

A Theme Landscape for Tagged Data

Evelyn Münster

Dep. of Visualization

Ludwig Boltzmann Institute Media.Art.Research.

Linz, Austria

muenster@wangenrot.de

Abstract—The wide variety of visualization methods for numerical values have, up until now, outweighed the relatively modest selection available for nominal dimensions. Theme Landscape is a visualization application designed for tagged data that typically contains a large number of nominal dimensions. It provides an overview of large data volumes, doing so by positioning objects in a landscape according to how similar their tags are. HexBoard technology has been used for dimension reduction. A further distinctive feature of Theme Landscape concerns the icons. The tags that belong to an object can be identified according to the shape of its icon. Each keyword is assigned to its own unique glyph and these glyphs form, in any combination, a hexagonal icon. For the prototype a data archive of interactive media artworks has been used, which had been tagged using a taxonomy.

Keywords— *Information Visualization, Tagged Data, Media Art Archives, Glyphs, Dimension Reduction*

I. INTRODUCTION

For the development of Theme Landscape we had access to a database of interactive art compiled by the Ludwig Boltzmann Institute Media.Art.Research, which contained approx. 300 entries on media art, a taxonomy with approx. 150 keywords, and multiple keyword selections (see section III. C.). For the Theme Landscape, a thematic overview of all of the works recorded was to be developed in the form of an interactive visualization. The goal was to find a means of presentation able to make the similarities and differences of the works visually recognizable, based on their keywords. However, Theme Landscape was also intended to be a generic solution that can be applied to any kind of tagged data that has a similar structure.

One difference between Theme Landscape and most of the existing tag visualisations (such as the semantic map of Netzspannung.org [1] and Figd't Visualize [2], and as discussed by Wu [3] or Sánchez-Zamora [4]) is that Theme Landscape does not show the tags themselves but rather the tagged objects. This leads to the question of how the tags can be visually represented as an object's feature vector. Usually, nominal feature vectors are mapped to colours. A maximum of six to eight colours should be used in order to allow the user to easily identify and search the categories [5]. However, keyword catalogues usually contain many more keywords, not to mention free tagging with random terms. So if more than just a few keywords are to be visually highlighted, a different, more scalable solution must be

found. Text labelling quickly becomes confusing when an object has more than one or two tags. For this reason, Theme Landscape defines a glyph for each keyword, and the combination of these glyphs generates icons¹ with a high density of information. First, the glyphs are automatically assigned to the keywords. Then for each tagged object, an individual icon is composed of the glyphs. The glyphs can be combined in any way, which means that they can theoretically be used for any kind of keyword catalogue.

For this visualization, a landscape metaphor has been used. Many of the brain's normative structures show a planar topology. It is created by multidimensional impressions from the real world being projected onto a surface. The way in which the HexBoard technology (see section V.) used for Theme Landscape functions is based on the same principle. This technology reduces any amount of keywords within a work, i.e. any number of dimensions, onto a two-dimensional surface. This allows a thematic landscape to be created in which similar works are positioned close to one another.

The Theme Landscape prototype is a client-server application with a Flash user interface, a Java backend and a MySQL database (see section VI.). The source code has been published as an open source that is subject to the GNU General Public License v3[6].

¹ The term "icon" is considered by the author as the most appropriate in this context and will be used hereafter.

II. THEME LANDSCAPE IN ACTION

Before going into detailed explanations of how the data is mapped to the visualization, let's first take a look at the outcome, the Theme Landscape user interface.

