *FacetClouds* with real data.

Our experimental tasks were chosen to represent typical low-level activities in information retrieval tasks with tag clouds, such as finding the most relevant tags according to some ordinal value or discriminating categories of tags. However, when working with real data, more complex compound tasks may have to be solved. Static *FacetClouds* do not support such compound activities well. We therefore plan an extension allowing dynamic filtering of the cloud, and better direct comparison between pairs of tags.

Acknowledgements

This work was supported by FFG Austria (projects 825345 'AIR' and 825352 'Softnet II') and the Vienna Science and Technology Fund (WWTF, project VRG11-010).

References

[1] S. Bateman, C. Gutwin, and M. Nacenta. Seeing things in the clouds: the effect of visual features on tag cloud selections. In *Proc. HT 2008*, pages 193–202. ACM Press, June 2008.

[2] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. 1984.

[3] N. Cao, J. Sun, Y.-R. Lin, D. Gotz, S. Liu, and H. Qu. FacetAtlas: multifaceted visualization for rich text corpora. *TVCG 2010*, 16(6):1172–1181, Nov. 2010.

[4] C. Collins, F. B. Viegas, and M. Wattenberg. Parallel Tag Clouds to explore and analyze faceted text corpora. In *Proc. VAST 2009*, pages 91–98. IEEE, 2009.

[5] W. Cui, Y. Wu, S. Liu, F. Wei, M. Zhou, and Q. Huamin. Context preserving dynamic word cloud visualization. *Computer Graphics and Applications*, 30(6):42 – 53, 2010.

[6] R. Dachselt, M. Frisch, and M. Weiland. FacetZoom: A Continuous Multi-Scale Widget for Navigating Hierarchical Metadata. In *Proc. CHI 2008*, pages 1353–1356. ACM, 2008.

[7] M. Dörk, N. H. Riche, G. Ramos, and S. Dumais. PivotPaths: Strolling through Faceted Information Spaces. *TVCG 2012*, 18(12):2709 – 2718, 2012.

[8] C. Dunne, N. Henry Riche, B. Lee, R. Metoyer, and G. Robertson. GraphTrail. In *Proc. CHI 2012*, pages 1663–1672. ACM Press, May 2012.

[9] K. Fujimura, S. Fujimura, T. Matsubayashi, T. Yamada, and H. Okuda. Topigraphy: Visualization for Large-scale Tag Clouds. In *Proc. WWW 2008*, pages 1087–1088. ACM Press, Apr. 2008.

[10] K. a. Ganesan, N. Sundaresan, and H. Deo. Mining tag clouds and emoticons behind community feedback. In *Proc. WWW 2008*, pages 1181–1182, New York, New York, USA, 2008. ACM Press.

[11] M. J. Halvey and M. T. Keane. An assessment of tag presentation techniques. In *Proc. WWW 2007*, pages 1313–1314. ACM Press, 2007.

[12] M. a. Hearst and D. Rosner. Tag Clouds: Data Analysis Tool or Social Signaller? In *Proc. HICSS 2008*. IEEE, Jan. 2008.

[13] K. Kim, S. Ko, N. Elmqvist, and D. S. Ebert. WordBridge: Using Composite Tag Clouds in Node-Link Diagrams for Visualizing Content and Relations in Text Corpora. In *Proc. HICSS 2011*, pages 1–8. IEEE, Jan. 2011.

[14] B. Y.-L. Kuo, T. Hentrich, B. M. . Good, and M. D. Wilkinson. Tag clouds for summarizing web search results. In *Proc. WWW 2007*, pages 1203–1204. ACM Press, May 2007.

[15] B. Lee, M. Czerwinski, G. Robertson, and B. B. Bederson. Understanding research trends in conferences using paperLens. In *Ext. Abstracts CHI 2005*, pages 1969–1972, New York, New York, USA, Apr. 2005. ACM Press.

[16] B. Lee, N. H. Riche, A. K. Karlson, and S. Carpendale. SparkClouds: Visualizing Trends in Tag Clouds. *TVCG 2010*, 16(6):1182–1189, Nov. 2010.

[17] B. Lee, G. Smith, G. G. Robertson, M. Czerwinski, and D. S. Tan. FacetLens. In *Proc. CHI 2009*, pages 1293–1302. ACM Press, Apr. 2009.

[18] J. Mackinlay. Automating the Design of Graphical Presentations of Relational Information. *ACM Transactions on Graphics*, 5(April 1986):110–141, 1987.

[19] A. W. Rivadeneira, D. M. Gruen, M. J. Muller, and D. R. Millen. Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proc. CHI 2008*, pages 995–998. ACM Press, Apr. 2007.

[20] J. Schrammel, S. Deutsch, and M. Tscheligi. Visual search strategies of tag clouds - results from an eyetracking study. In *Proc. INTERACT 2009*, pages 819–831, Berlin, Heidelberg, 2009. Springer-Verlag.

[21] J. Sinclair and M. Cardew-Hall. The folksonomy tag cloud: when is it useful? *Journal of Information Science*, 34(1):15–29, 2007.

[22] F. B. Viégas, M. Wattenberg, and J. Feinberg. Participatory visualization with Wordle. *TVCG 2009*, 15(6):1137–1144, 2009.

[23] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proc. CHI 2003*, pages 401–408. ACM Press, 2003.