

Zusammenfassung für InfoVis II

Sommersemester 2017

von Dagmar Sorg

DATENVORVERARBEITUNG

1 Datentypen

Nominal

- keine quantitative Relation zwischen Kategorien
- Klassifikation ohne Reihenfolge

Ordinal

- Anordnung der Attribute möglich
- Abstand zwischen zugeordneten Werten hat keine Bedeutung

Numerisch

- Anordnung der Attribute möglich
- Abstand zwischen zugeordneten Werten hat Bedeutung
- mathematische Operationen möglich

2 Fehlende Werte

- Wert manuell nachtragen
- Attribut-Durchschnitt
- ignorieren
- globale Konstante
- wahrscheinlichster Wert

3 Normalisierung

4 Sampling, Datenreduktion

- random sampling
- stratified random sampling
- systematic random sampling
- cluster random sampling

5 PCA

- Finden der zwei Hauptkomponenten mithilfe von Eigenvektoren und einer Kovarianzmatrix
- problematisch bei Datensätzen mit Gleichverteilung
- Outlier werden ignoriert
- Distanzen haben Bedeutung
- Richtungen haben Bedeutung
- Verlust der Information welche Dimension welche ist
- Landscapes: Dimensionen werden durch PCA bestimmt



6 MDS

- relative Entfernungen
- Projektion in 2D
- Distanzen haben Bedeutung
- Positionen entsprechen nicht unbedingt richtigen Positionen

VISUALISIERUNGSPROZESS

Ausdrucksstärke (Expressiveness) alle und nur die vorhandenen Informationen werden gezeigt

Effektivität Interpretation der Visualisierung kann schnell und akkurat erfolgen, kosteneffektiv

Visuelle Variablen

- Position
- Zeichen
- Größe (Länge, Fläche, Volumen)
- Helligkeit
- Farbe
- Orientierung
- Textur
- Motion (Flicker): in Kombination mit alle anderen visuellen Variablen

1 Menschliche Wahrnehmung

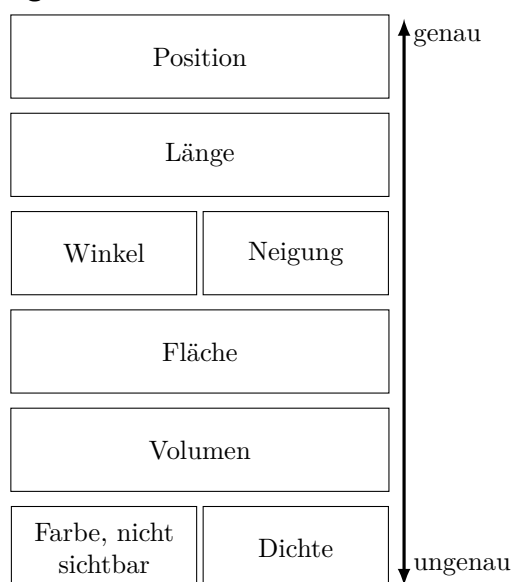
pre-attentive Vorverarbeitung

Gestalt-Laws

- Gesetz der Nachbarschaft (Proximity)
- Gesetz der Ähnlichkeit (Similarity)
- Gesetz der Verbundenheit (Connectedness)
- Gesetz der Fortführung (Continuity)
- Gesetz der Symmetrie
- Gesetz von Bild und Hintergrund (Figure and Ground)

Änderung der Blindheit bei Fokusverlust gehen Veränderungen verloren

Genauigkeit der visuellen Variablen



1 Hochdimensionale Daten

1.1 geometrisch-transformierte Displays

- Projektion von n Datendimensionen auf k Displaydimensionen
- jeder Record in der Projektion ist ein visuelles Zeichen (Mark)
- es gibt verschiedene Projektionstechniken (z.B. PCA, MDS)
- Beispiele:

Scatterplot-Matrizen

Landscapes

Force-Based Methoden / RaddViz

- schwierig für zu viele Dimensionen
- Beispiel: MDS
- möglicherweise Infos in Visualisierung aber nicht in den Daten

