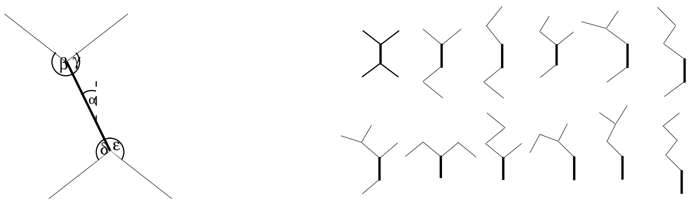

a. Strichmännchen b. Eine Familie von Strichmännchen

Abb. 6: Strichmännchen Technik

die resultierende Visualisierung Strukturmuster, die gemäß der Datencharakteristika variieren. Als Strichmännchen können verschiedene Icons mit unterschiedlicher Dimensionalität verwendet werden (vgl. Abb. 6). Die Strichmännchen-Technik eignet sich insbesondere für die Visualisierung von Daten mit 5 bis 10 Attributen und einer gleichmäßigen Anordnung bezüglich zwei der Attribute.

Das VisDB System ist in C++/Motif implementiert und läuft unter X11/R5 auf Unix-Workstations (HP 7xx).

Literatur

[ID 90] Inselberg A., Dimsdale B.: *'Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry'*, Visualization '90, San Francisco, CA, 1990, pp. 361-370.

[KK 94] Keim D. A., Kriegel H.-P.: *'VisDB: Database Exploration using Multidimensional Visualization'*, Computer Graphics & Applications, Sept. 1994, pp. 40-49.

[Kei 95] Keim D. A.: *'Visual Support for Query Specification and Data Mining'*, Shaker Verlag, 1995.

[KK 95] Keim D. A., Kriegel H.-P.: *'Visualisierungstechniken zur Exploration und Analyse sehr großer Datenbanken'*, Proc. Datenbanksysteme in Büro, Technik und Wissenschaft (BTW), Dresden, 1995, in: Informatik Aktuell, Springer, 1995.

[PG 88] Pickett R. M., Grinstein G. G.: *'Iconographic Displays for Visualizing Multidimensional Data'*, Proc. IEEE Conf. on Systems, Man and Cybernetics, IEEE Press, Piscataway, NJ, 1988, pp. 514-519.

Leitung:

Prof. Dr. Hans-Peter Kriegel Tel. (089) 2178 - 2191 / -2190

Mitarbeiter:

Dr. Daniel A. Keim Tel. (089) 2178 - 2225

Unsere Anschrift:

Lehr- und Forschungseinheit für Datenbanksysteme
Institut für Informatik, Ludwig-Maximilians-Universität München
Oettingenstr. 67, D-80538 München
Fax: (089) 2178 - 2192
E-Mail: { kriegel | keim }@informatik.uni-muenchen.de

VisDB:

Ein Visualisierungssystem zur Exploration und Analyse großer Datenbanken

Das VisDB System, das am Institut für Informatik der Universität München entwickelt wurde, ist ein Visualisierungssystem, das die Exploration und Analyse großer Datenbanken mit Hilfe neuartiger Visualisierungstechniken unterstützt. Neu sind nicht nur die Visualisierungstechniken, die eine gleichzeitige Visualisierung der größtmöglichen denkbaren Anzahl von Datenwerten erlauben, sondern auch die anfragebezogene Betrachtung der Daten.

Die prinzipielle Idee unseres Ansatzes ist die gleichzeitige Darstellung möglichst vieler Datenobjekte am Bildschirm, wobei jeder Datenwert durch ein Pixel des Bildschirms repräsentiert wird. Die Farbe eines Pixels entspricht dem Abstand des zugehörigen Datenwertes zum Anfragewert. Die Anordnung hängt von der Gesamtdistanz des Datenobjektes in Bezug auf die Anfrage und von der gewählten Visualisierungstechnik ab. Die Interaktion mit dem VisDB System erfolgt über eine graphische Benutzerschnittstelle, die es dem Benutzer u.a. erlaubt, Anfragen inkrementell zu ändern, Beziehungen zwischen Attributen herzustellen, sowie auf die zugehörigen Datenwerte zuzugreifen. Im folgenden sollen die durch die derzeitige Version des VisDB Systems unterstützten Visualisierungstechniken kurz vorgestellt werden.

