

Die Qualität von Visualisierungen: Eine Methode zum Vergleich zweidimensionaler Karten

Maximilian Eibl
GESIS - Berlin

Thomas Mandl
FB Informations- und Kommunikationswissenschaften - Universität Hildesheim

Abstract: Zweidimensionale Karten dienen im Information Retrieval (IR) immer häufiger der Visualisierung von Dokumentbeständen. Über die Qualität der verschiedenen Methoden zur Erzeugung dieser Karten ist wenig bekannt. In diesem Artikel werden nach einem Überblick über Stand der Forschung und Technik zu Visualisierungen und ihrer Bewertung zwei Verfahren hinsichtlich ihrer Unterschiede verglichen. Die Evaluierung basiert auf der Perspektive des Benutzers beim *Browsing* in einer Karte. Das Experiment zeigt, dass die Methoden sowohl bei großen als auch bei kleinen Dokumentenmengen zu sehr unterschiedlichen Karten führen und zwischen den Karten keine Korrelation besteht. Daher müssen Methoden zur Abbildung einer Dokumentenmenge auf eine zweidimensionale Karte anwendungsspezifisch ausgewählt werden.

1 Einführung

Visualisierung dient der Reduktion der Komplexität und ermöglicht Benutzern das schnelle Erkennen komplexer Zusammenhänge in Wissensstrukturen. Im Information Retrieval Prozess unterstützt die Visualisierung mehrere Stufen und kann etwa der Darstellung von Termen, logischen Beziehungen zwischen Anfrage-Operatoren oder Dokumenten dienen. Für letzteres werden oft zwei- oder dreidimensionale Räume entworfen, in denen etwa die Gravitationsmetapher wirkt. Zweidimensionale Karten drücken semantische Nähe von Dokumenten durch geringe räumliche Distanz aus. Am häufigsten wird die Kohonen *Self-Organizing Map* (SOM, cf. Kohonen 1998) benutzt, aber auch Latent Semantic Indexing (LSI, cf. Berry et al. 1995) oder Faktorenanalyse können zur Reduktion von Repräsentationen auf zwei Faktoren dienen. Diese Ansätze scheinen ausgereift; inzwischen finden sich bereits kommerzielle Produkte auf der Basis dieser Ideen. So bieten z.B. *newsmaps*¹ und *cartia*² zwei- und drei-dimensionale Karten im Internet an (siehe Abbildung 1) und *PersonalBrain* von *SER*³ lernt die Kategorisierung von Dokumenten des Benutzers in vorgegebene Klassen ohne eine Visualisierung direkt anzuschließen.

Obwohl die Theorien der Wahrnehmung Vorteile versprechen, bleibt unklar inwieweit Visualisierungen den Information Retrieval Prozess unterstützen. Auch die Softwareergonomie verspricht einen einfacheren Einstieg, da der Benutzer aufgrund des Vorlagecharakters selektieren kann und z.B. nicht in einem leeren Eingabefeld zunächst produzieren muss. Gewissheit können aber nur empirische Evaluierungen bringen und diese sind selten, auch für zweidimensionale Karten.

In den letzten Jahren haben zwei Implementierungen zweidimensionaler Karten im Internet große Mengen von Dokumenten analysiert und dargestellt. Chen et al. 1996 haben mit einer Kohonen-SOM die Dokumente aus der Yahoo-Kategorie *Entertainment* klassifiziert und dargestellt. Dabei werden die Dokumente anhand ihrer Eigenschaften (in der Regel das Vorkommen von Termen) und nach einem Algorithmus analysiert. Die große Anzahl von Dokumenten zwang zu einer schichtenweisen Realisierung, bei der man von einer groben Überblickskarte in mehreren Schritten hin zu einer detaillierten Karte mit Dokumenten navigiert. Die Interaktion ähnelt einem Zoom-Vorgang.

¹ <http://www.newsmaps.com/>

² <http://www.cartia.com/>

³ <http://www.serware.de/de/serware/download/download-000.html>

Eine noch umfangreichere Datenbasis liegt dem System WEBSOM⁴ (Kohonen 1998) zugrunde. Darin werden etwa eine Million Beiträge zu Diskussionsforen (Internet Newsgroups) organisiert und visualisiert. Ziel ist es, neben der hierarchischen Struktur der Diskussionsforen einen assoziativen Zugang über mehrere Newsgroups hinweg zu ermöglichen. Kohonen 1998 wählte als Strategie für die Bewältigung der großen Datenmenge wie Chen et al. 1996 einen schichtenweisen Aufbau.

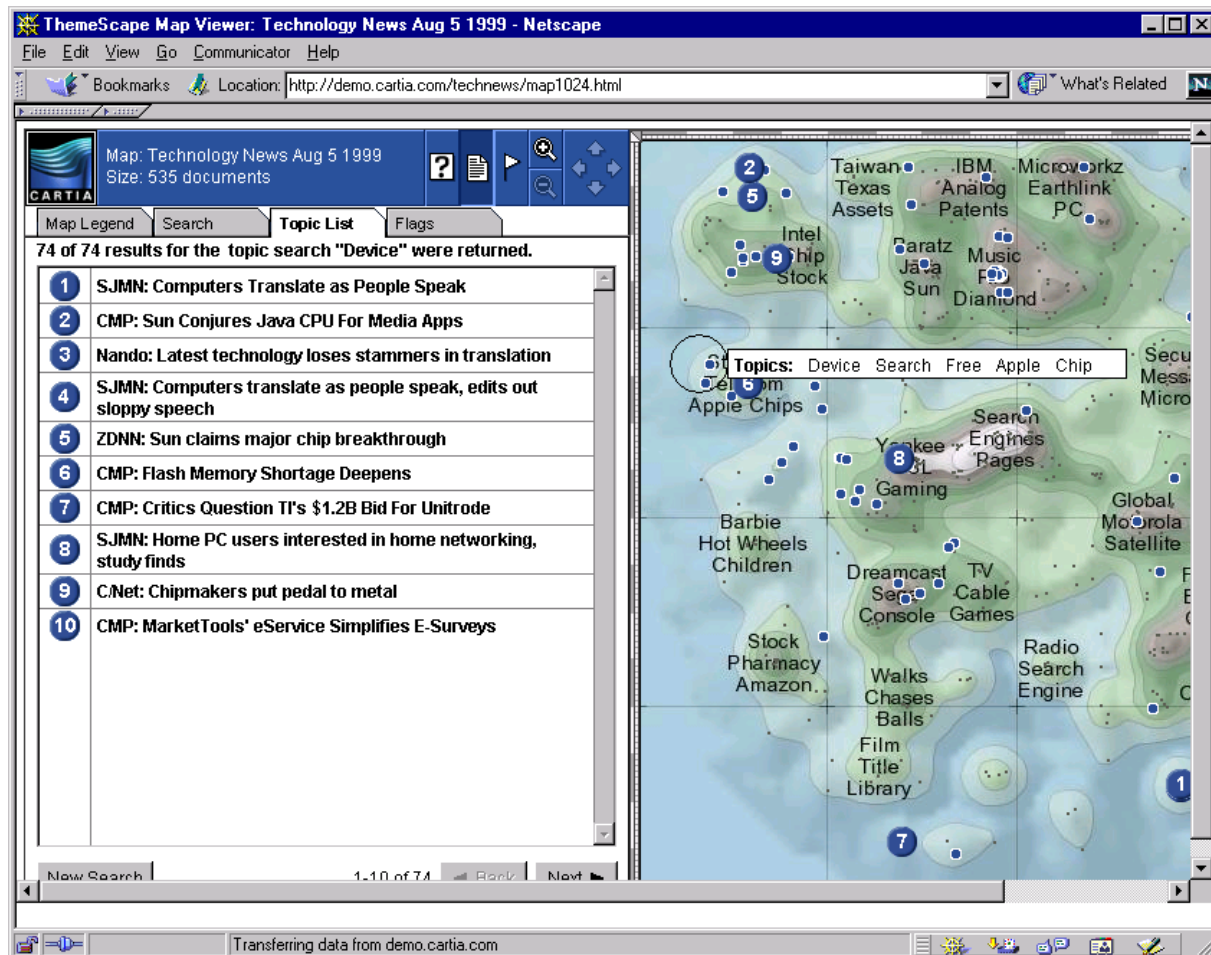


Abbildung 1: Benutzungsoberfläche von Cartia mit Suchmöglichkeiten

Die beiden Systeme zeigen auch Schwächen des Ansatzes. Lediglich Chen et al. 1996 berichtet von informellen Benutzertests, so dass keine allgemeine Aussagen über die Vorteile der SOM oder zweidimensionaler Karten bei der Lösung von Informationsproblemen möglich sind. Die Autoren begründen auch die Wahl ihrer Methode nicht und können für den jeweiligen Anwendungsfall die Überlegenheit der SOM gegenüber anderen Methoden nicht nachweisen. In diesem Beitrag soll gezeigt werden, dass dies unbedingt erforderlich ist, da die zwei Methoden zu völlig unterschiedlichen Karten führen können.

2 Visualisierung

Im Oktober 1986 veranstaltete die *National Science Foundation* eine Tagung mit Computer Graphik Experten, auf der der Begriff "Visualization in Scientific Computing" (ViSC) geschaffen wurde. Im darauffolgenden Februar wurde der *Workshop on Visualization in Scientific Computing* durchgeführt, auf dem die Grundlagen und Möglichkeiten der Visualisierung besprochen und Initiativen ihrer Erforschung angeregt wurden. Obwohl

⁴ <http://websom.hut.fi/>

auf dem Gebiet der Visualisierung bereits ausgiebig Forschung betrieben wurde, war es der erste Versuch, sie zu systematisieren. Der Begriff Visualisierung wurde wie folgt definiert:

"Visualization is a method of computing. It transforms the symbolic into the geometric, enabling researchers to *observe* their simulations and computations. Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights. In many fields it is already revolutionizing the way scientists do science.

Visualization embraces both image understanding and image synthesis. That is, visualization is a tool both for interpreting image data fed into a computer, and for generating images from complex multi-dimensional data sets. It studies those mechanisms in humans and computers which allow them in concert to perceive, use and communicate visual information. (...) The goal of visualization is to leverage existing scientific methods by providing new scientific insight through visual methods." (McCormick et al. 1987: 3)

Auch wenn diese Definition später als Definition für *scientific visualization* verwendet wurde, bezog sie sich zunächst allgemein auf Visualisierung. Auffällig ist, dass hier noch zwei gegensätzliche Richtungen einbezogen werden. Zunächst wird Visualisierung als Methode beschrieben, symbolische Darstellungen in geometrische umzuwandeln, und so dem Wissenschaftler eine bessere Möglichkeit zu geben, Simulationen zu beobachten. Als besonders wichtig wird dabei angesehen, dass Visualisierung zur wissenschaftlichen Entdeckung entscheidend beitragen kann.

Andererseits wird jedoch auch die Analyse bildlicher Daten durch den Computer unter dem Oberbegriff Visualisierung gefaßt. Neuere Definitionen lassen diesen Aspekt aus. Hier wird nur noch die Generierung einer graphischen Darstellung berücksichtigt. Betrachtet man jedoch die interaktiven Möglichkeiten computerbasierter Visualisierungen, so scheint der Wegfall des analytischen Teils ein Rückschritt in die *post-mortem* Zeit zu sein. Wolff 1996 beispielsweise beschreibt im Gegensatz dazu ein graphisches Recherchetool für Faktendaten, welches auf der Darstellung von Liniendiagrammen basiert. Hier können graphische Eingaben vorgenommen werden, die zunächst analysiert werden, bevor die entsprechende Ausgabe generiert wird.

Inzwischen haben sich in der Forschung zwei große Bereiche von Visualisierung herausgebildet: Neben der ursprünglichen *scientific visualization* etablierte sich Anfang der 90er zunächst bei XEROX-PARC *information visualization* (cf. Däßler&Palm 1998: 40). Während *scientific visualization* konkrete physische Daten verwendet, handelt es sich bei *information visualization* um abstrakte. Nimmt man diese Zweiteilung als Ausgangsbasis, würden oben genannte Konzepte wie *program visualization* und *knowledge visualization* unter das Konzept *information visualization* fallen.