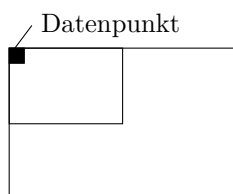
Tabular Displays

- Farbe, Größe/Länge als Datenkodierung
- Darstellung in Tabelle

Parallel Coordinates

Dimensionales Stacking

- n -dimensionale Daten werden auf 2-dimensionales Bild gemappt
- starten mit $2n + 1$ dimensionalen Daten
- feste Größe für jede Dimension festlegen
- eine Dimension: abhängige Variable
- alle weiteren Dimensionen: unabhängige Variablen



1.2 Ddense-Pixel Displays

- jeder gefärbte Pixel repräsentiert einen Wert
- alle Attributwerte in separaten Unterfenstern
- Anordnung der Pixel:
 - Peano-Hilbert
 - Morton (Z-Curve)
 - rekursive Pattern Technik
- Form der Unterfenster: Pixel des selben Datenpunktes sollen nah bei einander sein
- Anordnung der Dimensionen: ähnliche Dimensionen nah beieinander

1.3 Pixel Bar Chart

- Probleme:
 - bar charts zeigen nur einen Teil der Daten
 - Scatterplots haben viel overlap
 - pixel-displays brauchen Pixel-Zusammenhang
- Definition:
 - $\langle D_x, D_y, O_x, O_y, C \rangle$
 - D_x, D_y sind Teilungsattribute für x - und y -Achse
 - O_x, O_y sind Reihenfolgeattribute für x - und y -Achse
 - C ist (sind) Farbattribut(e)
- Formalisierung des Pixel-Positionierungsproblems:

Dense-Display-Constraint Display ist vollständig mit Pixeln gefüllt

No-Overlap-Constraint keine doppelte Belegung einer Position (jedes Pixel hat seine eigene Position)

Locality-Constraint ähnliche Pixel liegen nah bei einander

Ordering-Constraint Reihenfolge auf x - und y -Achse wird erzwungen
- Heuristik für Pixel-Positionierung:
 1. Quantile für x - und y -Achse bestimmen
 2. erstes Pixel unten links platzieren
 3. Pixel links und unten positionieren
 4. restliche Pixel positionieren

1.4 Color Icons

Array von Farbfeldern, die Attributwerte darstellen

1.5 Vergleich der Techniken

Kriterien

- Aufgabencharakteristiken (Clustering, multi-mannigfaltige Hot Spots, ...)
- Datencharakteristiken (Anzahl mannigfaltiger Datenitems, kategorische Daten, ...)
- Visualisierungscharakteristiken (visuelle Overlaps, Lernkurven, ...)

Vergleich

		cluster- ing	multi- variate hot spot	no. of variates	no. of data items	cate- gorical data	visual overlap	learning curve
Geometric Transformations	Scatterplot Matrices	++	++	+	+	-	0	++
	Landscapes	+	+	-	0	0	+	+
	Prosection Views	++	++	+	+	-	0	+
	Hyperslices	+	+	+	+	-	0	0
	Parallel Coordinates	0	++	++	-	0	--	0
Iconic Displays	Stick Figures	0	0	+	-	-	-	0
	Shape Coding	0	-	++	+	-	+	-
	Color Icon	0	-	++	+	-	+	-
Pixel Displays	Query-Independent	+	+	++	++	-	++	+
	Query-Dependent	+	+	++	++	-	++	-
Stacked Displays	Dimensional Stacking	+	+	0	0	++	0	0
	Worlds-within-Worlds	0	0	0	+	0	0	0
	Treemaps	+	0	+	0	++	+	0
	Cone Trees	+	+	0	+	0	+	+
	InfoCube	0	0	-	-	0	0	+

Hybrid-Ansätze

- möglicherweise zusätzliche Informationen darstellbar
- virtuell werden alle Visualisierungstechniken mit Dynamiken und Interaktionen kombiniert

Mehrfach-Ansichten

- wann?

