Semi-Automatic Information and Knowledge Systems

: Hierarchical Data Visualization & Ontology Visualization

Monika Lanzenberger
• Information Visualization
• Hierarchical Data Visualization Techniques
• Ontology Visualization
• Alignment Visualization
Information Visualization

InfoVis is ...

... the process of transforming data, information, and knowledge into visual form making use of humans’ natural visual capabilities.

... the computer-assisted use of visual processing to gain understanding.

... providing the user with an overview first and then details on demand (↔ text).

... based on pre-attentive features (< 200ms).

Information Visualization is ...

... based on pre-attentive features (< 200ms).

InfoVis: Using space

- Visualization of abstract data (e.g., financial transactions, insurance risks, etc.) means to find spatial representations (2D, 3D).

- No inherent spatial structure available, so the designer/ user needs to decide which dimensions are represented by space: Mapping.
InfoVis: Kinds of Data

- Entities (e.g., people, terms) and relations (e.g., part-of, is-a)
- Both can have sets of attributes (duration, color, time, etc.)

- Types of attributes
  1. nominal, ordinal, interval, ratio
  2. Category data (nominal),
     integer data (ordinal),
     real-number data (interval & ratio)

- High-frequency versus high-structural

[Ware: Information Visualization, 2000]
Classification

Data Visualization Techniques

- Geometric
- Icon-based
- Pixel-oriented
- Hierarchical
- Graph-based

Distortion Techniques

- Complex
- Simple

Interaction Techniques

- Mapping
- Projection
- Filtering
- Link & Brush
- Zooming

Recursive Pattern: FAZ-Index (Jan. ‘74 - Apr. ’95)

[Keim, 2001]
Coupling views by:

- **Slaving**
  movements in one view are automatically propagated in the other views

- **Linking**
  connects the data items of one view with the data items of the other views e.g., done by **brushing**: user selects and highlights items in one view and the corresponding items are highlighted automatically

[Baldonado, 2000]
Visual Encoding Techniques

Different ways in encoding information visually:

• Space
  (See details next slide)
• Marks (in space)
  Points, lines, areas, volumes
• Connections & enclosures
• Retinal properties
  Crispness, shape, resolution, transparency, color, grayscale
• Temporal changes
• Viewpoint transformations

[Card, Mackinlay & Shneiderman, 1999]
• Composition

The orthogonal placement of axes, creating a 2D metric space

[Card, Mackinlay & Shneiderman, 1999]
Visual Encoding Techniques

- **Composition**
  The orthogonal placement of axes, creating a 2D metric space

- **Alignment**
  The repetition of an axis at a different position in the space

[Card, Mackinlay & Shneiderman, 1999]
Visual Encoding Techniques

- **Composition**
  The orthogonal placement of axes, creating a 2D metric space

- **Alignment**
  The repetition of an axis at a different position in the space

- **Folding**
  The continuation of an axis in an orthogonal direction
Visual Encoding Techniques

• Composition
  The orthogonal placement of axes, creating a 2D metric space

• Alignment
  The repetition of an axis at a different position in the space

• Folding
  The continuation of an axis in an orthogonal direction

• Recursion
  The repeated subdivision of space

[Card, Mackinlay & Shneiderman, 1999]
Visual Encoding Techniques

• **Composition**
  The orthogonal placement of axes, creating a 2D metric space

• **Alignment**
  The repetition of an axis at a different position in the space

• **Folding**
  The continuation of an axis in an orthogonal direction

• **Recursion**
  The repeated subdivision of space

• **Overloading**
  The reuse of the same space

[Card, Mackinlay & Shneiderman, 1999]
• Information Visualization
• Hierarchical Data Visualization Techniques
• Ontology Visualization
• Alignment Visualization
Hierarchical Techniques

Basic Idea: Visualization of data using a hierarchical partitioning into subspaces

Examples are:

- Dimensional Stacking [LeBlance et al. 1990]
- Worlds-within-Worlds [Feiner & Besherss 1990]
- Treemaps [Shneiderman 1992; Johnson, 1993]
- Cone/Cam Trees [Robertson, Mackinlay, Card 1991]
- Cheops [Beaudoin et al., 1996]
- InfoCube [Rekimoto & Green 1993]
Treemap

Screen-Filling Methods
- Hierarchical partitioning of the screen depending on the attribute values
- Overcoming space limitations

Alternative Partitioning
- x- and y-dim of the screen

Attributes - User-Defined
- for partitioning and their ordering

Color Correspond to Add. Attributes

Overview over
- Large amount of hierarchical data (e.g., file system)
- Data with multiple ordinal/quant. attributes (e.g., census data)

[Shneiderman 1992; Johnson, 1993]
Hierarchical Data

Trees:
- Ordered
- Acyclical
- Hierarchical

[Shneiderman 1992; Johnson, 1993]
Horizontal vs. Vertical

Horizontal
  • Corresponding to Text

Vertical
  • Traditional

Figure 5: A Standard 2D Tree.

