FacetScape: A Visualization for Exploring the Search Space

Christin Seifert, Johannes Jurgovsky and Michael Granitzer
Media Computer Science, University of Passau
Passau, Germany
firstname.lastname@uni-passau.de

***Abstract*—Despite advancing search technologies, information overload has not yet been solved. Getting an overview of information or explorative access to information becomes increasingly difficult with the exponentially increasing amount of information. Search result visualizations, especially for faceted browsing, aim at supporting users to find their way through large document collections. We propose FacetScape, a novel visualization for navigation and refinement of search results allowing users to visually construct complex boolean search queries for narrowing down the search space. This visualization combines Voronoi subdivision and a tag cloud representation of the search facets. Further it includes a preview of action (query preview) and interactions to allow users to focus on important aspects of the data for the task at hand. In a comparative user study with 15 users we compared the visualization to a standard faceted browsing interface for different types of search tasks. The study revealed that participants used the unfamiliar interface as efficiently and effectively as the familiar tree-like display. Results indicate that the FacetScape is a promising way of supporting users in exploring the faceted search space.**

***Keywords*-Faceted navigation, search user interfaces, user evaluation**

I. Introduction

In the recent decade the amount of digital data has increased rapidly. In their latest study Bounie and Gille estimated a growth of 233% of information available on the Internet (from 5 million TB in 2003 to 1,8 billion TB in 2008) [1]. To make this information accessible, both specialized and multi-purpose search systems have emerged, Google Search being one of the most prominent. Due to the large amount of indexed data, those search systems almost always return a long list of search results, a fact that is even increased as users tend to formulate short search queries [2]. Approaches to structuring the search result space include hierarchical clustering of the results and – if the metadata is available in the search results – faceted browsing. Faceted browsing interfaces allow users to construct boolean search queries by selecting or deselecting values in facets and thus focus on specific search results (selecting a value) or remove the focus (deselecting a value). Selections in multiple facets then represent a complex boolean search query. A faceted browsing interface must satisfy the following requirements: First, it should fit into limited space on the screen and leave space for the search result list. Second, it should support intuitive query construction and visual feedback of the constructed query. Third, it should give a preview of action, i.e., allow previewing queries [3]. Fourth, it should allow users to focus on important aspects of the data (selecting/removing facets, and zooming into facets) and encourage explorative search.

In this paper we propose a novel visualization for faceted browsing, the FacetScape, that satisfies the above requirements[1]. The FacetScape combines Voronoi subdivision and tag clouds. Each Voronoi cell corresponds to a facet of the search results, and the values of the facets are placed inside the respective cell using a tag layout algorithm. By using an additively weighted power Voronoi diagram with appropriate interactions we allow users to focus on facets or let them remove facets from the view. The contribution of this work is the following:

1) We propose a novel visualization for overview and navigation of faceted search, allowing users to interactively focus on facets or remove facets from the visualization.
2) In a formative user study we compare this new visualization to a standard faceted browsing interface.
3) We propose an extended query preview, showing users the direct influence of the selection/deselection of a value on the result set.
4) We propose an interactive solution for the problem of degenerating, additively weighted power Voronoi diagrams depending on the weights.
5) We include a graphical method to generate complex boolean search queries, including fast (de-)selection of all values in one facet.

In the next section we review related work. In Section III we provide details of the proposed visualization and interactions. Section IV describes the comparative user study and discusses the results. Finally, Section V provides a summary and an outlook to future work.

II. Background and Related Work

In faceted browsing, a facet represents a feature dimension (e.g., author, date), and each document has a value for each facet (although there might be missing values). Thus the grouping of the search results is multi-dimensional, and

[1] Available at http://purl.org/eexcess/components/vis/facetscape

documents belong to multiple categories. Hearst et al. [3] coined the term "orthogonal sets" of categories for the dimensions/facets. Faceted browsing has been shown to be superior to navigating along a single dimension [4]. Jody Fagan advocated these findings in a more recent review on usability studies for faceted browsing and detailed more benefits, e.g. a higher participants' satisfaction with the user interface [5]. Examples of concrete faceted search user interfaces include the Flamenco system [3], the mSpace system [6] and TRIST [7], and commercial websites, such as booking.com and amazon.com. As the examples above, we aim to create a faceted browsing interface that supports an information seeking model similar to that of Marchionini and White [8], assuming a static information need.

Query previews have been identified as a success factor for faceted browsing [3] by enabling efficient browsing of large information systems. With query previews, browsing and querying can be combined by providing summary data to guide users in narrowing down the search space. The summary data varies with both the underlying data and the application and provides an overview of the data from several perspectives but is orders of magnitude smaller than the data itself [9]. Tanin et. al [10] propose a user interface architecture using generalized query previews incorporating data distribution information in the preview. Qvarfordt et al. [11] apply query previews in this scenario by introducing stacked bar charts visualizing the changes of documents in the result list over time. In our proposed visualization we focused on the compact summary aspect of query previews by introducing small visual marks that indicate changes in the list of retrieved results.

