Darstellung von Suchergebnissen in 3D-Dimensionen

Martha Argentina Barberena Najarro
Technische Universität Ilmenau
Institut Praktische Informatik und Medieninformatik
Fachgebiet Telematik
Email: Martha.Najarro@PrakInf.tu-ilmenau.de

Zusammenfassung

Im vorliegenden Beitrag wird ein neuer Lösungsansatz für die Darstellung von Informationen vorgestellt. Es wird in einer allgemeinen Form von dem Hauptproblem ausgegangen, daß von der Benutzerseite die Handhabung des Ergebnisses der Suche kompliziert und unübersichtlich ist. Für Verbesserungen stehen in der Computergrafik geeignete Mittel zur Verfügung, die nutzbar zu machen sind.

1. Einführung

Durch die textbasierte Informationsliste, die die Suchmaschinen nach einer Suche im WWW (World Wide Web) liefern, ist es für den Benutzer schwer zu ermitteln, welche Information wirklich für ihn relevant ist. Der Benutzer ist auch nicht mehr in der Lage, die für ihn relevante Informationen durch Navigieren anhand vorgegebener Links aufzufinden und zu nutzen. Das trifft besonders dann zu, wenn mit mehreren Suchbegriffen gearbeitet wurde oder eine unscharfe Suche zu einem ganzen Themenkomplex stattfand.

Im Vordergrund steht die Forderung nach neuen Techniken zur Informationsaufbereitung und Recherche. Hier wird deswegen einen neuen Ansatz vorgestellt, der das Navigieren erleichtern soll und somit die Informationssuche effektiver gestaltet. Die grundsätzlichen Prinzipien eines solchen Ansatzes sind in [1] vorgestellt worden.

In [2] wurde dargelegt, aus welchen Gründen dafür das Grafiksystem Java 3D geeignet ist. Java 3D bietet die Möglichkeit, Stand-alone-Applikation und Applets für das WWW zu generieren. Das bedeutet, daß der gesamte Zyklus der Informationssuche Abbildung 1 in einer Applikation realisiert werden kann, die dann ohne Anpassungen auf allen Plattformen lauffähig ist.

2. Schnittstelle Suchmaschine - Visualisierungstool

Eine der Hauptaufgaben dieses Themas ist es, eine vollständige Visualisierung der gesuchten Informationen zum Browsen in der Ergebnismenge und Mechanismen zum Navigieren innerhalb eines Informationsraumes [3] zu entwickeln.
Die Visualisierung soll nicht an eine bestimmte Suchmaschine gebunden sein. Daher ist die Definition einer Schnittstelle notwendig, über die eine Suchmaschine die für die Visualisierung erforderlichen Daten an das Visualisierungsmodul übergeben kann.
Aufgrund der allgemeinen Verfügbarkeit verwendet diese Schnittstelle die Form einer HTML-Seite zur Übermittlung der Informationen.

