

Semantik-gestütztes Hilfesystem für ein komposites Informationsvisualisierungssystem

Diplomarbeit
Technische Universität Dresden
November 2013

Nikolaus Piccolotto

Betreuer: Dipl.-Medieninf. Martin Voigt
Hochschullehrer: Prof. Dr.-Ing. Klaus Meißner

Fakultät Informatik
Institut für Software- und Multimediatechnik
Seniorprofessur für Multimediatechnik



Aufgabenstellung

Diese Seite muss vor dem Binden der gedruckten Fassung der Arbeit durch die von Herrn Meißner und dem Studenten eigenhändig unterschriebene originale Aufgabenstellung ersetzt werden. Das zweite abzugebende gebundene Exemplar soll stattdessen eine Kopie dieser originalen Aufgabenstellung enthalten.

Erklärung

Hiermit erkläre ich, Nikolaus Piccolotto, die vorliegende Diplomarbeit zum Thema

Semantik-gestütztes Hilfesystem für ein komposites Informationsvisualisierungssystem

selbstständig und ausschließlich unter Verwendung sowie korrekter Zitierung der im Literaturverzeichnis angegebenen Quellen und Hilfsmittel verfasst zu haben.

Dresden, 30. November 2013

Unterschrift

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Problemstellung und Zielsetzung	1
1.3	Aufbau der Arbeit	1
2	Stand der Forschung und Technik	2
2.1	Szenario	2
2.2	Anforderungsanalyse	4
2.2.1	Funktionale Anforderungen	4
2.2.2	Nichtfunktionale Anforderungen	5
2.3	Grundlagen	5
2.3.1	Semantische Datensätze	5
2.3.2	Informationsvisualisierung	6
2.3.3	Software Support	8
3	Konzeption	9
	Literaturverzeichnis	10

Abbildungsverzeichnis

1.1	VizBoard Workflow	1
2.1	Skizze der VizBoard Visualisierungskomponenten	3
2.2	Semantic Spectrum	6
2.3	Parallel Coordinates	7
2.4	Recursive Pattern Visualisierungstechnik	8

Tabellenverzeichnis

1 Einleitung

1.1 Motivation

1.2 Problemstellung und Zielsetzung

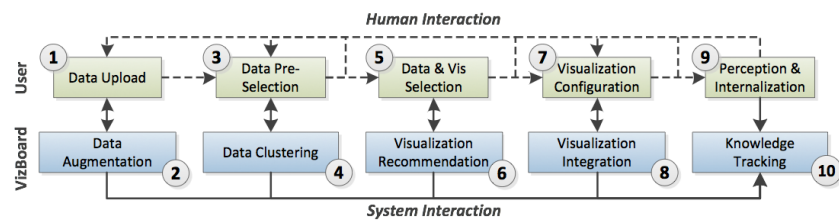


Abbildung 1.1: VizBoard Workflow

1.3 Aufbau der Arbeit

2 Stand der Forschung und Technik

Der folgende Abschnitt besteht aus drei Teilen. Zuerst wird die Aufgabenstellung in einem Szenario verdeutlicht (Abschnitt 2.1). Daraus werden Anforderungen an das Hilfesystem abgeleitet (Abschnitt 2.2) und danach die Grundlagen von semantischen Daten, Informationsvisualisierungen und Software Support erläutert (Abschnitt 2.3).

2.1 Szenario

Wie in Kapitel 1 erläutert, ist das komposite InfoVis-System Teil der webbasierten Anwendung VizBoard. Sie leitet den Benutzer in mehreren Schritten von der Auswahl eines Datensatzes zur finalen, kompositen Informationsvisualisierung. Im vorletzten Schritt wählt dieser mit Hilfe eines Facettenbrowsers geeignete Visualisierungskomponenten aus, welche danach angezeigt werden. Um die Problemstellung noch einmal zu verdeutlichen, wird im folgenden ein mögliches Szenario beschrieben.