Voronoi tessellation divides the space into polygonal areas [12]. The InfoSky visualization [13] uses a modified additively weighted (AW) power Voronoi diagram. AW Voronoi diagrams may degenerate, i.e., the generator points may lie outside their polygonal areas, without constraining the weights. In InfoSky the weights are globally constrained to avoid these cases. This constraint leads to cases where the area of a polygon is not related to its weight anymore, an issue that is solved heuristically by a modified initial force-directed placement algorithm in InfoSky. The Voronoi Treemaps introduced by Balzer and Deussen [14] applied centroidal Voronoi tessellations by iteratively (i) determining the Voronoi tessellation, (ii) moving each point into the center of mass of its polygon, until distance between point position and center of mass is below a given threshold. The first step also uses AW Voronoi subdivision with constraints imposed on the weights to avoid degeneration of the Voronoi diagram. We want to allow users to weight the size of different regions, thus we aim for an AW Voronoi diagram. Having users interacting with the visualization we can not impose static constraints on the weights for an optimal layout. We implement an AW Voronoi diagram with independent weights each of which controlling the radius of an imaginary circle within its Voronoi region. In order to prevent several regions from overlapping each other we dynamically constrain the user interactions.

Tag layout in rectangular boxes has been proposed in [15] and extended to polygonal boundary regions using simple heuristics [16]. Usability studies have shown that words in the centre of the tag cloud receive more attention and that font size is a strong visual clue for finding relevant tags [17], [18]. Our requirements for the tag layout were i) support of arbitrary shapes and ii) visual encoding of tag frequency. We chose a similar, but more computationally efficient layout technique as in [16], and chose to visually encode tag frequency with font size applying the findings from usability studies.

III. FacetScape Visualization

In this section we describe the construction of the FacetScape. First, the Voronoi generator points are positioned inside the available space in a uniform manner. Every generator point uniquely maps on a single facet. Then, the AW power Voronoi is constructed with all weights set to 1.0 (see section III-A). After that, the facet name and its values are laid out in the respective Voronoi area using a tag layout algorithm. Appropriate interactions enable fast selection and deselection while query previews allow users to judge the possible consequences of an interaction.

A. Voronoi Subdivision

A Voronoi diagram is a partition of the Euclidean space E^d into convex *Voronoi cells* $V(p)$ defined from a finite set of generating sites $p \in M \subseteq E^d$ with all points of $V(p)$ being closer to p than to any other site in M. Adjacent cells produce bisecting *Voronoi edges* defined by points with equal distance with respect to a certain function [12]. Assigning weights $w(p)$ to the sites $p \in M$ yields a more general version of the Voronoi diagram. The distance from p to a point $x \in E^d$ is given by a combination of the distance $d(x,p)$ and the weight function $w(p)$. For an additively weighted Power Voronoi diagram this function is defined as $pow(x,p) = (x-p)^T(x-p) - w(p)$ In E^2, the points of equal power with respect to two weighted sites form the *power line* of the corresponding sites. In the case of two sites having equal weights, the Voronoi edge is the bisector, whereas for sites with different weights it is a line parallel to the bisector moved by an offset $\Delta = \frac{w(p)-w(q)}{2d(p,q)}$ towards the site with lower weight [19]. In our implementation we consider the weight $w(p)$ of a site p as the radius of an imaginary circle defining the minimum size of a Voronoi cell. To prevent cells from overlapping the generator points of other cells we introduce a global constraint $w(p)+w(q) < d(p,q)$ for all pairs $(p,q) \in M^2$ which guarantees that arbitrary changes to the individual weights through user interactions will not lead to a degenerated Voronoi diagram. By using the power function the edges can be represented as straight lines instead of hyperbolic curves. Additionally, the non-uniformly weighting scheme allows us to provide the user a secondary parameter to control the Voronoi layout.

In our work, the generator points of the Voronoi regions are initially uniformly distributed across the available space. The number of regions that should be placed next to each other horizontally is fixed to a value C_h. The positions of the generating points are then calculated according to the width and height of the visualization area and the maximum horizontal capacity C_h. The default weight of each region is set to $\frac{1}{|M|} * a$ with a being a constant depending on the size of the visualization area. The weight of a region may change arbitrarily through user interaction by adding or subtracting a constant weight w_0.

B. Tag Layout

Each of the Voronoi regions corresponds to a facet, i.e., a nominal feature dimension in the result set. In this section we describe how we place the values for each facet and the facet name inside its Voronoi region.

