Forschungsverbund NRW
Bits ® Bilder ® Bedeutung

Zwischen Daten und Display

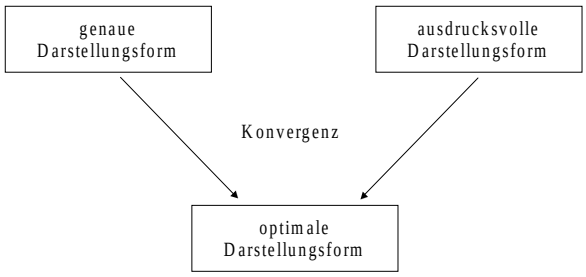
Prof. Dr. Mihai Nadin
Computational Design
Universität-GH Wuppertal

Wir wissen, wie man Bilder erzeugt; wir wissen weniger, wie man gute Bilder erzweugt; wir wissen sehr wenig, wie aus Bildern Wissen extrahiert wird.
Das Thema des Vortrages: Um Wissen aus Bildern zu extrahieren, müssen wir gute Bilder erzeugen.

Wissensaquisition ist das eigentliche Vorhaben der Forscher. Dazu bedienen Sie sich verschiedenster Methoden. Noch vor nicht all zu langer Zeit war die ganze Praxis der Erkenntnisgewinnung auf Sprache basiert und letztlich im wesentlichen in den Formalismen der Mathematik verankert war. Die Mathematik selber hat sich fundamental geändert und durch die Computation mehr als je zuvor die Ausdruckskraft des Visuellen in Anspruch genommen. Einige Bereiche der Mathematik würden heutzutage überhaupt nicht existieren, wenn die Visualisierung durch Computer nicht möglich wäre. Dazu gehören nicht nur die Theorien dynamischer Systeme (im Volksmund als Chaostheorie bekannt), sondern auch die ganze fraktale Mathematik, um nur zwei der bekanntesten Entwicklungen zu nennen.
Wenn man von den computationalen Wissenschaften spricht - die computationale Physik, Chemie, Astronomie, usw. - spricht man eigentlich von der Aneignung visueller Darstellungsmethoden und Techniken, die zum Geist dieser Zeit gehören. Heutzutage erreicht die visuelle Kommunikation bis zu 80% der gesamten Kommunikation - und das nicht nur durch Fernsehen und Werbung, sondern durch die visuelle Vermittlung der Information in der Arbeit, Freizeit, Gesellschaft. Aber darüber zu sprechen ist letztendlich nicht unsere Aufgabe.

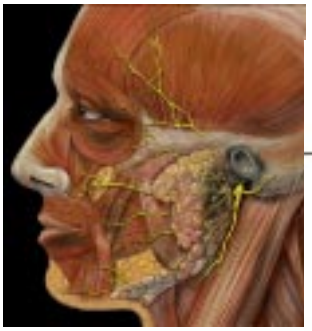
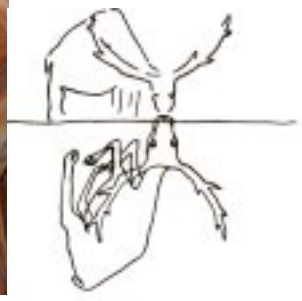
Genauigkeit

Wir wollen eher erneut versuchen, die Visualität als eine gestaltete Darstellung zu verstehen, die im wesentlichen uns durch Bestimmungen - analog zu einer Grammatik - erlauben, von Daten zu Bildern und von Bildern zu Wissen zu kommen. Diese Dimension der Gestaltung ist fundamental. Ohne in der eigenen Sache (Gestaltung) Partei zu ergreifen, muß man sich schon im klaren sein, daß Visualisierung genauso präzise sein muß, wie eine mathematische Formel, genauso nachvollziehbar, aber zusätzlich auch ausdrucksvoll sein muß. Kurzgefaßt: in der Visualisierung spiegelt sich ein fundamentales Gesetz wider: je präziser eine visuelle Darstellung ist, desto geringer die Ausdruckskraft (und sicherlich auch umgekehrt). Dieses Gesetz darf niemand, der visualisiert, ignorieren. Vom Gesetz an sich kann jedoch niemand Schlußfolgerungen für eine erfolgreiche Visualisierung ableiten.