Card et al. übernehmen diese Unterscheidung und definieren *scientific visualization* als „Use of interactive visual representations of *scientific* data, typically *physically* based, to amplify cognition“ (Card et al. 1999: 7) und *information visualization* als „Use of interactive visual representations of *abstract, nonphysically* data to amplify cognition“ (Card et al. 1999: 6). Sie weisen zur Problematik der strikten Trennung der beiden Formen jedoch darauf hin, dass der Gebrauch physischer Daten bei *scientific visualization* nicht in der ursprünglichen Definition der NSF (s.o.) impliziert war, sondern sich erst in der Praxis herausgestellt hat. Ferner können die in der *scientific visualization* dargestellten Informationen auch abstrakter Natur sein: „While [scientific] visualizations may derive from abstractions of this physical space, the information is nevertheless inherently geometrical.“ (Card et al. 1999:6)

Neben diesen Punkten wird die Problematik einer strikten Trennung der Konzepte durch die Möglichkeit ihrer gleichzeitigen Anwendung weiter verschärft. Däßler&Palm 1998 führen *scientific* und *information visualization* unter dem Begriff „Multiple Informationsräume“ wieder zusammen: „Multiple Informationsräume, eine Kombination aus physisch-raumbezogenen und abstrakt-symbolischen Daten, werden die Benutzerschnittstellen der nächsten Generation von Informationssystemen maßgeblich beeinflussen.“ (Däßler&Palm 1998: 41)

Abschließend läßt sich feststellen, dass computerbasierte Visualisierung die in der Regel interaktive graphische Umsetzung von Daten bezeichnet. Handelt es sich bei den Daten um physische, so spricht man von *scientific visualization*, handelt es sich um abstrakte bzw. nicht-physische, so spricht man von *information visualization*.

2.1 Verbesserung der Interaktion durch Visualisierung

Der Hauptnutzen der Visualisierung liegt in zwei Aspekten: Der Präsentation von Daten und ihrer Interpretation. Die Basis der Argumentation für Visualisierung bietet die Quantität der Daten moderner Datenbanken, die ohne Visualisierung weder darstellbar noch interpretierbar wären.

Will man große Mengen von Daten auf einem herkömmlichen Bildschirm präsentieren, so stößt man schnell an die Grenzen der Darstellbarkeit. Leung & Apperley 1993 sehen das Problem darin, wie viel der Datenbasis dargestellt werden soll. Bei der Darstellung des gesamten Datenbestandes ergeben sich sehr schnell vehemente Platzprobleme. Es muss sich schon um eine vergleichsweise kleine Datenbank handeln, wenn man nicht auf unhandliche Hilfsmittel wie Scrolling u.ä. zurückgreifen will. Will man das Platzproblem dadurch lösen, dass nur ein Teil der Daten präsentiert wird, so steht man vor der Entscheidung, welche Daten relevant sind. Sind es zu viele Daten, tritt erneut ein Platzproblem auf, sind es zu wenige oder die falschen, ist die Darstellung ineffektiv.

Je mehr Platz darzustellende Daten einnehmen, desto größer wird das Platzproblem. Zeigt man nun die Daten in ihrer numerischen Form, so ist die Darstellung beschränkt. Formuliert man die Daten jedoch in eine graphische Darstellung um, so können die einzelnen Daten und ihre Eigenschaften auf minimalen Raum präsentiert werden. Die einzelnen Eigenschaften werden in Form, Schattierung, Farbe, Position, Größe u.ä. umkodiert. Auf diese Art und Weise kann ein Vielfaches der herkömmlichen Darstellung gezeigt werden, wodurch sich oben beschriebenes Platzproblem zumindest reduziert.

Selbst wenn dieses Platzproblem nicht bestünde, so wäre eine numerische Darstellung großer Datenmengen nicht angemessen. Die Leistungsfähigkeit des menschlichen Gehirns reicht nicht aus. "The deluge of data generated by supercomputers and other high-volume data sources (...) makes it impossible for users to quantitatively examine more than a tiny fraction of a given solution." (DeFanti et al. 1989: 12) Das von Leung&Apperley 1993 geschilderte Phänomen wird hier auf die menschliche Kognition ausgeweitet. Es können nur Ausschnitte einer Datenmenge in numerischer Form interpretiert werden. Ist dieser Ausschnitt sehr klein, so scheint seine numerische Darstellung sinnvoller, da sie genauer ist als eine graphische. Überschreitet ein solcher Ausschnitt jedoch eine gewisse Größe, oder wird ein gesamter Datensatz dargestellt, so ist der Mensch nicht mehr in der Lage, mit den Daten effektiv umzugehen.

Hier setzt die Visualisierung ein, um auch die Handhabung großer Datenmengen zu ermöglichen: "Data visualization shifts the load from numerical reasoning (...) to visual reasoning." (Parsaye & Chignell 1993: 203) Das menschliche Gehirn ist für viele Aufgaben wesentlich besser dazu ausgebildet, mit Bildern umzugehen als mit Zahlen.

2.2 Visualisierung im Information Retrieval

Der Versuch, mit Hilfe von Visualisierung die Anfragekomponente von Dokument Retrievalsystemen anwenderfreundlich und effizient zu gestalten, ist keineswegs neu. Im Vordergrund stand dabei immer die Ambition, Schwierigkeiten im Umgang mit textbasierten Anfragesprachen auszuräumen und neue Möglichkeiten des Erkennens von Relevanzen und Relationen zu eröffnen. Hier soll ein Überblick über die vorhandenen Strategien und Systeme gegeben werden, um darauf aufbauend auf generelle Fragestellungen der Konzipierung einzugehen.

Die Strategien, die bei der Visualisierung im Textretrieval verfolgt werden, lassen sich in drei grobe Richtungen einteilen. In den 80er Jahren lag der Fokus der Visualisierung vor allem auf der Formulierung der Anfrage. Mit Hilfe einer geeigneten Visualisierung sollte die Problematik der Verwendung Boolescher Operatoren umgangen werden. Durch den Einsatz von Venn-Diagrammen wurde versucht, den Anwender bei der Formulierung einer Anfrage zu unterstützen. Ein erstes Beispiel für diese Strategie ist das System GQL von Michard 1982. Diese noch vergleichsweise einfache Form der Visualisierung besteht aus drei ineinander verschachtelten Kreisen, die mit Termen initialisiert werden können. Den Vorteil von GQL sieht Michard in der Möglichkeit, dem Anwender in der Verwendung Boolescher Operatoren entgegenzukommen: "Aid in the manipulation of boolean operators for set operations. The difficulty for 'naive' users to manage with some boolean operators, for example disjunctive AND and NOT, is well known. Our query language will provide very great help for these operations using Venn diagrams to materialize selected subsets." (Michard 1982: 280f.) Tatsächlich können mit Hilfe von GQL strikte Boolesche Anfragen formuliert werden, ohne dass dabei Boolesche Operatoren in Erscheinung treten. In einem Anwendertest vergleicht Michard GQL mit einer textbasierten Anfragesprache und kommt zu dem Ergebnis, dass Anwender mit der textbasierten Anfragesprache eine um ein vielfaches höhere Fehlerquote haben.

Seit Beginn der 90er Jahre hat sich der Fokus der Forschung hin zur Präsentation der Ergebnismenge verschoben. Dies liegt zum einen daran, dass sich das Interesse der IR-Forschung insgesamt von der Booleschen Recherche abgewandt hat, und somit die Problematik der Anfrageformulierung in den Hintergrund trat. Bei *partial match*-Systemen besteht die Anfrage meist nur aus einer Liste von Suchkriterien ohne logische Verknüpfung. Zum anderen waren aber nun die Computer leistungsfähig genug, um die wesentlich anspruchsvolleren graphischen Umsetzungen der Ergebnismenge zu verarbeiten. Die neuen Systeme sollten ein

graphisches Äquivalent zu den Relevanz-Urteilen von *partial match*-Systemen bilden. Bei diesen Systemen besteht die Anfrage selbst entweder aus einer einfachen Auflistung von Suchkriterien oder einer Zusammenstellung mehrerer Kriterien aus vorgegebenen Auswahlboxen, oder aber es wird wie bei der RelevanceSphere (Hemmje 1993) eine eigene Visualisierung vorgeschaltet, um eine Präselektion durchführen zu können. Die eigentliche Visualisierung aber wird erst erstellt, nachdem die Anfrage abgeschlossen ist. In der Regel basiert die graphische Umsetzung auf der Metapher des Informations- oder Dokumentenraumes. Hierbei wird das Konzept der Relevanz von *partial match*-Systemen auf Räumlichkeit transferiert. In einem Informationsraum nehmen sowohl ikonisierte Suchkriterien als auch ikonisierte Dokumente jeweils einen bestimmten Platz ein, so dass die Verteilung der Dokumente in Relation zu den Suchkriterien gezeigt wird. Beim System VIBE (Visual Information *Browsing* Environment) (Dubin 1995; Olsen et al. 1993) spannen sogenannte *points of interest* (POI) ein zweidimensionales Polygon auf, innerhalb dessen sich die Dokumente befinden. Dieses Polygon präsentiert einen durch die POIs definierten Ausschnitt des Informationsraumes. Die einzelnen POI werden durch jeweils eine Ikone repräsentiert und umfassen ein oder mehrere Suchterme, die einen für die Suche interessanten Aspekt eines Themas beschreiben. Bei der Definition eines POI werden dem Anwender dabei keinerlei Vorschriften gemacht. Er kann einen POI als einzelnen Term definieren, als Dokument oder aber als komplexe Suchanfrage. Die Dokumente haben keine feste Position, sondern bewegen sich entsprechend der aktuellen Positionen der POI. So lassen sich bei einer glücklichen Anordnung Informationen bezüglich der Relevanzen und Gruppierungen der Dokumente leicht ablesen.

Ähnlich arbeitet die LyberSphere (Hemmje 1993; Hemmje et al. 1994) des Systems LyberWorld besteht aus einer Kugel bzw. einem Planeten über dessen Oberfläche die Terme gleichmäßig verteilt gleichsam stationären Satelliten schweben. Innerhalb der Kugel befinden sich die zu den Termen gefundenen Dokumente, deren Positionen sich wie bei VIBE aus ihren Beziehungen zu den Termen ergeben. Die konsequente Erweiterung der planetarischen Gravitationsmetapher führt zum Galaxiemodell des VINETA-Systems (Krohn 1996). Zu einem eingegebenen Term erscheinen im dreidimensionalen Raum zusätzliche Terme als Pfeile, die einen Informationsraum aufspannen, in dem Dokumente als Kugeln schweben. Ähnlich operiert auch das System SENTINEL, welches zusätzlich zu einer traditionellen Rechercheoberfläche die Ansicht der Ergebnisliste als dreidimensionale Graphik ermöglicht (Fox et al. 1999).

Eine weitere Nutzung der Informationsraum-Metapher stellen Themenkarten dar. In der Literatur finden sich hier verschiedenen Namen wie *tree map*, *theme map*, *concept map*, *theme scape* u.a., zugrunde liegt jedoch immer das gleiche Prinzip: Die N-Dimensionalität der Dokumente wird in Form einer Karte dargestellt. Auf sie geht der nächste Abschnitt genauer ein.

Die dritte Strategie schließlich umfasst die beiden ersteren. Hier werden sowohl die Anfrage als auch die Ergebnismenge visualisiert. Im Vergleich zur Visualisierung der Ergebnismenge wird diese integrale Strategie erstaunlich selten verfolgt. Der InfoCrystal (Spoerri 1994a&b) löst das Problem der Geschlossenheit von Michards GQL indem er die Teilmengen des Diagramms trennt. Resultat ist eine Darstellung, in der die Terme in ikonisierter Form als Eckpunkte ein Polygon aufspannen, und die logischen Kombinationen der Terme innerhalb dieses Polygons ebenfalls als Ikonen dargestellt werden, die mit der Anzahl der entsprechenden Dokumente initialisiert sind. Die Mengenkategorien innerhalb des Polygons sind dabei mehrfach codiert. Ihre Positionierung folgt nach den Prinzipien: „Nähe zu Ecke bedeutet Erfüllung einer Bedingung“ und „Nähe zum Zentrum bedeutet Erfüllung vieler Bedingungen“ (Roppel 1996: 130)

Um Beziehungen eindeutig darzustellen, verwendet SPOERRI noch eine ganze Reihe zum Teil redundanter Codierungen: Form, Rang, Farbe, Textur, Ausrichtung, Größen, Helligkeit und Saturierung (cf. Spoerri 1994a: 689). Mit Hilfe dieser graphischen Mittel können die einzelnen Ikonen zwar eindeutig zugeordnet werden, jedoch müssen diese Codierungen zunächst vom Anwender erlernt und bei der Bearbeitung einer Anfrage einzeln analysiert werden. Durch die Erweiterung dieser Grundidee, durch die im Zuge komplexer Anfragen die Eckpunkte eines Polygons selbst wieder durch andere Polygone definiert werden können – wodurch tatsächlich so etwas wie eine Kristallstruktur entsteht – wird die Interpretation noch weiter erschwert.