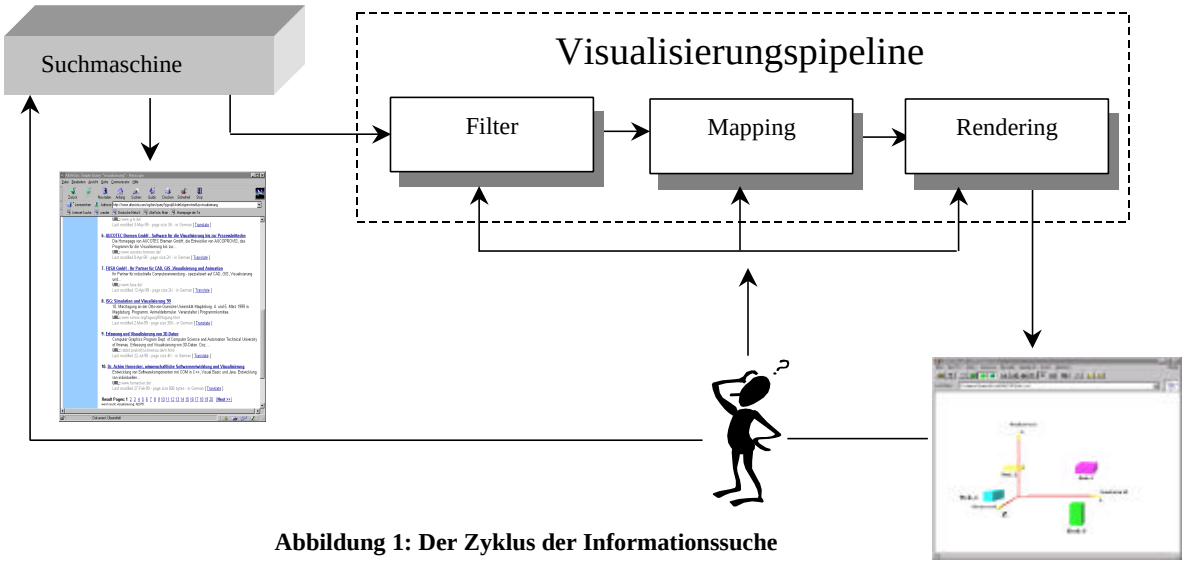


Abbildung 1: Der Zyklus der Informationssuche

Die Schnittstellen-Implementation ist über das <APPLET>-Tag in einer HTML-Seite realisiert. Dieses Tag bindet das Applet in eine HTML-Seite ein, das die Darstellung der Dokumente als geometrische Objekte enthält. Zusätzlich sind von dieser HTML-Seite aus dem Applet Parameter übergeben worden. Dies ist mit dem <PARAM>-Tag realisiert. Das <PARAM>-Tag wird innerhalb der <APPLET>-Umgebung der HTML-Seite plaziert.

<APPLET CODE="DocVis8.class" WIDTH="400" HEIGHT="400">

<PARAM NAME =„VIS_NUMBER_OF_KEYWORDS“ VALUE = „Suchbegriffanzahl“>
<PARAM NAME =„VIS_KEYWORD_1“ VALUE = „Suchbegriff“>
...
<PARAM NAME =„VIS_NUMBER_OF_DOCUMENT“ VALUE = „Dokumentanzahl“>
<PARAM NAME = VIS_DOCUMENT_1 VALUE = „Dokumentitel“>
...
<PARAM NAME = VIS_RELEVANT_1 VALUE = „Wert1“>
<PARAM NAME = VIS_RELEVANT_2 VALUE = „Wert2“>
<PARAM NAME = VIS_RELEVANT_3 VALUE = „Wert3“>
<PARAM NAME = VIS_RELEVANT_4 VALUE = „Wert4“>
...

In dieser Liste von Parametern ist die Anzahl der Suchbegriffe angegeben. Die Suchbegriffe werden mit genauem Namen und in einer Reihenfolge beschrieben. Die Anzahl der Dokumente und dessen Titel ist ebenfalls definiert. Zu jedem Dokument sind die Relevanzwerte bezüglich der einzelnen Suchbegriffe angegeben.

Das Einlesen von Parametern

Das Einlesen von Parametern wird dann mit Hilfe einer geeigneten Methode erfolgen. Über die Methode wird die Anzahl der Suchbegriffe und die entsprechenden Namen, die auf der HTML-Seite für die Parameter festgelegt sind, übergeben. Diese Methode liefert die

Suchbegriffe und die Dokumente mit den entsprechenden Titeln und Relevanzwerten der Suchbegriffe.

Wird ein Applet parametisiert, kann es universell eingesetzt werden. Das Applet besitzt eine Schnittstelle zur HTML-Seite, über die die Eigenschaften des Applets von außen verändert werden können. Man muß also nicht den Quellcode des Applets modifizieren und neu kompilieren, um ihm neue Eigenschaften zu verleihen.

Darstellung des Informationsraumes in Browsern

Nachdem die Daten die Schnittstelle zwischen Suchmaschinen und dem Applet passiert haben, soll ihre Darstellung durch das Applet in einem Browser durchgeführt werden.
Die Java-Anwendung ist wie oben beschrieben, als Applet geschrieben und das Applet kann sowohl im Appletviewer als auch im Browser ausgeführt werden und so über das WWW verfügbar sein. Jeder, der Zugang zum World Wide Web besitzt, kann diese Anwendung benutzen.

Dieses Prinzip gilt natürlich nicht nur für das World Wide Web. Auch innerhalb eines Intranets können die Java-Anwendungen zur Verfügung gestellt und somit genutzt werden. Die Software soll nicht mehr auf dem lokalen Rechner installiert werden, da die Anwendung innerhalb des Browsers oder Appletviewers unabhängig vom System des Benutzers ausgeführt wird.