The centre of a tag cloud is given by the centroid of its Voronoi region. For a layout without overlapping tags, the tags' bounding boxes are considered as the basic graphical entities that need to be positioned. As basis for the layout we use a Fermat's spiral sampled with a constant stepping. Starting at the centroid position, each sample position is tested if it lies inside the polygon and if the next relevant, but not yet positioned tag, would fit at that location. The decision if a tag fits to a particular location is constrained by the requirements that tags may not overlap other tags and that the bounding box of a tag has to fit inside the polygon's boundaries entirely. This process is repeated until all sample points of a full $360°$ degree rotation are outside the polygon or all tags have been positioned.

The Fermat's spiral $r^2 = a^2\phi$ provides multiple features that strongly support the peculiarities in the perception of tag clouds. An increase of the rotation angle ϕ for small values of ϕ causes high increases with respect to the radius, whereas for large ϕ the increasing of the radius is only marginal. Such a relation is well suited for tag clouds as important tags with large font size tend to consume big areas near the center but less important tags with small font size consume small areas at the peripheral region. Additionally, the implicit symmetry of a Fermat's spiral supports the idea of two tags with similar degree of importance being positioned at approximately the same distance to the center.

C. Interactions

Our FacetScape implementation offers multiple ways of interaction with both, the facet itself and its values. As tags may be of arbitrary length, the tag text is trimmed to a maximum number of characters in order to be presentable within a Voronoi region. This factor adds to the rather usual interaction techniques regarding tag selection:

- *Hovering over a Tag:* Moving the mouse pointer over a tag causes it to expand to its full length and thereby overlap elements that are beneath it.
- *Selection of a Tag:* Clicking on a tag adds a new constraint to the user's current selection. After each click the set of constraints is re-evaluated to reconcile the user's selection with the list of results (query construction).
- *Selection of a Facet:* Similar to selecting a single tag, one can select the whole facet by clicking on its name. The list of results is then restricted to those results that are annotated with any tag belonging to the facet.

Besides the above mentioned selection mechanisms which involve changes in the search result list, the user may want to rearrange the positions and to adjust the sizes of the Voronoi regions. These mechanics are derived from the properties of the Power Voronoi and are designed to support the user in getting a better overview of the available facets and their provided values.

- *Moving a Facet:* By holding down the left mouse button on a facet's name, the facet can be dragged across the visualization area or temporarily dropped on a spare area to give more space to other facets.
- *Resizing a Facet:* Turning the mouse wheel will change the size of the facet where the mouse pointer currently resides in by adapting the weight of the AW power Voronoi (see also Section III-A).

Both interaction mechanisms could possibly lead to a degeneration of the AW power Voronoi if they were not restricted by the global constraint introduced in Section III-A. A Voronoi region refuses to move further towards an adjacent cell or to increase its size if this interaction would violate the global constraint. The edges of those regions that would conflict with the global constraint are then highlighted in a noticeable different colour for a short amount of time to give visual feedback that the desired interaction is currently not possible. Figures 1 shows examples of the visualization and changes imposed by different types of interaction.

D. Dynamic Query Preview

A small number next to a value of a facet encodes the amount of documents in the result set that is annotated with this value. When hovering the mouse over a value, two visuals appear at the border of the value. The first visual inherits the amount of corresponding documents from the unhovered value and relates it to the total number of documents in the result set. The format is "x/N" and reflects the ratio of documents having the metadata value (x) and all documents N in the result set. Next to this visual we show a second visual to integrate the concept of query preview in terms of the quantity to expect when the user selects the tag. A positive number, e.g. "+17", indicates an increase of the result set by 17 documents and

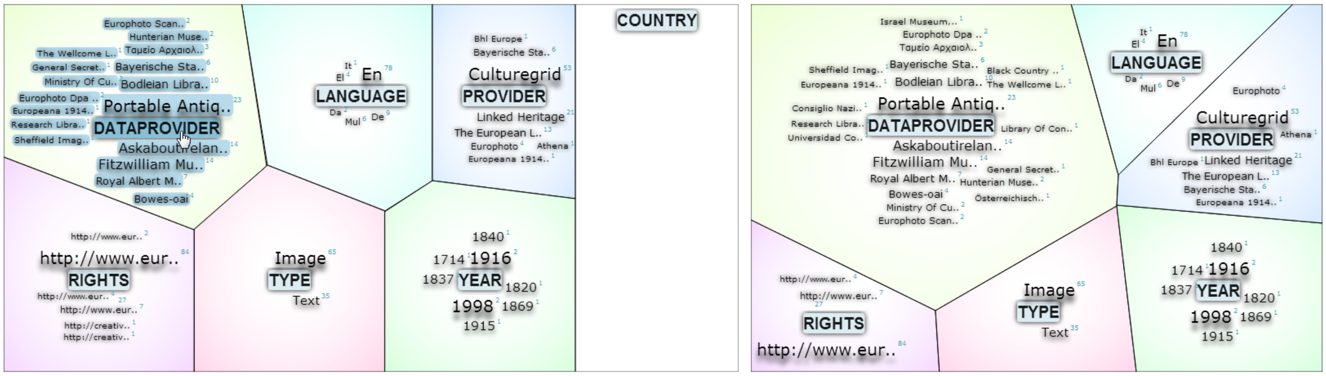

Figure 1. Layout after removing facet “country” and selecting facet “dataprovider”(left), and after enlarging facet “dataprovider” (right).