Figure 6: A Rotated 2D Tree.

[Shneiderman 1992; Johnson, 1993]
Standard Representations

[Shneiderman 1992; Johnson, 1993]
Treemaps: Venn-Diagrams

Nested Treemap

[Shneiderman 1992; Johnson, 1993]
Example: File Structure to Tree

File System:
- 3 Folders
- 6 Files

Root --> whole Screen

[Shneiderman 1992; Johnson, 1993]
Example: File Structure to Tree

File System:
- 3 Folders
- 6 Files

Cutting - according to the size (30% and 70% of the space)

[Shneiderman 1992; Johnson, 1993]
Example: File Structure to Tree

File System:
- 3 Folders
- 6 Files

Iteration: folder and subfolder

[Shneiderman 1992; Johnson, 1993]
Example: File Structure to Tree

File System:
- 3 Folders
- 6 Files

One Solution

[Shneiderman 1992; Johnson, 1993]
Treemap: View Large Trees with Node Values

- Space filling
- Color coding
- Size coding

- Requires learning

[Shneiderman 1992]
Treemaps: Layouts

Figure 1: Top-Down, Size by Weight

Figure 2: Slice-and-Dice, Size by Unit

Figure 3: Slice-and-Dice, Size by Weight

Figure 4: Slice-and-Dice, no offsets

[Shneiderman 1992; Johnson, 1993]
Treemaps Variants

Figure 6: 2 1/2-D Treemaps

Figure 7: Treemap UNIX Experiment Results
Treemaps: Finance Analysis

http://www.smartmoney.com

Losers

- Health Management As: 11.11%
- TMP Worldwide: 4.40%
- Circuit City: 4.24%
- Brightpoint: 10.86%
- Avon Products: -7.83%
Treemaps: Finance Analysis

http://www.smartmoney.com

Gainers

- Family Dollar Stores: +5.09%
- Baxter International: +4.61%
- Tjx Cos: +5.47%
SequoiaView

http://www.win.tue.nl/sequoiaview/

Squarified Treemaps
3D Hyperbolic Space

[Munzner, 1998]
Hierarchical Clustering Explorer

for Interactive Exploration of Multidimensional Data

http://www.cs.umd.edu/hcil/multi-cluster
• Information Visualization
• Hierarchical Data Visualization Techniques
• **Ontology Visualization**
• Alignment Visualization
Ontology Visualization Tools

Protégé plug-ins

- **OntoViz** tab [7]
- **Jambalaya** [8]
- **TGViz** [9]
- **OWLViz** [10]
Ontology tools applying unconventional visualization techniques

- **The cluster map** [2] applied in *Autofocus* [1], *Spectacle* [2], the *DOPE* project [3], and *SWAP* [4].

- **Ontorama** [5] is a hyperbolic-style browser designed to render RDF files derived from a web-accessible ontology server called WEBKB-2, which contains descriptions of over 74,500 object types from WORDNET.

- **Ontobroker** [6] utilizes a hyperbolic tree view and is an ontology-based semantic indexing and instance querying technology for the WWW.
Ontology Visualization Tools (3)

Graph-based visualization tools:

- **WebODE** [12] uses the tool called OntoDesigner to graphically edit ontologies using common node/edge to represent the concepts and the relations in a tree.

- **Tadzebao** [13], which is a tool for collaborative development of ontologies, includes the tree-tool WebOnto.

- **FCA** [14] uses simple node-link visualizations of the inherent structure.

- **Conzilla** [15] and **VizCo** [16] apply RDF-graphs to create and manipulate ontologies.

- **Vizigator** [17] represents topic maps using the **Touchgraph** technology [18].

- **ViSWeb** [19] is an OPM-based (Object-Process Methodology) layer on top of XML/RDF/OWL to express knowledge visually and in natural language.
Graph-based visualization tools:

- **ORIENT** (Ontology engineering Environment) [20] is an Eclipse-based system using RDF-graphs and includes ontology building, mapping, evolution, evaluation and visualization.

- **RDFAuthor** [21] supports the creation of RDF instance data by dragging the data into a graph and binding it together using a graphical and quite simple interface.

- **FRODO RDFSVis** tool [24], which provides class models of ontologies represented in RDF Schema using GraphViz

- Building ontology-based queries with different levels of guidance is the aim of **GODE** [25] (Graphical Ontology Design Environment)
Ontology Visualization Tools (5)

Graph-based visualization tools: **IsaViz** [22] relies on **GraphViz** [23] to browse and author RDF models presented as graphs.
Why Visualization?