3. Grafische Auswertung der Daten

Das grundlegende Ziel dieses Systems zur Informations-Visualisierung ist es, relevante, interessante Informationen aus einer großen Menge zur Verfügung stehender Daten zu selektieren und zu visualisieren. Der Ablauf der einzelnen Phasen für die Erzeugung einer visuellen Repräsentation von Informationen ist der aus der wissenschaftlich-technischen Visualisierung bekannten Visualisierungspipeline entnommen, der aus nicht-geometrischen Daten in mehreren Schritten oder Phasen eine visuelle Darstellung generiert [4]. Die modulare Aufbauweise der Visualisierungspipeline wird adaptiert und angewendet, um Informationsmengen darzustellen zu können.

Die Filterstufe

Für die Darstellung der Informationen im dreidimensionalen Raum liegen die dafür benötigten Daten in einer unaufbereiteten, abstrakten Form vor. Während die Filterstufe in der Pipeline der wissenschaftlich-technischen Visualiserung Aufgaben wie Glättung, Normierung oder Vervollständigung der Daten übernimmt, ist in der Informationsvisualisierung ausschließlich die Reduktion der Datenmenge interessant.

Die folgenden Varianten der Datenfilterung sollen genauer betrachtet werden. Filterung nach:

- dem Namen des Dokumentautors,
- dem eingetragenen Datum,
- der Größe des Dokumentes,
- der Erscheinungsart,
- einer Relevanz-Schwelle,
- Redundanzen.

Voraussetzung für das Funktionieren der einzelnen Filter ist das korrekte Vorhandensein der entsprechenden Parameter (Datum, Autor, ...) zu den jeweiligen Dokumenten.

Filterung nach dem Autor

Der Ausgangspunkt für diese Filterung ist folgende Situation: Der Nutzer hat die Menge der interessanten Dokumente bereits eingegrenzt, indem er sich in den entsprechenden Bereich des Informationsraumes navigiert hat. Mit Hilfe der vom System zur Verfügung gestellten Interaktionsmechanismen hat er den Inhalt einiger Dokumente näher untersucht. Ist ein relevantes Dokument gefunden worden, können nun alle vorhandenen Dokumente desselben Autors herausgefiltert werden. Diese Möglichkeit ist besonders dann nützlich, wenn das Ziel der Informationssuche unscharf ist.

Filterung nach dem Erscheinungsdatum

Ein großes Problem des WWW (World Wide Web) ist, daß ein Teil der Informationsbestände nicht gepflegt wird, was dazu führt, daß eine Reihe der abrufbaren Informationen bereits überholt ist. Einige Dokumente sind auch in verschiedenen Versionen an verschiedenen Orten des Netzes vorhanden.

Filterung nach der Dokumentengröße

Die von einer Suchmaschine gefundenen Dokumente können verschiedener Art sein: Kurzmeldungen im Umfang weniger Zeilen, Berichte, Aufzählungen, Werbung bis hin zu kompletten Büchern. Die Größe der Datei kann als Hinweis auf die Art des Dokumentes dienen. Es ist jedoch anzumerken, daß dies auf keinen Fall als scharfes Kriterium angesehen werden darf. Bei großen Datenmengen sollten dennoch alle Möglichkeiten zur Reduktion der Datenmenge ausgeschöpft werden.

Filterung nach dem Ort der Veröffentlichung

Dieser Filter kann benutzt werden, um in gewissem Maße Beziehungen zwischen Dokumenten zu visualisieren. Häufig sind wissenschaftliche Veröffentlichungen in „Sammlungen“ eingebunden, wie Bücher, Proceedings von Workshops u.ä. Die auf diese Weise zusammengefaßten Dokumente haben natürlicherweise eine enge thematische Beziehung, die sehr einfach visualisert und nutzbar gemacht werden kann. Hat der Anwender einmal ein solches, für ihn interessantes Dokument gefunden, können alle zur betreffenden Sammlung gehörigen Dokumente herausgefiltert werden. Damit steht dem Nutzer sofort ohne weitere Suche eine Menge von Dokumenten zur Verfügung, die für ihn mit sehr hoher Wahrscheinlichkeit ebenfalls interessant sind.

