

Visualization Ideas

More Detailed Requirements(MDR)

- I have compiled requirements for the potential visual representation of hierarchical data after taking all the information(including info in this document) I know into account.
- Saw some 3-D representations of hierarchical data but some of the papers I read dismissed 3-D representations because of occlusion.

implicit	visualization method representing hierarchy in a space-filling way [55]	tree-maps (see Fig. 5)	Sunburst (see Fig. 6)
----------	---	---------------------------	-----------------------------

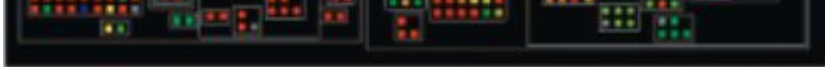
MDTA – How are Hierarchies Represented?

- From 2017 paper:
https://acris.aalto.fi/ws/portalfiles/portal/28853528/283_1_717_2_10_20180201.pdf
- Implicit vs Explicit, Axes vs Radial, 2 by 2 = 4 categories of visualizations of hierarchical data.
- Therefore, our visualization should fall under one of these categories.
- From experience, all the representations I saw were derived from this basic classification.



MDTA – Explicit Representation?

- Arguments against Explicit
 - Explicit Axes
 - Generally, less scalable when more data is added. We have lots of data.
 - Dendograms(**left pic**) and Indented Lists(**top right**) are examples of Explicit Axes representations.
 - https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9373115&casa_token=-HI7Z8uzFBUAAAAA:KFKb12-c9gwQpcrAugCJXsnB7oWf3XzZbiAU-5Zyt6cA23mw2Ci2CnOR9APtmeDdx8R3H9bHNA[paper from 2020] suggests that Explicit(axes or radial) methods are not good especially if there is a lot of hierarchical data. Circular tree is example of Explicit Radial representation(**middle right**).
- Arguments for Explicit
 - Easier than implicit to Trace Ancestry of leaf or node.
 - General Structure of Hierarchy Really apparent
 - Names of leaves can be seen well.



imported from METAREP and displayed using Krona. Taxonomy nodes are shown as nested sectors arranged from the top level of the hierarchy at the center and progressing outward. Navigational controls are at the top left, and details of the selected node are at the top right. The chart is zoomed to place the domain "Bacteria" at the root and the taxon "Gammaproteobacteria" is shown selected. An interactive version of this chart is available on the Krona website.

MDTA – Implicit Representation?

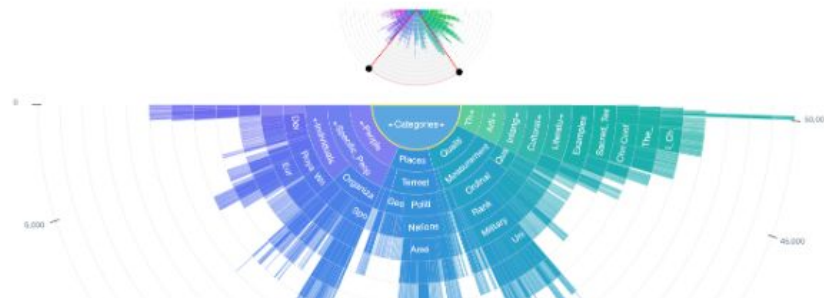
- Arguments against Implicit
 - Not so apparent when tracing ancestries.
 - Hierarchy not very apparent.
- Arguments for Implicit
 - <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9371413>[paper from 2022] suggests that if dataset has inner numeric values in the hierarchy, treemap(which is implicit) can be used.
 - https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9373115&casa_token=-HI7Z8uzFBUAAAAA:KFKb12-c9gwQpcrAugCJXsnB7oWf3XzZbiAU-5Zyt6cA23mw2Ci2CnOR9APtmeDdx8R3H9bHNA[paper from 2020] suggests that implicit representation is more suitable for lots of hierarchical data.
 - Treemap(**Left**) and Sunburst(**Right**) are examples of Implicit Representations.
- I saw more implicit representations than explicit.
- **VERDICT:** Implicit Representation and some explicit features and some interaction

Figure 2. Circular treemaps automatically generated by our variational algorithm for the hierarchical data with 6 levels and 2516 nodes. The color legend indicates the level of the hierarchy.