a negative number a decrease, respectively. The background of the visual is colour-coded according to the algebraic sign in order to support a quick estimate of the impact of the selection (orange: decreasing number of results, blue: increasing number of results). Figure 2 shows an example; a preview on what would happen if the value “Ireland” got selected (80 results would be removed) and deselected (80 results would be added).

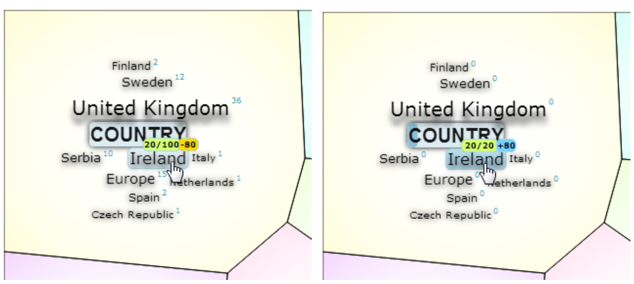

Figure 2. Query Preview for selecting (left) and deselecting (right) a value in a facet.

E. Interactive Query Construction

With the selection of tags in the cloud layout the user implicitly formulates a boolean query. Multiple selections can be performed both inside the same facet as well as across several different facets. However, the interpretation of inter-facet selections differs from those of intra-facet selections. For intra-facet selections, a document is part of the result set if it features either of the selected tags of the same facet (OR). And vice versa for inter-facet selections, only those documents are members of the result set which comply with all constraints given by multiple intra-facet selections (AND). Additionally, the user may define textual constraints on its own by specifying keywords that should be present in a document’s metadata. When the user enters a list of keywords separated by whitespaces, the list is considered as an OR-constraint of the individual keywords. If several lists of keywords are provided by the user, they are combined with a Boolean AND operation. The value that a tag represents is one particular stamping of a facet and in the following denoted as *word* w. For C facets, each facet $c \in \{1, ..., C\}$ provides N words denoted as $w_c^{(i)}$ with $i \in S_c \subseteq \{1, ..., N\}$ and S_c being the set of selected tags in facet c. A function $f(r, w_c^{(i)})$ validates the occurrence of the word $w_c^{(i)}$ in the document r and returns a boolean value. The overall expression that has to be evaluated for each document is $\delta_1(r) = \bigwedge_c \left(\bigvee_i f(r, w_c^{(i)}) \right)$. In case the user specified V additional constraints manually through keywords, we derive a similar expression from the expression above incorporating some minor changes. A constraint $v \in \{1, ..., V\}$ consists of a list of M keywords $w_v^{(j)}$ with $j \in \{1, ..., M\}$. Similarly, a function $k(r, w_v^{(j)})$ validates the occurrence of keyword $w_v^{(j)}$ against the whole metadata attached to document r. $\delta_2(r) = \bigwedge_v \left(\bigvee_j k(r, w_v^{(j)}) \right)$. A document will be a member of the result set if it complies both with the selections in the tag cloud and the user-created keyword constraints, i.e., $\delta(r) = \delta_1(r) \wedge \delta_2(r)$.

IV. User Study

In a user study we wanted to find out for which search tasks the FacetScape can efficiently and effectively be used. On this account, we compared the FacetScape to a standard faceted search user interface.

A. Design

For the comparative evaluation we used a within-subjects design. The independent variable, the type of user interface, has two values: i) standard faceted search interface (called “TreeView”) and ii) FacetScape visualization. Because both interfaces only show the metadata (facets) of the search results, we additionally display a standard search result list showing document surrogates, to both visualizations. In both conditions the visualizations occupy roughly the same screen space (920x920 pixels). We account for learning and fatigue effects by changing the sequence of the interfaces and provided different tasks with the same level of difficulty in both interfaces. We disabled the functionality of searching the result set to get a clearer picture of the navigability of the visualizations.

Dependent variables were task success, task completion time [ms] and the number of clicks. Task success was measured with the help of a ground truth, i.e. by comparing the participant's decision with the correct result. Task completion time was measured as the time difference of the participants clicking the start and the stop button. We counted the number of clicks while differentiating between clicks in the result list and clicks in either the TreeView or the FacetScape to be able to distinguish navigational clicks and clicks to assess resource details. In a questionnaire, we assessed subjective variables, e.g., perceived helpfulness, using a five-point Likert-scale (1 being "very positive" and 5 being "very negative").