Rule of Diversity Vielfalt in Attributen, Modellen, Benutzerprofilen, Abstraktionsleveln, Genres

Rule of Complementary unterschiedliche Sichten zeigen Korrelationen oder Ungleichheiten

Rule of Decomposition komplexe Daten in mehreren Sichten aufteilen für besseren Einblick in unterschiedliche Attribute

Rule of Parsimony (Sparsamkeit) minimale Anzahl an mehreren Sichten

- wie?

Rule of Space/Time Resource Optimization Nutzen der Darstellung muss mit dem Zeitaufwand zur Berechnung ausbalanciert sein

Rule of Self-Evidence Verknüpfung der einzelnen Ansichten

Rule of Consistency Interface und Sichten müssen konsistent sein

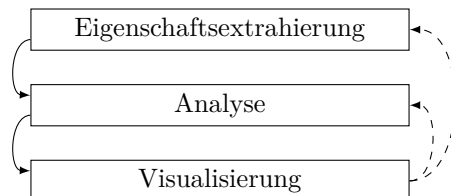
Rule of Attention Management Wahrnehmungstechniken verwenden, um Fokus des Benutzers richtig zu setzen

2 Techniken für Textdaten

Schwierigkeiten

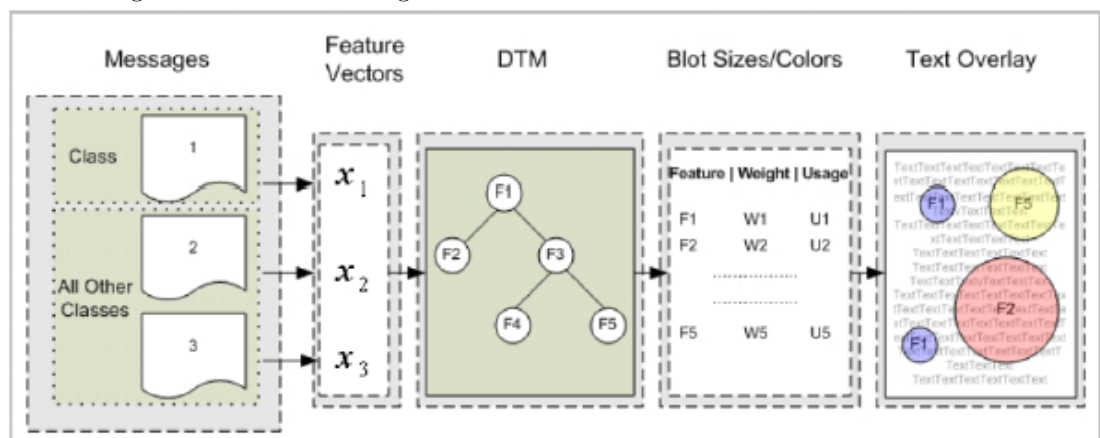
- nicht pre-attentive
- abstrakt
- sehr hoch-dimensional
- Aufbau (sehr unterschiedlich möglich)
- subtil

