Graphische Darstellung von Informationen

Martha A. Barberena N.
Technische Universität Ilmenau
Institut Praktische Informatik und Medieninformatik
Fachgebiet Telematik
Martha.Najarro@Prakinf.TU-Ilmenau.DE

Zusammenfassung

Im vorliegenden Beitrag werden Probleme bei der Arbeit mit Internet-Suchdiensten analysiert und Lösungsansätze vorgestellt. Das Hauptproblem auf der Benutzerseite ist die bisher komplizierte und unübersichtliche Handhabung der Ergebnismengen von Suchanfragen. Geeignete Mittel zur Bewältigung dieses Problems bietet die 3D-Computergrafik. Ausgehend von den Erfahrungen in der wissenschftlich-technischen Visualisierung wird nach allgemeingültigen Methoden zur Visualisierung von Informationen gesucht.

1 Einführung

Der exponentielle Anstieg von Informationsanbietern und Daten im Internet macht es zunehmend schwieriger, die wichtigen Informationen zu finden. Es wird noch schwieriger, ohne eine inhaltliche Erschließung und Systematisierung der vernetzten Informationen die unstrukturierte Ansammlung digitaler Dokumente in ein globales Informationssystem zu überführen.

Außerdem führte eine wachsende Anzahl von Suchdiensten, die ihren Service im Netz anbieten wollten, wie Bibliotheks-OPACs oder Campusnetze der Universitäten, zu einem wachsenden Bedarf an Systemen für die textbasierte Informationssuche. Es entstanden eine Reihe von Informationsdiensten (*Wide Area Information System: Wais, Gopher*) und Suchwerkzeugen (Archie, Veronica). Mit World Wide Web (WWW) kam die verstärkte private und kommerzielle Nutzung des Internet. Automatische Verfahren zur Indexierung, Klassifikation und Benutzerführung waren notwendig geworden. Neben den traditionellen oben genannten Internetsuchwerkzeugen werden derzeit zwei Recherchedienste im Web eingesetzt: die Katalog- und verzeichnisbasierte Suche (Browsing-Dienste) und die Roboterbasierte Suche bzw. Suchmaschine [1].

2 Katalogbasierte Suche und Suchmaschinen

2.1 Katalogbasierte Suche (Browsing-Dienste)

Browsing-Dienste (z.B. Yahoo, Infoseek Guide) sind hierarchisch strukturierte Verzeichnisse bzw. Sammlungen von Internetressourcen, in denen man gezielt blättern (browsen) kann. Sie sind in der Regel eine zusammengestellte und intellektuell ausgewählte Sammlung. Browsing-Dienste bieten die Möglichkeit, thematisch zusammengehörige Internetressourcen zu finden. Ein besonderer Vorteil von Browsing-Diensten liegt in der Auswahl der Ressourcen. Es werden nur solche Links aufgenommen, die für relevant gehalten werden. Aber Browsing-Dienste decken nur einen geringen Teil der Internetressourcen ab. Außerdem besteht die Gefahr, daß Sammlungen nicht mehr weitergepflegt werden und veralten.

2.2 Suchmaschinen

Bei Suchmaschinen (z.B. Alta Vista, Lycos) werden Suchbegriffe zum Auffinden der Internetressourcen benutzt. Sie bieten verbale Suchmöglichkeiten an. Über Suchmaschinen greift man auf große Datenbaken zu, deren Inhalt automatisch mit Hilfe von Analyseprogrammen (*Robots* oder *Spidern*) erstellt wird. Sie erfassen einen größeren Anteil der Internetressourcen als die Browsingdienste, da sie auf automatischen Verfahren basieren. Suchmaschinen bieten auch die Möglichkeit, nach anderen Internetangeboten zu suchen, wie FTP, Email, Newsgroups usw. Sie bieten ein benutzerfreundliches Suchformular, das die verbale Suche erleichtert. Ein Problem bei der Arbeit mit Suchmaschinen besteht darin, daß man oftmals viel zu viele Treffer bekommt, die sich zum Teil wiederholen und von denen ein großer Teil irrelevant ist.