Visualization techniques support by:

• direct manipulation of the classifications / concepts / instances
• providing with overview
• appropriate presentation of semantically rich query results
• visual support for exploration and querying
• focus on structure (metadata) or on data: different points of view
• efficiently comparing ontologies
• supporting creation of ontologies based on standards
Aduna AutoFocus

- Brings semantic, multi-dimensional information visualization (cluster map) to everyone’s desktop
- Lets users oversee and access the overwhelming amount of information
- Integration of different sources: local files, emails, websites, intranet resources
- Using a local Sesame RDF Repository + Aduna Metadata Server for sharing Metadata in enterprise environments
- Metadata: file type, size, date, author(s), sender, keywords
- [http://www.aduna-software.com/home/overview.view](http://www.aduna-software.com/home/overview.view)
Outline

• Information Visualization
• Hierarchical Data Visualization Techniques
• Ontology Visualization
• Alignment Visualization
OLA Visualization
PromptViz
[Lanzenberger et al., 2006]
URI: http://meh/tourism2#Erlebnisurlaub Entity label: Erlebnisurlaub URI: http://meh/tourism1#Erholungspausen Entity label: Erholungspausen Confidence = 0.547619047619048 Syntactic similarity: 0.6428571428571429 Similar Superclasses: 1.0 Similar Class Object Properties To: 1.0000000000000022 Correct Value = 0
URI: http://meh/tourism2#Erlebnisurlaub Entity label: Erlebnisurlaub URI: http://meh/tourism1#Aktivurlaub Entity label: Aktivurlaub Confidence = 0.45454545454545486 Syntactic similarity: 0.36363636363636365 Similar Class Object Properties To: 1.0000000000000022 Correct Value = 1
URI: http://meh/tourism2#Erlebnisurlaub Entity label: Erlebnisurlaub URI: http://meh/tourism1#Kremerfahrt Entity label: Kremerfahrt Confidence = 0.29584910972503153 Syntactic similarity: 0.16666666666666666 Similar Superclasses: 1.0 Similar Class Object Properties To: 1.0000000000000022 Correct Value = 0
URI: http://meh/tourism2#Erlebnisurlaub Entity label: Erlebnisurlaub URI: http://meh/tourism1#Schwimmen Entity label: Schwimmen Confidence = 0.166666666666666705 Similar Superclasses: 1.0 Similar Class Object Properties To: 1.0000000000000022 Correct Value = 0
URI: http://meh/tourism1#Immaterialies Entity label: Immaterialies URI: http://meh/tourism2#Immaterialies Entity label: Immaterialies Confidence = 1.0 Syntactic similarity: 1.0 Similar Superclasses: 1.0 Similar Class Object Properties To: 1.0000000000000022 Correct Value = 1
URI: http://meh/tourism1#Immaterialies Entity label: Immaterialies URI: http://meh/tourism2#Situation Entity label: Situation Confidence = 0.47256039045316767 Similar Superclasses: 1.0 Similar Subclasses: 0.835623427190036 Similar Class Object Properties To: 1.0000000000000022 Correct Value = 0
URI: http://meh/tourism1#Immaterialies Entity label: Immaterialies URI: http://meh/tourism2#Raumliches_Konzept Entity label: Raumliches_Konzept Confidence = 0.35915492957727985 Similar Superclasses: 1.0 Similar Subclasses: 0.15492957746367686 Similar Class Object Properties To: 1.0000000000000022 Correct Value = 0

[Lanzenberger et al., 2006]
### Introduction: Relations Among Concepts

<table>
<thead>
<tr>
<th>OWL Ontology Construct</th>
<th>Comparison Relationship</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept</td>
<td>Equal</td>
<td>URI’s equal.</td>
</tr>
<tr>
<td></td>
<td>Syntactically equal</td>
<td>Class member instances equal.</td>
</tr>
<tr>
<td></td>
<td>Similar</td>
<td>Labels are the same.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Superclasses are the same.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subclasses are the same.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data properties are the same.</td>
</tr>
<tr>
<td></td>
<td>Broader than</td>
<td>Object properties are the same.</td>
</tr>
<tr>
<td></td>
<td>Narrower than</td>
<td>Similar low/high fraction of instances.</td>
</tr>
<tr>
<td></td>
<td>Different</td>
<td>Subclass superclass comparison.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Superclass subclass comparison.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class is different from all classes of the second ontology.</td>
</tr>
</tbody>
</table>

[Lanzenberger et al., 2006]
Tasks in Visual Ontology Alignment

- Read / assess / correct alignment result
- Examine the context of entities for both source ontologies
- Manipulate source ontologies (change labels, URIs, etc.)
Neither completely regular nor completely random: Regular graphs 'rewired' to introduce increasing amounts of disorder

Two characteristic features: clustering coefficient high and average path length short

Variety of edge lengths, with shorter lengths for edges in tight clusters, longer lengths for random edges between clusters

[Watts & Strogatz, 1998]
Small-world phenomenon: according to Milgram each actor in a social network is linked to any other with a maximum of 6 intermediaries. Experiment in 1967 suggested that two random US citizens were connected on average by a chain of six acquaintances.