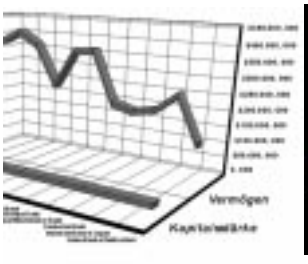
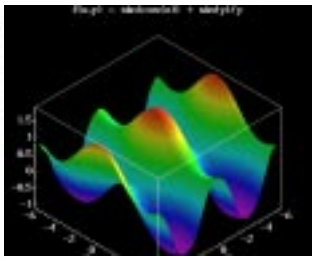
Wir vertreten - wie schon inzwischen im Forschungsverbund NRW bekannt - den Standpunkt, daß Design, daß heißt Gestaltung, ein Teil der Ausbildung der Informatiker sein muß. Daß die Integration des Design in der Informatikausbildung nicht von sich aus stattfindet wissen wir alle. Somit stehen wir vor der Aufgabe, Design in computationaler Form auszudrücken und zu vermitteln. Diese Aufgabe ist nicht einfach. Das liegt nicht daran, daß Design im Gegenteil zu den Wissenschaften nicht genau ist, sondern eher, weil Genauigkeit als Merkmal der Erkenntnis bisher nur auf Quantität reduziert, aber nicht auch auf Qualität ausgedehnt wurde. Quantitative und qualitative Aspekte der Phänomene müssen zusammengefaßt werden und die Erwartung, die an diese Genauigkeit geknüpft wird - auch wenn

diese als Fuzzy-Genauigkeit verstanden wird - muß die Gesamtheit erfassen.
Was in jeder Visualisierung passiert, ist die Übersetzung vom Quantitativen (Daten) ins Qualitative (Bilder). Dies gilt beispielsweise für einfache Illustration,

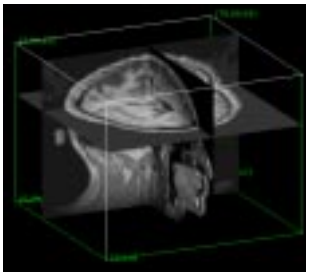



geometrische Darstellung,


computergraphische Rekonstruktion

Modellierung,



Simulation,

VR Darstellung.

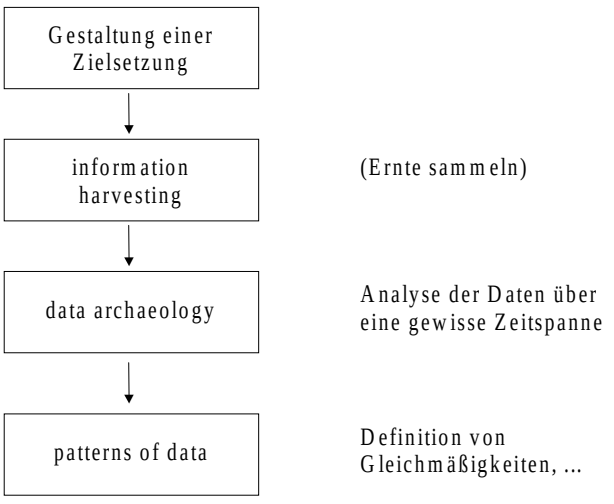
Zwischen Daten und Bildern soll ein grundlegender Isomorphismus der Genauigkeit bestehen bleiben.

Was wird extrahiert?

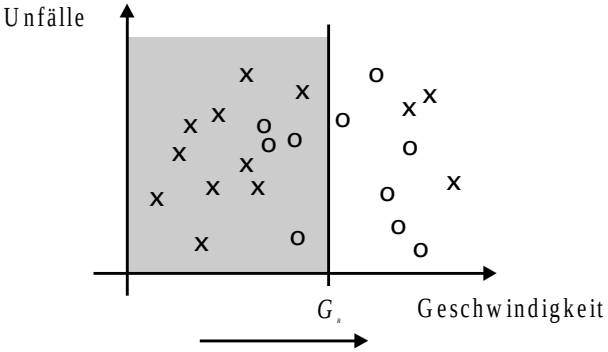
Nun aber zurück zu Wissensacquisition oder Erkenntnisentdeckung als Aspekte des Data Mining. Der erste und wichtigste Element ist dabei die Bestimmung der notwendigen Daten. Wir müssen uns im klaren sein: alles wird gemessen, alles wird durch Daten dargestellt. Meistens sind solche Daten bedeutungslos oder falsch, weil sie durch Messungsverfahren extrahiert wurden, die die Phänomene selbst gestört haben. Bevor Daten extrahiert werden, muß unbedingt eine klare Zielsetzung definiert werden: es wird im Zusammenhang mit einem Ziel gemessen, das wir auf abstrakter Ebene durch die Definition der notwendigen Variablen und deren uns bekannten Zusammenhänge beschreiben. Hier fängt die Schwierigkeit an: wir messen, weil wir eigentlich mit unserem Wissen weiterkommen möchten. Also, wir wissen noch nicht, welche Variablen wichtig sind, welche Zusammenhänge bedeutungsvoll oder bedeutungslos sind, usw. Diese Zusammenhänge sind im Nachhinein feststellbar.
Die klassische Induktion (immer als Teil-