Einen ähnlichen Ansatz verfolgt das System DEViD von Eibl 2000, welches boolesches Retrieval mit vagem Retrieval und probabilistischem Ranking kombiniert. Hier wird allerdings eine im Vergleich zum InfoCrystal stark zurückgenommene Codierung verwendet, die in Kooperation mit professionellen Medien Designern konzipiert wurde. Entsprechend unterscheidet sich diese Visualisierung schon rein optisch stark von den übrigen. DEViD ist außerdem eins der wenigen Systeme, zu denen auch Anwendertests durchgeführt wurden. Diese vergleichen die Visualisierung mit textbasierten booleschen und probabilistischen Rechercheoberflächen und zeigen eine deutliche Überlegenheit DEViDs.

3 Verfahren zur Dimensionalitätsreduktion

Dieser Abschnitt stellt exemplarisch und in Kürze zwei wichtige Verfahren zur Erstellung von zwei-dimensionalen Karten aus Objekt-Merkmalismatrizen vor. Darunter sind die zwei Verfahren, die unten verglichen werden.

3.1 Faktorenanalyse

Faktorenanalyse kommt traditionell in der statistischen Datenanalyse zum Einsatz. (cf. Rodeghier 1997:174ff; Haykin 1999:392ff.). Für die Erstellung einer Faktorenanalyse (z.B. *principal component analysis*) wird zunächst eine Korrelationsmatrix aller Merkmale erstellt, die analysiert, wie stark die beteiligten Merkmale korrelieren. Im ersten Schritt bestimmt die Faktorenanalyse wie viel jede Variable einer Matrix zur Erklärung der Varianz in einer Korrelationsmatrix beiträgt. Im zweiten Schritt wählt der Anwender die n wichtigsten Variablen in der originalen Matrix. Die Faktorenanalyse fasst alle Variablen zu n Faktorengruppen oder Hintergrundvariablen zusammen, wobei der Einfluss zueinander ähnlicher Faktoren in einer Gruppe gebündelt wird. Nachträglich lässt sich der Anteil jedes originalen Faktors an einer Hintergrundvariablen bestimmen und auch visualisieren (cf. Rodeghier 1997:174ff.).

3.2 Latent Semantic Indexing (LSI)

Latent Semantic Indexing (LSI) komprimiert die Dokument-Term-Matrix und nutzt dazu das mathematische Verfahren *Singular Value Decomposition* (SVD). Dabei wird der ursprüngliche Termraum auf in der Regel zwischen 100 und 300 Variablen reduziert. LSI wird in Deerwester et al. 1990, und Berry et al. 1995 beschrieben, einen Überblick über *Singular Value Decomposition* bieten Berry et al. 1993. *Latent Semantic Indexing* arbeitet ähnlich wie eine Faktorenanalyse, jedoch sind die Faktoren oder in diesem Fall die entstehenden LSI-Dimensionen nicht interpretierbar.

LSI nutzt aus, dass es sich bei der Dokument-Term-Matrix um eine spärlich besetzte Matrix handelt, bei der nur ein kleiner Teil der Zellen mit einem Gewicht belegt ist und die somit relativ wenig Information enthält. Die vorhandene Informationsmenge fasst auch eine kleinere Matrix. Durch das mathematische Verfahren der *Single Value Decomposition* wird aus der ursprünglichen eine neue und kleinere Matrix gewonnen. Dabei werden die *Singular Values* und zwei weitere Matrizen bestimmt, von denen eine einen reduzierten Term-Raum und die andere einen reduzierten Dokumenten-Raum darstellt. Die zweite Matrix stellt eine komprimierte Beschreibung der Terme durch die Dokumente dar.

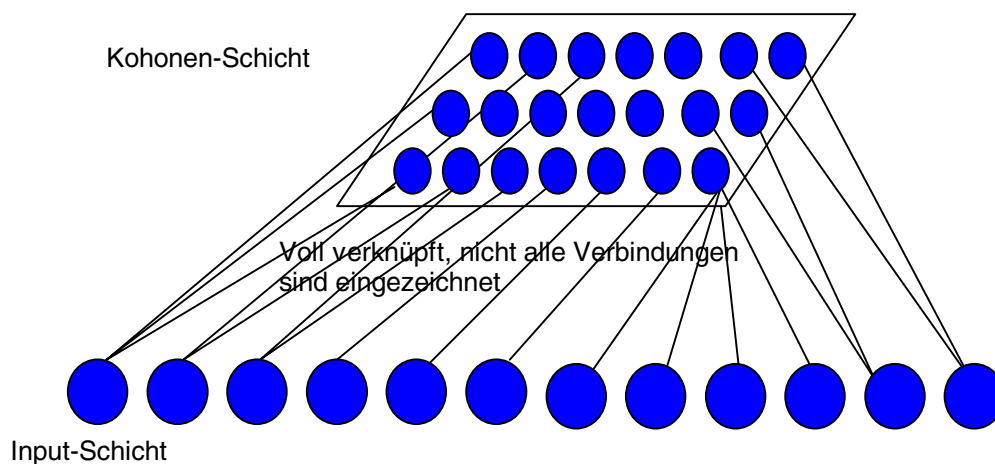


Abbildung 2: Schematisches Kohonen-Netzwerk

3.3 Kohonen SOM

Kohonen-Netze sind ein unüberwachter Lernalgorithmus, der Strukturen in Daten aufdeckt. Sie werden auch als selbstorganisierende Karten (*Self Organizing Maps*, SOM) bezeichnet (cf. Kohonen 1984, Scherer 1997:93ff.). Kohonen-Netze bestehen aus zwei Schichten, einer Eingabe-Schicht und einer Kohonen- oder Ausgabe-Schicht, in der eine topologische Karte und damit eine zweidimensionale Visualisierung der Daten entsteht. Die Anzahl der Neuronen in der Eingabe-Schicht ergibt sich aus der Dimensionalität der Eingangsdaten. Die Struktur der Kohonen-Schicht wird vom Anwender vorgegeben. Eine zwei- oder dreidimensionale Anordnung unterstützt die spätere Visualisierbarkeit der Daten. Abbildung 2 zeigt ein Kohonen-Netzwerk mit zweidimensionaler Anordnung der Output-Neuronen. Jede Unit der Eingabe-Schicht ist mit allen *Units* der Kohonen-Schicht verbunden und jedem Kohonen-Neuron ist ein Gewichtsvektor von der Größe des Eingabe-Vektors zugeordnet. Zudem sind die Kohonen-Neuronen untereinander verbunden.

Ein Ähnlichkeitsmaß wie die Euklidische Distanz vergleicht die Eingabe-Vektoren mit den Gewichtsvektoren. Das Neuron mit der geringsten Distanz oder der höchsten Ähnlichkeit zum Eingabemuster gewinnt und erhält die gesamte Aktivierung. Die Gewichtungen des Gewinner-Neurons in die Eingabe-Schicht werden so modifiziert, dass die Ähnlichkeit weiter steigt. Geometrisch betrachtet verschiebt der Algorithmus den Gewichtsvektor in Richtung des Eingabevektors. Soweit arbeitet das Kohonen-Netzwerk wie andere *Clustering*-Verfahren. Um die topologische Struktur zu erzeugen, verändern sich auch die Gewichtsvektoren der Nachbar-Neuronen des Gewinners. Dies erfordert eine Definition von Nachbarschaft, die verschiedene Funktionen wie etwa die Gauss-Funktion oder der Kosinus liefern. Diese Funktionen liefern ein Maß für die Entfernung jedes Neurons in der Kohonen-Schicht vom Gewinner-Neuron, das die Intensität der Gewichtsänderung beeinflusst. Je näher ein Neuron dem aktivierten Neuron ist, desto stärker wird sein Gewichtsvektor adaptiert. Die Vektoren sehr naher Neuronen werden somit immer in ähnliche Richtungen verschoben. Dadurch entstehen *Cluster*, in die ähnliche Muster abgebildet werden.

4 Zweidimensionale Karten für Dokumentkollektionen

Kohonen-Netze oder selbstorganisierende Karten (*Self-Organizing-Maps*, SOM) erfreuen sich in den letzten Jahren im Information Retrieval steigender Beliebtheit und sind damit eines der wenigen Modelle aus dem Bereich neuronaler Netze, die im Information Retrieval für größere Datenmengen eingesetzt wurden (cf. Mandl 2001). Kohonen-Netze dienen als unüberwachtes *Clustering*-Verfahren, das ähnliche Muster auf topologisch nahe Neuronen in der Kohonen-Schicht abbildet. Dahinter steht die Hoffnung, dass semantisch ähnliche Dokumente oder Terme in der Kohonen-Schicht nahe beieinander liegen und auf diese Weise assoziatives Browsen möglich wird. Gloor 1997 sieht Ähnlichkeit neben z.B. Verlinkung und Sequentialisierung als eines von sieben Konzepten für die Gestaltung von Navigation in Hypertext-Systemen wie dem Internet.

Bei der Reduktion gehen natürlich immer Aspekte des Ausgangsraums verloren wie Abbildung 3 illustriert. Bei LSI werden diese Verluste durch Weglassen kleiner *Singular Values* gesteuert, die bei der Rekonstruktion zu Verlusten führen. Bei SOM ist die Rekonstruktion nicht möglich und es gibt keine Abschätzung über die Qualität der Komprimierung. Während LSI und ähnliche Verfahren nach der Reduktion eine Ähnlichkeitsberechnung anschließen, dienen *Self Organizing Maps* vorwiegend der Visualisierung.

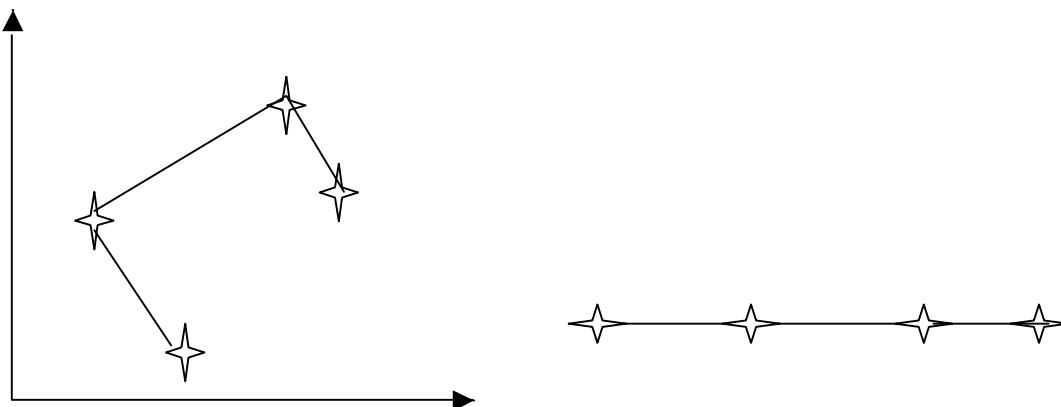


Abbildung 3: Reduktion durch Abbildung von einem zweidimensionalen auf einen eindimensionalen Raum. Bereits dieses einfache Beispiel zeigt, dass die Abbildung nicht völlig ähnlichkeitsertaltend ist. In der zwei-

dimensionalen Darstellung sind sich die Objekte A und D ähnlicher als B und C. In der eindimensionalen Darstellung rechts haben A und D dagegen minimale Ähnlichkeit.

Bei explorativen und assoziativen Suchen sollen Benutzer ausgehend von bekannten Objekten entlang der *Cluster* interessante Dokumente oder Terme finden. Die Verwendung des Begriffs Dimension für die Kohonen-Schicht in der Literatur ist zweideutig. Ein Kohonen-Netz bildet grundsätzlich einen n-dimensionalen Raum in einen m-dimensionalen ab. Während die Dimensionen der Eingabe-Schicht durch die Daten vorgegeben sind, ist die Anzahl der Klassen in der Kohonen-Schicht beliebig. Die Anzahl wird heuristisch festgelegt und sollte der Anzahl der gewünschten Klassen entsprechen. Die Kohonen-Schicht dient meist der Visualisierung und ordnet die Neuronen zweidimensional an. Die m Dimensionen oder *Cluster* werden zu Koordinaten von Punkten im zweidimensionalen Darstellungsraum. Grundsätzlich können die Ausgangs- oder Kohonen-Neuronen jedoch beliebig und in mehr als drei-dimensionalen Räumen angeordnet werden.