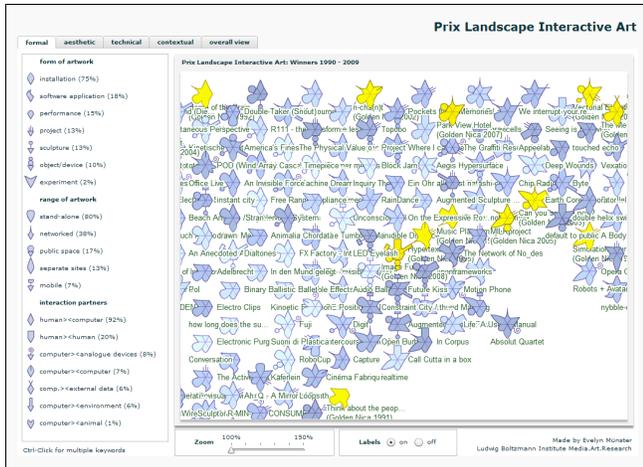


Figure 1. Overview of the User Interface

The prototype of Theme Landscape has been named "Prix Landscape Interactive Art" and shows approx. 300 media artworks that have been tagged with an interactive art taxonomy (see section III. C.). In order for changes to be immediately visible the data is dynamically loaded from the database at each request.

The application has multiple tabs whereby each tab represents a particular section of the taxonomy. The landscape of the media art projects is recalculated in each of the tabs based on the keywords of the section concerned. The graphical elements are identical in each tab view:

- *Legend* is on the left, comprising of a list of all the keywords in the current section and a distinct glyph for each keyword. Next to it, in brackets, is the assignment frequency as a percentage.
- To the right of the legend is the actual *Theme Landscape* itself, in which an icon has been placed for each artwork. These are positioned using the HexBoard algorithm in order that similar works are positioned close to one other.
- Each *icon* consists of six glyphs, which represent the tags of an artwork. The brightness of the icon colour shows the date of the artwork: the brighter the icon the older the artwork. In addition, the icons for those artworks that have won a Golden Nica, the highest award of the Prix Ars Electronica, are a golden yellow colour.
- The title of the artwork is below the icon, in the form of a *label*. The labels can also be hidden in order to maintain a clear overview.
- When the mouse is moved over an artworks icon, all the keywords in the selected section that have been assigned to this artwork will be highlighted in the

legend. The other way round, if the mouse is moved over a keyword then all the works that have been tagged with this keyword are highlighted. It is also possible to select several keywords simultaneously: in this case the intersection of the works is highlighted.

- When a work's icon is clicked on, a detailed view of the work appears with additional data: Artist, year, description and a small image.
- The size in which the icons are shown can be altered using the zoom slider positioned underneath the Theme Landscape.

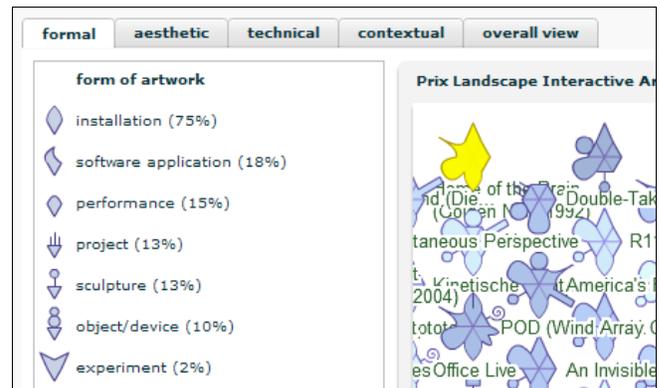


Figure 2. Detail View of the User Interface

Theme Landscape was designed as a module to be incorporated into an archive portal. This portal was meant to provide – alongside the traditional access via a menu – a search feature and detail views of the works. Theme Landscape is intended on the one hand to give a summary of the total amount of existing data available and on the other hand show search results, and when a work is clicked on redirect to that work's detail page. Unfortunately this portal has not yet been realized.

III. TAGGED DATA

A. Mapping the Data

For the visualization, a list of 25 unique glyphs has been designed. This list has been ordered by simplicity of the glyphs' design, i.e. the simplest glyphs come first. On the data side, the keywords have been ordered by their frequency of occurrence as tags, with the most frequently used coming first. These two lists are automatically and dynamically mapped, so that the first glyph is assigned to the first keyword, the second glyph to the second keyword, and so on. For the legend display, this list of glyph-keyword items has been grouped into existing taxonomy categories. Each artwork is tagged with 0 to 6 keywords and is visualised by a hexagonal icon that consists of 6 glyphs. A segment of a circle is used as default glyph to build icons with less than 6 glyphs. The assembly of an icon starts with

putting the glyph for the most frequently used keyword at the bottom and continuing clockwise. The resulting icons are placed in the landscape using the Hexboard algorithm (see section V.).