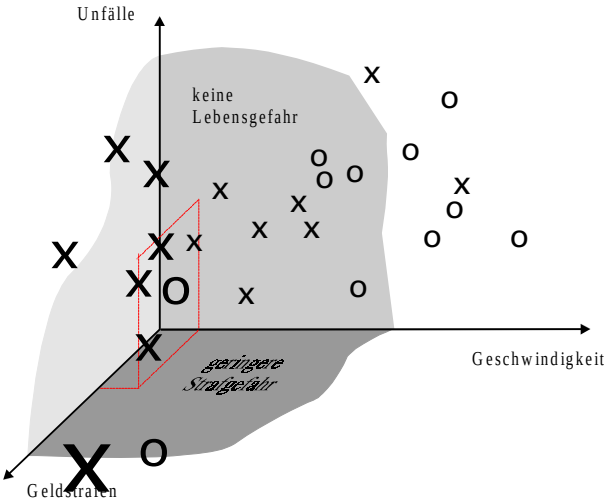
induktion zu verstehen) und die klassische Wertinduktion helfen dabei. Ohne jedoch ein abduktives Verfahren in Gang zu setzen, werden wir auch weiter in der Schlußfolgerung (der Induktion, Deduktion) das finden, was wir "hereinbeobachtet" haben.

Deswegen kann dies letztendlich nur zu Tautologien führen. Wir können heute dadurch, daß die Computation effektiver geworden ist, verschiedene Datensätze aus verschiedenen Quellen verknüpfen - jeder Datensatz als Prämisse einer effektiven Abduktion. Verschiedene Datenbanken (lokal gespeichert, verteilt, parallel verarbeitet,usw.) können kombiniert werden und gegenseitig auf Vollständigkeit, Widersprüchlichkeit, Kohärenz usw. geprüft werden. Die Integration auf Datenebene und die Integration auf Bildebene sind fundamental verschieden. Die extrem große Datenbank des *Human Genome Projektes* (1994, K.H. Fassmann, A.J. Cuticchia, D.T. Kingsbury, The GDB (TM) Human Genome Database), die Datenbank des Himmels (Astronomie U.M. Fayad, R. Wthurusanny, 1995 AAAi Press, 1995), usw., bestehen aus Eingaben im Bereich der Terrabytes. Bei der NASA werden im sogenannten Earth Observing System (EOS) pro Stunde ca. 50 GB gespeichert. Die intelligente automatische Bearbeitung, d.h. Visualisierung, die eventuell winzige Erkenntnisse ermöglicht, besteht aus Prozeduren, die Gestaltung – Bilderzeugung - Bildanalyse und Bildoptimierung und Interpretation kombinieren. Die Schritte die hier vorliegen sind folgende:

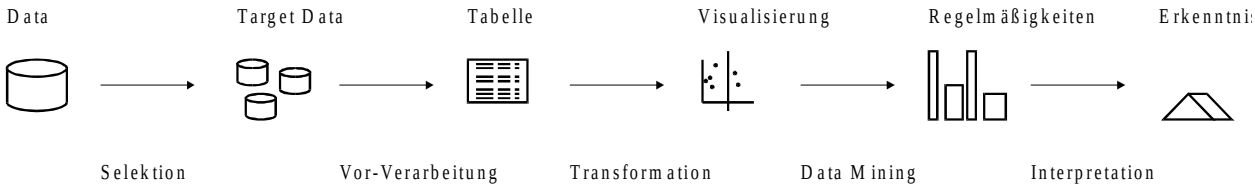


Die erste Methode (1989- Knowledge Discovery in Databases (KDD)) basiert auf Algorithmen zum extrahieren von Regelmäßigkeiten in den Daten. Diese Regelmäßigkeiten werden dadurch auf visueller Ebene geprüft. Es soll damit die alte Tendenz zum "data fishing" beendet werden, indem man definiert, was eigentlich zu messen ist, d.h. welche Daten machen für eine gewisse Zielsetzung Sinn. Diese erste Methode beinhaltet statistische Prozeduren für den Aufbau des Modells und für die Handhabung von Stördaten (noise). Datawarehousing gehört in diese Kategorie. OLAP (on-line-analytical processing, 1993, Codd) ermöglicht die Anwendung im Netzwerk.

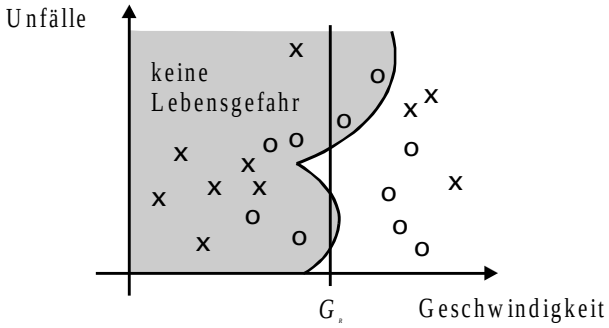


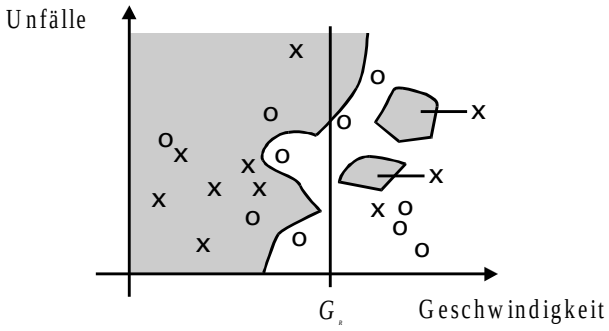


Daten, eine Menge von Fakten (X_n, Fälle mit 3 Feldern: Unfälle, Geschwindigkeit, Tempolimit). Muster (Pattern), ein Ausdruck E in einer Sprache L die Fakten beschreibt in einer Untermenge F_E.