Fig. 4. Package structure of the `ngnias` project as primary hierarchy in comparison to a clustered decomposition as secondary hierarchy; top: primary hierarchy; middle: nested secondary hierarchies retrieved on demand for the `tracker` module (zoomed); bottom: direct main submodules of the `tracker` module (zoomed and clipped).

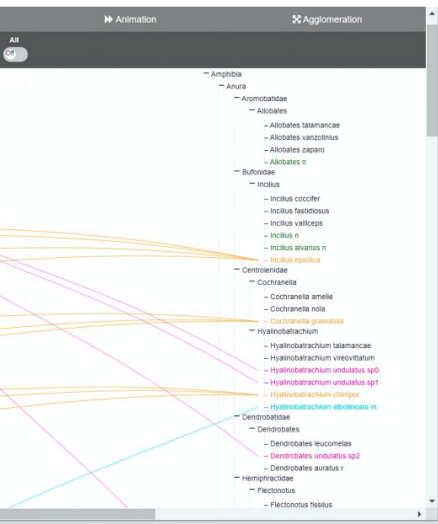
MDTA – Implicit Axes or Radial?

- https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9373115&casa_token=-HI7Z8uzFBUAAAAA:KFKb12-c9gwQpcrAugCJXsnB7oWf3XzZbiAU-5Zyt6cA23mw2Ci2CnOR9APtmeDdx8R3H9bHNA[paper from 2020] says that 17 people found:
 - for large hierarchies(**6 levels with 2-5 child nodes per level**), Sunburst and Circular Treemaps(**right**) are the best overall performers.
 - for small hierarchies(**2 levels with 10-20 child nodes per level**), Treemap, Icicle Plot and Circular treemaps(**left**) performed best.
- <https://arxiv.org/pdf/1908.01277>[paper from 2019] say that for hierarchical quantitative data, 12 users preferred a sundown chart(**down**)(semicircular variant of sunburst chart) and icicle plot over treemap.
- **VERDICT: Sunburst(or variant) or circular treemap**(Implicit) and explicit features.

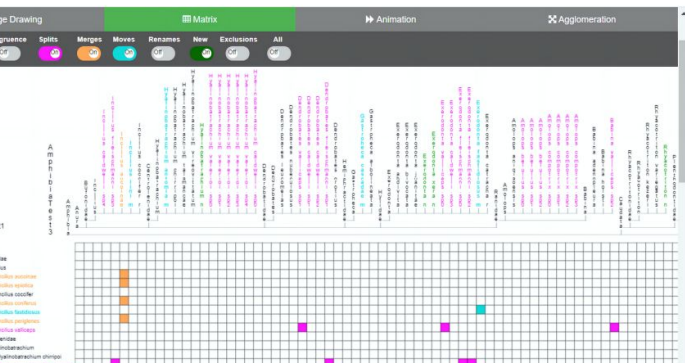


MDTA – Sunburst or Circular Treemap And Explicit Features. Multiple Views?

- <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f56e5596b634afab4d89c56dc26b4f69a3344f67> paper suggests that having different tree representations at the same time can be complementary(in regards to strengths and weaknesses of every visual representation) and users can walk away getting a full picture of the data.
- Verdict: **YES(for now)**



tion. Changes are indicated by relationships between the entation, relationships are visualized by color-coded edges ough the colored toggle buttons on the menu, users can es at once. Users can select a specific taxon in order to Full-size [DOI: 10.7171/peerj-cs.277/fig-3](https://doi.org/10.7171/peerj-cs.277/fig-3)



MDTA – Multiple Views of Sunburst/Circular Tree Explicit Features. Compare Hierarchies?