Stationen



Beispiele

InkBlots Texteigenschaft-Visualisierung



Plagiaterkennung Strukturanalyse

Cloudlines Text and Geo visuelle Event Analyse



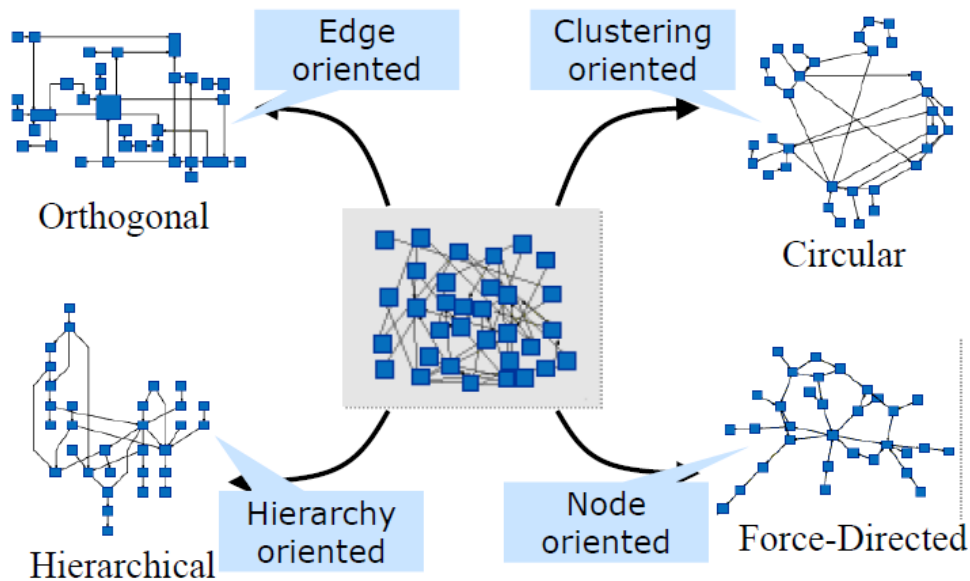


GLYPHEN

- unterscheiden sich von Icons, Indizes und Symbolen
- Mappings:
 - Many-to-One** alle Attribute werden auf die gleiche visuelle Variable gemappt
 - One-to-One** alle Attribute werden auf unterschiedliche visuelle Variablen gemappt
 - One-to-Many** manche Attribute werden auf mehrere visuelle Variablen gemappt
- Layout-Techniken (Positionierung der Datenpunkte):
 - strukturgetrieben
 - * Reihenfolge
 - * Hierarchien
 - * Netzwerke
 - datengetrieben
 - * Positionierung der Daten auf Grundlage der expliziten Werte
- Anwendungsgebiete
 - Linguistik
 - Sport-Analyse

NETZWERKE

- Zeichnungskonventionen



- Force-Directed Methoden (Spring Embedder)
 - Minimierung von Kantenkreuzungen
 - Minimierung der Gesamtlänge aller Kanten
 - maximale Kantenlänge einhalten
 - einheitliche Kantenlänge einhalten
 - Minimierung der Kantenknick
 - Minimierung von unterschiedlichen Winkeln/Krümmungen
 - Maximierung der Winkel zwischen zwei Kanten
 - Minimierung der Gesamtfläche der Zeichnung
 - Aspect Ratio einhalten
 - symmetrische Struktur
- hierarchisches Kantenbündeln
- NodeTrix: hybride Matrix Repräsentation (Matrizen als Knoten)

1 TreeMaps

Original

Squarified

- Rand proportional zur Fläche
- Aspect Ratio ist besser als bei *Original*
- Hierarchie schlechter erkennbar als bei *Original*

Cushion Form/Abdruck gibt Baumstruktur an

Ordered

- Items, die in der Baumstruktur nah bei einander sind, sind in der TreeMap adjazent
- Aspect Ratios sind besser als bei *Original*
- Glätte des Updates (smoothness of update) besser als bei *Squarified*
- Reihenfolge besser eingehalten als bei *Squarified*
- guter Kompromiss, wenn alles wichtig ist

Strip Variation von *Squarified*, geordnet

ZEITABHÄNGIGE DATEN

statisch historische, unveränderliche Daten

dynamisch Datenstrom, es können neue Daten hinzukommen (Wetterdaten)

1 Variation der x-Achse

Zeitpunkt diskret, keine Dauer

Zeitintervall zwei Zeitpunkte

lineare Zeit Startpunkt

zyklische Zeit Zeitabschnitte wiederholen sich

2 Fragen

Existenz Gibt es einen Datenpunkt zu einem gegebenen Zeitpunkt?

Lokalisierung Wann gab es den Datenpunkt?

zeitliches Interval Dauer eines Datenpunktes

Veränderungsrate Wie sehr hat sich ein über Zeit Attributswert verändert?

Sequenz In welcher Reihenfolge tauchen die Datenpunkte auf?

zeitliche Textur Wie oft kommt ein Datenpunkt vor?

Synchronisation Existieren zwei Datenpunkte gemeinsam?