Durch weitere Verarbeitungsmethoden, wie z.B. auch Backpropagation in Neuronalnetzwerken, werden neue Klassifizierungen möglich.





Aufgrund solcher Methoden wird nicht nur festgestellt was war, sondern es wird auch eine mögliche Handlung (in die Zukunft) projiziert, Die Aussagen (Prediction) die auf Algorithmen basieren, können weiter visuell differenziert werden.

Statistische graphische Darstellungen (Graphen, Charts) werden heutzutage immer mehr durch interaktive Visualisierungen ersetzt. Die Technik des Brushing (die Maus wird wie ein Pinsel (brush) benutzt, der farbige Spuren hinterläßt) um die Zusammenhänge zu visualisieren, die auch durch die Intensität der Farbe angezeigt werden, enstand durch diesen Versuch.

Durch Data Mining ist es gelungen, den Zusammenhang zwischen Bildern und Erkenntnis zu thematisieren. Deswegen glaube ich , daß wir etwas mehr über Bilder zu sagen haben.

Die ästhetische Grundlage

Eigentlich gibt es kein Bild, das nicht direkt oder indirekt ästhetisch bewertet wird, d. h. es gibt kein Bild, das nicht als „schön“, „gelungen“, „angenehm“ oder als „nicht schön“, „ nicht gelungen“, „unangenehm“ u.ä. wahrgenommen wird. Diese Wahrnehmung ist öfter eine implizite (man denkt darüber nach und handelt entsprechend, d.h. man akzeptiert es oder lehnt es ab), insbesondere wenn wir keine Wahl haben. Dies gilt auch für einen selbstreflektierenden Künstler in Anbetracht seiner eigenen Arbeit, Sonntagsmaler, Bastler und auch immer öfter die wachsende Anzahl derjenigen, die mittels unterschiedlicher Programme Bilder erzeugen.

Es entstand durch Einbeziehung der Wissenschaft und Technik des Computers tatsächlich eine Kultur des Visuellen, die sich von der Werbung bis hin zu den Versuchen der Visualisierung komplexer wissenschaftlicher Sachverhalte (inkl. virtuelle Räume und Handlungen) erstreckt.
Nur einige wenige, die sich im Visuellen bewegen, haben dabei Schwierigkeiten. Im Vergleich zum Schreiben und Lesen, zur Mathematik, und sogar zur Musik scheint alles, was sichtbar ist, genauso natürlich zu sein, wie das Essen und Schlafen, das Laufen und die Liebe. So mancher hat jedoch schon gemerkt, daß das Essen selber auch nicht so natürlich ist, daß nicht jeder ein guter Koch werden kann, daß an das Laufen und die Liebe auch noch Erwartungen gestellt werden, die über die Physik (Biometrik) hinausgehen, und daß das Schlafen, bzw. die Welt des Traumes zu relativ kompli-

zierten wissenschaftlichen Ausarbeitungen geführt hat.

Tatsächlich ist das Sichtbare – Himmel, Erde, Bäume, Wasser, Sterne u. ä. – das Fenster, die Schnittstelle zu unserer Umwelt. Aber damit ist sie noch kein Bild – weil Bilder, egal welche, gute, schlechte, zufällige – die Person beinhaltet, die das Bild erzeugt hat; und mit der Person, die Kultur, die diese Person geprägt hat. Anders ausgedrückt, Bilder sind genauso künstlich wie die geschriebene und gelesene Sprache, wie die Formeln der Mathematik, die musikalische Notation, oder die Laban – Kurzschrift des Tanzes, die Kineme (eines Eisensteins) der Kinematographie oder die Scriptsprache der Multimedien (z.B. Director, mit oder ohne Lingo-Erweiterungen).