Smaller communities, such as mathematicians, are densely connected: Mathematicians use the Erdös number to describe their distance from Paul Erdös based on their shared publications. The Erdös Number Project: http://www.oakland.edu/enp/
Small World Graphs

• Based on a spring-embedded algorithm that position tightly coupled groups of nodes closely together and loosely coupled groups of nodes far apart

• Uses clusters to group the nodes of a graph according to the selected level of detail (degree of abstraction DOA ∈ [0, 1])

• Distance between two clusters of nodes is inversely proportional to their coupling (LinLog)

• Average link uses the average distance between all members

[van Ham, et al. 2004, Noack 2003]
Small World Graphs

All spring-embedded algorithms bear the problem of high computational complexity - usually $O(N^3)$, Optimization: $O(N^2 \log(N))$

Clustering the graph improves program's interactivity: On average there are only $O(\log(N))$ clusters visible

Users’ Goals:
- Are there any distinct groups of items that are strongly interconnected (i.e. graph clusters)?
- How do these split into separate clusters?
- How do these clusters relate?

[van Ham, et al. 2004]
AlViz Implementation

- Tab widget plug-in for Protégé 3.2
- AlViz links four views in order achieve a better integration of overview and details
- Represents the entities linked together according to selected mutual properties such as IsA, IsPart, IsMember, locatedIn, hasOwner, isMadeBy, ...
- Color encodes alignment type

[Lanzenberger et al., 2006]
• Reduced saturation indicates mixed clusters

• Different levels of detail (degree of abstraction)

• Shape and size of cluster represents number of nodes

• Implementation: 2D graphs
  (based on implementation from Stephen Ingram)
[Lanzenberger et al., 2006]
[Lanzenberger et al., 2006]
Focus on a certain entity, visualization the entity and its context

Small world graph visualizations of two ontologies in the tourism domain: the focus of the graph is on the entity 'Urlaub' showing all related concepts for both ontologies

Labeling is activated

This view includes all sub-entities (transitive relation) and directly related entities (non-transitive relation), supplemented with all relations and entities among them within a beforehand defined number of hops (relations)

The nodes are not clustered meaning each node of the graph represents one entity
• The edges represent three different types of relations. In tourismA the depicted relations are: *IsA, hatReisedauer, hatZiel, hatReisemittel*. In tourismB *IsA, hatEineDauer, manBenoetigtAusruestung, hatEinZiel*.  

• The IsA paths are shorter than the other because we gave these edges a higher weight.  

• To distinguish different types of relations such as functional, transitive, or non-transitive we apply different weights, which can be modified by the user according to the exploration needs.
• By moving the cluster sliders next to the graph the user can zoom in or out
• The number of aggregated entities is shown next to the label
• This example shows the clustering along the ‘IsA’ relations - transitive relations are clustered first
• Clustering emphasizes the structure of the ontology
• An iterative process of zooming in and out allows to explore the ontology on different levels of detail.
• Here clustering fades out the 'IsA' relationships among the entities focusing on the non-transitive relations of the central entity 'Urlaub'
• In tourismA 'Urlaub' is related to 'Gebiet', 'Fortbewegungsmittel', and 'Zeitraum'
• In tourismB the related entities are: 'Ausruestung', 'Root', and 'Zeitraum'
Many strengths:

- **Location:**
  Where do most of the mappings between ontologies occur?

- **Impact:**
  Do the mapping choices directly or indirectly affect parts of the ontology the user is concerned about?

- **Type:**
  What kinds of mappings occur between the ontologies?

- **Reason:**
  Why do this mappings exist?

... open issues ...
AlViz: Open Issues

• Show multiple associations
  (emphasized the ‘relatedness’ of ontologies)

• Pre-define weights of edges for groups of properties
  (e.g., transitive, symmetric, functional, inverse functional)

• Consider confidence value or correct value

• Use methods of graph analysis
to support the analysis of the alignments
AlViz: Open Issues

• Include focus+context techniques (e.g., distortion or SDOF)

• Labeling / Coloring of edges

• Stronger integration of AlViz and the alignment algorithm: re-calculate alignments?

• Detailed user testing


Thanks to ... Silvia Miksch for making nice slides on hierarchical data visualization available.