2.3 Probleme von Suchdiensten

Die robotbasierte Suche nach Internetressourcen und die automatische Indexierung der gefundenen Webseiten sind nicht unproblematisch. Nicht alle Bereiche des Web werden erfaßt. Einige Dokumente wie geschützte Dokumente oder dynamische Dokumente können gar nicht erfaßt werden. Problematisch ist auch die automatische Indexierung nach statistischen Verfahren, wenn nur ein Teil der Webseite wie Titel oder die ersten Zeilen des Textes ausgewertet werden können. Lycos wertet z.B. nur die ersten 100 Wörter einer Webseite aus. Ein anderes Problem ist das Update der Indexdatenbank. Viele Links in der Trefferanzeige von Suchmaschinen sind nicht mehr aktuell.

Auch im Dialog direkt mit einer Suchmaschine ist der Benutzer mit Problemen konfrontiert. Er bekommt die Ergebnismenge der Anfrage in einer sequentiellen Liste von Dokumenttiteln. Bei großen Datenbanken tritt das Problem auf, daß nur ein Teil der Dokumente der Ergebnismenge dargestellt wird. Um die ganze Menge von Dokumenten darstellen zu können, muß man die Ergebnismenge auf mehrere Seiten verteilen. Der Benutzer verliert in der Form den Überblick über die Ergebnismenge. Viele Retrievalsysteme treffen deshalb eine Auswahl von besonders relevanten Dokumenten, um die Größe der Ergebnismenge zu begrenzen, ohne daß die individuellen Suchwünsche der Benutzer berücksichtigt werden. Dem Benutzer werden bei diesem Verfahren Informationen vorenthalten, die das System als ungenügend relevant beurteilt hat. Außerdem vermittelt diese sequentielle Dokumentliste dem Benutzer keine zusätzliche Information über die ausgewählten Dokumente. Bestenfalls kann die Reihenfolge einen schwachen Anhaltspunkt über das Relevanzmaß bestimmter Dokumente liefern. Es gibt keine Information über semantische Beziehungen zwischen einzelnen Dokumenten. Der Benutzer bekommt auch keine weitere Hilfe zur Verbesserung der Anfrage durch die Listendarstellung.

Ziele der Repräsentation der Ergebnismenge einer Suchanfrage sollten sein:

- mögliche Darstellung der vollständigen Ergebnismenge in grafischer Form
- Darstellungsformen, die es ermöglichen, sofort die Relevanz eines Dokumentes visuell zu erfassen
- Navigation durch die Dokumentenmenge und Selektion „wichtiger“ Dokumente
- Möglichkeit des Ausblendens „unwichtiger“ Dokumente
- Bereitstellung spezieller Informationen zu einem Dokument

Ähnliche Anforderungen an die Repräsentation von Daten bestehen auch im technischen Bereich, daher soll versucht werden, die dort vorhandenen Erfahrungen zu nutzen.

3 Wissenschaftlich-technische Visualisierung

3.1 Grundlagen

Die wissenschaftlich-technische Visualisierung (*Scientific Visualization*) ist ein relativ neuer Bereich der Informatik. Sie hat das Ziel, komplizierte Vorgänge in Wissenschaft und Technik zu visualisieren. Die Grundlagen der wissenschaftlich-technischen Visualisierung sind die graphische Datenverarbeitung speziell mit den Teilbereichen Computer-Graphik und Bildverarbeitung sowie das Bilderverstehen (*Computer-Vision*) [3].

Die klassische (*generative*) Computer-Graphik beschäftigt sich mit der Generierung von Bildern aus formalen Beschreibungen wie etwa der Modellierung dreidimensionaler Objekte aus vorliegenden Konstruktionsdaten. Dieser Bereich hat auch die größte Bedeutung für die Visualisierung, da es die Algorithmen erlauben, Daten visuell aufzubereiten und zu interpretieren.