Die Mächtigkeit von Kohonen-Netzen bestätigen z.B. Graupe/Kordylewski 1998, die ein System für medizinische Diagnose mit realen Daten vorstellen. Das System klassifiziert Patientendaten besser als statistische Verfahren und andere neuronale Netze. Es implementiert eine Datenbankabfrage mit unvollständigen Mustern.

Lin et al. 1991 nutzen ein Kohonen-Netz für eine Kollektion von 140 Artikeln zur Künstlichen Intelligenz. Die Dokumentvektoren mit der Länge 25 dienten als Input für das Training. Das heißt, es wurden nur 25 Terme analysiert. Die Punkte in der Karte repräsentieren Dokumente, die Regionen sind aber nach Termen benannt. Dazu erstellten Lin et al. 1991 Term-Vektoren, in diesem Fall Einheitsvektoren, die ein Dokument mit nur einem Term nachbilden. Der Term-Vektor mit der höchsten Ähnlichkeit zu einem Punkt oder Neuron in der Karte bzw. dessen Gewichtsvektor bestimmt die Benennung. Die Region eines Terms erstreckt sich nur über benachbarte Neuronen, da das Lernverfahren der Kohonen-Karte Ähnlichkeit als Nähe im zweidimensionalen Raum darstellt. Lin et al. 1991 beobachten, dass die Dokument-Vektoren neben den Termen, mit denen sie indiziert sind, weitere Terme mit Gewichten ungleich Null enthalten. Diese bilden assoziative Erweiterungen der Indexterme, da sie in der Kollektion häufig gemeinsam mit ihnen vorkommen. Aufgrund der starken Dimensionsreduktion eignet sich die Karte als Grundlage einer grafischen Benutzungsoberfläche für assoziative Navigation im Korpus. Die Autoren schlagen vor, die Karte zur Visualisierung des Ergebnisses eines IR-Systems einzusetzen.

Scholtes 1992 präsentiert eine SOM für Information Retrieval, die sich im Repräsentationsmechanismus vom Standard-IR-Modell unterscheidet. Das System für russische Texte nutzt N-Gramme von Buchstaben und nicht Wörter als Eigenschaften der Dokumente. Eine Häufigkeitsverteilung über die 225 vorkommenden Trigramme ersetzt die Terme in Form von Wörtern. Diese Repräsentationen werden zunächst mit der Repräsentation von Texten durch Terme verglichen. Eine interessante Eigenschaft ist, dass auch ohne morphologische Analyse eine hohe Ähnlichkeit zwischen verschiedenen Wortformen eines Begriffs besteht. Insgesamt scheint die Repräsentation durch Häufigkeitsverteilungen von N-Grammen besser für sprachliche Analysen, wie z.B. die Identifizierung von Sprachen oder die Zuordnung von Buchstaben zu Phonemen, geeignet zu sein. Für Probleme der Semantik wie im Information Retrieval scheinen Terme besser geeignet zu sein. Endgültig kann dies jedoch nur eine empirische Überprüfung klären.

Im Zusammenspiel mit dem Kohonen-Netz führt die Repräsentation durch Trigramme zu zahlreichen Detailproblemen. So sind z.B. die häufigsten Trigramme nicht in der 15x15 Neuronen umfassenden Kohonen-Karte vertreten. Da das Kohonen-Netz ein unüberwachtes Netz ist, ergeben sich die Klassen, die in der Karte repräsentiert sind, nur aus dem Lernalgorithmus und nicht aus äußeren Vorgaben. Nach der Einschätzung von Scholtes 1992 identifiziert das Netz die für die Aufgabe am besten geeigneten Trigramme. Die Trigramm-Analyse berücksichtigt die Leerzeichen (Blanks) zwischen den Wörtern nicht, so dass das Konzept Wort keinerlei Rolle spielt.

Der Autor räumt ein, dass er für den Kontext Prawda typische Stoppwörter wie *Held* oder *Kommunismus* nicht eliminiert. Die Aussagekraft der durchgeführten Tests ist sehr gering, da die Testkollektion nur 50 Texte umfasst. Die Aufgabe im Test ist unklar und es werden keine Standard-Maße wie Recall und Precision untersucht. Die Tests scheinen die Wiedererkennung von Texten und ihre Zuordnung zu anderen, ähnlichen Texten aus der Kollektion zu bewerten. Ähnlichkeit wird durch Themengleichheit in mindestens einem Abschnitt definiert. Reale Benutzer oder reale Anfragen fließen offensichtlich nicht ein. Es zeigte sich lediglich, dass die Qualität bei längeren N-Grammen besser wird, was auf die höhere Adäquatheit von Wörtern als Repräsentationsmechanismus hinzuweisen scheint.

Zavrel 1996 wendet eine Erweiterung des Kohonen-Netzes u.a. auf die Cranfield-Kollektion⁵ an. Nach einer Einordnung in die Forschungskontexte Visualisierung, Kohonen-SOM und neuronale Netze im IR präsentiert Zavrel 1996 seine Erweiterung des Kohonen-Algorithmus, bei der während des Trainings neue Neuronen in die Kohonen-Schicht eingefügt werden. Dazu misst jedes Kohonen-Neuron die Größe des von ihm abgedeckten Eingaberaums. Das Neuron, das den größten Bereich von potenziellen Eingabemustern vertritt, erhält einen Nachbarn, der diesen Raum mit ihm teilt und so eine bessere Differenzierung zulässt.

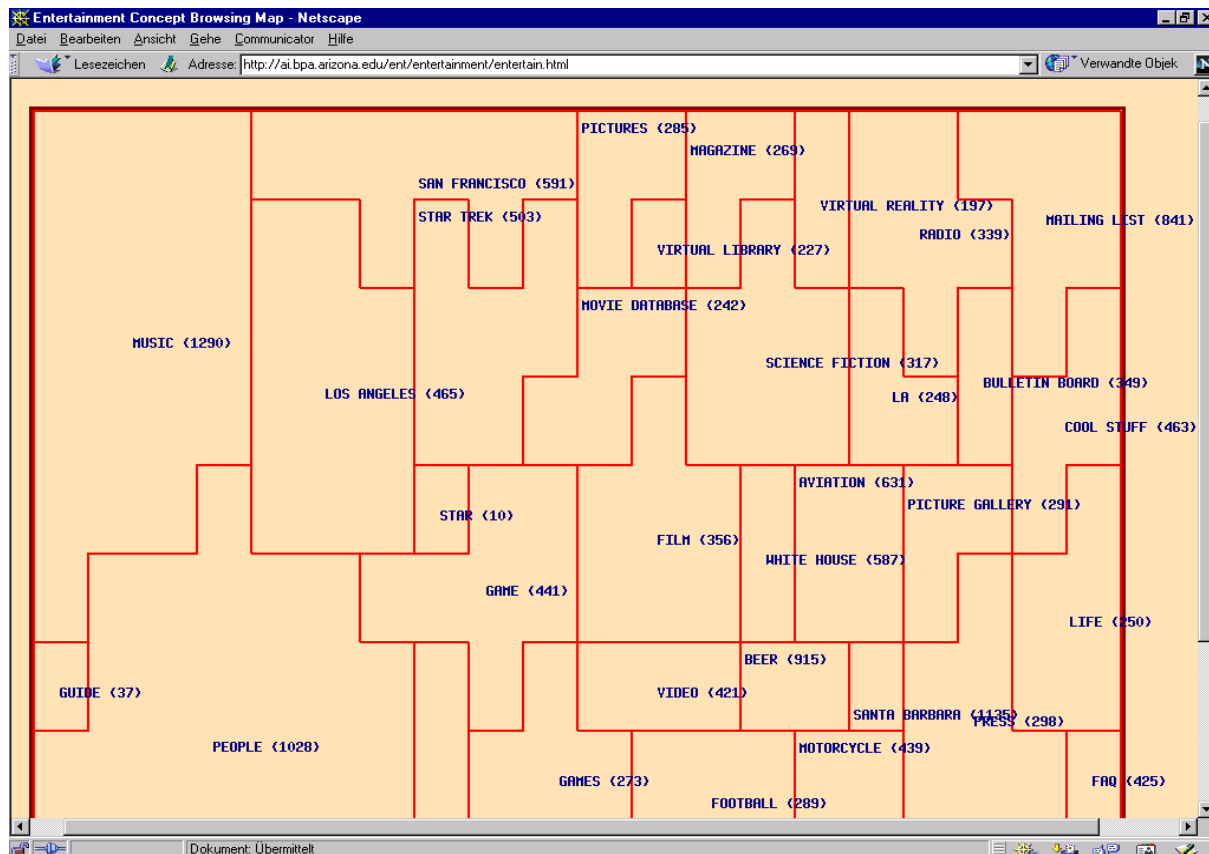


Abbildung 4: Ausschnitt aus der Visualisierung der Kohonen-Karte von Chen et al. 1996

Zavrel 1996 evaluiert die Qualität des Kohonen-Netzwerks als IR-System im Vergleich zu anderen *Clustering*-Methoden und dem Vektorraum-Modell. Dazu wird das *Clustering*-Verfahren als IR-Modell interpretiert und die Anfrage dient als Eingabe im Kohonen-Netz. Der *Cluster*, den das aktivierte Ausgabe-Neuron repräsentiert, ist die Ergebnismenge. Die Evaluierung erfolgt also auf der Basis einer booleschen Menge und kommt damit dem *Clustering*-Verfahren entgegen. Das Ergebnis des Vektorraum-Modells, eine gerankte Liste aller Dokumente (wird ebenfalls als boolesche Menge interpretiert). Das Vektorraum-Modell verwendet in diesem Fall das Kosinus-Ähnlichkeitsmaß. Die Anzahl der Dokumente bestimmt Zavrel 1996 aus der Anzahl der durchschnittlich in den *Clustern* enthaltenen Dokumenten. Die vom Vektorraum-Modell gelieferte Liste von Dokumenten wird bei einem Schwellenwert in zwei boolesche Mengen geteilt. Die *Clustering*-Verfahren geben damit das Retrieval-Verfahren vor und die im Vektorraum-Modell enthaltene Reihenfolge der Dokumente geht für die Bewertung verloren. Zudem wird die Größe der zu berücksichtigenden Menge aus dem Durchschnitt der *Clustering*-Verfahren berechnet. Trotzdem schneidet das Vektorraum-Modell in dem Vergleich von Zavrel 1996 besser ab als alle *Clustering*-Verfahren, unter denen das modifizierte Kohonen-Netz die beste Qualität erreicht. Dieses Evaluationsverfahren ignoriert jedoch die Aspekte der Visualisierung und der *Browsing*-Möglichkeiten für den Benutzer und erscheint somit für den Vergleich von zwei-dimensionalen Karten ungeeignet.

Chen et al. 1996 übertragen den Kohonen-Ansatz auf Internet-Dokumente (siehe Abbildung 4). Trotz der Beschränkung auf eine Kategorie des Internet-Katalogs Yahoo (*entertainment*) führen sie eine Vielzahl semantischer Karten ein. Sobald ein *Cluster* mehr als eine bestimmte Menge von Seiten enthält, wird er auf einer eigenen Karte dargestellt. Die auf Basis einer Trainingsmenge von 10.000 Seiten entstehenden Karten wurden in Benutzertests empirisch getestet. Dabei zeigte sich, dass viele Versuchspersonen im Umgang mit der Karte ein

⁵ <ftp://ftp.cs.cornell.edu/pub/smart/cran/>

Gefühl der Desorientierung befiel, das auch in assoziativ verknüpften Hypertexten vorkommt und als „lost in hyperspace“-Syndrom bekannt ist (cf. Kuhlen 1991). Die Orientierung fiel oft schwer und die Benutzer fragten nach geordneten Listen der enthaltenen Suchwörter. Die Karte eignet sich nach Einschätzung der Autoren vor allem für assoziative Suchen, bei denen Benutzer einen Überblick über die vorhandenen Suchbegriffe wünschen.

Eine kleinere Implementierung dieses Systems organisiert Äußerungen aus elektronischen Online-Besprechungen (cf. Orwig et al. 1997). Die Testmenge umfasst 202 Dokumente mit 190 Termen und damit Eingabedimensionen. Das Problem der Überladung einer Karte tritt in diesem Anwendungsfall kaum auf, da jede Besprechung nur eine begrenzte Anzahl von Redebeiträgen umfasst.