B. Suitable Data Structures

The technique of tagging, such as content indexing, has been used more and more frequently over the past few years. In a wide variety of fields entities are usually given keywords manually, in order to integrate them into a complex, non-hierarchical system. The aim is to make the content easier to search through and to spot similar objects, but also to gain an overview of the data available. The processes, tools and data structures used for tagging are just as diverse as the fields of application. In addition to a defined keyword catalogue, a taxonomy or a thesaurus, there are freely definable keywords that – through their usage in social tagging – in ideal cases create a folksonomy. Sometimes the keyword catalogue is very straightforward, however it might also contain several thousand entries. The keywords can be assigned to just a few objects, or to hundreds of them. The tagged data can be in the form of blog entries, bookmarks, news articles, cars, music albums, photos, or artworks, to name just a few. The tagging can be edited by specialists, or by anyone using web 2.0. We will summarize this in the following section under the term ‘tagging’.

Each tagged data pool consequently has its own individual structure, which has implications for the suitable method of visualization. The following should be observed when using Theme Landscape:

- The 1200 x 700 pixel window is only able to display a maximum of 300 objects. If there are more objects they should only be represented by points. The icons can then be displayed at a certain zoom level.
- Due to the hexagonal shape of the icons a maximum of six tags can be displayed at one time. The remaining tags will still be used to calculate the coordinates, but remain invisible.
- Twenty-five different glyphs have been developed for the prototype. Should there be more than twenty-five keywords they can either be divided up into several views, as was the case in the “Prix Landscape Interactive Art”, or further glyphs can be created, e.g. for an expert system. Alternatively the less important keywords can be given a default glyph.
- Keywords that are either almost always, or hardly ever used seem to be of little importance for the Theme Landscape. It might be possible to define a threshold value.
- If there are no more than eight different keywords it might be preferable to use a colour code.
- If the number of tags for an object is on average lower than six, it might be possible in certain circumstances to use decorated icons (see section IV.).
- If the objects in question are texts, images or videos, one should check whether or not it might be

preferable to use traditional forms of representation in the form of titles and subtitles or thumbnails instead of the icons presented here.

- Finally, if the keyword catalogue has a hierarchic structure or weighting then there is sure to be a more suitable method of visualisation.

C. Example: Taxonomy of Interactive Art

During the scholarly review of art-history at the Ludwig Boltzman Institute for Media.Art.Research, a taxonomy of interactive art [7] was created for the characterization of media art artworks using nine different categories, which contained a total of 150 possible keywords. In turn the categories have been summarized in a formal, an aesthetic, a technical and a contextual view or section [8]. This has provided us with a hierarchical taxonomy with three levels, whereby only the third level contains keywords that can be assigned.

The winning projects of the annual Prix Ars Electronica in the category “Interactive Art” for the years 1990 to 2009 were tagged by media art experts using this taxonomy. In doing so, not more than 6 keywords per section were assigned for each of the total 298 works, resulting as far as possible in an equal and in terms of content also consistent attribute allocation. In addition, the database contains further information on the title, artist and the Prix year, as well as a small image and description. This data has been used for the detail view of the works.

IV. INFORMATION-DENSE ICONS

A defining characteristic of Theme Landscape is the icons. In addition to the renowned and controversial Chernoff Faces there is a series of glyphs used for the information visualization of multivariate data [9]. However, these systems are hardly suitable for nominal data since they map above all numeric values to size, length, angle and quantity, etc.