Anna möchte für ihr Biologiestudium mehr über die geografische Verteilung verschiedener Genvariationen herausfinden. Dazu sucht sie im Internet nach einem Datensatz, welchen sie auch findet. Leider ist er in einem für Anna unbekannten Format abgespeichert, nämlich OWL. Sie versucht die Datei mit Microsoft Excel und SPSS zu öffnen, weil sie keine anderen Programme zur Datenverarbeitung kennt, aber scheitert. Anna stellt fest, dass nur ihr Texteditor OWL öffnen und vernünftig darstellen kann. Als sie die Datei überfliegt, kann sie den Inhalt zwar erahnen, aber es ist einfach zu viel Text um ihn vollständig zu lesen. Davon abgesehen sind geografische Breite und Länge als Zahlenkombination keine anschauliche Repräsentation von Orten, auch Verteilungen von Werten sind so schwer ersichtlich. Anna würde viel Zeit aufwenden müssen um sehr wenig des Inhalts zu verstehen. Aber selbst wenn sie die Datei in Excel hätte öffnen können, hätte sie nicht gewusst, mit welchen Diagrammen die vorhandenen Daten am Besten verstanden würden. Anna hört von einem Freund, dass VizBoard gut geeignet ist, um semantische Datensätze anzusehen und probiert es aus.

Anna hat ihren Datensatz auch bei VizBoard gefunden und ist neugierig: Sie wählt eine Karte, ein Balkendiagramm, eine Tabelle und eine Treemap aus (Abbildung 2.1); kurz darauf werden ihr die Visualisierungskomponenten angezeigt. Anna benutzt VizBoard zum ersten Mal und macht außer Facebook und YouTube auch sonst nicht viel im Internet, das heißt sie ist zunächst von den vier unterschiedlichen Fenstern etwas überfordert.

VizBoard bietet Anna aber sofort eine einführende Übersicht und erklärt kurz die Darstellungsform und den Inhalt jeder Komponente. Eine denkbare Erklärung der Treemap wäre zum Beispiel:

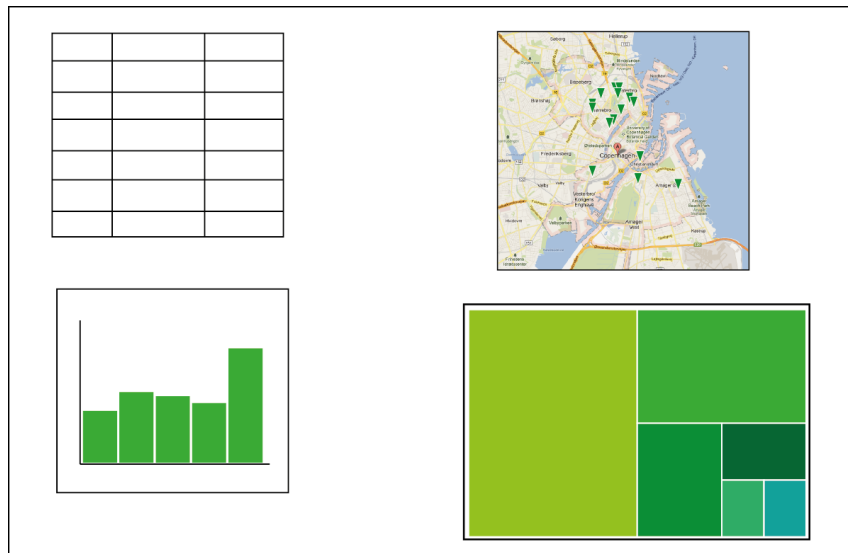


Abbildung 2.1: Skizze der VizBoard Visualisierungskomponenten

Eine Treemap ist eine hierarchische Visualisierung, um Größenverhältnisse anschaulich zu machen. In dieser werden die Anzahl von Genvariationen pro geografischer Region dargestellt.