Survey of Experts, Paper from 2020:

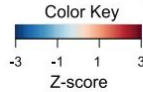
- <https://peerj.com/articles/cs-277.pdf>
- Paper evaluates four well known methods for the comparison of two hierarchies:
 - Edge Drawing
 - Matrix Representation
 - Animation
 - Agglomeration
- General themes:
 - Being able to identify the origin and destination of relations was very important to recognize changes when biological taxonomies.
 - Across exercises, experts referred to the importance of understanding what is going on at a glance.
 - Several experts commented that the methods could be complementary; for instance, that the edge drawing methods could be used to visualize all cases at once whereas the animation and the agglomeration methods when analyzing smaller groups of species. They explained that, by combining several methods, the advantages could overcome the disadvantages of another one. On the other hand, the experts also emphasized the need for both overview and detailed views; the first one to obtain a general understanding of changes and the second one to obtain detailed information on a focused part of the taxonomies.

This study contributes insights on the capacity of four visualization methods for hierarchy comparison in taxonomy curation tasks. Twelve expert taxonomists took part in a study and provided feedback. We performed a qualitative analysis. **The results clearly show differences among the methods, on both users' satisfaction: the edge drawing method was preferred over other methods.**

What should our hierarchical representation + comparison look like?

- Sunburst seems like the best candidate to represent one hierarchy given the data we have.
- Clustered Heat Map/Double Dendrogram will be good when comparing 2 hierarchies.
 - It has the 'explicit' element in that we will see labels of leaves very clearly while maintaining an understanding of the hierarchical structure and tracing ancestries.
 - We can cluster (like the namesake) parts of the tree to reduce what we are seeing. This abides by the “details on demand” principle in data viz.
 - Called a clustered heatmap, this is basically a matrix representation of when two hierarchies are compared.
- Basically, we have multiple views which can be complementary to the user.

Figure 2 The Krona RST display. The bacterioplankton metagenome from a vertical profiling of the North Pacific Subtropical Gyre [19] was imported from METAREP and displayed using Krona. Taxonomy nodes are shown as nested sectors arranged from the top level of the hierarchy at the center and progressing outward. Navigational controls are at the top left, and details of the selected node are at the top right. The chart is zoomed to place the domain "Bacteria" at the root and the taxon "Gammaproteobacteria" is shown selected. An interactive version of this chart is available on the Krona website.



Metabolite code

■ Lipid metabolism

■ Amino acid metabolism

■ Carbohydrate metabolism

□ Miscellaneous

■ Cofactors and Vitamins

Microbiome Hierarchy Viz 1 - 2011

● Use Case

- Link: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-385>
- Paper was not made specifically to show microbiome hierarchy but show metagenomic hierarchy in general.
- It just so happens that this paper has 1500+ citations on google scholar and there are countless papers that use the sunburst model(called **Krona**) that this paper produced.
 - Specifically, many papers(from what I gleaned) that cited this paper were/had
 - Biological in nature and not published at viz conferences.
 - Sunburst to represent microbiome hierarchy
 - Double Dendrogram(aka Clustered Heatmap) to compare one group vs another group in microbiome abundance. Assuming that there is an OG paper that put out the double dendrogram first.
 - This is a paper that cited **Krona**(there are many more like this):
<https://www.sciencedirect.com/science/article/pii/S2001037023003859>

Starting next slide, you will see my pros and cons for vizs I saw



General Hierarchy 1 – 2004

- Use Case

- Points/Leaves denote websites
 - Green: Updated recently
 - Red: Not updated recently
 - Height[in 3D]: Access Frequency

- Benefits

- Space-filling approach, aka space is not lost and everything is contained. Adding new nodes/leaves will not change size of container.

- Drawbacks

- Hierarchy not as explicit(will need some scanning to trace ancestry of leaves)
- Names not there
- Can't really tell percentages of nodes and/or leaves
- Will need to do 3D

- Link: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1272729>

hierarchy. The encircled area hints at relocation of part of the hierarchy; (b) an SE inspects the encircled area in more detail by crossing it. Zooming is subsequently performed to provide the SE with a more detailed view of the changes.