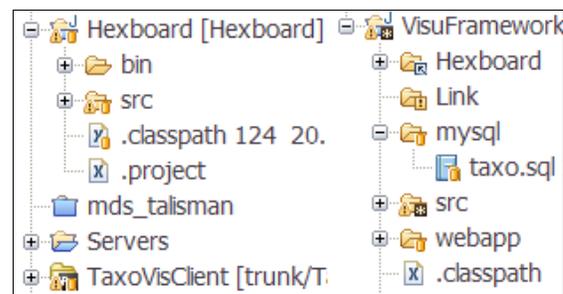


Figure 3. Decorated Icons of the Eclipse IDE

The Decorated Icons [10] used by the Eclipse IDE [11] however pursue a similar goal as the Theme Landscape. The decorations are used for nominal data attributes of files and include a large amount of glyphs and meanings, such as for example, whether the file is a link, a project directory, in version management, or has errors, has been altered or is binary. The decorated icons have proven themselves in practice, however they are only suitable for a relatively small

number of tags per object, because otherwise the space is quickly used up and the decoration elements overlap. Also, the elements can not necessarily be put together in any combination: This is why they are subdivided into main and additional categories, trying to ensure as much as possible that only mutually exclusive attributes share the same space in an icon.



Figure 4. Great variety of shapes by combining glyphs to make a hexagon

However, the glyphs in Theme Landscape can be combined in any order, as if in a building block system, to make a hexagon. Each glyph has its own individual form which shares only the same starting point and end point. Numeric variables such as length or height are not used to differentiate the glyphs, therefore they can be used for nominal dimensions. Besides, the capacity of the twenty-five glyphs currently available is far from being exhausted. The position of a glyph within a hexagon can vary because twenty-five or more glyphs have to share the six sections. The icons should, however, to be of a certain size² in order for the eye to be able to effortlessly differentiate between similar glyphs within a set of icons.

V. HEXBOARD

Each keyword in the keyword catalogue constitutes a dimension, and thereby the multidimensional vector space is created. Each dimension can only have two different values: 1 if a keyword is assigned to the object, 0 if an object does not have this keyword. This vector space serves as input for the classical multidimensional scaling algorithm (MDS) [12], which is used for dimension reduction. The initial attempt to display the absolute distances of the MDS directly in the visualization (Fig. 6) has not been satisfying, because the icons were overlapping, making it very difficult to compare the shapes.

To solve these problems, the innovative HexBoard visualization tool [13] was introduced³. The absolute distances, which have been computed by the MDS, serve as input. The HexBoard structures the elements into a hexagonal grid of fixed size cells. Each cell holds only one element, thus neighbouring elements are always placed at the same distance from each other. So we obtain a regular hexagonal grid where similar icons are placed adjacently. However, it must be taken into account that absolute distances and also clustering will be lost and thereby the information on how similar the elements are exactly. But this

² For the prototype, a 20 pixel radius stood the test.

³ There is a third alternative which is often used for tag visualizations: a SOM (self organizing map) with thematic regions. But this was no option for Theme Landscape, as the keyword allocation characteristics did not allow to choose main keywords to form the regions.

tradeoff gives Theme Landscape the big advantage of the icons not overlapping, making comparisons more feasible:

"The incBoard [the predecessor of HexBoard] visualization does not suffer from occlusion of elements, and allocating a pre-defined screen space to each cell enables using sophisticated and highly informative glyphs." [14]

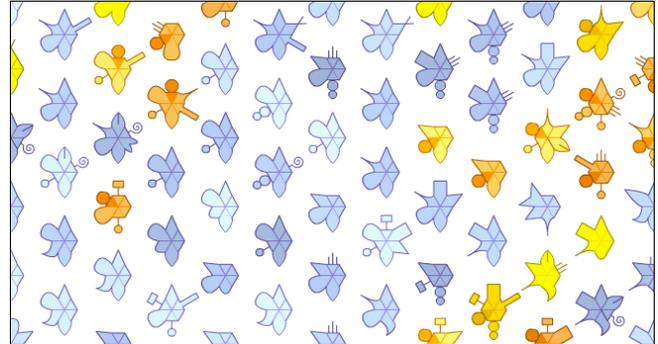


Figure 5. No overlapping using the Hexboard algorithm

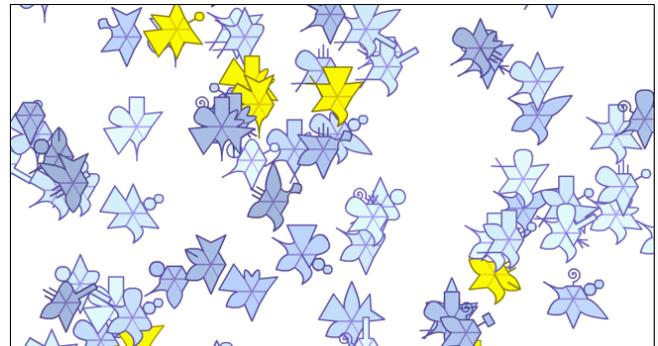


Figure 6. MDS positioning with absolute distances: occlusions, but clusters emerge