Damit bekommt Anna einen Überblick über die verfügbaren Visualisierungen und weiß, welches Fenster welche Visualisierung enthält und für was diese gut sind. Nun möchte sie die Tabelle, in der die durch die Treemap visualisierten Daten stehen, nach der Spalte »Anzahl« sortieren. Anna sieht aber nicht, wie sie das machen soll, da in der Tabelle kein offensichtliches Kontrollelement wie z. B. ein Button vorhanden ist. Sie bemerkt ein Fragezeichen in der Titelleiste des Fensters und klickt darauf. Der verfügbare Viewport wird abgedunkelt und es erscheint ein neues Fenster, welches die verfügbaren Aktionen mit Hilfe von Text, Bildern und Animationen erklärt. Anna lernt, dass sie mit einem einfachen Linksklick auf den jeweiligen Kopf einer Tabellenspalte nach dieser sortieren kann und außerdem eine oder mehrere Zeilen auswählen kann. Sie sortiert die Tabelle wie gewollt und wählt die ersten drei Zeilen aus. Plötzlich verkleinert die Karte ihr Zoomlevel und Anna ist verwirrt: Sie hat nur mit der Tabelle interagiert und es bestand keine sichtbare Verbindung zwischen den beiden Fenstern. Allerdings wurde nach der Zeilenauswahl ein Pfeil von der Tabelle zur Karte gezeichnet, welcher mit einem Icon in Form eines Briefes versehen ist. Anna vermutet, dass doch irgendeine Verbindung zwischen den beiden Visualisierungen besteht und klickt auf den Brief. Ähnlich wie vorhin bei der Hilfe zur Tabelle wird der Viewport abgedunkelt und ein neues Fenster wird eingeblendet. Es erklärt die Kommunikation zwischen den Visualisierungen mit Hilfe von Animationen, Text und Bildern. Nun weiß Anna auch, wie die verschiedenen Fenster zusammenhängen und kann sich ihrer eigentlichen Aufgabe widmen.

In der Tabelle findet sie auch eine Spalte »SNP«. Anna weiß zwar, dass sie die Abkürzung schon einmal gesehen hat, kennt aber im Moment ihre Bedeutung nicht. Praktischerweise ist der Spaltenkopf mit der Wikipedia verlinkt und sie wird sofort auf die entsprechende Seite weitergeleitet. Anna erinnert sich, dass »SNP« »Single-nucleotide polymorphism« bedeutet und sie bekommt auch gleich zusätzliche Infor-

mationen zu diesem Thema. Sie widmet sich weiter der Tabelle und stellt fest, dass die Ortsbezeichnung »Kopenhagen« nicht mit der Markierung in der Karte übereinstimmt. Außerdem ist sie erstaunt, wie hoch die Verbreitung eines bestimmten SNPs dort ist und würde gerne die Ursache dafür wissen. In der Hilfe zur Tabelle wurde sie auch über die Möglichkeit, Kommentare an den Daten vorzunehmen, aufgeklärt. Anna kommentiert sowohl die falschen Geokoordinaten als auch ihre Frage über die Verbreitung des SNPs, sodass sie später per Email über Antworten informiert wird. Nun kann Anna sich mit der vierten Visualisierung, dem Balkendiagramm, beschäftigen. Allerdings reagiert es auf keine Mausklicks und macht auch sonst nicht den Eindruck, die Daten akkurat darzustellen. Anna meldet die kaputte Visualisierung über die eingebaute Reporting-Funktion und schließt das Fenster, um sich den anderen drei Visualisierungskomponenten zuzuwenden.

2.2 Anforderungsanalyse

Aus dem Szenario (Kapitel 2.1) lassen sich nun verschiedene Anforderungen an ein Hilfesystem für komposite Informationsvisualisierungssysteme ableiten.

2.2.1 Funktionale Anforderungen

Blabla

- **Überblick:** Das Hilfesystem soll einen kurzen Überblick über das InfoVis-System geben und Darstellungsform sowie Inhalt jeder Komponente kurz erläutern.
- **Bedienung:** Das Hilfesystem soll erklären können, wie eine Komponente bedient wird. Diese Informationen umfassen beispielsweise welche Operationen welche Aktionen (eventuell auf welchen Daten) ausführen.
- **Reporting:** Fehler in Komponenten sollen über ein Reporting-System gemeldet werden können.
- **Verlinkung:** Das Hilfesystem soll nicht-triviale Begriffe mit der Wikipedia verlinken, sodass nicht nur auf die Begriffsbedeutung hingewiesen werden, sondern dem Benutzer auch zusätzliche Informationen zur Verfügung gestellt werden können.
- **Kommunikation:** Das Hilfesystem soll erklären können, wie gegebene Komponenten miteinander kommunizieren.
- **Kommentare:** Der Benutzer soll die Möglichkeit haben Daten zu kommentieren und Bereiche der Visualisierung zu markieren und mit ebenfalls mit einem Kommentar zu versehen, sodass auch auf fehlende Daten hingewiesen werden kann.