Im wissenschaftlich-technischen Bereich profitieren Simulationen von der Möglichkeit, interaktiv Einflußgrößen zu ändern und zu visualisieren. Strömungsdynamische Untersuchungen, klimatologische Applikationen, Molekül-Design oder die Entwicklung neuartiger Materialien sind Beispiele für die wissenschaftliche Visualisierung rechnergestützter Simulation. Die Visualisierung dient neben der visuellen Verarbeitung von Simulationsergebnisssen auch als Analyseinstrument bei der Auswertung von Meßdaten. Beispiele dafür sind in den Bereichen Medizin (Computer-Tomografie), Satellitenbildanalyse (Aufbereitung der Daten von Planeten), Geländedatenanalyse (Auswertung von Bodenpotentialen) aber auch in der Wirtschaftswissenschaft *(*Portfolioanalyse) zu finden.

3.2 Der Visualisierungsprozeß

3.2.1 Die Visualisierungs-Pipeline

Ausgehend von zugrundeliegenden Meß- oder Simulationsdaten müssen die interessierenden wissenschaftlich-technischen Phänome in geeigneter Weise sichtbar gemacht werden. Dies erfordert problemspezifische Visualisierungstechniken.

Simulationen produzieren z.B. im Bereich des *Scientific Computing* große Mengen von Ergebnisdaten, da häufig mit Punktmengen in hochdimensionalen Räumen gearbeitet wird. Deswegen müssen die Daten zuerst aufbereitet und entsprechenden graphischen Grundelementen zugewiesen werden, bevor sie unter Berücksichtigung der gewählten Visualisierungsmethode graphisch dargestellt werden. Das Modell in der Abbildung 3.1 konkretisiert die Aufgabenverteilung bei einer visuellen Analyse von Simulationsdaten [3] .

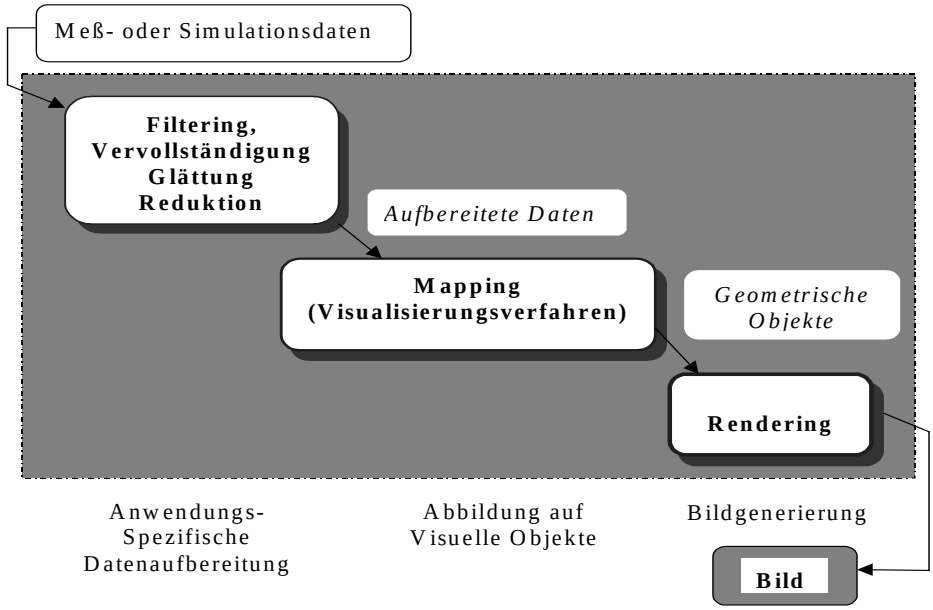


Abbildung 3.1: Schematische Darstellung der Visualisierungs-Pipeline bei der graphischen Auswertung von Simulationsergebnissen

Der Visualisierungsvorgang besteht aus drei Transformationstufen. In der ersten Stufe findet eine Aufbereitung der Rohdaten statt. Das bedeutet, die Ergebnisdaten werden beispielsweise durch Selektion oder Projektion in einen niedrigerdimensionalen Raum transformiert. Die zweite Phase und entscheidende für den Visualisierungsprozeß ist die problemspezifische Zuordnung der Darstellungselemente und folglich die Visualisierungstechnik. Die letzte Phase ist die eigentliche Darstellung. In dieser Phase findet die graphische Umsetzung der Objekte in visuelle Informationen statt.