In addition, in the HexBoard technology each comb is assigned a maximum of one element. Therefore, identical icons do not form a stack but are positioned next to each other. First, this has the advantage of there being no three-dimensional stacks, and the resulting more complex user interface in terms of labelling and detail views can be omitted. Second, the number of elements always corresponds to the size of the landscape.

We decided to use HexBoard because with the MDS version of the theme landscape – despite having a sophisticated zoom function – it was not possible to recognize the icon forms satisfactorily. In addition, there were only a few convincing clusters in the data pool of interactive media art artworks available. The similarity structures are more complex and much more easily recognizable in the HexBoard version.

VI. TECHNICAL IMPLEMENTATION

The application has a classic three-tier architecture, comprising a front end, an application server and a

persistence layer. The persistence layer is a MySQL database, which is accessed by the application server using Hibernate. The data is read dynamically from the database so that changes are immediately visible in the visualization. Further data sources can be integrated modularly into the system if required. For the application server Java, Spring and Hibernate have been used on a Tomcat webserver. The domain objects are processed here by various logic components, for example the HexBoardService calculates the coordinates. The front end is a Flex based Flash client, which sends its requests to services of the application server via BlazeDS [15] and receives the responses in the form of objects that have been prepared for the presentation. This means the client need only make a minimal number of computation-intensive processes. In custom Flex components, glyphs are created and dynamically assigned to the keywords, the icons are composed, and the objects are displayed in the landscape, to name but a few tasks. For the graphic display of the visualization Flare API, a Flash version of the Prefuse Visualization Library is used [16]. Finally, the detail data is called up from the server on demand.

VII. USE AND USABILITY

The prototype has been used and evaluated by various different people over a period of several months, internally at the institute by a team, but also within the exhibition Mapping the Archive⁴, where the Ars Electronica audience had access to the prototype. Two typical workflows were observed:

In the first workflow, the users select a work that they are interested in. They look at the surrounding landscape to see if there are similar works, and also in what aspects they are similar. In this way the work can be explored using different tabs, or respectively different thematic perspectives.

In the second workflow, the user calls up an overview of the works available. By clicking on different keywords and their multiple selections the user is able to recognize structures. For example there may be typical category combinations or also exclusions. There might be works with a very unique selection of tags but also many that have the same tags. The proportions can of course also be easily ascertained. As a result, the user develops a topological mental map that makes the data space conceivable.

In addition, Theme Landscape could be used for the evaluation of keyword catalogues. A visual analysis helps ascertain when and how often keywords are used as tags. And whether there are mutually exclusive or correlating tags, or to what degree the amount of tags per object fluctuates, and much more.

All the test users initially had difficulties understanding the presentation because Theme Landscape breaks with three learnt principles:

- Tags in the form of nominal attributes are usually coded with colour values. The use of glyphs does not meet this expectation. In addition, the icons are harder to decode, which makes them less visually ascertainable than colour codes.
- In a topological landscape map the user also expects, in addition to the individual elements, regions that are differentiated from one another. Due to the distribution characteristics of the tags, and the fact that the keywords are not weighted within the taxonomy, this is not possible with the data pool used here. A hierarchic classification into regions requires an explicit or implicit hierarchy of the tags, and is therefore contrary to the concept of tagging, which in fact overcomes the weaknesses of a strictly hierarchical structural system through precisely these multiple classifications.
- The user expects the tags themselves to be visualized, in the form of a tag cloud or as a network. The representation of tagged data with its implicit connections is an additional complex step that requires explanation.