Filterung nach Relevanz-Schwellwerten

Zu jedem Dokument sind zu jedem Suchbegriff die entsprechenden Relevanzen bekannt. Um die Menge der zu visualiserenden Dokumente zu reduzieren, kann festgelegt werden, daß ein Dokument dessen Relevanz bezüglich eines bestimmten Schlüsselwortes unterhalb eines spezifizierten Schwellwertes liegt, von der Weiterverarbeitung ausgeschlossen wird. Auf diese Weise kann außerdem eine Wichtung der Suchergebnisse vorgenommen werden.

Verhindern von Redundanzen

Bei der Verwendung von Meta-Suchmaschinen erscheinen besonders viele Dokumente mehrfach in der Ergebnisliste und sind damit auch mehrfach im Informationsraum vorhanden. Mit Hilfe dieses Filters ist es möglich, redundante Dokumente von der Weiterverarbeitung auszuschließen.

Die grafische Darstellung der Dokumente

Das Mapping der gefilterten Daten (Informationen) in einen virtuellen Informationsraum kann wesentlich komplizierter sein als bei der wissenschaftlichen Visualisierung. 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. Weiterhin hängt die Informationsrepräsentation, die erfolgen soll, vom Datenmodell und von spezifischen Benutzeranforderungen ab. Die im Visualisierungssystem existierenden Dokumente besitzen nur eine mathematisch erfaßbare Größe, die zur Abbildung herangezogen werden kann - die Relevanz bezüglich der Schlüsselwörter.
Ein anderes Problem besteht darin, daß Informationsräume multidimensional sind. Das bedeutet, sie besitzen abstrakte Objekte mit oft mehr als drei Objektdimensionen. Da maximal drei der Dimensionen auf die Raumachsen abgebildet werden können, müssen verstärkt grafische Attribute benutzt werden.

Mapping mit Glyphs

Als Glyphs [5], oder auch Ikonen, werden geometrische Objekte bezeichnet, deren Erscheinungsbild von bestimmten Parametern oder Zuständen abhängig ist.

Position

Jeder Glyph wird initial mit der Position P(x, y, z)=(0.0, 0.0, 0.0) erzeugt. Die abzubildenden Relevanzen sind zu diesem Zeitpunkt bereits normiert. Das System verfügt über die Information, welche Relevanz R auf welche Raumachse abzubilden ist. Die Abbildung selbst erfolgt über eine einfache Translation des Glyphs. Hier müssen - je nach Anzahl der abzubildenden Schlüsselworte - drei Fälle unterschieden werden:

- Ein Schlüsselwort:
  Es wird ein Relevanz-Parameter auf eine Raumachse abgebildet, so daß der Glyph auf einer Achse verschoben wird.

- Zwei Schlüsselworte:
  Das Abbildungsverfahren des ersten Falles wird um eine Raumdimension erweitert, die neue
  Das Erscheinungsbild ist das eines gebräuchlichen X-Y-Diagramms.

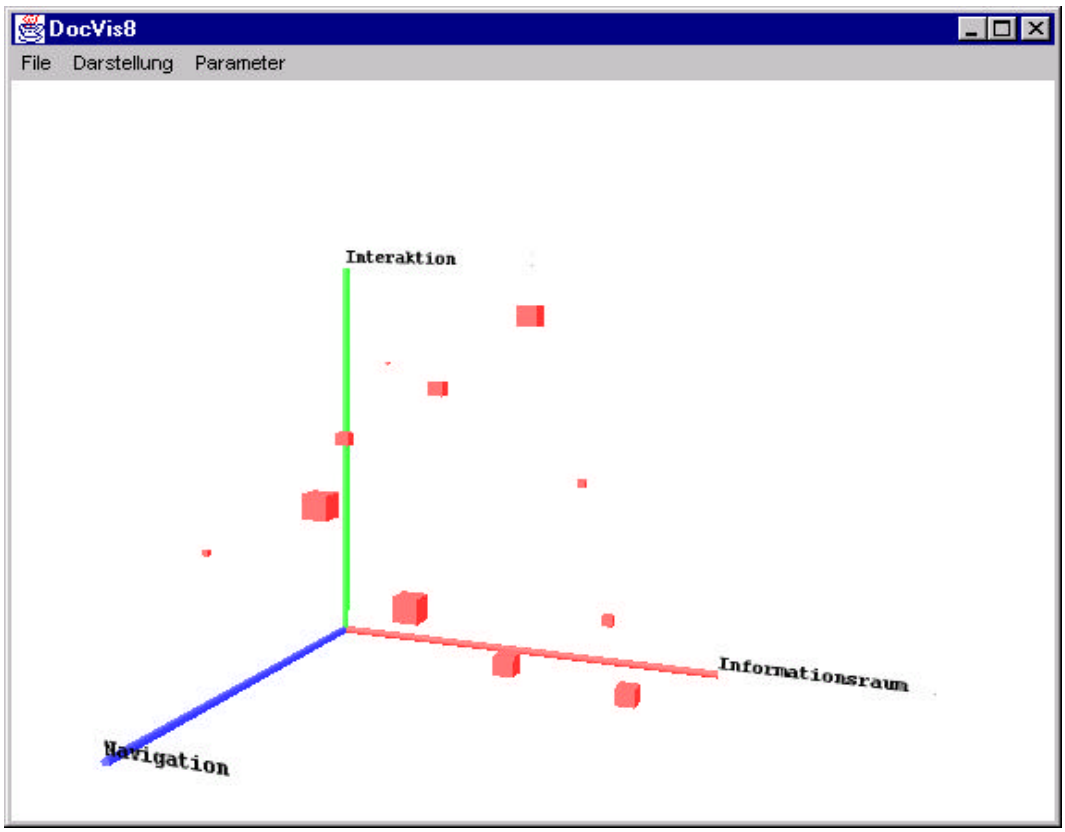
- Drei Schlüsselworte:
  Es werden alle drei Raumachsen für die Abbildung der Relevanzen verwendet.