Spiralanordnung

Bei der Spiralanordnung wird jeder Datenwert (bzw. sein Abstand zum Anfragewert) durch ein Pixel repräsentiert. Die Distanzwerte für die einzelnen Attribute sowie das Gesamtergebnis werden in separaten Fenstern dargestellt (vgl. Abb. 1). Die Anordnung der Pixel geschieht spiralförmig um die Mitte der Fenster herum (vgl. Abb. 1). Die Reihenfolge der Pixel entspricht dabei der Sortierung entsprechend der Gesamtdistanz. Im Fenster für das Gesamtergebnis sind in der Mitte die gelben Pixel; weiter nach außen verlaufen die Farben kontinuierlich von hellgrün bis dunkelbraun. Die Fenster für die einzelnen Attribute weisen keine kontinuierlichen Farbübergänge auf, da die Pixel der Attribute in derselben Reihenfolge angeordnet sind wie im Fenster für das Gesamtergebnis. Die Visualisierung einer Datenbank mit k Attributen besteht insgesamt aus k+1 Fenstern der gleichen Größe. Die Pixel, die zu den Attributwerten eines Datensatzes gehören, liegen in verschiedenen Fenstern (vgl. Abb. 1).

Achsenanordnung

Die Achsenanordnung ist eine Weiterentwicklung der Spiralanordnung und beruht auf der Idee, die Daten entsprechend ihrer Distanz bezüglich zweier

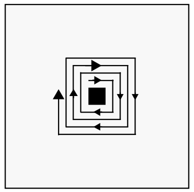

Abb. 1: Spiralanordnung eines Attributs

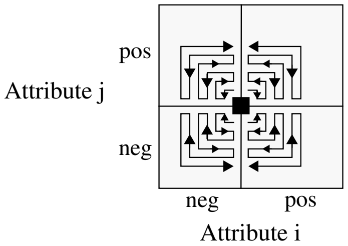

Abb. 2: Achsenanordnung eines Attributs

ausgewählter Attribute auf dem Bildschirm anzuordnen. Da Attributwerte im allgemeinen kleiner oder größer als das vorgegebene Anfrageintervall sind, ergeben sich bei der Berechnung der Distanzen positive und negative Werte. Diese zusätzliche Information wird bei der Achsenanordnung ausgenutzt, um die Daten in vier Quadranten anzuordnen. Dazu werden zwei Attribute ausgewählt, und die Teilbilder für die Attribute bzw. die Gesamtdistanz durch zwei orthogonale Achsen in jeweils vier Quadranten aufgeteilt. Datensätze mit positiven Distanzen werden rechts bzw. oberhalb, Datensätze mit negativen Distanzen links bzw. unterhalb der jeweiligen Achse gezeichnet. Die Anordnung der Datensätze innerhalb der Quadranten erfolgt rechtwinklig um das Zentrum herum (vgl. Abb. 1). Die Datensätze, die die Anfrage erfüllen, liegen wie bei der Spiralanordnung im Zentrum der Achsen.

Gruppenanordnung

Bei der Spiral- und Achsenanordnung sind die Pixel, die die Distanzen eines Datensatzes bezüglich seiner Attribute darstellen, in mehrere Teilbilder aufgeteilt. Im Gegensatz dazu werden bei der Gruppenanordnung die zu einem Datensatz gehörenden Distanzen in einem zusammenhängenden Bereich dargestellt. Die Bereiche sind wie bei der Spiralanordnung spiralförmig um die Mitte des Fensters angeordnet. Im Gegensatz zur Spiralanordnung besteht die Visualisierung aus nur einem Fenster, das die Distanzwerte sämtlicher Datensätze darstellt, d.h. die Visualisierung ist nicht in Teilbilder für die Attribute und die Gesamtdistanz unterteilt.