2.2.2 Nichtfunktionale Anforderungen

Bla bla

- **Korrektheit:** Eine gegebene Hilfestellung darf keine Fehlinformationen enthalten, weil sie sonst mehr verwirrt als hilft.
- **Vollständigkeit:** Eine gegebene Hilfestellung muss alle Informationen enthalten, die der Nutzer benötigt um danach seine gewünschte Aufgabe ausführen zu können.
- **Verständlichkeit:** Hilfestellungen müssen in einer Form präsentiert werden, die der Benutzer schnell und mit geringem mentalen Aufwand verarbeiten kann.
- **Einheitlichkeit:** Das Look & Feel von Teilen des Hilfesystems (z.B. Kommentare) muss komponentenübergreifend einheitlich sein, damit der Benutzer einmal gelerntes wiederverwenden kann.
- **Minimalität:** Der Komponentenentwickler soll seine Komponente mit möglichst wenig Aufwand zum Hilfesystem kompatibel machen können, ansonsten werden nur sehr wenige Komponenten – und damit der Benutzer – davon profitieren.
- **Universalität:** Das Hilfesystem soll für alle Komponenten und Visualisierungen in gleicher Qualität funktionieren.
- **Wiederverwendbarkeit:** Die Kommentare sollen möglichst in allen Visualisierungen wiederverwendet werden, damit viele Benutzer von den Erkenntnissen anderer profitieren können.
- **Unaufdringlichkeit:** Das Hilfesystem soll den Benutzer nicht von seinen Aufgaben ablenken und nur auf Anfrage zum Einsatz kommen oder es soll selbstständig erkennen, wenn der Benutzer Hilfe benötigt.

2.3 Grundlagen

2.3.1 Semantische Datensätze

Eine Ontologie ist »an explicit specification of a conceptualization« [Gru+95] und wird benutzt um domänenspezifisches Wissen abzubilden [CJB99]. Sie besteht aus mehreren Elementen:

- Eine Klasse repräsentiert ein Konzept, eine Entität, ein Ding, beispielsweise ein *Smartphone*.
- Eine Instanz ist ein konkretes Objekt einer Klasse, zum Beispiel das *iPhone mit der Seriennummer XYZ-ABC*.
- Datenattribute beschreiben eine Instanz näher, zum Beispiel die *Seriennummer* oder *Bildschirmgröße* des iPhones.

- Objektattribute beschreiben Beziehungen zwischen Klassen und deren Instanzen, beispielsweise eine Person *besitzt* ein Smartphone.
- Außerdem existieren noch Axiome, Regeln, Funktionen und Einschränkungen, welche die Logik einer Ontologie beschreiben.

Um eine Ontologie maschinenlesbar darzustellen, hat das World Wide Web Consortium (W3C) verschiedene Beschreibungssprachen eingeführt. Die bekanntesten sind Resource Description Framework Schema (RDFS) und Web Ontology Language (OWL). Mit diesen Sprachen lässt sich unterschiedlich viel Semantik u. a. in Form von Ontologien, Thesauren oder Vokabularen ausdrücken; die Komplexität der Dokumente und damit der Aufwand, sie zu erstellen, verhalten sich aber direkt proportional (siehe Abbildung 2.2) [Ber07].