Daß Bilder als künstliche Ausdrucksformen der Menschen zu verstehen sind, dürfte niemanden überraschen.
Jede Form des menschlichen Ausdrucks erfordert Erfahrungen, die weit über das Dargestellte (in schriftlicher Form, in mathematischer Notation, in der Musik, im Tanz, in Bildern, usw.) hinausgehen. Diese Erfahrungen mußte man sich vorher angeeignet haben.
In erster Linie ist in jedem Ausdruck unsere biologische Wirklichkeit in komprimierter Form enthalten. Tatsächlich können nie Elemente benutzt werden, die unsere Sinnesorgane nicht wahrnehmen können. Das klingt so einfach, daß man darüber weiter kein Wort verlieren muß. Jedoch am Computer könnte man es einrichten, daß Farben außerhalb des sichtbaren Spektrums benutzt werden, oder Töne außerhalb des hörbaren Spektrums, usw. Damit dürfte - unter Umständen – ein anderer Teil der Welt angesprochen sein (Elefanten-Liebesrufe befinden sich bekanntlich im Ultraschallbereich, Fledermäuse operieren mit Signalen, die viel höher als die 20.000 Hz der theoretischen Hörgrenze des Menschen liegen. Gewisse Bilder, die wir als Darstellungen von uns bekannten Objekten sehen, erscheinen als Stimulus für schwer erklärbare Handlungen von Leuten, die wir als geistig gestört etikettiert haben (weil diese in einem anderen Spektrum der Sinnesorgane operieren). Aber nicht nur die biologische Wirklichkeit ist in Bildern wiederzuentdecken, sondern auch unsere kulturelle. Wobei der Begriff Kultur hier eigentlich etwas zu einfach, zu grob gewählt ist. Die Kultur selber ist als eine Verflechtung von mehreren menschlichen Dimensionen zu verstehen: die Erfahrung des Machens, des Wissens und der Interaktion (mit der Umwelt, mit den Mitmenschen, mit Objekten und sogar mit Begriffen).
Um mich aber hier nicht in Einzelheiten zu verlieren - egal wie aufregend und aufschlußreich diese auch sind - möchte ich einfach behaupten, daß alle diese Erfahrungen stets konstituieren und gleichzeitig eine grundlegende Struktur widerspiegeln, und zwar die des Ästhetischen.

Wer meint, daß man Bilder (auf die wir uns von diesen Zeitpunkt an beschränken wollen) außerhalb der ästhetischen Erfahrung erzeugen kann, der bewegt sich auf dem selben Grat wie die Alchimisten, oder diejenigen, die nach dem Perpetuum Mobile suchen. Aber was heißt nun ästhetische Erfahrung? Ich kann auf eine unendliche Bibliographie verweisen, aber mein Beitrag soll hier nicht die Wiederholung der Geschichte der Ästhetik sein, sondern eher eine Zusammenfassung in einem Begriff, den wir in computionaler Form benutzen können. Ästhetische Erfahrung ist die Selbstkonstituierung des Einzelnen in der Praxis des Lebens (egal in welcher Form dieser Praxis - als Jäger, Sammler, Bauer, Arbeiter, Wissenschaftler, Lehrer, Pfarrer, Politiker, usw.) gemäß des universellen Gesetzes der Selbstoptimierung.
Tatsächlich ist jede menschliche Handlung – egal welcher Natur – eine durch Optimierung, (d.h. der besten Relation zwischen Nutzen und Anstrengung) geprägte und zwar, weil innerhalb des Auslesens das Überleben, und das Erreichen eines immer höheren Lebensstandards alles, was nicht optimal ist, zum verschwinden bringt. Die optimale Selbstkonstituierung ist nicht eine für immer gegebene; sie ist vielmehr dynamisch zu verstehen.

Zu Elementen dieser Optimierung zählt alles, was unsere Sinnesorgane beeinflußt, oder durch sie weitergegeben werden kann.

Darum scheint mir auch heute noch die geniale Definition von Baumgarten (1750), “Ästhetik als Logik der Sinne”, die am besten angepaßte und durch die ästhetische Praxis mehrmals bestätigte, die erklärt, was Ästhetik ist. Wie schon gesagt, habe ich nicht die Absicht, die Geschichte zu erzählen, sondern möchte ihr sogar eher widersprechen.

In seinem im Jahre 1974 veröffentlichten Buch „Ästhetik als Informationsverarbeitung - Grundlagen und Anwendung der Informatik im Bereich ästhetischer Produktion und Kritik” (Springer Verlag, Wien, New York)“ versuchte Frieder Nake, die Ästhetik zu definieren, indem er die Kunst, bzw. Kunstwerke als Referenz nahm. Vielleicht ist dies der einzige Punkt, in dem ich nicht mit Frieder Nake übereinstimme, und zwar, weil für mich die grundlegende ästhetische Struktur eine Voraussetzung für die praktische Erfahrung der Kunst (Produktion, Wahrnehmung, etc.) ist, und nicht umgekehrt die Ursache dieser tiefgreifenden, strukturierenden Merkmale der menschlichen Existenz und des menschlichen Handelns.
Aber das ist ein anderes Thema, was nur insofern hier und heute wichtig ist, als jeder Versuch, Bilder zu erzeugen - ob als Kunst oder als wissenschaftliches Mittel der Kommunikation - unausweichlich ästhetisch geprägt ist.