B. Procedure

An overview of the evaluation procedure is depicted in Figure 3. The evaluation started with a general introduction to faceted search result navigation (I). Then the trials started, in either the sequence FacetScape - TreeView or TreeView - FacetScape. Before each trial, the participants got an introduction to the respective visualization (IFS or ITV), in which all interaction possibilities were explained. They could test each visualization on a test data set and could ask questions. After that introduction the participants were given six tasks in each visualization. Note that they received a different set of tasks for each visualization, and the task sets and the visualizations where given in different sequence for the participants. After participants had read the task instruction for one task, they clicked on the "start" button, the visualization was shown and the measurements started. To avoid measuring typing speed and correction of misspellings, the participants selected the respective search queries from a drop down list. When participants had finished the task, they clicked the "finish" button, the timing stopped and the experimenter noted the result. At the end, participants were asked to fill out the questionnaire.

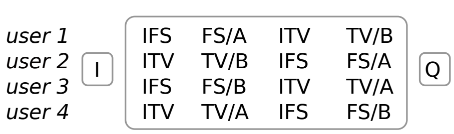

Figure 3. Overview of the evaluation procedure. I - introduction; IFS, ITV - introduction FacetScape or TreeView; FS, TV - trials with FacetScape or trials with TreeView; A, B - task set A or task set B; Q - questions.

C. Participants and Environment

Both visualizations were evaluated by 15 (10 male, 5 female) German speaking participants. The participants were on average 31 years ranging from 20 to 59 years. 11 participants described themselves as computer experts and 4 as normal computer users. We conducted the evaluation in a calm environment. The technical setup included a 13.3" MacBook Air with a 1,3 GHz Dual-Core Intel Core i5 processor, 4 GB of RAM, and a USB mouse. The display resolution was 1440 x 900 pixels.

D. Test Material

To test the visualization we used the Europeana collection of cultural heritage objects provided by the Europeana Foundation. We access the collection using the Europeana REST API[2]. The search query for each task was defined in the task description and provided via a drop-down menu to the participants. We showed the first 100 (most relevant) search results returned by Europeana sorted descendingly by the provided relevance score. As of the time of the evaluation (late 2013) Europeana comprised of 30 million digitized objects (images, videos, texts) and provided six facets (e.g., year, description language, provider) by their API. Due to Europeana data integration issues not all documents had a value for each facet assigned. Figure 4 shows the results for search query "loom" in the FacetScape and the TreeView.

E. Tasks

We created two different task sets, A and B, each comprised 6 tasks. The tasks in each set were of equal difficulty, e.g. Task 1 in set B was equally difficult as Task 1 in set A. Two tasks are considered to be equally difficult if they required the same minimal number of clicks to be solved. The minimal number of clicks was defined as the minimal number of interactions with the facet values until at most five results remained in the result set (following the findings that most users do not much consider search results that are ranked lower [20]). For each task, the initial search query was given in the task description to ensure the same initial set of results across different users.

With the first three tasks in each set we tested explicit navigation, i. e., the required facets and values were explicitly given in the task description. The correct answer set consisted of a particular subset of the initial set of results. The size of the correct answer sets ranged from 3 to 77 results.

Task 1: Selection of all but one value in one facet is required to find the result (complex OR query).

Task 2: Selection of two high-frequent values in two facets is required to find the result (AND/OR query).

Task 3: Selection of two low-frequent values in two facets is required to find the result (AND/OR query).

An example task description of Task 1 is: "You are interested in machines for paper production. You are not interested in the machines from the year 2005. How many relevant results do you find?".

With the fourth to sixth task implicit navigation within the visualizations is tested. This means, the facets for navigation