General Hierarchy 2 – 2008

- Use Case
 - Top and Bottom Structures are 2 hierarchies/2 versions of Azureus dataset.
 - Relevant to Software Engineers since directories and files can be represented as hierarchies
 - 2 hierarchies have 2283 and 3179 leaves respectively.(assuming leaves are files)
- Benefits:
 - Comparison of 2 hierarchies
 - Space filling approach
 - Apparent about similarities and differences in hierarchies
 - Red signals difference, Green Similarity
 - Edges show how hierarchies differ in structure, even though the hierarchies share majority elements
- Drawbacks:
 - Our dataset assumes that 2 hierarchies are very different, so we probably don't even need edges.
 - Animal kingdom hierarchy structure does not change with time. It is just a matter of which person has which subset of the hierarchy. For that reason, we don't need edges.
 - Not obvious about percentages of nodes and/or leaves.
 - Can't tell names of lower hierarchies immediately
- Link:
https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2008.01205.x?casa_token=USq6oahVzwIAAAAA%3ADSmWQmIEn8O7Cbi6VXoVt9_JI4jrBSMRfnispAPjsiwHGH_0065HjUQ7v-T_PTvEsA8e6zudqQ0Jugw

positionally equivalent nodes that form the number of overlapping pro and con units at the respective position as one-bar histogram (c). Adding d extends the aggregation resulting in an AOUT (e) surveying the argument structures of 3 essays.

General Hierarchy 3 – 2021

- Use Case

- Relevant to comparison of multiple essays and their structures where essays are represented as hierarchical structures

- Benefits:

- Compression of multiple hierarchical structures into one – Scalable
- Similarities and Differences are very apparent.
 - Red = Con, Green = Pro, White = None
 - For instance the node at the very left in Fig 6e shows that the majority of essays(in this case 2, since 3 essays are being compared) have a pro unit(green) at that node.
 - The node to the left of the above node shows that the majority of essays(in this case 2, since 3 essays are being compared) have no unit(red) at that node.

- Drawbacks:

- Not space filling approach so if one tree has many nodes and/or leaves, it won't look good.
- Assumes that nodes and/or leaves are similar across hierarchies. We may have hierarchies where nodes and/or leaves are **very** different.

Fig 2. a) Phylogenetic tree with 16 OTUs. b) Topo-phylogeny chart based on the same structure. The equivalent branches are highlighted red in both figures.

General Hierarchy 5 – 2017

- **Use Case:**
 - Used to represent phylogenetic hierarchy.
- **Benefits:**
 - Removes perception bias regarding leaf to leaf relationship in trees.
 - Intuitive and Unique
 - Space filling
 - Hierarchy and ancestral tracing very apparent
- **Drawbacks:**
 - Percentage of nodes and/or leaves not clear.
 - Will need to scan structure to find sub-hierarchies – especially when large trees are concerned.
 - Can't tell names immediately.
- **Link:**
<https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0175895&type=printable>

Figure 3. Here is a screenshot of the prototype. The blue numbers highlight the main parts of this interface whose roles are detailed in the paper. The central part contains a Stacked Tree (4,5,6). The left (1,2,3) and right (7,8,9) widgets allow to explore and to control this structure but are independent of the Stacked Trees. While this screenshot has been done on a small 13" screen, 2200 molecule are displayed with for each one up to three properties.

General Hierarchy 6 – 2012

- Use Case
 - Used to represent chemical hierarchies
 - Idea being that if main structure and leaves are the most important things in a given hierarchy, ignore/don't call much attention to stuff in the middle.
- Benefits
 - Fits more leaves on screen than space filling and not filling approaches.
 - Effective at high level overview without going into too much detail.
 - **UNIQUE: Neither space filling nor node-link**
- Drawbacks:
 - Does not communicate percentages of nodes and/or leaves.
- Link: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6295817>

and high-detail instance view. Observe that drama novels make less mention of fuel than adventure novels.

identification of overall abundant words. For example, most novels contain less words that describe something physical than those that describe something abstract, and words that refer to some form of matter mostly refer to a substance.