3 Serielle Daten

- alle Daten können zeitabhängig gemacht werden
- Darstellung parallel oder sequenziell (Animation)

GEOSPATIAL DATEN

1 Phenomena

Punkt

- 0-dimensional
- definiert durch Längen- und Breitengrad

Linie

- 1-dimensional
- Länge, keine Breite
- Paar von ungeschlossenen Längen- und Breitengrad

Fläche

- 2-dimensional
- Länge, Breite, Längen- und Breitengrad begrenzen die Fläche
- Menge für jedes Phänomenon

Oberfläche

- 2,5-dimensional
- Länge, Breite, Höhe
- Höhenkoordinatenvektor
- Menge von Attributen für jedes Paar von Längen- und Breitengraden

2 Kartenprojektionen

winkelgetreu Form bleibt erhalten

gleichgroße Fläche Flächenproportionen bleiben erhalten

identische Entfernungen Distanzen bleiben erhalten

zentral kürzester Weg zwischen zwei Punkten bleibt erhalten

scheitelwinkelig Richtung vom Zentralpunkt bleibt erhalten

3 Mapping Probleme

- relative und absolutes Mapping ergibt unterschiedliches Ergebnis
- Klassenaufteilung
- räumliche Aggregationen

4 Punktdaten

DotMaps

- viel Overlap
- Kreis repräsentiert die räumliche Position eines Events
- keine Skalierung

PixelMaps

- Overlap-frei
- Punkt so nah wie möglich an Originalposition
- skaliert Karte neu

visuelle Explorationsziele

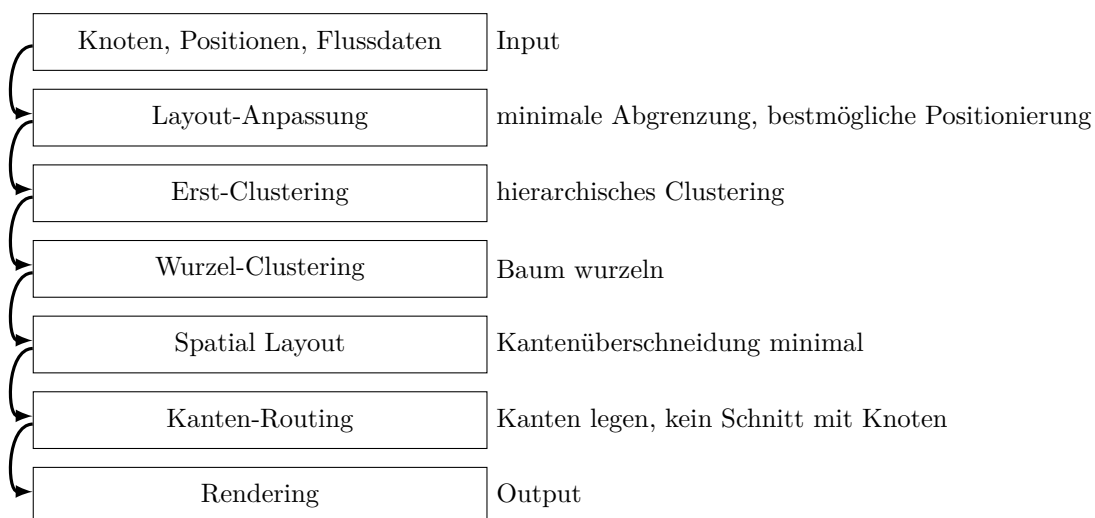
- kein Overlap für „normale“ Bildschirmauflösung
 - ergibt effektive Visualisierung
 - kein Informationsverlust
- Clustering
 - generelle geo-spatial Relationen sichtbar
 - geo-Pattern werden sichtbar
- Positionserhaltung
 - Pixel-Zusammenhänge erkennbar
 - kleine Cluster sichtbar

effiziente Implementierung Neuskalierung der Kartenregionen, damit die 3D-Punkt-Wolken besser auf den Bildschirm passen

5 Liniendaten

- müssen nicht linear sein (Bezier-Kurven)
- Symbole auf Knoten oder Schnittpunkten für Dateneigenschaften