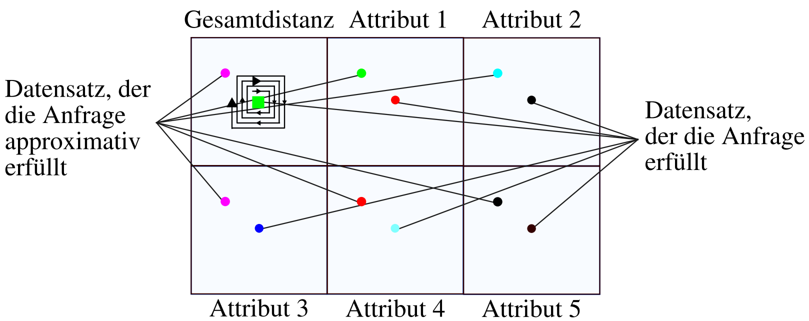

Abb. 3: Anordnung der Fenster bei der Visualisierung fünfdimensionaler Daten

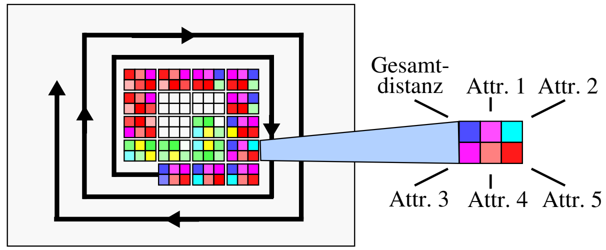

Abb. 4: Gruppenanordnung von Daten mit fünf Attributen

Technik der Parallelen Koordinaten

Eine weitere Visualisierungstechnik ist die Technik der Parallelen Koordinaten (parallel coordinates (ID 90)). Diese Technik stellt die k Attribute mit Hilfe von k äquidistanten Achsen dar, die parallel zu einer der Bildschirmachsen liegen. Die Achsen entsprechen den Attributen und sind vom Minimum- bis zum Maximumwert der Attribute linear skaliert. Jeder Datensatz wird als polygonale Linie dargestellt, die jede Achse an dem Punkt schneidet, dessen Wert dem jeweiligen Attributwert entspricht (vgl. Abb. 5). Obwohl die Grundidee der `parallel coordinates'-Technik einfach ist, ermöglicht sie das Erkennen eines weiten Spektrums von Datencharakteristika, wie z.B. verschiedene Datenverteilungen und funktionale Abhängigkeiten. Wegen der Überlappungen der Linien ist die Anzahl der Datensätze, die gleichzeitig visuell darstellbar ist, jedoch auf ca. 1.000 begrenzt.

Strichmännchen Technik

Eine weitere bekannte Visualisierungstechnik ist die sog. Strichmännchen (stick figure)-Technik (PG 88). Bei dieser Technik wird jeder Datensatz durch ein Strichmännchen dargestellt, wobei die Winkel und Strichlängen die Attributwerte repräsentieren. Wenn die Datensätze im Bezug auf die Bildschirmdimensionen verhältnismäßig dicht beieinander liegen, zeigt

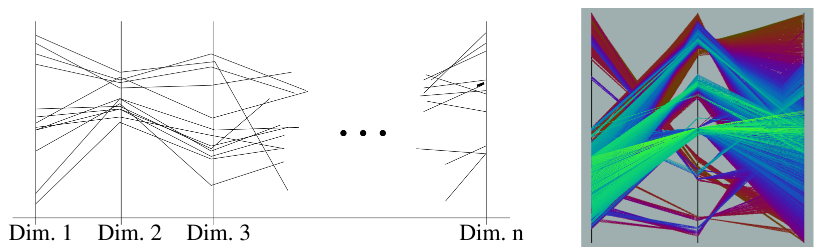

Abb. 5: Technik der Parallelen Koordinaten