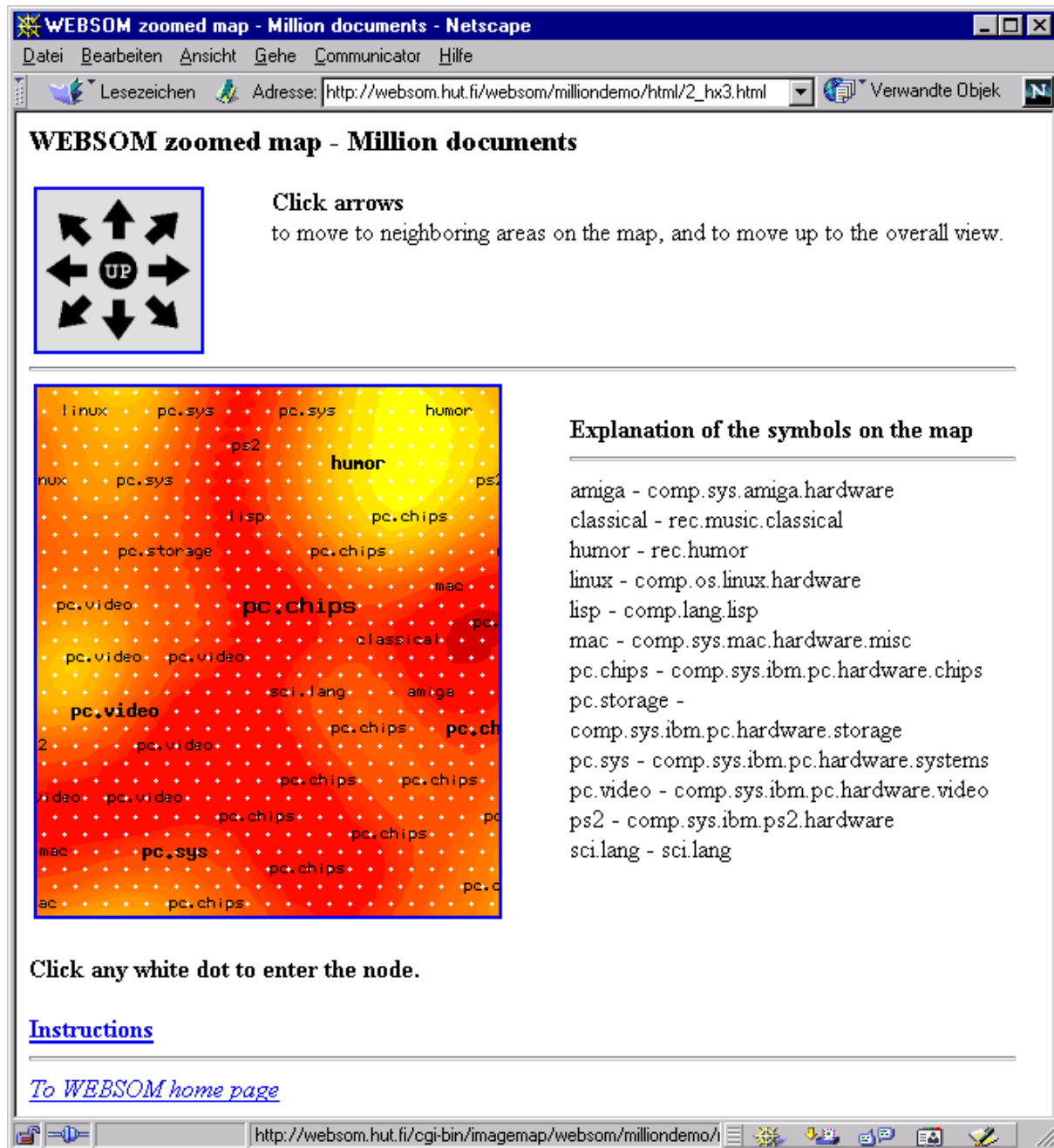


Abbildung 5: Benutzungsoberfläche der WEBSOM-Karte mit News-Artikeln

Eine weitere Implementierung einer Kohonen-Karte, die als Interface im Internet zur Verfügung steht, ist WEBSOM (cf. Kohonen 1998). WEBSOM stellt Karten mit großen Mengen von Newsgroup-Artikeln zur Verfügung. Aufgrund der großen Anzahl von Dokumenten ordnen die Autoren wie im Ansatz von Chen et al. 1996 die Karten in mehreren Schichten an. Die *Cluster* in der ersten Karte führen also zu weiteren Karten und erst beim dritten oder vierten Schritt gelangt der Benutzer zu einem *Cluster* von Newsgroup-Artikeln.

Abbildung 5 zeigt eine Karte vor diesem letzten Schritt. Die Interaktion beschränkt sich bei mehrschichtigen Systemen nicht auf das Klicken in einer Karte. Zusätzlich ist in der Seite ein Kontrollelement integriert, um zu einer anderen Karte zu wechseln. Die Beschriftung der *Cluster* ist stark verkürzt, was die Interaktion erschwert. Der Benutzer muss die Langform der Einträge rechts außerhalb der Visualisierung lesen. Dies stört die angestrebte Direktheit der Interaktion.

Schatz 1998 überträgt die Input-Muster mit einer adaptierten SOM in einem dreidimensionalen Raum, in dem der Benutzer navigiert (*space flight*). Abbildung 6 zeigt die dreidimensionale Darstellung, in der jedoch die dritte Dimension nicht homogen besetzt ist. Nur die am Rand der zweidimensionalen Oberfläche liegenden Bereiche verlaufen nennenswert in der dritten Dimension. Dadurch entsteht eine Art Terrain, das sicher leichter zu verstehen ist, als ein homogen mit Dokumenten besetzter dreidimensionaler Raum. Allerdings ist unklar inwieweit, ein Vorteil gegenüber den obigen zweidimensionalen Darstellungen besteht.

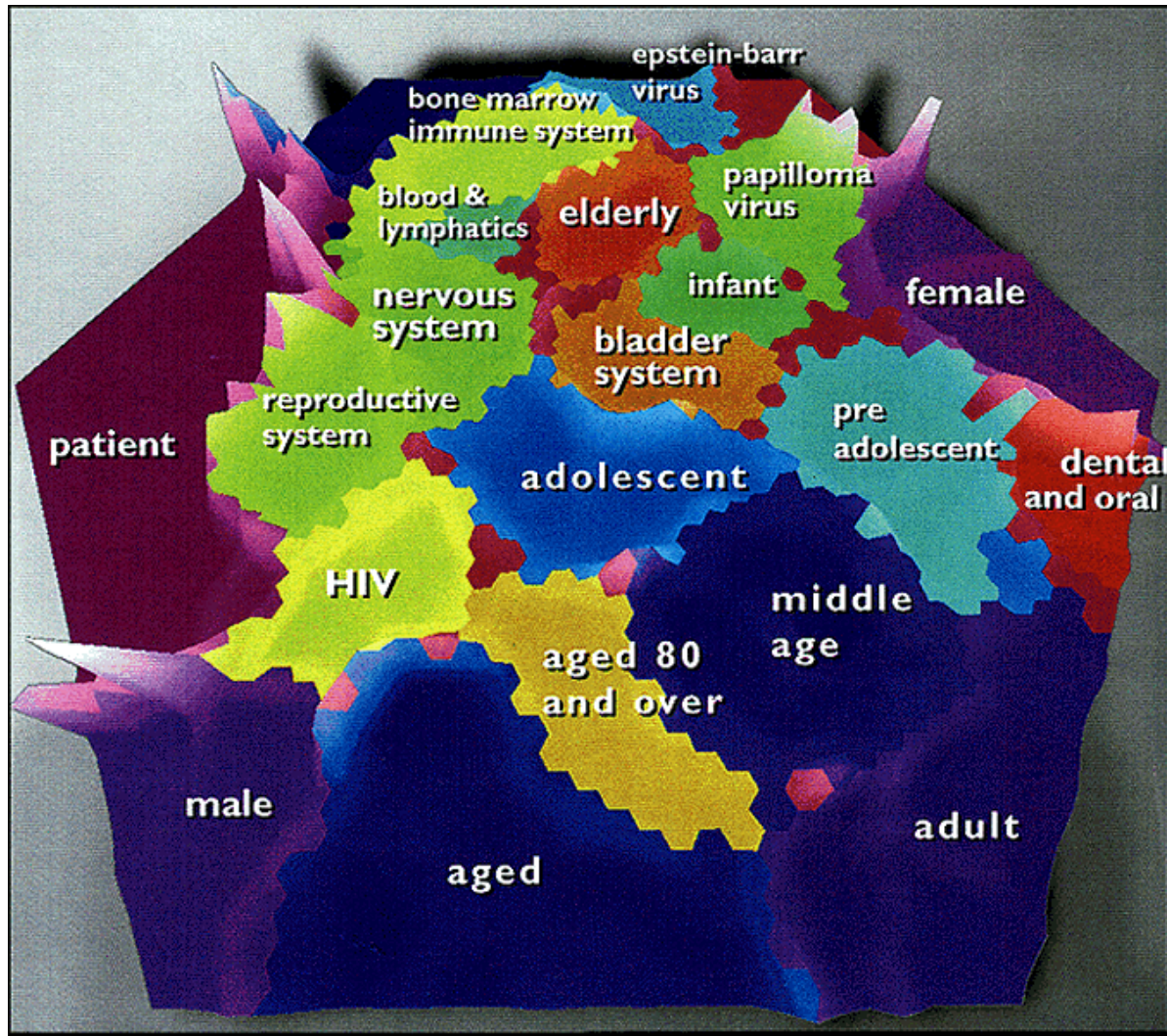


Abbildung 6: Visualisierung einer SOM als dreidimensionales Terrain (aus Schatz 1998)

Hui/Goh 1996 implementieren ein vollständiges IR-System, dessen Kern ein Kohonen-Netz bildet. Ihr System lernt aus Relevanz-Feedback, indem es seine Gewichte verändert, hat jedoch einige Schwächen. In der Trainingsphase extrahiert es Terme, die dann in eine kodierte Repräsentation umgeformt werden. Das Trainingsverfahren präsentiert sie dem Netz einzeln und der Kohonen-Algorithmus ordnet sie in *Cluster*. Beim Retrieval werden ebenfalls die Terme der Anfrage einzeln an das Netz gelegt und einem *Cluster* zugeordnet. Die Terme aller *Cluster*, die die Anfrage-Terme aktivieren, bilden dann die eigentliche Anfrage. Das Kohonen-Netz dient also der Termexpansion. Alle Dokumente, in denen die Terme in den aktivierten *Clustern* vorkommen, bilden die Ergebnismenge. Wie bei anderen Kohonen-Modellen entsteht dadurch ein unflexibles Retrieval-System.

Der hohe Zeitbedarf der Kohonen *Self-Organizing Map* gilt als einer ihrer größten Nachteile. Dem begegnen z.B. Kaski 1998 und Kohonen 1998 mit einer Komprimierung der Eingangsdaten. Merkl 1995 komprimiert mit einem neuronalen Backpropagation-Netzwerk. Diese Komprimierung reduziert bei Merkl 1995 die Trainingszeit des Netzes auf unter 20% der ursprünglichen Zeit. Jedoch erfolgt kein Vergleich der Qualität der komprimierten Repräsentation mit der ursprünglichen Repräsentation. Die Eignung für *software-reuse* in realen Szenarien müsste genauer untersucht werden.

Dieser Überblick über vorhandene Information Retrieval Systeme auf der Basis von Kohonen-Netzen zeigt vor allem, dass in realen Umgebungen der Umfang der zu verarbeitenden Daten für eine zweidimensionale Karte zu groß ist. Die meisten Karten erlauben weniger als 1000 Eingangsdimensionen und Muster. Chen et al. 1996 und WEBSOM (cf. Kohonen 1998) testen mit realen Daten und lösen das Problem durch die Einführung mehrerer Schichten von Kohonen-Netzen, was weitere Interaktionsmechanismen zum Wechseln von einer Karte zur anderen erfordert. Das ursprüngliche Grundprinzip einer einfachen und scheinbar natürlichen Visualisierung wird dadurch überlagert. Eine weitere Schwäche besteht darin, dass das Kohonen-Netz nicht den Kern eines IR-Systems implementiert.

Die meisten Systeme bieten ein vollständiges *Cluster* von Dokumenten als Ergebnismenge. Die statische Zuordnung der Dokumente in die *Cluster* bildet damit die Grundlage für das eigentliche Retrieval. In diesem Modell ist der dynamische Aspekt von Retrieval-Prozessen und die Integration von Relevanz-Feedback schwierig zu modellieren. Selbst wenn ein Benutzer einige Dokumente des *Clusters* als relevant und andere als nicht relevant einordnet, findet eine automatisch neu formulierte Anfrage in der Regel nur wieder das gleiche *Cluster*. Gelingt es einem Verfahren, dann ein zweites *Cluster* zu treffen, sind die relevanten Dokumente aus dem ersten gefundenen *Cluster* nicht mehr in der Treffermenge.