General Hierarchy 7 – 2011

- Use Case:
 - Allows for concurrent comparison of hundreds of cumulative weighted instances of a single hierarchy
 - Applied to novels structures
- Benefits:
 - Can be used to compare many hierarchies at once(similar to General Hierarchy 3)
 - Effective at showing commonalities
 - For instance, we could put 100 healthy people data into this hierarchy and if one node/leaf shared the rest, we can correlate healthiness(not necessarily causal) to the existence of that node/leaf
- Drawbacks:
 - Assumption of one single hierarchy is very different from our assumption that 2 hierarchies will be similar or different.
 - Does not communicate percentages of nodes/leaves.
 - Individuality of person is basically gone because hierarchy is geared towards finding high level agreement between individuals rather than focusing on individuals. However, there are workarounds to this.[See the left]
- Link:

https://online.library.wiley.com/doi/pdf/10.1111/j.1467-8659.2011.01862.x?open_access=true&url=/media/asset/10.1111/j.1467-8659.2011.01862.x/asset/BWPlsYc/m0AAAAA-25danYBf_dDiKifQ4tdK6

Enterobacter #2 28 2.45% 0.94%/2 genus

Fig 2. Visualization of Gammaproteobacteria present in two human gut samples. (a) Menu bar. (b) Treemap view of Gammaproteobacteria from samples F1-S and F1-T [6]. All phylogenetic ranks are displayed, nodes are colored by taxon name. (c) Synchronized table view of the same data.

General Hierarchy 8 – 2016

- **Use Case:**
 - Allows users to see phylogenetic trees in treemap version, either of an individual or an aggregation of individuals.
- **Benefits:**
 - Can be used to compare many individuals.
 - Space-filling
 - Two views cover each other's weaknesses
 - For instance, treemap not good at showing percentages but the table will do that – color coding is similar across views.
- **Drawbacks:**
 - Individual rectangles too small in treemap.
 - Ancestry can be traced but not too easily.
- **Link:**
<https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0158261&type=printable>

decomposition as secondary hierarchy; top: primary hierarchy; middle: nested secondary hierarchies retrieved on demand for the tracker module (zoomed); bottom: direct main submodules of the tracker module (zoomed and clipped).

General Hierarchy 9 – 2014

- Use Case:
 - Says that this viz will come in handy **IF** there is a *gradient of importance* of hierarchies: aka there is a reason why one hierarchy is more important than another.
 - Applied to Azureus dataset(see General Hierarchy 2, which uses a version of the Azureus dataset)
- Benefits:
 - Space filling
 - Ancestry can be traced
 - Good at showing structure of hierarchy and showing details if user asks for it.
- Drawbacks:
 - We don't have the same assumptions. We consider hierarchies equally important for now(**Correct me if I am wrong since I am assuming this on our part**)
 - Does not communicate percentages of nodes/leaves.
 - Will be hard to show taxonomic names
- Link: <https://ceur-ws.org/Vol-1244/GViP-paper3.pdf>

General Hierarchy 10 – 2019

- Use Case
 - Visualization of Biological Taxonomies
- Benefits:
 - Indented List. therefore, names can be seen
 - Edge Drawing, which was a recommendation/preference by experts over other hierarchical comparison methods
 - Multiple views, see the super small matrix on the top right.
 - Colors are used efficiently.
 - Ancestry can be traced in indiv trees
- Drawbacks:
 - Scrolling, which is impractical. Will not scale well with lots of data
 - Assumes overlap in hierarchies which we don't.
 - Not space filling
 - Does not communicate percentages.
- Link:
<https://www.vis.wiwi.uni-due.de/fileadmin/migratedchart3assets/file/carla2019-diafora.pdf>