Größe

Häufig ist es sinnvoll, einen Parameter auf die Größe der grafischen Erscheinungsform der Dokumente abzubilden. Dieses Mapping ist sehr stark von der Art der ausgewählten Repräsentationsform abhängig.

Transparenz

Das Abbilden einer Relevanz auf die Transparenz der Dokumentrepräsentation ist besonders bei Informationsräumen mit sehr hoher Dokumentendichte von Bedeutung. Dabei sollte so vorgegangen werden, daß eine niedrige Relevanz eine hohe „Durchsichtigkeit“ der entsprechenden grafischen Objekte bedeutet.



5. Ausblick

Definition von Meta-Tags

MetaTags sind eines der wichtigsten Hilfsmittel für eine wirklich erfolgreiche Registrierung in Suchmaschinen, die die Dokumente enthalten können.
Es ist zu untersuchen, ob sich ein fest definierter Satz von MetaTags nutzen läßt, um die Suche nach wissenschaftliche Dokumenten effizienter zu gestalten.

Prüfung der Möglichkeiten von XHTML

Die XHTML verspricht zwei entscheidende Vorteile:

- Erweiterbarkeit: In einer XML –basierten Daten-Typ-Definition (DTD) können neue Elemente einfach zur DTD hinzugefügt werden, solange die Konsistenz gewahrt bleibt
- Portabilität: Dokumente können in zunehmender Zahl auf alternativen Plattformen, also Nicht-Desktop-Sytemen, dargestellt werden.

Eine der zentralen Anforderungen an XHTML-Dokumente ist es, daß die Dokumente exakt an die Regeln der Spezifikation halten müssen.

Literatur:

[1] Martha Barberena N. *Graphische Darstellung von Informationen*, MIK98 4. Workshop Multimediale Informations- und Kommunikationssysteme 8.10 und 9.10.1998, Ilmenau.

[2] Martha Barberena N. *Graphische Informationsdarstellung auf der Basis dreidimensionaler Visualisierung*, STJA 99 - MIK99 5. Workshop Multimediale Informations- und Kommunikationssysteme 28-30. Sept, Kongreßzentrum Messe Erfurt AG.

[3] Däßler R., Palm H. *Virtuelle Informationsräume mit VRML: Informationen recherchieren und präsentieren in 3D.* Heidelberg: dpunkt-Verl.; ISBN 3-9209993-78-0, 1998.

[4] Müller U., Leissler M., Hemmje M. *Entwurf und Implementierung eines generischen Mechanismus zur dynamischen Einbettung multimedialer Daten in VRML-Szenen auf der Basis eines objektrelationalen DBMS* GMD Forschungszentrum und Informationstechnik GmbH.-Sankt Augustin; 1998 ISBN 3-88457347-0

[5] Barth R. *Vorlesungsunterlagen: Graphische Datenverarbeitung III, Spezielle Verfahren der Computer Graphik.* Institut für Praktische Informatik und Medieninformatik, Techniche Universität Ilmenau, 1996