In seinem Buch definiert Frieder Nake, und damit nähern wir uns dem Thema, den Raum der möglichen Bilder, und läßt die Zahl der möglichen Bilder offen. Hier finde ich den Ansatz, der im heutigen Kontext des Interesses für das, was die Computerwissenschaftler „Data Mining“ nennen, sehr vielversprechend ist. Aber bleiben wir bei der Sache. In Übereinstimmung mit Azriel Rosenfeld wird ein Bild als eine reelle Funktion über der Ebene definiert.

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}$$

Der Wert f(x,y) eines Bildes f auf einem innerhalb der Fläche definierten Punkt (x, y) steht für die dort tatsächlich bestehende Farbe.

Sicherlich könnte man an dieser Stelle verschiedene Einwände formulieren - es gibt Bilder, die sich im 3-dimensionalen Raum entfalten, Farbton, Intensität und Farbsättigung, was durch Vektoren beschreibbar ist (wie es auch Nake gemerkt hat), sind eher durch fuzzy Werte als durch reelle Zahlen definierbar. Es fehlt auch noch die Zeitkomponente, d.h. Farbe und Licht, unter denen man Farbe wahrnimmt, sind nicht trennbar, usw. Aber an und für sich ist diese Beschreibung ein guter Ausgangspunkt, besonders in Hinblick auf die weitere Definition des digitales Bildes - dem eigentlichen Objekt dieser Analyse (Computervisualisierungen entstehen ausschließlich als digitale Bilder): Hier wird der Unterschied zwischen der Rasterung des *Definitionsbereiches*, die zu einem digitalen Bild (daß heißt, einer reellen Matrix) führt und der des *Wertebereiches*, die zu einem quantitativ diskreten Bild führt, sehr genau deutlich. Im ersten Fall handelt es sich um eine Funktion, die für jeden Punkt des Rasterfeldes das Paar (x_i,y_j) liefert, das heißt, die Farbwerte innerhalb des Feldes so daß

$$f: \mathbb{N}^2 \rightarrow \mathbb{R}$$

ein digitales Bild beschreibt.
Im zweiten Fall sind an den Koordinatenpaaren (x,y) nur diskrete Werte Z_1, Z_2, ... feststellbar. Dementsprechend ist ein diskretes Bild durch

$$f: \mathbb{R}^2 \rightarrow \mathbb{N}$$

definierbar.
Das digitale diskrete Bild ist schließlich durch eine Funktion

$$f: \mathbb{N}^2 \rightarrow \mathbb{N}$$

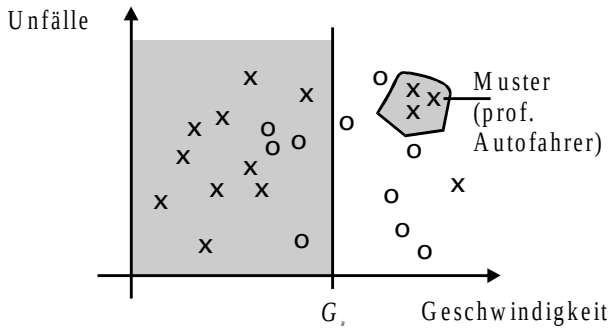
dargestellt, wobei eine Matrix mit endlichem Wertebereich eigentlich die Form ist, die bei der Computergenerierung von Bildern benutzt wird (man legt der Berechnung eine endliche Menge N Ì N von möglichen Werten zugrunde).

Weitere, viel bessere Einzelheiten als mein Kommentar es schaffen kann, liefert der Autor selber. Von Linienzeichnung, Rasterbild, Textbild hin zu Halbtonbild werden die Bildkomponenten sorgfältig definiert. Ich will mir an dieser Stelle nur die Schlußfolgerung aufheben, daß hinter jedem computergenerierten

Bild eine Matrix steht, die verschiedene Daten beinhaltet. Inzwischen wissen wir eine ganze Menge mehr über die Komplexität solcher Matrizen und über die Matrixoperationen (gewichtete Summierungen, Multiplikationen usw.), als zum Zeitpunkt des Entstehens der Arbeit von Nake. Was wir aber nicht wissen ist, inwiefern aus einer solchen Matrix mit Mühe und Aufwand verarbeiteten Daten Bilder entstehen, die auch die Voraussetzungen der ästhetischen Signifikanz erfüllen. Auch wissen wir nicht, wie man ästhetische Signifikanz ins computer-erzeugte Bild einprogrammiert.
Nake schrieb: „Dennoch sei erwähnt, daß wir im Prinzip einen simplen Algorithmus aufstellen können, der alle Objekte einer Klasse erzeugt, indem er alle möglichen Kombinationen durchspielt. So einfach ein solcher Algorithmus wäre, so unbrauchbar wäre er andererseits. Aber er würde ja angesichts des riesigen Umfangs dieser Klassen Jahrtausende benötigen, bevor ein erstes "interessantes" Objekt erzeugt wäre."
Was aber ist mit interessant im Text von Nake gemeint? Auf einem Umweg über Data Mining Prozeduren möchte ich den Versuch machen, das Interessante als Computations-Aufgabe zu definieren. Ich bin mir dabei bewußt, daß es nur ein anfänglicher Ansatz sein kann und zwar einer, der zu der Kategorie der Computation gehört, die man Wissensacquise nennt.