Filtering-Prozeß

Eine Datenquelle bildet den Anfang der Visualisierungs-Pipeline (siehe Abb. 3.1). Sie stellt einen Rechenprozeß zur Auswertung eines physikalischen Modells dar.
Im Vordergrund steht generell die Repräsentation von Modellen (Phänomenen). So müssen die Rohdaten in geeigneter Weise aufbereitet werden. Dieser Prozeß wird Filtering genannt. Die Datenaufbereitung beschränkt sich auf die Konstruktion eines empirischen Modells, das die geometrische Struktur des interessierenden Phänomens charakterisiert. So können diskrete Punktmengen in den zwei- oder dreidimensionalen Raum abgebildet werden.

Mapping-Prozeß

Der Mapping-Prozeß ist für die Komposition des Bildes, das erzeugt wird, verantwortlich. Hier wird die eigentliche Visualisierungstechnik festgelegt. Der Mapping-Prozeß bildet die gefilterten Daten auf abstrakte visuelle Objekte und ihre Attribute (Farbe, Erscheinung,...) ab. Das sind allgemein geometrische Primitive wie Linie, Kreis, Tetraeder und Kugel oder andere Darstellungselemente wie Volumenelemente (Voxel).

Rendering-Prozeß

Der Rendering-Prozeß verkapselt die Transformationsstufe der Bildgenerierung. Er rechnet das abstrakte Visualisierungsobjekt nach vorheriger Zusammensetzung der graphischen Primitive in eine zweidimensionale bildliche Darstellung unter Berücksichtigung der Licht- und Schattierungsverhältnisse, der optischen Reduzierung des Treppeneffektes sowie der Sichtbarkeitsklärung um.

3.2.2 Der Visualisierungszyklus

Die vorgestellte modulare Aufbauweise der Visualisierungs-Pipeline suggeriert eine sequentielle Arbeitsweise des Visualisierungsprozesses. Die Parameter der Filtering-, Mapping- und Rendering-Transformationsstufe können beispielsweise interaktiv geändert oder gegen andere, besser geeignete, ausgetauscht werden. Wenn man die Abbildung 3.1 in der Form erweitert, daß die Eingriffe in den Visualisierungsprozeß interaktiv sind, bekommt man einen zirkulären Visualisierungsablauf, der in Abbildung 3.2 gezeigt wird.

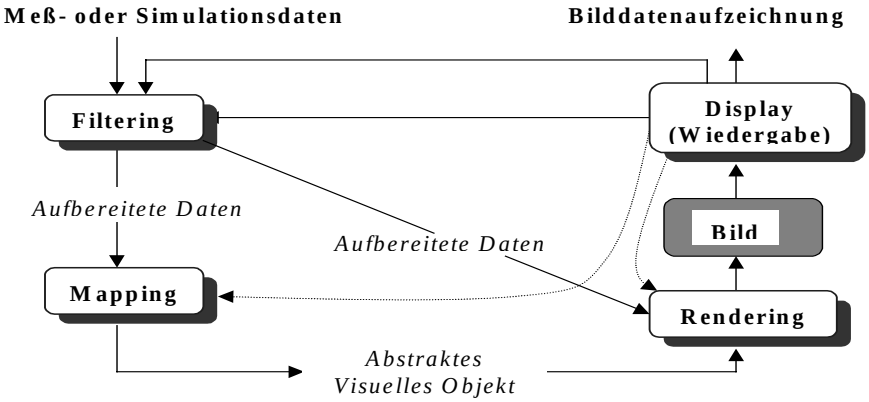


Abbildung 3.2 Visualisierungszyklus

Die Simulationsdaten werden von einem Visualisierungsprogramm gelesen, gefiltert und visuell dargestellt. Die gestrichelten Linien in der Abbildung 3.2 deuten Eingriffe in den Visualisierungszyklus an. Im Prinzip sind interaktive Eingriffsmöglichkeiten in allen Transformationsstufen möglich und erforderlich. Interaktive Eingriffe in den Mapping-Prozeß erfüllen die folgende Ziele:

- Auswahl bzw. Modifikation einer Visualsierungsmethode
- Anpassung an Dateigenerierung und -aufbereitung
- Spezifikation von Darstellungsattributen

Die Interaktiven Eingriffe in den Rendering-Prozeß erlauben:

- Auswahl der Rendering-Methode
- Spezifikation von Darstellungsattributen:
  - Farbe, Transparenz, Linientypen
  - Oberflächeneigenschaften
  - Lichtquellen

4 Informations-Visualisierung

Betrachtet man den Zyklus „Suchanfrage - Suche - Betrachten der Ergebnisse - Verbesserung der Suchanfrage - ...“, ergeben sich durchaus Ähnlichkeiten zum oben beschriebenen Visualisierungszyklus. Die darin enthaltene Visualisierungspipeline sollte grundsätzlich übertragbar sein, es sind jedoch völlig andere Filter- und Mapping-Stufen nötig.

Der Begriff wissenschaftlich-technische Visualisierung bezeichnet die visuelle Darstellung eines physikalischen Phänomens, die dem Benutzer zum besseren Verständnis und zur Erkenntnisgewinnung dienen soll. Die Informations-Visualisierung hat ähnliche Ziele, ist jedoch mit Aufgaben verbunden die bei der wissenschaftlich-technischen Visualisierung nicht auftreten.
Im technischen Bereich ist die graphische Darstellung meistens von einem physikalischen Sachverhalt abgeleitet. Bei der Visualisierung von Informationen haben die darzustellenden Elemente typischerweise keine physikalischen Eigenschaften, die die Grundlage für die Visualisierung bilden könnten. Die Elemente haben oft nur semantische Eigenschaften.

Um abstrakte Informationen, d.h. Dokumente, Begriffe oder andere textbasierte Informationen aus Datenbanken, Retrievalsystemen oder im WWW zu visualisieren, ist es nötig, zunächst einen entsprechenden Informationsraum zu definieren. Der Informationsraum hat eine spezifische Raumstruktur, in die die Objekte und Objektbeziehungen abgebildet werden (Mapping).

Das Mapping von Informationen in einen virtuellen Informationsraum kann wesentlich komplizierter sein als bei der wissenschaftlichen Visualisierung. Wie die Informationsrepräsentation erfolgen soll, hängt vom Datenmodell und von spezifischen Benutzeranforderungen ab. Ein anderes Problem besteht darin, daß Informationsräume multidimensional sind. Das bedeutet, sie besitzen abstrakte Objekte mit oft mehr als drei Objektdimensionen (siehe Abb. 3.3). Da maximal drei der Dimensionen auf die Raumachsen abgebildet werden können, müssen verstärkt grafische Attribute benutzt werden. Dazu gehören:

- Darstellungsform des Objektes (Dokumentes)
- Farbe
- Transparenz
- Parameter für Beleuchtung/Schattierung

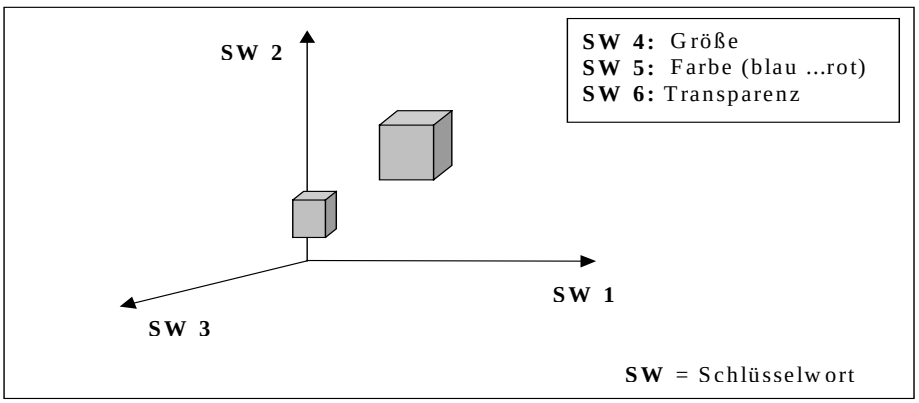


Abbildung 4.1 Dokumente im 3D-Koordinatensystem

Die Visualisierung stellt eine starke Abstraktion des Datenmodells dar. Das Gebiet der Informations-Visualisierung ist noch komplexer als das der wissenschaftlichen Visualisierung, aber die Informations-Visualisierung bietet die Möglichkeit, durch frei definierbare visuelle Metaphern die Informationsräume optimal den Erfordernissen des Benutzers anzupassen.