[2]http://europeana.eu/portal/api-introduction.html

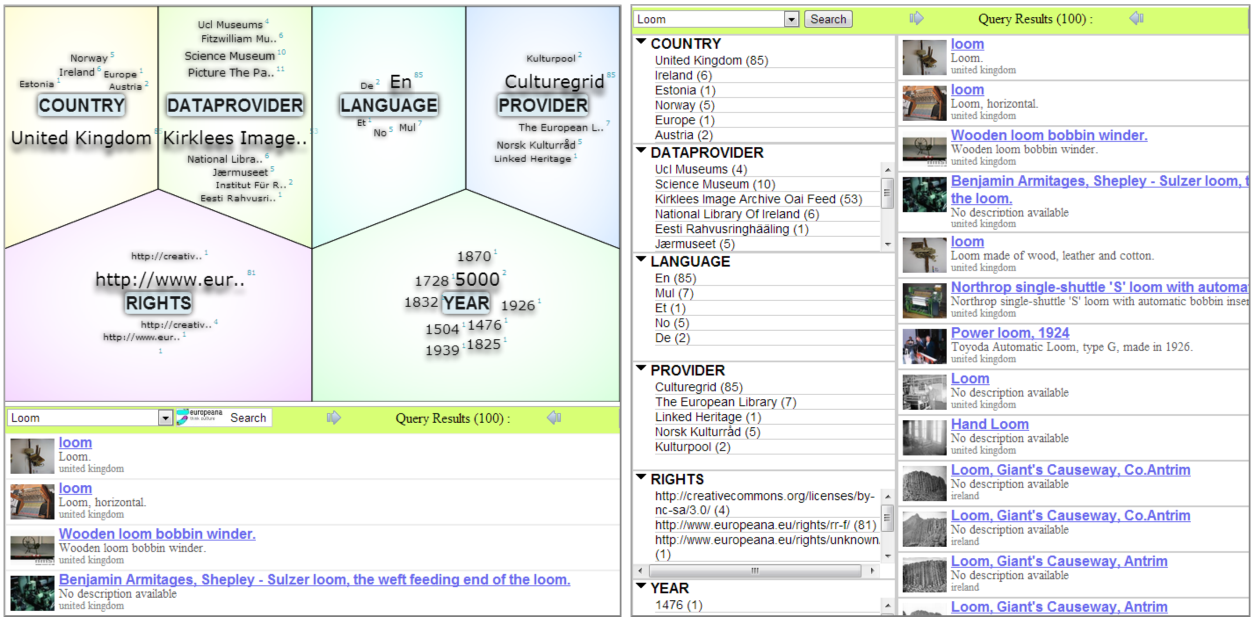

Figure 4. Compared interfaces for faceted navigation of search results: FacetScape (left) and TreeView (right) both using the same screen space.

are not explicitly named in the task description. The correct answer sets comprised exactly one item. The tasks differ in the level of difficulty in the following way:

Task 4: Selection of only one value in one facet is required to find the result.

Task 5: Selection of one value in two facets is required to find the result (simple AND query).

Task 6: Similar to Task 5, but task description contains additional, necessary information for identifying the result, but this information does not refer to facets.

An example task description of Task 6 is: "Find a 114 page document about steam engines written in English". Besides implicit vs. explicit task descriptions the tasks span another, orthogonal dimension, namely the required search behaviour. In Tasks 1 and 4 users needed to exploit the search space, whereas Task 2 and 5 required more explorative search behavior. Task 3 and 6 then again needed an even larger rate of explorative behavior: In Task 3 the low frequent facet values were not initially visible in the interface and in Task 6 the additional, necessary information was given but was not reflected in the facets.

F. Results and Discussion

Table I shows an overview of the task completion time, the number of clicks and task success for all tasks for FacetScape (FS) and TreeView (TV). As can be seen in Figure 5 for task completion time, the data contains outliers, therefore we report the median in the table. Better values for success and statistically significant better values for time and clicks are marked bold.

Table I
OVERVIEW OF RESULTS FOR ALL TASKS. STATISTICALLY SIGNIFICANT DIFFERENCES ARE MARKED BOLD.

	time [sec]		clicks		sucess rate [%]	
	FS	**TV**	**FS**	**TV**	**FS**	**TV**
Task 1	34.1	37.6	3	16	67	67
Task 2	51.0	40.3	4	4	87	100
Task 3	76.3	**48.4**	5	4	73	67
Task 4	15.3	18.0	2	2	93	100
Task 5	24.6	24.0	3	3	93	100
Task 6	80.3	55.4	5	3	67	87

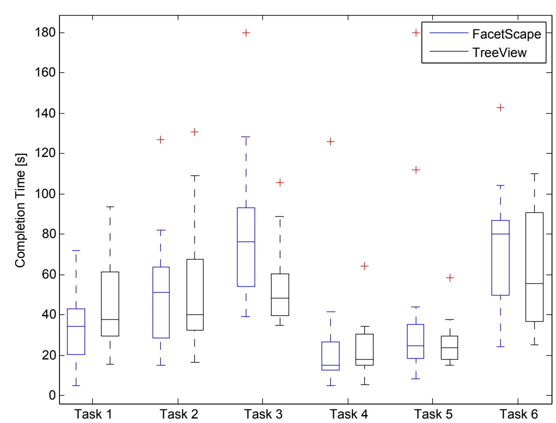

Figure 5. Overview of task completion time averaged over all users.