Interessantheitsgrad

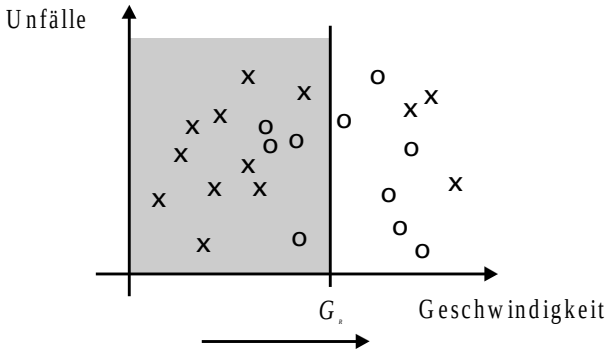
Daten sind eine Gruppe von Fakten F (z.B. Fälle in einer Datenbank).
Ein **Muster** ist ein Ausdruck E in einer Sprache L, der Fakten in einer Untermenge F_E von F beschreibt.
E wird ein Muster genannt, wenn es einfacher als die Menge aller Fakten in F_E ist.



Wenn eine Person mit einer Geschwindigkeit < G_t fährt, ist die Unfallgefahr geringer.
Das Muster stellt keine komplette Beschreibung dar (Es gibt Autofahrer, die trotz einer Geschwindigkeit > G_t keine Unfälle verursachen.)

Prozeß: Die Integrierung der Datenvorbereitung, Suche nach Mustern, Wissensauswertung, Verfeinerung durch Iteration nach der Modifikation.
Der Vorgang, den wir als Prozeß definieren, besteht aus mehreren Schritten.

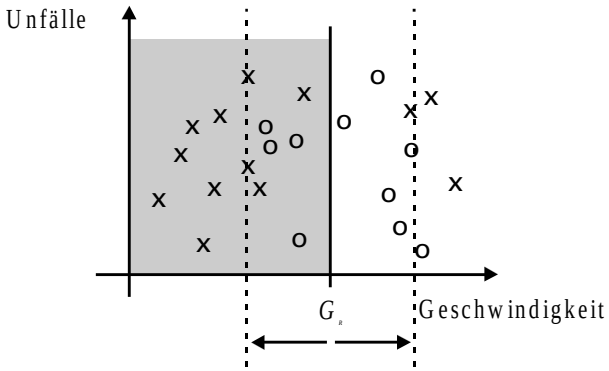
Gültigkeit: Entdeckte Muster sollten zu einem hohen Grad an Wahrscheinlichkeit auch für neue Daten gültig sein. Eine Funktion C stellt ein Maß an Wahrscheinlichkeit dar, das Ausdrücke L auf einem teilweise oder komplett geordneten Maß-Raum M_c abbildet.
Einem Ausdruck E in L über einer Untermenge F_E (Muster) Ì F (Menge von Tatsachen) kann ein Wahrscheinlichkeits-Maß c = C (E,F) zugewiesen werden.



Wenn der im Faktenfeld F definierte Schwellenwert nach rechts bewegt wird, dürften mehr Unfälle geschehen

Neuartigkeit: Muster sind von der Definition her neuartig für das System.
Neuartigkeit wird bezüglich der Änderungen der Daten (durch Vergleich von gegenwärtigen mit bisherigen oder erwarteten Werten) oder des Wissens (wie ein neues Ergebnis mit alten verwandt ist) gemessen.
N kann durch eine Funktion N(E,F) dargestellt werden, die eine Boolesche ist oder ein Maß an Neuheit oder – das Unerwartete - vermittelt.

möglicherweise nützlich
Muster sollten möglicherweise zu nützlichen Handlungen führen. Diese sollte man durch eine Nutzfunktion messen können.
Solch eine Funktion U bildet Ausdrücke in L auf einen teilweise oder komplett geordneten Maß-Raum M_u ab; folglich gilt u = U(E,F).
Es ist beispielsweise nützlich, nicht über Tempolimit zu fahren, um Strafen zu vermeiden.



letztlich verständlich

- Das Ziel ist, Muster verständlich zu machen.
- Es ist immer schwer, präzise zu messen.
- Es gibt viele Einfachheits-Maße. Man kann sich auf eine:
- syntaktische
  (Größe der Muster, ausgedrückt in Bits!) oder
- semantische (leicht verständliche) Ebene beschränken.

Dies wird, falls möglich, von einer Funktion S gemessen, die die Ausdrücke E aus L auf einen teilweise oder komplett geordneten Maß-Raum M_S abbildet, folglich gilt s=S(E,F).