However, after examining and using it more intensively, these difficulties understanding the tool were quickly overcome. If one can remember which glyph goes with which keyword this makes it much easier to comprehend the works icons. For this reason it would be wise to match the shapes of the glyphs as far as possible with the meanings of the keywords. It is essential that the coordinates of the icons remain static so that the users can develop their own mental map over time. This means that for a specific object selection and tag selection the landscape must always be identical. In the case of small changes to the data there should only be a moderate shift in the landscape. This has been achieved with the HexBoard algorithm used. All in all Theme Landscape provides a good overview of a tagged data pool. The traceability of the similarities between objects using icons with a higher information density supports the development of the user's own mental maps.

VIII. CONCLUSIONS AND FUTURE WORK

With the combination of a two-dimensional landscape and icons with high information density Theme Landscape is a new tool for the visualization of tagged data. The results so far are encouraging enough for further development of the concept. The software should be transformed from the prototype into a stable release that is configurable and can be used for a wide range of data. The adoption of Theme Landscape in other domains is sure to bring further insights. The visualization of music albums at last.fm or Bookmarks at Delicious is also conceivable.

The following functional extensions and improvements appear to be particularly promising:

- Integration of Theme Landscape in a search function as a search results display alternative to a hit list.

⁴ A huge poster themed "30 year ars" was exhibited at the History Lounge of the Ars Electronica Festival 2009 in Linz, Austria. Besides several other visualizations by the Ludwig Boltzmann Institute Media.Art.Research a print version of Theme Landscape was displayed. Additionally the interactive screen version was accessible.

- Assignment of the glyphs to a tag according to content, i.e. by hand, while at the same time expanding the glyph library if necessary.
- Improvement of labelling in order that the labels of an icon are completely readable without the zoom function having to be used.

ACKNOWLEDGEMENTS

My heartfelt thanks go to: Roberto Pinho and Maria Cristina Ferreira de Oliveira for providing the Hexboard API and their committed support, Katja Kwastek for providing the data and the evaluation of Theme Landscape, Dietmar Offenhuber and Moritz Stefaner for their feedback.

REFERENCES

- [1] The semantic map interface to the Netzspannung.org archive: <http://netzspannung.org/archive/>
- [2] Figd't Visualize: <http://figdt.com/visualize>
- [3] Ch. Wu. Visualization of Tag Sequence. In International Conference on Computer Science and Software Engineering, pages 341 - 344, 2008.
- [4] F. Sánchez-Zamora, M. Llamas-Nistal. Visualizing Tags as a Network of Relatedness. In ASEE/IEEE Frontiers in Education Conference, 2009.
- [5] L. T. Nowell. Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size To Convey Nominal and Quantitative Data. 1997
- [6] The Sourcecode is provided by Google Code: <http://code.google.com/p/visframework/>
- [7] K. Kwastek. Classification vs. Diversification. The Value of Taxonomies for New Media Art. In P. Gendolla, J. Schaefer: Beyond the Screen, Bielefeld, 2010.
- [8] Table of Taxonomy of Interaktive Art: http://vis.mediaartresearch.at/webarchive/images/projects/themelandscape/table_kwastek.pdf
- [9] Ch. Chen, W. Härdle, A. Unwin. Handbook of data visualization. pages 181-183, 2008, Springer.
- [10] The Eclipse Decorators have until now not been subject to scientific research. <http://www.eclipse.org/articles/ArticleDecorators/decorators.html>
- [11] The Eclipse IDE: <http://www.eclipse.org/>
- [12] J. Zhang. Visualization for Information Retrieval. Pages 143-163, 2008, Springer.
- [13] R. Pinho, M. C. F. Oliveira. HexBoard: Conveying Pairwise in an Incremental Visualization Space. In Proceedings Information Visualization IV09, pages 32-37, Barcelona, 2009
- [14] R. Pinho, M. C. F. Oliveira, and A. de A. Lopes. Incremental board: a grid-based space for visualizing dynamic data sets. In SAC 09: Proceedings of the 2009 ACM symposium on Applied Computing, pages 1757-1764, New York, 2009. ACM.
- [15] BazeDS: <http://opensource.adobe.com/wiki/display/blazeds/BlazeDS/>
- [16] Flare: <http://flare.prefuse.org/>