Grundsätzlich erscheint die Kohonen-Karte als Repräsentation einer Dokumenten-Kollektion problematisch. Um eine einfache und benutzerfreundliche Visualisierung zu erhalten, werden die zahlreichen Dimensionen des Term-Raums auf zwei reduziert. Dass eine solche Reduktion die Struktur der Kollektion noch adäquat widerspiegelt, ist eher unwahrscheinlich. Trotz der visuellen Adäquatheit ist die kognitive Repräsentation des Term-Raums wohl komplexer als eine zweidimensionale Karte. Die *Cluster*bildung ist stark kontextabhängig. Die zweidimensionale Darstellung erlaubt in einer größeren Kollektion insgesamt nur eine begrenzte Anzahl von Assoziationen in Form räumlicher Nachbarschaft. Zwar bietet die Kohonen-Karte für kleinere Mengen von Dokumenten einen möglichen Einstieg für assoziative und explorative Informationsbedürfnisse, als einziger Zugang ist sie in jedem Fall ungeeignet.

Einen interessanten Anwendungsfall aus Lin et al. 1991 greift Lin 1995 in seinen Tests mit verschiedenen kartenorientierten Darstellungen auf. Die Tests von Lin 1995 basieren auf einer sehr kleinen Menge von Daten. Die Aufgabe für die Benutzer bestand darin, aus 133 dargestellten Dokumenten zehn zufällig ausgewählte zu finden, was keine reale Retrievalsituation widerspiegelt. Die Grundidee der Beschränkung auf kleine Mengen ist jedoch verfolgenswert. Dient die Karte nur der Ergebnisanzeige, ist das Problem der großen Datenmengen gelöst. Durch die Beschränkung auf eine kleine Aufgabe innerhalb des Retrievalprozesses könnten sich die Vorteile der Visualisierung in einem ausgewogenen Gesamtsystem besser entfalten.

5 Evaluierung von Visualisierungen

Zur Bewertung von Mensch-Maschine-Interaktion werden objektive und subjektive Verfahren eingesetzt (cf. Eberleh et al. 1994). Letztlich liefern nur objektive Tests mit realen Benutzern verwertbare Daten für die Analyse, Bewertung und Verbesserung der Mensch-Maschine-Interaktion.

In bezug auf Anwendertests sind die Vertreter von Visualisierung erstaunlich zurückhaltend. In der Regel wird die Vorteilhaftigkeit von Visualisierung gegenüber textuellen Retrievalsystemen als allgemeingültig vorausgesetzt. Eine nähere Hinterfragung durch Anwendertests scheint als nicht notwendig betrachtet zu werden. Die *Conference Notes* des „2. Workshop on Visual Information Retrieval Interfaces“ (VIRIs) während der SIGIR'96 stellen fest: „There are now a large number of these experimental visualization systems which have been constructed by both private industry and government. Nevertheless, it was surprising that participants of this workshop believed that the time has arrived for evaluation studies.“ (Rorvig & Hemmje 1999:835)

Seit 1989 läuft in den USA das inzwischen wichtigste Projekt zur Bewertung der Qualität von Information Retrieval Systemen. Die vom National Institute of Standards and Technology (NIST), Gaithersburg, Maryland jährlich durchgeführte *Text Retrieval Conference* (TREC⁶) stellt eine Infrastruktur für die Evaluierung von

⁶ <http://trec.nist.org>

Systemen mit Massendaten zur Verfügung. Da Forscher für ihre IR-Tests verschiedenste Textkollektionen benutzten, waren die Ergebnisse bisher oft nicht vergleichbar. TREC hat diese Situation verbessert. Die TREC-Konferenzen bieten eine einheitliche Testumgebung und sind als gemeinsame Plattform zum Leistungsvergleich organisiert. Die Initiative findet großen Anklang; so beteiligen sich an TREC 6 bereits 58 Retrieval-Systeme (cf. Voorhees & Harman 1998). TREC bietet den Teilnehmern eine große Dokument-Kollektion, Anfragen und übernimmt für eingereichte Ergebnisse die Auswertung. Als Trainingsdaten stehen die Kollektionen der Vorjahre zur Verfügung, die intellektuelle Relevanzurteile enthalten. Die Ergebnisse der Systeme werden untereinander verglichen.

Während der letzten Jahre führte TREC eine Kategorie ein, in der die Interaktion zwischen System und Benutzer bewertet wird. Selbst in diesem *interactive track* spielen Visualisierungen noch keine herausragende Rolle.

Auch in einer anderen Evaluierungsinitiative spielen Visualisierungen eine Rolle, dabei kommen u.a. Kohonen SOM zum Einsatz. Das *Information Exploration Shootout Project* (cf. Grinstein et al. 1997) stellt eine große Menge von realen Log-Daten eines WebServers bereit, an dem Teilnehmer versuchen, Anomalitäten zu finden und so sicherheitsrelevante Angriffe zu entdecken. Dafür eignen sich *Data Mining*-Techniken wie nicht zuletzt die Visualisierung durch SOM (Girardin 1999). Die Initiative hat allerdings den Umfang von TREC noch nicht erreicht und liefert keine Anhaltspunkte für die Frage, ob zweidimensionale Visualisierungen vorteilhaft sind und wenn ja, welche.

In der Literatur finden sich drei Visualisierungen von Information Retrieval Systemen, die evaluiert wurden: SENTINAL, J24 und DEViD. Die ersten beiden nahmen 1998 am TREC-7 teil, DEViD wurde anhand der GIRT (*German Indexing and Retrieval Testdatabase*) getestet. Die Kollektion von GIRT besteht aus Fachtexten zu den Sozialwissenschaften und bildet inzwischen einen Teil der Testkollektion des *Cross Language Evaluation Forum* (CLEF⁷), das im europäischen Rahmen eine Evaluierungsumgebung für multilinguales Information Retrieval zur Verfügung stellt.

Abbildung 7 zeigt die Recall- und Precisionwerte von SENTINAL und anderen Teilnehmern des Ad-Hoc-Tests in Abhängigkeit voneinander. In der Abbildung sind die Ergebnisse sämtlicher Systeme verzeichnet, die am TREC-7 Ad Hoc (manuell) teilnahmen. Die Ergebnisse von SENTINEL (Harris Information Systems) sind rot eingezeichnet:

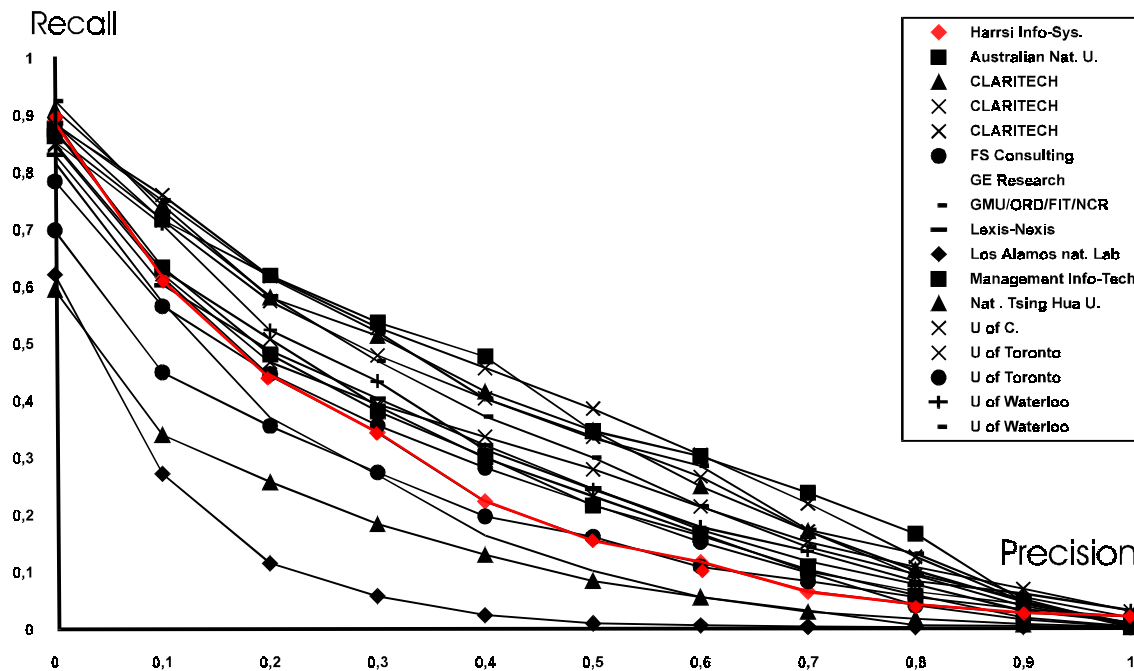


Abbildung 7: SENTINEL: Recall / Precision für Ad-Hoc-Tests (manual) TREC-7 (cf. TREC-7 Appendix A)

⁷ <http://www.clef-campaign.org>

SENTINEL war beim Ad-Hoc-Test von Trec-7 das einzige System, welches Visualisierung nutzte. Die Ergebnisse im Vergleich zu anderen Systemen scheinen wenig überzeugend. Es lässt sich jedoch nicht eindeutig feststellen, inwieweit die Visualisierung für die Güte des Ergebnisses verantwortlich ist, da SENTINEL auch über ein textuelles Interface recherchierbar ist. Zwar schnitt SENTINEL bei Trec-6, als noch keine Visualisierung implementiert war, deutlich schlechter ab (cf. Knepper et al. 1998: Fig. 1&4), jedoch wurden in der Zwischenzeit ebenfalls umfassende Verbesserungen an der Retrievaltechnik vorgenommen, so dass die Visualisierung allein nicht verantwortlich sein kann.

Etwas klarer können dagegen die Ergebnisse des Retrievaltests von J24 ausgelegt werden. Abbildung 8.a zeigt die durchschnittlichen Recallwerte, welche die einzelnen Teilnehmer mit J24 und dem Vergleichssystem ZPRISE⁸ erreichten. Hier entsteht der Eindruck, dass einige Teilnehmer mit J24 deutlich besser zurecht kamen, während andere Teilnehmer offensichtlich mit ZPRISE bessere Ergebnisse erzielten. Dieser Eindruck täuscht: Tatsächlich bearbeiteten die Teilnehmer, die mit J24 bessere Ergebnisse erzielten mit ZPRISE die Themen 352-355 und mit J24 die Themen 356-359. Entsprechend umgekehrt vielen die Ergebnisse aus, wenn die Teilnehmer mit J24 den ersten Satz Themen bearbeiteten und mit ZPRISE den zweiten.

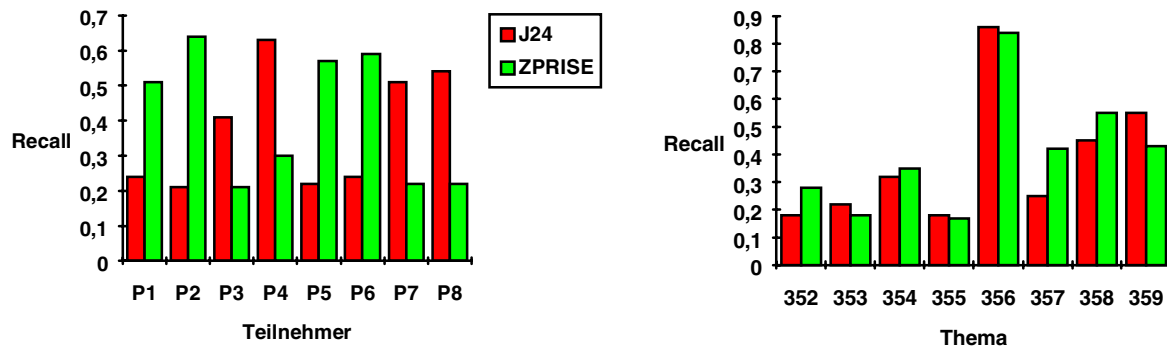


Abbildung 8.a: Recall pro Teilnehmer (Odgen et al. 1998: Fig.2) Abbildung 8.b: Recall pro Thema (Odgen et al. 1998: Fig.3)

Abbildung 8.b verdeutlicht den Zusammenhang zwischen der Güte der Retrievalergebnisse und Thema. Tatsächlich wurde der zweite Satz Themen von beiden Systemen besser bearbeitet als der erste Satz. Vergleicht man die Systeme weiter miteinander, so stellt man jedoch fest, dass auch hier kein System als das generell geeignetere hervortritt.