1) Efficiency and Effectiveness: Because the data is not normally distributed (Shapiro-Wilks test) we tested for statistically significant difference of median values using the two-

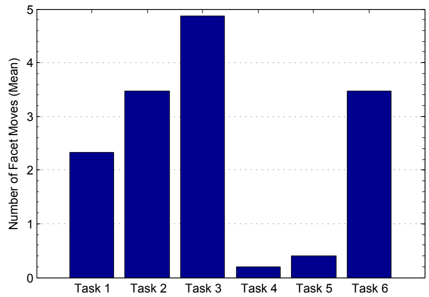

Figure 6. Number of clicks not related to query construction averaged over all users.

sided Wilcoxon rank sum test at significance level $\alpha = 0.05$. For all tasks but Task 3 we found no significant difference in the number of clicks, task completion time and task success rate. For Task 3 the completion time was significantly larger with the FacetScape (p=0.014). The FacetScape offers more interaction possibilities than the TreeView, which are not directly related to query construction, most importantly moving of a facet. We investigated how much the task completion time for the FacetScape depends on the number of facet moves as an estimate of the time spent for moving facets. Figure 6 depicts the number of average facet moves for each task. As can be seen, the number is largest for Task 3, in which the task completion time was significantly larger with the FacetScape. Also the nature of explorative search behaviour can be seen in this figure, with Tasks 3 and 6 having larger numbers of clicks for moving facets and exploring the search space. The Pearson correlation of completion time and number of facet moves is 0.66 (across tasks), a tendency that is also reflected when comparing Figures 5 and 6.

Summing up, we found that there is not much difference in the efficiency and effectiveness of users between the two interfaces. The difference in task completion time for Task 3 (FacetScape significantly slower) can be explained by spatial re-arrangements necessary to find the low-frequent facet values for solving this task. This indicates the necessity for better, initial placement of the Voronoi generator points, a finding that is confirmed by the quantitative evaluation.

Table II
OVERVIEW OF QUALITATIVE QUESTIONNAIRE RESULTS. STATISTICALLY SIGNIFICANT DIFFERENCES ARE MARKED BOLD.

		FacetScape	**TreeView**
Performance	Efficiency	2.5	2.5
	Accuracy	2.1	1.8
Presentation	Familiarity	4.1	**1.1**
	Helpfulness	2.6	2.0
	Beauty	2.5	3.2
General	Trust	2.4	1.9

2) Quantitative User Feedback: Table II summarizes the main results from the questionnaire. Because the data was not normally distributed, we tested for equal medians using Wilcoxon rank sum test at significance level $\alpha = 0.05$. Statistically significant differences are marked bold in the table. In terms of performance users did not perceive any difference in their efficiency between the two visualizations ($p = 0.81$). This is in accordance with the measurements on task completion time. Further, users felt that they could perform their tasks accurately, with no difference between the visualizations ($p = 0.46$), found both visualizations equally helpful ($p = 0.06$) and trusted both presentations equally ($p = 0.13$). However, we found that they were more accurate using the TreeView visualization in 4 of 6 tasks. Not surprisingly, users rated the TreeView as being much more familiar than the FacetScape ($p = 0.00$). We think this is because the TreeView is a standard faceted browsing interface in well-known websites and different file system browsers use the same visual approach.

In summary, the quantitative data of the questionnaire revealed nearly no difference between the two visualizations, and only confirmed the assumption of the TreeView being more familiar.

3) Qualitative User Feedback: In this section we report qualitative feedback and user suggestions from the questionnaire. Almost half of the participants (7/15) had suggestions how to improve the moving of facets. Most of them wanted to be able to drag and drop and move a facet by selecting any point inside the respective area. One participant would favor to move area edges instead of the area itself. Two of the participants did not like the blocking we implemented to avoid degeneration of the Voronoi diagram. One participant expected a fish-eye like distortion while zooming. Further, participants noted, that they missed visual clues for both interactions. Regarding the layout, 4 participants would prefer an overall layout similar to the TreeView with the FacetScape located on the left of the search result list. Two participants expected the search field to be on top of the display area. For the Voronoi layout 3 participants noted that it would be helpful to initially fit the size of the regions to the number of facet values. Finally, 3 participants would like to have the facet values sorted alphabetically instead of arranged by relevance.

In summary, the suggestions for the FacetScape were mostly directed towards an improved facet movement and an area distribution that better reflects the number of facet values.

V. CONCLUSION

In this paper we proposed a visualization for faceted browsing, which combines Voronoi subdivision and a tag cloud representation of the search facets. A comparative user study revealed that participants used the unfamiliar interface as efficient and effective as the familiar tree-like display. Thus, the greater variety of possible interactions in the visualization does not hinder search performance even for untrained users. The significantly higher completion time in one task (Task 3) can be explained with the non-optimal initial positioning of the Voronoi generator points, which we will address in the future. In a next step we will address scalability issues. While Europeana provides only six facets, other digital libraries provide richer metadata for their content. We expect the feature of moving facets out of focus to be of greater importance when the number of facets increases. For instance, the econbiz online library for economic content[3] currently provides 14 different facets, some of which are likely to be unimportant for the task at hand. When applying the visualization to other data we will take user comments into account which pointed towards missing visual clues for possible interactions. For example, the removal of facets was not obvious for many users and therefore the feature was not used.