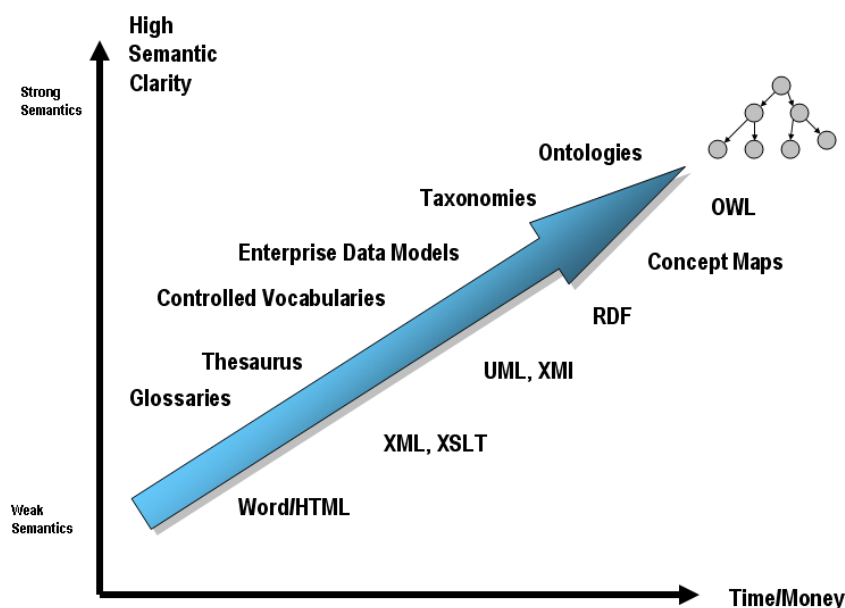


Abbildung 2.2: Semantic Spectrum

Noch auf OWL oder RDF/RDFS eingehen?

2.3.2 Informationsvisualisierung

Card et al. [CMS99] definieren »Informationsvisualisierung« wie folgt:

The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.

Beispiele dafür sind Balkendiagramme, Treemaps [Shn92] und Parallel Coordinates (Abbildung 2.3) [ID91].

Keim [Kei02] klassifiziert Informationsvisualisierungen nach dargestellten Daten, Visualisierungs- und Interaktionstechnik:

- Dargestellte Daten
 - Eindimensional: Zum Beispiel Zahlenreihen (1992, 1994, 2001, 2002...)

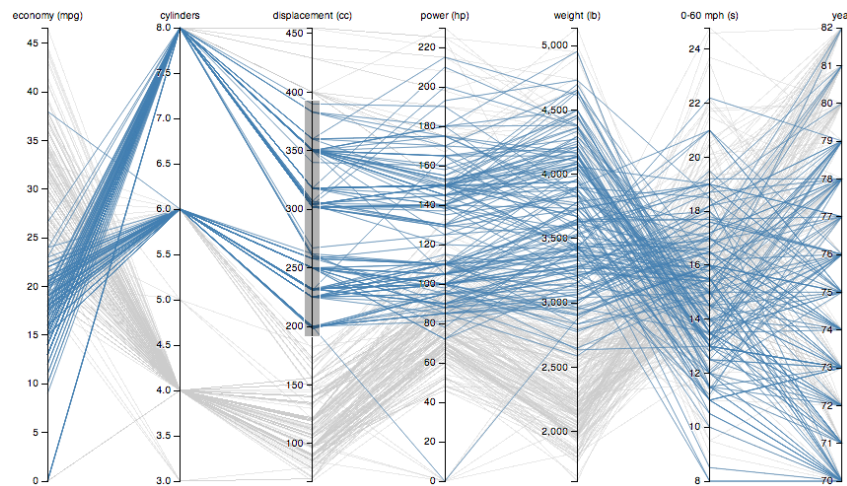


Abbildung 2.3: Parallel Coordinates

- Zweidimensional: Beispielsweise Key/Value Paare (Apple: 445, Google: 873, Microsoft: 34, ...)
- Multidimensional: Etwa verschachtelte Key/Value Paare (Apple: {Stock: 445, Founded: 1976}, ...)
- Text: Zum Beispiel Blogposts.
- Netzwerke/Graphen, wie zum Beispiel die Hyperlinks im World Wide Web.
- Software/Algorithmen, beispielsweise welche Prozesse welche Locks halten.
- Visualisierungstechniken
 - Standard 2D/3D: Balkendiagramme, Liniendiagramme, Karten etc.
 - Geometrically transformed: Zum Beispiel Parallel Coordinates.
 - Iconic: Hier werden Bilder zur Informationsvisualisierung benutzt, beispielsweise Smileys [Che73].
 - Dense Pixel: Wie zum Beispiel Recursive Patterns (Abbildung 2.4).
 - Stacked: Zum Beispiel Treemaps.
- Interaktionstechniken
 - Standard
 - Projection
 - Filtering
 - Zoom
 - Distortion
 - Link & Brush