Ein weiteres System, das sich einer Evaluation unterzog ist DEViD (Eibl 2000). Anhand der Standardisierten Testdatenbank GIRT, die mittlerweile in die TREC-Evaluationen integriert wurde, konnten objektive Vergleiche zu den ohne Visualisierungskomponente arbeitenden Systemen Messenger und freeWAIS gezogen werden. Dabei schnitt DEViD deutlich besser ab, wie Abbildung 9 verdeutlicht:

⁸ ZPRISE ist ein am NIST eingesetztes Information Retrieval System (<http://www-nlpir.nist.gov/works/papers/zp2/zp2.html>).

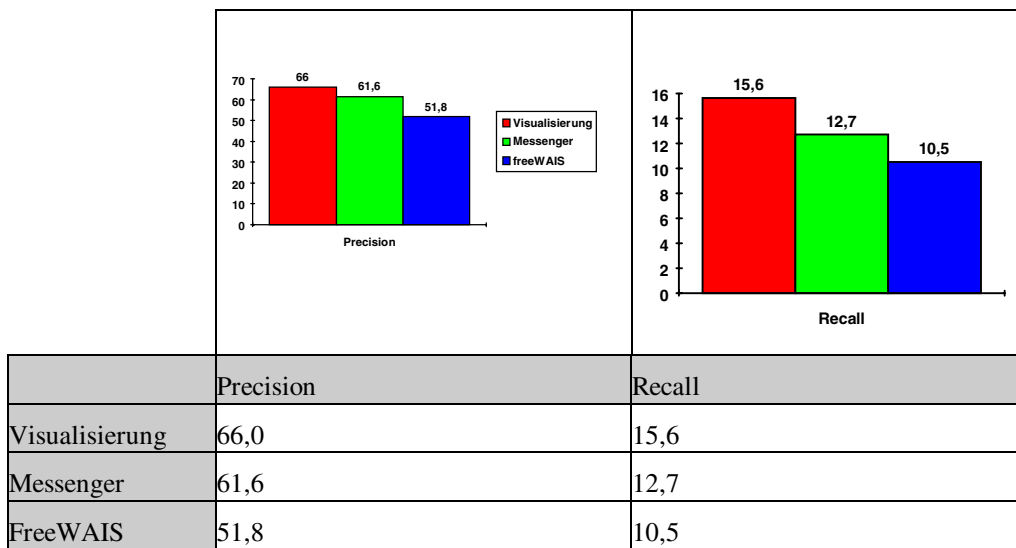


Abbildung 9: Testergebnisse von DEViD in Prozent

Von den hier vorgestellten Visualisierungen im Dokument Retrieval haben sich nur drei öffentlichen Anwendertests unterzogen. Inwieweit Tests an den übrigen Systemen vorgenommen wurden, lässt sich nicht definitiv sagen. Angesichts der Tatsache jedoch, dass die hier besprochenen Tests wenig überzeugen und dass im Gebiet der Informationsvisualisierung, welche Anfang der 90er Jahre so euphorisch angegangen wurde, inzwischen viele Projekte, u.a. bei Netscape, Apple, XEROX, MIT oder Microsoft, eingestellt wurden, lässt sich schließen, dass die Prämisse der unbedingten Vorteilhaftigkeit einer Visualisierung gegenüber einer textuellen Darstellung nicht standhält. Zu einem ähnlichen Ergebnis kommen auch informelle Nutzerumfragen, wie sie etwa auch Krohn 1996: 105 berichtet:

„Discussions with people who were faced with Vineta’s galaxy confirm indeed that it is difficult to interpret the spatial layout of the galaxy interface. In particular, they were not able to perceive depth in the galaxy. Moreover people had several other difficulties in interpreting the galaxy display:

- The color of the spheres did not intuitively indicate relevance.
- The direction of the shooting stars did not clearly show in which documents the terms of the shooting stars occur.
- Because the proposed terms are spread all over the screen, evaluation and comparison of the terms are made more difficult than if the terms were grouped together in one part of the screen. (...)
- Not only the relation between spikes and the corresponding selected term was unclear but also that a relation exists was unclear.
- The abrupt change in the display after selecting a new term could be understood only with difficulty.“

Für Mischsysteme wie Cartia (siehe Abbildung 1), die das Konzept zweidimensionaler Karten mit traditionellen Schlagwortsuchen verknüpfen fehlen Evaluierungen völlig.

6 Experiment und Ergebnisse

Für die Evaluierung von zweidimensionalen Karten gibt es keine Standard-Verfahren wie für die Bewertung von Information Retrieval Systemen. Diese statistischen Methoden, die z.B. im Rahmen der *Text Retrieval Conference* (TREC) eingesetzt werden, basieren auf Ranking-Listen und lassen sich so nicht auf Ansätze auf der Grundlage von *Browsing*-Techniken anwenden. Dieser Abschnitt stellt eine Methode zum Vergleich zweier

Karten vor, die den Benutzerstandpunkt beim *Browsing* in den Mittelpunkt stellt, aber ohne aufwendige Benutzertests auskommt.

6.1 Evaluierungsmethode

Eine Karte von Dokumenten hilft dem Benutzer dann, wenn semantisch ähnliche Dokumente nahe zusammen gruppiert sind. Nach der Fokussierung eines Dokuments wird der Benutzer die am nächsten gelegenen betrachten. Die Evaluierungsmethode nimmt daher ein Dokument als Ausgangspunkt und bestimmt die Euklidische Distanz zu allen anderen. Durch Berücksichtigung aller Dokumente entsteht so eine Ähnlichkeitsmatrix. Das Verfahren wird für jede Visualisierungsmethode wiederholt und es ergeben sich so mehrere Matrizen. Die Korrelation zwischen diesen Matrizen dient als Maß für den Vergleich der Methoden. Für die Berechnung der Korrelation wird die Matrix in ihre Spalten aufgeteilt und jede Spalte wird nach Ähnlichkeit sortiert. Der Spearman-Rangfolgen-Koeffizient bestimmt die Korrelation zwischen den Ranking-Listen der verschiedenen Methoden. Der Durchschnitt der einzelnen Listen bildet das Resultat für die Matrizen.

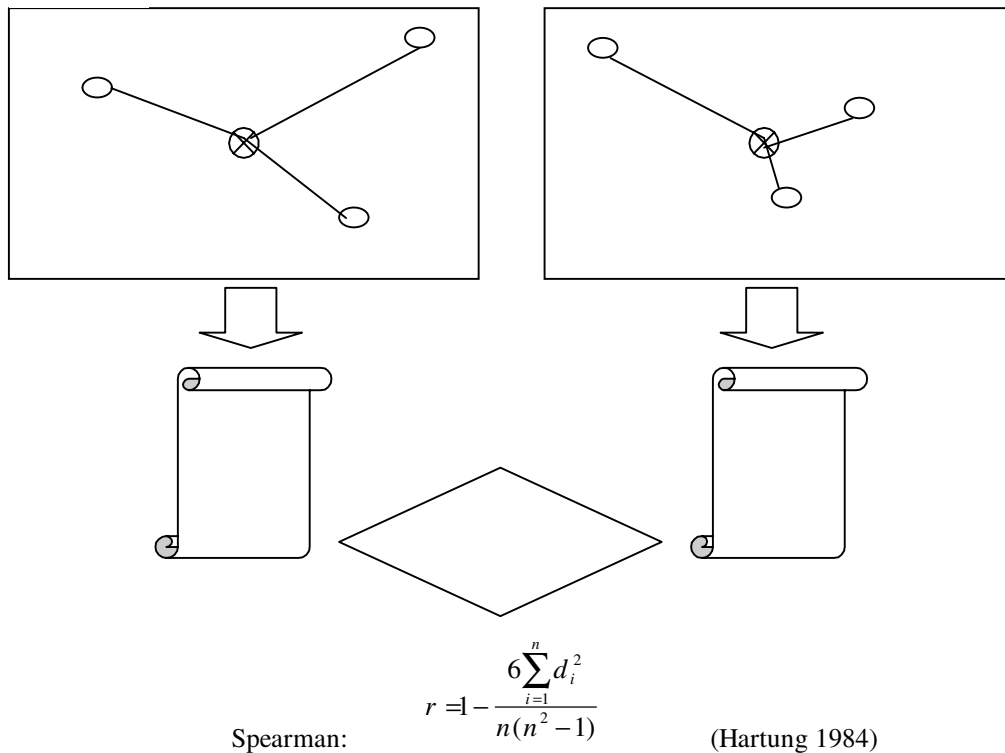


Abbildung 10: Der Vergleich für ein Ausgangsdokument

Der in Abbildung 10 skizzierte Evaluierungsalgorithmus umfasst die folgenden in Pseudo-Code formulierten Schritte:

Für alle Dokumente d

 Für alle Dokumente j

 Berechne Ähnlichkeit von d zu j mit Verfahren 1

 Berechne Ähnlichkeit von d zu j mit Verfahren 2

 Bilde Rangliste für alle Dokumente nach ihrer Ähnlichkeit zu d nach Verfahren 1

Bilde Rangliste für alle Dokumente nach ihrer Ähnlichkeit zu d nach Verfahren 2

Vergleiche die zwei Ranglisten mit dem Spaerman-Rangfolgen-Koeffizienten

Bilde den Durchschnitt aller Rangfolge-Koeffizienten

Zwar erfasst dieses Verfahren nur die Entfernungen von einem Dokument ausgehend und vernachlässigt so die potentielle „*Browsing*-Ketten“. Durch die Iteration über alle Dokumente wird dies aber ausgeglichen. Ein ähnliches Verfahren wurde bereits in Mandl 2001 für den Vergleich von Rangfolgen im Bereich Information Retrieval benutzt. Es scheint sich als Ergänzung zu traditionellen Bewertungsverfahren im Information Retrieval wie *recall-precision*-Graphen und die Berechnung der durchschnittlichen *precision* zu eignen.

6.2 Daten und Ergebnisse

Die Daten für das Experiment stammen aus den Datenbanken des Informationszentrums Sozialwissenschaften (IZ) in Bonn⁹. Indexierer des IZ verschlagworteten intellektuell sozialwissenschaftliche Fachliteratur und Projektinformationen.

Mit einem Ausschnitt von 1000 Dokumenten, der auch in CLEF zur Verfügung steht, wurden mit LSI und SOM zwei zweidimensionale Dokumentenkarten gebildet. Der Vergleich nach der oben geschilderten Methode führte zu einer Korrelation von 0,053. Dies zeigt, dass zwischen den Matrizen keine Ähnlichkeit besteht. Das Experiment wurde auch für kleinere Mengen von zehn und hundert Dokumenten durchgeführt. Auch in diesen Fällen ergab sich keine Korrelation wie die folgende Tabelle zeigt.

Tabelle 1: Ergebnisse

Anzahl der Dokumente	1000	100 (Durchschnitt aus zehn Mengen)	10 (Durchschnitt aus 100 Mengen)	Höchste Korrelation für eine Menge von 10 Dokumenten
Korrelation	0,053	0,037	0,013	0,34

7 Fazit

In der Literatur wird oftmals davon ausgegangen, dass Visualisierungen textuellen Formen der Datenpräsentation überlegen sind. Die kognitionswissenschaftlich durchaus begründbare Vermutung führt bei der Gestaltung von Visualisierungen oftmals scheinbar zu einer nicht weiter hinterfragten Annahme, dass die geschaffene Visualisierung von Vorteil sei. Auf die Durchführung von Evaluationen wird oft verzichtet. Dabei zeigt sich übrigens immer wieder, dass eine Überlegenheit von 2- oder 3-D-Visualisierungen gegenüber textbasierten Anzeigen nicht nachgewiesen werden kann (cf. z.B. Sebrechts et al. 1999).

In diesem Artikel wird gezeigt, wie wichtig jedoch solche Evaluationen sind, kann doch schon die Verwendung unterschiedlicher Konstruktionsprinzipien ein und desselben Visualisierungstyps zu erheblich unterschiedlichen Ergebnissen führen. Solange die Ergebnisse verschiedener Techniken zu solch divergenten Ergebnissen führen, sind zusätzliche Anwendertests notwendig, um die geeignetere Technik zu bestimmen. Zusätzlich hängt die Qualität von Information Retrieval Verfahren stark vom Korpus ab, d.h. was sich für eine Menge von Dokumenten, Benutzern und Anfragen eignet, muss bei anderen Kollektionen nicht unbedingt zu guten Ergebnissen führen (cf. z.B. Krause 1996; Mandl 2001; Womser-Hacker 1997). Auch für die Qualität Visualisierungen ist eine starke Abhängigkeit vom jeweiligen Korpus zu erwarten.