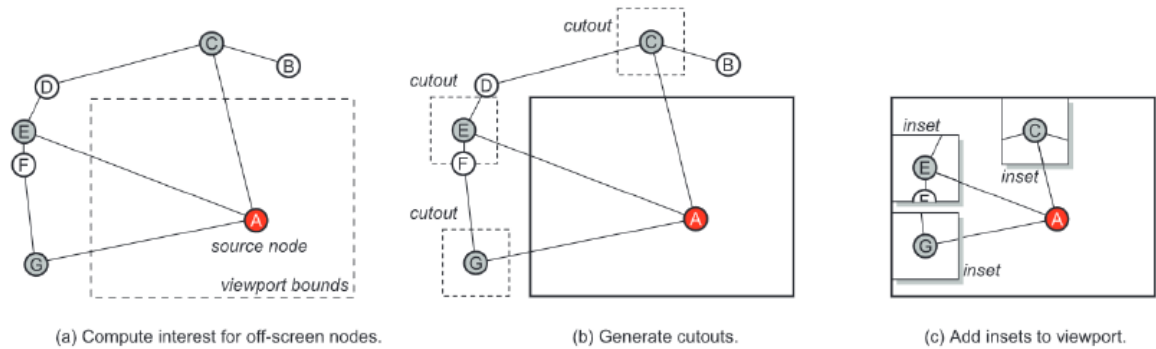
5.1 Flow Maps



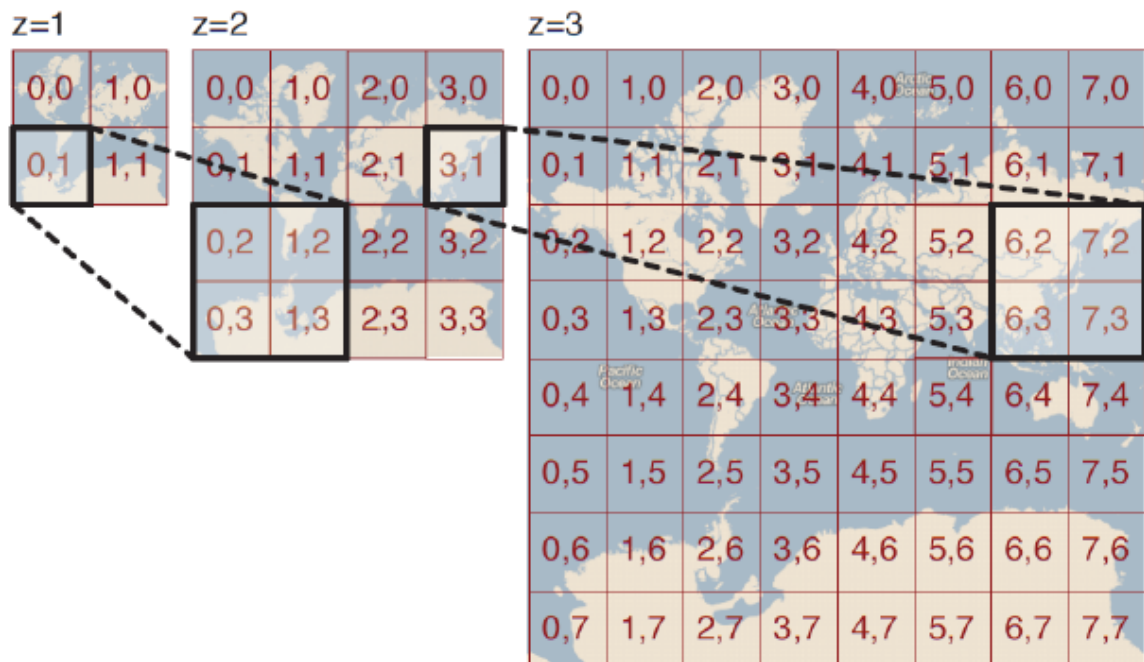
- können verbessert werden durch Kantenbündelung