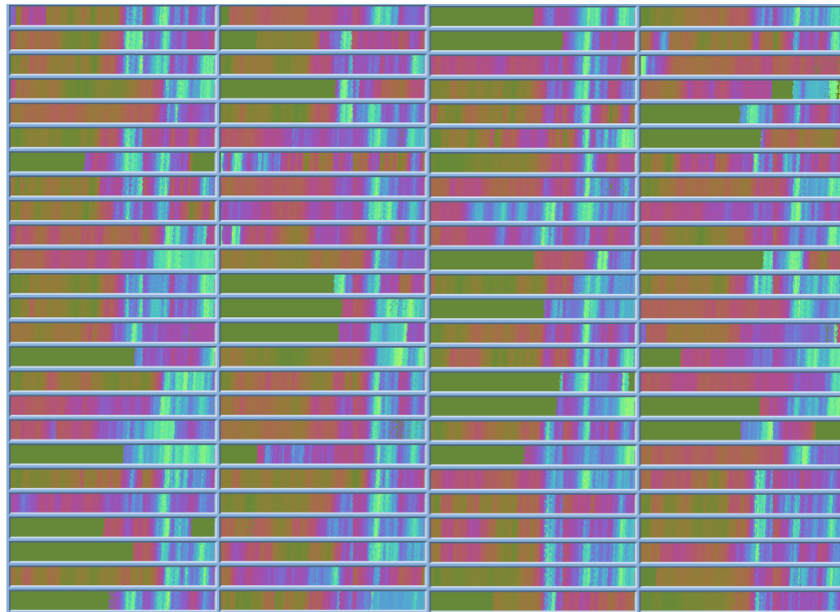


Abbildung 2.4: Recursive Pattern Visualisierungstechnik

Interaktionstechniken

2.3.3 Software Support

3 Konzeption

Literaturverzeichnis

- [Ber07] M. Bergman. *An intrepid guide to ontologies*. Mai 2007. URL: <http://www.mkbergman.com/374/an-intrepid-guide-to-ontologies/> (siehe Seite 6).
- [Che73] Herman Chernoff. »The use of faces to represent points in k-dimensional space graphically«. In: *Journal of the American Statistical Association* 68.342 (1973), Seiten 361–368.
- [CJB99] Balakrishnan Chandrasekaran, John R Josephson und V Richard Benjamins. »What are ontologies, and why do we need them?«. In: *Intelligent Systems and Their Applications, IEEE* 14.1 (1999), Seiten 20–26 (siehe Seite 5).
- [CMS99] Stuart K Card, Jock D Mackinlay und Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann Pub, 1999 (siehe Seite 6).
- [Gru+95] Thomas R Gruber u. a. »Toward principles for the design of ontologies used for knowledge sharing«. In: *International journal of human computer studies* 43.5 (1995), Seiten 907–928 (siehe Seite 5).
- [ID91] Alfred Inselberg und Bernard Dimsdale. »Parallel Coordinates«. In: *Human-Machine Interactive Systems*. Springer, 1991, Seiten 199–233 (siehe Seite 6).
- [Kei02] Daniel A. Keim. »Information visualization and visual data mining«. In: *Visualization and Computer Graphics, IEEE Transactions on* 8.1 (2002), Seiten 1–8.
- [Shn92] Ben Shneiderman. »Tree visualization with tree-maps: 2-d space-filling approach«. In: *ACM Transactions on graphics (TOG)* 11.1 (1992), Seiten 92–99 (siehe Seite 6).