Im Bereich der Visualisierung zeigt sich, dass das Fehlen einer etablierten Methode zur Evaluierung zu einem Mangel an experimentellen Arbeiten führt. Der Erfolg von Initiativen zur Bewertung von Information Retrieval Verfahren beruht nicht zuletzt auf einfachen Kennzahlen zur Interpretation der Ergebnisse. Solche Kennzahlen wurden dagegen z.B. für maschinelle Übersetzung (King 1997) und eben für Visualisierung noch nicht gefunden. Die Entwicklung von geeigneten Evaluierungsmethoden, die auf den Benutzer fokussieren sind entscheidend.

⁹ <http://www.bonn.iz-soz.de>

Literatur

- Berry, Michael; Do, Theresa; O'Brien, Gavin; Krishna, Vijay; Varadhan, Sowmini (1993). SVDPACKC (Version 1.0) User's Guide. Arbeitspapier. Computer Science Department. University of Tennessee in Knoxville, USA. <http://www.netlib.org/svdpack/svdpackc.tgz>
- Berry, Michael; Dumais, Susan; Letsche, Todd (1995). Computational Methods for Intelligent Information Access. In: Proceedings of ACM Supercomputing '95. San Diego, CA. S. 1-38.
- Card, Stuart; Mackinley, Jock; Shneiderman, Ben (Hrsg.) (1999). Readings in Information Visualization. Using Vision to Think. San Francisco.
- Chen, Hsinchun; Schuffels, Chris; Orwig, Richard (1996). Internet Categorization and Search: A Self-Organizing Approach. In: Journal of Visual Communication and Image Representation. 7(1). S. 88-101.
- Däßler, Rolf; Palm, Hartmut (1998). Virtuelle Informationsräume mit VRML. Information recherchieren und präsentieren in 3D. Heidelberg.
- Deerwester, Scott; Dumais, Susan T.; Harshman, Richard (1990). Indexing by Latent Semantic Analysis. In: Journal of The American Society For Information Science (JASIS) 41(6). S. 391-407.
- DeFanti, Thomas; Brown, Maxine; McCormick, Bruce (1989). Visualization. Expanding Scientific and Engineering Research Opportunities. In: IEEE Computer 22-8, Aug. 1989, S.12-25.
- Dubin, David (1995). Document Analysis for Visualization. In: Proceedings of the SIGIR'95, S.199-204.
- Eberleh, Edmund; Oberquelle, Horst; Oppermann, Reinhard (1994) (Hrsg.): Einführung in die Softwareergonomie. Gestaltung graphisch-interaktiver Systeme : Prinzipien, Wahrnehmungen, Lösungen.
- Eibl, Maximilian (2000). Visualisierung im Document Retrieval - Theoretische und praktische Zusammenführung von Softwareergonomie und Graphik Design. Informationszentrum Sozialwissenschaften Forschungsberichte 3, 241 Seiten, 122 Abbildungen, CD-ROM, Bonn.
- Fox, Kevin; Frieder, Phir; Knepper, Margret; Snowberg, Eric (1999) SENTINEL: A Multiple Engine Information Retrieval and Visualization System. In: Journal of the American Society for Information Systems (JASIS) Vol.50, Nr. 7, S.615-625.
- Girardin, Luc (1999). An eye on network intruder administrator shootouts. In: Proceedings Workshop on Intrusion Detection and Network Monitoring. Santa Clara, California, April 9-12, 1999. http://www.usenix.org/publications/library/proceedings/detection99/full_papers/girardin/girardin.pdf
- Gloor, Peter (1997). Elements of Hypermedia Design: Techniques for Navigation & Visualization in Cyberspace. Birkhäuser: Boston.
- Graupe, D.; Kordylewski, H. (1998). A Large Memory Storage and Retrieval Neural Network for Adaptive Retrieval and Diagnosis. In: International Journal of Software Engineering and Knowledge Engineering 8(1). S. 115-138.
- Grinstein, Georges; Laskowski, Sharon; Wills, Graham; Rogowitz, Bernice (1997). Information exploration shootout project and benchmark data sets (panel). evaluating how visualization does in analyzing real-world data analysis problems. In: Proceedings of the conference on Visualization '97. Oct. 18 - 24, 1997, Phoenix, AZ USA. S. 511-513
- Hartung, Joachim (1984). Lehr- und Handbuch der angewandten Statistik. München, Wien.
- Haykin, Simon (1999). Neural Networks: A Comprehensive Foundation. London et al.
- Hemmje, Matthias; Kunkel, Clemens; Willett, Alexander (1994). LyberWorld - A Visualization User Interface Supporting Fulltext Retrieval. In: Proceedings SIGIR '94 Dublin, 3.-6.7.1994. S.249-259.
- Hemmje, Matthias (1993). LyberWorld - Eine 3D-basierte Benutzerschnittstelle für die Computerunterstützte Informationssuche in Dokumentmengen. In: GMD-Spiegel. Hrsg.: Gesellschaft für Mathematik und Datenverarbeitung mbH (GMD). Nr.1, März 1993, S.56-63.
- Hersh, W.; Over, P. (2000). TREC-8 Interactive Track Report. In: Voorhees, Ellen; Harman, Donna (Hrsg). The Eighth Text REtrieval Conference (TREC-8). National Institute of Standards and Technology. Gaithersburg, Maryland. S.57. <http://trec.nist.gov/pubs/trec8/papers/t8irep.pdf>

- Hui, Siu; Goh, Angela (1996). Incorporating Fuzzy Logic with Neural Networks for Document Retrieval. In: Engineering Applications of Artificial Intelligence 9(5). S. 551-560.
- Kaski, Samuel (1998). Dimensionality Reduction by Random Mapping: Fast Similarity Computation for *Clustering*. In Proceedings of IJCNN'98, International Joint Conference on Neural Networks, Piscataway, NJ. vol. 1. S. 413-418.
- King, Margaret (1997). Evaluating Translation. In: Hauenschild, Christa; Heizmann, Susanne (Hrsg.). Machine Translation and Translation Theory. Berlin, New York: Mouton de Gruyter. S. 251-263.
- Kohonen, Teuvo (1984). Self-Organization and Associative Memory. Berlin et al.
- Kohonen, Teuvo (1998). Self-organization of Very Large Document Collections: State of the art. In Niklasson, L.; Bodén, M.; Ziemke, T. (Hrsg.). Proceedings of ICANN '98, 8th International Conference on Artificial Neural Networks, Springer: London. vol. 1, S. 65-74.
- Krause, Jürgen (1996). Principles of Content Analysis for Information Retrieval Systems: An Overview. In: Zuell, C.; Harkness, J.; Hoffmeyer-Zlotnik, J. (Hrsg.). Text Analysis and Computer. ZUMA-Nachrichten Spezial. Mai. Mannheim. S.77-100.
ftp://ftp.zuma-mannheim.de/pub/zuma/zuma-nachrichten_spezial/znspezial1.pdf
- Krohn, Uwe (1996). Visualization for Retrieval of Scientific and Technical Information. Dissertation, Technische Universität Clausthal.
- Kuhlen, Rainer (1991). Hypertext: ein nicht-lineares Medium zwischen Buch und Wissensbank. Berlin.
- Leung, Ying; Apperley, Mark (1993). E3: Towards the Metrication of Graphical Presentation Techniques for large Data Sets. In: Bass et al. 1993: 125-140.
- Lin, Xia (1995). Searching and *Browsing* on Map Displays. In: Kinney, Thomas (Hrsg.). Proceedings of the 58th Annual Meeting of the American Association for Information Science (Vol. 32) ASIS'95. Chicago, Oktober 1995. S. 13-18.
- Lin, Xia; Soergel, Dagobert; Marchionini, Gary (1991). A Self-Organizing Semantic Map for Information Retrieval. In: Bookstein, Abraham; Chiaramella, Yves; Salton, Gerard; Raghavan, Vijay V. (Hrsg.). Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '91). Chicago, IL, USA 13.-16.10.91. New York. S. 262-269.
- Mandl, Thomas (2001). Tolerantes Information Retrieval: Neuronale Netze zur Erhöhung der Adaptivität und Flexibilität bei der Informationssuche. Erscheint.
- McCormick, Bruce; DeFanti, Thomas, Brown, Maxine D. (Hrsg.) (1987). Visualization in Scientific Computing. In: Computer Graphics Vol.21, Nr.6, Nov. 1987.
- Merkel, Dieter (1995). Content-Based Document Classification with Highly Compressed Input Data. In: Proceedings of the International Conference on Artificial Neural Networks ICANN '95. Paris. Oktober 9-13 1995. Bd. 2, S. 239-244.
- Odgen, William; Davis, Mark; Rice, Sean (1998): Document Thumbnail Visualizations for Rapid Relevance Judgements: When do They Pay off? In: Voorhees & Harman 1998. S. 599-612.
- Olsen, Kai A.; Korfhage, Robert; Sochats, Kenneth; Spring, Michael; Williams, James (1993). Visualization of a Document Collection: The Vibe System. In: Information Processing & Management, Vol.29, Nr.1, S.69-81.
- Orwig, Richard; Chen, Hsinchun; Nunamaker, Jay (1997). A Graphical, Self-Organizing Approach to Classifying Electronic Meeting Output. In: Journal of the American Society for Information Science. JASIS 48(2). S. 157-170.
- Parsaye, Kamran; Chignell, Mark (1993). Intelligent Database Tools&Applications: hyperinformation access, data quality, visualization, automatic discovery. New York et al.
- Rodeghier, Mark (1997). Marktforschung mit SPSS: Analyse, Datenerhebung und Auswertung. Bonn et al.
- Roppel, Stephan (1996). Visualisierung und Adaption: Techniken zur Verbesserung der Interaktion mit hierarchisch strukturierter Information. Konstanz. [Schriften zur Informationswissenschaft Bd. 33]
- Schatz, Bruce (1998). High-Performance Distributed Digital Libraries: Building the Interspace on the Grid. In: 7th IEEE Int Symp High-Performance Distributed Computing (July). S. 224-234.
http://www.canis.uiuc.edu/archive/papers/hpdc.pdf

- Scherer, Andreas (1997). Neuronale Netze: Grundlagen und Anwendungen. Braunschweig, Wiesbaden.
- Scholtes, J.C. (1992). Neural Nets in Information Retrieval: A Case Study of the 1987 Pravda. In: Ruck, Dennis (Hrsg.). Science of Artificial Neural Networks: Proceedings of the International Society for Optical Engineering, 1710 (SPIE Conf., Orlando, Fl.) S. 631-641.
- Sebrechts, Marc; Vasilakis, Joana; Miller, Michael; Cugini, John; Laskowski, Sharon (1999). Visualization of Search Results: A Comparative Evaluation of text, 2D and 3D Interfaces. In: Hearst, Marti; Gey, Frederic; Tong, Richard (Hrsg.). Proceedings of the 22nd Annual International ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR '99). Berkeley, CA 15-19.8.99. New York.
- Spoerri, Anselm (1994a). InfoCrystal: Integrating Exact and Partial Matching Approaches through Visualization. In: RIAO'94 Proceedings, Oct.11-13, 1994 New York (NY), S.687-696.
- Spoerri, Anselm (1994b). InfoCrystal: A visual tool for information retrieval & management. Conference Companion, CHI'94, 24-28.4.1994, Boston (MA), S.11-12.
- Voorhees, Ellen; Harman, Donna (Hrsg.) (1998). The Seventh Text REtrieval Conference (TREC-7), NIST Special Publication 500-242. http://trec.nist.gov/pubs/trec7/t7_proceedings.html
- Womser-Hacker, Christa (1997). Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift. Universität Regensburg, Informationswissenschaft.
- Wolff, Christian (1996). Graphisches Faktenretrieval mit Liniendiagrammen: Gestaltung und Evaluierung eines experimentellen Rechercheverfahrens auf Grundlage kognitiver Theorien der Graphenwahrnehmung. [Schriften zur Informationswissenschaft 24] Konstanz.
- Zavrel, Jakob (1996). Neural Navigation Interfaces for Information Retrieval: Are They More than an Appealing Idea? In: Artificial Intelligence Review 10. S. 477-504.