4.1 Bekannte Techniken der Informations-Visualisierung

In diesem Abschnitt werden Techniken, die derzeit zur Anzeige verschiedener Informationstechniken verfügbar sind, erläutert. Die Techniken können grob in drei Gruppen aufgeteilt werden:

Direkte Abbildung
Alle Eigenschaften oder der Wert der darzustellenden Datenelemente werden zur Abbildung auf Objekte im Raum benutzt. Diese Technik wird u.a. im System *LyberWorld* [4] eingesetzt.

Informationspräsentationstechnik
Diese Technik konzentriert sich auf die Präsentation, den Zugriff und die Nutzbarkeit der Daten. Das Ziel ist benutzerfreundliche und intuitive Schnittstelle. Beispiele dafür sind: *Cone Trees* und *Cam Trees* zur Visualisierung von hierarchischen Daten (z.B. Verzeichnisbäume)[5].

Dynamische Informationsvisualisierungstechnik
Die Visualisierungsverfahren werden mit Verhaltensmustern und dynamischen Eigenschaften ausgestattet, die der Applikation automatisches Reagieren auf Änderungen von Daten bzw. Interaktionen erlauben. Zu dieser Gruppe gehört z.B.: *Perpective Walls* zum Betrachten und Navigieren in großen, linear strukturierten Informationsmengen [5].
Diese Einteilung ist nur grob und es können in den Applikationen Überlappungen auftreten.

5 Ausblick

Mit der Hilfe der Informations-Visualisierung kann man die relevanten Dokumente wesentlich schneller als mit der traditionellen „Liste“ finden und die Menge an Informationen erhöhen, die ein Benutzer gleichzeitig handhaben kann. Die Dokumentenmengen können interaktiv manipuliert werden, da die grafischen Techniken eine direkte und intuitive Interaktion mit den interessierenden dargestellten Elementen erlauben können. Diese Mengen von Dokumenten oder Daten werden in einer wesentlich intuitiveren Form repräsentiert. Mittel dafür sind: Farben, Schattierung, Beleuchtung, Transparenz, Bewegungsmöglichkeiten und 3D-Perspektive.

Die Anordnung der Dokumente wird in der grafischen Darstellung nicht unproblematisch bleiben. Eine vollständige Visualisierung des Dialoges zur Informationssuche, zum Browsen in der Ergebnismenge und zum Navigieren innerhalb eines Informationsraumes ist jedoch anzustreben. Die existierenden Systeme zur Informations-Visualisierung enthalten spezielle, pragmatische Ansätze. In Zukunft werden jedoch allgemeingültige, übertragbare und integrierbare Konzepte benötigt.

Literatur

[1] R. Däßler, H. Palm: *Virtuelle Informationsräume mit VRML*: Informationen recherchieren und präsentieren in 3D. Heidelberg: dpunkt-Verl. ISBN 3-920993-78-0, 1998.

[2] S. Rudolph, M. Hemmje: *Visualisierung von Thesauri zur interaktiven Unterstützung von visuellen Anfragen an Textdatenbanken*. In: GMD-Studien Nr. 247, Juni 1994, Gesellschaft für Mathematik und Datenverarbeitung, Sankt Augustin.

[3] S. Ortmanns: *Modelle der Kooperation von Graphik-Workstations und Höchstleistungsrechnern*. Bericht des Forschungszentrums Jülich. ISBN 0366-0885, 1993.

[4] Hemmje, M.: *LyberWorld - eine 3D-basierte Benutzerschnittstelle für die computerunterstützte Informationssuche in Dokumentmengen.* Der GMD-Spiegel, 1993.

[5] XEROX Xsoft Central Europe: *Visual Recall: Document Management System für MS Windows*, XEROX Xsoft CE: Düsseldorf, 1994.