Interessantheitsgrad
Der Grad an Informativität kann über eine Ordnung gefundener Muster definiert werden.
Der Interessantheitsgrad ist das Gesamtmaß von Muster-Werten, das Gültigkeit, Neuheit, Nutzen und Einfachheit verbindet.
Die Funktion für den Grad an Informativität, die Ausdrücke in L auf einen Maß-Raum Mi abbildet, lautet:
i= I (E,F,C,N,U,S)
Man müßte herausfinden, was ein benutzter Algorithmus als Wissen identifiziert.
Als Data Mining kann eine Versicherung feststellen, welche 11 Risiken sich in jedem Versicherungsvertrag verbergen.

Beispiel: Eine Beschreibung eines Bileds von Rembrandt dürfte zu einer Replikation eines Rembrandt-Bildes führen. Frieder Nake (aber nicht nur er) hat Mondrian-Bilder auf diese Art und Weise erzeugt. Inwieweit bietet ein Algorithmus die Mlglichekt, das Wissen über die Rembrandt-Ästhetik (oder die von Mondrian) zu erkennen?

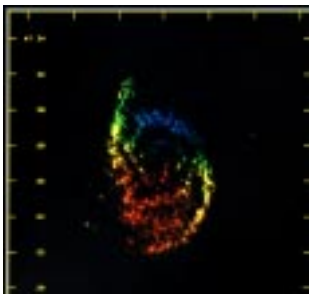
Wissen

Ein Muster E Î L wird als Wissen bezeichnet, wenn eine benutzer-definierte Schwelle i Î M_I, I(E,F,C,N,U,S) > i definierbar ist. Wissen ist immer benutzer-orientiert und wird durch beliebige Funktionen und Schwellen, die vom Benutzer gewählt werden, bestimmt.

Zum Beispiel:
Wir können irgendeine Schwelle c Î M_C, s Î M_S, u Î M_U wählen und die Wissensmuster Wissen nennen, wenn - und nur wenn - C (E,F) > c, S(E,F) >s, U (S,F)> u ist.

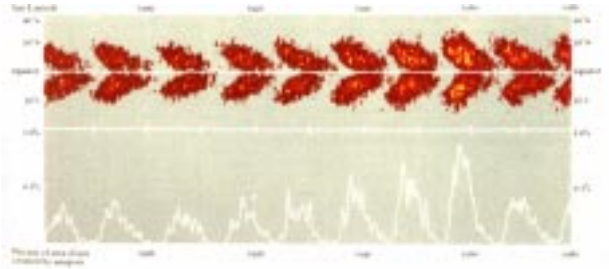
Durch entsprechende Einstellungen von Schwellen kann man zu genauen Vorhersagen oder nützlichen Muster (über Andere) kommen. Es gibt einen unendlichen Raum von Möglichkeiten, wie die Abbildung I definiert werden kann.

Data Mining



Ein Schritt in der Wissensentdeckung in Datenbanken besteht aus speziellen Data Mining Algorithmen, die eine besondere Aufzählung von Mustern Ej über F (unter akzeptabler begrenzter computationaler Effektivität) erzeugen.

Der Raum von Mustern ist üblicherweise ein unendlicher Raum. Eine beliebige Aufzählung von Mustern kann durch begrenzte Suchvorgänge in diesem entstehen. Es wird fast immer nur in Unterräumen nach Mustern gesucht.



Der KDD Prozeß des Entdeckens von Wissen in Datenbanken ist nur ein Anfang.
Es werden Algorithmen zum extrahieren (identifizieren) verwendet, was gemäß der Spezifikationen von Maßen und Schwellen als Wissen erachtet wird. Dabei wird die Datenbank F zusammen mit jeder notwendigen Vorberechnung, Untermusterung (subsampling) und Transformation von F genutzt.

Nehmen wir an, daß alle diese Definitionen operativ eingesetzt werden. Nehmen wir weiter an, daß Computation möglich ist. Können wir uns tatsächlich erhoffen, von allen möglichen Bildern die es gibt – computergeneriert oder nicht – die interessantesten zu finden? Und können wir weiter davon ausgehen, daß das Interessante auch universell ist?
Ein Experte in Geodaten definiert das Interessante aus dem Blickwinkel der Handlungen, die zu seinem Gebiet gehören. Ein Physiker, ein Chemiker, ein Biologe werden auch von der Interessantheit – und zwar in Bezug zu einem klaren Zeitpunkt – der eigenen Wissensdomäne ausgehen.
Das ästhetisch Interessante mag außerhalb des Ästhetischen nicht anerkannt werden. Dementsprechend möchte ich nach diesem ersten Versuch, Interessantheit zu defnieren, die Schlußfolgerung ziehen, daß nur ein dynamischer Begriff in Frage kommt. Dafür reichen meine heutigen Kenntnisse nicht – somit ist vielleicht das Ende meines Vortrages ein Anfang für Andere, die diese Gedanken weiterentwickeln können.