6 Karteninteraktionen

- Zoom and pan

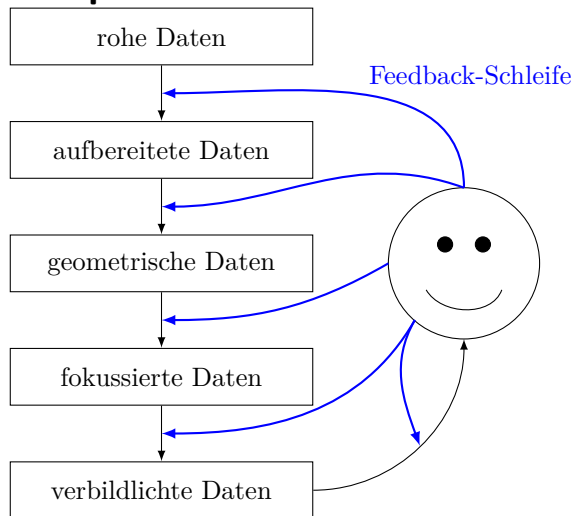


- Animation
- Scaling (Quadtree)



VISUALISIERUNGSKONZEPTE

1 Pipeline



2 Operatoren

Navigationsooperatoren

Selektionsoperatoren Teilmenge der Daten wählen

Filteroperatoren Slider, Filter-Flow-Modell (Visualisierung baut sich beim Filtern auf)

Neukonfigurationsoperatoren Filtern, dann Neuordnen der Dimensionen, Sortieren von Zeilen und Spalten in Tabellen

Kodierungsoperatoren unterschiedliche visuelle Variablen, Parametermodifikation (z.B. Farbe ändern)

Verbindungsoperatoren Interaktion mit einer Visualisierung hebt Datenpunkte auch in allen anderen Visualisierungen hervor

ABstraktions-/Elborationsoperatoren Fokus auf Teilmenge für Details, Reduktion des Detaillevels anderer Datenteile, Verzerrung in interaktiver Erkundung der Daten

3 Räume

Bildschirmraum (Pixel)

- Pixel-Regionen werden vergrößert/verkleinert für detailreichere Darstellung
- (semantisches) Zoomen

Datenwertraum (Multivariate Daten) Transformationen werden auf Dimensionalität angewendet

Datenstrukturraum (Teile der Datenorganisation)

- Auswahl von Knoten in Hierarchie
- Hervorheben in allen relevanten Knoten durch Benutzerinteraktion (nicht nur ein Knoten auf einmal)
- resampeln der Datenpunkte bei zoomen

Attributraum (Teile/Eigenschaften von grafischen Einheiten)

- modifizieren von einem oder mehreren Attributen in grafischen Objekten
- Kontrast anpassen

Objektraum (3D-Oberflächen)

- Verzerrung eines Objektes durch Auswahl von Daten
- Bsp.: Parallel-Coordinate wird auf Wände gemappt, eine Wand als Hervorhebung ausgewählt, alle anderen verzerrt

Visualisierungsraum

- modifizieren des unterliegenden Strukturelementes der Visualisierung
- Bsp.: Anordnung der Parallel-Coordinate Dimensionen

VISUALISIERUNGSSYSTEME

Graph Views zirkulär, radial, force-directed, hierarchisch, Adjazenz

Tree Layout Views horizontal/vertikal, radial, Ballons, TreeMap, Eiszapfenplots

Tabular Data Views bar/line/pie charts, gestapelt, box plots, scatter plots, parallel coordinates

Geospatial, spatiotemporal Views Karten-Overlays, Symbolkarten, Chloropleth-Karten

verschiedene Systeme für

- Visualisierung und Interaktivität
- automatische Analyse
- Datenkomprimierung, Speicheroptimierung
- Präsentationsorientierte Eigenschaften
- Netzwerkvisualisierung

zukünftige Richtungen

- semi- und unstrukturierte Daten
- verbesserte Visualisierung
- veränderliche (anpassbare) Visualisierungen
- linguistische Analyse auf Textdokumenten