Acknowledgment

The presented work was developed within the EEXCESS project funded by the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number 600601.

References

[1] D. Bounie and L. Gille, "Info capacity— international production and dissemination of information: Results, methodological issues and statistical perspectives," *International Journal of Communication*, vol. 6, no. 0, 2012.

[2] M. Taghavi, A. Patel, N. Schmidt, C. Wills, and Y. Tew, "An analysis of web proxy logs with query distribution pattern approach for search engines," *Computer Standards & Interfaces*, vol. 34, no. 1, pp. 162 – 170, 2012.

[3] M. Hearst, A. Elliott, J. English, R. Sinha, K. Swearingen, and K.-P. Yee, "Finding the flow in web site search," *Commun. ACM*, vol. 45, no. 9, pp. 42–49, Sep. 2002.

[4] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst, "Faceted metadata for image search and browsing," in *Proc. SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2003, pp. 401–408.

[5] J. C. Fagan, "Usability studies of faceted browsing: A literature review," *Information Technology and Libraries*, vol. 29, no. 2, 2010.

[6] M. C. Schraefel, D. A. Smith, A. Owens, A. Russell, C. Harris, and M. Wilson, "The evolving mspace platform: leveraging the semantic web on the trail of the memex," in *Proc. ACM conference on Hypertext and Hypermedia*. New York, NY, USA: ACM, 2005, pp. 174–183.

[7] D. Jonker, W. Wright, D. Schroh, and P. Proulx, "Information triage with trist," in *Proceedings of International Conference on Intelligence Analysis*, 2005, pp. 2–4.

[3]https://api.econbiz.de/v1/fields?scope=facet, accessed 2014-05-15

[8] G. Marchionini and R. White, "Find what you need, understand what you find," *Int. J. Hum. Comput. Interaction*, vol. 23, no. 3, pp. 205–237, 2007.

[9] C. Plaisant, B. Shneiderman, K. Doan, and T. Bruns, "Interface and data architecture for query preview in networked information systems," *ACM Trans. Inf. Syst.*, vol. 17, no. 3, pp. 320–341, Jul. 1999.

[10] E. Tanin, B. Shneiderman, and H. Xie, "Browsing large online data tables using generalized query previews," *Inf. Syst.*, vol. 32, no. 3, pp. 402–423, May 2007.

[11] P. Qvarfordt, G. Golovchinsky, T. Dunnigan, and E. Agapie, "Looking ahead: Query preview in exploratory search," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '13. New York, NY, USA: ACM, 2013, pp. 243–252.

[12] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tesselations: Concepts and Applications of Voronoi Diagrams*, 2nd ed., ser. Probability and Statistics. NYC: John Wiley, 2000.

[13] K. Andrews, W. Kienreich, V. Sabol, J. Becker, G. Droschl, F. Kappe, M. Granitzer, P. Auer, and K. Tochtermann, "The InfoSky Visual Explorer: Exploiting hierarchical structure and document similarities," *Information Visualization*, vol. 1, no. 3–4, pp. 166–181, Dec 2002.

[14] M. Balzer and O. Deussen, "Voronoi treemaps," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, 10 2005, pp. 49–56.

[15] O. Kaser and D. Lemire, "Tag-cloud drawing: Algorithms for cloud visualization," in *Proc. Workshop on Tagging and Metadata for Social Information Organization (WWW2007)*, Banff, Alberta, Canada, May 2007.

[16] C. Seifert, B. Kump, W. Kienreich, G. Granitzer, and M. Granitzer, "On the beauty and usability of tag clouds," in *Proceedings of the 12th International Conference on Information Visualisation (IV)*. Los Alamitos, CA, USA: IEEE Computer Society, 7 2008, pp. 17–25.

[17] S. Bateman, C. Gutwin, and M. Nacenta, "Seeing things in the clouds: the effect of visual features on tag cloud selections," in *Proc ACM conference on Hypertext and Hypermedia*. New York, NY, USA: ACM, 2008, pp. 193–202.

[18] 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 Computer/Human Interaction Conference (CHI)*. New York, NY, USA: ACM Press, 2007, pp. 995–998.

[19] M. Gavrilova and J. Rokne, "An efficient algorithm for construction of the power diagram from the voronoi diagram in the plane," *International Journal of Computer Mathematics*, vol. 61, no. 1-2, pp. 49–61, 1996.

[20] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *Proc. ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2005, pp. 154–161.